

**UNIVERSIDAD DE CÓRDOBA**  
DEPARTAMENTO DE PRODUCCIÓN ANIMAL

---

**DESARROLLO Y EVALUACIÓN DE MODELOS  
PARA LA TOMA DE DECISIONES.  
CARACTERIZACIÓN DE LA PRODUCCIÓN DE  
ANGUILAS (*Anguilla anguilla* L.) EN  
SISTEMAS INTENSIVOS**

Tesis Doctoral presentada por Juan Carlos Gutiérrez Estrada, en satisfacción de los requisitos necesarios para optar al grado de Doctor en Ciencias Biológicas. Dirigida por el Dr. Emiliano de Pedro Sanz, Profesor Titular de la Universidad de Córdoba, y por el Dr. Rafael López Luque, Profesor Titular de la Universidad de Córdoba

Los directores,

El doctorando,

Emiliano de Pedro Sanz

Rafael López Luque

Juan Carlos Gutiérrez Estrada

Córdoba, 2003

*A Inma... evidentemente*

*¿Qué sabe el pez del agua donde nada toda su vida?*

*A. Einstein*

## Agradecimientos

Como suele ocurrir con casi cualquier aspecto de la vida, *‘al final todo se acaba’*. Tras esa meta y al echar la vista atrás, es inevitable recordar vivencias, situaciones y personas que, de una forma u otra, han contribuido a la finalización de esta memoria de Tesis. No me gustaría olvidarme de ninguno de los que me han asesorado y ayudado de muy diversas formas, entre otras cosas, porque a pesar de ser yo el autor de este trabajo, ellos forman parte íntegra de él.

En primer lugar mi reconocimiento a los directores de este trabajo, Emiliano de Pedro Sanz y Rafael López Luque, sin los cuales no se podría haber concebido, desarrollado y terminado. Sus meditaciones aportaciones, comentarios, correcciones y críticas constructivas, me han servido de aliciente y motor para centrar y afinar muchos aspectos, que de otra forma hubieran quedado oscuros o demasiado simplistas. Quiero destacar el entusiasmo de ambos al afrontar el desafío de la dirección de un trabajo enmarcado en una incipiente línea de investigación, con todo lo que ello supone. No dudaron en apoyarme en mi intento de expresarles la ilusión, aprecio e importancia que yo sentía por el tema de esta Tesis.

En la misma línea, es inevitable recordar a las personas que dirigían y gestionaban la empresa Hidrorecursos S.A en el año 1998, cuya grata actitud facilitó la concesión de la beca de investigación. Carlos Blanco (Director General) acogió con expectación mis ideas, retándome en multitud de ocasiones a llevarlas a la práctica. Con Ana Landa (Directora Técnica) tuve la oportunidad de aprender los aspectos básicos de la producción de anguilas en sistemas intensivos. A los dos mi agradecimiento por su generosa disposición a proporcionarme sin reservas un espacio donde trabajar y cuanto material necesité de la piscifactoría.

Como olvidarme del resto de trabajadores de Hidrorecursos. Rafael Membrives, los otros ‘Rafas’, José Antonio, Juan, Pepe y Julián. Siempre estuvieron dispuestos a ayudarme cuándo y cuánto les necesité. Nunca tuve un ‘no’ por respuesta a pesar de la ‘apretada’ jornada de trabajo. Buena parte del desarrollo de esta investigación dependía de entender el funcionamiento de las instalaciones, y sin ellos no hubiera sido posible. Hago extensivo también este sentimiento al conjunto de la familia Jurado: Antonio (padre), Carmen (madre), Antonio, Juan, Rogelio, Guadalupe, Diego y los ‘peques’.

La veterinaria de Hidrorecursos, Soledad González, se interesó por este trabajo durante su elaboración y contribuyó en el desarrollo del sistema experto. Tuve además, la magnífica oportunidad de aprender y discutir amablemente con ella sobre aspectos relacionados con la patología de los peces. En la fase de validación del sistema experto, tengo que destacar la colaboración de Noelia Molina, ‘una auténtica todo terreno’. No me puedo olvidar de los muchos becarios que durante mi estancia en la piscifactoría aportaron su granito de arena, especialmente de Marta.

Un agradecimiento especial y un reconocimiento personal se lo debo a Pieter de Visser Van Bloemen, cuya paciencia conmigo ha sido ilimitada. Pieter es una de las pocas personas que he conocido que poseen esa virtud que consiste en comprender las

cosas a la primera y hacer fácil lo difícil. No me equivoco al afirmar que es uno de los mejores especialistas en producción de anguilas en Europa y yo he tenido la suerte de trabajar con él.

Las gestiones para la obtención de los datos procedentes de la Central Térmica de Puente Nuevo han sido complejas y difíciles, al extremo de haberme llegado a aburrir... ¡Al final no he conseguido todos los datos que quería!. Debo reconocer que nunca he entendido cuál ha sido la razón última de este comportamiento, aunque..., ¡lo sospecho!. No sería justo, de todas maneras, si no reconociera que también he encontrado a técnicos dispuestos a ayudarme. Debo agradecer su comprensión a Angel Zafra, Juan Luís Pina y Sixto Rodríguez.

También deseo resaltar mi gratitud al Ministerio de Educación y Cultura, que me concedió una Beca de Formación de Personal Investigador para la realización de esta Tesis Doctoral, en el Departamento de Producción Animal de la Escuela Técnica Superior de Ingenieros Agrónomos y de Montes de la Universidad de Córdoba. A todos los profesores de este Departamento, y en particular a Ana Garrido y Pablo Lara, por su entusiasmo y amabilidad. Los amigos y compañeros que he tenido durante mi estancia en el Departamento merecen recuerdo por los buenos ratos y experiencias que hemos compartido: Juanito, Víctor y Nieves.

A la Universidad de Huelva que me ha permitido comenzar mi labor docente y continuar mi investigación. Mención aparte merecen los profesores, amigos y compañeros del Departamento de Ciencias Agroforestales que de una u otra manera, técnica o humana, me han brindado su apoyo, y entre ellos quiero destacar la ayuda en la validación del sistema experto de Ignacio de la Rosa y Luís Angel Zarazaga.

Mi asistencia al ‘V Curso de Introducción a las Redes Neuronales y Sistemas Borrosos’ en la Universidad de Verano de Teruel, marcó un punto de inflexión en el tratamiento de algunos de los modelos utilizados. El intercambio de impresiones con el profesor Bonifacio Martín del Brío (Universidad de Zaragoza), me facilitó la comprensión de los aspectos más oscuros de las Redes Neuronales. Asimismo, su asesoramiento sobre la predicción de algunas de las variables ha sido particularmente valioso. En este mismo sentido, también recordar la gentileza del profesor Víctor Prybotuk (Universidad de Texas).

La buena disposición de los profesores Sebastián Ventura y César Hervás (Universidad de Córdoba), que me proporcionaron muchas de las referencias y me resolvieron diversos aspectos relacionados con las Redes Neuronales Computacionales.

No quisiera dejar de citar al Profesor Titular de la Universidad de Córdoba, Dr. Carlos Fernández Delgado, quien me introdujo en el mundo de la ictiología, y del que aprendí que la investigación y la educación consisten en enseñar no lo que se debe pensar, sino a pensar. El Catedrático de Escuelas Universitarias de la Universidad de Huelva, Dr. José Prenda Marín, me inició en el estudio de las relaciones entre las poblaciones de peces y su entorno, aunque de nuestras horas de trabajo en común su

principal enseñanza fue que la amistad termina donde empieza el interés. A ambos por igual, muchas gracias.

A Unicaja y Analistas Económicos de Andalucía por la concesión de un accésit en la convocatoria del IV Premio Unicaja de Investigación sobre Desarrollo Económico y Estudios Agrarios, al trabajo ‘SEDPA: Sistema Experto de Ayuda al Diagnóstico de Patologías y de Asesoramiento Técnico en Piscifactorías de Anguilas’, inspirado en el capítulo 3 de la presente Tesis. Por favorecer y potenciar la investigación la cual es, sin duda alguna, la base del desarrollo de cualquier sociedad.

Mis amigos y compañeros de piso, Ramón y Salas, y sus respectivas ‘flores’, Isabel e Inmaculada, merecen una especial dedicatoria por esos días de trabajo y charlas compartiendo alegrías y desilusiones. Y por supuesto, a pesar de la distancia, no me quiero olvidar de los amigos de siempre, con los que he compartido muchas horas y cosas: José Carlos (el ingeniero industrial) a quien pude convencer de que un sistema biológico no se comporta como una máquina, Víctor, Jesús,... y todos los demás que estuvieron o están en algún momento u otro se interesaron por mi trabajo.

A mi familia, biológica y política, por su apoyo, estímulo, paciencia y confianza sin límites a lo largo de demasiados años. Por ese escaso tiempo que he tenido para dedicarles, ya que toda Tesis es en cierto modo como un hijo, y como tal exige su cuota de participación.

Por último, y sin embargo, el más importante de mis agradecimientos va dirigido a Inmaculada. Ella ha conocido, vivido y sufrido todas y cada una de las etapas de la presente Tesis. Ha soportado estoicamente mis divagaciones (a veces casi metafísicas) sobre los resultados, los interminables enfados con el soporte informático, mis constantes despistes, esos días en los que el ánimo se encuentra a ras de suelo y por qué no decirlo, aquellos en los que se encontraba demasiado elevado. Nunca hubo una palabra de queja después de escuchar miles de veces la canción ‘*Hide in your shell*’, y siempre estuvo a mi lado cuando las cosas salían bien y cuando salían mal. Es evidente que este trabajo no sería lo mismo sin ti.

A todos ellos, las más sinceras gracias, como expresión de la extraordinaria valoración que doy a su ayuda.

# Índice

<b>CAPÍTULO 1. INTRODUCCIÓN.....</b>	<b>1</b>
1.1. Objetivos generales .....	5
1.2. Objetivos específicos.....	6
1.3. Referencias .....	7
<b>CAPÍTULO 2. HIDRORECURSOS S.A. Y SU PRODUCTO.....</b>	<b>9</b>
2.1. La anguila europea .....	11
2.1.1. Descripción y posición taxonómica .....	11
2.1.2. Ciclo biológico.....	12
2.1.3. Anguicultura .....	13
2.2. Hidrorecursos S.A. ....	14
2.2.1. Marco geográfico .....	14
2.2.2. Antecedentes .....	17
2.2.3. Situación actual .....	19
2.2.3.1. Zona de alevinaje .....	20
2.2.3.2. Zona de preengorde.....	21
2.2.3.3. Zona de engorde.....	22
2.2.3.4. Biofiltros .....	23
2.2.3.5. Sistemas de control.....	27
2.3. El programa de control en diferido <b>CONTROLPLANT 1.0.</b> ....	27
2.3.1. Inicio del programa .....	29
2.3.2. Ventana principal de <b>CONTROLPLANT</b> .....	29
2.3.2.1. Barra de herramientas.....	30
2.3.2.2. Sección superior de la ventana principal.....	35
2.3.2.3. Sección inferior de la ventana principal.....	36
2.4. Referencias .....	36
<b>CAPÍTULO 3. SISTEMA EXPERTO DE DIAGNÓSTICO PATOLÓGICO PARA ANGUILAS (SEDPA).....</b>	<b>39</b>
3.1. Introducción .....	41
3.1.1. Los sistemas expertos como métodos de diagnóstico .....	42
3.1.2. Formulación del problema del diseño de un sistema experto para el diagnóstico clínico.....	44
3.2. Descripción del sistema en estudio.....	45
3.3. Agentes patológicos reconocidos por el sistema .....	45
3.3.1. Componentes del medio ambiente .....	48
3.3.2. Virus.....	49
3.3.3. Bacterias.....	49
3.3.4. Hongos .....	51
3.3.5. Parásitos .....	51
3.4. Material y métodos.....	52
3.4.1. Bases de conocimiento del dominio.....	52
3.4.2. El motor de inferencia.....	54
3.4.2.1. Análisis morfológico-sintáctico. Red de Transición Aumentada ( <b>RTA</b> ).....	55
3.4.2.2. El controlador borroso.....	59
3.4.2.2.1. Lógica borrosa y teoría de conjuntos borrosos. Conjuntos borrosos y funciones de inclusión.....	59

3.4.2.2.2. Reglas de inferencia borrosa y Memoria Asociativa Borrosa ( <b>MAB</b> ). Asignación de niveles de creencia individuales y globales.....	62
3.4.2.3. Teoría de Dempster-Shafer para la transmisión de incertidumbre.....	68
3.4.2.4. Test $\chi^2$ modificado como elemento diferenciador de los grupos patológicos con mayor aportación a la creencia global.....	72
3.4.3. Subsistema explicativo.....	73
3.4.4. Subsistema de sugerencia o motor proposicional.....	75
3.4.5. Subsistema de aprendizaje.....	77
3.4.6. Valoración del diagnóstico.....	77
3.4.7. Validación del sistema.....	78
3.5. Resultados.....	81
3.5.1. Información contenida en las bases de datos del dominio principal y secundario.....	81
3.5.2. Número de casos utilizados para la evaluación.....	83
3.5.3. Análisis de las respuestas aportadas por los expertos humanos.....	83
3.5.4. Análisis preliminar de las Memorias Asociativas Borrosas.....	96
3.5.5. Análisis de las respuestas de <b>SEDPA</b> .....	97
3.5.6. Agrupación de diagnósticos individuales de <b>SEDPA</b> .....	109
3.5.7. Diagnóstico de los expertos frente al diagnóstico de <b>SEDPA</b> .....	115
3.5.8. Selección de la mejor configuración de <b>SEDPA</b> .....	119
3.5.9. Análisis de las respuestas de <b>SEDPA</b> con reducción de información.....	121
3.5.10. Respuesta del subsistema sugerencia o motor proposicional.....	123
3.6. Discusión.....	125
3.6.1. Respuestas de los expertos humanos.....	126
3.6.2. Lenguaje natural.....	128
3.6.3. Respuestas de <b>SEDPA</b> : creencias y valoraciones. Conclusiones del motor de inferencia y proposicional.....	129
3.6.4. Expertos humanos frente a <b>SEDPA</b> .....	131
3.7. Referencias.....	133
3.8. ANEXO III.1. Glosario.....	143
3.9. ANEXO III.2. Manual del usuario.....	147
<b>CAPÍTULO 4. CARACTERIZACIÓN Y PREDICCIÓN DE LOS PRINCIPALES PARÁMETROS FÍSICO-QUÍMICOS DEL AGUA.....</b>	<b>165</b>
4.1. Introducción.....	167
4.1.1. Descripción de los factores físico-químicos críticos.....	167
4.1.1.1. Oxígeno disuelto.....	167
4.1.1.2. Amoníaco, nitritos y nitratos.....	169
4.1.1.3. Temperatura.....	171
4.1.1.4. Carácter ácido-básico (pH).....	172
4.1.2. Aprovechamiento de los efluentes de refrigeración de las centrales térmicas.....	173
4.1.3. El problema de la predicción.....	174
4.1.4. Desarrollo histórico de las Redes Neuronales Computacionales ( <b>RNCs</b> ).....	174
4.1.5. Aplicaciones de las Redes Neuronales Computacionales.....	176
4.1.5.1. Biología, producción vegetal, producción animal, química y medicina.....	177
4.1.5.2. Hidrología, hidráulica e ingeniería.....	178
4.1.5.3. Comparación con otras metodologías.....	178
4.2. Material y métodos.....	179
4.2.1. Fuente de los datos.....	179
4.2.2. Métodos tradicionales de predicción: análisis multivariante (regresión múltiple) y análisis univariante (descomposición y <b>ARIMA</b> ).....	179



4.2.2.1. Análisis de regresión múltiple.....	179
4.2.2.2. Métodos de suavizado y modelos <b>ARIMA</b> .....	180
4.2.3. Predicción de los principales factores físico-químicos críticos mediante la aplicación de Redes Neuronales Computacionales .....	182
4.2.3.1. Función de transferencia .....	184
4.2.3.2. Entrenamiento de la Red Neuronal .....	184
4.2.3.3. Arquitectura de las Redes Neuronales .....	187
4.2.3.4. Patrones de entrenamiento durante la calibración.....	188
4.2.4. Generalización de los modelos.....	188
4.2.5. Pretratamiento de la información .....	191
4.3. Resultados .....	192
4.3.1. Variables de la Central Térmica de Puente Nuevo. Análisis de producción bruta.....	192
4.3.2. Variables de planta .....	194
4.3.2.1. Temperatura en la serie 'A'.....	194
4.3.2.2. Temperatura en la serie 'B'.....	203
4.3.2.3. Temperatura en las series 'E' y 'F' .....	208
4.3.2.4. Amoníaco en la serie 'A'. Estimación a partir de las concentraciones de amoníaco en los días previos .....	214
4.3.2.5. Amoníaco en la serie 'A'. Estimación a partir de otras variables implicadas .....	222
4.3.2.6. Amoníaco en la serie 'A'. Estimación a partir de la serie filtrada .....	226
4.3.2.7. Amoníaco en las series 'B', 'E' y 'F' .....	233
4.3.2.8. Nitritos en la serie 'A' .....	237
4.3.2.9. Nitritos en las series 'B', 'E' y 'F' .....	245
4.3.2.10. Nitratos en la serie 'A' .....	250
4.3.2.11. Nitratos en las series 'B', 'E' y 'F' .....	259
4.3.2.12. pH en la serie 'A' .....	263
4.4. Discusión.....	266
4.4.1. Temperatura .....	266
4.4.2. Amoníaco .....	269
4.4.3. Nitritos y nitratos.....	271
4.4.4. pH.....	272
4.5. Referencias .....	272
4.6. ANEXO IV.1. Glosario .....	289
4.7. ANEXO IV.2. Manual del usuario del simulador de Redes Neuronales Computacionales ( <b>REDGEN 1.0</b> ).....	293

<b>CAPÍTULO 5. CARACTERIZACIÓN Y PREDICCIÓN DE PARÁMETROS BIOLÓGICOS. ESTRATEGIA DE PRODUCCIÓN.....</b>	<b>303</b>
5.1. Introducción .....	305
5.1.1. Modelación del crecimiento y las estrategias de producción .....	306
5.1.2. Crecimiento de la anguila.....	307
5.2. Material y métodos.....	308
5.2.1. Caracterización de la biomasa y mortalidad real.....	308
5.2.2. Estimación de la desviación típica o estándar. Derivación de los modelos de varianza .....	310
5.2.3. La tasa de crecimiento.....	313
5.2.4. La tasa de mortalidad .....	316
5.2.5. Modelación de la cosecha .....	317
5.2.6. Proceso de simulación dinámica. Valoración económica inicial .....	319
5.3. Resultados .....	321

5.3.1. Error de predicción del gestor .....	321
5.3.2. Análisis de las muestras del tanque 'A1'. Relación longitud-peso .....	322
5.3.3. Estimación de los rangos de variación de la desviación típica. Aplicación de los modelos de varianza.....	324
5.3.4. Caracterización de la biomasa.....	325
5.3.5. Caracterización de la mortalidad.....	335
5.3.6. Análisis de los resultados obtenidos con el Programa Operativo de Simulación (POS).....	338
5.3.6.1. Simulaciones de referencia.....	339
5.3.6.2. Simulación de referencia 1 .....	340
5.3.6.3. Simulación de referencia 2 .....	342
5.3.6.4. Simulación 3. Incorporación de una nueva cohorte.....	344
5.3.6.5. Simulación 4. Tres cohortes, tres años e incorporación del factor de crecimiento debido a la temperatura .....	347
5.3.6.6. Simulación 5. Incorporación del factor de crecimiento debido a la concentración de amoníaco .....	350
5.3.6.7. Simulación 6. Incorporación del factor de crecimiento debido a la concentración de nitritos .....	350
5.3.6.8. Simulación 7. Incorporación del factor de crecimiento debido a la concentración de nitratos.....	353
5.3.6.9. Simulación 8. Incorporación del factor de crecimiento debido al pH .....	357
5.3.6.10. Simulación 9. Factores de crecimiento debidos a la temperatura y amoníaco. Influencia de los porcentajes iniciales en cada subgrupo de crecimiento .....	357
5.3.6.11. Simulación 10. Influencia del lote de cosecha.....	362
5.3.6.12. Simulación 11. Influencia del cambio del lote de cosecha durante la simulación .....	362
5.3.6.13. Simulación 12. Influencia del tamaño de cosecha.....	364
5.3.6.14. Simulación 13. Aplicación de los modelos de predicción de las variables temperatura y amoníaco .....	366
5.3.6.15. Simulación 14. Conexión con <b>SEDPA</b> 1.0 a través de las reglas A-A de <b>CONTROLPLANT</b> 1.0.....	367
5.4. Discusión.....	369
5.4.1. Variación de la biomasa y mortalidad.....	370
5.4.2. Comportamiento y robustez del modelo .....	370
5.4.3. Inclusión de factores de crecimiento debidos a parámetros físico-químicos .....	371
5.4.4. Conexión con <b>SEDPA</b> y reglas A-A de <b>CONTROLPLANT</b> .....	373
5.5. Referencias .....	373
5.6. ANEXO V.1. Glosario .....	379
5.7. ANEXO V.2. Manual del usuario del Programa Operativo de Simulación (POS).....	381
<b>CAPÍTULO 6. CONCLUSIONES .....</b>	<b>389</b>



# CAPÍTULO 1

---

## INTRODUCCIÓN

- 1.1. Objetivos generales
- 1.2. Objetivos específicos
- 1.3. Referencias



## CAPÍTULO 1. INTRODUCCIÓN

A principios de siglo abrigábamos la ilusión de que el mar podría alimentar a todo el mundo. Después de la Segunda Guerra Mundial, la pesca marítima creció rápidamente como consecuencia en parte de las nuevas mejoras tecnológicas de los sistemas de captura. Sin embargo este crecimiento, llevado a cabo a costa de muchas poblaciones de peces que pensábamos podrían suministrar una pesca estable, empezó a declinar en los años setenta a causa de la intensa explotación. Este estancamiento de la pesca tradicional coincidió con un período en el cual la demanda de proteínas de origen animal se multiplicaba a medida que la población humana crecía. De esta forma, la acuicultura se erige en esta época como una actividad de futuro que se desarrolla y crece frente a otro tipo de explotaciones gracias a la experiencia adquirida a nivel experimental y de laboratorio en años precedentes.

La idea de que la acuicultura podría convertirse en una de las principales actividades capaz de paliar la falta de alimento puede deberse a que es posible afirmar que los peces poseen ciertas ventajas para la cría frente a los animales terrestres. A pesar de que las comparaciones de productividad entre peces y animales terrestres sean poco significativas, se puede decir que esta afirmación se fundamenta básicamente en tres razones: a) los peces son animales poiquiloterms, por lo que no consumen energía en el mantenimiento de la temperatura corporal, b) sus ciclos vitales se desarrollan dentro del agua, no teniendo por tanto que desperdiciar energía soportando su propio peso y, c) por regla general presentan una menor proporción de huesos y partes no comestibles. Todo esto hace que la conversión del alimento en carne sea mucho más eficiente que en los animales terrestres (Bone *et al.*, 1995; Shepherd y Bromage, 1999). El conjunto de estos factores junto con el estancamiento del crecimiento de la industria pesquera o como en el caso de la anguila, el descenso de las capturas de la pesca tradicional, el aumento de la demanda humana de proteínas y el avance en el conocimiento de la biología de determinadas especies, ha provocado que la producción piscícola esté contribuyendo en los últimos años en una gran proporción al suministro mundial de pescado. De esta forma se ha pasado de los 10.1 millones de toneladas producidas en el año 1986, lo que suponía un 7.9 % de la pesca total mundial, a los 19.3 millones de toneladas producidas en 1992 (16.2 %), lo que indica un incremento medio anual del 6 % (FAO, 1992 y 1994). De esta producción, la mayor parte corresponde a la llevada a cabo en aguas continentales (65 %), mientras que el resto (35 %) corresponde a agua salada o salobre (Buxadé, 1997).

En España la tendencia de crecimiento de la producción acuícola se ha invertido en los últimos años, descendiendo desde el 17 % sobre la producción total en 1984 hasta el 10 % en 1995. Sin embargo esta disminución está íntimamente ligada a la producción de mejillón, cuyo cultivo representaba en 1984 el 93 % de la producción y que pasó a ser en 1995 del 67 %. Así, cuando se excluye el cultivo del mejillón, la producción acuícola presenta un comportamiento totalmente diferente, pasando de 20900 toneladas en 1984 a 46000 toneladas en 1995, lo que supone un crecimiento del 120 % en el período

considerado. Esto equivale a una tasa de crecimiento anual media del 6.8 %. La mayor tasa de crecimiento se ha realizado en la acuicultura marina, con un 14.5 % de media al año, respecto a la continental, que sólo ha alcanzado el 2.6 % (Cobo *et al.*, 1999).

En acuicultura, dependiendo de la especie y las instalaciones se diferencian multitud de sistemas de cultivos. Éstos se pueden agrupar en dos grandes bloques en función del tipo de alimentación: acuicultura de producción o extensiva, donde la dieta de los animales está basada en alimento natural y, por tanto, el aporte de alimento manufacturado es escaso o nulo (p.e.: el cultivo de mejillones), y acuicultura de transformación o intensiva, donde los productos y subproductos de bajo valor comercial de la industria pesquera transformados en pienso de alto poder energético, son utilizados para la alimentación de especies muy apreciadas y de gran valor en el mercado.

En las explotaciones intensivas conforme aumenta la densidad de cría se ejerce un mayor y más preciso control de las operaciones llevadas a cabo sobre la población existente. Un análisis más profundo del manejo permite optimizar los niveles de nutrición dando estimaciones más predecibles de su crecimiento, así como abarcar los regímenes de retirada de los productos finales. Como resultado hay menos alimento perdido, las pérdidas de individuos no contabilizados son menores y el piscicultor puede controlar y responder más fácilmente a problemas tales como la enfermedad, las variaciones físico-químicas, etc. En conjunto se tiende a un uso más eficiente del capital ligado a los peces, alimentos y la planta en sí. Esta optimización puede llevarse a cabo mediante la introducción de mejoras técnicas, nuevas maquinarias e importantes inversiones estructurales. Pero al mismo tiempo, factores tales como el incremento de las producciones o la competencia generan bajadas en el precio del producto, lo que unido a un incremento paralelo del precio de los suministros, obliga a los empresarios a reducir costes de producción. Esto lleva a considerar en muchas explotaciones acuícolas otro tipo de planteamientos de gestión técnica y económica.

Este fenómeno se ve magnificado en aquellos sectores productivos con mercados temporales y localizados como el de la anguila. La producción de anguilas en España no supera en la actualidad las 300 toneladas anuales y su producción se destina casi por completo a la zona levantina, Galicia y como exportación al centro y norte de Europa, donde existe una gran tradición de su consumo en determinadas épocas del año. Se podría pensar, por tanto, que la anguicultura se encuentra en una etapa de crecimiento temprano caracterizada por una producción baja, un precio alto y estable y unas previsiones de venta futuras crecientes. Sin embargo éste es un mercado ya estabilizado donde el precio por kilogramo es establecido por las grandes explotaciones intensivas del norte de Europa que presentan amplios márgenes de beneficios. Estos márgenes les permiten fijar un precio más o menos constante con ligera tendencia al alza, a pesar de que la anguicultura sea un sector dedicado al engorde y absolutamente dependiente de poblaciones naturales cada vez

más esquilgadas. De esta forma, el precio de la anguila de peso comercial sufrió pequeñas variaciones en el período comprendido entre 1988 y 1996 (5.39-13.08 €/kg) con un incremento interanual del 23.6 %. Sin embargo, si deflactamos estos precios en función del incremento del **IPC** anual encontramos que la tasa de crecimiento del precio del kilogramo de anguilas fue del -18.6 % (Cobo *et al.*, 1999). Esto obliga a los anguicultores españoles a aproximar aun más las predicciones sobre mortalidad, consumo de pienso y fármacos, incremento de los crecimientos y, en general, a tener un mayor control sobre todos los parámetros que influyen en la producción final para así conseguir una reducción de costes.

La reducción de los costes en su conjunto pasa por conseguir la modelación y caracterización de la producción. El gran número de factores que intervienen y afectan a las producciones en el medio acuático, sobre todo en explotaciones de carácter intensivo, así como las complejas relaciones existentes entre ellos, hacen que el proceso de modelado sea un objetivo difícil. La combinación de técnicas estadísticas clásicas con sistemas englobados tradicionalmente dentro del campo de la inteligencia artificial (**IA**) (sistemas expertos basados en conocimiento y lógica borrosa, y redes neuronales computacionales), puede proporcionar una buena aproximación a la solución del problema. El uso de la **IA** ha llegado a ser común en la industria y en los sistemas de control de procesos en los últimos años (Rock y Guerin, 1992). Lee (1995 y 2000) destaca que entre las principales ventajas que la **IA** tiene para la acuicultura frente a otro tipo de métodos se encuentran: a) la rápida transferencia del conocimiento del experto a lo largo de un sector industrial, especialmente a aquellas industrias que no tienen disponibles el número suficiente de expertos, b) una posible reducción de costes como consecuencia de la automatización de la mayor parte de las funciones primarias, c) mejora del proceso de producción en estabilidad y eficiencia y, d) mejora del entendimiento del proceso de producción a través del desarrollo y evaluación de reglas y algoritmos.

El objetivo principal de este estudio es conseguir una mejora del régimen de explotación a través de unas herramientas de predicción y simulación lo suficientemente fiables, que permitan al acuicultor prever a corto y medio plazo los acontecimientos que van a suceder desde un punto de vista patológico, así como a nivel de control de parámetros físicos, químicos y biológicos.

## **1.1. OBJETIVOS GENERALES**

De manera global, los principales beneficios que se obtendrán de la aplicación del sistema de predicción y simulación son:

1. Desde el punto de vista patológico, se dispondrá de un sistema que permitirá un diagnóstico precoz de la enfermedad o agente patológico, gracias al cual será



posible la aplicación del tratamiento preventivo adecuado, con el consecuente ahorro de productos terapéuticos y disminución de la mortalidad.

2. La predicción de las condiciones fisico-químicas permitirá al mismo tiempo evaluar el comportamiento de la planta en su totalidad, de tal forma que el gestor podrá adecuar las actividades de los operarios para, en el caso de un peligro significativo, establecer un sistema de control que modifique la desviación de los parámetros afectados hasta sus niveles normales. Esto se puede reflejar en la disminución del estrés inducido por la anormalidad de las condiciones de habitabilidad y, por tanto, en la reducción de la posibilidad de aparición de focos infecciosos favorecidos por factores estresantes.
3. La predicción de la evolución de los parámetros poblacionales facilitará el control de la producción, mortalidad, distribución de las raciones diarias de alimento, seguimiento del *stock*, y puede servir como indicativo de la estabilidad o variabilidad de las condiciones ambientales dentro de la explotación.

## **1.2. OBJETIVOS ESPECÍFICOS**

Esta memoria se fundamenta en los siguientes objetivos específicos:

1. La construcción de un sistema de decisión soporte que integre la base de datos generada diariamente en la planta, el sistema experto de diagnóstico patológico, los modelos estadísticos clásicos encontrados como consecuencia del análisis de los datos históricos, las redes neuronales computacionales igualmente generadas, así como los modelos que definen la estrategia de producción.
2. La programación de un sistema experto para el diagnóstico de las posibles enfermedades y agentes patógenos que pueden afectar a una planta de cultivo de anguilas en régimen de explotación intensiva. Este incluirá un módulo de introducción de información en forma de lenguaje natural para facilitar su uso por parte de todos los operarios de la planta. Asimismo, será capaz de proporcionar conclusiones a partir de información escasa e imprecisa gracias al tratamiento de la incertidumbre por parte del motor de inferencia. Presentará además un módulo de adquisición de conocimiento, un módulo proposicional y un módulo explicativo.
3. La modelación de los principales parámetros fisico-químicos y poblacionales que permita predecir su comportamiento a corto y medio plazo. Se propone junto con técnicas tradicionales de predicción, como los análisis univariantes

(métodos de descomposición y ARIMA) y multivariantes (regresión múltiple) de series temporales, el uso de Redes Neuronales Computacionales que han demostrado conseguir muy buenos resultados en el campo de las técnicas de control.

4. La simulación de la estrategia de producción mediante el uso de modelos de crecimiento y de extensión de la varianza que incluyan además los modelos calibrados y validados en apartados anteriores.

### 1.3. REFERENCIAS

- Bone, Q., N.B. Marshall y J.H.S. Blaxter. 1995. *Biology of fishes*. Edita: Chapman & Hall. Londres.
- Buxadé, C. 1997. *Zootecnia. Bases de la producción animal. Producción animal acuática*. Edita: Mundi-Prensa. Madrid.
- Cobo, A., D. González-López, V. Iglesias-Argüelles, F. Ledesma, L. Luna-Sotorrio, F.J. Martínez, A. Martín-Hernández, J. Fernández-Polanco, S.M. Ruesga, F. Somohano y J.A. Trespacios-Gutiérrez. 1999. *La acuicultura. Biología, regulación, fomento, nuevas tendencias y estrategia comercial. Tomo II: Economía y gestión de la acuicultura*. Edita: Mundi-Prensa. Madrid.
- FAO. 1992. *Aquaculture production 1984-1990*. FAO Fisheries Circular (Nº 815-4). Roma.
- FAO. 1994. *Aquaculture production 1986-1992*. FAO Fisheries Circular (Nº 815-6). Roma.
- Lee, G.P. 1995. A review of automated control systems for aquaculture and design criteria for their implementation. *Aquacultural Engineering*, 14: 205-227.
- Lee, G.P. 2000. Process control and artificial intelligence software for aquaculture. *Aquacultural Engineering*, 23: 13-26.
- Rock, D. y D. Guerin. 1992. Applying AI to statistical process control. *AI Expert*, 7: 30-35.
- Shepherd, J. y N. Bromage. 1999. *Piscicultura intensiva*. Edita: ACRIBIA, S.A. Zaragoza.



## CAPÍTULO 2

---

### **HIDRORECURSOS S.A. Y SU PRODUCTO**

- 2.1. La anguila europea
  - 2.1.1. Descripción y posición taxonómica
  - 2.1.2. Ciclo biológico
  - 2.1.3. Anguicultura
- 2.2. Hidrorecursos S.A.
  - 2.2.1. Marco geográfico
  - 2.2.2. Antecedentes
  - 2.2.3. Situación actual
    - 2.2.3.1. Zona de alevinaje
    - 2.2.3.2. Zona de preengorde
    - 2.2.3.3. Zona de engorde
    - 2.2.3.4. Biofiltros
    - 2.2.3.5. Sistemas de control
- 2.3. El programa de control en diferido **CONTROLPLANT 1.0**
  - 2.3.1. Inicio del programa
  - 2.3.2. Ventana principal de **CONTROLPLANT**
    - 2.3.2.1. Barra de herramientas
    - 2.3.2.2. Sección superior de la ventana principal
    - 2.3.2.3. Sección inferior de la ventana principal
- 2.4. Referencias



## **CAPÍTULO 2. HIDRORECURSOS S.A. Y SU PRODUCTO**

### **2.1. LA ANGUILA EUROPEA**

#### **2.1.1. Descripción y posición taxonómica**

La anguila europea (*Anguilla anguilla* L., 1758) es un pez eurihalino que se caracteriza por el cuerpo alargado, serpentiforme, de sección redondeada en la parte anterior y con el extremo posterior comprimido lateralmente. Carecen de aletas pélvicas aunque poseen una larga aleta impar continua de unos 500 radios resultante de la fusión de la aleta dorsal con la anal. En la cabeza se pueden distinguir los orificios branquiales, los ojos, los orificios nasales y la boca.

Las aberturas branquiales son cortas, estrechas y están protegidas por los opérculos. Son animales lucífugos que presentan una visión deficiente que es compensada mediante el órgano olfativo. Éste no tiene comunicación con la cavidad oral, sino que se compone de dos sacos ciegos nasales separados, localizados uno a cada lado de la parte superior del rostro. Existen dos orificios nasales por saco. La boca se abre hasta la parte inferior de los ojos estando compuesta por fuertes mandíbulas y teniendo carácter prognático.

Presentan escamas muy pequeñas sobre una piel rica en glándulas mucosas que la hace viscosa y la protege de las variaciones osmóticas. Pueden alcanzar hasta un 60 % de respiración cutánea y permanecen vivas fuera del agua varias horas siempre y cuando su piel se mantenga húmeda. Esta característica les permite flanquear importantes obstáculos durante sus migraciones. El esqueleto está completamente osificado. Existe una gran irregularidad en el crecimiento, variando tanto su tamaño como el peso en ejemplares de la misma edad, por lo que ésta es difícil de establecer mediante escalimetría. Para conocer la edad de las anguilas se recurre al estudio de los otolitos.

El sexo no se puede reconocer por caracteres externos, exceptuando la longitud. En las anguilas adultas los machos son de menor tamaño, midiendo entre 20 y 40 centímetros, mientras que las hembras pueden sobrepasar los 150 centímetros. Tanto testículos como ovarios ocupan dos bandas que se extienden a lo largo del cuerpo, desde las aberturas branquiales hasta poco después del ano. Los óvulos presentan un tamaño de unas 300 micras y califican a la anguila como uno de los peces más prolíficos al alcanzar la fecundidad cifras de entre 5 y 20 millones de huevos.

La posición taxonómica de la anguila europea es la siguiente:

Clase OSTEICTIOS

Orden ANGUILIFORMES

Familia ANGUILLIDAE

Género *Anguilla*

Especie *A. anguilla* (Linnaeus, 1758)

### 2.1.2. Ciclo biológico

Las anguilas son peces catádromos que desovan en localizaciones oceánicas específicas, dependiendo de las especies implicadas. El desarrollo de los leptocéfalos de la anguila europea comienza en el Mar de los Sargazos. Durante su viaje a través del océano, estas larvas transparentes nadan activamente alimentándose de plancton. En esta fase tienen una forma similar a una hoja de sauce. El proceso de metamorfosis hacia la típica forma redonda y larga que asociamos con las anguilas, comienza cuando los leptocéfalos alcanzan el borde de la plataforma continental después de un viaje que dura entre 2 y 7 años. Se convierten entonces en anguilas transparentes, listas para entrar en las aguas continentales (Brown, 2000).

La entrada en las aguas salobres de los estuarios, deltas y lagunas del litoral mediterráneo se produce con mayor intensidad entre los meses de agosto y marzo durante las noches de luna nueva, aunque pueden ser capturadas durante todo el año (Arias-García y Drake-Moyano, 1990). Remontan los cauces en largos cordones, cercanos a los márgenes y al fondo, en cada marea favorable. Algunas permanecen en zonas próximas a la costa, como en aguas de albuferas o marjales, de salinidad variable mientras que otras remontan hasta la cabecera de los ríos. La selección de la zona en aguas salobres parece estar relacionada con la capacidad de aclimatación a distintas temperaturas, y en aguas dulces con las preferencias de los estímulos olorosos independientemente de la temperatura (Tongiorgi *et al.*, 1986; Tosi *et al.*, 1988, 1990; Sola y Tosi, 1993; Tosi y Sola, 1993; Sola, 1995; Sola y Tongiorgi, 1996).

Una vez establecidas se produce una reactivación del crecimiento como consecuencia de la actividad trófica. Su dieta está compuesta por cualquier presa animal de tamaño adecuado, siguiendo un comportamiento oportunista y adaptándose a lo más abundante en cada zona.

El tiempo necesario para alcanzar la madurez sexual es variable, dependiendo básicamente del tamaño del individuo (Beullens *et al.*, 1997). Los dos sexos exigen para la maduración un mínimo de talla que garantice una capacidad energética suficiente para el

desarrollo gonadal, la migración y el éxito en la reproducción. La migración reproductiva comienza en los meses de otoño. Las anguilas abandonan las aguas continentales, desplazándose desde las orillas hasta las corrientes de los cursos de agua que conducen a las costas. Desde allí, tardan aproximadamente unos 18 meses en llegar hasta el Mar de los Sargazos donde tiene lugar la puesta en primavera o verano. Tras ésta se supone que los ejemplares adultos mueren, ya que no se ha encontrado ninguno en viaje de retorno (Usui, 1978).

### 2.1.3. Anguicultura

La anguicultura como tal comienza en Japón a principios de la *Era Meiji*, en la segunda mitad del siglo XIX con el cultivo de la *Anguilla japonica* (L. 1758). El objetivo era conseguir anguilas de entre 150 y 200 gramos que permitieran confeccionar el tradicional 'Kabayaki con soyu'. De pequeñas explotaciones feudales se pasó en poco tiempo a grandes explotaciones semiintensivas en el sector del lago Hamana y llanuras vecinas. La asociación de estas explotaciones constituyó la Prefectura de Shizuoka, que sigue siendo en la actualidad la zona de mayor producción de Japón. El sistema de producción consistía en llevar las angulas hasta estado de anguilón durante el primer año y engordarlas durante dos o tres años más. Después de un período de progreso constante impulsado por una fuerte demanda, se produce un importante retroceso como consecuencia de la II Guerra Mundial y la reforma agraria que la sucedió, debido a la cual desaparecieron el 80 % de las explotaciones (Tesch, 1977). Durante el principio de la segunda mitad del siglo XX, el progreso de la hidráulica, las mejoras en el manejo de las explotaciones y la alimentación, y la fuerte demanda hace que el número de granjas se multiplique hasta principios de los años 70. En esta década las capturas de angulas para abastecimiento disminuyen de forma drástica, derrumbándose la producción y originando una profunda crisis, a raíz de la cual nace la anguicultura moderna.

En Europa el cultivo de modo extensivo se realiza en Polonia, Suecia, Francia y sobre todo Italia, donde existe una larga tradición del cultivo de la anguila conocida con el nombre de 'vallicultura'. Las explotaciones intensivas se han desarrollado a partir de principios de los años 80, llevándose a cabo numerosos ensayos a nivel experimental. La experiencia adquirida sirvió para que hoy en día funcionen multitud de explotaciones en Francia, Portugal, Italia, Alemania, Gran Bretaña, Países Bajos y Dinamarca. Esta última figura en cabeza en lo que atañe a técnicas de recirculación y reciclaje de agua (Carrillo, 1999).

En España la explotación de la anguila ya aparece en escritos del siglo XIII, cuando se crea la Comunidad de Pescadores de la Albufera de Valencia (Jover-Cerdá, 1993). En la



actualidad existen dos piscifactorías con regímenes de explotación intensiva: ‘Valenciana de Acuicultura S.A.’ en Valencia e ‘Hidrorecursos S.A.’ en la provincia de Córdoba, objeto de este estudio.

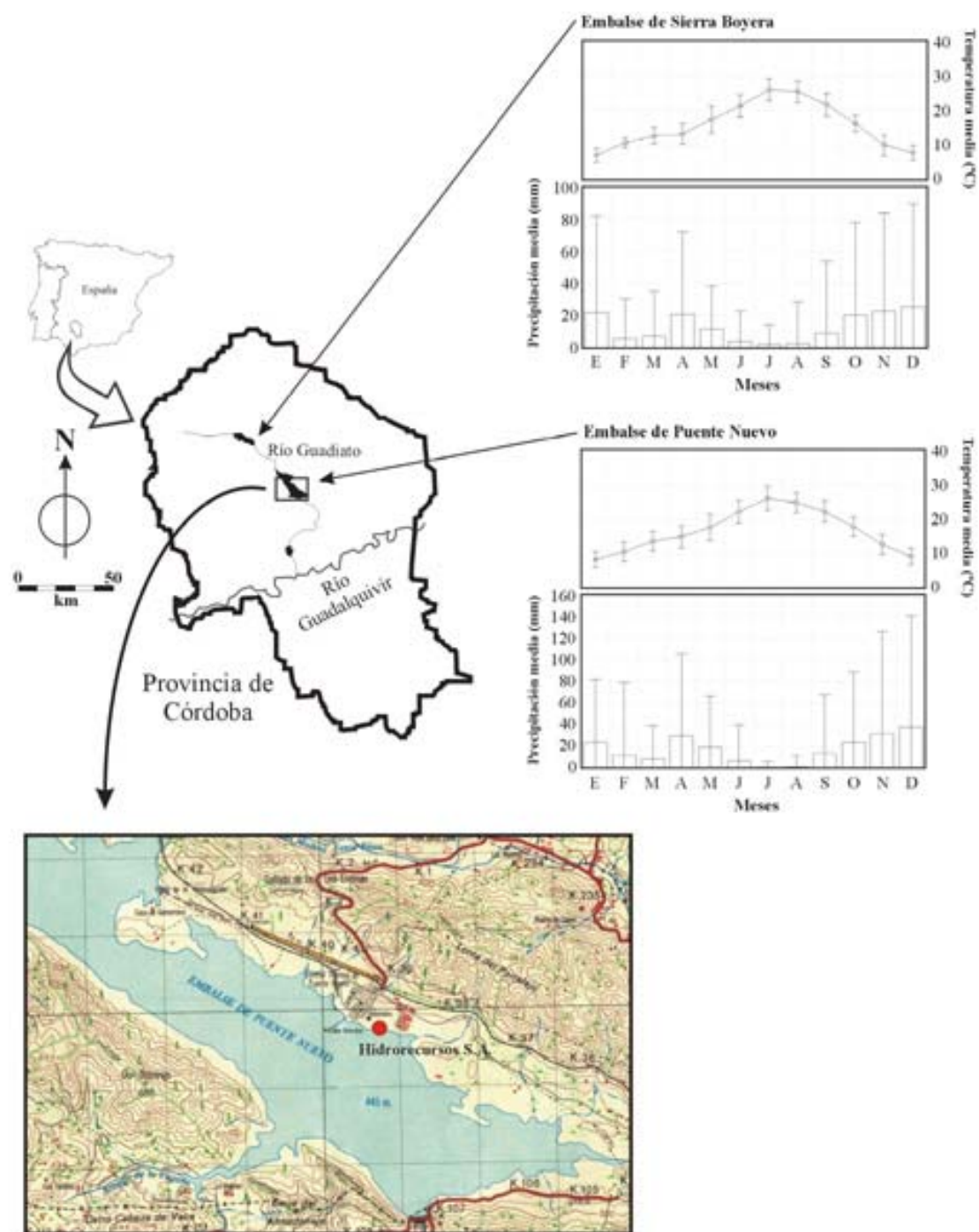
## **2.2. HIDRORECURSOS S.A.**

### **2.2.1. Marco geográfico**

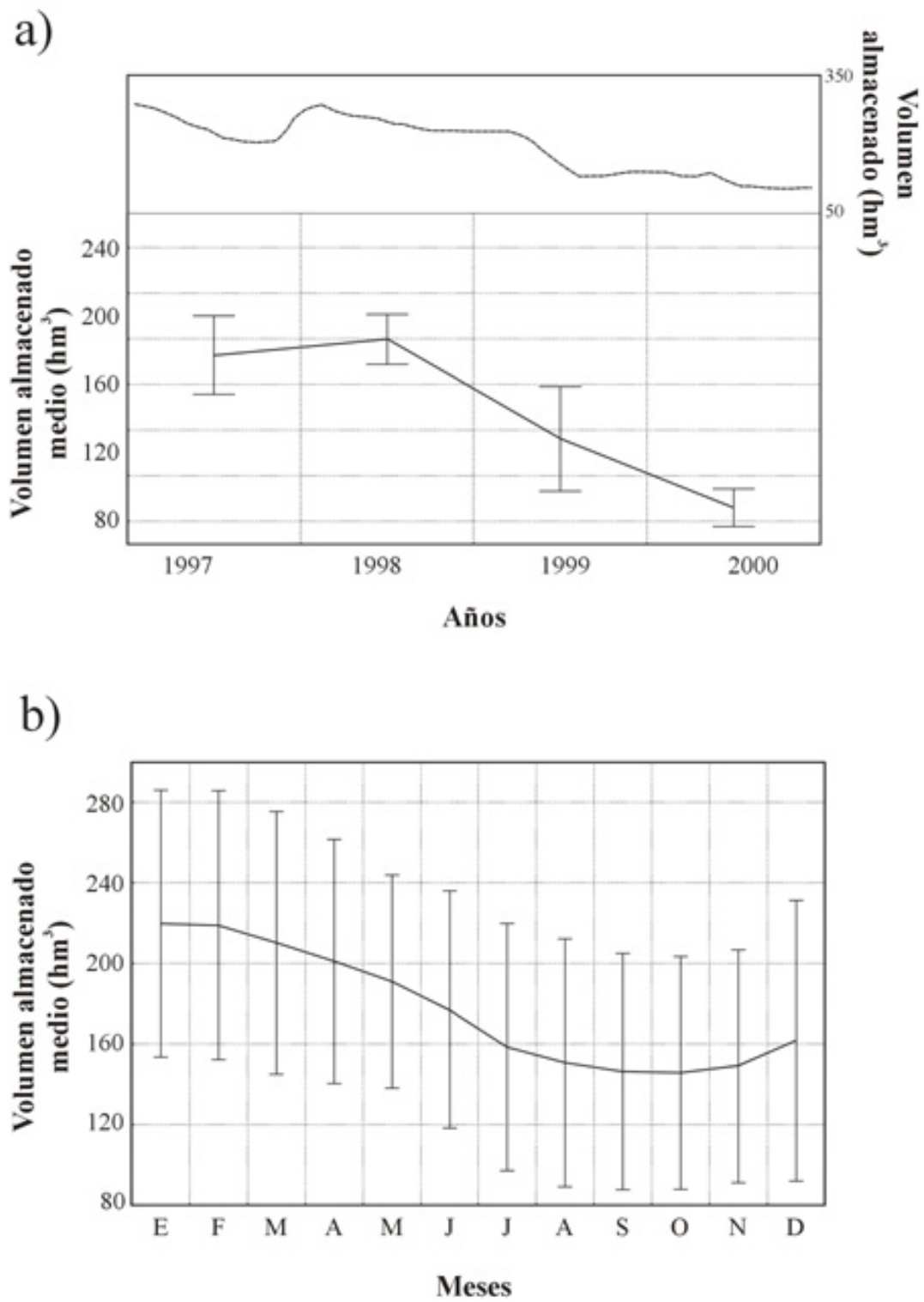
La piscifactoría Hidrorecursos S.A. se encuentra situada en la mitad septentrional de la provincia de Córdoba (30S UH 305198), a una altura de 450 metros y a orillas del embalse de Puente Nuevo (figura II.1). Éste se encuentra englobado en la cuenca del río Guadiato, el cual discurre por una amplia zona del norte de la misma provincia. Según Cabanás (1975) la zona es parte del conjunto apalechense que en la provincia de Córdoba aparece formada por la típica sucesión paralela de valles y sierras con rumbos NW-SE, estando los primeros fraguados en materiales pizarrosos blandos, en tanto que los segundos se corresponden con corridas de cuarcitas y otros materiales duros.

Ferreras (1980) incluye toda la cuenca del río Guadiato en la región mesomediterránea, a excepción de la zona de Peñarroya Pueblonuevo, que pertenece a la subregión termomediterránea. En su totalidad se puede decir que cuenta con períodos húmedos y suaves, con temperaturas que en pocas ocasiones descienden de los 0 °C y períodos con temperaturas muy altas que pueden llegar a sobrepasar los 40 °C (figura II.1). Estas condiciones climáticas tienen una fuerte influencia en el régimen hídrico del río Guadiato. Así pues, como en casi todos los ríos de carácter mediterráneo, en el período de lluvias abundantes se producen fuertes crecidas e incluso desbordamientos. Por el contrario, durante el estío la corriente desaparece y el agua queda confinada en pozas sin conexión entre sí donde se refugia la fauna acuática. Estas condiciones de habitabilidad se vuelven más extremas si cabe con la regulación artificial del caudal (Castelló, 1980; Herrera, 1991; Gutiérrez-Estrada, 1997; Gutiérrez-Estrada *et al.*, 2000).

La cantidad de agua diaria que la piscifactoría necesita para su funcionamiento es obtenida del embalse de Puente Nuevo. Construido entre los años 1956 y 1973 para paliar las necesidades de refrigeración de la central térmica (instalada en la margen izquierda del mismo), y como reservorio para el embalse de la Breña (situado unos kilómetros aguas abajo), presenta una capacidad total de 282 hm<sup>3</sup>. Desde el año 1997 el volumen almacenado ha variado significativamente (ANOVA: F=1709.5, p<sub>α</sub><0.05) alcanzándose valores medios máximos en el año 1998 y mínimos en el 2000 (figura II.2a). El análisis de la varianza de los volúmenes almacenados por meses indica que enero y febrero son los meses en los que habitualmente se tiene una mayor cantidad de agua disponible, mientras que es de septiembre a noviembre donde el embalse alcanza sus mínimos medios anuales (ANOVA: F=27.4, p<sub>α</sub><0.05) (figura II.2b).



**Figura II.1.** Precipitación y temperatura media correspondientes a las estaciones meteorológicas de los embalses de Sierra Boyera y Puente Nuevo para los años 1997, 1998, 1999 y 2000, y localización de la piscifactoría Hidrorecursos S.A. (E. 1:75.000) (Hoja 901 de la Cartografía Militar de España, 1981)



**Figura II.2.** a) Evolución de los volúmenes medios anuales almacenados en el embalse de Puente Nuevo junto con la variación diaria real (línea discontinua). b) Valores promedio por meses almacenados en el embalse durante los años 1997, 1998, 1999 y 2000

### 2.2.2. Antecedentes

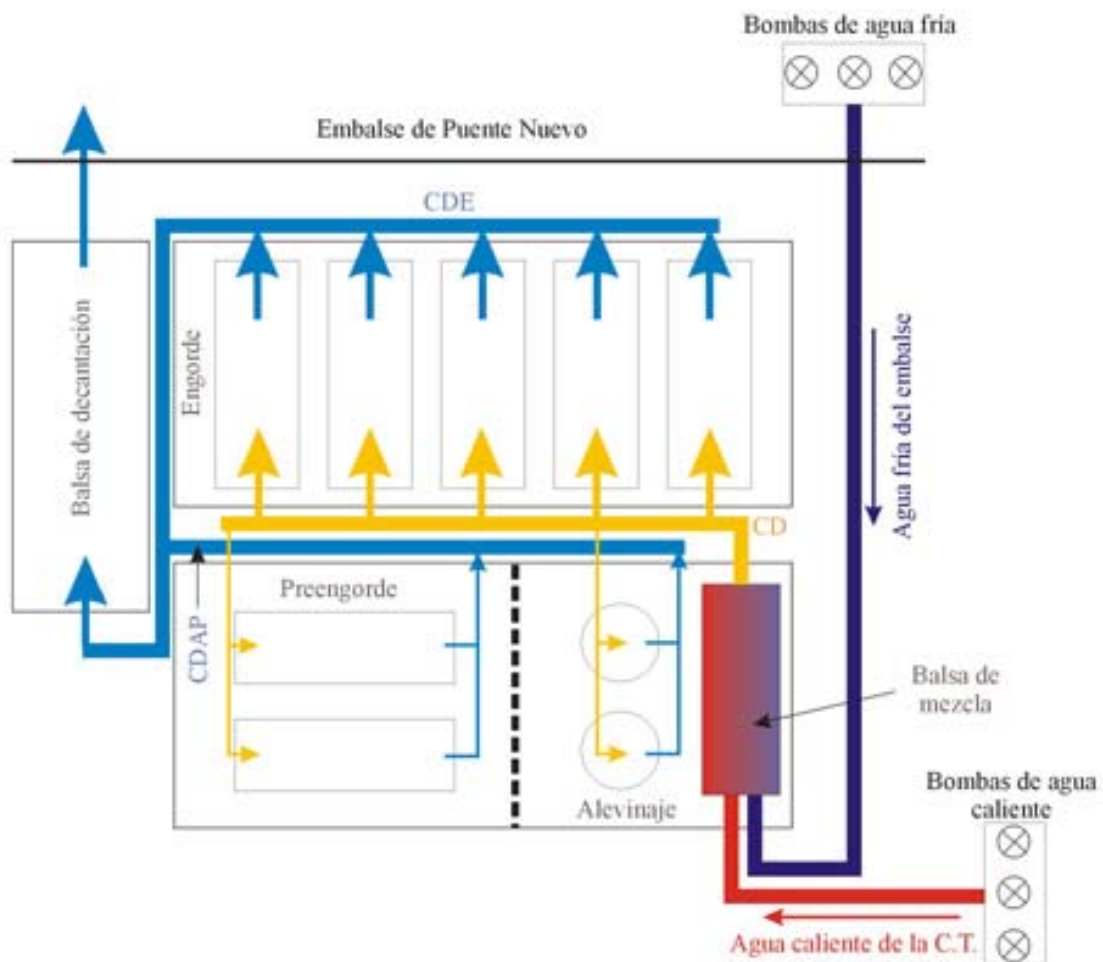
La oportunidad de aprovechar el agua caliente del sistema de refrigeración que la Central Térmica de Puente Nuevo vierte al embalse, fue el factor principal que decidió la inversión en la planta, ya que el agua de refrigeración de las centrales térmicas presenta importantes ventajas para los sistemas de acuicultura intensiva (Nash y Paulsen, 1981). Estas ventajas son básicamente el gran volumen de agua utilizado, la presión de salida, y las elevadas temperaturas que se alcanzan, permitiendo crecimientos muy rápidos de los organismos acuáticos. Por otra parte, los cambios que puede sufrir la calidad del agua como consecuencia de su paso a través del sistema de refrigeración de la central térmica, puede introducir potenciales desventajas que impidan la actividad acuícola. A pesar de ello, existen múltiples ejemplos de este tipo de sistemas (Veil, 1998).

Después de la visita a otros centros de producción con las mismas características y de una serie de estudios de viabilidad realizados a mediados de los ochenta, en los que se comprobó el estado del mercado europeo y nacional de la anguila (tabla II.1), se constituye en 1988 Hidrorecursos S.A., empresa dedicada al cultivo intensivo de anguilas europeas. Promovida inicialmente por Naviera Fierro S.A. y E.S. Brazatortas S.A., la construcción de las instalaciones comienza en septiembre de 1991 y finaliza en agosto de 1992.

**Tabla II.1.** *Distribución de las capturas y producciones anuales, consumos e importación de los principales países productores de anguila europea en 1985 (Más-Alvarez y Barrera-Orozco, 1989)*

<b>País</b>	<b>Captura y producción (tonelada/año)</b>	<b>Consumo (tonelada/año)</b>	<b>Importación</b>
Alemania Oriental	725	4000	82 %
Dinamarca	2120	3000	30 %
España	600	300	0 %
Francia	4000	-	-
Gran Bretaña	955	-	-
Holanda	1249	6500	81 %
Italia	2076	-	-
Polonia	1103	-	-
Suecia	1199	-	-
<b>Total</b>	<b>14027</b>		

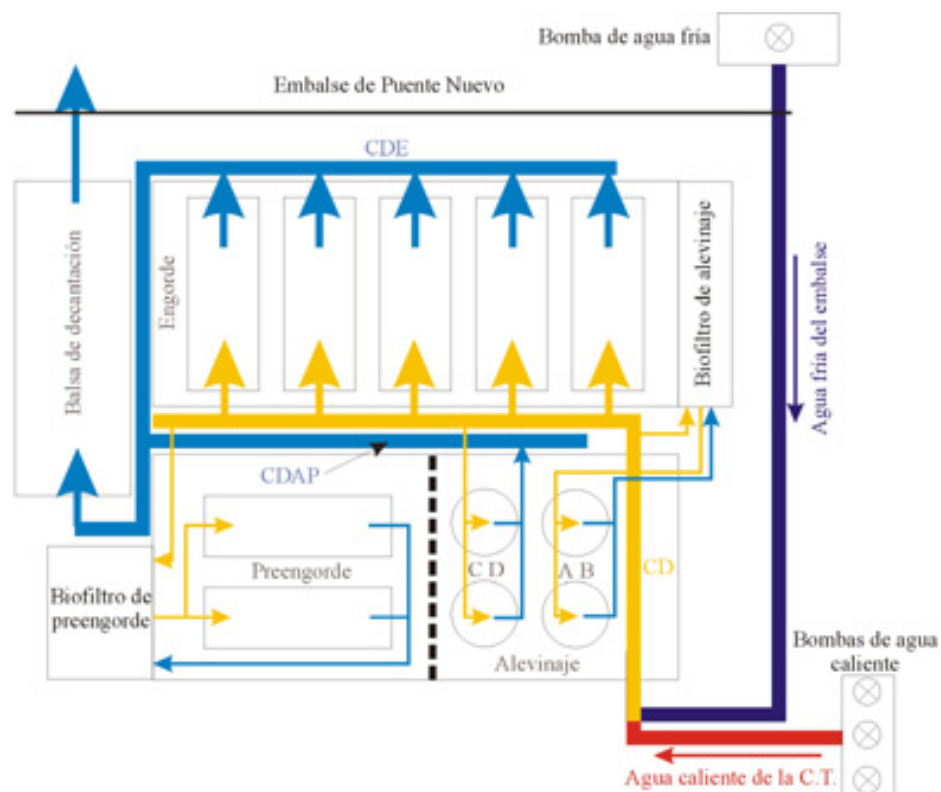
En el proyecto inicial dichas instalaciones constaban de dos naves adosadas de 1600 m<sup>2</sup> cada una, para una capacidad total de 96 tanques en los que se podía almacenar hasta 1204 m<sup>3</sup> de agua. Para permitir la captación continua de agua se instalaron seis bombas, con una potencia total de 600 kW, que introducían permanentemente 800 l/s y permitían la mezcla de agua fría procedente del embalse y agua caliente de la central térmica. El esquema básico de funcionamiento de la planta con la configuración inicial era la de unas instalaciones con circulación libre del agua, más conocido como circuito abierto. De esta forma, el agua que era bombeada desde el embalse y el canal de desagüe de la central térmica se mezclaba en una balsa, se repartía desde el canal de distribución principal, discurría a través de los tanques hacia el canal de desagüe y balsas de decantación de la planta y de allí pasaba nuevamente al embalse (figura II.3).



**Figura II.3.** Esquema de funcionamiento inicial de Hidrorecursos S.A (circuito abierto). El agua fría procedente del embalse se mezclaba con el agua caliente de la central térmica (C.T.) en una balsa de mezcla. De esta forma se conseguía una temperatura próxima al óptimo fisiológico de la anguila europea. Desde la balsa de mezcla, el agua se repartía a través del canal de distribución (CD) a la zona de alevinaje, preengorde y engorde. Una vez que pasaba por los tanques se recogía por medio de dos canales de desagüe (alevinaje y preengorde: CDAP y engorde: CDE). Ambos canales terminaban en la balsa de decantación y desde allí pasaba nuevamente al embalse

### 2.2.3. Situación actual

En la actualidad, Hidrorecursos S.A. ha pasado a ser propiedad de tres empresas: Promotora de Minas de Carbón S.A., ACS S.A. y Naviera Fierro S.A. También en las propias instalaciones se han llevado a cabo importantes modificaciones. Las más significativas han sido, por una parte, la transformación de gran parte de la planta (la mitad de la zona de alevinaje y preengorde en su totalidad) en un sistema de recirculación de agua o sistema cerrado (figura II.4), y por otra, la consecuente variación del sistema de bombeo. Estos cambios supusieron la construcción entre los años 1996 y 1997, por parte de la empresa danesa Billund, de tres módulos de filtración biológica mediante los cuales gran parte del agua bombeada podía ser reutilizada nuevamente en la planta. La entrada en funcionamiento de los nuevos módulos tuvo dos consecuencias inmediatas: una reducción de los costes energéticos, ya que desde ese momento era necesario bombear un menor volumen de agua, y un mantenimiento más regular de la temperatura del agua en aquellas zonas de la piscifactoría donde ésta se recirculaba.



**Figura II.4.** Esquema de funcionamiento actual de Hidrorecursos S.A. (circuito mixto). El agua fría procedente del embalse e impulsada ahora por tan solo una bomba, se mezcla directamente en la tubería con la caliente procedente de la central térmica (C.T.). Este agua se reparte a través del canal de distribución a todos los tanques que aún permanecen en régimen de circuito abierto (series 'C' y 'D' de alevinaje y engorde), pasando a través de ellos y llegando a los canales de desagüe (CDAP y CDE). De la misma forma, el agua de renovación del circuito cerrado se obtiene del mismo canal de distribución. En las series 'A' y 'B' de alevinaje y en preengorde el agua es depurada en los biofiltros, reintroduciéndose nuevamente en los tanques

### 2.2.3.1. Zona de alevinaje

La zona de alevinaje ocupa la mitad de la primera de las dos naves principales con las que cuenta la piscifactoría. Esta primera nave se caracteriza por encontrarse completamente cerrada. Consta de 48 tanques autolimpiables, cilíndricos con el fondo trococónico y fabricados en fibra de vidrio. Con un diámetro de 2.9 metros y una altura de 1.5 metros, pueden contener hasta 3.2 m<sup>3</sup> de agua (figura II.5). Todos los tanques se encuentran equipados por un comedero de reloj, un descansadero de rejilla, una sonda de nivel de agua, una sonda de nivel de oxígeno y un oxigenador de emergencia, constituido por una tubería porosa sumergida conectada directamente a un tanque de oxígeno líquido ubicado en el exterior de la planta.

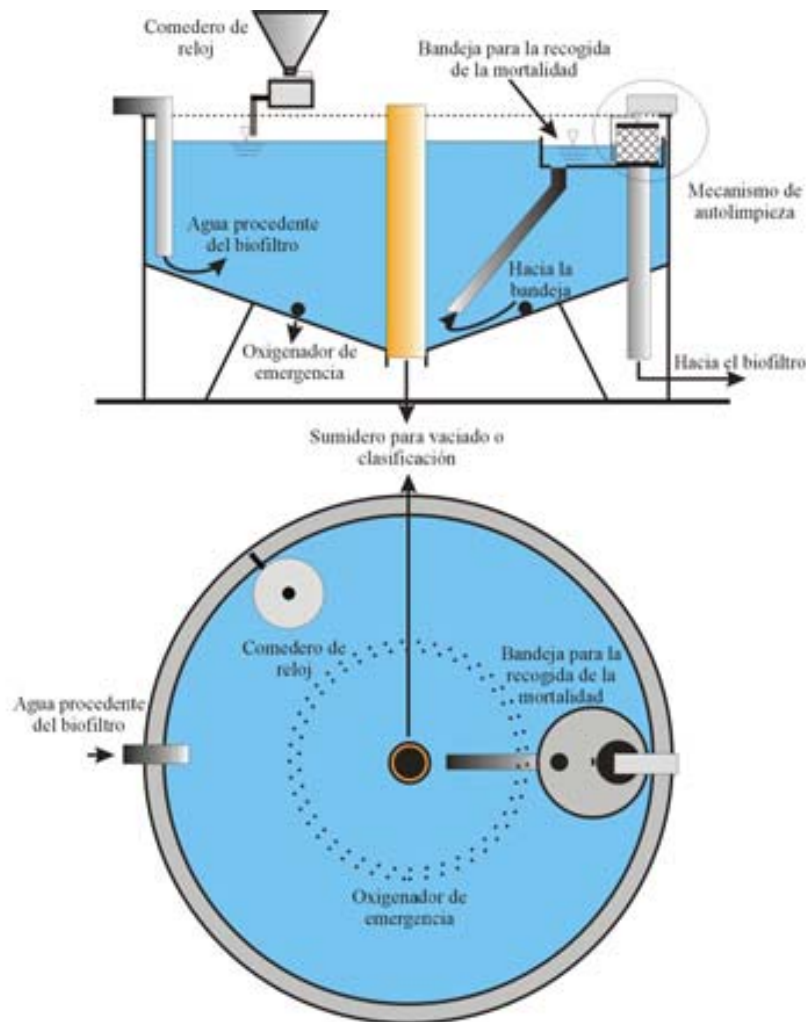
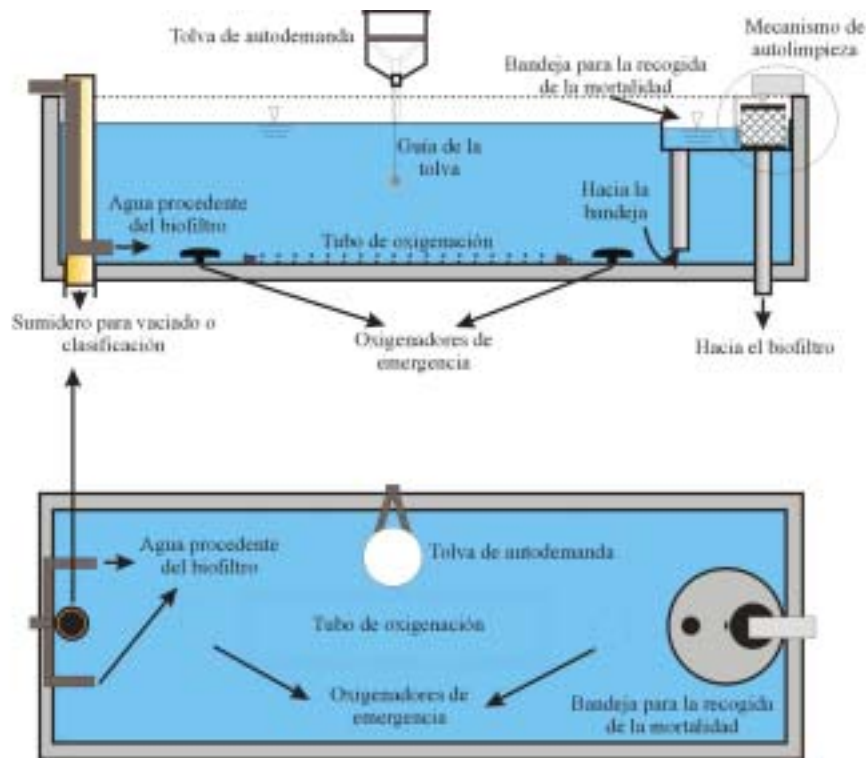


Figura II.5. Vista lateral y cenital de los tanques de fibra de vidrio de la zona de alevinaje

Los 48 tanques se dividen en series de 12 (series 'A', 'B', 'C' y 'D') con condiciones físico-químicas que se mantienen relativamente constantes. Las dos primeras series ('A' y 'B') se encuentran en régimen de circuito cerrado, mientras que las series 'C' y 'D' están en circuito abierto.

### 2.2.3.2. Zona de preengorde

El área de preengorde ocupa la segunda mitad de la primera nave. Está formada por 20 tanques de hormigón, de formato rectangular. Se encuentran divididos en dos series de 10 tanques cada una (series 'E' y 'F') con dos dimensiones básicas diferentes: 8 tanques de 8 metros de largo por 2 metros de ancho y 2 tanques de 8 metros de largo por 4 metros de ancho (en todos los casos la altura es de 1 metro). De forma individual todos se encuentran equipados con un comedero de autodemanda tipo tolva, una sonda de oxígeno, una sonda de nivel de agua, un oxigenador compuesto por una tubería porosa sumergida y dos oxigenadores de emergencia en forma de disco (figura II.6). En esta serie todos los tanques se encuentran en régimen de circuito cerrado.

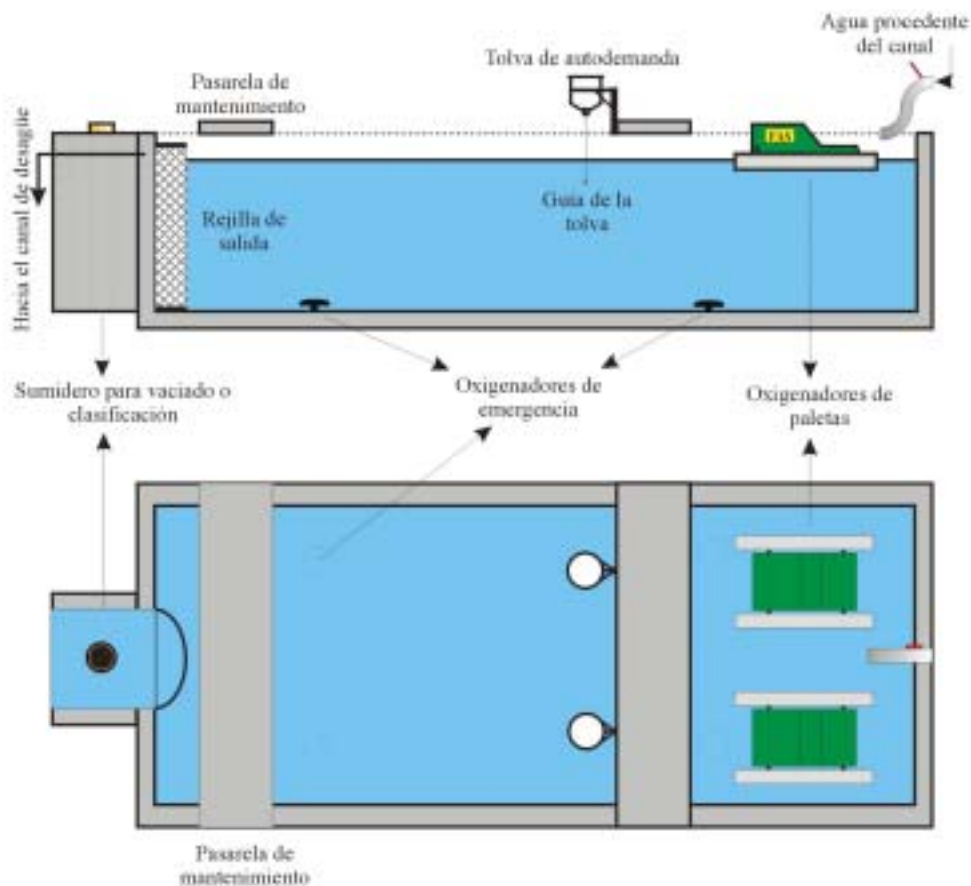


**Figura II.6.** Vista lateral y cenital de los tanques de hormigón de la zona de preengorde. Además de las diferencias estructurales con los tanques de la zona de alevinaje destaca el método de distribución de alimento. Las tolvas de autodemanda distribuyen el alimento en función del apetito de los individuos



### 2.2.3.3. Zona de engorde

El área de engorde ocupa en su totalidad la segunda nave de la planta. Al contrario de lo ocurre con la anterior, ésta se encuentra abierta, es decir, se puede considerar que más que una nave es un techo protector. Aquí encontramos un total de 14 tanques rectangulares de hormigón de 16 metros de largo por 4.5 metros de ancho y 1.6 metros de alto, que en su conjunto forman la serie 'G'. En cada tanque nos encontramos 2 comederos tipo tolva de autodemanda, dos oxigenadores flotantes de paletas FAS® y cuatro de discos sumergidos con sus respectivos reguladores, una sonda de nivel de agua y una sonda de oxígeno (figura II.7). La serie 'G' constituye en la actualidad la única área que en su totalidad se encuentra en régimen de circuito abierto.

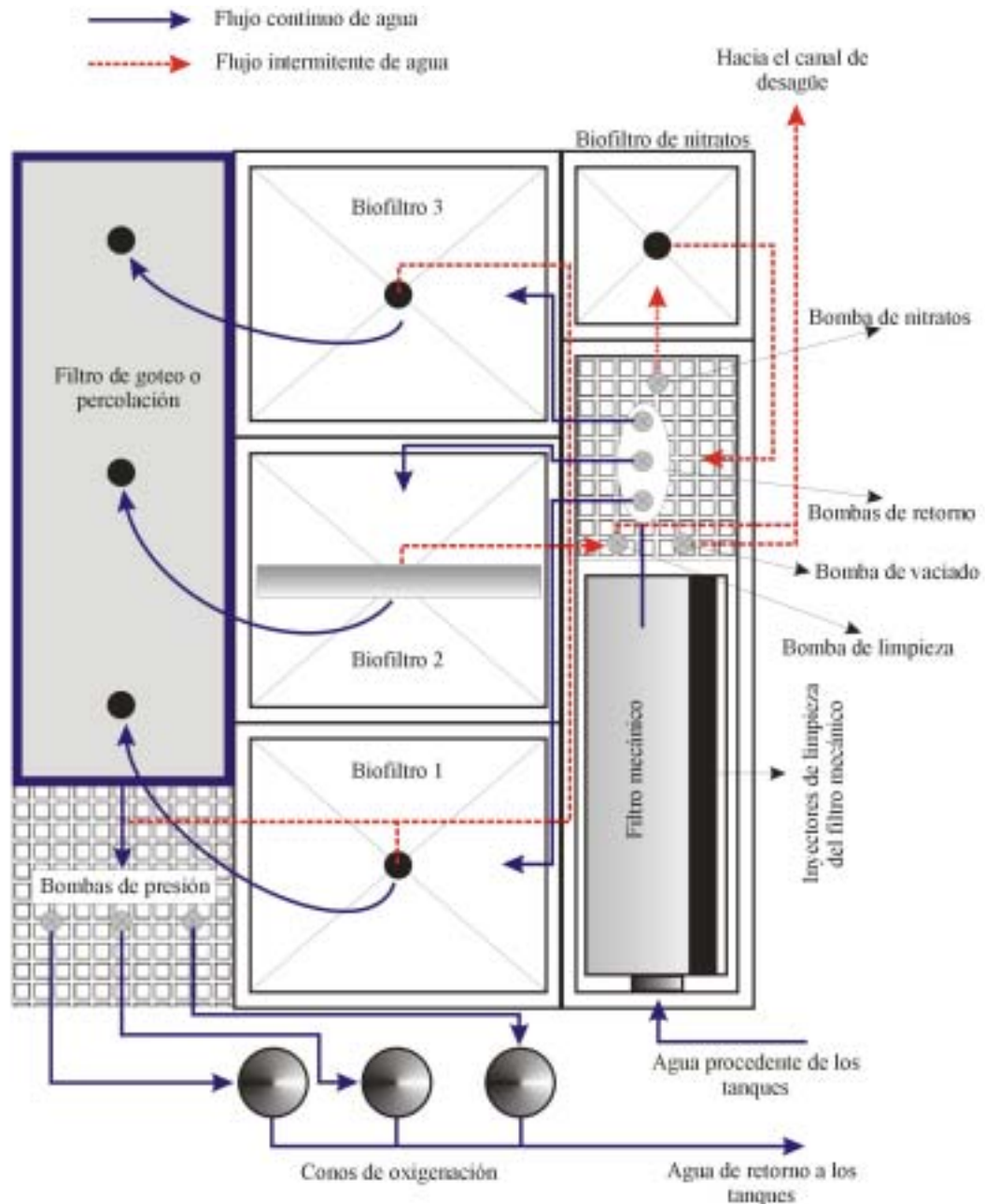


**Figura II.7.** Vista lateral y cenital de los tanques de hormigón de la zona de engorde. Debido a que en la zona de engorde todos los tanques se encuentran en régimen de circuito abierto, el agua procede directamente del canal de distribución principal y se evacua a través del canal de desagüe de la zona de engorde

#### 2.2.3.4. Biofiltros

Como ya se ha comentado en secciones anteriores (2.3.2.1 y 2.3.2.2) parte de la planta se encuentra en régimen de circuito cerrado. Para hacer esto posible es necesario depurar el agua e introducirla nuevamente en el sistema. Dicho proceso de depuración se lleva a cabo en los módulos de filtración o biofiltros. De forma general, se puede decir que Hidrorecursos cuenta con tres módulos de filtrado. Los dos primeros se encargan de depurar el agua procedente de las series 'A' y 'B', respectivamente, y el tercero trata el agua de toda la zona de preengorde (series 'E' y 'F') (figura II.4). Desde un punto de vista funcional ambos módulos constan de los mismos elementos, aunque el que se encuentra asociado a la zona de preengorde es de mayores dimensiones ya que se encarga de tratar un mayor volumen de agua. Cada módulo consta de los siguientes elementos (figura II.8):

- a) Un filtro mecánico de tambor rotatorio autolimpiante Hydrotech<sup>®</sup> con una malla de entre 40 y 60 micras. Con este filtro se extraen del agua todas las partículas superiores a 40-60 micras tales como restos de heces, restos de anguilas muertas, pienso no consumido, etc.
- b) Un aireador de esponjas sumergido bajo el filtro mecánico cuya misión es oxigenar ligeramente el agua.
- c) Primera estación de bombeo. Está formada por:
  - c.1) Pozo de bombas y poceta para el depósito de lodos.
  - c.2) Bomba de vaciado. Esta bomba toma el agua de la poceta de lodos y la expulsa a través de una tubería sumergida hasta el canal de desagüe. Su misión es retirar una cantidad fija diaria de agua del sistema (aproximadamente un 10 %) que es sustituida por agua procedente del canal de distribución. Esto supone que el agua del sistema cerrado es renovada completamente cada diez días. La puesta en marcha de esta bomba se realiza de forma automática mediante un temporizador.
  - c.3) Bomba de limpieza. A ella se conectan a través de tuberías subterráneas los filtros aerobios, el filtro de goteo o percolación y los depósitos de las bombas de presión. Tiene su salida en el canal de desagüe y su misión, como su nombre indica, es la de permitir extraer toda el agua del sistema para seguidamente proceder a su limpieza.
  - c.4) Bombas de retorno. Son las bombas encargadas de elevar el agua desde la poceta de aspiración hasta el interior de los filtros aerobios. Su número varía entre dos y tres dependiendo de la zona donde nos encontremos (dos en



**Figura II.8.** Esquema del biofiltro de la zona de preengorde. Desde un punto de vista funcional el biofiltro de la zona de alevinaje es idéntico, variando el tamaño de la estructura debido al menor volumen de filtrado. El proceso de filtración biológica comienza con la llegada del agua procedente de los tanques de cultivo hasta el filtro mecánico. Desde allí el agua es bombeada hasta los filtros aerobios o parte de ella puede ser elevada hasta el filtro anaerobio si es necesario, volviendo en este último caso a la poza de aspiración de las bombas de retorno. Una vez que pasa por los filtros aerobios llega hasta el filtro de goteo o percolación, donde se libera de gran parte del  $\text{CO}_2$  que lleva disuelto. Desde aquí, se acumula en la poza de aspiración de las bombas de presión donde es bombeada hasta los conos de oxigenación. Finalmente y una vez oxigenada vuelve a los tanques de cultivo

alevinaje por módulo y tres en preengorde). En cualquier caso encontramos una para cada filtro. Su parada y puesta en marcha se lleva a cabo de forma automática gracias a la presencia de una sonda de nivel bajo y otra de nivel alto.

c.5) Bomba de nitratos. Esta bomba se encarga de llevar parte del agua desde la poceta de aspiración hasta el filtro de nitratos o anaerobio. Este funciona de forma ocasional cuando la concentración de nitratos en agua es excesivamente elevada.

- d) Biofiltros sumergidos o aerobios. Encontramos dos de estos biofiltros en la zona de alevinaje (por módulo) y tres en la de preengorde. En ambos casos, cada biofiltro se encuentra formado por un tanque rectangular fabricado en hormigón con el fondo en forma de pirámide invertida. Desde el vértice de la pirámide parte una tubería en forma de 'U' que se comunica con el filtro de goteo o percolación. Junto a la tubería de comunicación encontramos cuatro aireadores de esponja que permiten la introducción de aire a través del fondo de la estructura. Todos los biofiltros excepto el número dos de la zona de preengorde, están rellenos de una malla consistente en tubos de **PVC** perforados y rugosos en su interior, empaquetados en forma de cubo (biomalla). De esta forma se consigue aumentar la superficie de contacto con el agua ( $150 \text{ m}^2/\text{m}^3$ ) y el sustrato sobre el que se asientan las bacterias aerobias encargadas de la oxidación del amonio. El rendimiento de este tipo de sistemas es relativamente bajo, alcanzando tasas de eliminación de amonio cercanas a  $0.12 \text{ g m}^{-2} \text{ día}^{-1}$ . Esto es debido a que la mayor parte de la transformación del amonio en nitritos y nitratos se produce en la parte superior del biofiltro donde el agua permanece oxigenada.

El biofiltro número dos de la zona de preengorde presenta una configuración distinta. Éste se encuentra relleno de partículas de unos 6 milímetros de diámetro de polipropileno granulado comúnmente denominado **SAMM** (*Suspended Aerated Mobile Medium*) que tiene la propiedad de presentar la misma densidad que el agua y la capacidad de proporcionar una superficie de soporte bacteriano mucho mayor que el obtenido gracias a las biomallas ( $300 \text{ m}^2/\text{m}^3$ ) (Visser van Bloemen, 1999). También encontramos una tubería en forma de 'T' con la parte superior perforada, permitiéndose de esta forma el paso de agua desde el biofiltro al filtro de percolación, y un mayor número de aireadores (ocho en total) gracias a los cuales se provocan células de convección que arrastran a las partículas (**SAMM**). Con este sistema se aumenta en gran medida el rendimiento y la eficacia en el manejo de la instalación, ya que debido a la circulación de las partículas por toda la estructura del tanque se obtienen dos efectos positivos: un proceso de autolimpieza como consecuencia del roce de las partículas (**SAMM**) con las paredes del tanque, y un aumento de la tasa de oxidación de amonio que alcanza niveles de  $0.63 \text{ g m}^{-2} \text{ día}^{-1}$  (Visser van Bloemen, 1999).

- e) Filtro de nitratos o anaerobio. Formado por un depósito rectangular con el fondo en forma de pirámide invertida y construido en hormigón, tiene como misión la eliminación de nitratos cuando su concentración se hace tóxica para los peces. Al igual que en el caso de los filtros aerobios, se encuentra relleno de biomalla ( $150 \text{ m}^2/\text{m}^3$ ). Sin embargo, al contrario de lo que sucede en éstos, la mayor eficacia de transformación de los nitratos en otras sustancia nitrogenadas por las bacterias se encuentra en la parte más profunda del filtro, donde el agua presenta una baja concentración de oxígeno o es totalmente anóxica.
- f) Filtro de goteo o percolación. Está formado por un tanque fabricado en nailon impermeabilizado sujeto por una estructura de acero galvanizado. Ya que los filtros de goteo se encuentran a una altura sensiblemente inferior a la que se encuentran los filtros aerobios y a que ambos están comunicados a través de un tubo en forma de 'U', el agua pasa de uno a otro por efecto de vaso comunicante. Toda la estructura está llena de las mismas biomallas que se encuentran en el interior del resto de los biofiltros pero con una relación superficie-volumen de  $200 \text{ m}^2/\text{m}^3$ . A modo de tapadera encontramos una plancha de **PVC** perforada a través de la cual fluye el agua. Esto tiene como objetivo la caída del agua en forma de gotas, liberándose así de gran parte del  $\text{CO}_2$  que contiene debido a la acción bacteriana. Unos ventiladores que extraen continuamente aire del interior del filtro de goteo favorecen el intercambio gaseoso, impidiendo que la atmósfera interior del filtro se sature de  $\text{CO}_2$ .
- g) Segunda estación de bombeo. Está formada por:
- g.1) Depósito de aspiración de las bombas. Éste se encuentra situado justo debajo del filtro de goteo.
  - g.2.) Bombas de presión. Son las bombas encargadas de devolver el agua ya depurada hasta los tanques de cultivo. Son dos o tres dependiendo de la zona (dos por módulo en alevinaje y tres en preengorde).
- h) Conos de oxigenación. Permiten la oxigenación óptima del agua. Cuando el agua es bombeada desde la balsa de aspiración de las bombas de presión, presenta una concentración de oxígeno disuelto de entre 1 y 2 mg/l. Cuando llega a los conos de oxigenación se produce por una parte un aumento de presión y por otra una inyección directa de oxígeno. De esta forma cuando el agua sale de los conos de oxigenación presenta una concentración de oxígeno disuelto cercana a la supersaturación (18-20 mg/l). La disminución de presión en la entrada a los tanques de cultivo libera oxígeno dejando la concentración en el agua cercana a los 7-9 mg/l.

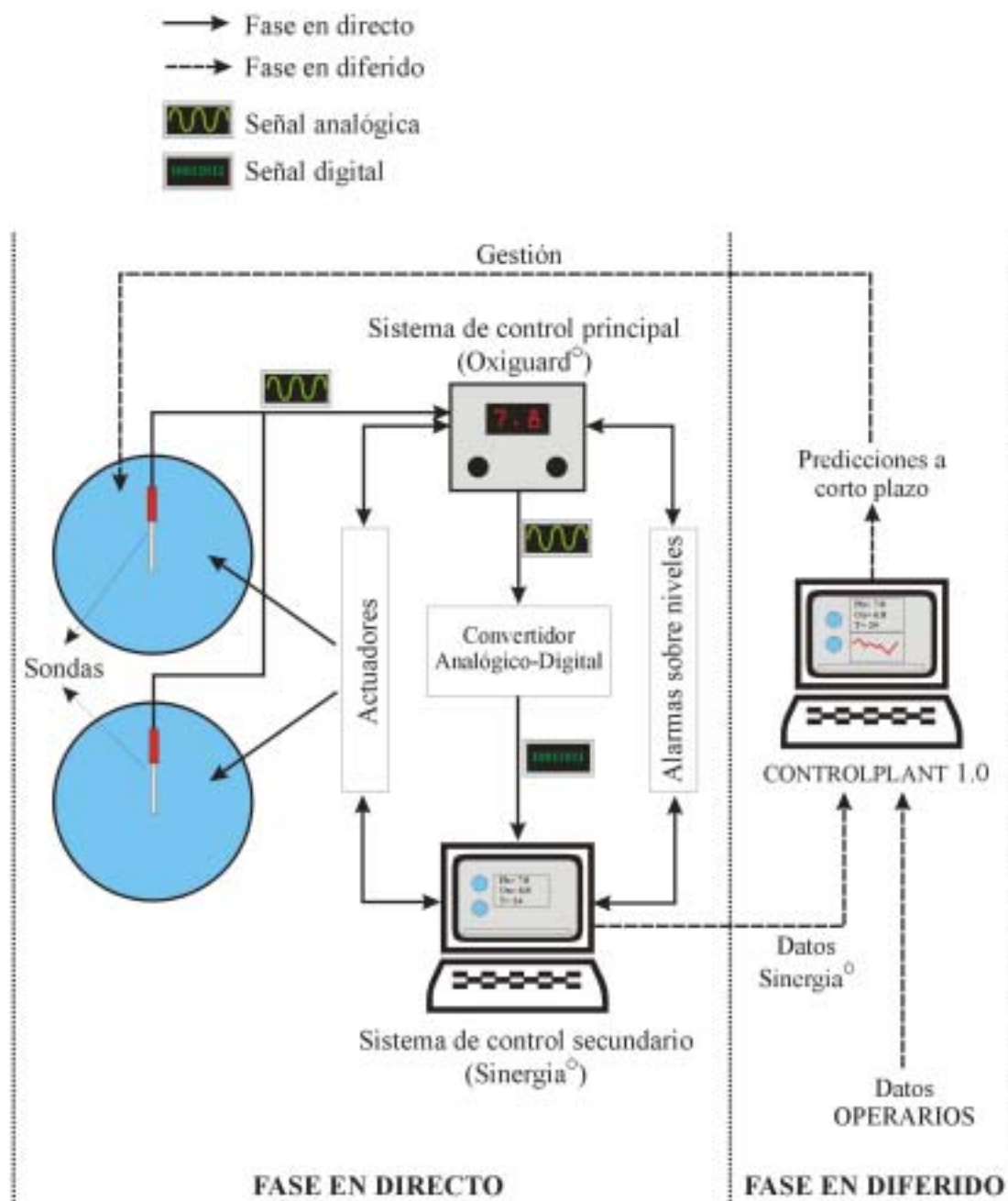
#### 2.2.3.5. Sistemas de control

Hidrorecursos mantiene en sus instalaciones actuales un alto grado de automatización, lo que se debe fundamentalmente a que los procesos de filtración biológica se llevan a cabo bajo unas determinadas condiciones físico-químicas cuyo margen de variación es muy pequeño. Esto obliga a tener bajo un control continuo variables como la concentración de oxígeno disuelto (mg/l), temperatura (°C) o pH. Para comprobar los valores de estas variables existen repartidas, tanto por los tanques de cultivo como en los propios filtros, sondas que permiten hacer una medición automática. Las mediciones son recogidas en tiempo real por un ordenador central donde se almacenan y se establecen las medidas de emergencia. El control sobre el ordenador central se lleva a cabo mediante dos sistemas automáticos diferentes: Oxyguard<sup>®</sup> y Sinergia<sup>®</sup>. Para complementar estos dos sistemas de control en 'directo' se ha programado un tercero (**CONTROLPLANT 1.0**) que funciona en 'diferido' y que permite, además del manejo avanzado de las bases de datos creadas a partir de los dos sistemas anteriores y los operarios de la planta, un análisis de los datos que proporciona predicciones sobre el comportamiento del sistema a corto y medio plazo (figura II.9).

El primero de ellos (Oxyguard<sup>®</sup>) es el sistema de control principal. Está fabricado por la empresa danesa Billund y se compone de un cuadro de control a nivel de tanque que recibe señales de tipo analógico. Esto tiene dos implicaciones importantes: el sistema es muy robusto y, es poco intuitivo. Esto es debido a que está basado en *Hardware*. El sistema (Sinergia<sup>®</sup>) o sistema de control secundario está fabricado por la empresa italiana Sinergia. Éste está basado en *Software* que corre bajo el sistema operativo Windows 95<sup>®</sup>. Mediante este sistema se pueden controlar los niveles de las variables y almacenar los datos para un posterior análisis. Tiene además la ventaja de que toda la planta puede ser identificada en la pantalla del monitor, asociándose gamas de colores con distintos niveles de las variables. Sin embargo tiene dos graves inconvenientes: dependencia de las actualizaciones por falta de depuración del programa y, dependencia de un sistema operativo que puede presentar problemas de estabilidad.

### 2.3. EL PROGRAMA DE CONTROL EN DIFERIDO CONTROLPLANT 1.0

**CONTROLPLANT 1.0** tiene como objetivo facilitar el almacenamiento de todos los datos procedentes del sistema de control secundario (Sinergia<sup>®</sup>) y de los obtenidos por los operarios a lo largo de la jornada de trabajo. Esta información va a ser codificada en unas bases de datos que van a conformar la base de Conocimiento del Dominio Secundaria (**CDS**). Ésta va a ser empleada como fuente numérica para la elaboración de los modelos de gestión y valoración de los diagnósticos realizados a través del Sistema Experto de



**Figura II.9.** Esquema de funcionamiento de los sistemas de control de Hidrorecursos. Oxiguard<sup>®</sup> es el sistema de control principal. Sobre éste se apoya el secundario (Sinergia<sup>®</sup>). Ambos sistemas se engloban en la fase de control en 'directo', de tal forma que una variación de los niveles normales de alguna de las variables puede ser modificada mediante la acción de los actuadores (p.e.: un descenso de la cantidad de oxígeno disuelto hace saltar una alarma y conecta automáticamente los oxigenadores de emergencia). En la fase en 'diferido' los datos almacenados por Sinergia<sup>®</sup> pueden ser utilizados junto con los introducidos por los operarios para la elaboración de modelos de predicción para una gestión más eficiente de la planta

Diagnóstico Patológico para Anguilas (**SEDPA**). Cuenta además con un módulo de simulación denominado Programa Operativo de Simulación (**POS**) que utiliza los modelos obtenidos durante el estudio gracias al cual es posible obtener patrones de comportamiento de la planta a nivel de producción.

Para facilitar la entrada de datos y la interpretación de los mismos por parte de los operarios de la planta, se ha dotado a **CONTROLPLANT** de un aspecto similar a la pantalla principal del sistema de control Sinergia<sup>®</sup>, con el que los operarios ya se encuentra familiarizados. De la misma forma que en Sinergia<sup>®</sup>, se ha asociado la visualización de los parámetros controlados por el programa con una gama de colores fácilmente reconocibles, reproduciéndose fielmente la estructura de la propia planta en la pantalla.

### 2.3.1. Inicio del programa

Para poder acceder al programa el usuario tan solo tiene que hacer doble 'click' sobre el icono de acceso directo del escritorio de Windows<sup>®</sup> (figura II.10). El resultado de la acción es la aparición de una ventana de presentación donde se le pide al usuario una clave de acceso. Esta clave de acceso permite el registro de todos los usuarios que utilizan el programa, el tiempo de acceso y comprobar que tipo de datos ha introducido. Si la clave de acceso es aceptada por el programa, aparecerá la ventana principal de la aplicación (figura II.11).



**Figura II.10.** Icono de acceso a **CONTROLPLANT 1.0** desde el escritorio de Windows<sup>®</sup>, y ventana de acceso y verificación de contraseña

### 2.3.2. Ventana principal de CONTROLPLANT

En la ventana principal de la aplicación **CONTROLPLANT** podemos distinguir tres partes bien diferenciadas: a) una barra de menú donde tenemos las opciones *Archivo*, *Ver*,



*Mantenimiento, SPI y Ayuda*, b) una sección superior que nos proporciona información preliminar básica y, c) una zona inferior donde se presenta un esquema de la totalidad de la planta.



**Figura II.11.** Ventana principal de CONTROLPLANT 1.0

### 2.3.2.1. Barra de herramientas

La primera opción que nos encontramos en la barra de herramientas es *Archivo*. Si se hace 'click' en esta opción se despliega un menú donde se puede optar entre: *Nueva incorporación de ejemplares...Ctrl+N*, *Abrir tanque (editar)*, *Cerrar-transferir...Ctrl+I*, *Imprimir...Ctrl+I* y *Salir...Ctrl+S*.

La primera de las opciones (*Nueva incorporación de ejemplares*) está siempre activa por defecto. Si la seleccionamos aparece una ventana (figura II.12) que nos permite incorporar a la base de datos una nueva entrada de ejemplares a la planta. Para ello se le pregunta al usuario el tanque en el que van a quedar depositados los individuos, un nombre identificativo del lote, el peso total (kg) y peso medio (g/unidad). Seguidamente y mediante el botón *OK*, los datos pasan de forma provisional a la matriz situada en la parte inferior de la ventana. Este proceso se puede repetir cuantas veces sea necesario. Cuando se haya finalizado y mediante el botón *Fin* los datos pasan definitivamente a formar parte de la base

de datos. Los efectos de los cambios en las bases de datos se refleja inmediatamente sobre la pantalla principal del programa.



**Figura II.12.** Ventana para la incorporación de nuevos ejemplares

La opción *Abrir tanque (editar)* permite ver todas las características del tanque seleccionado. Éstas se muestran en una nueva ventana donde se pueden distinguir dos zonas: a) la zona izquierda de la ventana donde se distinguen un total de cinco bloques (*Tanque, Biometría, Mortalidad, Alimentación y P. Físico-químicos*) y, b) la zona derecha donde se puede observar de arriba hacia abajo una representación tridimensional del tipo de tanque en el que nos encontramos junto a una ventana de hipertexto, que nos transmite informes sobre determinados acontecimientos importantes que hayan sucedido en el tanque desde que este se encuentra habilitado (tratamientos, patologías, etc.), una matriz donde se indican todos los movimientos de individuos que se han llevado a cabo en dicho tanque (fecha de llegada, origen, salida, destino, peso total o biomasa y peso medio), y un cuadro de gráficos donde se puede ver la evolución que han sufrido parámetros tales como la mortalidad, la cantidad de pienso suministrado, temperatura, oxígeno, pH, amoníaco, nitritos, nitratos y caudal (figura II.13).

*Cerrar-transferir* es la opción mediante la cual se actualizan las bases de datos cuando un tanque se vacía completamente o cuando parte de los individuos se trasladan a otro tanque. Al hacer 'click' en esta opción aparece una nueva ventana (figura II.14) en la que se puede distinguir un primer bloque donde se pide al usuario el tanque que se quiere modificar (cerrar o transferir parte de su carga). Sólo será posible seleccionar aquellos tanques que en ese momento se encuentren habilitados. En un segundo bloque el usuario debe seleccionar *Cerrar* o *Transferir*. Finalmente, un tercer bloque pide al usuario el destino de los individuos (otro tanque o venta), el peso total de todos los individuos

transferidos, y el peso medio de esos individuos. Cuando los datos son introducidos, se habilita un botón que presenta una flecha hacia la derecha. Al presionar dicho botón los datos introducidos pasan a la sección de confirmación para que el usuario compruebe que los datos son correctos. Una vez confirmados, mediante el botón *Aceptar*, los datos introducidos por el usuario pasan a la malla que se encuentra en la parte inferior de la ventana. El proceso se puede repetir tantas veces como sea necesario. Para modificar definitivamente las bases de datos es necesario presionar el botón *Fin*.



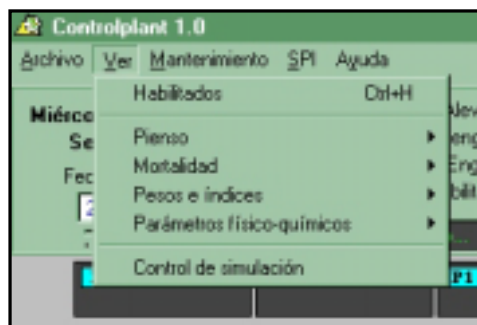
**Figura II.13.** Ventana que se visualiza cuando se hace 'click' sobre la opción Abrir tanque (editar)



**Figura II.14.** Ventana Cerrar-transferir

Las dos últimas opciones que aparecen cuando se selecciona *Archivo* son las típicas de cualquier aplicación en entorno Windows®. Éstas son por una parte *Imprimir*, la cual permite la obtención de informes sobre todas las actividades llevadas a cabo durante la sesión, y por otra, la opción *Salir* que cierra definitivamente la aplicación. Cuando se selecciona esta última opción, el programa pregunta al usuario si desea que **CONTROLPLANT** compruebe todo el sistema de ficheros y programas de operaciones de la aplicación. Si se elige *Si*, se comprueba que todos los archivos y matrices se encuentran cerradas, y que las matrices temporales han descargado la información sobre los ficheros definitivos. Si se opta por *No* se sale de la aplicación.

La segunda opción de la barra de herramientas (*Ver*) permite al usuario visualizar todos los parámetros físico-químicos introducidos. Para ello, dependiendo del tipo de dato se visualiza en cada tanque el valor o el color asociado correspondiente (figura II.15).



**Figura II.15.** Opción *Ver* de la barra de herramientas

De esta forma si seleccionamos *Pienso*  $\rightarrow$  *kg por día* aparecerá la cantidad suministrada en cada tanque para el día escogido. Si se elige *Parámetros físico-químicos*  $\rightarrow$  *Series A, B, E y F*  $\rightarrow$  *Temperatura*, los tanques correspondientes a dichas series adquirirán una coloración relacionada con la temperatura que para la fecha escogida presentaban (figura II.16). La primera opción del menú *Ver* es *Habilitados*. Esta opción permite volver al estado por defecto del programa, en la cual se presentan los tanques que se encuentran llenos (blancos) y los que se encuentran vacíos (negros). La última opción del menú es *Control de simulación*. Ésta hace posible la visualización de los controles mediante los cuales da comienzo y se detiene la simulación cuando se controla desde **CONTROLPLANT**.

La tercera opción de la barra de herramientas es *Mantenimiento*. Mediante ésta el operario de planta puede introducir todos los datos que habitualmente se toman en la piscifactoría y que hasta el momento del desarrollo de la aplicación se registraban en papel. Los datos son tanto parámetros físico-químicos por series, como variables biológicas

(alimentación, mortalidad). También a través de este menú se permite la impresión de los partes de alimentación que siguen los operarios para distribuir diariamente la cantidad de pienso correspondiente a cada tanque, e informes sobre las patologías y tratamientos específicos para cada tanque o serie (figura II.17).



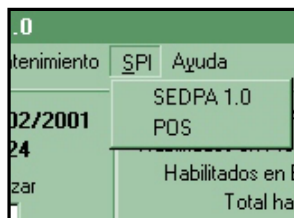
**Figura II.16.** Ventana principal en la que se muestra la temperatura y gamas de colores asociados para las series 'A', 'B', 'E' y 'F' en una fecha determinada



**Figura II.17.** Menú de mantenimiento

**SPI** son las siglas de Sistema de Predicción Integrado. Al hacer 'click' en esta opción se despliega un menú donde tan solo es posible escoger entre **SEDPA 1.0** y **POS**. **SEDPA** o Sistema Experto de Diagnostico Patológico para Anguilas es un programa que permite el diagnóstico de posibles patologías en la planta (se describe en el capítulo 3). Por otra parte, **POS** o Programa Operativo de Simulación (capítulo 5) permite integrar todos los modelos generados a partir de los datos suministrados por Sinergia® y los operarios. Agrupa a **SEDPA** y todos los modelos estadísticos y heurísticos (capítulo 4) (figura II.18).

Finalmente, la última opción de la barra de herramientas (*Ayuda*) despliega un pequeño menú que permite al usuario obtener información respecto al funcionamiento de la aplicación desarrollada. La opción *Acerca de...* muestra una ventana con información general del programa: nombre, autor, organismo donde se ha desarrollado la aplicación, versión y año (figura II.19).



**Figura II.18.** Menú del Sistema de Predicción Integrado (SPI)



**Figura II.19.** Ventana *Acerca de...* donde se muestra información adicional sobre el programa

### 2.3.2.2. Sección superior de la ventana principal

En la sección superior de la ventana principal encontramos varios bloques claramente diferenciados. El primero de ellos es un bloque temporal donde se informa al usuario del día, semana y fecha de visualización. Esta fecha de visualización se encuentra asociada a una flecha de desplazamiento que tiene un rango de actuación de dos meses. El usuario tiene la opción de utilizar la flecha de desplazamiento o bien introducir la fecha deseada. Hacia la derecha del bloque temporal encontramos una zona donde se suministra

información preliminar: habilitados en alevinaje, preengorde, engorde y total. Esta información varía en función del parámetro visualizado, de tal forma que si lo que se quiere es ver temperatura (o cualquier otra variable) para una fecha en concreto, además de la asociación de colores correspondientes aparecerá en esta misma zona el valor numérico.

Justo debajo de esta zona de información encontramos una ventana que va indicando las distintas zonas por las que se va desplazando con el ratón. Así, si el usuario coloca el ratón encima de cualquier tanque de la serie 'A', la ventana de mensajes indicará 'Serie A'. En la parte más derecha de esta sección encontramos una fotografía del operario que ese momento esté utilizando el programa. Al hacer 'click' con el botón derecho del ratón sobre la fotografía aparecerá un menú flotante que pregunta *¿Qué es esto?*. Si escogemos esta opción aparecerá una pequeña ventana con toda la información referente al usuario actual: número de accesos al programa, días de accesos y tipo de datos introducidos. También en esta zona, siempre y cuando se tenga activada la opción *Control de simulación*, podremos ver los controles de simulación. Estos consisten en dos botones relacionados con reproducción (una flecha dirigida hacia la derecha) y parada (un cuadrado negro).

### 2.3.2.3. Sección inferior de la ventana principal

En la sección inferior de la ventana principal se muestra una reproducción de todos los tanques de la planta, de tal forma que el operario puede tener una visión rápida del estado general de la instalación. Esto se ve reforzado, como ya se ha comentado en secciones anteriores, por la asociación de colores con los distintos niveles de las variables tratadas. Desde esta sección se puede además acceder rápidamente a las características concretas de cada tanque. Cada tanque tiene asociado en la parte superior una etiqueta identificativa. Cuando el tanque está habilitado, al colocar el ratón sobre la etiqueta aparece una pequeña lupa. Si en ese momento se hace 'click' nos aparecerá la ventana que se mostraba al seleccionar *Abrir tanque (editar)* de la opción de la barra de herramientas *Archivo* (figura II.13).

## 2.4. REFERENCIAS

Arias-García, A.M. y P. Drake-Moyano. 1990. *Estados juveniles de la ictiofauna en los caños de las salinas de la bahía de Cádiz*. Edita: Instituto de Ciencias Marinas de Andalucía. CSIC. Cádiz.

- Cabanás, R. 1975. Los afluentes del Guadalquivir por la derecha. Notas sobre morfología fluvial. *Estudios Geográficos*. Números 136-139.
- Carrillo, M. 1999. *Valoración alimentaria de la producción de anguilas en un sistema intensivo*. Trabajo Profesional Fin de Carrera. Escuela Técnica Superior de Ingenieros Agrónomos y de Montes. Universidad de Córdoba.
- Castelló, V. 1980. *Ictiología del río Guadiato. Estudio de la población de barbos (Barbus barbus)*. Tesis de Licenciatura. Universidad de Córdoba.
- Beullens, K., E.H. Eding, F. Ollevier, J. Komen y C.J.J. Richter. 1997. Sex differentiation, changes in length, weight and eye size before and after metamorphosis of European eel (*Anguilla anguilla* L.) maintained in captivity. *Aquaculture*, 153: 151-162.
- Brown, L. 2000. *Acuicultura para veterinarios. Producción y clínica de peces*. Edita: ACRIBIA, S.A. Zaragoza.
- Ferreras, M. 1980. *Los odonatos de Sierra Morena Central*. Tesis Doctoral. Facultad de Biología. Universidad de Sevilla.
- Herrera, M. 1991. *Estrategias en los ciclos de vida de una comunidad de ciprínidos en un arroyo de primer orden en la cuenca del río Guadalquivir*. Tesis Doctoral. Facultad de Ciencias Biológicas. Universidad de Córdoba.
- Gutiérrez-Estrada, J.C. 1997. *Estructura por edades, crecimiento y biología reproductiva del pez sol (Lepomis gibbosa, L. 1758) en el río Guadiato (Córdoba)*. Tesis de Licenciatura. Facultad de Ciencias Biológicas. Universidad de Córdoba.
- Gutiérrez-Estrada, J.C., I. Pulido-Calvo y C. Fernández-Delgado. 2000. Age-structure, growth and reproduction of the introduced Pumpkinseed (*Lepomis gibbosa*, Linnaeus 1758) in a tributary of the Guadalquivir River basin (southern Spain). *Limnetica*, 19: 21-29.
- Jover-Cerdá, M. 1993. Producción de anguilas. *Productos del Mar*, Junio: 78-84.
- Más-Alvarez, B. y R. Barrera-Orozco. 1989. *Producción de anguilas*. Hojas divulgadoras (Nº: 14/89 HD) del Ministerio de Agricultura Pesca y Alimentación. Madrid.
- Nash, C.E. y C.L. Paulsen. 1981. Water quality changes relevant to heated effluents and intensive aquaculture. En *Proceedings of the World Symposium on Aquaculture 'Heated Effluents and Recirculation System'*. Berlín (Alemania).



- Sola, C. y L. Tosi. 1993. Bile salt taurine as chemical stimuli for glass eels, *Anguilla anguilla*: a behavioural study. *Environmental Biology of Fishes*, 37: 197-204.
- Sola, C. 1995. Chemoattraction of upstream migrating glass eels *Anguilla anguilla* to earthy and green odorants. *Environmental Biology of Fishes*, 43: 179-185.
- Sola, C. y P. Tongiorgi. 1996. The effect of salinity on the chemotaxis of glass eels, *Anguilla anguilla*, to organic earthy and green odorants. *Environmental Biology of Fishes*, 47: 213-218.
- Tesch, F.W. 1977. *The eel biology and management of anguillid eels (Anguilla anguilla L.)*. Edita: Chapman and Hall. Londres.
- Tongiorgi, P., L. Tosi y M. Balsamo. 1986. Thermal preferences in upstream migrating glass-eels of *Anguilla anguilla* (L.). *Journal of Fish Biology*, 28: 501-510.
- Tosi, L., L. Sala, C. Sola, A. Spampanato y P. Tongiorgi. 1988. Experimental analysis of the thermal and salinity preferences of glass-eels, *Anguilla anguilla* (L.), before and during the upstream migration. *Journal of Fish Biology*, 33: 721-733.
- Tosi, L., A. Spampanato, C. Sola y P. Tongiorgi. 1990. Relation of water odour, salinity and temperature to ascent of glass-eels, *Anguilla anguilla* (L.): a laboratory study. *Journal of Fish Biology*, 36: 327-340.
- Tosi, L. y C. Sola. 1993. Role of Geosmin, a typical inland water odour, in guiding glass eel *Anguilla anguilla* (L.) migration. *Ethology*, 95: 177-185.
- Usui, A. 1978. *El cultivo de la anguila*. Edita: ACRIBIA S.A. Zaragoza.
- Veil, A.J. 1998. The potential for effluent trading in the energy industries. *Environmental Science and Policy*, 1: 39-49.
- Visser van Bloemen, P.J. 1999. Recirculation system for unit 'G'. Informe interno de Hidrorecursos S.A.

## CAPÍTULO 3

---

### SISTEMA EXPERTO DE DIAGNÓSTICO PATOLÓGICO PARA ANGUILAS (SEDPA)

#### 3.1. Introducción

- 3.1.1. Los sistemas expertos como métodos de diagnóstico
- 3.1.2. Formulación del problema del diseño de un sistema experto para diagnóstico clínico

#### 3.2. Descripción del sistema en estudio

#### 3.3. Agentes patológicos reconocidos por el sistema

- 3.3.1. Componentes del medio ambiente
- 3.3.2. Virus
- 3.3.3. Bacterias
- 3.3.4. Hongos
- 3.3.5. Parásitos

#### 3.4. Material y métodos

- 3.4.1. Bases de conocimiento del dominio
- 3.4.2. El motor de inferencia
  - 3.4.2.1. Análisis morfológico-sintáctico. Red de Transición Aumentada (**RTA**)
  - 3.4.2.2. El controlador borroso
    - 3.4.2.2.1. Lógica borrosa y teoría de conjuntos borrosos. Conjuntos borrosos y funciones de inclusión
    - 3.4.2.2.2. Reglas de inferencia borrosa y Memoria Asociativa Borrosa (**MAB**). Asignación de niveles de creencia individuales y globales
  - 3.4.2.3. Teoría de Dempster-Shafer para la transmisión de incertidumbre
  - 3.4.2.4. Test  $\chi^2$  modificado como elemento diferenciador de los grupos patológicos con mayor aportación a la creencia global
- 3.4.3. Subsistema explicativo
- 3.4.4. Subsistema de sugerencia o motor proposicional
- 3.4.5. Subsistema de aprendizaje
- 3.4.6. Valoración del diagnóstico
- 3.4.7. Validación del sistema

#### 3.5. Resultados

- 3.5.1. Información contenida en la bases de datos del dominio principal y secundario
  - 3.5.2. Número de casos utilizados para la evaluación
  - 3.5.3. Análisis de las respuestas aportadas por los expertos humanos
  - 3.5.4. Análisis preliminar de las Memorias Asociativas Borrosas
  - 3.5.5. Análisis de las respuestas de **SEDPA**
  - 3.5.6. Agrupación de diagnósticos individuales de **SEDPA**
  - 3.5.7. Diagnóstico de los expertos frente al diagnóstico de **SEDPA**
  - 3.5.8. Selección de la mejor configuración de **SEDPA**
  - 3.5.9. Análisis de las respuestas de **SEDPA** con reducción de información
  - 3.5.10. Respuesta del subsistema sugerencia o motor proposicional
- 3.6. Discusión
- 3.6.1. Respuestas de los expertos humanos
  - 3.6.2. Lenguaje natural
  - 3.6.3. Respuestas de **SEDPA**: creencias y valoraciones. Conclusiones del motor de inferencia y proposicional
  - 3.6.4. Expertos humanos frente a **SEDPA**
- 3.7. Referencias
- 3.8. ANEXO III.1. Glosario
- 3.9. ANEXO III.2. Manual del usuario

## **CAPÍTULO 3. SISTEMA EXPERTO PARA EL DIAGNÓSTICO PATOLÓGICO PARA ANGUILAS (SEDPA)**

### **3.1. INTRODUCCIÓN**

El desarrollo de la acuicultura y de la pesca deportiva ha puesto de manifiesto la incidencia creciente de las enfermedades de los peces en estas actividades piscícolas y a menudo ha sido el vector de aparición, o al menos de propagación, de ciertas enfermedades (Neumann, 1985; Peters y Hartmann, 1986; Höglund y Andersson, 1993). La sanidad piscícola adquiere por tanto, una gran importancia económica que hace necesaria la lucha contra los problemas patológicos.

El estado de enfermedad se manifiesta en los peces mediante la aparición de lesiones y/o anomalías en el comportamiento (síntomas), lo que supone un descenso de la calidad física, rendimientos y, a menudo, la muerte de los sujetos afectados. Estas manifestaciones mórbidas son debidas a causas de orden físico, químico o biológico, actuando solas o en asociación, perturbando las funciones fisiológicas del animal, y poniéndose de manifiesto de manera natural o inducida (Kinkelin *et al.*, 1991). De esta forma, los constituyentes del medio ambiente acuático son susceptibles de ser peligrosos para los peces (Wedemeyer, 1981). Del mismo modo, los bioagresores, seres vivos que subsisten a expensas de los peces, están siempre presentes allá donde se encuentren sus huéspedes para los cuales representan una amenaza permanente. Así, el medio acuático encierra una población en la cual conviven animales y vegetales, con predadores intermediarios de parásitos, algas, etc. Todos estos elementos de forma aislada o conjuntamente pueden engendrar enfermedades (Roberts, 1981; Møllergaard y Dalsgaard, 1989; Kinkelin *et al.*, 1991; Shepherd y Bromage, 1999).

La actividad humana viene a añadir nuevos riesgos mediante la introducción de sustancias peligrosas (antifloculantes, clorines, supersaturación de determinados gases, etc.), o desarrollando prácticas que aumentan los efectos dañinos de los factores físicos, químicos o biológicos presentes en el medio, haciéndolos patógenos para los peces (Nash y Paulsen, 1981; Wickins, 1981). El primero de los casos es el de la polución, permanente o accidental, cuyos daños se hacen sentir principalmente sobre la fauna salvaje, pero que pueden igualmente alterar directa o indirectamente la salud de las explotaciones. En segundo lugar encontramos las prácticas de la tecnología piscícola, es decir, las operaciones referentes a la explotación (clasificaciones, cálculo de raciones, etc.), el comercio, transporte y la repoblación de las aguas.

La explotación intensiva de la anguila es uno de los ejemplos más contundentes de inducción a la acción de los factores patógenos. En este tipo de explotaciones, las enfermedades pueden aparecer como consecuencia de una calidad del agua inadecuada, falta de higiene, una mala nutrición, etc., o por combinación de dos o más de estos factores (Munro y Fijan, 1981), con la consecuente disminución de la producción y pérdida económica. Esas enfermedades pueden ser erradicadas, en la mayor parte de los casos, mediante la aplicación de un tratamiento farmacológico adecuado, lo que implica un proceso de diagnóstico. Sin embargo, este proceso se ve dificultado generalmente por la coincidencia de uno o varios factores como la escasez de expertos en ictiopatología o la abundancia/ausencia y dispersión de la información disponible. Esto influye además sobre una variable crítica en el diagnóstico, que es el tiempo transcurrido desde la observación de los síntomas/lesiones hasta la determinación de los agentes patógenos responsables de la enfermedad.

Estas características justifican la aplicación de métodos de diagnóstico englobados tradicionalmente dentro del campo de la inteligencia artificial conocidos como Sistemas Expertos. Según Castillo *et al.* (1996) y González-Andujar y Recio-Aguado (1997) el uso de los sistemas expertos se recomienda especialmente:

- Cuando el conocimiento es difícil de adquirir o se basa en reglas que sólo pueden ser aprendidas de la experiencia.
- Cuando la mejora continua del conocimiento es esencial y/o cuando el problema está sujeto a reglas o códigos cambiantes.
- Cuando los expertos humanos son caros o difíciles de encontrar.
- Cuando el conocimiento de los usuarios sobre el tema es limitado.
- Cuando se espera que el tiempo de respuesta frente a un determinado problema sea mínimo.

### **3.1.1. Los sistemas expertos como métodos de diagnóstico**

Un sistema experto es un programa de ordenador que emplea la información codificada en una base de conocimiento para resolver problemas de un determinado dominio calificados como difíciles, de la misma forma en que lo haría un experto humano en ese dominio. Un sistema experto que esté correctamente diseñado imita el proceso de razonamiento del experto humano para resolver problemas específicos, por lo que pueden ser usados tanto

por operarios no expertos para mejorar su capacidad de resolución de problemas, así como por expertos como sistemas inteligentes de ayuda.

Los primeros sistemas expertos fueron desarrollados a mediados de los años 60. Este período de investigación de la inteligencia artificial estuvo dominado por la creencia de que la determinación de leyes de razonamiento generales unidas a potentes computadoras harían posible la creación de una máquina 'superhumana' capaz de resolver cualquier tipo de problema. Debido a esto, se generaron una gran cantidad de programas de ordenador con grandes expectativas que, no obstante, no llegaron a cumplir los sueños de sus programadores. La evolución desde el propósito general hasta el particular de los sistemas expertos ocurrió con el desarrollo de MYCIN (Shortliffe y Buchanan, 1975) y DENDRAL (Lindsay *et al.*, 1980). El primero de ellos, un sistema experto basado en reglas, fue desarrollado para ayudar en el diagnóstico de infecciones bacterianas durante el período crítico que va desde las 24 hasta las 48 horas tras un ingreso hospitalario. DENDRAL fue diseñado para inferir la estructura molecular de un compuesto desconocido a partir de la masa espectral y su magnetismo nuclear. El sistema usaba un código que enumeraba sistemáticamente todas las posibles estructuras moleculares para posteriormente llevar a cabo una selección de las más probables mediante el empleo de datos químicos. Desde entonces hasta la actualidad muchos han sido los sistemas expertos diseñados como aplicaciones a distintos campos de la ciencia y la industria (Campbell *et al.*, 1982; Hawkes, 1992 y 1993; Afgan *et al.*, 1994; Malczewski, 1995; Tarifa y Scenna, 1997) con especial atención al diagnóstico clínico humano (Kahn y Anderson, 1994; Van Diest *et al.*, 1994; Andreassen *et al.*, 1995; Birndorf *et al.*, 1996; Firriolo y Levy, 1996; Hernando *et al.*, 1996; Kahn, 1996; Kentala *et al.*, 1996; Nguyen *et al.*, 1996; Díez *et al.*, 1997; Ohayon *et al.*, 1997; Weyn *et al.*, 1998; Weyn *et al.*, 1999), habiéndose obtenido en la mayor parte de los casos resultados muy significativos.

El campo de la producción animal también ha sido objeto de aplicación de los sistemas expertos (González-Andújar y Recio-Aguado, 1997). De esta forma, GTEP-X, MAMMITRON, TRBVENT y DXMAX, fueron diseñados para ayudar en la resolución de distintos problemas relacionados con las explotaciones porcinas y bovinas (Amador-Hidalgo, 1996). Más concretamente para la producción en medio acuático, Hoskins *et al.* (1983) presentaron dos programas de ordenador interconectados llamados CASE REPORTER y FISH HEALTH específicamente diseñados para ayudar en la mejora de la producción del salmón. CASE REPORTER intentaba determinar cuales eran las principales causas que generaban un problema de salud en la explotación, solicitando para ello información precisa sobre las condiciones de producción, comportamiento de los animales y signos de enfermedad. Por su parte, FISH HEALTH intervenía en el control y prevención de enfermedades haciendo uso de la información disponible sobre las veinticuatro enfermedades más frecuentes del salmón así como de los datos computados para CASE REPORTER. Posteriormente, Garnerin y Tuffery (1988) programaron un sistema experto

(SCHUBERT 3000) que era capaz de ofrecer información útil sobre patología, prevención de enfermedades, manejo y producción de peces. Más recientemente otros sistemas expertos programados para diagnosticar enfermedades en peces tropicales como FISHVET (Zeldis y Prescott, 2000) o HAMES de la Universidad de Hawai son capaces de realizar la misma función.

### **3.1.2. Formulación del problema del diseño de un sistema experto para diagnóstico clínico**

El tipo de problemas que pueden solucionar los sistemas expertos pueden clasificarse como deterministas y estocásticos. Los primeros se caracterizan por que pueden ser formulados usando una serie de reglas bien definidas que relacionan varios objetos y por obtener conclusiones utilizando un mecanismo de razonamiento lógico (Castillo *et al.*, 1996). Normalmente este tipo de reglas están caracterizadas por una estructura básica del tipo 'SI...ENTONCES'. Por otra parte, en situaciones en las cuales las relaciones entre los objetos no está claramente definida (como en el diagnóstico clínico), es necesario introducir medios que traten la incertidumbre asociada al propio diagnóstico. De esta forma algunos autores como Kentala *et al.* (1996) usan la misma estructura de los sistemas basados en reglas, pero incorporando una medida de incertidumbre mediante la adición a la base de conocimiento de límites superiores e inferiores de la puntuación asociada a síntomas y utilizando fórmulas de propagación para calcular la incertidumbre asignada a las conclusiones. Otros investigadores (Shortlife y Buchanan, 1975; Fatma *et al.*, 1993; Nguyen *et al.*, 1996) siguen un esquema semejante incluyendo factores de certeza. De esta forma a cada atributo o síntoma se le hace corresponder un peso que es una medida de la importancia relativa que tiene ese síntoma sobre una determinada enfermedad.

Cuando los datos asociados al problema a resolver son abundantes y se dispone de información objetiva sobre las frecuencias de enfermedades y síntomas de la población, se puede utilizar como medida de incertidumbre la probabilidad, en la que la distribución conjunta de un grupo de variables se usa para describir las relaciones de dependencia entre ellas (Castillo *et al.*, 1996). Bryan *et al.* (1994), Chang *et al.* (1994), Haddawy *et al.* (1994) Díez *et al.* (1997) y Kahn *et al.* (1997) entre otros, emplean un esquema de representación del conocimiento en forma de red probabilística o bayesiana en la cual las relaciones entre variables quedan definidas gráficamente. Esta forma de representación del conocimiento permite una propagación de la incertidumbre muy eficiente constituyendo un modelo causal del cual es posible obtener inferencias y conclusiones. Al igual que en los casos anteriores, la base de funcionamiento de las redes probabilísticas está construida sobre la consideración de proposiciones individuales a las que se asigna una estimación del grado de

creencia que se satisface dada una evidencia. En aquellos casos en los que la falta de datos dificulta la aplicación de funciones de probabilidad bayesianas es posible la aplicación de otras técnicas alternativas, de forma individual o conjunta, como la denominada teoría de Dempster-Shafer (Dempster, 1968; Shafer, 1982; Hájek, 1994) o la lógica borrosa (Zadeh, 1979; Klir *et al.*, 1997).

### **3.2. DESCRIPCIÓN DEL SISTEMA EN ESTUDIO**

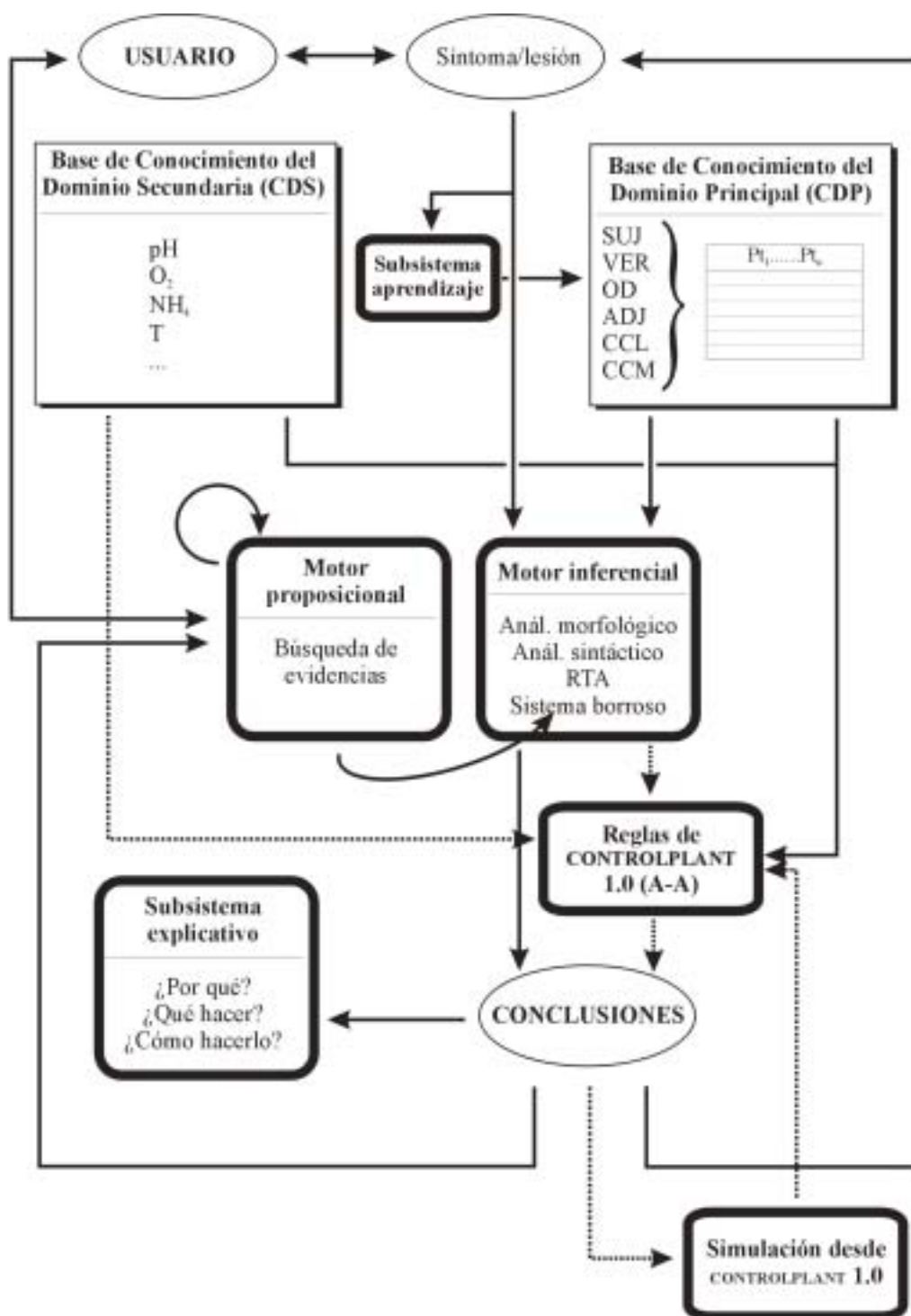
Los sistemas expertos suelen estar constituidos por varios módulos íntimamente relacionados entre sí, cuya interacción junto con los datos proporcionados por el usuario, permite al sistema la obtención de conclusiones. De esta forma, los sistemas expertos básicamente constan de una o varias bases de datos o bases de conocimiento, un motor de inferencia o núcleo del sistema, un subsistema explicativo, un motor proposicional, un subsistema de adquisición de conocimiento y una interfaz usuario-máquina (figura III.1).

El objetivo principal de este capítulo es conseguir el correcto funcionamiento de la estructura de control básica del sistema. Para ello, el sistema tiene que proporcionar la información suficiente en forma de conclusiones al usuario, entendiéndose éstas como el diagnóstico de la/s patología/s responsable/s del deterioro de la salud animal en la planta y la determinación de las posibles soluciones mediante la aplicación de medidas correctivas (fármacos, dosis y duración del tratamiento). Las sugerencias aportadas por el sistema permitirán tomar una decisión a corto plazo sobre el estado de salud de la explotación. Por otra parte, la interacción del sistema a través del conjunto de reglas **A-A** con el módulo de gestión de la planta (**CONTROLPLANT**) en estado de simulación debe suministrar indicios de la aparición de nuevos focos patológicos.

### **3.3. AGENTES PATOLÓGICOS RECONOCIDOS POR EL SISTEMA**

Los procesos patológicos que pueden afectar a la producción de anguilas y que, por tanto, deben ser reconocidos por el sistema pueden ser calificados de forma genérica como: enfermedades provocadas por bioagresores (virus, bacterias, hongos y parásitos animales) (tabla III.1a) y componentes del medio ambiente (temperatura, oxígeno, productos nitrogenados, etc.) (tabla III.1b).





**Figura III.1.** Diagrama de flujo del funcionamiento del Sistema Experto para el Diagnóstico Patológico en Anguillas (SEDPA). Las líneas continuas indican los posibles caminos seguidos por el programa cuando funciona de forma independiente. Las líneas discontinuas señalan los caminos alternativos seguidos cuando el programa de gestión CONTROLPLANT interactúa con SEDPA en modo simulación

**Tabla III.1a.** Principales bioagresores que atacan a la anguila europea (*A. anguilla*) en cultivos intensivos

BIOAGRESORES	DISTRIBUCIÓN GEOGRÁFICA	PRESENCIA
<b>VIRUS</b>		
NPI-EVE	Cosmopolita	Muy rara
<b>BACTERIAS</b>		
<i>Aeromonas hydrophila</i>	Cosmopolita	Común
<i>Edwardsiella tarda</i>	Norte de África, América del Norte y Asia	Rara
<i>Flexibacter columnaris</i>	Cosmopolita	Rara
<i>Pseudomonas fluorescens</i>	Cosmopolita	Común
<i>Pseudomonas anguilliseptica</i>	Cosmopolita	Común
<i>Vibrio anguillarum</i>	Cosmopolita	Rara
<i>Yersinia ruckeri</i>	América del Norte, Oceanía y Europa	Muy rara
<b>HONGOS</b>		
<i>Saprolegnia sp.</i>	Cosmopolita	Rara
<b>PROTOZOARIOS</b>		
<i>Eimeria sp.</i>	Cosmopolita	Muy rara
<i>Ichthyobodo sp.</i>	Cosmopolita	Rara
<i>Ichthyophthirius multifiliis</i>	Cosmopolita	Muy común
<i>Trichodina sp.</i>	Cosmopolita	Común
<b>MYXOZOARIOS</b>		
<i>Henneguya sp.</i>	Cosmopolita	Muy rara
<i>Myxidium sp.</i>	Cosmopolita	Común
<i>Myxobolus sp.</i>	Cosmopolita	Muy rara
<b>HELMINTOS</b>		
<i>Dactylogyrus anguillae</i>	Cosmopolita	Muy común
<i>Gyrodactylus sp.</i>	Cosmopolita	Muy común
<b>ARTRÓPODOS</b>		
<i>Argulus sp.</i>	Cosmopolita	Rara
<i>Ergasilus sp.</i>	Cosmopolita	Común
<b>SIN POSICIÓN TAXONÓMICA CLÁRAMENTE ESTABLECIDA</b>		
<i>Dermocystidium anguillae</i>	---	Rara
Papilomatosis	---	Muy rara

**Tabla III.1b.** Principales componentes del medio ambiente que pueden afectar a la anguila europea (*A. anguilla*) en cultivos intensivos

COMPONENTE AMBIENTAL	PRESENCIA
Niveles altos de Amoniac	Muy común
Niveles altos de Nitritos	Muy común
Niveles altos de Oxígeno	Común
Niveles bajos de Oxígeno	Muy común
Alta turbidez del agua	Muy común
Temperatura por debajo del óptimo fisiológico	Común
Temperatura por encima del óptimo fisiológico	Común
Niveles altos de pH	Rara
Niveles bajos de pH	Rara

### 3.3.1. Componentes del medio ambiente

Los componentes del medio ambiente que pueden ser causa de enfermedades en los peces pueden actuar generalmente siguiendo dos modelos distintos: un modelo directo por el cual es el propio componente el que promueve la aparición de la enfermedad, o un modelo indirecto mediante el cual la variación de un determinado parámetro puede aumentar o disminuir los niveles de otros factores ambientales hasta alcanzar valores tóxicos para la población, lo que puede facilitar además la proliferación de organismos como, virus, hongos, bacterias, etc. Estos componentes son propiedades físicas del agua, básicamente temperatura y turbidez, y propiedades químicas como el carácter ácido o básico, los gases disueltos, sustancias nitrogenadas, etc.

Los efectos directos de alguno de estos componentes sobre los peces pueden manifestarse como una fuerte irritación, que lleva consigo secreción de moco, hemorragias cutáneas y/o la muerte de los animales, por ejemplo como consecuencia de una elevación o disminución del pH. También pueden ser el origen de enfisemas cutáneos o embolias gaseosas que se producen después de la exposición continuada a hipersaturaciones de gases como el oxígeno. Esta saturación de gases puede tener dos causas: una elevación de más de 10 °C de la temperatura del agua en muy poco tiempo (efecto indirecto de la variación de temperatura) o por compresión del aire en el agua (efecto indirecto de la variación de presión) (Kinkelin *et al.*, 1991). Según Smart (1976), las dosis subletales de amonio son responsables de hiperplasias branquiales asociadas a una importante secreción mucosa y a hemorragias puntiformes como consecuencia de la telangiectasia. Efectos similares se han observado bajo la exposición permanente de los peces a concentraciones de 0.1 mg/l de nitritos. Smith y Williams (1974) señalan la aparición de focos necróticos en animales intoxicados por nitritos. Mellergaard y Dalsgaard (1989) indican además que el efecto

individual o combinado de alguno de estos factores ambientales pueden ser el motivo de la aparición de estrés asociado a una disminución de la inmunidad.

### 3.3.2. Virus

En la actualidad, el tipo de virus que provoca mayores pérdidas económicas en las explotaciones intensivas de anguilas son los conocidos como virus **ARN** con doble cadena bisegmentadas. Dentro de este grupo encontramos los virus de la **NPI** (Necrosis Pancreática Infecciosa) que afecta principalmente a especies de la familia *Salmonidae* (Underwood *et al.*, 1977) y que pueden ser reunidos en tres grupos serológicos (Okamoto *et al.*, 1983): el grupo I o americano, representado por el virus West Biston; el grupo II, europeo conocido con el nombre de Sp; y el grupo III, que corresponde al virus de tipo Ab y al virus de la branquionefritis de la anguila también conocido con el nombre de **EVE** (European Virus Eel) (Sano *et al.*, 1981; Sano *et al.*, 1992). Con respecto a este último virus, Hedrick *et al.* (1983) consideran que aunque serológica y morfológicamente está muy próximo a un virus **NPI** (tipo Ab) puede ser considerado como una entidad viral distinta, ya que la movilidad electroforética de su genoma difiere significativamente de la del virus **NPI**.

Las primeras manifestaciones clínicas de la aparición de la enfermedad en la explotación son la inapetencia y la anorexia, seguidas de astenia. Rápidamente aparecen alteraciones en el equilibrio que se traducen por natación helicoidal que se alterna con manifestaciones periódicas de excitación y postración. Paralelamente a los síntomas descritos, se observa una serie de modificaciones de aspecto corporal. Las hemorragias son las primeras alteraciones visibles a nivel de la piel y las mucosas. Éstas se localizan en el tejido subcutáneo y en las branquias. Se exteriorizan entonces exoftalmias como consecuencia del efecto conjugado de los edemas y las hemorragias, asociadas a menudo a una decoloración branquial. La evolución puede ser rápida ya que pueden transcurrir tan solo 12 horas desde la aparición de las exoftalmias y la muerte del sujeto (Roberts, 1981; Kinkelin *et al.*, 1991; Shepherd y Bromage, 1999).

### 3.3.3. Bacterias

Las enfermedades bacterianas constituyen un conjunto considerable en la patología de los peces (Kinkelin *et al.*, 1991). Concretamente en el caso de la anguila, siete especies acaparan prácticamente la totalidad de los casos registrados de enfermedades provocadas por bacterias. Desde un punto de vista fisiológico estas especies pueden ser separadas en dos grupos claramente diferenciados: anaerobios estrictos y anaerobios facultativos. Dentro

del primero nos encontramos con un representante de la familia Cytophagaceae (*Flexibacter columnaris* –sin.: *Flavobacterium columnare*-) y con dos representantes de la familia Pseudomonadaceae (*Pseudomonas fluorescens* y *P. anguilliseptica*). Con respecto al segundo, la única familia representada (Enterobacteriaceae) agrupa a *Aeromonas hydrophila*, *Edwardsiella tarda*, *Vibrio anguillarum* (sin.: *Lisnonella anguillarum*) y *Yersinia ruckeri*.

Las enfermedades provocadas por estos organismos son con mucho las más frecuentes y espectaculares. Se caracterizan fundamentalmente por su carácter septicémico con dominante necro-hemorrágico aunque pueden producir infecciones locales. Los primeros aspectos lesionales se manifiestan por la presencia de hemorragias externas que se localizan en puntos preferenciales, tales como la base de inserción de las aletas y la zona perianal. Al mismo tiempo se forman petequias sobre los flancos y en la región cefálica: labios, opérculos y cavidad bucal. La extensión de las hemorragias al conjunto del cuerpo no es rara y explica el apelativo de ‘pestes rojas’ que se aplica a las aeromoniasis, pseudomoniasis, y más recientemente a la edwardsielosis (Wakabayashi y Egusa, 1972; Møllergaard y Dalsgaard, 1989; Michel *et al.*, 1992; Lönnström *et al.* 1994; Wiklund y Lönnström, 1994; Berthe *et al.*, 1995). Al mismo tiempo pueden producirse focos necróticos que se caracterizan por la sobreelevación de la piel de tipo flicteno. Estas vesículas, más o menos discretas, se rompen con facilidad evolucionando hacia una ulceración superficial en la vibriosis (tan solo cuando el porcentaje de salinidad se encuentra entre el 3 y el 5 %, Pazos *et al.*, 1993) (Kitao *et al.*, 1983; Egidius, 1987; Larsen *et al.*, 1988; Møllergaard y Dalsgaard, 1989; Myhr *et al.*, 1991), edwardsielosis y pseudomoniasis. En el caso de *F. columnaris*, la infección puede extenderse también a las branquias donde pueden observarse necrosis amarillentas asociadas a sufusiones hemorrágicas (Kinkelin *et al.*, 1991; Bernardet, 1998; Shepherd y Bromage, 1999).

Internamente estas enfermedades se caracterizan por infecciones de tipo entérico, que se traducen en la congestión de la parte terminal del tubo digestivo, sobre todo en yersiniosis y aeromoniasis. Asimismo es frecuente observar la presencia de líquido ascítico, resultado de la acumulación ligada al mal funcionamiento circulatorio y renal. Este líquido, seroso y amarillento en su forma inicial, puede tomar un aspecto fibrinoso y hemorrágico si las rupturas son importantes. Las lesiones necróticas evolucionan a menudo en pequeños focos o en el tejido intersticial del riñón, bazo e hígado (Kinkelin *et al.*, 1991; Shepherd y Bromage, 1999).

### 3.3.4. Hongos

Las micosis de los peces constituyen ciertamente el aspecto más confuso y el menos explorado de la ictiopatología. Tan solo se conocen una docena de especies cuya posición taxonómica al menos en su mayor parte, es dudosa y arbitraria. En la actualidad, los hongos parásitos de peces se agrupan principalmente dentro de dos grupos: los oomycetos y los hyfomicetos. En el primero de los grupos se encuentran englobadas las especies de *Saprolegnia*, un género ubicuo de hongos acuáticos encontrados frecuentemente en piscifactorías (Marking *et al.*, 1994), posiblemente responsables de la mayor parte de infecciones fúngicas registradas en explotaciones de anguilas (Shepherd y Bromage, 1999). Comúnmente son designadas con el nombre de ‘algodones’, haciendo referencia al aspecto coposo de las formaciones micelares. Los patólogos las clasifican globalmente con el nombre de saprolegniosis, pero el término es poco adecuado ya que por regla general en las infecciones se ven implicados varios géneros (Pickering y Willoughby, 1977; Kinkelin *et al.*, 1991) como *Branquiomycetes* e *Ichthyophonus*.

Las infecciones fúngicas evolucionan generalmente a partir de focos de colonización de los tegumentos o de las mucosas. Parece ser que la instalación y el desarrollo de estos agentes está condicionado por factores de estrés, o por la existencia de heridas iniciales. Los primeros estadios de la infección se manifiestan a simple vista por formaciones algodonosas que pueden variar del blanco al marrón según el número de partículas sedimentarias que estén atrapadas entre los micelios. Las localizaciones más comunes son la piel y las branquias donde pueden formar crecimientos de color blanco en forma de cigarro. Producen lesiones básicamente extensivas por lo que el pez acaba por manifestar signos de agotamiento lo que acelera su muerte (Mellergaard y Dalsgaard, 1989; Kinkelin *et al.*, 1991; Shepherd y Bromage, 1999).

### 3.3.5. Parásitos

El parasitismo es un fenómeno frecuente en los peces, sin embargo las enfermedades parasitarias no se manifiestan más que cuando las condiciones del medio ambiente permiten la proliferación del parásito. Es por ello por lo que, salvo raras excepciones (Molnár *et al.*, 1991; Székely, 1994), las parasitosis clínicas son escasas en estado natural mientras que en las explotaciones están muy extendidas (Kinkelin *et al.*, 1991). Desde un punto de vista económico, los parásitos animales más importantes para una explotación intensiva de anguilas pueden agruparse en: Subtipo Sarcomastigophora o Rizoflagelados que agrupa a las especies del género *Ichthyobodo* (sin.: *Costia*); Subtipo Sporozoa representado por el género *Eimeria*; Subtipo Ciliophora dentro del cual encontramos a los géneros *Ichthyophthirius* y *Trichodina*; Subtipo Cnidospora caracterizado por los géneros

*Myxidium*, *Myxobolus* y *Henneguya*; Tipo Platelminfos que agrupa a *Gyrodactylus* y *Dactylogyrus*; Tipo Nematelminfos en el que se engloba el género *Anguillicola* y finalmente el Tipo Arthropoda donde encontramos los géneros *Ergasilus* y *Argulus*.

Los cuadros clínicos y lesiones macroscópicas más importantes están caracterizadas por afecciones cutáneo-branquiales pruriginosas y deteriorantes que aparecen fundamentalmente en individuos jóvenes, en los que se llega a producir una descamación epidérmica y necrosis tisular que conduce en la mayor parte de los casos a la formación ulcerosa. Esto generalmente provoca la disminución del espesor de la piel, facilitándose así la penetración del agua y la aparición de edemas superficiales. En el origen de estas alteraciones encontramos a *Argulus*, *Ergasilus* y microparásitos como *Ichthyophthirius* y *Trichodina* (Kinkelin *et al.*, 1991; Lobo-da-Cunha y Azevedo, 1993; Shepherd y Bromage, 1999).

Las enfermedades respiratorias también pueden ser provocadas por parásitos externos con tropismo branquial. Las branquioparasitosis externas tienen unas sintomatologías de tendencia irritativa con secreción de moco y necrosis de las laminillas branquiales. Entre los parásitos branquiales más peligrosos figuran *Ichthyobodo*, *Ichthyophthirius*, *Dactylogyrus*, *Gyrodactylus* y *Ergasilus*. Las afecciones respiratorias quísticas son el resultado de microsporidios y, sobre todo, de mixozoos. Los más frecuentes son *Myxobolus*, *Henneguya* y *Myxidium* (Kent *et al.*, 1993). Por regla general, éstas se manifiestan en animales adultos por la presencia de quistes de color blanco que miden varios milímetros y presentan una forma esférica u oval (Roberts, 1981; Mellergaard y Dalsgaard, 1989; Kinkelin *et al.*, 1991; Shepherd y Bromage, 1999).

Entre las afecciones de tipo digestivo cabe destacar las provocadas por *Eimeria*, caracterizadas por enteritis fatales (Roberts, 1981), *Anguillicola* cuya instalación en el hospedador origina una aguda peritonitis, inflamación de la vejiga natatoria y por tanto un aumento significativo de la cavidad abdominal (Mellergaard y Dalsgaard, 1989; Székely *et al.*, 1991), y *Myxidium* que se asocia a enfermedades renales proliferativas (PKD) (Clifton-Hadley *et al.*, 1984; MacConnell *et al.*, 1989).

### **3.4. MATERIAL Y MÉTODOS**

#### **3.4.1. Bases de conocimiento del dominio**

La base de conocimiento del dominio constituye la memoria permanente del sistema en la que se encuentra recogida toda la información que, junto con los datos proporcionados por el usuario, necesita el sistema para obtener sus conclusiones. Esto no significa que esta base

de datos sea estática o permanente, ya que puede ser modificada continuamente como consecuencia del aumento o cambio de la información existente sobre el dominio.

La información contenida en la base de datos ha sido extraída de dos fuentes principales (Haddawy *et al.*, 1994; Kahn *et al.*, 1997): a) publicaciones especializadas (libros y revistas) (Roberts, 1981; Clifton-Hadley *et al.*, 1984; MacConnell *et al.*, 1989; Mellergaard y Dalsgaard, 1989; Kinkelin *et al.*, 1991; Kentala *et al.*, 1996; Shepherd y Bromage, 1999, entre otros) y, b) los registros históricos de parámetros físico-químicos, biológicos y de producción recogidos en la planta desde el año 1997.

Para dar flexibilidad y potencia de cálculo al sistema la base de conocimiento se ha dividido en dos, atendiendo básicamente a la distinta naturaleza de la información codificada. En una de ellas, que llamaremos base de Conocimiento del Dominio Principal (**CDP**), se recoge la información extraída de las publicaciones especializadas y de los casos reales acaecidos en la planta. En ésta se han establecido mediante unos y ceros las relaciones existentes entre las distintas patologías que el sistema es capaz de reconocer, con estructuras o etiquetas sintácticas (Van Diest *et al.*, 1994) (tabla III.2), lo que facilita la detección de la información relevante introducida por el usuario, el proceso de búsqueda secuencial y la formación de la Red de Transición Aumentada (**RTA**). Para la organización del léxico usado para el análisis, se lleva a cabo una adaptación de la aproximación léxica (Steffens, 1994) mediante la aplicación de un diccionario estructurado jerárquicamente. Por tanto, tan solo se acumulan las formas canónicas de las palabras, asignándoles las figuras sintácticas mediante su acumulación en distintas bases de datos (Winiwarter, 2000). Así, la **CDP** ha sido dividida en seis subbases denominadas: **SUJ, VER, OD, ADJ, CCL y CCM**. De esta forma se consigue una eliminación de ambigüedad mediante la separación de términos en diferentes significados (Losee, 2001). El criterio de asignación de los términos a cada una de las subbases ha sido la función sintáctica típica (sujeto, verbo, objeto directo, adjetivo, complemento circunstancial de lugar y modo) que los autores y técnicos que describen las patologías y los síntomas/lesiones asociados han asignado a dichos términos.

Por otra parte, tenemos la base de Conocimiento del Dominio Secundaria (**CDS**) (tabla III.3) que está constituida por los registros históricos recogidos en la planta. De la misma forma, la **CDS** se encuentra dividida en tantas subbases como parámetros son medidos en la explotación: O<sub>2</sub> disuelto (mg/l), temperatura del agua (°C), pH, NH<sub>3</sub> (mg/l), NO<sub>2</sub><sup>-</sup> (mg/l), NO<sub>3</sub><sup>-</sup> (mg/l), tipo de pienso suministrado (marca y grosor), cantidad de pienso suministrado (kg/tanque), incremento de biomasa (% biomasa/día), factor de conversión o **FCR** esperado y mortalidad (individuos muertos/día).



**Tabla III.2.** Distribución de la información codificada en la base de Conocimiento del Dominio Principal (CDP) para aquellas estructuras sintácticas que funcionan normalmente como adjetivos dentro de la oración. En la primera fila se representan las distintas patologías reconocidas por el sistema (Pt). En la primera columna se muestran a modo de ejemplo algunos de los adjetivos que podemos encontrar relacionados con dichas patologías

ADJ	Pt <sub>1</sub>	Pt <sub>2</sub>	Pt <sub>3</sub>	...	Pt <sub>n</sub>
Roj-	1	1	0	-	1
Grand-	1	0	1	-	0
Inflama-	0	0	1	-	0
...	-	-	-	-	-
Irrita-	1	1	0	-	1

**Tabla III.3.** Distribución de la información codificada en la base de Conocimiento del Dominio Secundaria (CDS). A modo de ejemplo se muestran los registros de O<sub>2</sub> disuelto (mg/l) en distintas fechas para cuatro tanques

Fecha	Tanque A4	Tanque B2	Tanque C12	...	Tanque G14
13/01/1998	7.2	8.3	6.5	...	7.0
14/01/1998	8.6	8.9	7.0	...	6.5
15/01/1998	9.0	9.9	6.8	...	6.7
16/01/1998	7.9	9.8	6.5	...	7.0
...	...	...	...	...	...
20/05/1999	6.9	10.0	6.3	...	6.8
21/05/1999	7.8	12.0	5.6	...	10.0
22/05/1999	7.0	10.2	7.1	...	9.6

### 3.4.2. El motor de inferencia

El motor de inferencia de un sistema experto es el auténtico cerebro del sistema, permitiendo la obtención de conclusiones a través de la gestión y control, de forma lógica, de todo el proceso relacionado con el manejo y utilización eficiente de la información incorporada a la base de conocimiento y los datos aportados por el usuario.

En **SEDPA**, el motor de inferencia consta de cuatro partes claramente diferenciadas: a) un analizador morfológico-sintáctico a través del cual se constituye la Red de Transición Aumentada (**RTA**), b) un controlador borroso que asigna niveles de creencia individuales a cada una de las patologías seleccionadas, c) un módulo de propagación de la incertidumbre basado en la teoría de Dempster-Shafer y, d) una rutina del test estadístico Chi-cuadrado ( $\chi^2$ ) modificado que es utilizado como elemento separador de los grupos patológicos con mayor aportación a la creencia global.

### 3.4.2.1. Análisis morfológico-sintáctico. Red de Transición Aumentada (RTA)

Como ya se comentó en el apartado 3.1.1, un sistema experto que esté bien diseñado debe poder ser utilizado por un usuario inexperto. Siguiendo esta premisa se ha dotado al sistema de un módulo de extracción de información a partir de datos con estructura de lenguaje natural. De esta forma, las entradas al sistema son del tipo: *‘Las angulas del tanque A4 presentan manchas blancas en la cabeza’*. Para convertir esta oración lineal de entrada en una estructura jerárquica en la que se relacionen cada una de las unidades de la oración con su significado, es necesario llevar a cabo un preprocesamiento de la información introducida conocido como análisis morfológico-sintáctico.

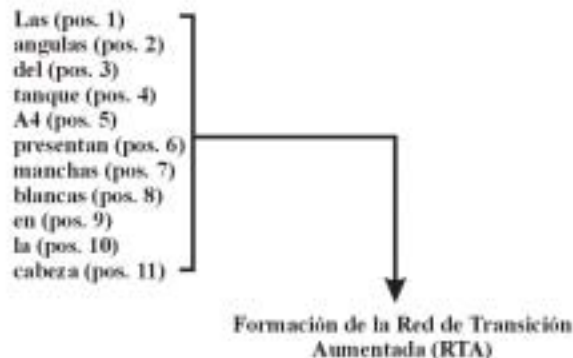
El primer paso del análisis morfológico-sintáctico consiste en la separación física de cada componente (palabra-etiqueta) de la oración. Para ello, una rutina examina uno a uno todos los caracteres que la componen, determinando la posición de los espacios en blanco comprendidos entre la primera letra (en el ejemplo: ‘L’) y el punto final, detectándose de esta forma el número total de palabras (en el ejemplo 11) y la posición que ocupa cada una de ellas dentro de la oración. Seguidamente se procede a llevar a cabo el análisis sintáctico mediante la formación de una **RTA** (figura III.2). **RTA** es un procedimiento de análisis sintáctico descendiente que permite incorporar al sistema varias clases de conocimientos de forma que pueda funcionar eficientemente. Debido a la temprana utilización de **RTAs** (Woods, 1973), este mecanismo se ha usado en varios sistemas de comprensión del lenguaje (De Carolis *et al.*, 1996). Una **RTA** es similar a una máquina de estados finitos en la que se ha ampliado el tipo de etiquetas que se pueden asociar a los arcos que definen las transiciones (Rich y Knight, 1994; Cazorla-Quevedo *et al.*, 1999). Los arcos pueden etiquetarse como: categorías de palabras como DETERMINANTE o NOMBRE y llamadas a rutinas que reconozcan componentes importantes de la oración como el significado de ‘A4’.

Así la formación de la **RTA** asociada al ejemplo anterior se llevaría a cabo de la siguiente forma:

- I. Comprobar si la primera palabra-etiqueta es igual a ‘NO’ (*La comprobación no tiene éxito, por lo que se hace una llamada a la subrutina SINTAGMA NOMINAL*).
- II. Hacer una comprobación de categorías para ver si ‘Las’ es un determinante (*La comprobación tiene éxito, por lo que el registro DETERMINANTE se pone en DEFINIDO*).
- III. Hacer una comprobación de categorías para ver si ‘angulas’ es un nombre (*La comprobación tiene éxito, por lo que el registro SUJETO se pone en DEFINIDO*).



Número de pasos hasta completar el proceso=Número de caracteres (n=65)  
 Número de espacios encontrados=10  
 Número de palabras en la oración=11



**Figura III.2.** En el esquema se representa una entrada tipo llevada a cabo por parte de un usuario que ha detectado un síntoma en el tanque A4, así como el funcionamiento del algoritmo que permite realizar el análisis morfológico. En el ejemplo se muestra el resultado de la búsqueda del algoritmo en el paso 3 (s) y 36 ( ' ' ) así como la determinación de la posición de cada palabra dentro de la oración. Después de la finalización de este proceso se hace una llamada a la subrutina que permite la formación de la Red de Transición Aumentada (RTA)

- IV. Hacer una comprobación de categorías para ver si 'del' es un verbo (La comprobación no tiene éxito por lo que se hace una llamada a la subrutina QUE TANQUE).
- V. Comprobar si 'del' presenta una estructura del tipo 'A1, A2,...,G13, G14' (La comprobación no tiene éxito).
- VI. Comprobar si 'tanque' presenta una estructura del tipo 'A1, A2,...,G13, G14' (La comprobación no tiene éxito).
- VII. Comprobar si 'A4' presenta una estructura del tipo 'A1, A2,...,G13, G14' (La comprobación tiene éxito por lo que el registro QUE TANQUE se pone en DEFINIDO).
- VIII. Parcialmente se devuelve [SINTAGMA NOMINAL(angulas(A4))].
- IX. Hacer una llamada a la subrutina SINTAGMA VERBAL.

- X. Hacer una comprobación de categorías para ver si ‘*presentan*’ es un verbo (*La comprobación tiene éxito por lo que el registro VERBO se pone en DEFINIDO*).
- XI. Hacer una comprobación de categorías para ver si ‘*manchas*’ es un nombre que no hace referencia a una localización (*La comprobación tiene éxito por lo que el registro OBJETO DIRECTO se pone en DEFINIDO*).
- XII. Hacer una comprobación de categorías para ver si ‘*blancas*’ es un adjetivo (*La comprobación tiene éxito por lo que el registro ADJETIVO se pone en DEFINIDO*).
- XIII. Hacer una comprobación de categorías para ver si ‘*en*’ es una preposición (*La comprobación tiene éxito*).
- XIV. Hacer una comprobación de categorías para ver si ‘*la*’ es un determinante (*La comprobación tiene éxito*).
- XV. Hacer una comprobación de categorías para ver si ‘*cabeza*’ es un nombre que hace referencia a una localización (*La comprobación tiene éxito por lo que el registro COMPLEMENTO CIRCUNSTANCIAL DE LUGAR se pone en DEFINIDO*).
- XVI. Parcialmente se devuelve: [SINTAGMA VERBAL(*manchas-blancas(cabeza)*)].
- XVII. Estado final del análisis y formación de la **RTA**:

{ [SN(*angulas(A4)*)] presentan [SV(*manchas-blancas(cabeza)*)] }

Una vez asignada la **RTA** se procede a determinar las asociaciones establecidas en cada una de las subbases de Conocimiento del Dominio Principal entre las etiquetas sintácticas y las distintas patologías codificadas. Tras la introducción del síntoma/lesión obtenemos por tanto el número de patologías asociadas a cada una de las etiquetas sintácticas. Cotejando aquellas patologías que son comunes a todas la etiquetas, es decir, siguiendo el concepto de consistencia (Fatma *et al.*, 1993) por el que los candidatos del diagnóstico son aquellos que no entran en conflicto con los datos, obtenemos el grupo de patologías que se relaciona con la **RTA** recién formada (figura III.3).

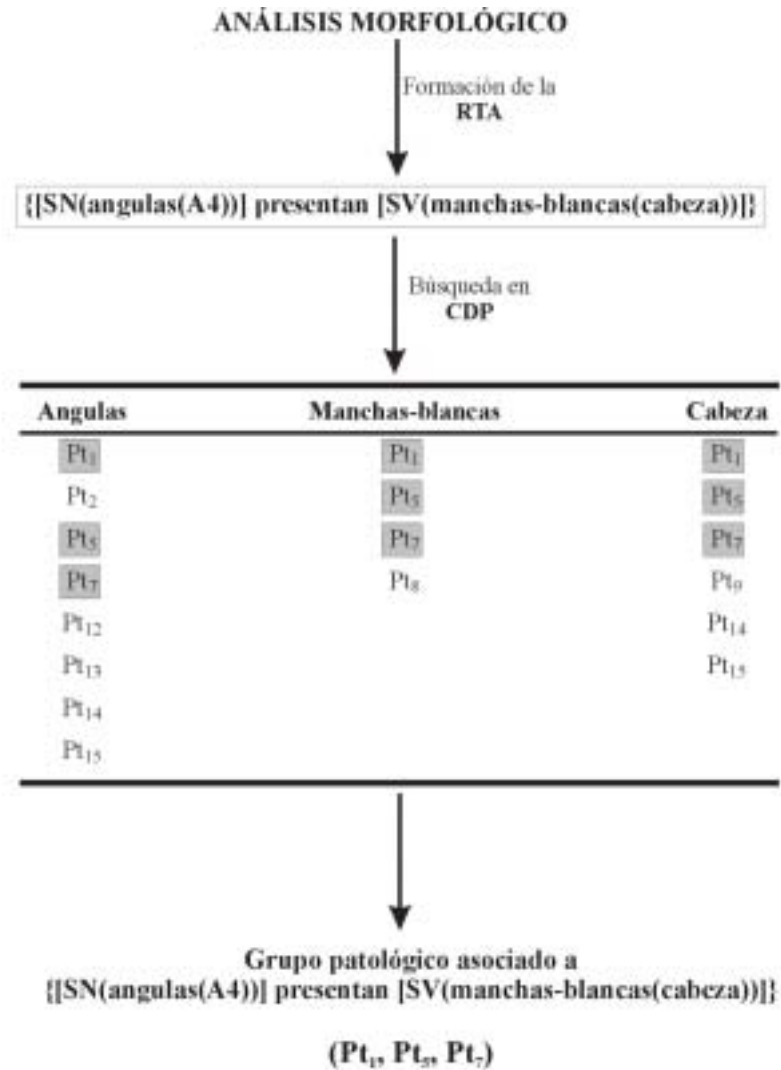
Paralelamente a este proceso se obtiene el valor de la variable porcentaje compartido (**PCOM<sub>p,i</sub>**) para las **p** etiquetas e **i** subbases de **CDP**, es decir:

$$PCOM_{p,i} = \frac{\sum_{j=1}^n P_{t_{p,j}}}{n} \quad (3.1)$$

donde  $j$  es la fila que ocupa la etiqueta  $p$  en su correspondiente  $CDP$  y  $n$  es el número total de patologías reconocidas por el sistema. Finalmente de los  $PCOM_{p,i}$  obtenidos se escoge como entrada al controlador borroso el más restrictivo, es decir:

$$[MPC] = \text{Mínimo}[PCOM_{p,i}] \quad (3.2)$$

donde  $[MPC]$  es la variable Mínimo Porcentaje Compartido.



**Figura III.3.** Tras el análisis morfológico y la formación de la RTA se procede a llevar a cabo la búsqueda de las patologías comunes a todas las etiquetas sintácticas. En el ejemplo se muestra como para la etiqueta 'angulas' se han encontrado en la CDP hasta ocho patologías asociadas, para 'manchas-blancas' se asocian cuatro y para 'cabeza' un total de seis. De todo ese conjunto de patologías, tan solo las designadas con los subíndices 1, 5 y 7 son comunes a todas las etiquetas

### 3.4.2.2. El controlador borroso

Dentro de los sistemas borrosos se incluyen diversas teorías, como la teoría de los conjuntos borrosos, extensión de la teoría de conjuntos clásica, o la lógica borrosa, que pueden ser consideradas una ampliación de la lógica trivaluada (verdadero, falso e indeterminado). No obstante, la principal aplicación actual de la lógica borrosa son los sistemas de control que utilizan sus expresiones para formular reglas orientadas al control de sistemas (Martín del Brío y Sanz-Molina, 1997).

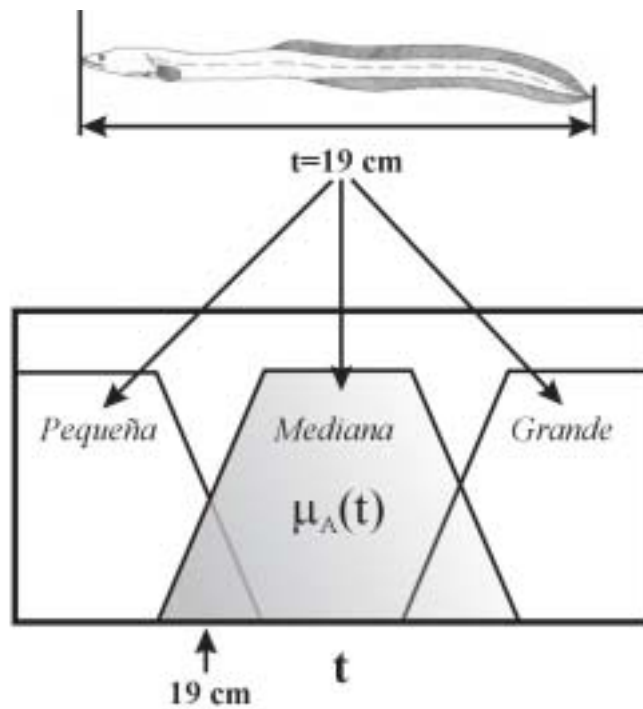
#### 3.4.2.2.1. Lógica borrosa y teoría de conjuntos borrosos. Conjuntos borrosos y funciones de inclusión

La teoría de conjuntos borrosos tiene su origen en la teoría clásica de conjuntos, con la diferencia de que a cada conjunto se le asigna una función de pertenencia definida como un número real entre 0 y 1. De esta forma, se introduce el concepto de conjunto borroso asociado a un determinado valor lingüístico, definido por una palabra, adjetivo o etiqueta lingüística. En nuestro caso, la etiqueta lingüística a la que vamos a asociar los conjuntos borrosos es la variable **[MPC]** y la frecuencia de aparición de las distintas patologías (**[FA]**).

Para cada conjunto borroso se define una función de pertenencia  $\mu_A(t)$ , que determina el grado en que la variable **t** está incluida en el concepto representado por la etiqueta **A** (Hilera y Martínez, 1995; Martín del Brío y Sanz-Molina, 1997) (figura III.4). Para la definición de estas funciones de pertenencia se utilizan convencionalmente familias de formas estándar. De entre todas las posibles (para más detalles Martín del Brío y Sanz-Molina, 1997), **SEDPA** es capaz de trabajar con las formas básicas triangular, trapezoidal, gaussiana y singleton.

La función triangular (**T**) puede definirse como (figura III.5a):

$$T(t;a,b,c)=\begin{cases} 0 & t < a \\ \frac{t-a}{b-a} & a \leq t \leq b \\ \frac{c-t}{c-b} & b \leq t \leq c \\ 0 & t > c \end{cases} \quad (3.3)$$



**Figura III.4.** Los conjuntos borrosos permiten agrupar objetos o sucesos por el valor de una cierta magnitud. Por ejemplo, las anguilas pueden ser agrupadas por tamaños. Así, si definimos el conjunto clásico de las anguilas de tamaño pequeño como las que miden menos de 20 cm, resulta que una anguila de 19 cm es pequeña mientras que una de 21 cm no lo es. Esta descripción que proporciona la teoría clásica de conjuntos no resulta perfectamente satisfactoria ya que los tamaños tan solo se diferencian en 2 cm. Una descripción en términos de conjuntos borrosos resulta más adecuada en casos de este tipo. Por ejemplo, podrían introducirse los términos pequeña, mediana y grande, y definirse mediante funciones de pertenencia, que al variar de forma continua en el rango 0 a 1 nos indicaría si una anguila es pequeña, pequeña tirando a mediana, claramente grande, etc.

La función trapezoidal (**D**) se define como (figura III.5b):

$$D(t;a,b,c,d)= \begin{cases} 0 & t < a \\ \frac{t-a}{b-a} & a \leq t \leq b \\ 1 & b \leq t \leq c \\ \frac{d-t}{d-c} & c \leq t \leq d \\ 0 & t > d \end{cases} \quad (3.4)$$

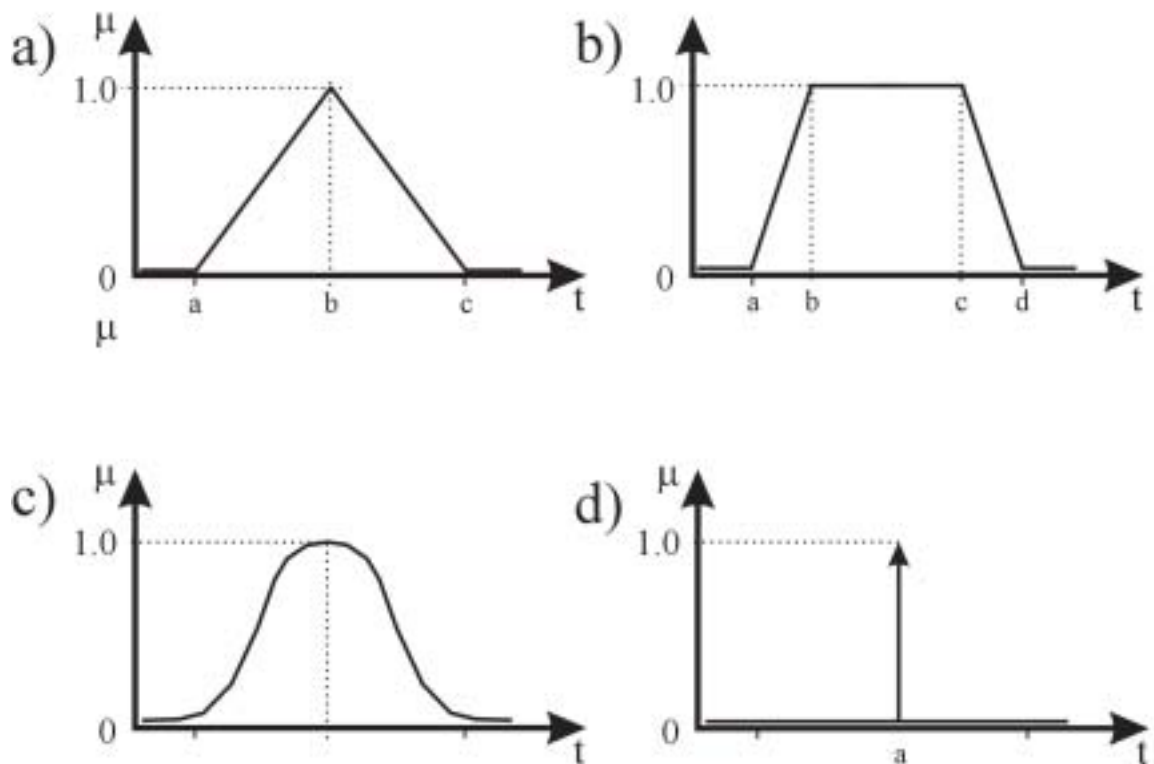
La función gaussiana (**G**) se define como (figura III.5c):

$$G(t) = e^{(-\lambda \cdot t^2)} \quad (3.5)$$

donde  $\lambda$  es una constante.

Finalmente, la función singleton (**S**) (figura III.5d) obtiene el valor 1 para un punto **a** y cero para el resto, es decir:

$$S(t) = \begin{cases} 1 & t = a \\ 0 & t \neq a \end{cases} \quad (3.6)$$



**Figura III.5.** a) Función de pertenencia de tipo triangular (T); b) función de pertenencia de tipo trapezoidal (D); c) función de pertenencia de tipo gaussiana (G) y d) función de tipo singleton (S)

Para que el controlador borroso funcione como tal, es necesario definir las particiones de las variables del controlador, entendiéndose por partición un conjunto de los conjuntos que se han definido para la variable **A**. Normalmente estas particiones son completas, impares y con niveles de solapamiento que oscilan entre el 20 % y el 50 %. El número de particiones también es variable, empleándose por regla general de 3 a 7 conjuntos ya que la complejidad no es excesiva y permiten una precisión suficiente para la descripción de los valores de la variable (Hilera y Martínez, 1995; Martín del Brío y Sanz-Molina, 1997; Tsoukalas y Uhrig, 1997). De esta forma en **SEDPA**, la variable



[MPC] se ha particionado como ‘Muy Poco Compartido’ (MP), ‘Poco Compartido’ (PC), ‘Normalmente compartido’ (NC), ‘Bastante Compartido’ (BC) y ‘Muy Compartido’ (MC), pudiéndose elegir la forma geométrica del conjunto (triangular, trapezoidal o gaussiano) y establecer un total de hasta 40 niveles distintos de solapamiento entre conjuntos al poder variar éste entre el 10 % y el 50 % (figura III.6a). Por otra parte, la variable [FA] se ha asociado a un conjunto de tipo singleton con cuatro particiones: ‘Muy Rara’ (MR), ‘Rara’ (RR), ‘Común’ (CM) y ‘Muy Común’ (MM) (figura III.6b).

### 3.4.2.2.2. Reglas de inferencia borrosa y Memoria Asociativa Borrosa (MAB). Asignación de niveles de creencia individuales y globales

Las reglas borrosas permiten la combinación de uno o más conjuntos borrosos de entrada ([MPC] y [FA]) a los cuales se asocia un conjunto borroso de salida, que en este caso es ‘la creencia’ de que la patología o patologías seleccionadas sea correcta [CR]. Al igual que [MPC], la variable de salida [CR] se ha particionado en cinco conjuntos borrosos (‘creencia Muy Alta’ -MA-, ‘creencia Alta’ -A-, ‘creencia Normal’ -N-, ‘creencia Baja’ -B- y ‘creencia Muy Baja’ -MB-) (figura III.6c) cuyos niveles de solapamiento pueden variar entre el 10 % y el 50 % y con geometrías de tipo triangular, trapezoidal o gaussiana. Los conjuntos borrosos de entrada se relacionan a través de expresiones lógicas como y, o, etc. Estas reglas permiten expresar el conocimiento que se dispone sobre la relación existente entre entradas y salidas. Para expresar este conocimiento de forma completa se necesitan generalmente varias reglas que en su conjunto forman lo que se conoce como base de reglas.

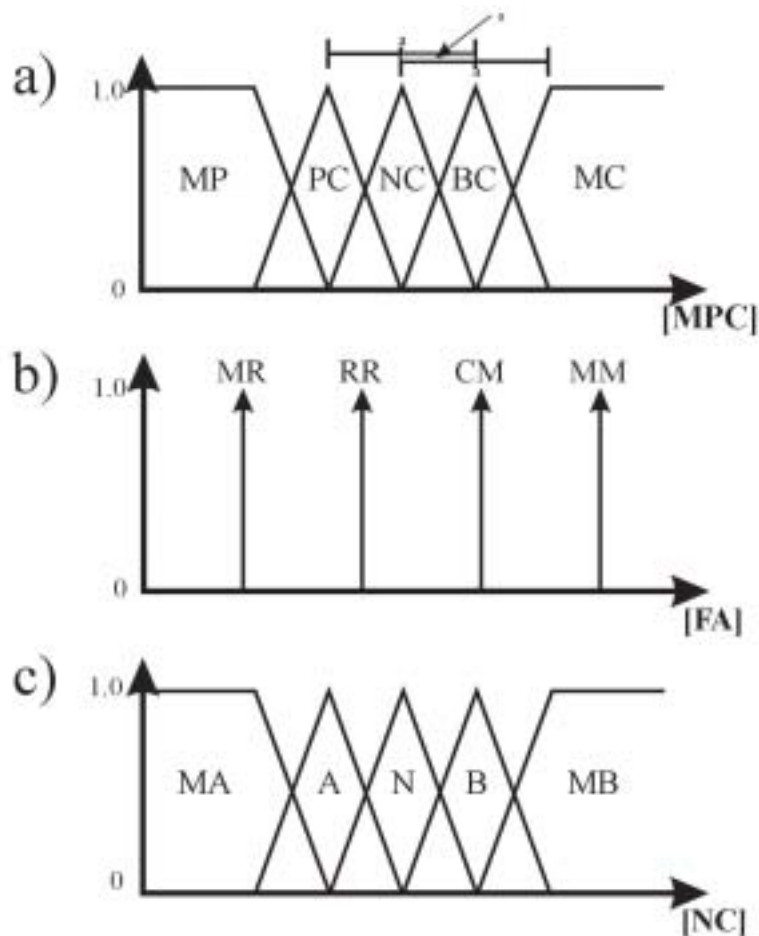
La base de reglas se puede representar bien como una tabla o como una memoria asociativa borrosa (MAB). Las MABs son matrices que representan la consecuencia de cada regla definida para cada combinación de dos entradas y permiten realizar una clara representación gráfica de las relaciones existentes entre dos variables lingüísticas de entrada y la variable lingüística de salida. Para ello requieren que se indiquen explícitamente todas las reglas que se pueden formar con estas dos variables de entrada.

Formalmente, una base de reglas borrosas es un conjunto de reglas  $R^{(L)}$  de la forma:

$$R^{(L)} : \text{SI } l_1 \text{ es } H_1^L \text{ y...y } l_n \text{ es } H_n^L \text{ ENTONCES } f \text{ es } K^L \quad (3.7)$$

donde  $H_1^L \dots H_n^L$  y  $K^L$  son conjuntos borrosos y  $l=(l_1\dots l_n)$  y  $f$  son variables lingüísticas.

SEDPA en este caso puede trabajar con MABs que no estén incluidas como código en la fuente de control del programa. Esto significa que la MAB puede ser cargada independientemente desde el exterior del programa, facilitando de esta forma la expresión



**Figura III.6.** a) Partición de la variable de entrada [MPC] en cinco conjuntos borrosos de tipo triangular: 'Muy Poco Compartido' (MP), 'Poco Compartido' (PC), 'Normalmente Compartido' (NC), 'Bastante Compartido' (BC) y 'Muy Compartido' (MC). Asimismo se muestra el nivel de solapamiento ( $s=0.5$  a  $50\%$ ) de los cinco conjuntos; b) partición de la variable de entrada [FA] en cuatro conjuntos de tipo singleton: 'Muy Rara' (MR), 'Rara' (RR), 'Común' (CM) y 'Muy Común' (MM) y; c) partición de la variable de salida [NC] en cinco conjuntos borrosos de tipo triangular: 'creencia Muy Alta' (MA), 'creencia Alta' (A), 'creencia Normal' (N), 'creencia Baja' (B) y 'creencia Muy Baja' (MB)

del conocimiento de distintos expertos y la obtención de resultados distintos en función de la **MAB** cargada. Concretamente, la versión inicial de **SEDPA** cuenta con tres **MABs** distintas construidas por otros tantos expertos (figura III.7). Ya que una **MAB** representa el conocimiento de un experto y la percepción que tiene éste sobre el dominio de trabajo, la comparación de las memorias asociativas nos puede permitir detectar *a priori* si van a existir diferencias significativas en los diagnósticos. El análisis puede ser abordado como un problema de distribución geográfica de poblaciones, debido a que 'la creencia' así como

su posición dentro de la matriz presenta una influencia definitiva sobre las conclusiones finales del sistema. De esta forma, se ha diseñado un test que es sensible a las diferencias de distribución de las creencias contenidas en las **MABs**.

		Mínimo Porcentaje Compartido MPC					
		MP	PC	NC	BC	MC	
Frecuencia de Aparición (FA)	MR	MA	A	N	B	MB	<b>Visser MAB</b>
	RR	A	N	B	B	B	
	CM	A	N	N	A	A	
	MM	MA	A	A	MA	MA	
		MP	PC	NC	BC	MC	
	MR	MA	A	N	B	MB	<b>Sole MAB</b>
	RR	MA	N	B	MB	MB	
	CM	MA	N	B	MB	MB	
	MM	MA	N	B	MB	MB	
		MP	PC	NC	BC	MC	
	MR	MA	A	B	MB	MB	<b>Carlos MAB</b>
	RR	MA	MA	N	B	MB	
	CM	MA	A	N	N	A	
	MM	MA	MA	N	MA	MA	

**Figura III.7.** Memorias Asociativas Borrosas (MABs) con las que cuenta la versión original de SEDPA. Cada MAB presenta una doble entrada: por una parte se encuentra el 'Mínimo Porcentaje Compartido' ([MPC]) que hace referencia a como de común es el síntoma lesión introducido al sistema, y por otra se encuentra la 'Frecuencia de Aparición' ([FA]) de la/s patología/s seleccionada/s en función de dicho síntoma lesión. La combinación de [MPC] y [FA] nos aporta 'la creencia' ([CR]) de que la/s patología/s seleccionada/s sean correctas. Por ejemplo, en la Visser-MAB para un [MPC]-BC (esto es, un síntoma lesión Bastante Compartido entre todas las patologías posibles) y una patología con una [FA]-RR (es decir, una patología cuya frecuencia de aparición es Rara), obtendremos una [CR]=B (la creencia de que esa patología sea correcta es Baja). Obsérvese además que en función de la MAB seleccionada, las salidas serán distintas para las mismas entradas

El test propuesto está basado en una generalización bivalente del test no paramétrico de Cramér Von Mises (Conover, 1980; Syrjala, 1996). La hipótesis nula del test supone que a lo largo de la matriz, la distribución normalizada de las creencias es la misma, mientras que la hipótesis alternativa supone que hay diferencias significativas en dicha distribución. Para construir el test en primer lugar se debe asumir que éste es independiente del nivel medio de creencia, por lo que se lleva a cabo la normalización de las creencias observadas dividiendo cada creencia por la suma de todas las creencias contenidas en la matriz, para lo cual se asigna a cada creencia un valor ( $MA=5$ ;  $A=4$ ;  $N=3$ ;  $B=2$ ;  $MB=1$ ). De esta forma (3.8) define las creencias observadas normalizadas, donde  $\mathbf{d}_i(\mathbf{x}_j, \mathbf{y}_k)$  es el valor asignado a la creencia asociada a los conjuntos borrosos de entrada  $\mathbf{x}_j$  e  $\mathbf{y}_k$  en la memoria asociativa  $\mathbf{i}$ .

$$\gamma_i(\mathbf{x}_j, \mathbf{y}_k) = \frac{\mathbf{d}_i(\mathbf{x}_j, \mathbf{y}_k)}{D_i} \quad D_i = \sum_{\forall j, \forall k} \mathbf{d}_i(\mathbf{x}_j, \mathbf{y}_k) \quad (3.8)$$

El valor de la función de distribución acumulada  $\Gamma_i(\mathbf{x}_j, \mathbf{y}_k)$  para el elemento  $(\mathbf{x}_j, \mathbf{y}_k)$  en la matriz  $\mathbf{i}$  es la suma de las creencias normalizadas  $\gamma_i(\mathbf{x}, \mathbf{y})$ , cuya localización  $(\mathbf{x}, \mathbf{y})$  es tal que  $\mathbf{x} \leq \mathbf{x}_j$  e  $\mathbf{y} \leq \mathbf{y}_k$ :

$$\Gamma_i(\mathbf{x}_j, \mathbf{y}_k) = \sum_{\forall \mathbf{x} \leq \mathbf{x}_j, \forall \mathbf{y} \leq \mathbf{y}_k} \gamma_i(\mathbf{x}, \mathbf{y}) \quad (3.9)$$

Según Cramér-von Mises, el estadístico del test ( $\Psi$ ) es la diferencia al cuadrado entre las dos funciones de distribución acumuladas:

$$\Psi = \sum_{\forall \mathbf{x}, \forall \mathbf{y}} [\Gamma_h(\mathbf{x}_j, \mathbf{y}_k) - \Gamma_i(\mathbf{x}_j, \mathbf{y}_k)]^2 \quad (3.10)$$

Ya que el estadístico  $\Psi$  es variable en función de la posición a partir de la cual se genera la función de distribución acumulada  $\Gamma_i(\mathbf{x}_j, \mathbf{y}_k)$ , es necesario introducir una modificación que considere el cálculo del estadístico  $\Psi$  a partir del valor promedio de los  $\Psi_c$  correspondientes a cada una de las cuatro esquinas de la memoria asociativa (Zimmerman, 1993):

$$\Psi = \frac{1}{4} \sum_{c=1}^4 \Psi_c \quad (3.11)$$

El nivel de significación del estadístico  $\Psi$  puede ser determinado usando la metodología propuesta por Edgington (1980). Bajo la hipótesis nula, para una localización en la memoria asociativa  $(\mathbf{x}_j, \mathbf{y}_k)$  la creencia acumulada  $\gamma_i(\mathbf{x}_j, \mathbf{y}_k)$ , siendo  $\mathbf{i}=1, 2$ , debe ser la misma. Así para un grupo de datos  $\mathbf{K}$ , la distribución del estadístico puede ser construida mediante el cálculo del valor del estadístico para los  $2^{\mathbf{K}}$  pares de permutaciones del grupo

de datos, de tal forma que, el valor de  $p_\alpha$  será la proporción de veces que el estadístico recalculado  $\Psi_r$  ha sido mayor o igual que el observado  $\Psi$ . Debido a que el número de permutaciones a realizar puede llegar a ser inabordable en la práctica (para una matriz con 25 datos:  $2^{25} > 33 \cdot 10^6$  permutaciones) es posible utilizar una aproximación a la metodología propuesta por Edgington (1980), mediante el cálculo del estadístico para un número lo suficientemente elevado de permutaciones elegidas al azar. Según Syrjala (1996) el número de permutaciones debe ser aproximadamente igual a mil. De esta forma el nivel de significación se calcula como:

$$p_\alpha = \frac{\text{Número veces que } \Psi_r > \Psi}{1000} \quad (3.12)$$

Una vez establecida la base de reglas es necesario definir lo que se conoce como dispositivo de inferencia borrosa, que permite la interpretación de las reglas SI...ENTONCES con el fin de obtener los valores de salida a partir de los valores actuales de las variables lingüísticas de entrada al sistema. En nuestro caso el método de obtención del valor de salida ha sido la regla del mínimo:

$$\mu_{A \rightarrow B}(t, y) = \text{Mínimo}[\mu_A(t), \mu_B(y)] \quad (3.13)$$

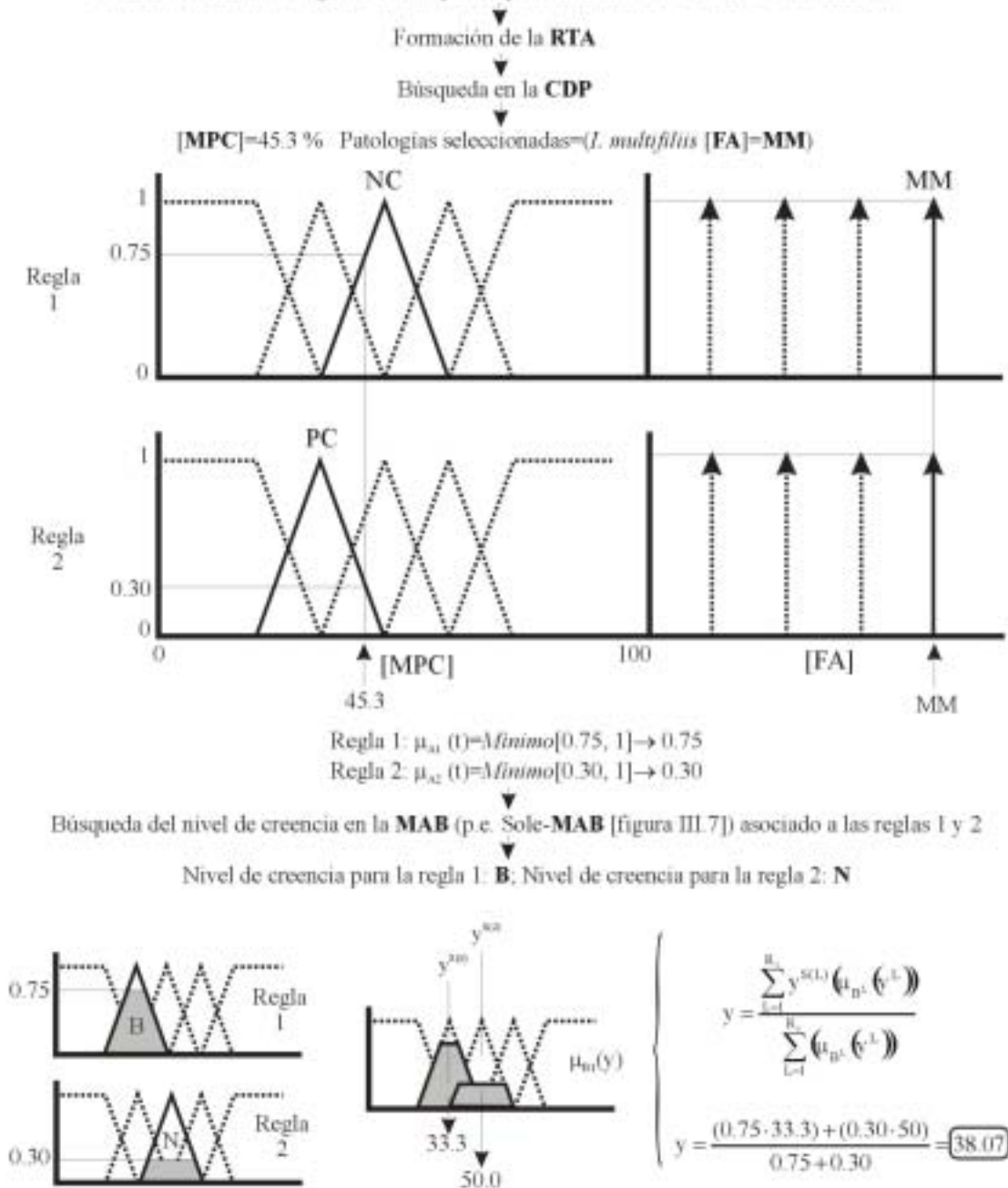
donde  $A \rightarrow B$  significa lo mismo que (3.7). Del mismo modo, también necesitamos una función que nos permita transformar un conjunto borroso, es decir, la salida del dispositivo de inferencia borrosa, en un valor no borroso. Para ello la función o desborrosificador utilizado ha sido la media de centros que se define como:

$$y = \frac{\sum_{L=1}^{R_c} y^{S(L)} (\mu_{B^L}(y^L))}{\sum_{L=1}^{R_c} (\mu_{B^L}(y^L))} \quad (3.14)$$

donde  $y$  es el valor no borroso de salida,  $R_c$  es el número de reglas de la base de reglas satisfechas,  $y^{S(L)}$  representa el centro del conjunto borroso de salida y  $\mu_{B^L}(y^L)$  viene definido según (3.13). El desarrollo completo de un ejemplo de lo anteriormente expuesto se muestra en la figura III.8.

Para un síntoma/lesión determinado, este proceso se repite tantas veces como número de patologías se hayan seleccionado para dicho síntoma/lesión, ya que para un mismo síntoma/lesión pueden encontrarse relacionadas más de una patología. De esta forma obtenemos lo que denominamos como grupo sintomático ( $gs$ ) asociado a un síntoma/lesión ( $gs = \{Pt_1, Pt_2, \dots, Pt_k\}$ ). El nivel de creencia global del grupo sintomático es el valor promedio de los niveles de creencia individuales de las patologías que lo componen.

**Síntoma/lesión:** Las angulas del tanque A4 presentan manchas blancas en la cabeza.



**Figura III.8.** Ejemplo en el que se muestra el proceso a través del cual es posible asignar niveles de creencia a patologías seleccionadas a partir de un síntoma/lesión. Para facilitar la comprensión se ha asignado tan solo una patología al síntoma introducido, a pesar de que por regla general, debido a lo poco específico de los síntomas, el número de patologías asociadas suelen ser más de una. En este caso el resultado final (un nivel de creencia de 38.07 sobre 100) se ha obtenido utilizando conjuntos borrosos de geometría triangular y con solapamientos del 50 %

Así:

$$m_n = \frac{\sum_{i=1}^k y_i}{k} \pm \sqrt{\frac{\sum_{i=1}^k (m_n - y_i)^2}{k-1}} \quad (3.15)$$

donde  $m_n$  es la creencia global del grupo sintomático para el síntoma/lesión  $n$  e  $y_i$  es la creencia individual asignada a las  $k$  patologías seleccionadas para ese grupo sintomático.

### 3.4.2.3. Teoría de Dempster-Shafer para la transmisión de la incertidumbre

El manejo de la incertidumbre es un punto crítico en el diseño de sistemas inteligentes para el diagnóstico (Wong y Yao, 1992; Yao y Wong, 1994; Szolovits, 1995). Como ya se comentó en el 3.1.2, la ausencia o dispersión de la información dificulta la aplicación de funciones de probabilidad bayesianas para la toma de decisiones, debido a la dependencia de éstas de probabilidades *a priori*, que a su vez, dependen del volumen de información disponible. Cuando no se dan estas condiciones lo recomendable es la aplicación de técnicas alternativas como la teoría de Dempster-Shafer o teoría de la evidencia (Dempster, 1968; Shafer, 1976; Denœux, 1995; Denœux 1997; Fixen y Mahler, 1997; Le Hégarat-Masclé *et al.*, 1997). Esta teoría asume que: a) la evidencia o creencia proporcionada por una proposición es un número real que oscila entre cero y uno y, b) las creencias no son necesariamente aditivas. De esta forma, la teoría de la evidencia puede entenderse como una extensión de la teoría de probabilidad bayesiana con la diferencia de que en vez de asignar una única estimación a una proposición individual, se asigna el intervalo [Creencia, Verosimilitud] donde la ‘Creencia’ mide la fuerza de la evidencia a favor de un conjunto de proposiciones y la ‘Verosimilitud’ definida con (3.16) mide el alcance con que la evidencia a favor de  $\neg gs$  deja espacio para la creencia en  $gs$  (Rich y Knight, 1994).

$$\text{Verosimilitud}(gs) = 1 - \text{Creencia}(\neg gs) \quad (3.16)$$

Al igual que en la teoría de Bayes, es necesario definir un universo de hipótesis mutuamente excluyentes o marco de discernimiento ( $\Theta$ ), que en este caso queda delimitado por las  $n$  patologías que el sistema es capaz de reconocer. Por otra parte, la función que permite la asignación de niveles de creencia es la función  $m$  de densidad de probabilidades que se define no sólo para los elementos de  $\Theta$ , sino para todos sus subconjuntos, de tal forma que:

$$\sum_{\forall i} m_n(gs_i) = 1 \quad (3.17)$$

donde  $m_n(\mathbf{gs}_i)$  mide la cantidad de creencia asignada a cada uno de los grupos sintomáticos  $i$  obtenidos tras la evidencia  $n$ . Así,  $m_n(\mathbf{gs}_i)$  es la creencia global de que la respuesta correcta esté en alguna parte del conjunto  $\mathbf{gs}$ . Para poder utilizar estos niveles de creencia es necesario definir una función que nos capacite para combinar las  $m_n(\mathbf{gs}_i)$  que surjan de múltiples fuentes de evidencia. La función que nos permite combinar dos funciones de creencia es la regla de combinación de Dempster que se define como:

$$m_3(Z) = \sum_{X \cap Y = Z} m_1(X) \cdot m_2(Y) \quad (3.18)$$

donde  $\mathbf{X}$ ,  $\mathbf{Y}$  y  $\mathbf{Z}$  son conjuntos de grupos sintomáticos. Esta expresión proporciona una nueva función de creencia que puede combinarse con otras asociadas a grupos sintomáticos que surgen como consecuencia de la aparición de nuevas evidencias.

Cuando alguno de los subconjuntos creados por combinación es vacío ( $\mathbf{X} \cap \mathbf{Y} = \emptyset$ ) la regla de Dempster proporciona resultados ‘intuitivos’ (Voorbraak, 1991; Zadeh, 1986), lo que se explica como una infrautilización de la regla (Shafer, 1982; Weiru y Bundy, 1994). Esto es debido a que la forma de calcular  $m_3(\mathbf{Z})$  garantiza que la suma de los valores individuales de cada grupo sintomático agrupado en  $\mathbf{Z}$  es uno, lo que implica que si se crea un conjunto vacío, parte de la creencia se asocia a dicho conjunto. Así, es necesario hacer una redistribución de la creencia asociada al conjunto vacío entre los no vacíos mediante la introducción de un factor de escala. De esta forma la nueva creencia de  $\mathbf{Z}$  sería:

$$m_3(Z) = \frac{\sum_{X \cap Y = Z} m_1(X) \cdot m_2(Y)}{1 - \sum_{X \cap Y = \emptyset} m_1(X) \cdot m_2(Y)} \quad (3.19)$$

donde (3.20) es la creencia asociada al conjunto vacío. Shafer (1976) y Murphy (1998) definen este fenómeno como Conflicto de Evidencia (**Con**) que puede ser medido con (3.21).

$$\sum_{X \cap Y = \emptyset} m_1(X) \cdot m_2(Y) \quad (3.20)$$

$$\text{Con}[m_1(X), m_2(Y)] = \text{Ln} \left( \frac{1}{1 - \sum_{X \cap Y = \emptyset} m_1(X) \cdot m_2(Y)} \right) \quad (3.21)$$

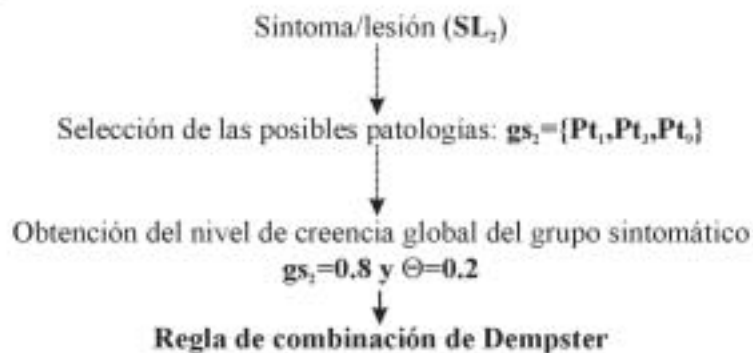
donde **Con** toma valores entre 0 y  $\infty$ . Un ejemplo de la metodología expuesta se muestra en las figuras III.9a y III.9b.



## Paso 1:



## Paso 2:



$$m_3(Z) = \sum_{X \cap Y = Z} m_1(X) \cdot m_2(Y)$$

SL <sub>1</sub> \ SL <sub>2</sub>	SL <sub>2</sub>	{Pt <sub>1</sub> , Pt <sub>2</sub> , Pt <sub>3</sub> }	(0.8)	{Θ}	(0.2)
{Pt <sub>1</sub> , Pt <sub>2</sub> , Pt <sub>3</sub> }	(0.6)	gs <sub>2</sub> ={Pt <sub>1</sub> , Pt <sub>2</sub> }	(0.48)	gs <sub>2</sub> ={Pt <sub>1</sub> , Pt <sub>2</sub> , Pt <sub>3</sub> }	(0.12)
{Θ}	(0.4)	gs <sub>2</sub> ={Pt <sub>1</sub> , Pt <sub>2</sub> , Pt <sub>3</sub> }	(0.32)	{Θ}	(0.08)

**Figura III.9a.** Ejemplo en el que se muestra la metodología utilizada para la transmisión de la incertidumbre. El paso 1 se puede desglosar en seis etapas, a través de las cuales llegamos a obtener un nivel de creencia global asociado a un grupo sintomático ( $gs_1$ ) a partir de un sintoma/lesión ( $SL_1$ ) introducido por el usuario. El paso 2 es igual al paso 1, con la diferencia de que el sintoma/lesión ( $SL_2$ ) introducido por el usuario es distinto al anterior, por lo que se obtiene un segundo grupo sintomático ( $gs_2$ ) (en este caso distinto al anterior). Tras el paso 2 tenemos por tanto dos fuentes de evidencia que debemos combinar para obtener grupos sintomáticos en cuyos niveles de creencia estén recogidas ambas. Este proceso se lleva a cabo mediante la regla de combinación de Dempster. De los resultados obtenidos se deduce que el grupo sintomático con un nivel de creencia mayor es el compuesto por las patologías  $Pt_1$  y  $Pt_2$  ( $gs_2=0.48$ ). Las flechas punteadas indican la existencia de las mismas etapas que en el paso 1

### Paso 3:



$SL_1 \cap SL_2$	$SL_3$	$\{Pt_1\}$ (0.9)	$\{\Theta\}$ (0.1)
$\{Pt_1, Pt_1\}$ (0.48)		$gs_2 = \{\emptyset\}$ (0.432)	$gs_{10} = \{Pt_1, Pt_1\}$ (0.048)
$\{Pt_1, Pt_1, Pt_1\}$ (0.32)		$gs_2 = \{Pt_1\}$ (0.288)	$gs_{10} = \{Pt_1, Pt_1, Pt_1\}$ (0.032)
$\{Pt_1, Pt_1, Pt_1\}$ (0.12)		$gs_2 = \{\emptyset\}$ (0.108)	$gs_{10} = \{Pt_1, Pt_1, Pt_1\}$ (0.012)
$\{\Theta\}$ (0.08)		$gs_2 = \{Pt_1\}$ (0.072)	$\{\Theta\}$ (0.008)

Conjuntos vacíos:  $gs_2$ ,  $gs_{10}$

$$\sum_{X \cap Y = \emptyset} m_2(X) \cdot m_3(Y) = 0.432 + 0.108 = 0.54 \quad \text{Con}[m_2(X) \cdot m_3(Y)] = 0.77$$

Conjuntos no vacíos semejantes:  $gs_2$  y  $gs_{10}$

$$\{Pt_1\} = 0.288 + 0.072 = 0.36$$

Escalado de las creencias de los conjuntos no vacíos

$$m_4(Z) = \frac{\sum_{X \cap Y = Z} m_2(X) \cdot m_3(Y)}{1 - \sum_{X \cap Y = \emptyset} m_2(X) \cdot m_3(Y)} \longrightarrow$$

$\{Pt_1\}$	(0.783)
$\{Pt_1, Pt_1\}$	(0.104)
$\{Pt_1, Pt_1, Pt_1\}$	(0.070)
$\{Pt_1, Pt_1, Pt_1\}$	(0.026)
$\{\Theta\}$	(0.017)

**Figura III.9b.** Tras la introducción del último sintoma/lesión ( $SL_3$ ) y por combinación con los grupos anteriores a través de la regla de Dempster, se produce la aparición de grupos sintomáticos vacíos ( $gs_2$  y  $gs_{10}$ ), lo que se explica como la aparición de conflicto entre los síntomas introducidos, por lo que se procede a escalar las creencias asignadas a los grupos no vacíos en función de la suma de creencias asignadas a grupos vacíos

3.4.2.4. *Test  $\chi^2$  modificado como elemento diferenciador de los grupos patológicos con mayor aportación a la creencia global*

Un problema que plantea la aplicación de la teoría de Dempster-Shafer en sistemas de diagnóstico es el reparto de la creencia total entre todos los grupos sintomáticos formados. Así, cuando el número de evidencias encontradas es grande y la información que contienen se encuentra muy solapada, la respuesta del sistema puede ser un gran número de grupos sintomáticos distintos a los que se asigna niveles de creencia bajos. De esta forma un nivel de creencia de 0.6 asignado a un grupo entre un total de dos tras la primera evidencia, puede ser significativamente menos importante que 0.35 tras la cuarta evidencia teniendo un total de diez grupos.

Para resaltar la importancia real de los niveles de creencia se ha empleado un test chi-cuadrado ( $\chi^2$ ) modificado, que consiste en considerar que las variables que se tratan de contrastar (niveles de creencia) son continuas y no discretas. La metodología se basa en tres aspectos: a) la comparación de las creencias obtenidas tras la introducción de una evidencia y las creencias asignadas a cada grupo si la creencia total se repartiese de forma equitativa entre todos ellos (ambas creencias expresadas en tanto por ciento), b) el cálculo del incremento de  $\chi^2$  entre una evidencia y otra y, c) el reescalado de la creencia asignada a un grupo en función del incremento porcentual de  $\chi^2$  y el número de patologías contenidas en el grupo, siendo cada uno de los sumandos de  $\chi^2$  las contribuciones de cada grupo a la discrepancia global. Es decir, dadas un par de evidencias  $\mathbf{i}$ ,  $\mathbf{j}$  siendo  $\mathbf{j}=\mathbf{i}+1$  podemos definir el incremento porcentual de chi-cuadrado ( $\Delta\chi^2$ ) como:

$$\Delta\chi^2 = 100 - \left( \frac{\left[ \sum_{\Delta g} \frac{(O_g(i) - E_g(i))^2}{E_g(i)} \right] \cdot 100}{\sum_{\Delta k} \frac{(O_k(j) - E_k(j))^2}{E_k(j)}} \right) \quad (3.22)$$

$$E_g(i) = \frac{100}{g} \quad \text{y} \quad E_k(j) = \frac{100}{k} \quad (3.23)$$

donde  $O_g(i)$  son las creencias originales en tanto por ciento asignadas a través de la regla de Dempster a los  $g$  grupos sintomáticos creados tras la evidencia  $\mathbf{i}$  y  $O_k(j)$  son las creencias originales asignadas en tanto por ciento a los  $k$  grupos sintomáticos creados tras la evidencia  $\mathbf{j}$ . Este escalado será aplicado siempre y cuando el test sea significativo ( $p_\alpha < 0.05$ ).

Una vez calculado  $\Delta\chi^2$  podemos establecer el nuevo valor escalado [ $\mathbf{m}_j^*(\mathbf{gs}_k)$ ] para cada grupo establecido por la evidencia  $\mathbf{j}$ , de tal forma que:

$$m_j^*(gs_k) = \left[ (1 - m_j(gs_k)) \cdot \left( \frac{\Delta\chi^2}{L_k} \cdot 0.01 \right) \right] + m_j(gs_k) \quad (3.24)$$

donde  $m_j(gs_k)$  es la creencia original asignada al grupo  $k$  tras la evidencia  $j$  y  $L_k$  es el número de patologías que compone el grupo sintomático  $k$  (figura III.10). Para este nuevo valor de creencia los límites de confianza al 95 % quedan definidos como:

$$LC_s = \frac{m_j^*(gs_k)}{(100 - m_j^*(gs_k) + 1) \cdot F_s + m_j^*(gs_k)} \quad (3.25)$$

$$LC_i = \frac{(m_j^*(gs_k) + 1) \cdot F_i}{(100 - m_j^*(gs_k) + 1) \cdot F_i + (m_j^*(gs_k) + 1) \cdot F_i} \quad (3.26)$$

donde  $LC_s$  es el límite superior,  $LC_i$  es el límite inferior y:

$$F_s = F_{\alpha/2} [2 \cdot (100 - m_j^*(gs_k) + 1); 2 \cdot m_j^*(gs_k)] \quad (3.27)$$

$$F_i = F_{\alpha/2} [2 \cdot (m_j^*(gs_k) + 1); 2 \cdot (100 - m_j^*(gs_k))] \quad (3.28)$$

donde  $F_{\alpha/2}$  es el valor de la  $F$  de Snedecor para los grados de libertad especificados (Martín-Andrés y Luna del Castillo, 1994).

### 3.4.3. Subsistema explicativo

La capacidad de brindar explicaciones es una cualidad importante que facilita la plena aceptación de los sistemas expertos por parte de los usuarios. Debido a ello, empieza a ser frecuente que sistemas expertos de alto nivel incluyan, como una utilidad más, un subsistema que permite al usuario obtener una justificación adecuada del razonamiento seguido por el motor de inferencia (Amador-Hidalgo, 1996; De Carolis *et al.*, 1996). Sin embargo hasta el momento, en la mayoría de los sistemas el subsistema explicativo no existe o sólo ofrece un nivel de explicación para la persona que lo diseña o para un experto en el dominio. En estos casos, el dispositivo de justificación es una simple traza del proceso de resolución y razonamiento que se ha seguido hasta el momento en el que el usuario requiere la información.

En **SEDPA** el subsistema explicativo está constituido por dos módulos independientes. El primero de ellos establece el ¿por qué? determinadas patologías han sido favorecidas o perjudicadas durante el proceso de valoración. Esto es posible gracias a la existencia de un grupo de reglas (**A-A**) del tipo SI...ENTONCES que establecen si los datos

Sintoma/lesión (SI)

Grupo	$O_p(i)$ $m_i(\%)$	$E_p(i)$	$\frac{[O_p(i)-E_p(i)]^2}{E_p(i)}$
$gs_{p1}$	37.58	25.00	6.33
$gs_{p2}$	27.87	25.00	0.32
$gs_{p3}$	19.84	25.00	1.06
$gs_p$	14.71	100/4	4.23

$\Sigma \downarrow$   
100       $\Sigma \downarrow$   
100

Número de grupos=4

$\chi^2=11.96$   
 $p=0.0075$   
 $gl=3$

Sintoma/lesión(SL<sub>j</sub>)

Grupo	$O_p(j)$ $m_j(\%)$	$E_p(j)$	$\frac{[O_p(j)-E_p(j)]^2}{E_p(j)}$
$gs_{k-5}$ $L_{k-5}=2$	37.62	16.66	26.36
$gs_{k-4}$	27.90	16.66	7.58
$gs_{k-3}$	12.90	16.66	0.82
$gs_{k-2}$	9.61	16.66	2.98
$gs_{k-1}$	6.84	16.66	5.78
$gs_k$	5.07	100/6	8.06

$\Sigma \downarrow$   
100       $\Sigma \downarrow$   
100

Número de grupos=6

$\chi^2=51.60$   
 $p<0.001$   
 $gl=5$

$$\Delta\chi^2 = 100 - \left( \frac{\sum_{p \in i} \frac{(O_p(i) - E_p(i))^2}{E_p(i)} \cdot 100}{\sum_{k \in j} \frac{(O_k(j) - E_k(j))^2}{E_k(j)}} \right) = 76.82\%$$

En este caso, el nivel de creencia compensado para el grupo  $gs_{k-5}$ , suponiendo  $L_{k-5}=2$

$$m_j^*(gs_{k-5}) = \left[ (-m_j(gs_{k-5})) \left( \frac{\Delta\chi^2}{L_{k-5}} \cdot 0.01 \right) \right] + m_j(gs_{k-5}) = 0.6158$$

**Figura III.10.** Una vez aplicada la regla de Dempster y obtenidas las creencias de todos los grupos se procede a reescalar las creencias en función de la evidencia total introducida y del número de patologías que compone cada grupo.  $E_p(i)$  y  $E_p(j)$  representan las creencias (expresadas en tanto por ciento) asignadas a cada grupo si la creencia global se repartiese de forma equitativa entre todos los grupos obtenidos tras la evidencia  $i$  y  $j$ , respectivamente. Para facilitar la comprensión del proceso, el cálculo del nivel de creencia compensado que se muestra en el ejemplo se ha llevado a cabo tan solo para el grupo sintomático  $gs_{k-5}$  obtenido tras la introducción del sintoma/lesión (SL<sub>j</sub>). De esta forma, si originalmente el nivel de creencia para el grupo  $gs_{k-5}$  es de 0.3762 (37.62 % de la creencia total), teniendo en cuenta el incremento de chi-cuadrado éste es equivalente a 0.6158

incluidos en la Base de Conocimiento del Dominio Secundaria (**CDS**) soportan la aparición de unas patologías u otras (apartado 3.4.6). Cada una de esas reglas está asociada a una breve explicación que justifica la regla. Por otra parte, una vez restringido el rango de posibles patologías es posible determinar cuales son las posibles medidas correctoras (¿qué hacer?) y como llevarlas a cabo (¿cómo hacerlo?) (figura III.1). El primer bloque de información se suministra a través de una ventana hipertexto (Jimison *et al.*, 1992; Firriolo *et al.*, 1996) que permite además la visualización de fotografías tipo para facilitar la identificación. El segundo bloque requiere nuevamente la información contenida en la **CDS** (tamaño medio, número de individuos, tratamientos previos y determinadas variables ambientales dependiendo de la patología) para poder establecer los fármacos a aplicar durante el tratamiento así como las dosis adecuadas.

#### 3.4.4. Subsistema de sugerencia o motor proposicional

En algunas ocasiones el proceso de diagnóstico puede verse interrumpido lo que supone la obtención de conclusiones poco satisfactorias. Esto puede ser debido a que el usuario no tenga la experiencia necesaria para extraer la información contenida en las muestras o bien que, a pesar de que el usuario sea un experto, no se observen aparentemente otros síntomas/lesiones que puedan ser aportados al sistema. Para paliar dentro de lo posible este fenómeno, **SEDPA** incorpora un subsistema que llegado el diagnóstico a este punto puede proporcionar una guía que el usuario opcionalmente puede seguir. Esta guía se traduce en dos tipos de preguntas que el sistema puede hacer al usuario: ¿Se observa algún síntoma/lesión localizado en....? ó ¿Se observa el síntoma/lesión....?. Es decir, el sistema aprovecha la información introducida hasta el momento construyendo dos matrices **CD<sub>sel</sub>** (cuya dimensión es: Número de filas en **CD** x **n+1**) y **CCL<sub>sel</sub>** (cuya dimensión es: Número de filas en **CCL** x **n+1**) para las **n** patologías seleccionadas tras el último síntoma introducido. De ambas matrices podemos obtener los lugares que son susceptibles de presentar algún síntoma/lesión así como los síntomas/lesiones que de forma más restrictiva se asocian con una patología en concreto. Así, para un grupo sintomático compuesto por **n** patologías, siendo **n>1**, la propuesta del sistema tras la introducción de un determinado síntoma será:

Si la matriz seleccionada es **CD<sub>sel</sub>**: ¿Se observa el síntoma/lesión **cd<sub>i,0</sub>**?

sí y sólo sí:

$$\sum_{j=1}^n cd_{i,j} = 1 \quad (3.29)$$

donde  $cd_{i,0}$  es la etiqueta sintáctica correspondiente a la fila  $i$  de la matriz  $CD_{sel}$ , y  $cd_{i,j}$  para  $j>0$  es el valor (0 ó 1) que asocia la etiqueta  $cd_{i,0}$  con la patología  $j$ .

Si la matriz seleccionada es  $CCL_{sel}$ : ¿Se observa algún síntoma/lesión localizado en  $ccl_{i,0}$ ?  
sí y sólo sí:

$$\sum_{j=1}^n ccl_{i,j} = 1 \quad (3.30)$$

donde  $ccl_{i,0}$  es la etiqueta sintáctica correspondiente a la fila  $i$  de la matriz  $CCL_{sel}$ , y  $ccl_{i,j}$  para  $j>0$  es el valor (0 ó 1) que asocia la etiqueta  $ccl_{i,0}$  con la patología  $j$ .

Frente a este tipo de pregunta, el usuario puede responder SI o NO, teniendo en cuenta que sólo cuando se cumple (3.31) la respuesta aportada por el usuario es totalmente eliminatoria:

$$\sum_{h=1}^{nt} cd_{i,h} = 1 \Rightarrow \sum_{j=1}^n cd_{i,j} = 1 \quad \text{ó} \quad \sum_{h=1}^{nt} ccl_{i,h} = 1 \Rightarrow \sum_{j=1}^n ccl_{i,j} = 1 \quad (3.31)$$

siendo  $nt$  el total de patologías reconocidas por el sistema. Esto quiere decir que cuando la asociación entre una etiqueta lingüística y el total de patologías es única y, a su vez, es encontrada por el subsistema proposicional, la respuesta proporcionada por el usuario se traduce en eliminar todas las patologías del grupo sintomático, excepto aquella que se encuentra asociada a la etiqueta lingüística cuando la respuesta del usuario es SI, o se elimina dicha patología si la respuesta es NO. Sin embargo, si se cumple (3.32) la respuesta no es totalmente eliminatoria por lo que hay que establecer un factor de impacto de la respuesta sobre las creencias individuales del grupo sintomático. Ya que la importancia de la respuesta depende de lo común que sea la etiqueta lingüística entre todas las patologías podemos definir el factor de impacto con (3.33):

$$\sum_{j=1}^n cd_{i,j} = 1 \quad \text{y} \quad \sum_{h=1}^{nt} cd_{i,h} = s \quad \text{ó} \quad \sum_{j=1}^n ccl_{i,j} = 1 \quad \text{y} \quad \sum_{h=1}^{nt} ccl_{i,h} = r \quad (3.32)$$

$$\text{factor de impacto} = 100 - \left( \frac{s}{nt} \cdot 100 \right) \quad \text{ó} \quad 100 - \left( \frac{r}{nt} \cdot 100 \right) \quad (3.33)$$

siendo  $1 < s \leq (nt - n - 1)$  y  $1 < r \leq (nt - n - 1)$ .

### 3.4.5. Subsistema de aprendizaje

En la creación de cualquier sistema experto resulta ser de gran importancia el proceso mediante el cual se transfiere la experiencia desde el experto humano al programa. En infinidad de sistemas expertos el módulo de adquisición de conocimiento suele presentarse, por regla general, como una herramienta segregada de la interfaz del experto (Pesonen *et al.*, 1994; Kentala *et al.*, 1995; Firriolo *et al.*, 1996; Ivandic *et al.*, 1996; Nguyen *et al.*, 1996; Kahn *et al.*, 1997; Ohayon *et al.*, 1997). Resulta incuestionable el hecho de que disponer de un buen módulo de adquisición facilita de manera considerable la modificación y actualización de las bases de conocimiento. Con la utilización de este módulo es posible la creación y posterior edición de la base de conocimiento, además de poder efectuar una supervisión continuada de la sintaxis incorporada a las unidades en las que se ha estructurado el conocimiento (Amador-Hidalgo, 1996).

El subsistema de aprendizaje de **SEDPA** tiene básicamente dos propósitos: a) encontrar a través del módulo de interpretación de lenguaje natural términos que no se encuentren recogidos en las bases de conocimiento del dominio principal y, b) la posibilidad de actualizar las bases de datos frente a la aparición de nuevas enfermedades que sean susceptibles de afectar a la producción de anguilas o la modificación de los parámetros que definen a las ya existentes. El primero de los casos se lleva a cabo en el momento de la introducción de los síntomas/lesiones y la formación de la **RTA**. Cuando no es posible situar sintácticamente alguna de las palabras detectadas tras el análisis morfológico, el sistema pregunta al usuario si la palabra en cuestión es nueva, si presenta algún error gramatical o si es sinónimo de alguna ya conocida. En este último caso, el programa presenta un listado de posibles sinónimos que ocupan la misma posición sintáctica dentro de la oración. Cuando lo que se pretende es aumentar o disminuir la base de datos existente, **SEDPA** dispone de una interfaz amigable para poder introducir el conocimiento en la base.

### 3.4.6. Valoración del diagnóstico

Como ya se comentó en el apartado 3.4.3 una herramienta más con la que cuenta **SEDPA** para establecer con precisión el diagnóstico es el análisis de los datos contenidos en la base de conocimiento del dominio secundaria (**CDS**). Así, cualquier diagnóstico con un nivel de creencia alto que se establezca a través de los datos suministrados por el usuario puede tener una valoración baja debido a que las características físico-químicas que hacen posible la aparición en la explotación del agente o agentes causantes de la enfermedad no se cumplen según la **CDS**, o al menos los datos no se encuentran disponibles en ésta.



En **SEDPA** la valoración del diagnóstico responde a un esquema de funcionamiento similar al de otros sistemas expertos tales como **GOLDFINDER** (Hawkes, 1992). Así cada condición de aparición de la enfermedad se asocia a una puntuación de tal forma que se compara la puntuación máxima ( $PT_{Máxima}$ ) que podría alcanzar la enfermedad si se cumpliesen todas las condiciones de aparición con la suma de la puntuación asociada a todas las condiciones de aparición satisfechas ( $PT_{Satisfecha}$ ):

$$PT = \frac{\sum PT_{Satisfecha}}{PT_{Máxima}} \cdot 100 \quad (3.34)$$

Esta puntuación está relacionada con una escala de valoración y se trata de la misma forma que durante el proceso de asignación de creencias individuales (apartado 3.4.2.2.1). Por tanto, dependiendo del valor de **PT** la valoración podrá pertenecer a una única categoría o a varias en un determinado porcentaje. Las categorías establecidas son: Muy Prometedora, Prometedora, Especulativa, Muy Especulativa y Nada Prometedora.

### 3.4.7. Validación del sistema

Un requerimiento básico para la aceptación por parte de los especialistas de un sistema de diagnóstico clínico es que el sistema sea capaz de proporcionar un grupo de posibles patologías con probabilidades altas sobre las que el usuario pueda elegir en función de su propia experiencia (Molino *et al.*, 1996). Se ha seguido esta estrategia en la validación del sistema, de tal forma que se asume como condición de fiabilidad que la patología establecida como causante del problema se encuentre entre las seleccionadas como compatibles.

De acuerdo con tal estrategia, las conclusiones del sistema han sido clasificadas como incorrectas cuando la patología de referencia o ‘estándar oro’ (Wyatt y Spiegelhalter, 1990; Van Diest *et al.*, 1994) fue omitida mientras todas las hipótesis seleccionadas como compatibles fueron aceptadas como fiables, correctas cuando el ‘estándar oro’ se encontraba incluida entre las patologías seleccionadas por el sistema en el grupo sintomático con el nivel de creencia más alto, y aproximadas cuando el ‘estándar oro’ fue englobada en grupos sintomáticos creados tras la última evidencia pero con niveles de creencia inferiores. Para determinar el poder de discriminación del sistema en los dos niveles de aceptación se ha empleado la tasa de acierto (**TA**) (Molino *et al.*, 1996) definida como:

$$TA = \frac{1}{N_c} \sum_{\forall i=j} C_i \quad (3.35)$$

donde  $N_c$  es el número casos presentados al sistema y  $C_i$  es el número de respuestas aportadas por el sistema que coinciden con el 'estándar oro'  $j$  (Turing, 1950; Cohen, 1968; Firriolo *et al.*, 1996; Ohayon *et al.*, 1997).

Un paso esencial en la validación de un sistema experto es la correcta elección del número de casos a utilizar durante la validación. Ya que el objetivo del diseñador del sistema es mostrar que éste alcanzará al menos un determinado umbral de acierto (lo que desde un punto de vista estadístico significa que con un determinado límite de confianza  $LC$ , una especificación es satisfecha en al menos la fracción  $F$  de una muestra), la validación del sistema aportará afirmaciones del tipo: '*El sistema experto diagnosticó correctamente la/s patología/s presentes en la planta en el 80 % de las ocasiones con unos límites de confianza del 95 %*'. Dado que el diseñador o usuario final puede desear que el sistema alcance un nivel de acierto mínimo *a priori*, se puede utilizar este valor para averiguar el número de casos necesarios para validar el sistema. De esta forma si  $F$  es el nivel de acierto mínimo aceptable,  $F_0$  es el nivel de acierto medio deseado y  $LC$  es el límite de confianza requerido, se puede definir el error estándar experimental ( $ES_{exp}$ ) como:

$$ES_{exp} = \frac{F_0 - F}{z(LC)} \quad (3.36)$$

donde  $z(LC)$  es el número tal que el área comprendida bajo la distribución normal entre  $z(LC)$  y  $-z(LC)$  es  $LC$ . A partir de aquí se puede obtener el número mínimo de casos a evaluar ( $N_c$ ):

$$N_c \geq \frac{F_0 \cdot (1 - F)}{ES_{exp}^2} \quad (3.37)$$

Los  $N_c$  casos necesarios para validar el sistema fueron juzgados por tres expertos con distinta experiencia en el diagnóstico patológico de peces. A estos se les permitió la consulta de libros, atlas y/o artículos científicos simulándose, por tanto, una situación de diagnóstico diaria. Para normalizar la información extraída por los expertos durante el proceso de diagnóstico y hacerla comparable con las conclusiones del sistema, se empleó un informe que fue entregado al experto (figura III.11).

Dicho informe constaba de cuatro secciones: a) Características físico-químicas y ambientales donde se desenvolvían los individuos, incluyendo la fecha del examen, condiciones de confinamiento, tanque, fecha de entrada a la planta, temperatura, oxígeno disuelto, amoniaco, pH, salinidad, turbidez, tamaño medio de los individuos, peso medio y densidad del tanque, b) signos clínicos, lesiones macroscópicas e histológicas. En esta sección se especificaban las observaciones llevadas a cabo por un experto que no intervendría durante el proceso de validación, c) entrada de síntomas/lesiones en **SEDPA**,

CASO 10  
 FECHA DEL EXAMEN: 15/02/99  
 CONDICIONES DE CONFINAMIENTO: Cerrado control  
 TANQUE: 41  
 FECHA DE ENTRADA EN LA PLANTA:

TEMPERATURA (°C):  
 OXÍGENO DISUELTO (mg/l):  
 AMONÍACO (mg/l):  
 PH: 7  
 SALINIDAD (‰):  
 TURBIDEZ (npt):

TAMAÑO MEDIO (mm):  
 PESO MEDIO (g):  
 DENSIDAD DEL TANQUE (individuos/l):

**SEÑALES CLÍNICAS, LESIONES MACROSCÓPICAS E HISTOLÓGICAS**

*El análisis de los individuos contenidos en el tanque al morir, la presencia de gusanos adultos en la redada filar y róbica, lo que probablemente es responsable de una disentería en la misma zona. Es importante que solo se pudieran distinguir hemerogonias inmaduras.*

**ENTRADA DE SÍNTOMAS/LESIONES EN SEDPA:**

- LAS ANGLAS DEL TANQUE A1 PRESENTAN OJITOS ESFERICOS EN LA VERICELA HERRIAR
- LAS ANGLAS DEL TANQUE A1 PRESENTAN OJITOS ESFERICOS EN EL REJES
- LAS ANGLAS DEL TANQUE A1 PRESENTAN UNA HERRIPLATA EN LA VERICELA
- LAS ANGLAS DEL TANQUE A1 TIENEN HERRIPLATAS EN LAS BRANQUEAS

**CUESTIONARIO:**

Por favor escriba las conclusiones que puede ser hechas en base a los datos aportados anteriormente usando la terminología de la parte referida.

1. Tipo de agente patógeno:

Bacterias (especificar género o especie cuando sea posible):

Virus:

Parasitos: *Platyhelmin*

Parásitos fisis-químicos:

Otros:

2. En función de los síntomas/lesiones descritos, ¿cómo es la creencia en cuanto a su diagnóstico? (Marque con un círculo la respuesta que crea más apropiada)

Página 1

Página 2

----- Muy Baja Baja Normal Alta Muy Alta -----

1. Asigne un rango cualitativa a su nivel de creencia en una escala del 1 al 100 (por ejemplo, la creencia de que su diagnóstico es correcto se encuentra entre el 80-90 %)

*80 - 90 %*

4. En función de los síntomas/lesiones descritos así como en función de las condiciones físico-químicas que se dan cuando el/los agente(s) patógeno(s) apropiado(s), ¿cómo valoraría su diagnóstico? (Marque con un círculo la respuesta que crea más apropiada)

Nada Probable    Muy Especulativo    Especulativo    Probable    Muy Probable

5. ¿Cree que los datos introducidos en el sistema se corresponden con los descritos por el experto?

SI

NO

Comentarios:

**OBSERVACIONES:**

*Por los síntomas podría tratarse de una parasitosis causada por Platyhelmin, aunque está sujeto en el futuro a más grandes pruebas.  
 Podría tratarse de una enfermedad causada por parásitos parásitos o quistes.*

Figura III.11. Tipo de encuestas suministrada a los expertos humanos

es decir, los datos suministrados al sistema y, d) un cuestionarios que constaba de cinco apartados.

El primero indicaba al experto que señalase el tipo de agente patológico que según los datos señalados en la sección 'b' era el más probable según su propia experiencia, teniendo que especificar si era posible, el género y la especie. El segundo apartado preguntaba al experto '¿cómo es la creencia de que su diagnóstico sea correcto?' sólo en función de los síntomas o lesiones aportados. Para facilitar la respuesta y el posterior análisis, el experto tuvo que elegir entre las siguientes categorías: muy baja, baja, normal, alta y muy alta. De la misma forma, en el tercer apartado se le pedía al experto que indicase un rango numérico comprendido entre 0 y 100. En el cuarto apartado se le pidió que en función de los síntomas y lesiones descritos así como en función de las condiciones físico-químicas reinantes en el momento del examen, valorase finalmente su diagnóstico. Al igual que en el apartado 2, para facilitar el proceso se le permitió escoger entre cinco posibles opciones: nada prometedor, muy especulativo, especulativo, prometedor y muy prometedor. Finalmente se le preguntó sobre la semejanza entre la descripción de los signos clínicos y los datos introducidos al sistema. Se ha optado por el uso de esta metodología ya que ha sido utilizada con éxito en la evaluación de otros sistemas expertos (Berry y Hart, 1990; Hernando *et al.*, 1996; Díez *et al.*, 1997; Kentala *et al.*, 1998).

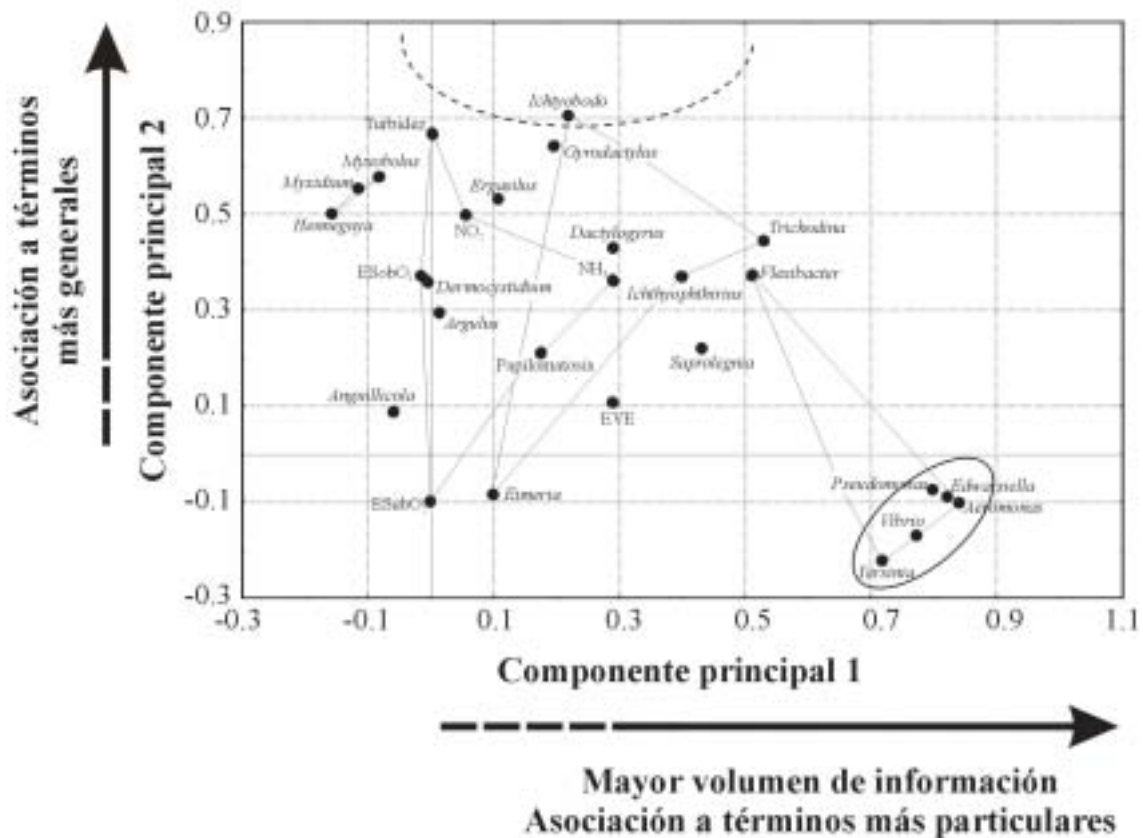
### **3.5. RESULTADOS**

#### **3.5.1. Información contenida en las bases de datos del dominio principal y secundario**

Como ya se comentó en la sección 3.4.1, una vez extraída la información de las fuentes bibliográficas (libros especializados en patología de peces, revistas especializadas y datos recogidos en la propia planta) la base de Conocimiento del Dominio Principal en su versión inicial ha quedado constituida por un total de 517 términos relacionados con 27 patologías (tablas III.1a y III.1b), de los cuales el 23.2 % son términos básicos y el resto (76.8 %) se encuentran englobados dentro del diccionario.

La información codificada en la base de Conocimiento del Dominio Principal puede ser analizada testeando si las proporciones o frecuencias de etiquetas sintácticas difieren significativamente entre las distintas patologías. Para ello y debido a la arquitectura de las matrices puede utilizarse el test de Cochran que es una extensión del test  $\chi^2$  de McNemar (Ruiz-Maya-Pérez *et al.*, 1990; StatSoft, 1995). De esta forma, el test de Cochran mostró la existencia de marcadas diferencias significativas ( $Q=185.592$ , g.l.=26,  $p_\alpha < 0.001$ ) en cuanto a la cantidad de información asociada a cada patología. Por otra parte un análisis de componentes principales (PCA) indicó la existencia de un gradiente de volumen y distribución de información que permitió diferenciar grupos de patologías asociadas a

etiquetas sintácticas semejantes. Asimismo se pudieron identificar planos de patologías pertenecientes a grupos etiológicos parecidos. Se puede observar como exceptuando a *F. columnaris*, el resto de bacterias se encuentran agrupadas constituyendo el único conjunto de patologías que presenta una correlación significativa con el componente principal 1. Aunque para el resto de patologías (excepto *Ichtyobodo* sp.) las correlaciones no fueron significativas con ninguno de los dos componentes principales, se pudieron establecer planos de asociación para los factores ambientales, protozoarios, myxozoarios, helmintos y artrópodos. Dichos planos se encontraron más o menos solapados indicando la similitud de los términos asociados a todas las patologías, a pesar de que en el componente principal 2 se puede establecer un gradiente sobre el tipo de estructuras sintácticas asociadas a patologías (figura III.12).



**Figura III.12.** Análisis de componentes principales sobre la totalidad de información contenida en las bases de Conocimiento del Dominio Principal en función del tipo de patología. El óvalo con línea continua indica correlación significativa con el componente principal 2. El semióvalo con línea discontinua indica correlación significativa con el componente principal 1. Los polígonos de línea continua indican planos de patologías pertenecientes al mismo grupo. Las flechas paralelas a los componentes indican gradientes.

### 3.5.2. Número de casos utilizados para la evaluación

Según la ecuación (3.36) es necesario establecer *a priori* el nivel de acierto mínimo aceptable  $F$ , el nivel de acierto medio deseable  $F_0$  así como los límites de confianza requeridos (LC) para calcular el número de casos mínimos para evaluar el sistema. El nivel de acierto mínimo aceptable se fijó al 65 %, mientras que el nivel de acierto medio deseable se estableció en un 80 %. De esta forma, para unos límites de confianza del 95 % y aplicando la ecuación 3.36 se obtiene que el error estándar experimental es del 0.0765. Establecido el error estándar experimental y sustituyéndolo en la ecuación (3.37) obtenemos que el número mínimo de casos a evaluar es de 27. Así, para evaluar el sistema se han seleccionado un total de 29 casos, de tal forma que nos aseguramos que los resultados se encuentren dentro de los límites de aceptación. En la tabla III.4 se detallan cada uno de los 29 casos o ‘estándar oro’ presentados tanto a **SEDPA** como a cada uno de los expertos humanos.

### 3.5.3. Análisis de las respuestas aportadas por los expertos humanos

Los resultados indican que la mayor tasa de acierto (86.2 %) es alcanzada por el *experto 2* en el tipo de agente patógeno (virus, bacteria, hongo, parásito, factor físico-químico y otros). Tanto el *experto 1* como el *experto 3* obtuvieron tasas de acierto inferiores a las conseguidas por el *experto 2* (Exp.1=79.3 %; Exp.3=65.5 %). En los tres expertos las tasas de acierto disminuyen significativamente (hasta 45 puntos para el *experto 2*) cuando, una vez seleccionado el tipo de agente patógeno se intenta determinar el género y la especie, alcanzándose de esta forma en el mejor de los casos (*experto 1*) tasas de acierto cercanas al 50 %. A pesar de que a los tres expertos se les suministró el mismo tipo de encuestas, la forma de responder fue distinta. Mientras que el *experto 1* en todos los casos llevó a cabo una respuesta concreta incluso en aquellos casos en los que el número de síntomas era escaso o la información suministrada era claramente ambigua o insuficiente, los *expertos 2* y *3* respondieron de forma más general llegando incluso en ocasiones a omitir su respuesta cuando se le requería el posible género del agente etiológico.

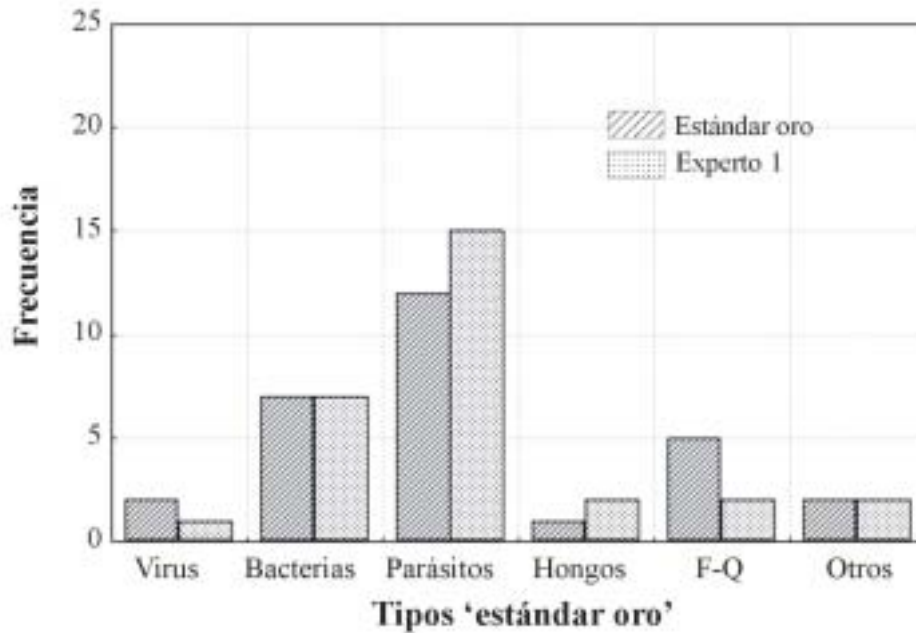
En el caso del *experto 1*, un test  $\chi^2$  no mostró diferencias significativas entre las frecuencias de agentes patógenos reales (frecuencia de cada tipo de agente patógeno en el conjunto ‘estándar oro’) y las frecuencias obtenidas como consecuencia de la respuesta del experto ( $\chi^2=6.600$ , g.l.=5,  $p_\alpha=0.252$ ) (figura III.13a). Resultados similares se obtienen

**Tabla III.4.** Número de casos presentados tanto a los expertos humanos como a SEDPA. Se indica el 'estándar oro' para cada caso así como el número de síntomas dentro de cada caso

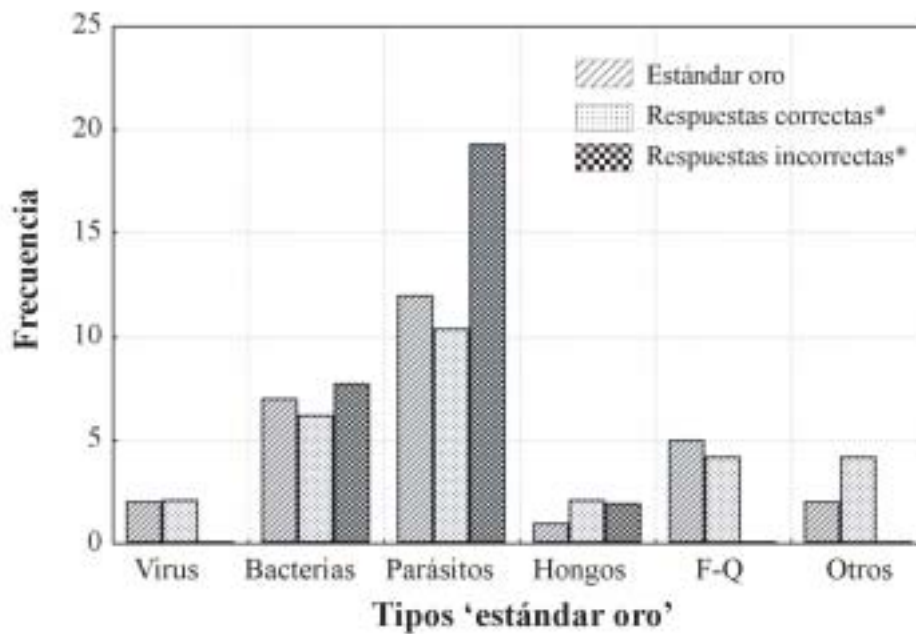
Caso	Estándar oro (tipo-género)	Número de síntomas
1	Virus-EVE	6
2	Virus-EVE	4
3	Bacteria- <i>Vibrio</i>	7
4	Bacteria- <i>Vibrio</i>	3
5	Bacteria- <i>Aeromonas</i>	8
6	Bacteria- <i>Pseudomonas</i>	5
7	Bacteria- <i>Flexibacter</i>	6
8	Bacteria- <i>Edwardsiella</i>	7
9	Bacteria- <i>Yersinia</i>	5
10	Otro- <i>Dermocystidium</i>	4
11	Hongo- <i>Saprolegnia</i>	3
12	Otro-Papilomatosis	2
13	Parásito- <i>Dactylogyrus</i>	4
14	Parásito- <i>Trichodina</i>	3
15	Parásito- <i>Myxidium</i>	4
16	Parásito- <i>Myxobolus</i>	4
17	Parásito- <i>Ichtyobodo</i>	4
18	Parásito- <i>Henneguya</i>	4
19	Parásito- <i>Eimeria</i>	1
20	Parásito- <i>Ergasilus</i>	6
21	Parásito- <i>Argulus</i>	5
22	Parámetro físico-químico-NO <sub>2</sub>	2
23	Parámetro físico-químico-NH <sub>3</sub>	3
24	Parámetro físico-químico-Saturación de O <sub>2</sub>	4
25	Parámetro físico-químico-Subsaturación de O <sub>2</sub>	1
26	Parámetro físico-químico-Turbidez	1
27	Parásito- <i>Gyrodactylus</i>	4
28	Parásito- <i>Ichthyophthirius</i>	3
29	Parásito- <i>Anguillicola</i>	3

cuando se comparan las frecuencias de respuesta del experto y las frecuencias obtenidas cuando se seleccionan las respuestas correctas e incorrectas de forma independiente (Frecuencia de acierto:  $\chi^2=2.202$ , g.l.=5,  $p_\alpha=0.821$ ; Frecuencia de fallo:  $\chi^2=3.302$ , g.l.=5,  $p_\alpha=0.654$ ) (figura III.13b). Asimismo, al contrastar las frecuencias de las creencias de acierto en el diagnóstico del *experto 1* frente a las frecuencias de creencias de acierto en el diagnóstico en las respuestas acertadas y falladas, tampoco se encontraron diferencias significativas (Respuestas acertadas:  $\chi^2=5.333$ , g.l.=4,  $p_\alpha=0.255$ ; Respuestas falladas:

a)



b)



**Figura III.13.** a) Frecuencia de respuestas del experto 1 frente a la frecuencia del 'estándar oro'. b) Frecuencia del 'estándar oro' frente a la frecuencia de respuestas correctas e incorrectas del experto 1. \* Valores escalados a 29 casos para permitir la comparación



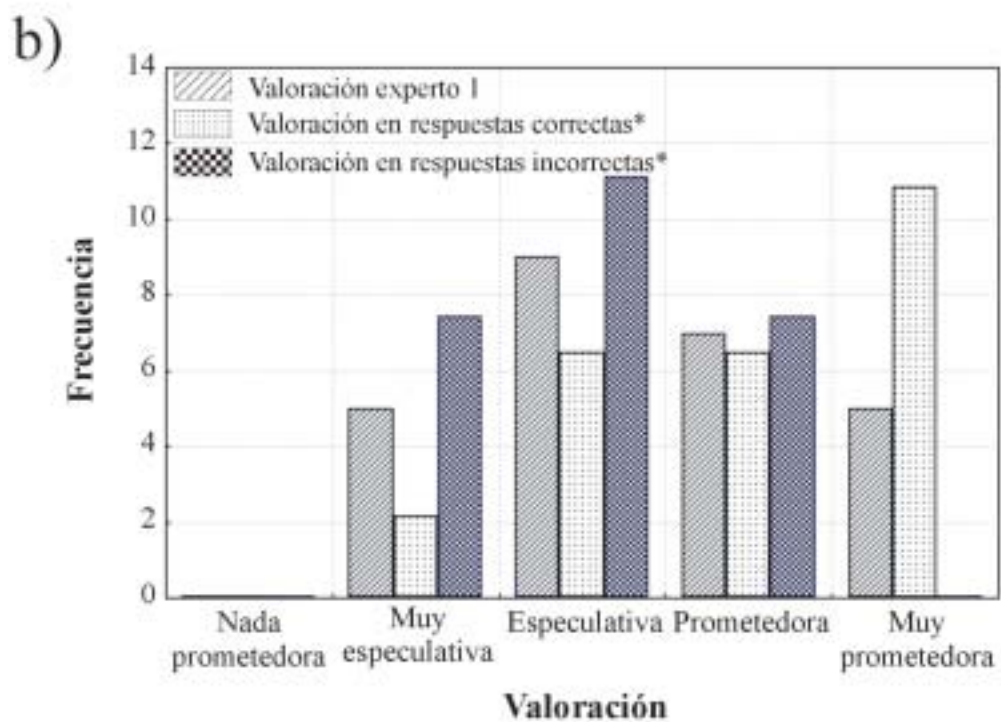
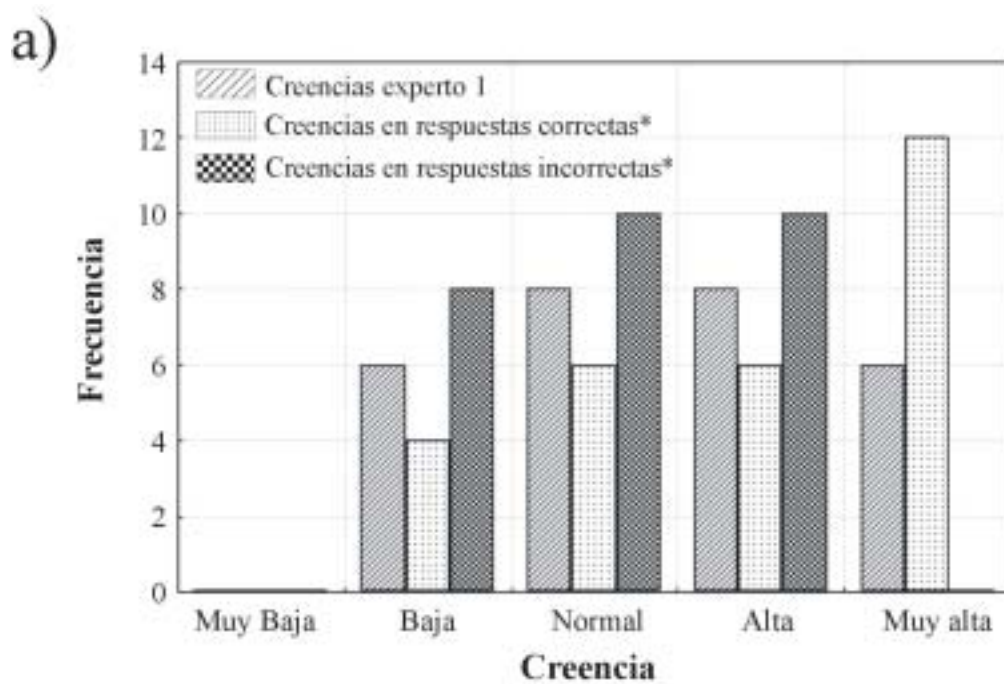
$\chi^2=1.300$ , g.l.=4,  $p_\alpha=0.861$ ) (figura III.14a). El mismo resultado se obtiene cuando lo que se compara son las frecuencias de la valoración final del diagnóstico frente a las frecuencias de la valoración final del diagnóstico en las respuestas acertadas y falladas (Respuestas acertadas:  $\chi^2=7.846$ , g.l.=4,  $p_\alpha=0.097$ ; Respuestas falladas:  $\chi^2=1.231$ , g.l.=4,  $p_\alpha=0.873$ ) (figura III.14b).

Al llevar a cabo la regresión lineal entre los valores de creencia cualitativa asignados por el *experto 1* en función de tan solo síntomas o lesiones y los valores medios de creencia cuantitativa, se obtienen resultados bastante aceptables ( $R^2=0.898$ ,  $F(1,27)=228.19$ ,  $p_\alpha<0.001$ ,  $n=29$ ) (figura III.15a). De la misma forma, también se obtienen resultados estadísticamente significativos en la regresión lineal entre los valores medios de creencia cuantitativa y los valores cualitativos de la valoración final ( $R^2=0.927$ ,  $F(1,27)=331.32$ ,  $p_\alpha<0.001$ ,  $n=29$ ) (figura III.15b).

Por último, el *experto 1* marcó en todas las ocasiones la casilla SI cuando respondió a la pregunta indicada en el apartado cinco del cuestionario.

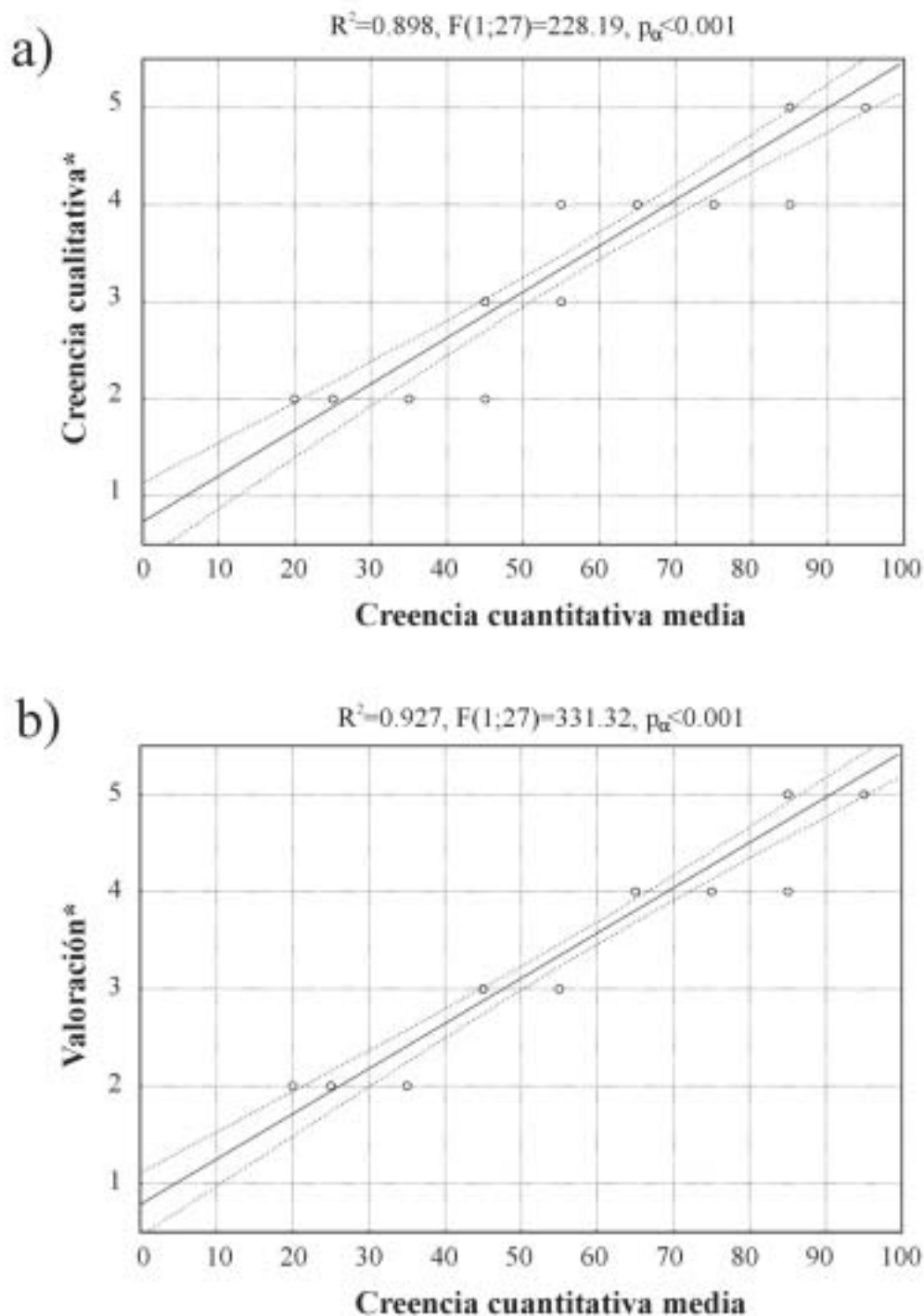
Los análisis llevados a cabo sobre los resultados del *experto 2* mostraron la presencia de diferencias significativas cuando se comparan las frecuencias de los agentes patógenos o etiológicos reales con los obtenidos tras el diagnóstico ( $\chi^2=37.167$ , g.l.=6,  $p_\alpha<0.001$ ) (figura III.16a). El *experto 2* introduce un nuevo grupo en el que engloba aquellos casos en los que el diagnóstico estaba compuesto por más de un agente patógeno (combinación de grupos). A pesar de que este nuevo grupo está constituido por un total de 14 casos, la mayor contribución a la discrepancia general del test se debe al grupo de los factores físico-químicos. La discrepancia de este grupo junto con el de nueva inclusión supone un 81.1 % del valor final del estadístico  $\chi^2$ . Por otra parte, las frecuencias de las respuestas del *experto 2* no mostraron diferencias significativas cuando éstas se compararon con las frecuencias de las repuestas acertadas ( $\chi^2=11.414$ , g.l.=6,  $p_\alpha=0.076$ ). Sin embargo, al repetir el proceso pero con las respuestas falladas, el test si mostró diferencias significativas provocadas fundamentalmente por la contribución del grupo de los parásitos y el grupo compuesto ( $\chi^2=41.126$ , g.l.=6,  $p_\alpha<0.001$ ) (figura III.16b).

Las frecuencias de las creencias frente a la frecuencias de las creencias de las respuestas acertadas no mostraron diferencias significativas en el caso del *experto 2* ( $\chi^2=8.517$ , g.l.=6,  $p_\alpha=0.074$ ). El mismo resultado se obtiene cuando la comparación se lleva a cabo con las respuestas falladas ( $\chi^2=3.896$ , g.l.=6,  $p_\alpha=0.420$ ) (figura III.17a). Sin embargo, cuando el mismo proceso se realiza con la valoración final del diagnóstico, es decir, contrastar frecuencias de la valoración del *experto 2* frente a las frecuencias de las valoraciones en las respuestas correctas e incorrectas, en ambos casos se obtienen diferencias significativas (Respuestas acertadas:  $\chi^2=12.892$ , g.l.=6,  $p_\alpha=0.012$ ; Respuestas falladas:  $\chi^2=10.960$ , g.l.=6,  $p_\alpha=0.027$ ) (figura III.17b).

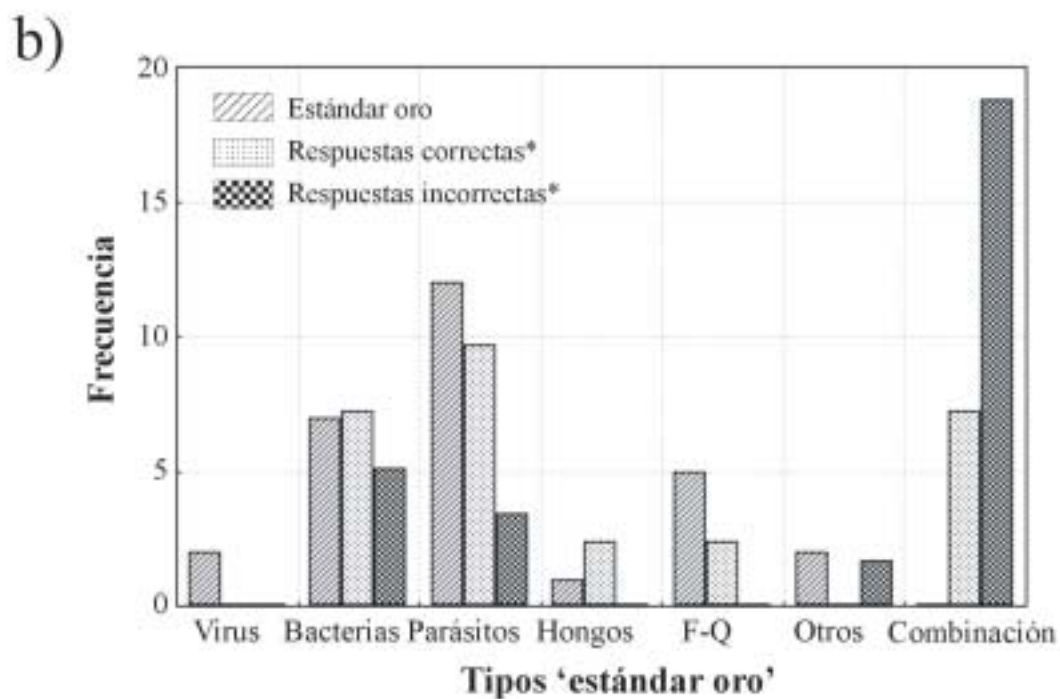
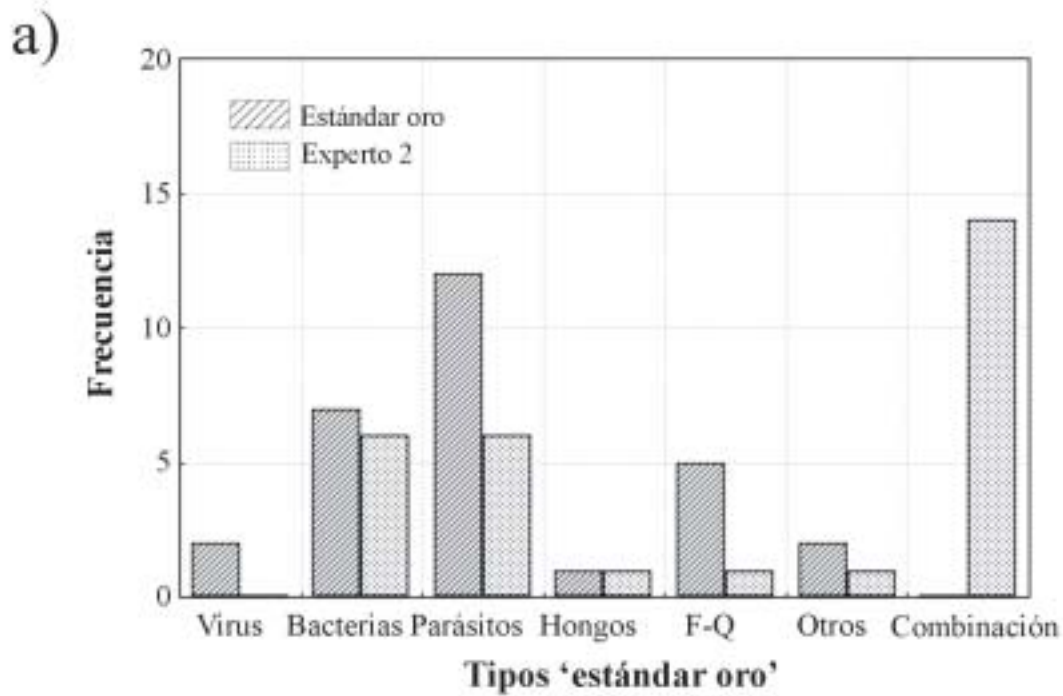


**Figura III.14.** a) Frecuencias de las creencias del experto 1 frente a las frecuencias de creencias en las respuestas correctas e incorrectas. b) Frecuencias de las valoraciones del experto 1 frente a las frecuencias de las valoraciones en respuestas correctas e incorrectas.

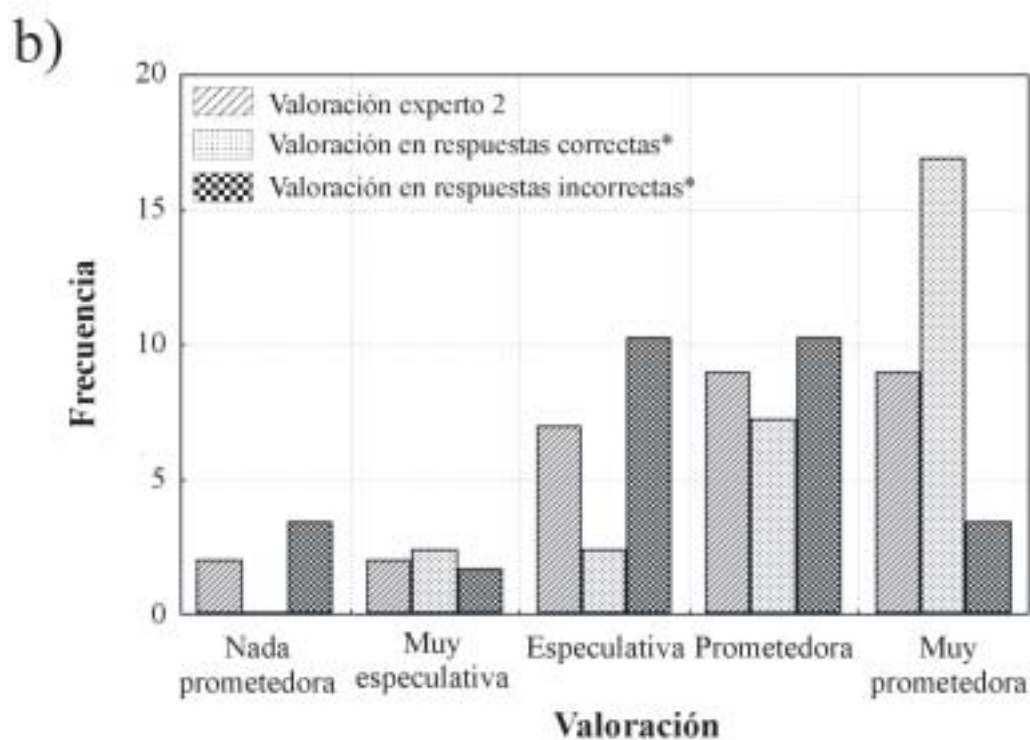
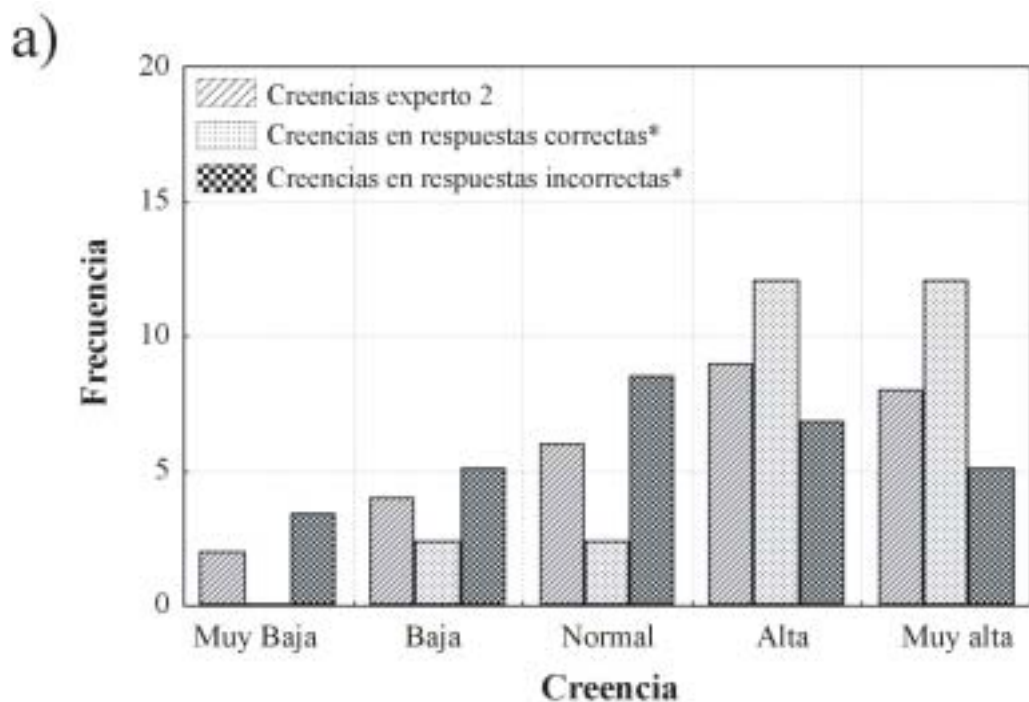
\* Valores escalados a 29 casos para permitir la comparación



**Figura III.15.** a) Regresión entre la creencia cualitativa (muy baja, baja, normal, alta y muy alta) y la creencia cuantitativa media aportada por el experto 1. b) Regresión entre la valoración final del diagnóstico (nada prometedor, muy especulativo, especulativo, prometedor y muy prometedor) y la creencia cuantitativa media aportada por el experto 1. \* Para permitir los cálculos se ha asignado a cada etiqueta cualitativa un valor cuantitativo de tal forma que muy baja=nada prometedor=1, baja=muy especulativa=2, normal=especulativa=3, alta=prometedor=4 y muy alta=muy prometedor=5



**Figura III.16.** a) Frecuencia de respuestas del experto 2 frente a la frecuencia del 'estándar oro'.  
 b) Frecuencia del 'estándar oro' frente a la frecuencia de respuestas correctas e incorrectas del experto 2. \* Valores escalados a 29 casos para permitir la comparación



**Figura III.17.** a) Frecuencias de las creencias del experto 2 frente a las frecuencias de creencias en las respuestas correctas e incorrectas. b) Frecuencias de las valoraciones del experto 2 frente a las frecuencias de las valoraciones en respuestas correctas e incorrectas.  
\* Valores escalados a 29 casos para permitir la comparación

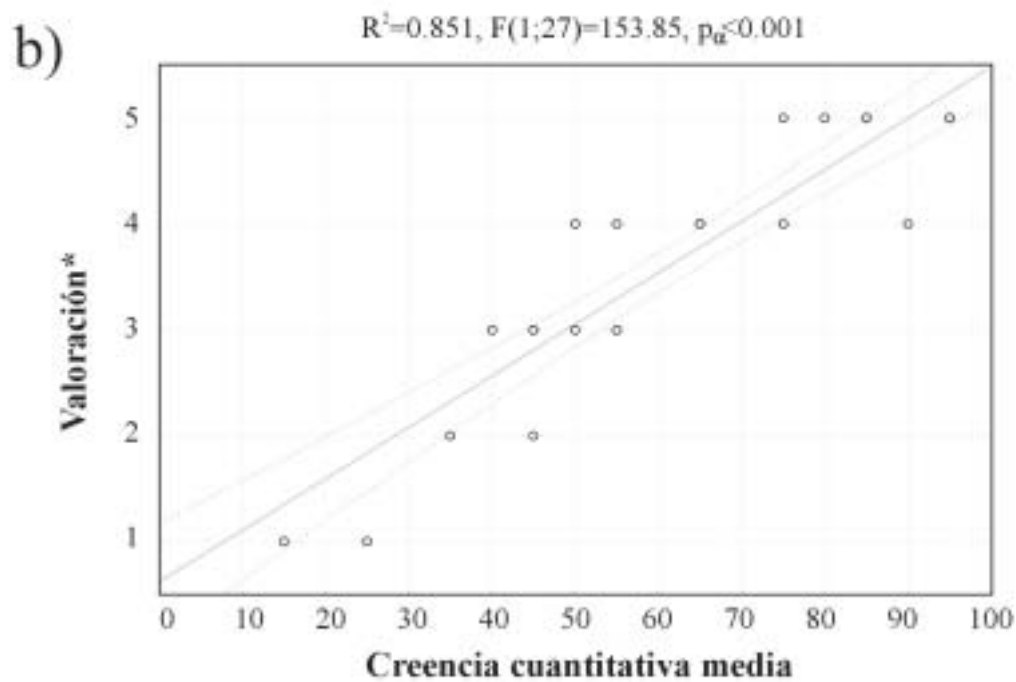
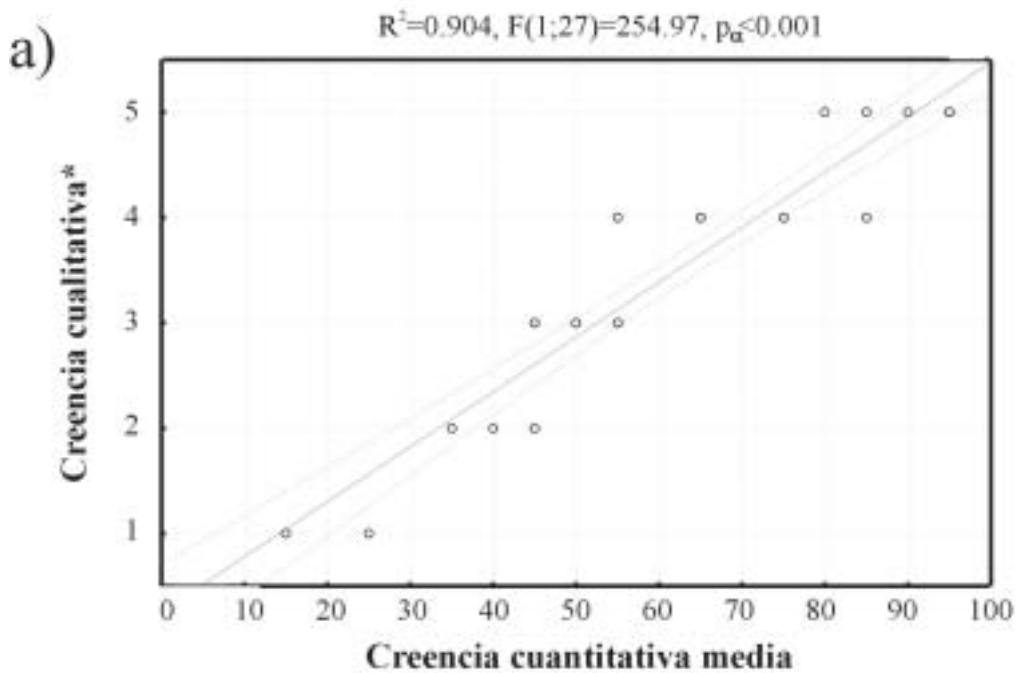
Al igual que en el *experto 1*, las regresiones entre las creencias cuantitativas y cualitativas así como entre las creencias cuantitativas y la valoración final del diagnóstico, fueron estadísticamente aceptables (Cuantitativa-Cualitativa:  $R^2=0.904$ ,  $F(1,27)=254.97$ ,  $p_\alpha<0.001$ ,  $n=29$ ; Cuantitativa-Valoración final:  $R^2=0.851$ ,  $F(1,27)=153.85$ ,  $p_\alpha<0.001$ ,  $n=29$ ) (figuras III.18a y III.18b).

Finalmente de la misma forma que en el caso del *experto 1*, el *experto 2* marcó en el 100 % de las ocasiones la opción 'SI' cuando se le preguntó sobre la similitud entre la descripción de los síntomas y lesiones hechas por el experto humano y la introducción de los datos a **SEDPA**.

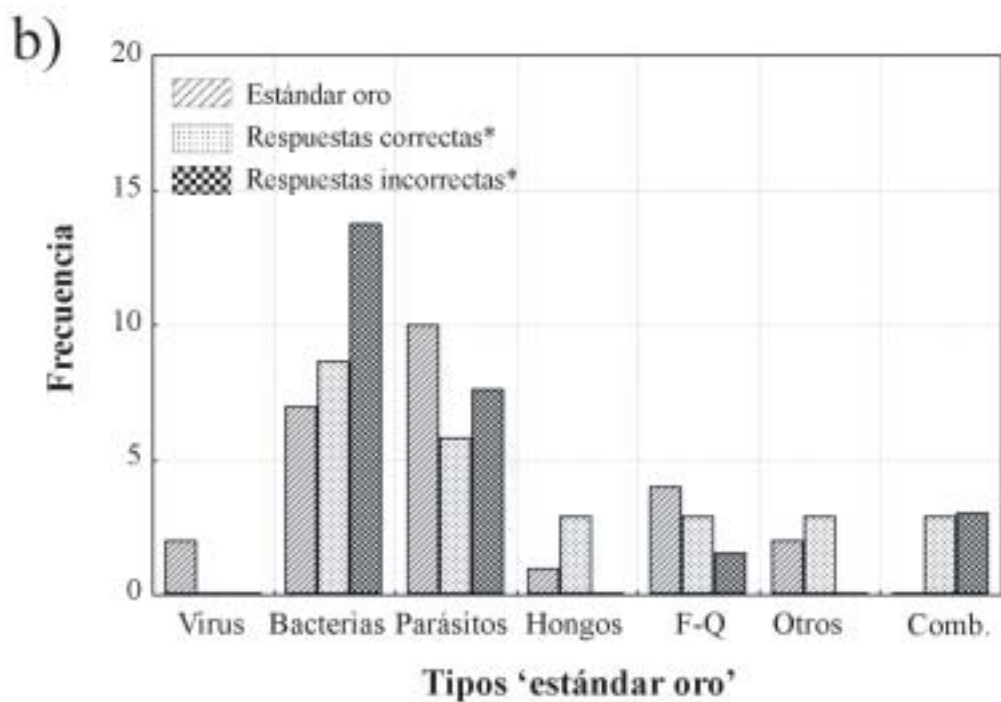
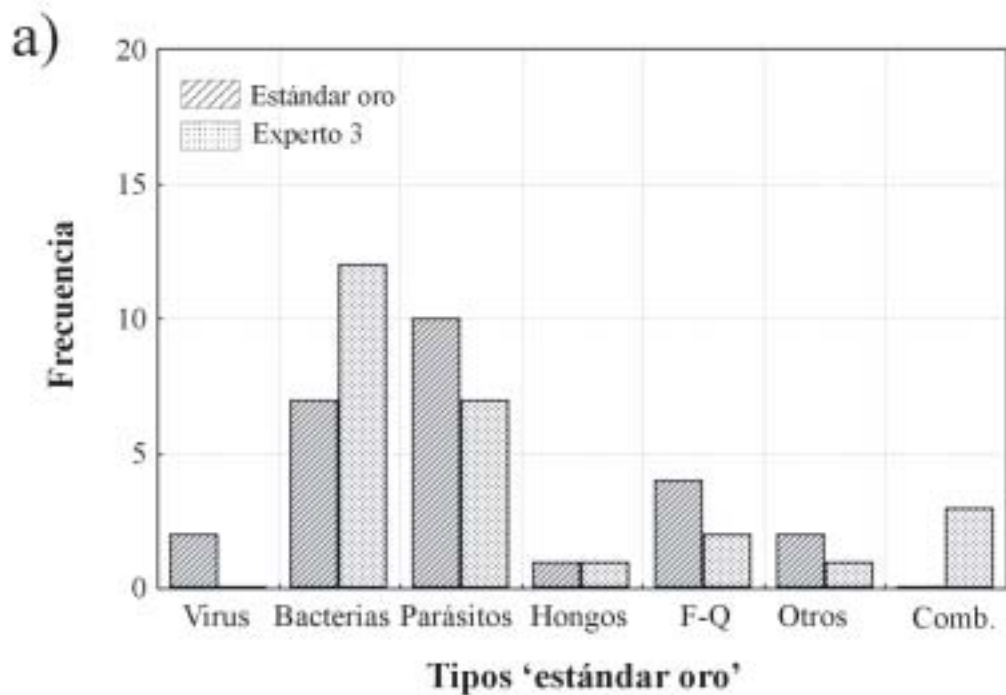
Para el *experto 3*, al igual que el *experto 1*, encontramos que las frecuencias correspondientes a cada tipo de agente etiológico no mostraron diferencias significativas con las frecuencias obtenidas tras la respuesta del experto ( $\chi^2=9.369$ , g.l.=6,  $p_\alpha=0.154$ ) (figura III.19a) denotando una tendencia a responder de forma correcta. En este caso la mayor contribución a la discrepancia total se lleva a cabo por el tipo etiológico combinación (32 %) que no había sido incluido como 'estándar oro'. Cuando comparamos las frecuencias de los tipos etiológicos con las frecuencias de los tipos etiológicos en las respuestas correctas e incorrectas del *experto 3* tampoco se encontraron diferencias significativas (Respuestas correctas:  $\chi^2=8.231$ , g.l.=6,  $p_\alpha=0.222$ ; Respuestas incorrectas:  $\chi^2=11.098$ , g.l.=6,  $p_\alpha=0.085$ ) (figura III.19b), indicándonos que este experto repartió las respuestas tanto correctas como incorrectas por todos los tipos etiológicos sin centrarse en ninguno de ellos.

La comparación de las frecuencias de las creencias del *experto 3* frente a las frecuencias de las creencias en las respuestas correctas e incorrectas no mostró la presencia de diferencias significativas (Respuestas correctas:  $\chi^2=5.625$ , g.l.=4,  $p_\alpha=0.229$ ; Respuestas incorrectas:  $\chi^2=1.871$ , g.l.=4,  $p_\alpha=0.760$ ) (figura III.20a). Sin embargo, en este caso sí que se encontraron diferencias significativas cuando lo que se comparaban eran las frecuencias de la valoración frente a las frecuencias de la valoración en las respuestas correctas e incorrectas (Respuestas correctas:  $\chi^2=34.223$ , g.l.=4,  $p_\alpha<0.001$ ; Respuestas incorrectas:  $\chi^2=14.925$ , g.l.=4,  $p_\alpha=0.005$ ) (figura III.20b), manifestando este experto una clara actitud optimista frente a su respuesta.

De la misma forma que en los dos expertos anteriores, cuando se compararon las creencias cuantitativas y cualitativas así como las creencias cuantitativas y la valoración final del diagnóstico, los resultados fueron estadísticamente aceptables aunque los coeficientes de determinación fueron en ambos casos inferiores a los obtenidos por los *expertos 1* y *2* (Cuantitativa-Cualitativa:  $R^2=0.655$ ,  $F(1,24)=45.659$ ,  $p_\alpha<0.001$ ,  $n=26$ ; Cuantitativa-Valoración final:  $R^2=0.758$ ,  $F(1,27)=75.102$ ,  $p_\alpha<0.001$ ,  $n=26$ ) (figuras III.21a y III.21b).

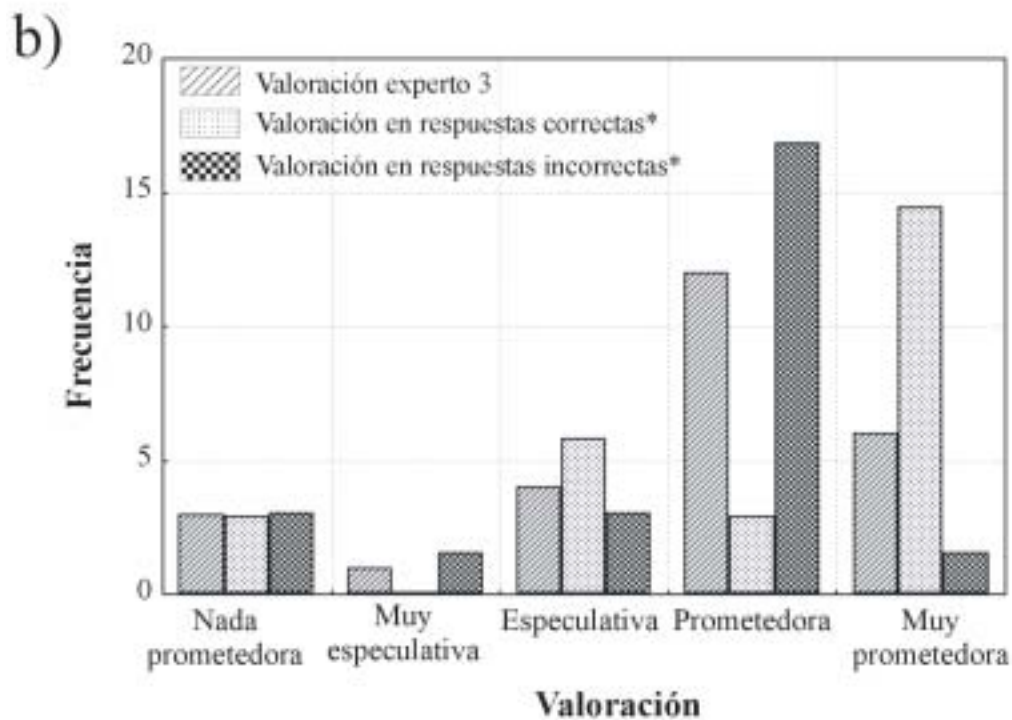
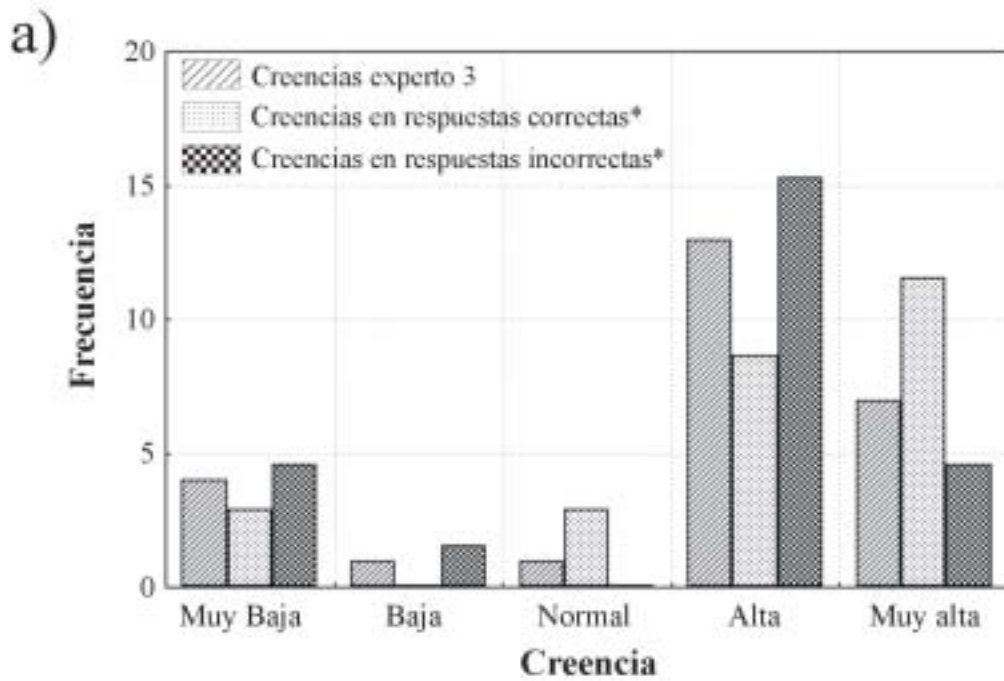


**Figura III.18.** a) Regresión entre la creencia cualitativa (muy baja, baja, normal, alta y muy alta) y la creencia cuantitativa media aportada por el experto 2. b) Regresión entre la valoración final del diagnóstico (nada prometedor, muy especulativo, especulativo, prometedor y muy prometedor) y la creencia cuantitativa media aportada por el experto 2. \* Para permitir los cálculos se ha asignado a cada etiqueta cualitativa un valor cuantitativo de tal forma que muy baja=nada prometedora=1, baja=muy especulativa=2, normal=especulativa=3, alta=prometedora=4 y muy alta=muy prometedora=5

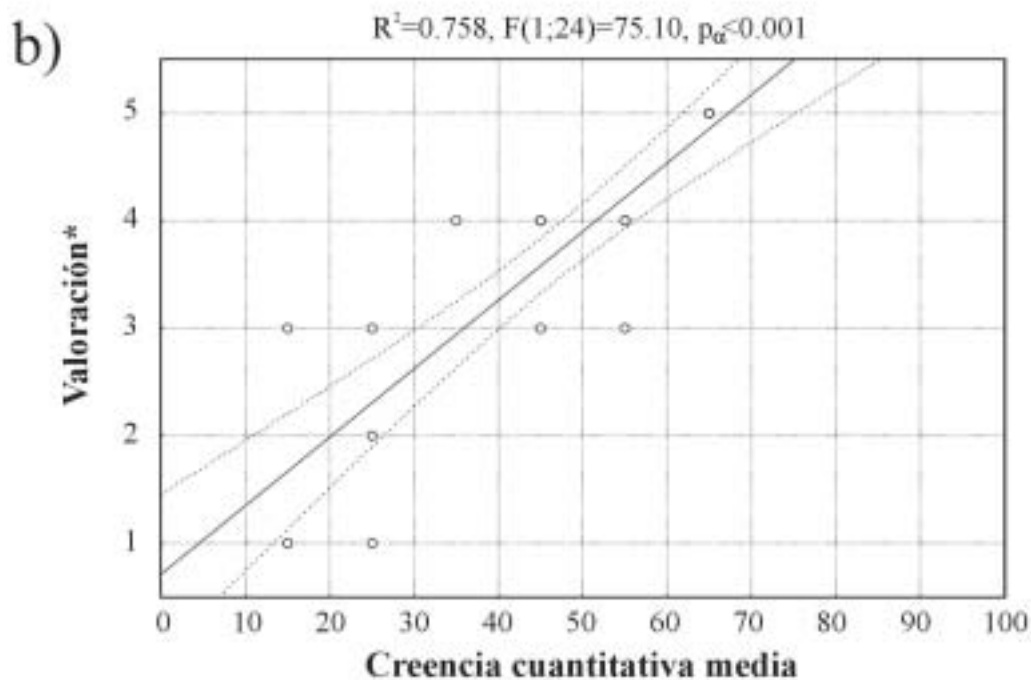
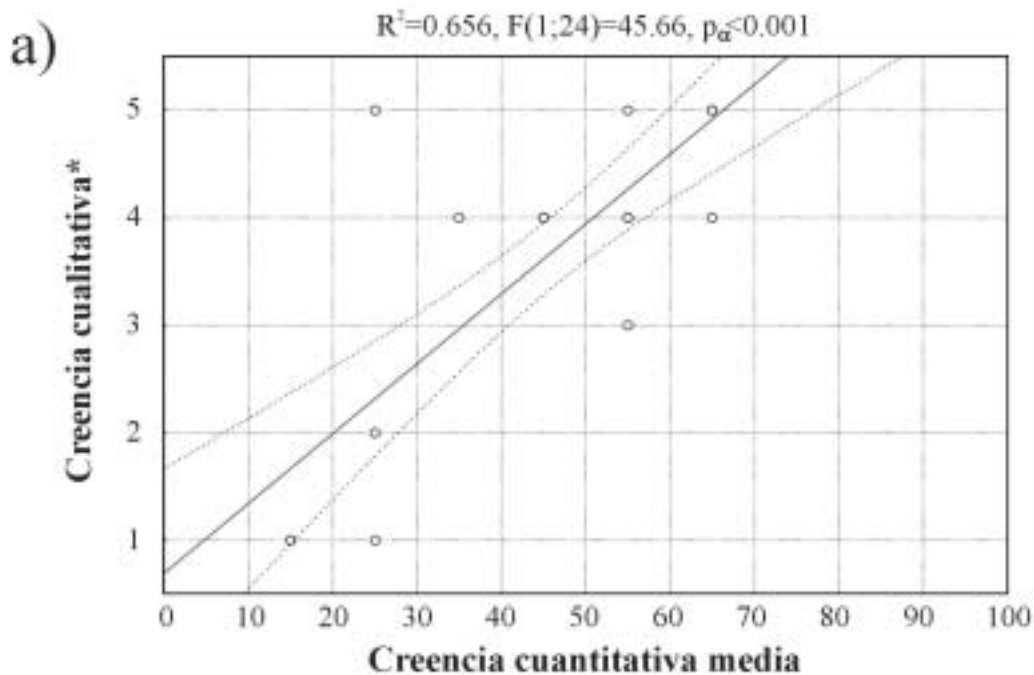


**Figura 19.** a) Frecuencia de respuestas del experto 3 frente a la frecuencia del 'estándar oro'.  
 b) Frecuencia del 'estándar oro' frente a la frecuencia de respuestas correctas e incorrectas del experto 3. \* Valores escalados a 26 casos para permitir la comparación





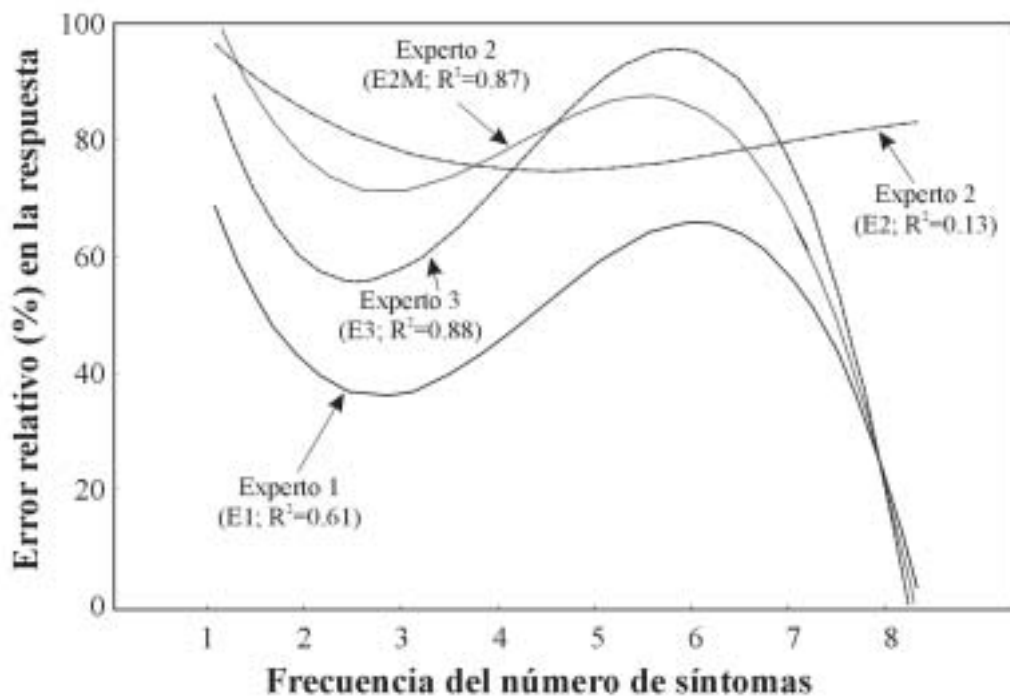
**Figura III.20.** a) Frecuencias de las creencias del experto 3 frente a las frecuencias de creencias en las respuestas correctas e incorrectas. b) Frecuencias de las valoraciones del experto 3 frente a las frecuencias de las valoraciones en respuestas correctas e incorrectas. \* Valores escalados a 26 casos para permitir la comparación



**Figura III.21.** a) Regresión entre la creencia cualitativa (muy baja, baja, normal, alta y muy alta) y la creencia cuantitativa media aportada por el experto 3. b) Regresión entre la valoración final del diagnóstico (nada prometedor, muy especulativo, especulativo, prometedor y muy prometedor) y la creencia cuantitativa media aportada por el experto 3. \* Para permitir los cálculos se ha asignado a cada etiqueta cualitativa un valor cuantitativo de tal forma que muy baja=nada prometedor=1, baja=muy especulativa=2, normal=especulativa=3, alta=prometedor=4 y muy alta=muy prometedor=5

A diferencia de los *expertos 1 y 2*, el *experto 3* tan solo respondió ‘SI’ en el apartado 5 del cuestionario en el 35 % de las ocasiones.

Se ha encontrado un patrón de comportamiento común a dos de los tres expertos cuando se relaciona el número de síntomas que compone cada caso con la frecuencia relativa de las respuestas erróneas. Esta relación viene determinada por una ecuación polinómica de tercer grado cuyo coeficiente de determinación fue superior a 0.59 en el caso de los *expertos 1 y 3*. En el *experto 2* no se encontró una relación clara entre las variables, obteniéndose una correlación muy baja (0.13). Esto fue debido a que el *experto 2* fue el único que respondió erróneamente al caso con ocho síntomas (Caso 5). Sin embargo, tras la eliminación de este caso se obtuvo el mismo patrón que para los *expertos 1 y 3*, alcanzándose un coeficiente de determinación de 0.87 (figura III.22).



**Figura III.22.** Regresiones polinómicas de tercer grado entre la frecuencia del número de síntomas por caso y el error relativo de los expertos. Se especifica la curva correspondiente a cada experto así como el coeficiente de determinación asociado. La curva E2M se corresponde con la regresión del experto 2 suponiendo que la respuesta al Caso 5 hubiese sido correcta

### 3.5.4. Análisis preliminar de las Memorias Asociativas Borrosas

El análisis preliminar de las Memorias Asociativas Borrosas se inicia con el cálculo del estadístico  $\Psi$ . Ya que la versión 1.0 de **SEDPA** cuenta con tres **MABs** independientes, es necesario llevar a cabo un total de tres cálculos para el estadístico ( $\Psi$ [Visser-Sole],

$\Psi$ [Visser-Carlos] y  $\Psi$ [Carlos-Sole]). Por otra parte, ya que el nivel de significación del estadístico es estimado mediante la elección aleatoria de mil pseudopermutaciones, se hace necesario repetir el test varias veces para así determinar un nivel de significación medio. De esta forma se ha llevado a cabo el proceso del cálculo de  $\Psi$  y su nivel de significación diez veces en cada comparación (tabla III.5).

**Tabla III.5.** Valores del estadístico  $\Psi$  para cada una de las comparaciones así como los niveles de significación en cada uno de los diez test realizados. Asimismo, se indica el valor medio del nivel de significación ( $p_{\alpha m}$ ) para cada comparación y su desviación estándar

	[Visser-Sole]	[Visser-Carlos]	[Carlos-Sole]
Estadístico $\Psi$	0.304	0.025	0.200
$p_{\alpha 1}$	0.002	0.001	0.079*
$p_{\alpha 2}$	0.001	0.005	0.039
$p_{\alpha 3}$	0.001	0.002	0.047
$p_{\alpha 4}$	0.002	0.000	0.042
$p_{\alpha 5}$	0.004	0.007	0.051*
$p_{\alpha 6}$	0.004	0.005	0.035
$p_{\alpha 7}$	0.001	0.002	0.034
$p_{\alpha 8}$	0.001	0.002	0.035
$p_{\alpha 9}$	0.002	0.000	0.084*
$p_{\alpha 10}$	0.002	0.009	0.057*
$p_{\alpha m}$	0.009	0.003	0.050
Desviación estándar	0.015	0.003	0.017

\* No existen diferencias significativas

Los resultados muestran que las mayores diferencias medias en las distribuciones de las creencias ( $\Psi=0.025$ ,  $p_{\alpha m}=0.003\pm 0.015$ ) así como las diferencias más acusadas de forma individual ( $\Psi=0.025$ ,  $p_{\alpha 4}=0.000$ ,  $p_{\alpha 9}=0.000$ ) las encontramos entre las memorias asociativas borrosas Visser y Carlos. Por otra parte, las memorias asociativas más parecidas son Carlos y Sole entre las cuales se alcanza un nivel de significación marginal ( $\Psi=0.200$ ,  $p_{\alpha m}=0.050\pm 0.017$ ) lo que puede deberse a que en cuatro de los diez test llevados a cabo en esta comparación no se encontraron diferencias significativas ( $\Psi=0.200$ ,  $p_{\alpha 1}=0.079$ ,  $p_{\alpha 5}=0.051$ ,  $p_{\alpha 9}=0.084$ ,  $p_{\alpha 10}=0.057$ ).

### 3.5.5. Análisis de las respuestas de SEDPA

Las tasas de acierto de **SEDPA** tras la introducción de los 29 casos, considerando las tres memorias asociativas, los conjuntos borrosos triangular, gaussiano y trapezoidal, y los niveles de solapamiento entre conjuntos (50 % y 10 %) se muestran en la tabla III.6.

**Tabla III.6.** Tasas de acierto, acierto parcial y fallo del diagnóstico del tipo y el género del agente patógeno de SEDPA mediante la combinación de tres memorias asociativas borrosas (Visser, Sole y Carlos), tres tipos de conjuntos borrosos (Triangular, Gausiano y Trapezoidal) y dos niveles de solapamiento (50 % y 10 %)

Combinación de MAB, conjunto borroso y nivel de solapamiento	Tasa de acierto (Tipo-Género) %	Tasa de acierto parcial (Tipo-Género) %	Tasa de fallo (Tipo-Género) %
Visser-Triangular-50	86.2-69.0	10.3-27.6	3.4-3.4
Sole-Triangular-50	82.8-62.1	10.3-31.0	6.9-6.9
Carlos-Triangular-50	86.2-69.0	10.3-27.6	3.4-3.4
Visser-Triangular-10	86.2-69.0	10.3-27.6	3.4-3.4
Sole-Triangular-10	82.8-62.1	10.3-31.0	6.9-6.9
Carlos-Triangular-10	86.2-69.0	10.3-27.6	3.4-3.4
Visser-Gausiano-50	86.2-69.0	10.3-27.6	3.4-3.4
Sole-Gausiano-50	82.8-62.1	10.3-31.0	6.9-6.9
Carlos-Gausiano-50	86.2-69.0	10.3-27.6	3.4-3.4
Visser-Gausiano-10	86.2-69.0	10.3-27.6	3.4-3.4
Sole-Gausiano-10	82.8-62.1	10.3-31.0	6.9-6.9
Carlos-Gausiano-10	86.2-69.0	10.3-27.6	3.4-3.4
Visser-Trapezoidal-50	86.2-69.0	10.3-27.6	3.4-3.4
Sole-Trapezoidal-50	82.8-62.1	13.8-34.5	3.4-3.4
Carlos-Trapezoidal-50	86.2-69.0	10.3-27.6	3.4-3.4
Visser-Trapezoidal-10	86.2-69.0	10.3-27.6	3.4-3.4
Sole-Trapezoidal-10	82.8-62.1	10.3-31.0	6.9-6.9
Carlos-Trapezoidal-10	86.2-69.0	10.3-27.6	3.4-3.4

A través de estos resultados se puede observar como la mayor tasa de acierto es obtenida para el tipo de agente patógeno (virus, bacteria, parásito, parámetro físico-químico y otros) cuando la memoria asociativa utilizada es Visser o Carlos (86.2 %). Cuando la memoria asociativa utilizada es Sole la tasa de acierto disminuye hasta un 82.8 %. Por otra parte, cuando se consideran aquellas respuestas en las que el sistema respondió correctamente de forma parcial, es decir, cuando además de diagnosticar el tipo del estándar oro también era diagnosticado otro tipo de agente patógeno, la tasa de acierto aumentó hasta el 96.6 % si las memorias asociativas utilizadas eran Visser o Carlos para los tres tipos de conjuntos borrosos y ambos niveles de solapamiento. En el caso de la **MAB** Sole dicha tasa de acierto se alcanzó cuando el conjunto borroso era trapezoidal y el solapamiento era del 50 %. Con esta memoria asociativa y considerando las respuestas aproximadas para el resto de conjuntos borrosos y niveles de solapamiento, el mejor resultado obtenido fue del 93.1 %. De este modo se obtiene que de forma general con Visser y Carlos el error cometido es del 3.4 % mientras que con Sole el error aumenta hasta

el 6.9 %, exceptuando el caso en el que se utiliza el conjunto trapezoidal al 50 % de solapamiento donde el error disminuye hasta el 3.4 %. Esto es debido a que un caso que en el resto de las configuraciones era clasificado como incorrecto con esta configuración de funcionamiento pasa a ser clasificado como aproximado.

Cuando el parámetro que tenemos en cuenta es el género del agente patológico responsable de la enfermedad los porcentajes varían ligeramente. En este caso, la tasa de acierto disminuye hasta un 69.0 % al utilizar las **MABs** Viser y Carlos y hasta un 62.1 % con la **MAB** Sole. De forma similar al diagnóstico del tipo, cuando consideramos los aciertos parciales los porcentajes son del 96.6 % en el primer caso y del 93.1 % en el segundo (también exceptuando el caso Sole-Trapezoidal-50 donde la tasa de acierto es del 96.6 %).

Al llevar a cabo la comparación de las creencias medias no compensadas para el conjunto de los 29 casos en función de las memorias asociativas borrosas, conjuntos borrosos de forma triangular y un nivel de solapamiento del 50 % por medio de un test t de Student, encontramos que en todos los casos existen diferencias estadísticas significativas (tabla III.7). Sin embargo, cuando fueron comparadas las creencias medias compensadas por el incremento de discrepancia detectadas a través del test  $\chi^2$  modificado, no se encontraron en ningún caso diferencias significativas (tabla III.8).

Resultados parecidos se obtienen cuando se repite el proceso pero cambiando el nivel de solapamiento de los conjuntos borrosos a un 10 %. Al igual que en el caso anterior, el contraste de las tres creencias medias no compensadas mostró diferencias significativas (tabla III.9). Asimismo la comparación entre las creencias medias compensadas para las **MABs** Sole y Carlos aportó el mismo resultado (test-t,  $t = -2.0872$ , g.l. = 25,  $p_\alpha = 0.0472$ ) (tabla III.10).

**Tabla III.7.** Resultados de la comparación de las creencias medias no compensadas en función de las memorias asociativas con conjuntos borrosos triangulares y niveles de solapamiento del 50 % a través de un test t-Student. En todos los casos g.l.=27

Variables	Creencia media	Desviación estándar	t	$p_\alpha$
Visser-Triangular-50	0.6945	0.2483	-3.1401	0.0041 *
Sole-Triangular-50	0.7332	0.2540		
Visser-Triangular-50	0.6945	0.2483	-6.6912	<0.0001*
Carlos-Triangular-50	0.8054	0.1864		
Sole-Triangular-50	0.7332	0.2540	-3.8893	0.0006*
Carlos-Triangular-50	0.8054	0.1864		

\* Diferencias significativas

**Tabla III.8.** Resultados de la comparación de las creencias medias compensadas en función de las memorias asociativas con conjuntos borrosos triangulares y niveles de solapamiento del 50 % a través de un test t-Student. En todos los casos g.l.=25

Variables	Creencia media	Desviación estándar	t	$p_{\alpha}$
Visser-Triangular-50	0.8054	0.2234	0.0888	0.9299
Sole-Triangular-50	0.8034	0.2729		
Visser-Triangular-50	0.8054	0.2234	-1.6448	0.1125
Carlos-Triangular-50	0.8582	0.1856		
Sole-Triangular-50	0.8034	0.2729	-1.8205	0.0806
Carlos-Triangular-50	0.8582	0.1856		

**Tabla III.9.** Resultados de la comparación de las creencias medias no compensadas en función de las memorias asociativas con conjuntos borrosos triangulares y niveles de solapamiento del 10 % a través de un test t-Student. En todos los casos g.l.=27

Variables	Creencia media	Desviación estándar	t	$p_{\alpha}$
Visser-Triangular-10	0.6879	0.2495	-2.9164	0.0070*
Sole-Triangular-10	0.7296	0.2577		
Visser-Triangular-10	0.6879	0.2495	-6.7302	<0.0001*
Carlos-Triangular-10	0.8130	0.1744		
Sole-Triangular-10	0.7296	0.2577	-3.9065	0.0006*
Carlos-Triangular-10	0.8130	0.1744		

\* Diferencias significativas

En cambio, estos resultados varían cuando llevamos a cabo el mismo grupo de comparaciones pero cambiando el conjunto borroso triangular por un gaussiano. De esta forma se puede observar como en el caso de las creencias medias no compensadas, no encontramos diferencias significativas cuando las **MABs** utilizadas son Visser y Sole (test-t,  $t = -0.4223$ , g.l. = 27,  $p_{\alpha} = 0.6761$ ) (tabla III.11). Con este tipo de conjunto borroso y utilizando las creencias medias compensadas, los resultados son similares a los obtenidos con los conjuntos borrosos triangulares aunque en el caso **MAB** Sole frente a **MAB** Carlos se obtienen diferencias significativas marginales (test-t,  $t = -1.9341$ , g.l. = 25,  $p_{\alpha} = 0.0645$ ) (tabla III.12).

**Tabla III.10.** Resultados de la comparación de las creencias medias compensadas en función de las memorias asociativas con conjuntos borrosos triangulares y niveles de solapamiento del 10 % a través de un test *t*-Student. En todos los casos g.l.=25

<b>Variables</b>	<b>Creencia media</b>	<b>Desviación estándar</b>	<b>t</b>	<b>p<sub>α</sub></b>
Visser-Triangular-10	0.8354	0.1935	0.7395	0.4665
Sole-Triangular-10	0.8108	0.2498		
Visser-Triangular-10	0.8354	0.1935	-0.9092	0.3719
Carlos-Triangular-10	0.8680	0.1666		
Sole-Triangular-10	0.8108	0.2498	-2.0872	0.0472*
Carlos-Triangular-10	0.8680	0.1666		

\* Diferencia significativas

**Tabla III.11.** Resultados de la comparación de las creencias medias no compensadas en función de las memorias asociativas con conjuntos borrosos gaussianos y niveles de solapamiento del 50 % a través de un test *t*-Student. En todos los casos g.l.=27

<b>Variables</b>	<b>Creencia media</b>	<b>Desviación estándar</b>	<b>t</b>	<b>p<sub>α</sub></b>
Visser-Gausiano-50	0.7196	0.2449	-0.4223	0.6761
Sole-Gausiano-50	0.7326	0.2535		
Visser-Gausiano-50	0.7196	0.2449	-2.7784	0.0098*
Carlos-Gausiano-50	0.8090	0.1813		
Sole-Gausiano-50	0.7326	0.2535	-3.8801	0.0006*
Carlos-Gausiano-50	0.8090	0.1813		

\*Diferencias significativas

El mismo efecto se consigue cuando los conjuntos borrosos gaussianos se solapan un 10 % y se tienen en cuenta tanto creencias medias compensadas (tabla III.13) y no compensadas, aunque en este caso las diferencias significativas marginales se obtienen con las **MAB** Visser y Carlos (test-t,  $t = -2.0527$ , g.l. = 25,  $p_{\alpha} = 0.0507$ ) (tabla III.14).



**Tabla III.12.** Resultados de la comparación de las creencias medias compensadas en función de las memorias asociativas con conjuntos borrosos gaussianos y niveles de solapamiento del 50 % a través de un test t-Student. En todos los casos g.l.=25

Variables	Creencia media	Desviación estándar	t	p $_{\alpha}$
Visser-Gausiano-50	0.8088	0.2290	-0.0692	0.9454
Sole-Gausiano-50	0.8102	0.2541		
Visser-Gausiano-50	0.8088	0.2290	-1.7537	0.0917
Carlos-Gausiano-50	0.8639	0.1755		
Sole-Gausiano-50	0.8102	0.2541	-1.9341	0.0645 <sup>1</sup>
Carlos-Gausiano-50	0.8639	0.1755		

<sup>1</sup>Diferencias significativas marginales

**Tabla III.13.** Resultados de la comparación de las creencias medias no compensadas en función de las memorias asociativas con conjuntos borrosos gaussianos y niveles de solapamiento del 10 % a través de un test t-Student. En todos los casos g.l.=27

Variables	Creencia media	Desviación estándar	t	p $_{\alpha}$
Visser-Gausiano-10	0.7178	0.2478	-0.0886	0.9300
Sole-Gausiano-10	0.7208	0.2639		
Visser-Gausiano-10	0.7178	0.2478	-3.0578	0.0050*
Carlos-Gausiano-10	0.8183	0.1660		
Sole-Gausiano-10	0.7208	0.2639	-4.2230	0.0002*
Carlos-Gausiano-10	0.8183	0.1660		

\* Diferencias significativas

**Tabla III.14.** Resultados de la comparación de las creencias medias compensadas en función de las memorias asociativas con conjuntos borrosos gaussianos y niveles de solapamiento del 10 % a través de un test t-Student. En todos los casos g.l.=25

Variables	Creencia media	Desviación estándar	t	p $_{\alpha}$
Visser-Gausiano-10	0.8067	0.2323	-1.3766	0.1809
Sole-Gausiano-10	0.8245	0.2501		
Visser-Gausiano-10	0.8067	0.2323	-2.0527	0.0507 <sup>1</sup>
Carlos-Gausiano-10	0.8710	0.1568		
Sole-Gausiano-10	0.8245	0.2501	-1.2954	0.2070
Carlos-Gausiano-10	0.8710	0.1568		

<sup>1</sup>Diferencias significativas marginales

Resultados semejantes a los obtenidos con conjuntos borrosos triangulares se consiguen con conjuntos de forma trapezoidal tanto para las creencias no compensadas como compensadas (tablas III.15, III.16, III.17 y III.18), excepto en la comparación entre las **MABs** Sole y Carlos, creencias compensadas y niveles de solapamiento del 10 % donde al contrario que en el caso de conjuntos borrosos triangulares, el contraste no mostró la presencia de diferencias significativas (test-t,  $t = -1.7463$ , g.l. = 25,  $p_{\alpha} = 0.0930$ ) (tabla III.18).

**Tabla III.15.** Resultados de la comparación de las creencias medias no compensadas en función de las memorias asociativas con conjuntos borrosos trapezoidales y niveles de solapamiento del 50 % a través de un test t-Student. En todos los casos g.l.=27

Variables	Creencia media	Desviación estándar	t	$p_{\alpha}$
Visser-Trapezoidal-50	0.6945	0.2484	-3.2833	0.0028*
Sole- Trapezoidal -50	0.7340	0.2532		
Visser- Trapezoidal -50	0.6945	0.2484	-6.7281	<0.0001*
Carlos- Trapezoidal -50	0.8035	0.1889		
Sole- Trapezoidal -50	0.7340	0.2532	-3.9310	0.0005*
Carlos- Trapezoidal -50	0.8035	0.1889		

\* Diferencias significativas

**Tabla III.16.** Resultados de la comparación de las creencias medias compensadas en función de las memorias asociativas con conjuntos borrosos trapezoidales y niveles de solapamiento del 50 % a través de un test t-Student. En todos los casos g.l.=25

Variables	Creencia media	Desviación estándar	t	$p_{\alpha}$
Visser-Trapezoidal-50	0.8068	0.2272	0.0926	0.9269
Sole- Trapezoidal -50	0.8049	0.2669		
Visser- Trapezoidal -50	0.8068	0.2272	-1.4906	0.1486
Carlos- Trapezoidal -50	0.8548	0.1918		
Sole- Trapezoidal -50	0.8049	0.2669	-1.7233	0.0972
Carlos- Trapezoidal -50	0.8548	0.1918		

**Tabla III.17.** Resultados de la comparación de las creencias medias no compensadas en función de las memorias asociativas con conjuntos borrosos trapezoidales y niveles de solapamiento del 10 % a través de un test t-Student. En todos los casos g.l.=27

VARIABLES	Creencia media	Desviación estándar	t	p $\alpha$
Visser-Trapezoidal-10	0.6924	0.2491	-2.5282	0.0176*
Sole- Trapezoidal -10	0.7376	0.2581		
Visser- Trapezoidal -10	0.6924	0.2491	-6.5952	<0.0001*
Carlos- Trapezoidal -10	0.8122	0.1761		
Sole- Trapezoidal -10	0.7376	0.2581	-3.1925	0.0036*
Carlos- Trapezoidal -10	0.8122	0.1761		

\* Diferencias significativas

**Tabla III.18.** Resultados de la comparación de las creencias medias compensadas en función de las memorias asociativas con conjuntos borrosos trapezoidales y niveles de solapamiento del 10 % a través de un test t-Student. En todos los casos g.l.=25

VARIABLES	Creencia media	Desviación estándar	t	p $\alpha$
Visser- Trapezoidal -10	0.8085	0.2302	-0.4052	0.6888
Sole- Trapezoidal -10	0.8179	0.2495		
Visser- Trapezoidal -10	0.8085	0.2302	-1.8887	0.0706
Carlos- Trapezoidal -10	0.8673	0.1685		
Sole- Trapezoidal -10	0.8179	0.2495	-1.7463	0.0930
Carlos- Trapezoidal -10	0.8673	0.1685		

Al comparar las creencias medias no compensadas variando el tipo de conjunto borroso y dejando fijos tanto la memorias asociativas borrosas como el nivel de solapamiento tan solo se encontraron diferencias significativas entre la **MAB** Carlos y niveles de solapamiento del 50 % en conjuntos borrosos de tipo triangular, gaussiano y trapezoidal (tabla III.19). Las diferencias se acentuaron ligeramente cuando los niveles de solapamiento se redujeron al 10 %, además de que de forma general las desviaciones estándar fueron mayores en todos los casos excepto cuando se utilizaba la **MAB** Carlos (tabla III.20). Estas diferencias no se detectaron cuando se compararon las creencias medias compensadas (tablas III.21 y III.22).

**Tabla III.19.** Resultados de la comparación de las creencias medias no compensadas en función de los conjuntos borrosos para las memorias asociativas borrosas Visser, Sole y Carlos y niveles de solapamiento del 50 % a través de un test t-Student. En todos los casos g.l.=27

<b>Variables</b>	<b>Creencia media</b>	<b>Desviación estándar</b>	<b>t</b>	<b>p<math>\alpha</math></b>
Visser-Triangular-50	0.6945	0.2483	0.6691	0.5091
Visser-Gausiano-50	0.6936	0.2486		
Visser-Triangular-50	0.6945	0.2483	-0.0273	0.9784
Visser-Trapezoidal-50	0.6945	0.2484		
Visser-Gausiano-50	0.6936	0.2486	-0.4256	0.6738
Visser-Trapezoidal-50	0.6945	0.2484		
Sole-Triangular-50	0.7332	0.2540	0.2665	0.7919
Sole-Gausiano-50	0.7326	0.2535		
Sole-Triangular-50	0.7332	0.2540	-0.4633	0.6468
Sole-Trapezoidal-50	0.7340	0.2532		
Sole-Gausiano-50	0.7326	0.2535	-0.3776	0.7087
Sole-Trapezoidal-50	0.7340	0.2532		
Carlos-Triangular-50	0.7998	0.1855	-2.3066	0.0287*
Carlos-Gausiano-50	0.8038	0.1802		
Carlos-Triangular-50	0.7998	0.1855	2.1391	0.0413*
Carlos-Trapezoidal-50	0.7977	0.1882		
Carlos-Gausiano-50	0.8038	0.1802	2.2612	0.0317*
Carlos-Trapezoidal-50	0.7977	0.1882		

\* Diferencias significativas

**Tabla III.20.** Resultados de la comparación de las creencias medias no compensadas en función de los conjuntos borrosos para las memorias asociativas borrosas Visser, Sole y Carlos y niveles de solapamiento del 10 % a través de un test t-Student. En todos los casos g.l.=27

Variables	Creencia media	Desviación estándar	t	p $\alpha$
Visser-Triangular-10	0.6879	0.2495	-0.4596	0.6495
Visser-Gausiano-10	0.6899	0.2499		
Visser-Triangular-10	0.6879	0.2495	-1.1298	0.2685
Visser-Trapezoidal-10	0.6924	0.2491		
Visser-Gausiano-10	0.6899	0.2499	-1.3112	0.2007
Visser-Trapezoidal-10	0.6924	0.2407		
Sole-Triangular-10	0.7296	0.2577	2.0828	0.0469*
Sole-Gausiano-10	0.7208	0.2639		
Sole-Triangular-10	0.7296	0.2577	-0.9494	0.3508
Sole-Trapezoidal-10	0.7376	0.2582		
Sole-Gausiano-10	0.7208	0.2639	-1.8590	0.0740
Sole-Trapezoidal-10	0.7376	0.2582		
Carlos-Triangular-10	0.8091	0.1726	-2.1515	0.0402*
Carlos-Gausiano-10	0.8163	0.1633		
Carlos-Triangular-10	0.8091	0.1726	1.8667	0.0723
Carlos-Trapezoidal-10	0.8079	0.1745		
Carlos-Gausiano-10	0.8163	0.1633	2.1119	0.0438*
Carlos-Trapezoidal-10	0.8079	0.1745		

\* Diferencias significativas

**Tabla III.21.** Resultados de la comparación de las creencias medias compensadas en función de los conjuntos borrosos para las memorias asociativas borrosas Visser, Sole y Carlos y niveles de solapamiento del 50 % a través de un test t-Student. En todos los casos g.l.=25

<b>Variables</b>	<b>Creencia media</b>	<b>Desviación estándar</b>	<b>t</b>	<b>p<sub>α</sub></b>
Visser-Triangular-50	0.8054	0.2234	-0.4480	0.6580
Visser-Gausiano-50	0.8088	0.2290		
Visser-Triangular-50	0.8054	0.2234	-0.1994	0.8436
Visser-Trapezoidal-50	0.8068	0.2272		
Visser-Gausiano-50	0.8088	0.2290	0.6846	0.4999
Visser-Trapezoidal-50	0.8068	0.2272		
Sole-Triangular-50	0.8034	0.2729	-1.1655	0.2548
Sole-Gausiano-50	0.8102	0.2541		
Sole-Triangular-50	0.8034	0.2729	-0.4950	0.6249
Sole-Trapezoidal-50	0.8049	0.2668		
Sole-Gausiano-50	0.8102	0.2541	1.1953	0.2432
Sole-Trapezoidal-50	0.8049	0.2668		
Carlos-Triangular-50	0.8582	0.1856	-1.7404	0.0941
Carlos-Gausiano-50	0.8639	0.1755		
Carlos-Triangular-50	0.8582	0.1856	1.6698	0.1074
Carlos-Trapezoidal-50	0.8548	0.1918		
Carlos-Gausiano-50	0.8639	0.1755	1.7163	0.0984
Carlos-Trapezoidal-50	0.8548	0.1918		

**Tabla III.22.** Resultados de la comparación de las creencias medias compensadas en función de los conjuntos borrosos para las memorias asociativas borrosas Visser, Sole y Carlos y niveles de solapamiento del 10 % a través de un test t-Student. En todos los casos g.l.=25

<b>Variables</b>	<b>Creencia media</b>	<b>Desviación estándar</b>	<b>t</b>	<b>p<math>\alpha</math></b>
Visser-Triangular-10	0.8354	0.1935	1.0622	0.2983
Visser-Gausiano-10	0.8067	0.2324		
Visser-Triangular-10	0.8354	0.1935	0.9859	0.3336
Visser-Trapezoidal-10	0.8085	0.2302		
Visser-Gausiano-10	0.8067	0.2324	-0.8936	0.3801
Visser-Trapezoidal-10	0.8085	0.2302		
Sole-Triangular-10	0.8108	0.2498	-0.8379	0.4100
Sole-Gausiano-10	0.8245	0.2501		
Sole-Triangular-10	0.8108	0.2498	-1.1690	0.2534
Sole-Trapezoidal-10	0.8179	0.2496		
Sole-Gausiano-10	0.8245	0.2501	0.3684	0.7156
Sole-Trapezoidal-10	0.8179	0.2496		
Carlos-Triangular-10	0.8582	0.1666	-1.0049	0.3245
Carlos-Gausiano-10	0.8710	0.1568		
Carlos-Triangular-10	0.8679	0.1666	1.0424	0.3072
Carlos-Trapezoidal-10	0.8673	0.1685		
Carlos-Gausiano-10	0.8710	0.1568	1.0113	0.3216
Carlos-Trapezoidal-10	0.8673	0.1685		

Para encontrar el tipo de relación existente entre creencias asignadas a través de la teoría de Dempster-Shafer y las creencias reescaladas mediante el cálculo del incremento porcentual de la discrepancia entre dos eventos consecutivos, se han llevado a cabo regresiones lineales y no lineales entre ambos tipos de creencias en función de la memoria asociativa borrosa, tipo de conjunto borroso y nivel de solapamiento. En todos los casos se han encontrado mayores porcentajes de varianza explicada cuando se han utilizado regresiones no lineales del tipo (figuras III.23, III.24 y III.25):

$$\text{Creencia Compensada} = K + p \text{Log}_{10}(\text{Creencia no Compensada}) \quad (3.38)$$

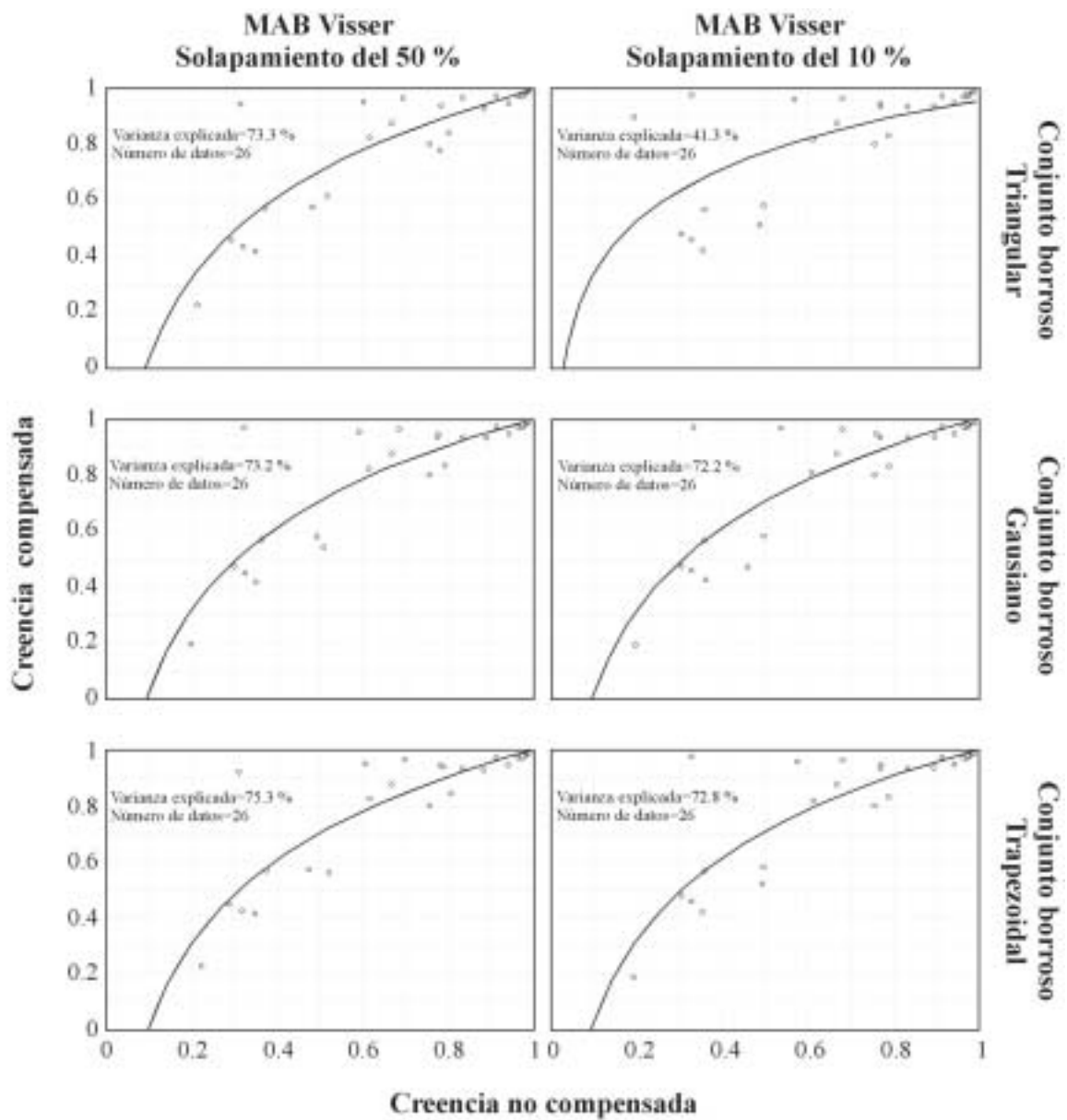
donde **K** es una constante y **p** es un parámetro que multiplica a la variable independiente. En todos los casos excepto en uno, el porcentaje de varianza explicada superó el 70 %. Cuando se utilizó la **MAB** Visser, conjunto borroso triangular y un nivel de solapamiento del 10 %, la varianza explicada alcanzó el valor mínimo de 41.3 % (figura III.23). En general, la varianza explicada cuando el nivel de solapamiento de los conjuntos borrosos era del 50 % fue mayor que cuando el solapamiento era del 10 %.

De la misma forma que en los expertos humanos se ha comparado la creencia no compensada y compensada con la valoración final del diagnóstico. En **SEDPA** al utilizar los datos de las condiciones de confinamiento de los individuos aportados a los expertos humanos, el 100 % de las valoraciones se han encontrado entre ‘Nada Prometedoras’ y ‘Muy Especulativas’, estableciéndose correlaciones cercanas al cero y no significativas entre creencias y valoraciones. De esta forma, el sistema perjudica a todos los grupos sintomáticos seleccionados al obtener puntuaciones a través de la expresión (3.34) en todos los casos inferiores al 50 %. Esto contrasta con los resultados obtenidos para los expertos humanos donde en todos los casos se encontraron correlaciones significativas y coeficientes de determinación siempre por encima de 0.75 (figuras III.15, III.18 y III.21).

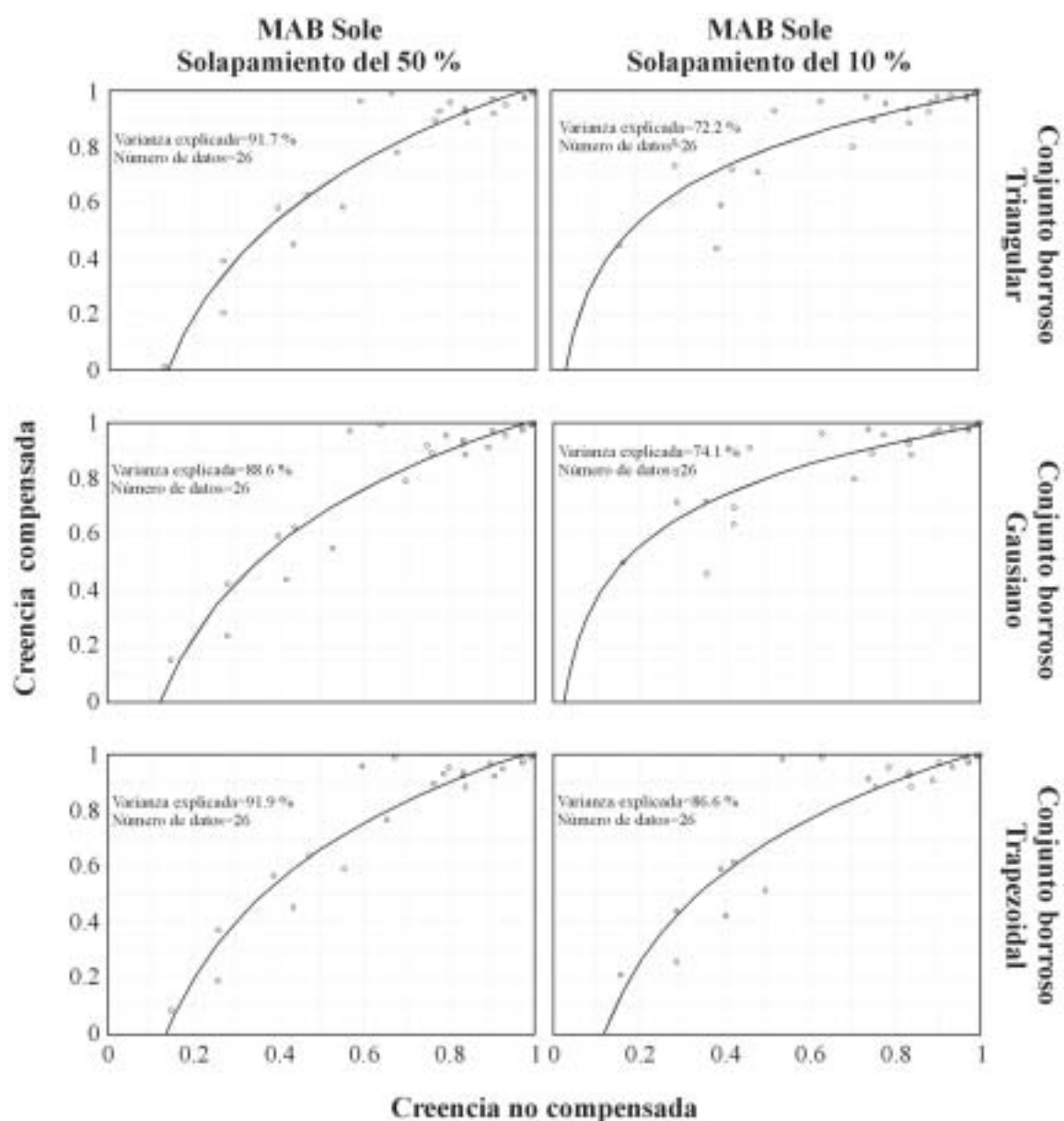
### 3.5.6. Agrupación de diagnósticos individuales de SEDPA

La agrupación de los diagnósticos individuales en función del conjunto borroso y el nivel de solapamiento se ha llevado a cabo mediante la aplicación de un método de clasificación automática arborescente de datos conocida como análisis jerárquico de *clusters*. De este modo se unen en cada nivel de jerarquía los dos *clusters* más cercanos, utilizando como distancia multivariante la euclídea. Asimismo, para la unión en otro nivel jerárquico de los *clusters* más próximos, se han probado los algoritmos de clasificación jerárquica del enlace promedio y del enlace por mínima varianza (Ward) (Hair *et al.*, 1999) obteniéndose resultados similares. La utilización del análisis de *clusters* muestra como en todos los casos excepto en uno, el grupo principal queda constituido por el diagnóstico llevado a cabo a través de las **MABs** Sole y Visser (figuras III.26 y III.27).

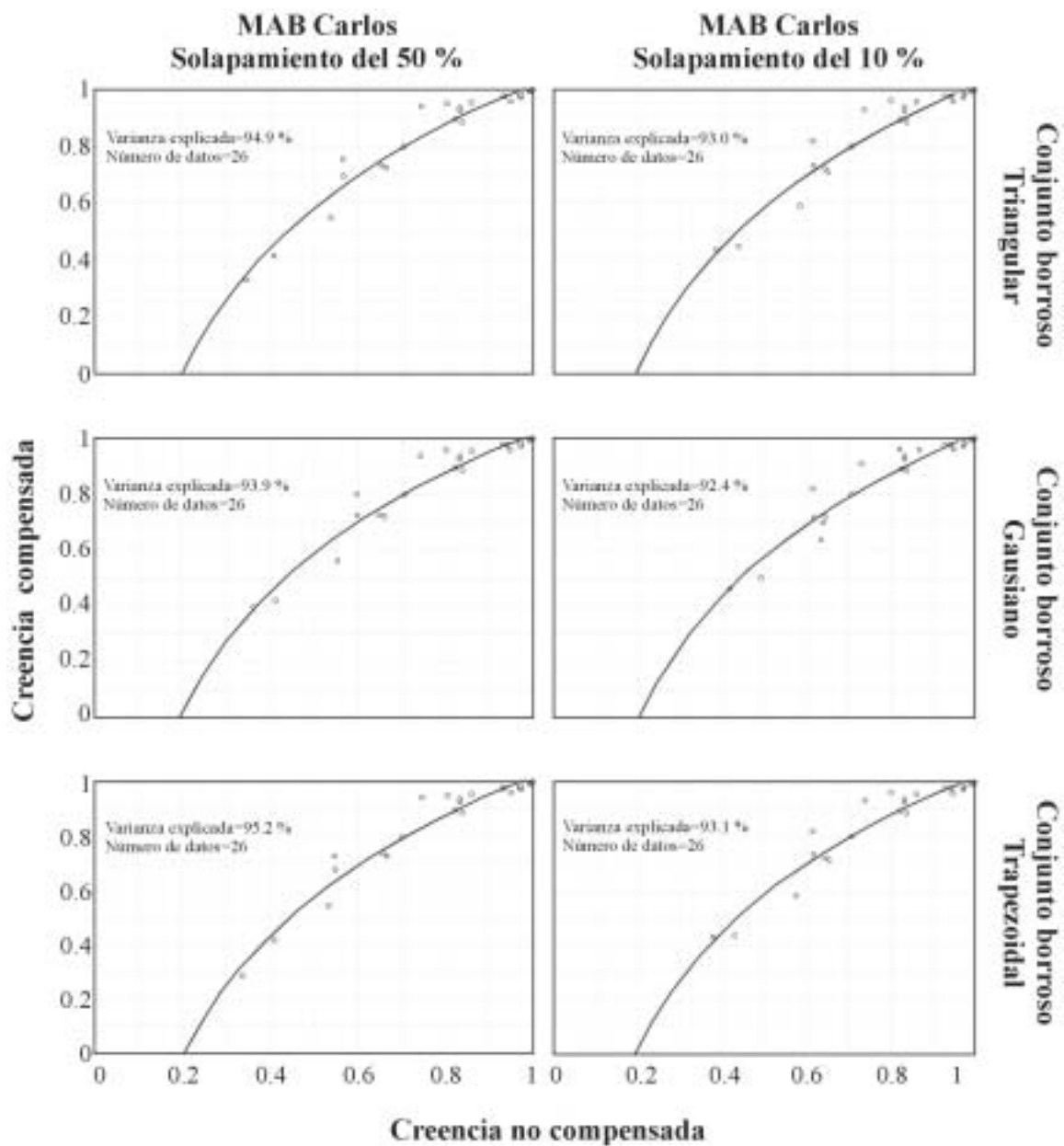




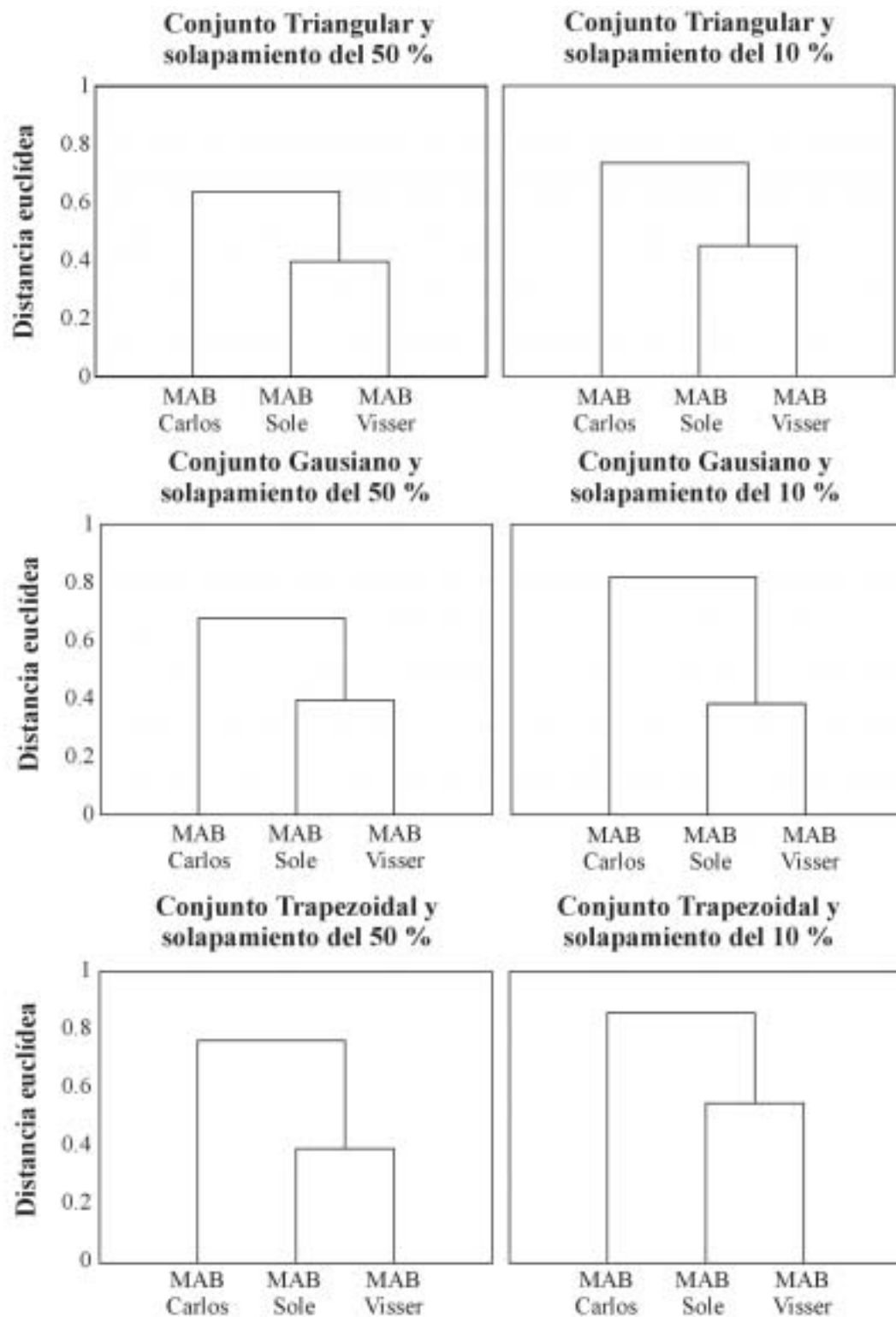
**Figura III.23.** Regresiones no lineales entre creencias no compensadas y compensadas para la memoria asociativa Visser; conjuntos borrosos triangulares, gaussianos y trapezoidales y niveles de solapamiento del 50 y 10 %. En cada regresión se indica el porcentaje de varianza explicada y el número de datos utilizados



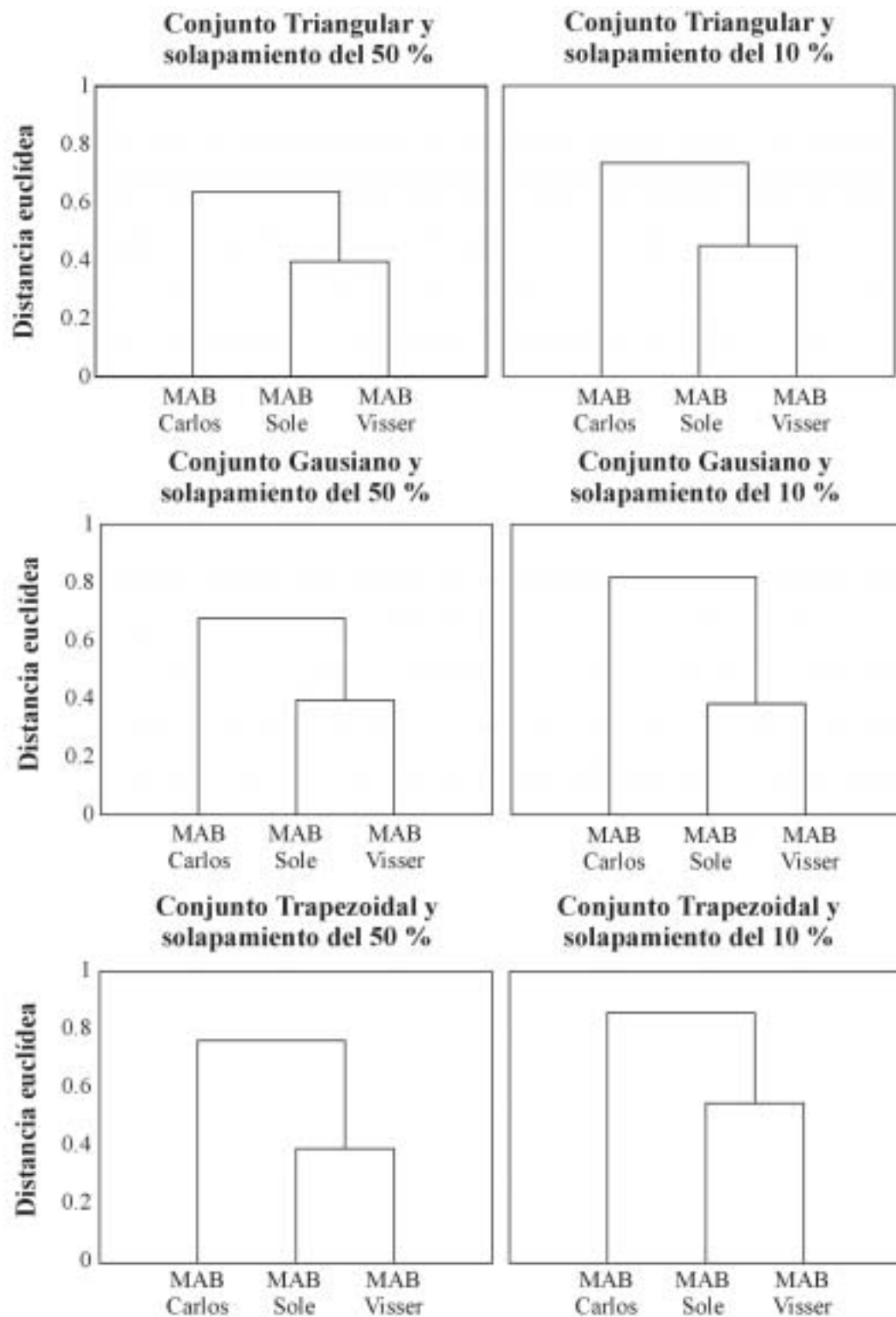
**Figura III.24.** Regresiones no lineales entre creencias no compensadas y compensadas para la memoria asociativa Sole, conjuntos borrosos triangulares, gaussianos y trapezoidales y niveles de solapamiento del 50 y 10 %. En cada regresión se indica el porcentaje de varianza explicada y el número de datos utilizados



**Figura III.25.** Regresiones no lineales entre crecencias no compensadas y compensadas para la memoria asociativa Carlos, conjuntos borrosos triangulares, gaussianos y trapezoidales y niveles de solapamiento del 50 y 10 %. En cada regresión se indica el porcentaje de varianza explicada y el número de datos utilizados



**Figura III.26.** Dendrogramas que agrupan las creencias individuales no compensadas obtenidas para tres memorias asociativas borrosas distintas (Carlos, Sole y Visser) en función del conjunto borroso (triangular, gaussiano y trapezoidal) y niveles de solapamiento del 50 y 10 %



**Figura III.26.** Dendrogramas que agrupan las creencias individuales no compensadas obtenidas para tres memorias asociativas borrosas distintas (Carlos, Sole y Visser) en función del conjunto borroso (triangular, gaussiano y trapezoidal) y niveles de solapamiento del 50 y 10 %

En función del tipo de conjunto borroso y un nivel de solapamiento del 50 %, podemos observar como las mayores distancias se encuentran entre el conjunto de creencias no compensadas para las **MAB** Sole y Visser o grupo principal y las creencias no compensadas para la **MAB** Carlos cuando el conjunto borroso es de tipo trapezoidal. Sin embargo cuando empleamos este tipo de conjunto borroso, las distancias entre las creencias no compensadas para las **MAB** Sole y Visser se hacen mínimas. El mismo efecto se produce cuando el nivel de solapamiento es del 10 % aunque tanto las distancias entre el grupo principal y el grupo constituido por las creencias no compensadas para la **MAB** Carlos, como la distancia existente entre los dos constituyentes del grupo principal aumentan. Si la comparación la llevamos a cabo en función del nivel de solapamiento, encontramos que para los tres tipos de conjuntos borrosos las distancias entre las creencias del grupo principal así como la existente entre el grupo principal y las creencias para la **MAB** Carlos aumentan para un nivel de solapamiento del 10 % (figura III.26).

En el caso de las creencias compensadas, los resultados varían ligeramente cuando las comparamos en función del conjunto y un solapamiento del 50 %. Así, las distancias tanto dentro del grupo principal como entre el grupo principal y las creencias compensadas para la **MAB** Carlos se ven incrementadas y son mayores a las encontradas para las creencias no compensadas. El mismo resultado lo encontramos cuando el nivel de solapamiento es del 10 %, con la diferencia de que en el caso de utilizar conjuntos triangulares, el grupo principal queda conformado por las creencias compensadas para las **MABs** Carlos y Sole. En este caso, la distancia euclídea dentro del grupo principal (0.76) es la más alta para el conjunto de todos los grupos principales formados. En función del nivel de solapamiento, al igual que sucedía con las creencias no compensadas, la mayor distancia entre el grupo principal y el grupo **MAB** Carlos la encontramos cuando el solapamiento es del 10 %. Sin embargo, no sucede lo mismo con las distancias entre los componentes del grupo principal, de tal forma que la menor distancia (0.34) se encuentra en el nivel de solapamiento del 10 % con un conjunto borroso de tipo gaussiano (figura III.27).

### 3.5.7. Diagnóstico de los expertos frente al diagnóstico de SEDPA

Para comprobar la agudeza de diagnóstico de **SEDPA** se han comparado los resultados obtenidos por el sistema con los aportados por los expertos a través del un test exacto de Fisher. De esta forma tanto para el *experto 1* como para el *experto 2* se puede observar que no se encuentran diferencias significativas en el diagnóstico del tipo de agente etiológico teniendo en cuenta para el sistema las respuestas consideradas como correctas, las memorias asociativas borrosas Visser, Sole y Carlos, conjuntos borrosos triangulares, gaussianos y trapezoidales y niveles de solapamiento del 50 y 10 %, y considerando tanto las respuestas correctas como aproximadas de los expertos. En el caso del *experto 3*

tampoco se encontraron diferencias significativas cuando la **MAB** utilizada era Sole, conjuntos borrosos triangulares y gaussianos al 50 y 10 % y trapezoidales al 10 % (tabla III.23).

**Tabla III.23.** Resultados del test exacto de Fisher entre el diagnóstico del tipo de agente patógeno por parte de SEDPA para las MABs Visser, Sole y Carlos, conjuntos borrosos triangulares, gaussianos y trapezoidales, y niveles de solapamiento del 50 y 10 %, y los expertos 1, 2 y 3 considerando sólo las respuestas correctas para el sistema y las respuestas correctas y aproximadas para los expertos

	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	25-22 <sup>†</sup>	23	24	19
<i>Fallo</i>	1-1 <sup>†</sup>	6	5	7
<b>MAB Visser</b> , Triangular-50-10, Gaussiano-50-10, Trapezoidal-50-10 (Exp. 1: n=55, $p_{\alpha}=0.0686$ ; Exp. 2: n=55, $p_{\alpha}=0.1229$ ; <sup>†</sup> Exp. 3: n=49, $p_{\alpha}=0.0370$ *)				
	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	24-21 <sup>†</sup>	23	24	19
<i>Fallo</i>	2-2 <sup>†</sup>	6	5	7
<b>MAB Sole</b> , Triangular-50-10, Gaussiano-50-10, Trapezoidal-10 (Exp. 1: n=55, $p_{\alpha}=0.1636$ ; Exp. 2: n=55, $p_{\alpha}=0.2587$ ; <sup>†</sup> Exp. 3: n=49, $p_{\alpha}=0.1000$ )				
	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	24-21 <sup>†</sup>	23	24	19
<i>Fallo</i>	1-1 <sup>†</sup>	6	5	7
<b>MAB Sole</b> , Trapezoidal-50 (Exp. 1: n=54, $p_{\alpha}=0.0759$ ; Exp. 2: n=54, $p_{\alpha}=0.1434$ ; <sup>†</sup> Exp. 3: n=48, $p_{\alpha}=0.0425$ *)				
	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	25-22 <sup>†</sup>	23	24	19
<i>Fallo</i>	1-1 <sup>†</sup>	6	5	7
<b>MAB Carlos</b> , Triangular-50-10, Gaussiano-50-10, Trapezoidal-50-10 (Exp. 1: n=55, $p_{\alpha}=0.0686$ ; Exp. 2: n=55, $p_{\alpha}=0.1229$ ; <sup>†</sup> Exp. 3: n=49, $p_{\alpha}=0.0370$ *)				

\*Diferencias significativas

<sup>†</sup>Sólo se consideran 26 casos

Sin embargo, si además de las respuestas correctas de **SEDPA**, se consideran las aproximadas, comprobamos que para el *experto 1* y utilizando tanto la **MAB Visser** como Carlos, se detectan diferencias significativas marginales para todos los conjuntos borrosos y todos los niveles de solapamiento (**MAB Visser** y Carlos; n=58;  $p_{\alpha}=0.0510$ ) así como cuando se utiliza la **MAB Sole** con conjuntos trapezoidales y niveles de solapamiento del 50 %. En el caso del *experto 2* en ningún caso se encuentran diferencias significativas,

mientras que con el *experto 3* se detectan en todos los casos excepto cuando se utiliza la **MAB Sole**, conjuntos triangulares y gaussianos al 50 y 10 % y trapezoidales al 10 % (tabla III.24).

**Tabla III.24.** Resultados del test exacto de Fisher entre el diagnóstico del tipo por parte de SEDPA para las MABs Visser, Sole y Carlos, conjuntos borrosos triangulares, gaussianos y trapezoidales, y niveles de solapamiento del 50 y 10 % y los expertos 1, 2 y 3, considerando las respuestas correctas y aproximadas del sistema y los expertos

	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	28-25 <sup>†</sup>	23	24	19
<i>Fallo</i>	1-1 <sup>†</sup>	6	5	7
<b>MAB Visser</b> , Triangular-50-10, Gaussiano-50-10, Trapezoidal-50-10 (Exp. 1: n=58, $p_{\alpha}=0.0510^1$ ; Exp. 2: n=58, $p_{\alpha}=0.0968$ ; <sup>†</sup> Exp. 3: n=52, $p_{\alpha}=0.0248^*$ )				
	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	27-24 <sup>†</sup>	23	24	19
<i>Fallo</i>	2-2 <sup>†</sup>	6	5	7
<b>MAB Sole</b> , Triangular-50-10, Gaussiano-50-10, Trapezoidal-10 (Exp. 1: n=58, $p_{\alpha}=0.1265$ ; Exp. 2: n=58; $p_{\alpha}=0.2114$ ; <sup>†</sup> Exp. 3: n=52, $p_{\alpha}=0.0700$ )				
	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	28-25 <sup>†</sup>	23	24	19
<i>Fallo</i>	1-1 <sup>†</sup>	6	5	7
<b>MAB Sole</b> , Trapezoidal-50 (Exp. 1: n=58, $p_{\alpha}=0.0510^1$ ; Exp. 2: n=58, $p_{\alpha}=0.0968$ ; <sup>†</sup> Exp. 3: n=52, $p_{\alpha}=0.0248^*$ )				
	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	28-25 <sup>†</sup>	23	24	19
<i>Fallo</i>	1-1 <sup>†</sup>	6	5	7
<b>MAB Carlos</b> , Triangular-50-10, Gaussiano-50-10, Trapezoidal-50-10 (Exp. 1: n=58, $p_{\alpha}=0.0510^1$ ; Exp. 2: n=58, $p_{\alpha}=0.0968$ ; <sup>†</sup> Exp. 3: n=52, $p_{\alpha}=0.0248^*$ )				

<sup>1</sup>Diferencias significativas marginales

<sup>\*</sup>Diferencias significativas

<sup>†</sup>Sólo se consideran 26 casos

Las diferencias se ven acentuadas cuando se comparan los resultados del diagnóstico del agente patológico específico. Se comprueba en todos los casos que **SEDPA** es más sensible que los *expertos 1, 2 y 3*. Asimismo, al igual que ocurre con el diagnóstico del tipo, encontramos que el nivel de sensibilidad del sistema varía en función de la memoria asociativa borrosa, tipo de conjunto borroso y nivel de solapamiento. De esta forma, las mayores diferencias significativas para los tres expertos las encontramos cuando



se utilizan las **MABs** Visser (Experto 3: n=45,  $p_{\alpha}<0.0001$ ) o Carlos (Experto 3: n=45,  $p_{\alpha}<0.0001$ ) con todas las posibles configuraciones, mientras que el sistema es menos sensible cuando se emplea la **MAB** Sole, conjunto borroso triangular y nivel de solapamiento del 10 % (n=48;  $p_{\alpha}=0.0034$ ) (tabla III.25).

**Tabla III.25.** Resultados del test exacto de Fisher entre el diagnóstico del agente patológico específico por parte de SEDPA para las MABs Visser, Sole y Carlos, conjuntos borrosos triangulares, gaussianos y trapezoidales, y niveles de solapamiento del 50 y 10 %, y expertos 1, 2 y 3 considerando sólo las respuestas correctas para el sistema y las respuestas correctas y aproximadas para los expertos

	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	20-18 <sup>†</sup>	14	13	9
<i>Fallo</i>	1-1 <sup>†</sup>	15	16	17
<b>MAB</b> Visser, Triangular-50-10, Gausiano-50-10, Trapezoidal-50-10 (Exp. 1: n=50, $p_{\alpha}=0.0003^*$ ; Exp. 2: n=50, $p_{\alpha}=0.0001^*$ ; <sup>†</sup> Exp. 3: n=45, $p_{\alpha}<0.0001^*$ )				
	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	19-16 <sup>†</sup>	14	13	9
<i>Fallo</i>	2-2 <sup>†</sup>	15	16	17
<b>MAB</b> Sole, Triangular-50, Gausiano-50, Trapezoidal-10 (Exp. 1: n=50, $p_{\alpha}=0.0018^*$ ; Exp. 2: n=50, $p_{\alpha}=0.0009^*$ ; <sup>†</sup> Exp. 3: n=44, $p_{\alpha}=0.0004^*$ )				
	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	17-15 <sup>†</sup>	14	13	9
<i>Fallo</i>	2-2 <sup>†</sup>	15	16	17
<b>MAB</b> Sole, Triangular-10 (Exp. 1: n=48; $p_{\alpha}=0.0034^*$ ; Exp. 2: n=48, $p_{\alpha}=0.0017^*$ ; <sup>†</sup> Exp. 3: n=43, $p_{\alpha}=0.0006^*$ )				
	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	19-17 <sup>†</sup>	14	13	9
<i>Fallo</i>	2-2 <sup>†</sup>	15	16	17
<b>MAB</b> Sole, Gausiano-10 (Exp. 1: n=50, $p_{\alpha}=0.0018^*$ ; Exp. 2: n=50, $p_{\alpha}=0.0009^*$ ; <sup>†</sup> Exp. 3: n=45, $p_{\alpha}=0.0002^*$ )				
	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	18-16 <sup>†</sup>	14	13	9
<i>Fallo</i>	1-1 <sup>†</sup>	15	16	17
<b>MAB</b> Sole, Trapezoidal-50 (Exp. 1: n=48, $p_{\alpha}=0.0007^*$ ; Exp. 2: n=48, $p_{\alpha}=0.0003^*$ ; <sup>†</sup> Exp. 3: n=43, $p_{\alpha}=0.0001^*$ )				
	<b>Diagnóstico SEDPA</b>	<b>Experto 1</b>	<b>Experto 2</b>	<b>Experto 3<sup>†</sup></b>
<i>Acierto</i>	20-18 <sup>†</sup>	14	13	9
<i>Fallo</i>	1-1 <sup>†</sup>	15	16	17
<b>MAB</b> Carlos, Triangular-50-10, Gausiano-50-10, Trapezoidal-50-10 (Exp. 1: n=50, $p_{\alpha}=0.0003^*$ ; Exp. 2: n=50, $p_{\alpha}=0.0001^*$ ; <sup>†</sup> Exp. 3: n=45, $p_{\alpha}<0.0001^*$ )				

\* Diferencias significativas

<sup>†</sup>Sólo se consideran 26 casos

De la misma forma que se hizo con el tipo de agente patógeno, cuando se tienen en cuenta al mismo tiempo las respuestas clasificadas como correctas así como las aproximadas, en todos los casos las diferencias encontradas entre las respuestas de **SEDPA** y los *expertos 1, 2 y 3* son significativas (tabla III.26).

**Tabla III.26.** Resultados del test exacto de Fisher entre el diagnóstico del agente patológico específico por parte de SEDPA para las MABs Visser, Sole y Carlos, conjuntos borrosos triangulares, gaussianos y trapezoidales, y niveles de solapamiento del 50 y 10 %, considerando las respuestas correctas y aproximadas del sistema y los expertos 1, 2 y 3

	Diagnóstico SEDPA	Experto 1	Experto 2	Experto 3 <sup>†</sup>
<i>Acierto</i>	28-25 <sup>†</sup>	14	13	9
<i>Fallo</i>	1-1 <sup>†</sup>	15	16	17
<b>MAB Visser</b> , Triangular-50-10, Gausiano-50-10, Trapezoidal-50-10 (Exp. 1: n=58, $p_{\alpha}<0.0001^*$ ; Exp. 2: n=58, $p_{\alpha}<0.0001^*$ ; <sup>†</sup> Exp. 3: n=52, $p_{\alpha}<0.0001^*$ )				
	Diagnóstico SEDPA	Experto 1	Experto 2	Experto 3 <sup>†</sup>
<i>Acierto</i>	27-24 <sup>†</sup>	14	13	9
<i>Fallo</i>	2-2 <sup>†</sup>	15	16	17
<b>MAB Sole</b> , Triangular-50-10, Gausiano-50-10, Trapezoidal-10 (Exp. 1: n=58, $p_{\alpha}=0.0002^*$ ; Exp. 2: n=58, $p_{\alpha}=0.0001^*$ ; <sup>†</sup> Exp. 3: n=52, $p_{\alpha}<0.0001^*$ )				
	Diagnóstico SEDPA	Experto 1	Experto 2	Experto 3 <sup>†</sup>
<i>Acierto</i>	28-25 <sup>†</sup>	14	13	9
<i>Fallo</i>	1-1 <sup>†</sup>	15	16	17
<b>MAB Sole</b> , Trapezoidal-50 (Exp. 1: n=58, $p_{\alpha}<0.0001^*$ ; Exp. 2: n=58, $p_{\alpha}<0.0001^*$ ; <sup>†</sup> Exp. 3: n=52, $p_{\alpha}<0.0001^*$ )				
	Diagnóstico SEDPA	Experto 1	Experto 2	Experto 3 <sup>†</sup>
<i>Acierto</i>	28-25 <sup>†</sup>	14	13	9
<i>Fallo</i>	1-1 <sup>†</sup>	15	16	17
<b>MAB Carlos</b> , Triangular-50-10, Gausiano-50-10, Trapezoidal-50-10 (Exp. 1: n=58, $p_{\alpha}<0.0001^*$ ; Exp. 2: n=58, $p_{\alpha}<0.0001^*$ ; <sup>†</sup> Exp. 3: n=52, $p_{\alpha}<0.0001^*$ )				

\* Diferencias significativas

<sup>†</sup>Sólo se consideran 26 casos

### 3.5.8. Selección de la mejor configuración de SEDPA

La selección de la mejor configuración de funcionamiento de **SEDPA** que proporcione al sistema un comportamiento estable tiene que llevarse a cabo a tres niveles distintos. El primero de ellos viene definido por la memoria asociativa. Así, si observamos los

resultados obtenidos tras el análisis preliminar de las memorias asociativas como el de la agrupación de los diagnósticos individuales de **SEDPA**, podemos comprobar como las **MABs** que presentan unas respuestas de la creencia más alejadas son Visser y Carlos, quedando las respuestas aportadas mediante la utilización de la **MAB** Sole más cercanas a Visser que a Carlos. Por otra parte, la **MAB** que hace que el sistema se comporte de una forma más caótica es Sole como se puede comprobar después del análisis de las respuestas de **SEDPA**. A pesar de que el comportamiento del programa es parecido cuando se utilizan las **MABs** Carlos y Visser a nivel de diagnóstico, también se ha comprobado que al comparar las creencias medias para los 29 casos con distintos conjuntos borrosos y niveles de solapamiento, siempre se obtienen marcadas diferencias significativas. Estas diferencias se deben a que las creencias obtenidas cuando se utiliza la **MAB** Visser son las más bajas mientras que cuando se utiliza la **MAB** Carlos las creencias son las más altas, lo que quiere decir que para un diagnóstico cualquiera, la **MAB** Carlos proporciona un nivel de creencia mayor. Esto podría clarificar la respuesta del usuario del programa cuando el diagnóstico es correcto, pero también puede hacer que se equivoque en mayor medida. Así podemos establecer que teniendo en cuenta tan sólo el primer nivel de estabilidad del sistema, la memoria asociativa borrosa que presenta un comportamiento más constante y conservador frente a su respuesta es Visser.

El segundo nivel de estabilidad viene determinado por el tipo de conjunto borroso. De esta forma el análisis de los resultados tras la comparación de las tasas de acierto para todas las memorias asociativas, conjuntos borrosos y niveles de solapamiento nos proporciona como solución óptima el empleo de conjuntos borrosos de forma trapezoidal, ya que con este tipo de conjuntos se consiguen las tasas de acierto más altas incluso cuando la memoria asociativa borrosa utilizada es Sole.

Por último es necesario establecer el nivel de solapamiento que proporcione los mejores resultados. Este puede ser determinado analizando las respuestas del sistema cuando utiliza conjuntos borrosos trapezoidales y la memoria asociativa que ofrece peores resultados (Sole). Esto es así debido a que los resultados conseguidos cuando se utiliza la **MAB** Visser son constantes para cualquier conjunto borroso y nivel de solapamiento. De esta forma se comprueba que al utilizar la **MAB** Sole, con conjuntos borrosos trapezoidales, las mejores tasas de acierto se obtienen con un nivel de solapamiento del 50 %. De este modo se puede establecer que el mejor comportamiento del sistema lo vamos a conseguir utilizando la **MAB** Visser, con conjuntos borrosos trapezoidales y niveles de solapamiento del 50 %.

La selección de la configuración óptima de funcionamiento se ve reforzada mediante el análisis del conflicto de evidencia. Así cuando se calcula el valor promedio del conflicto para cada una de las seis configuraciones por memoria asociativa y para cada caso que presenta conflicto (Casos 14 y 20), se observa que el menor valor de conflicto medio

para creencias se obtienen con la **MAB** Visser ([con=media  $\pm$  desviación estándar] Visser-Caso 14: con=0.5443  $\pm$  0.0294; Visser-Caso 20: con=0.9311  $\pm$  0.0185; Sole-Caso14: con=0.6014  $\pm$  0.0694; Sole-Caso 20: con=1.1871  $\pm$  0.0779; Carlos-Caso 14: con=1.0487  $\pm$  0.0208; Carlos-Caso 20: con=1.6985  $\pm$  0.0025).

### 3.5.9. Análisis de las respuestas de SEDPA con reducción de información

Para afinar aun más el diagnóstico de **SEDPA** se ha repetido el proceso llevado a cabo en apartados anteriores pero haciendo que la información suministrada al sistema sea más confusa. Así se han eliminado por una parte aquellos casos en los que el número de síntomas era menor o igual a 2 y por otra, en el resto de casos, aquellos síntomas asociados a una única patología. De este modo evitamos la preselección y la alta creencia asociada a estas patologías a través de la teoría de Dempster-Shafer. En estas condiciones el número de casos a evaluar se reduce por el número de síntomas a 24, de los que en tan solo en 5 ocasiones se produce eliminación de síntomas asociados a una única patología (Casos 1, 3, 5, 7 y 13) (tabla III.27). En esta nueva evaluación de **SEDPA**, la configuración del programa es la establecida como óptima en el apartado 3.5.8, es decir, la memoria asociativa Visser, conjunto borroso trapezoidal y nivel de solapamiento del 50 %.

Los resultados del nuevo diagnóstico de **SEDPA** para los cinco casos seleccionados muestran como excepto en el caso número 1, en el que el diagnóstico correcto era Virus-EVE y se obtiene Virus/Bacteria-EVE/*Vibrio*, las conclusiones obtenidas por el sistema son parecidas a las conseguidas a través de los síntomas originales. La diferencia fundamental en el segundo diagnóstico la encontramos en el valor de la creencia. Mientras que en el diagnóstico original la creencia en todos los grupos se encontraba entre el 0.8818 y 0.9788, en el diagnóstico con eliminación de síntomas no superó el 0.8080 del caso 3, alcanzándose el valor mínimo en el caso 13 (0.4668). Es en este último caso donde se produce una mayor regresión de la creencia. Este mismo efecto se produce en el caso 1 aunque no es tan pronunciado como en el caso 13 a pesar de una reducción de dos síntomas de un total de seis. Las diferencias en las creencias son menos acentuadas cuando comparamos las creencias compensadas a través del incremento de discrepancia entre grupos sintomáticos. Así, se puede observar como el único caso en el que no se produce un solapamiento de los límites de confianza es en el caso 13. A pesar de ello, la creencia compensada más optimista alcanza el valor de 0.6278 (tabla III.28).

**Tabla III. 27.** Selección de casos para el análisis de las respuestas de SEDPA bajo reducción de información

Caso	Estándar oro (tipo-género)	Nº de síntomas originales	Nº de síntomas finales	Eliminación	Selección
<b>1</b>	<b>Virus-EVE</b>	<b>6</b>	<b>4</b>	<b>No</b>	<b>Si</b>
2	Virus-EVE	4	4	No	No
<b>3</b>	<b>Bacteria-Vibrio</b>	<b>7</b>	<b>5</b>	<b>No</b>	<b>Si</b>
4	Bacteria-Vibrio	3	3	No	No
<b>5</b>	<b>Bacteria-Aeromonas</b>	<b>8</b>	<b>6</b>	<b>No</b>	<b>Si</b>
6	Bacteria-Pseudomonas	5	5	No	No
<b>7</b>	<b>Bacteria-Flexibacter</b>	<b>6</b>	<b>5</b>	<b>No</b>	<b>Si</b>
8	Bacteria-Edwardsiella	7	7	No	No
9	Bacteria-Yersinia	5	5	No	No
10	Otro-Dermocystidium	4	4	No	No
11	Hongo-Saprolegnia	3	0	Si	No
12	Otro-Papilomatosis	2		Si	No
<b>13</b>	<b>Parásito-Dactylogyrus</b>	<b>4</b>	<b>3</b>	<b>No</b>	<b>Si</b>
14	Parásito-Trichodina	3	3	No	No
15	Parásito-Myxidium	4	4	No	No
16	Parásito-Myxobolus	4	2	Si	No
17	Parásito-Ichtyobodo	4	4	No	No
18	Parásito-Henneguya	4	4	No	No
19	Parásito-Eimeria	1		Si	No
20	Parásito-Ergasilus	6	6	No	No
21	Parásito-Argulus	5	5	No	No
22	Parámetro físico-químico-NO <sub>2</sub>	2		Si	No
23	Parámetro físico-químico-NH <sub>3</sub>	3	3	No	No
24	Parámetro físico-químico-Saturación de O <sub>2</sub>	4	0	Si	No
25	Parámetro físico-químico-Subsaturación de O <sub>2</sub>	1		Si	No
26	Parámetro físico-químico-Turbidez	1		Si	No
27	Parásito-Gyrodactylus	4	4	No	No
28	Parásito-Ichthyophthirius	3	3	No	No
29	Parásito-Anguillicola	3	2	Si	No

**Tabla III.28.** Comparación de los resultados obtenidos por SEDPA para cinco casos en los que se utilizan los síntomas originales y aquellos asociados a más de una patología.  $LC_i$  = Límite Inferior de Confianza al 95 %;  $LC_s$  = Límite Superior de Confianza al 95 %

<b>Diagnóstico de SEDPA con los síntomas originales</b>						
Caso	Estándar oro (tipo-género)	Diagnóstico SEDPA	Creencia	Creencia compensada	$LC_i$	$LC_s$
1	Virus-EVE	Virus-EVE	0.9721	0.9721	0.9148	0.9953
3	Bacteria-Vibrio	Bacteria-Vibrio	0.9788	0.9834	0.9326	0.9988
5	Bacteria-Aeromonas	Bacteria-Aeromonas	0.9779	0.9800	0.9245	0.9986
7	Bacteria-Flexibacter	Bacteria-Flexibacter	0.8818	0.9297	0.8558	0.9715
13	Parásito-Dactylogyrus	Parásito-Dactylogyrus	0.9110	0.9743	0.9194	0.9957
<b>Diagnóstico de SEDPA tras la eliminación de síntomas asociados a una única patología</b>						
Caso	Estándar oro (tipo-género)	Diagnóstico SEDPA	Creencia	Creencia compensada	$LC_i$	$LC_s$
1	Virus-EVE	Virus/Bacteria-EVE/Vibrio	0.5674	0.7441	0.6348	0.8353
3	Bacteria-Vibrio	Bacteria-Vibrio	0.8080	0.9321	0.8598	0.9736
5	Bacteria-Aeromonas	Bacteria-Aeromonas	0.8011	0.9442	0.8750	0.9805
7	Bacteria-Flexibacter	Bacteria-Flexibacter	0.6448	0.9244	0.8470	0.9681
13	Parásito-Dactylogyrus	Parásito-Dactylogyrus	0.4668	0.4975	0.3676	0.6278

### 3.5.10. Respuesta del subsistema sugerencia o motor proposicional

Para comprobar la respuesta del subsistema sugerencia o motor proposicional se han seleccionado casos en los que el número de síntomas tras la eliminación de aquellos que se encontrasen asociados a una única patología fuese mayor a dos y a través de los cuales la respuesta del sistema fue tan solo aproximada. Bajo estas condiciones se seleccionaron cinco casos (4, 6, 8, 17 y 21). Con estos cinco casos y con la configuración óptima para el programa, se llevó a cabo nuevamente el diagnóstico con SEDPA preguntándole al final del proceso su sugerencia para poder seguir.

Para el primero de los casos (Caso 4, Bacteria-Vibrio) SEDPA selecciona cuatro posibles patologías asociadas al último síntoma introducido: *V. anguillarum*, *P.*

*anguilliseptica*, *E. tarda* y *Y. ruckeri*. Cuando se le pide al sistema una sugerencia para poder continuar el proceso de diagnóstico, éste construye las matrices  $\mathbf{CCL}_{sel}$  y  $\mathbf{CD}_{sel}$  (apartado 3.4.4) de la que se extraen seis y siete etiquetas sintácticas respectivamente que cumplen las ecuaciones (3.29) y (3.30). De la matriz  $\mathbf{CCL}_{sel}$ , cuatro etiquetas sintácticas quedan asociadas a *P. anguilliseptica* con un valor promedio del factor de impacto (ecuación 3.33) de  $87.96 \pm 11.86$ , una se relaciona con *V. anguillarum* (factor de impacto=88.89) y otra con *E. tarda* (factor de impacto=85.19). En la segunda matriz ( $\mathbf{CD}_{sel}$ ) **SEDPA** extrae un total de siete etiquetas de las cuales cuatro se agrupan con *V. anguillarum* (factor de impacto promedio= $92.6 \pm 5.24$ ), dos con *Y. ruckeri* (factor de impacto promedio= $90.75 \pm 7.86$ ) y una con *P. anguilliseptica* (factor de impacto=85.19). Teniendo en cuenta las dos matrices, se puede observar que la respuesta del usuario va a tener una mayor importancia sobre el diagnóstico final del sistema y sobre la creencia global del grupo sintomático si selecciona las etiquetas asociadas a *V. anguillarum* y *P. anguilliseptica*, tanto si la respuesta es SI como si es NO.

En el Caso 6 (Bacteria-*Pseudomonas*) el motor proposicional extrae cinco etiquetas de la matriz  $\mathbf{CCL}_{sel}$  y ocho de la  $\mathbf{CD}_{sel}$ . Estas etiquetas se reparten entre cinco patologías (*V. anguillarum*, *P. anguilliseptica*, *E. tarda*, *F. columnaris* y *A. hydrophila*) seleccionadas después de la introducción del último síntoma. En esta ocasión el subsistema de sugerencia se comporta de forma parecida al caso anterior aunque las diferencias entre factores de impacto promedio se reducen. En la primera matriz el factor de impacto más alto lo alcanza *P. anguilliseptica* (3 etiquetas;  $93.83 \pm 2.14$ ) mientras que en la segunda el máximo valor es alcanzado por *V. anguillarum* (3 etiquetas;  $95.06 \pm 2.14$ ). Como se puede observar, en el Caso 6 la decisión del usuario va a tener incluso más importancia que en el Caso 4.

En el Caso 8 (Bacteria-*Edwardsiella*) encontramos que el número de patologías seleccionadas después del último síntoma coinciden con el diagnóstico final de **SEDPA** (*E. tarda* y *V. anguillarum*). De la misma forma que en los dos casos anteriores, el factor de impacto promedio más elevado se asocia a *V. anguillarum* (3 etiquetas en la matriz  $\mathbf{CCL}_{sel}$ :  $87.04 \pm 2.62$ ; 6 etiquetas en la matriz  $\mathbf{CD}_{sel}$ :  $85.92 \pm 15.58$ ) frente al valor más bajo de *E. tarda* (1 etiqueta en la matriz  $\mathbf{CCL}_{sel}$ : 85.19).

En los Casos 17 y 21 (Parásito-*Ichthyobodo*; Parásito-*Argulus*) se produce una circunstancia que no había sucedido en los casos anteriores. Después de la introducción del último síntoma el número de patologías seleccionadas para el Caso 17 son seis: *A. hydrophila*, *V. anguillarum*, *P. anguilliseptica*, *F. columnaris*, *E. tarda*, *Ichthyobodo* sp. y *Ergasilus* sp., y para el Caso 21 ocho: *D. anguillae*, *A. hydrophila*, *V. anguillarum*, *P. anguilliseptica*, **EVE**, *Ergasilus* sp., *Argulus* sp., y Efecto Saturación de Oxígeno. Sin embargo, en ambos casos cuando se le requiere al sistema una sugerencia, las etiquetas seleccionadas se reparten entre patologías entre las que no se encuentra ni *Ichthyobodo* sp. (Caso 17) ni *Argulus* sp. (Caso 21). Así, el factor de impacto promedio más alto para el

Caso 17 lo encontramos en *Ergasilus* sp. asociado a la matriz  $\mathbf{CCL}_{sel}$  ( $96.06 \pm 4.90$ ) y el más bajo es el de *F. columnaris* asociado a la matriz  $\mathbf{CD}_{sel}$  ( $85.19 \pm 11.12$ ), mientras que en el Caso 21 el factor de impacto medio más alto se encuentra compartido entre *V. anguillarum*, **EVE** y Efecto Saturación de Oxígeno ( $96.30 \pm 0.00$ ) mientras que el más bajo se asocia a *Ergasilus* sp. (1 etiqueta en la matriz  $\mathbf{CD}_{sel}$ : 85.19).

### 3.6. DISCUSIÓN

La realización de trabajos previos en el campo de los sistemas expertos aplicados a la acuicultura han demostrado que este tipo de programas pueden ser de utilidad en una gran variedad de propósitos como el análisis financiero, manejo de la producción, formulación de raciones alimentarias, monitorización, control, diagnóstico y tratamiento de enfermedades, las cuales están íntimamente relacionadas con la correcta gestión de los sistemas de producción en medio acuático.

Zeldis y Prescott (2000) desarrollaron el sistema FISH-VET que es capaz de trabajar con multitud de especies y diagnosticar gran cantidad de patologías asociadas al manejo y cultivo de peces. Este tipo de sistemas abiertos a una gran cantidad de posibles entradas y salidas tienen el inconveniente de que es difícil que alcancen una solución particular, ya que este tipo de eventos complica la construcción de las bases de datos o bases de conocimiento del cual se nutre el sistema. De esta forma, un sistema con un dominio de funcionamiento más restringido puede alcanzar soluciones que faciliten la labor de los técnicos encargados de la política sanitaria de la explotación. Sin embargo, sistemas muy particularizados y contruidos bajo hipótesis de trabajo especiales (falta de datos, información confusa o redundante) pueden implicar una inversión económica y temporal excesivamente alta. Además, en este tipo de sistemas el conocimiento suele incorporarse en forma de reglas por lo que la ampliación o modificación del conocimiento suele ser compleja e implica la intervención del diseñador del programa, lo que puede suponer importantes gastos de mantenimiento. Estos factores hacen que el número de programas con estas características en el mercado no sea excesivamente elevado.

**SEDPA** conjuga las ventajas de los dos tipos de sistemas descritos anteriormente y evita alguno de sus problemas. Por una parte es un programa cuyo dominio de trabajo se reduce a una única especie. Esto implica respuestas muy concretas y particularizadas por lo que la agudeza del diagnóstico probablemente es mayor que la alcanzada por sistemas de propósito general. Además **SEDPA** va a ser utilizado por usuarios cuyo ámbito de trabajo es la anguila por lo que, tanto los posibles fallos en el diseño así como los errores contenidos en las bases de conocimiento van a ser más fácilmente detectables.



Por otro lado, el conocimiento adquirido por **SEDPA** parte del propósito general y no se encuentra almacenado en forma de reglas. Esto le confiere al sistema una gran versatilidad ya que para la ampliación o modificación del conocimiento no es necesario la intervención del diseñador, sino que es el propio usuario el que lleva a cabo la ampliación o modificación a través del módulo de aprendizaje. Esto quiere decir que el comportamiento y las conclusiones obtenidas por **SEDPA** implementado en dos explotaciones pueden ser totalmente distintas, simplemente por una adaptación del sistema por los técnicos correspondientes a condiciones de explotación diferentes (localización, productividad, mortalidad, etc). Aun más, el sistema puede ser útil a otras explotaciones que no sean de anguilas cuyos niveles de información sean semejantes con tal de sustituir las relaciones binarias de las bases de Conocimiento del Dominio Principal. Esto es posible gracias a que la obtención de conclusiones por el motor de inferencia del programa es independiente de reglas que relacionen a éste con la especie en concreto. Es decir, el sistema es capaz de obtener conclusiones a partir de la propia información contenida en las bases de conocimiento, en lugar de utilizar esta información para hacer cumplir reglas predeterminadas en el motor de inferencia. Así, como en cualquier otro sistema experto, cuanta más información cierta sea incorporada a las bases de conocimiento más aproximadas serán las conclusiones aunque no a costa de la construcción de nuevas reglas.

Como se puede comprobar, la independencia de reglas aporta importantes ventajas aunque desgraciadamente, también grandes inconvenientes. De esta forma, un sistema basado en reglas como GOLDFINDER (Hawkes, 1992), MUNIN (Andreassen *et al.*, 1995) o XPHEMO (Nguyen *et al.*, 1996) puede proporcionar al usuario valiosa información sobre el proceso de razonamiento que se ha llevado a cabo hasta obtener las conclusiones, simplemente retrocediendo a través de las reglas satisfechas. Este tipo de razonamiento hacia atrás no es posible en **SEDPA** debido a la inexistencia de dichas reglas, convirtiendo al sistema en una especie de caja negra.

### **3.6.1. Respuestas de los expertos humanos**

El análisis de las respuestas aportadas por los tres expertos evaluadores mostraron distintos comportamientos frente a las encuestas. De la forma de responder se puede establecer una jerarquía en cuanto al conocimiento soportado por cada experto. Tanto para el *experto 1* como para el *experto 3* no hubo diferencias significativas entre las frecuencias de agentes patógenos reales o ‘estándar oro’ y las frecuencias obtenidas como consecuencia de sus respuestas, mientras que para el *experto 2* si se obtuvieron diferencias. Esto indica que tanto el *experto 1* como el *3* no centraron su respuesta en ninguno de los tipos de agentes patógenos. En el caso del *experto 2* estas diferencias se debieron fundamentalmente a la incorporación de un nuevo grupo (además de los establecidos por los estándares oro) que se

encontraba constituido por aquellos casos en los que el diagnóstico estaba compuesto por más de un agente patógeno. Sin embargo, a pesar de que el *experto 3* también incorpora este nuevo grupo, el test  $\chi^2$  no mostró la presencia de diferencias significativas ya que éste sitúa en ese nuevo grupo tan solo tres casos. Por tanto, se muestra un carácter más arriesgado en el diagnóstico del *experto 1* y *3* frente al *experto 2*, lo que posiblemente puede deberse a la mayor experiencia de los *expertos 1* y *3*. Esto queda de manifiesto cuando al responder sobre el agente etiológico específico, la mayor disminución en la tasa de acierto se produce en el *experto 2*.

Cuando se comparan las frecuencias de los agentes etiológicos y las frecuencias de los agentes en las respuestas acertadas y falladas nuevamente se puede observar que para los *expertos 1* y *3* no se encuentran diferencias significativas. Es decir, ambos expertos tienden a equivocarse y a acertar más en aquellos agentes patológicos donde encontramos más casos. Sin embargo cuando se lleva a cabo el mismo proceso con el *experto 2* se encuentran marcadas diferencias significativas. De este modo, el *experto 2* tiende a responder con un grupo de agentes (normalmente 2) cuando se encuentra con síntomas o con combinación de síntomas que desconoce.

Cuando los expertos asignaron niveles de creencias a sus diagnósticos, los tres respondieron de la misma forma, ya que no se encontraron asociaciones estadísticas entre tipos de creencia y respuestas correctas e incorrectas. Sin embargo, la valoración final del diagnóstico fue tratada de forma distinta por los tres expertos. Mientras que para los *expertos 2* y *3* se obtuvieron diferencias significativas entre las frecuencias de los tipos de valoraciones y las frecuencias de las valoraciones en respuestas correctas e incorrectas, en el *experto 1* estas diferencias no llegaron a ser estadísticamente aceptables. En el caso del *experto 3*, estas diferencias se debieron a que una gran parte de los casos diagnosticados incorrectamente fueron asignados a la categoría valoración Prometedora. En el *experto 2*, la asignación prioritaria de los diagnósticos incorrectos se llevó a cabo en la categoría valoración Especulativa, siendo la mayor parte de los diagnósticos clasificados como valoración Muy Prometedora correctos. Es decir, tanto el *experto 2* como el *3* llevan a cabo valoraciones finales optimistas mientras que en el *experto 1* no se detecta una tendencia clara a valorar positiva o negativamente en función de que la respuesta sea correcta o incorrecta, por lo que en este aspecto se presenta más conservador.

Quizás algo más sorprendente sea la valoración final, ya que prácticamente los tres coincidieron en la dificultad que supone llevar a cabo un diagnóstico a través de este tipo de encuestas. Más concretamente apuntaron de forma general la ausencia de datos precisos sobre la calidad del agua, historial clínico, manejo de animales previo a la observación de la enfermedad (clasificaciones, trasposos, etc.), evolución de la enfermedad, tratamientos recibidos, número de muertos y consumo de alimento. Todos esos aspectos son necesarios para una correcta valoración final del diagnóstico. **SEDPA** tiene en cuenta estos factores a

través de las reglas **A-A** de **CONTROLPLANT**, pero para los casos en estudio los datos disponibles eran muy limitados. Sin embargo, se han encontrado en los tres casos una alta correlación significativa entre el nivel de creencia asignado teniendo en cuenta tan solo los síntomas y la valoración final teniendo en cuenta además las condiciones físico-químicas y ambientales. De esta forma, cuando la creencia asignada por el experto era Alta la valoración final era Prometedora, y cuando la creencia era Muy Baja la valoración era Nada Prometedora, indicando que la valoración final (apartado 4 de la encuesta) estaba condicionada básicamente por el nivel de creencia asignado en apartados anteriores (2 y 3). **SEDPA** en cambio, evalúa en todo momento la valoración para todos los casos con los datos disponibles entre Nada Prometedora y Muy Especulativa denotando una clara independencia entre creencias y valoraciones.

Los tres expertos muestran un comportamiento frente a eventos difusos (como los síntomas que describen los casos) ya comentados por otros autores (Sanz-Molina -comentario personal-). Según esta pauta, cuando un experto se encuentra frente a uno o dos acontecimientos difusos la probabilidad de error en la toma de decisiones o en el diagnóstico es alta. La probabilidad de error disminuye a medida que un mayor número de eventos se incorporan al marco de discernimiento, pero si el número de eventos aumenta excesivamente, el experto tiene dificultades para obtener conclusiones aumentando nuevamente la probabilidad de error. Se cumple asimismo que el número de eventos difusos que los expertos son capaces de procesar correctamente, es decir, con tasas de error menores, se encuentra entre tres y cinco (exceptuando el Caso 5 con ocho síntomas). Esto se corresponde con la teoría de conjuntos borrosos en la que se recomienda un número de particiones para los conjuntos comprendido entre tres y siete (Martín del Brío y Sanz-Molina, 1997; Tsoukalas y Uhrig, 1997).

### **3.6.2. Lenguaje natural**

La introducción de datos por parte de los usuarios es un aspecto fundamental del funcionamiento de cualquier sistema experto. Este problema ha sido resuelto por varios autores de maneras muy diversas, sin embargo, son raros los programas a los que se incorpora información a través de un módulo de interpretación del lenguaje natural. Esto puede estar relacionado con el hecho de que hasta ahora los sistemas expertos desarrollados han sido programas con dominios de trabajo relativamente amplios, o más probablemente, con que es un problema que puede ser abordado de muchas formas diferentes o que de por sí constituya un sector totalmente segregado de la inteligencia artificial (McKeown, 1985). Sin embargo, en sistemas como **SEDPA** con un espectro de trabajo localizado, la introducción de datos a través del lenguaje natural aporta importantes ventajas. Así un primer aspecto a considerar es el tipo de usuario que puede manejar el programa. Sistemas

como los propuestos por Fatma, *et al.*, 1993, Haddawy *et al.*, 1994, Nguyen *et al.*, 1996, Kahn *et al.*, 1997 entre otros, sólo pueden ser utilizados por expertos en el campo de trabajo del programa. Esto aplicado a un sistema como **SEDPA** puede implicar su infrautilización temporal perdiéndose por tanto rapidez en el diagnóstico. Sin embargo **SEDPA** puede ser utilizado por técnicos y operarios, ya que tanto para unos como para otros la entrada de datos al sistema se limita a escribir oraciones simples que describen lo que están viendo. Esto es posible gracias al control de coherencia y a la estructura del análisis morfológico-sintáctico que impone la formación de la Red de Transición Aumentada.

Otro aspecto importante a considerar es que la respuesta del usuario no se encuentra condicionada al no tener que elegir de una lista de términos o una selección dicotómica. Esto puede suceder en sistemas como CLICHE (Fatma *et al.*, 1993) donde para avanzar en el diagnóstico es necesario hacer una selección de entre todos los posibles síntomas recogidos en las bases de datos.

### **3.6.3. Respuestas de SEDPA: creencias y valoraciones. Conclusiones del motor de inferencia y proposional**

Las respuestas de **SEDPA** frente a los mismos casos planteados a los expertos humanos han mostrado que el sistema es una herramienta potencialmente válida en la evaluación y diagnóstico de los procesos patológicos que pueden producirse en la explotación, y que por tanto, puede utilizarse como un complemento más por los técnicos y operarios responsables de la política sanitaria de la planta.

La asignación de niveles de creencia no compensada sobre la fiabilidad del diagnóstico así como las tasas de acierto encontradas, indican que el proceso de razonamiento llevado a cabo por el motor de inferencia a través de un grupo de memorias asociativas borrosas y conjuntos borrosos con sus correspondientes niveles de solapamiento, es capaz de aportar resultados cuantitativos a partir de la introducción por el usuario de evidencias altamente difusas. Una importante ventaja que proporcionan los sistemas basados en lógica borrosa es la amortiguación de las respuestas debido a la habilidad de trabajar con cambios continuos y de generar pequeños o grandes controles en función de los síntomas asociados al problema. Así, podemos observar como el nivel de creencia asignado por el programa tras la introducción de una serie de síntomas puede no llegar a ser muy alto (con la configuración óptima), a pesar de que la respuesta sea correcta. Incluso si se utiliza un término asociado a una única patología, la creencia individual que el sistema asigna a dicha patología no tiene por que ser uno. Se podría pensar que esta propiedad puede ser más un inconveniente que una ventaja, pero de esta forma se consigue que el sistema se convierta en un elemento más de la explotación cuya misión es aportar

‘algo más’ al técnico, que en última instancia es quien toma las decisiones. Por otra parte también se consigue que cuando el sistema se equivoca, el error cometido sea bajo. En definitiva, la utilización de la lógica borrosa hace que **SEDPA** se comporte como un buen experto que al mismo tiempo presenta una actitud cauta con sus conclusiones.

Lee *et al.* (2000) señalan, en su estudio sobre un sistema de desnitrificación para acuicultura en el que emplean un controlador borroso, que el sistema experto diseñado a tal efecto fue capaz de controlar las acciones generadas de una forma conservativa y que las acciones derivadas de la actuación del sistema experto estuvieron altamente correlacionadas con las tomadas por expertos humanos. Los resultados indicaron además que dada la complejidad del sistema, las reglas basadas en lógica borrosa fue una elección apropiada superando en todo momento al proceso de control mediante técnicas clásicas. Estos resultados se encuentran en la misma línea que los obtenidos por otros autores como Mamdani (1977), Takagi y Sugeno (1985) y Whitsell y Lee (1994).

A pesar de que la teoría de Dempster-Shafer es un método reconocido como sistema de transmisión de la incertidumbre (Hájek, 1994), su utilización en casos prácticos ha sido hasta la actualidad bastante rara. Una posible explicación a este hecho puede ser la dificultad del problema de representación del conocimiento de los expertos en el marco de la teoría y su posterior proceso de validación. Otra posible causa puede atribuirse a que en condiciones normales el volumen de información disponible es grande, por lo que otros métodos de propagación de la incertidumbre son más recomendables. En **SEDPA**, la información obtenida de múltiples fuentes de evidencia es combinada mediante la teoría de Dempster-Shafer. Esta forma de utilizar la información ha proporcionado al sistema un alto nivel de discriminación de los distintos procesos patológicos que pueden incidir en la explotación, así como indicios de conflicto generados por algunos de los síntomas introducidos. Murphy (1998) explota estos dos componentes de la teoría de Dempster-Shafer en la construcción de un robot móvil autónomo: el conflicto y el agrandamiento del marco de discernimiento. De esta forma, el conflicto es usado como una medida de la cantidad de consenso entre los diferentes sensores del robot mientras que agrandar el marco de discernimiento permite una descomposición modular de la evidencia. Esta descomposición hace que el experto pueda integrar su conocimiento en el marco de discernimiento, simplificando la construcción y mantenimiento de la base de conocimiento.

Como se comentó en el apartado 3.4.2.3 un grave inconveniente con el que se enfrenta el método es que, frente a una gran cantidad evidencias asociadas a síntomas con niveles altos de solapamiento, las conclusiones del sistema pueden ser poco intuitivas. De esta forma una respuesta que esté compuesta por cinco grupos sintomáticos en la cual la creencia más alta asignada sea de 0.26 puede parecer poco satisfactoria a pesar de que la respuesta sea correcta. Esto posiblemente sea debido a la tendencia de asociar probabilidades con creencias. Así, una probabilidad de 0.26 es una probabilidad baja. Sin

embargo, una creencia en el marco de la teoría de Dempster-Shafer de 0.26 puede corresponderse con una probabilidad baja o alta dependiendo del número de grupos. El reescalado de la creencia medido como un incremento porcentual de discrepancia entre eventos consecutivos a través de un test  $\chi^2$  modificado, permite obtener una medida de creencia de más fácil interpretación por parte del usuario.

El reescalado de las creencias no se llevó a cabo linealmente, ya que éstas aumentaron en aquellos casos a los que se asignó creencias bajas a través de la teoría de Dempster-Shafer y estaban soportados por un número de evidencias suficientes. Este incremento fue más importante cuando se detectó conflicto entre los últimos síntomas introducidos por el usuario (por ejemplo el Caso 14). Por otra parte, cuando las creencias asignadas inicialmente eran altas los incrementos fueron proporcionalmente inferiores indicando un comportamiento distinto del sistema de reescalado en estos niveles. Es decir, el sistema se comporta como un potencial de acción en el que el umbral de actuación viene marcado por la creencia original, el número de síntomas y nivel de entrada en el que se detecta un conflicto entre síntomas. El sistema también es sensible a la configuración de **SEDPA** ya que se han encontrado mayores desajustes cuando el nivel de solapamiento era del 10 %.

Las conclusiones del motor proposicional han sido relativamente satisfactorias ya que de los cinco casos probados tan solo en tres de ellos se seleccionó el agente responsable de la enfermedad. Esta falta de especificidad en la selección de los agente patógenos puede ser debida al intento de modelar el lenguaje desde un punto de vista estadístico, tratando un término como una entidad individual cuya frecuencia es independiente del resto de términos que le rodean. Biber *et al.* (1998), Oakes (1998) y Manning y Schutze (1999) señalan que la respuesta de este tipo de modelos puede llegar a ser dependiente del volumen de información manejada. Esta dependencia queda patente en **SEDPA** donde se observa una tendencia en los cinco casos a seleccionar las patologías que el análisis de componentes principales había encontrado como las asociadas a un mayor volumen de información y con términos más particulares.

#### **3.6.4. Expertos humanos frente a SEDPA**

A través del proceso de evaluación se ha comprobado como **SEDPA** se comporta correctamente cuando se compara su respuesta frente a la aportada por expertos humanos. Esto queda confirmado al comprobar que de los 29 casos mostrados tanto al sistema como a los expertos, el sistema diagnosticó correctamente el tipo de agente etiológico en el 86.2 % de las ocasiones, alcanzándose de esta forma el mejor resultado obtenido por los expertos (86.2 %). Sin embargo hay que tener en cuenta que en la mejor tasa de acierto de los

expertos humanos se consideran tanto las respuestas completamente correctas como las aproximadas. Así, si consideramos también las respuestas aproximadas en **SEDPA**, ésta aumenta hasta el 96.6 %. En estas condiciones se observan diferencias singnificativas marginales cuando la memoria asociativa borrosa utilizada es Visser o Carlos. También se puede comprobar como al bajar a nivel de definición del agente etiológico responsable de la enfermedad, el sistema también responde correctamente encontrándose en todos los casos diferencias significativas importantes ( $p_{\alpha} < 0.01$ ) con las respuestas de los expertos. De esta forma obtenemos que considerando tan sólo las respuestas correctas, el nivel de acierto se sitúa para **SEDPA** alrededor del 70 % mientras que el mejor resultado conseguido por el experto humano no supera el 50 %.

Las diferencias encontradas entre **SEDPA** y los expertos humanos pueden venir determinadas por la distinta forma de utilización de la información. Los expertos humanos hacen un uso tradicional de esta información asumiendo que ésta se compone de un grupo de términos que pueden ser combinados para formar progresivamente componentes más complejos que son entendidos bajo nociones convencionales del lenguaje. Según Chomsky (1959) en estos modelos los términos, que en su conjunto definen la información, pueden ser agrupados en categorías específicas y pueden combinarse usando tan solo reglas gramaticales para producir sentencias. Se podría decir por tanto que la información transmitida por la descripción de un síntoma adquiere un carácter cualitativo y sirve como catapulta hacia la búsqueda de una confirmación cuantitativa (análisis químicos del agua, histórico de las condiciones ambientales, cultivos, etc.). Este efecto se ve acentuado cuando los agentes etiológicos presentan síntomas muy parecidos y el número de eventos observados por el experto es más de uno. Si a esto le unimos la falta de información cuantitativa se consigue que el diagnóstico sea difícilmente definible. Sin embargo **SEDPA** no lleva a cabo un uso tradicional de la información aportada por el usuario, sino que ésta es tratada de forma estadística, calculando la frecuencia de la ocurrencia de los términos y estableciendo relaciones entre los distintos grupos de términos. Así, combinando la información en este primer nivel es posible extraer la asociación de eventos difusos con grupos de patologías y la asignación de creencias. Como consecuencia, el sistema es capaz de aportar una respuesta aceptable cuando se eliminan aquellos síntomas que en principio se encuentran asociados a un único tipo de enfermedad.

Otro factor que ha podido contribuir a las diferencias encontradas es la falta de confirmación visual. Un buen diagnóstico debe cotejar toda la información disponible con las percepciones visuales del experto. Hay que tener en cuenta que, aunque la descripción de los síntomas y lesiones en cada uno de los 29 casos se llevó a cabo por un experto humano lo más sistemáticamente posible, puede ser que la información descrita llegue a ser confusa para otro experto si los efectos patológicos no pueden ser observados. De esta forma estructuras sintácticas tales como grande o pequeño, rosado o anaranjado que *in situ* pueden confirmar la existencia de una u otra enfermedad en base a la experiencia del

experto, pierden importancia cuando dichas características no pueden ser confirmados visualmente.

### 3.7. REFERENCIAS

- Afgan, N.H., L.M. Radanovic y A.I. Tikhonov. 1994. SRMES: an expert system performance analysis of small refrigerating machines. *International Journal of Refrigeration*, 17(8): 532-537.
- Amador-Hidalgo, L. 1996. *Inteligencia artificial y sistemas expertos*. Edita: Universidad de Córdoba. Córdoba.
- Andreassen, S., A. Rosenfalck, B. Falck, K.G. Olesen y S.K. Andersen. 1995. Evaluation of the diagnostic performance of the expert EMG assistant MUNIN. *Electroencephalography and Clinical Neurophysiology*, 101: 129-144.
- Bernardet, J.F. 1998. *Cytophaga, Flavobacterium, Flexibacter* and *Chryseobacterium* infections in cultured marine fish. *Fish Pathology*, 33(4): 229-238
- Berry, D.C. y A.E. Hart. 1990. Evaluating expert systems. *Expert Systems*, 7: 199-208.
- Berthe, F.C.J., C. Michel y J.F. Bernardet. 1995. Identification of *Pseudomonas anguilliseptica* isolated from several fish species in France. *Diseases of Aquatic Organisms*, 21: 151-155.
- Biber, D., S. Conrad y R. Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Edita: Cambridge University Press. Londres.
- Birndorf, N.I., J.O. Pentecost, J.R. Coakley y K.A. Spackman. 1996. An expert system to diagnose anemia and report results directly on hematology forms. *Computers and Biomedical Research*, 29: 16-26.
- Bryan S.T., R. Stamper y P. MacPherson. 1994. The design and construction of a medical simulation model. *Computer Methods and Programs in Biomedicine*, 42: 77-91.
- Campbell, A.N., V.F. Hollister, R.O. Dura y P.E. Hart. 1982. Recognition of a hidden mineral deposit by an artificial intelligence program. *Science*, 217: 927-929.
- Castillo, E., J.M. Gutiérrez y A.S. Hadi. 1996. *Sistemas expertos y modelos de redes probabilísticas*. *Monografías de la Academia de Ingeniería*. Edita: Academia de Ingeniería. Madrid.



- Cazorla-Quevedo, M.A., O. Colomina-Pardo, F. Escolano-Ruiz, D. Gallardo-López, R. Rizo-Aldeguer y R. Satorre-Cuerda. 1999. *Técnicas de inteligencia artificial*. Textos Docentes de la Universidad de Alicante. Edita: Universidad de Alicante. Alicante.
- Chang, P.L., Y.C. Li, C.J. Wu, M.H. Huang y P.J. Haug. 1994. Clinical evaluation of a renal mass diagnostic expert system. *Computers in Biology and Medicine*, 24(4): 315-322.
- Chomsky, N. 1959. On certain formal properties fo grammars. *Information and Control*, 2(2): 137-167.
- Clifton-Hadley, R.S., D. Bucke y R.H. Richards. 1984. Proliferative kidney disease of salmonid fish: a review. *Journal of Fish Diseases*, 7: 363-377.
- Cohen, J. 1968. Weighted Kappa: nominal scale agreement with provision for scaled disagreement of partial credit. *Psychological Bulletin*, 70: 213-220.
- Conover, W.J. 1980. *Practical nonparametric statistics*. Edita: John Wiley and Son. New York.
- De Carolis, B., F. Derosis, F. Grasso, A. Rossiello, D.C. Berry y T.Gillie. 1996. Generating recipient-centered explanations about drug prescription. *Artificial Intelligence in Medicine*, 8(2): 123-145.
- Dempster, A.P. 1968. A generalization of bayesian inference. *Journal of the Royal Statistical Society*, 30(B): 205-247.
- Denæux, T. 1995. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transctions on Systems, Man, and Cybernetics*, 25(5): 804-813.
- Denæux, T. 1997. Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition*, 30(7): 1095-1107.
- Díez, F.J., J. Mira, E. Iturralde y S. Zubillaga. 1997. DIAVAL, a bayesian expert system for echocardiography. *Artificial Intelligence in Medicine*, 10: 59-73.
- Edgington, E.S. 1980. *Randomization tests*. Edita: Marcel Dekker. New York.
- Egidius, E. 1987. Vibriosis: pathogenicity and pathology. A review. *Aquaculture*, 67: 15-28.

- Fatma, F.A., B.V. Faltings y C. Bachmann. 1993. A diagnostic assistant in clinical chemistry. *Chemometrics and Intelligent Laboratory Systems: Laboratory Information Management*, 21: 267-273.
- Firriolo, F.J. y B.A. Levy. 1996. Computer expert system for the histopathologic diagnosis of salivary gland neoplasms. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontics*, 82(2): 179-186.
- Fixen, D. y R.P.S. Mahler. 1997. The modified Dempster-Shafer approach to classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 27(1): 96-104.
- Garnerin, P. y G. Tuffery. 1988. SCHUBERT 3000: a interactive system to aid production and health management of fish farms. En *Proceedings of the first international workshop 'Applications of artificial intelligence to Agricultural, Agrochemical and Food Processing Industries'*. Caen (Francia).
- González-Andújar, J.L. y B. Recio-Aguado. 1997. *Aplicación de los sistemas expertos en la agricultura*. Edita: Ministerio de Agricultura, Pesca y Alimentación y Mundi-Prensa. Madrid.
- Haddawy, P., C.E. Kahn y M. Butarbutar. 1994. A bayesian network model for radiological diagnosis and procedure selection: work-up of suspected gallbladder diseases. *Medical Physics*, 21(7): 1185-1192.
- Hájek, P. 1994. Systems of conditional beliefs in Dempster-Shafer theory and expert systems. *International Journal of General Systems*, 22: 113-124.
- Hair, J.P., R.E. Anderson, R.L. Tatham y W.C. Black. 1999. *Análisis multivariante*. Edita: Prentice Hall. Madrid.
- Hawkes, D.D. 1992. GOLDFINDER: a knowledge-based system for mineral prospecting. *Journal of the Geological Society*, 149: 465-471.
- Hawkes, D.D. 1993. Expert system to aid mineral prospecting in epithermal gold terrain. *Transactions of the Institution of Mining and Metallurgy*, 102:135-140.
- Hedrick, R.P., N. Okamoto, T. Sano y J.L. Fryer. 1983. Biochemical characterization of eel virus european. *Journal of General Virology*, 64: 1421-1426.
- Hernando, M.E., E.J. Gómez, F. del Pozo y R. Corcoy. 1996. DIABNET: a qualitative model-based advisory system for therapy planning in gestational diabetes. *Medical Informatics*, 21(4): 359-374.

- Hilera, J.R. y V.J. Martínez. 1995. *Redes neuronales artificiales. Fundamentos, modelos y aplicaciones*. Edita: Ra-Ma. Madrid.
- Höglund, J. y J. Andersson. 1993. Prevalence and abundance of *Anguillicola crassus* in the european eel (*Anguilla anguilla*) at a thermal discharge site on the Swedish coast. *Journal of Applied Ichthyology*, 9: 115-112.
- Hoskins, G.E., C.J. West y A.E. Kling. 1983. Computer management of fish health problems in hatcheries. *Canadian Special Publication on Fisheries and Aquatic Sciences*, 68: 1-42.
- Ivandic, M., W. Hofmann y W.G. Guder. 1996. Development and evaluation of a urine protein expert system. *Clinical Chemistry*, 42(8): 1214-1222.
- Jimison, H.B., L.M. Fagan, L.D. Schachter y E.H. Shortlife. 1992. Patient-specific explanation in models of chronic diseases. *Artificial Intelligence in Medicine*, 4: 191-205.
- Kahn, C.E. 1996. Decision aids in radiology. *Radiologic Clinics of North America*, 34(3): 607-628.
- Kahn, C.E. y G. Anderson. 1994. Case-based reasoning and imaging procedure selection. *Investigate Radiology*, 29(6): 643-647.
- Kahn, C.E., L.M. Roberts, K.A. Shaffer y P. Haddawy. 1997. Construction of a bayesian network for mammographic diagnosis of breast cancer. *Computers in Biology and Medicine*, 27(1): 19-29.
- Kent, M.L., L. Margolis, D.J. Whitaker, G.E. Hoskins y T.E. McDonall. 1993. Review of Myxosporea of importance in salmonid fisheries and aquaculture in British Columbia. *Folia Parasitologica*, 41: 27-37.
- Kentala, E., I. Pyykkö, Y. Auramo y M. Juhola. 1995. Database for vertigo. *Otolaryngology Head and Neck Surgery*, 112: 383-390.
- Kentala, E., I. Pyykkö, Y. Auramo y M. Juhola. 1996. Otoneurological expert system. *Annals of Otolaryngology, Rhinology and Laryngology*, 105(8): 654-658.
- Kentala, E., Y. Auramo, M. Juhola y I. Pyykkö. 1998. Comparison between diagnoses of human experts and a neurologic expert system. *Annals of Otolaryngology, Rhinology and Laryngology*, 107(2): 135-140.

- Kinkelin, P., Ch. Michel y P. Ghittino. 1991. *Tratado de las enfermedades de los peces*. Edita: ACRIBIA, S.A. Zaragoza.
- Kitao, T., T. Aoki, M. Fukudome, K. Kawano, Y. Wada y Y. Mizumo. 1983. Serotyping of *Vibrio anguillarum* from diseased freshwater fish in Japan. *Journal of Fish Diseases*, 6: 175-181.
- Klir, G.J., Z. Wang y D. Harmanec. 1997. Constructing fuzzy measures in expert systems. *Fuzzy sets and systems*, 92: 251-264.
- Larsen, J.L., H.B. Rasmussen y I. Dalsgaard. 1988. Study of *Vibrio anguillarum* strains from different sources with emphasis on ecological and pathobiological properties. *Applied of Environmental Microbiology*, 54: 2264-2267.
- Le Hégarat-Masclé, S., I. Bloch y D. Vidal-Madjar. 1997. Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 35(4): 1018-1031.
- Lee, P.G., R.N. Lea, E. Dohmann, W. Prebilsky, P.E. Turk, H. Ying y J.L. Whitson. 2000. Denitrification in aquaculture systems: an example of a fuzzy logic control problem. *Aquacultural Engineering*, 23: 37-59.
- Lindsay, R.K., B.G. Buchanan, E.A. Feigenbaum y J. Lederberg. 1980. *Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL project*. Edita: McGraw-Hill. Nueva York.
- Lobo-da-Cunha, A. y C. Azevedo. 1993. Processing of food vacuoles in the parasitic ciliate *Ichthyophthirius multifiliis* after exit from the host. *Parasitology Research*, 79: 272-278.
- Lönström, L., T. Wiklund y G. Bylund. 1994. *Pseudomonas anguilliseptica* isolated from baltic herring *Clupea harengus membras* with eye lesions. *Diseases of Organisms*, 18: 143-147.
- Losee, R.M. 2001. Natural language processing in support of decision-making: phrases and part-of-speech tagging. *Information Processing and Management*, 37: 769-787.
- MacConnell, E., C.E. Smith, R.P. Hedrick y C.A. Speer. 1989. Cellular inflammatory response of rainbow trout to the protozoan parasite that causes proliferative kidney diseases. *Journal of Aquatic and Animal Health*, 1: 108-118.
- Malczewski, M.L. 1995. Use of expert system for data validation in UHP gas analysis. *Expert Systems*, October: 28-30.

- Mamdani, E. 1977. Application of fuzzy logic to approximate reasoning using linguistic synthesis. *IEEE Transactions on Computation*, 26(12): 1182-1191.
- Manning, C.D. y H. Schutze. 1999. *Foundations of statistical natural language processing*. Edita: MIT Press. Cambridge.
- Marking, L.L., J.J. Rach y T.M. Schreier. 1994. Evaluation of antifungal agents for fish culture. *The Progressive Fish-Culture*, 56: 225-231.
- Martín-Andrés, A. y J. D. Luna del Castillo. 1994. *Bioestadística para las ciencias de salud*. Edita: Norma. Madrid.
- Martín del Brío, B. y A. Sanz-Molina. 1997. *Redes neuronales y sistemas borrosos*. Edita: Ra-Ma. Madrid.
- McKeown, K. 1985. Discourse strategies for generating natural language text. *Artificial Intelligence*, 27.
- Møllergaard, S. y I. Dalsgaard. 1989. Handbook of eel diseases. *Dann Fisk og Havunders Rapport*, 293: 1-47.
- Michel, C., J.F. Bernardet y D. Dinand. 1992. Phenotypic and genotypic studies of *Pseudomonas anguilliseptica* strains isolated from farmed European eels (*Anguilla anguilla*) in France. *Fish Pathology*, 27: 229-232.
- Molino, G., F. Molino, D. Furia, F. Bar, S. Battista y N. Cappello. 1996. Computer-aided diagnosis in jaundice: comparison of knowledge-based and probabilistic approaches. *Methods of Information in Medicine*, 35: 41-51
- Molnár, K., C.S. Székely y F. Baska. 1991. Mass mortality of eel in Lake Balaton due to *Anguillicola crassus* infection. *Bulletin of the European Association of Fish Pathology*, 11: 211-212.
- Munro, A.L.S. y N. Fijan. 1981. Disease prevention and control. En *Proceedings of the World Symposium on Aquaculture 'Heated Effluents and Recirculation System'*. Berlín (Alemania).
- Murphy, R.R. 1998. Dempster-Shafer theory for sensor fusion in autonomous mobile robots. *IEEE Transactions on Robotics and Automation*, 14(2): 197-206.
- Myhr, E., J.L. Larsen, A. Lillehaug, R. Gudding, M. Heum y T. Hastein. 1991. Characterization of *Vibrio anguillarum* and closely related species isolated from farmed fish in Norway. *Applied of Environmental Microbiology*, 57: 2750-2757.

- Nash, C.E. y C.L. Paulsen. 1981. Water quality changes relevant to heated effluents and intensive aquaculture. En *Proceedings of the World Symposium on Aquaculture 'Heated Effluents and Recirculation System'*. Berlín (Alemania).
- Neumann, W. 1985. Schwimmblasenparasit *Anguillicola* bei Aalen. *Fischer und Teichwirt*, 11: 322.
- Nguyen, A.N.D., E.A. Hartwell y J.D. Milam. 1996. A rule-based expert system for laboratory diagnosis of hemoglobin disorders. *Archive of Pathology and Laboratories of Medicine*, 120: 817-827.
- Oakes, M.P. 1998. *Statistics for corpus linguistics*. Edita: Edinburgh University Press. Edinburgh.
- Ohayon, M.M., C. Guilleminault, T. Paiva, R.G. Priest, D.M. Rapoport, T. Sagales, S. Smirne y J. Zully. 1997. An international study on sleep disorders in the general population: methodological aspects of the use of the sleep-EVAL system. *Sleep*, 20(12): 1086-1092.
- Okamoto, N., T. Sano, R.P. Hedrick y J.L. Fryer. 1983. Antigenic relationships of selected strains of infectious pancreatic necrosis virus and european eel virus. *Journal of Fish Diseases*, 6: 19-25.
- Pazos, F., Y. Santos, B. Magariños, I. Bandín, S. Núñez y A.E. Toranzo. 1993. Phenotypic characteristics and virulence of *Vibrio anguillarum*-related organisms. *Applied and Environmental Microbiology*, sept.: 2969-2976.
- Pesonen, E., J. Ikonen, M. Juhola y M. Eskelinen. 1994. Parameters for a knowledge base for acute appendicitis. *Methods of Information in Medicine*, 33: 220-226.
- Peters, G. y F. Hartmann. 1986. *Anguillicola*, a parasitic nematode of the swim bladder spreading among eel populations in Europe. *Diseases of Aquatic Organisms*, 1: 229-230.
- Pickering, A.D. y L.G. Willoughby. 1977. Epidermal lesions and fungal infection on the perch, *Perca fluviatilis* L., in Windermere. *Journal of Fish Biology*, 11: 349-354.
- Rich, E. y K. Knight. 1994. *Inteligencia artificial*. Edita: McGraw-Hill. Madrid.
- Roberts, R.J. 1981. *Patología de los peces*. Edita: Mundi-Prensa. Madrid.
- Ruiz-Maya-Pérez, L., J. Martín-Pliego, J. López-Ortega, J.M. Montero-Lorenzo y P. Uriz-Tomé. 1990. *Metodología estadística para el análisis de datos cualitativos*.

Edita: Centro de Investigaciones Sociológicas (CIS) y Banco de Crédito Local. Madrid.

Sano, M., N. Okamoto, H. Fukuda, M. Sneyoshi y T. Sano. 1992. Virulence of Infectious Pancreatic Necrosis Virus Is Associated with the Larger RNA Segment (RNA Segment-A). *Journal of Fish Diseases*, 15(4): 283-293.

Sano, T., N. Okamoto y T. Nishimura. 1981. A new viral epizootic of *Anguilla japonica* Temminck et Schlegel. *Journal of Fish Diseases*, 4: 127-139.

Shafer, G. 1976. *A mathematical theory of evidence*. Edita: Princeton University Press. Princeton.

Shafer, G. 1982. Belief function and parametric models. *Journal of the Royal Statistical Society*, B44: 322-352.

Shepherd, J. y N. Bromage. 1999. *Piscicultura intensiva*. Edita: ACRIBIA, S.A. Zaragoza.

Shortliffe, E.H. y B.G. Buchanan. 1975. A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23: 351-379.

Smart, G.R. 1976. The effect of ammonia exposure on gill structure of the rainbow trout (*Salmo gairdneri*). *Journal of Fish Biology*, 8: 471-475.

Smith, C.E. y W.G. Williams. 1974. Experimental nitrite toxicity in rainbow trout and chinook salmon. *Transactions of the American Fisheries Society*, 103: 389-390.

StatSoft. 1995. *Statistica 5.0 User manual*. Edita: StatSoft. Tulsa.

Steffens, P. 1994. Machine translation and the lexicon. *Proceedings of the International EAMT Workshop*.

Syrjala, S.E. 1996. A statistical test for a difference between the spatial distributions of two populations. *Ecology*, 77(1): 75-80.

Székely, C.S. 1994. Paratenic hosts for the parasitic nematode *Anguillicola crassus* in Lake Balaton, Hungary. *Diseases of Aquatic Organisms*, 18: 11-20.

Székely, C.S., M. Láng y G.Y. Csaba. 1991. First occurrence of *Anguillicola crassus* in Hungary. *Bulletin of the European Association of Fish Pathology*, 11(4): 162-163.

Szolovits, P. 1995. Uncertainty and decisions in medical informatics. *Methods of Informations in Medicine*, 34: 111-121.

- Tarifa, E.E. y N.J. Scenna. 1997. Fault diagnosis, direct graphs, and fuzzy logic. *Computers and Chemical engineering*, 21: 649-654.
- Takagi, T. y M. Sugeno. 1985. Fuzzy identification of system and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetic*, 15: 116-132.
- Tsoukalas, L.H. y R.E. Uhrig. 1997. *Fuzzy and neural approaches in engineering*. Edita: Wiley-Interscience. New York.
- Turing, A.M. 1950. Computing machinery and intelligence. *Mind*, 59: 433-460.
- Underwood, B.O., C.J. Smale, F. Brown y B.J. Hill. 1977. Relationship of a virus from *Tellina tenuis* to infectious pancreatic necrosis virus. *Journal of General Virology*, 36: 93-109.
- Van Diest, P.J., J.A.M. Beliën, P.E. Zanstra, W.W. Wilhelm y J.P.A. Baak. 1994. Integrated decision support system/image archive for histological typing of breast cancer using a relation oriented inference system. *Histopatology*, 25: 253-259.
- Voorbraak, F. 1991. On the justification of Dempster's rule of combination. *Artificial Intelligence*, 48: 171-197.
- Wakabayashi, H. y S. Egusa. 1972. Characteristics of a *Pseudomonas sp.* from an epizootic of pond cultured eels (*Anguilla japonica*). *Bulletin of the Japanese Society Scientific Fish*, 38: 577-587.
- Wedemeyer, G.A. 1981. The physiological responses of fishes to the stress of intensive aquaculture in recirculation system. En *Proceedings of the World Symposium on Aquaculture 'Heated Effluents and Recirculation System'*. Berlín (Alemania).
- Weiru, L. y A. Bundy. 1994. A comprehensive comparison between generalized incidence calculus and the Dempster-Shafer theory of evidence. *International Journal of Human Computer Studies*, 40: 1009-1032.
- Weyn, B., G. Van de Wouwer, A. Van Daele, P. Sheunders, D. Van Dyck, E. Van Marck y W. Jacob. 1998. Automated breast tumor diagnosis and grading based on wavelet chromatin texture description. *Cytometry*, 33: 32-40.
- Weyn, B., G. Van de Wouwer, S. Kumar-Singh, A. Van Daele, P. Sheunders, E. Van Marck y W. Jacob. 1999. Computer-assisted differential diagnosis of malignant mesothelioma based on syntactic structure analysis. *Cytometry*, 35: 23-29.



- Whitsell, A. y P.G. Lee. 1994. A plug-and-play machine vision application for aquaculture. *Science, Computer and Automatic*, 10(8): 29-32.
- Wickins, J.F. 1981. Water quality requirements for intensive aquaculture: A review. En *Proceedings of the World Symposium on Aquaculture 'Heated Effluents and Recirculation System'*. Berlín (Alemania).
- Wiklund, T. y L. Lönnström. 1994. Occurrence of *Pseudomonas anguilliseptica* in finnish fish farm during 1986-1991. *Aquaculture*, 126: 211-217.
- Winiwarter, W. 2000. Adaptive natural language interfaces to FAQ knowledge bases. *Data and Knowledge Engineering*, 35: 181-199.
- Wong, S.K.M. y Y.Y. Yao. 1992. Characterization of comparative belief structures. *International Journal of Man-Machine Studies*, 37: 123-133.
- Woods, W.A. 1973. Progress in natural language understanding: An application to Lunar geology. En *Proceedings of the AFIPS conference 42*.
- Wyatt, J. y D. Spiegelhalter. 1990. Evaluating medical decision-aids: what to test, and how?. *Medical Informatics*, 15: 205-217.
- Yao, Y.Y. y K.M. Wong. 1994. Representation, propagation and combination of uncertain information. *International Journal of General Systems*, 23: 59-83.
- Zadeh, L.A. 1979. A theory of approximate reasoning. En *Machine Intelligence 9*, ed. J. Hayes. D. Michie, y L.I. Mikulich, 149-194. Nueva York.
- Zadeh, L.A. 1986. A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. *IA Magazine*, 7: 85-90.
- Zeldis, D. y S. Prescott. 2000. Fish disease diagnosis program –problems and some solutions. *Aquacultural Engineering*, 23: 3-11.
- Zimmerman, D.L. 1993. A bivariate Cramér-von Mises type of test for spatial randomness. *Applied Statistics*, 42: 43-54.

### 3.8. ANEXO III.1. GLOSARIO

$\Delta\chi^2$	Incremento porcentual del estadístico Chi-cuadrado modificado
$\neg$	Complementario de...
$\gamma_i(x_j, y_k)$	Creencia normalizada asociada a los conjuntos borrosos de entrada $x_j$ e $y_k$
$\Gamma_i$	Función de distribución acumulada para la matriz $\mathbf{i}$
$\Theta$	Marco de discernimiento
$\chi^2$	Valor del estadístico Chi-cuadrado
$\mu_A(t)$	Función de pertenencia borrosa. Determina el grado en que $t$ está incluida en el concepto representado por $\mathbf{A}$
$\Psi$	Estadístico de Cramér-von Mises
[CR]	Creencia de que la patología seleccionada sea correcta
[FA]	Frecuencia de Aparición
[MPC]	Mínimo Porcentaje Compartido
A	Creencia Alta
A-A	Reglas Arriba-Abajo
ADJ	Subbase ‘Adjetivo’ de la <b>CDP</b>
B	Creencia Baja
BC	Bastante Compartido
CCL	Subbase ‘Complemento Circunstancial de Lugar’
$ccl_{i,0}$	Etiqueta sintáctica correspondiente a la fila $\mathbf{i}$ de la matriz $\mathbf{CCL}_{sel}$
$ccl_{i,j}$	Cuando $\mathbf{j} > \mathbf{0}$ es el valor que asocia la etiqueta $ccl_{i,0}$ con la patología $\mathbf{j}$
$\mathbf{CCL}_{sel}$	Matriz que contiene parte de la información presente en la <b>CDS</b> utilizada por el subsistema de sugerencia

CCM	Subbase ‘Complemento Circunstancial de Modo’
$cd_{i,0}$	Etiqueta sintáctica correspondiente a la fila $i$ de la matriz $CD_{sel}$
$cd_{i,j}$	Cuando $j > 0$ es el valor que asocia la etiqueta $cd_{i,0}$ con la patología $j$
CDP	Base de Conocimiento del Dominio Principal
CDS	Base de Conocimiento del Dominio Secundario
$CD_{sel}$	Matriz que contiene parte de la información presente en la <b>CDS</b> utilizada por el subsistema de sugerencia
$C_i$	Número de respuestas dadas por el sistema que coinciden con el ‘Estándar Oro’ durante la validación
CM	Común
Con	Conflicto de evidencia
D	Función Trapezoidal
$d_i(x_j, y_k)$	Creencia asociada a los conjuntos borrosos de entrada $x_j$ e $y_k$
$ES_{exp}$	Error estándar experimental
F	Nivel de acierto mínimo aceptable
$F_0$	Nivel de acierto medio
FCR	Factor de conversión del alimento
$F_{\alpha/2}$	Valor del estadístico <b>F</b> de Snedecor
G	Función Gaussiana
gs	Grupo sintomático
$H_n^L$	Conjunto borroso de entrada asociado a la Memoria Asociativa Borrosa <b>L</b>
l, f	Variabes lingüísticas
LC	Límite de confianza
$LC_s$	Límite superior de confianza al 95 %

LC <sub>i</sub>	Límite inferior de confianza al 95 %
L <sub>k</sub>	Número de patologías que compone el grupo sintomático <b>k</b>
MA	Creencia Muy Alta
MAB	Memoria Asociativa Borrosa
MB	Creencia Muy Baja
MC	Muy Compartido
m <sub>j</sub> <sup>*</sup> (g <sub>s<sub>k</sub></sub> )	Cantidad de creencia escalada en función de $\chi^2$ asignada a cada uno de los grupos sintomáticos <b>k</b> obtenidos tras la evidencia <b>j</b>
MM	Muy Común
m <sub>n</sub> (g <sub>s<sub>i</sub></sub> )	Cantidad de creencia asociada a cada uno de los grupos sintomáticos <b>i</b> obtenidos tras la evidencia <b>n</b>
m <sub>n</sub>	Creencia global asociada a un grupo sintomático
MP	Muy Poco Compartido
MR	Muy Rara
N	Creencia Normal
NC	Normalmente Compartido
N <sub>c</sub>	Casos aportados al sistema durante el proceso de validación
N <sub>t</sub>	Número total de patologías reconocidas por el sistema
OD	Subbase ‘Objeto Directo’ de la <b>CDP</b>
O <sub>g</sub> (i)	Creencia original en tanto por ciento asignada a través de la regla de Dempster a los <b>g</b> grupos sintomáticos creados tras la evidencia <b>i</b>
p <sub>α</sub>	Nivel de significación estadístico
PC	Poco Compartido
PCOM <sub>p,i</sub>	‘Porcentaje Compartido’ para las <b>p</b> etiquetas e <b>i</b> subbases de <b>CDP</b>

Pt	Patología
PT	Puntuación total de la valoración de un diagnóstico
PT <sub>Máxima</sub>	Puntuación máxima que puede alcanzar una determinada patología en una valoración
PT <sub>Satisfecha</sub>	Puntuación alcanzada en una valoración dadas un número de condiciones a cumplir
R <sup>(L)</sup>	Conjunto de reglas que componen una Memoria Asociativa Borrosa
R <sub>c</sub>	Número de reglas de la ‘base de reglas’ satisfechas durante el proceso de búsqueda
RR	Rara
RTA	Red de Transición Aumentada
S	Función Singleton
SUJ	Subbase ‘Sujeto’ de la <b>CDP</b>
T	Función Triangular
VER	Subbase ‘Verbo’ de la <b>CDP</b>
X, Y, Z	Conjuntos de grupos sintomáticos
Y	Valor no borroso de salida
y <sup>S(L)</sup>	Centro del conjunto borroso
Z(LC)	Un número tal que el área comprendida bajo la distribución normal entre <b>z(LC)</b> y <b>-z(LC)</b> es <b>LC</b>

### **3.9. ANEXO III.2. MANUAL DEL USUARIO**

La aplicación de ordenador desarrollada tiene como fin servir de ayuda a los expertos humanos en el proceso de diagnóstico de las patologías que hasta el momento pueden ser causa de importantes pérdidas económicas en las explotaciones intensivas de anguilas.

La información de la que va a hacer uso el programa proviene de dos fuentes claramente diferenciadas: la información codificada en sus bases de conocimiento procedente de publicaciones especializadas y la que introduce el usuario consecuencia de las observaciones que éste hace sobre las muestras. Tal y como corresponde a la entidad de este tipo de programas, es posible modificar fácilmente la información contenida en sus bases de datos, lo que permite la ampliación o reducción de éstas cuando se obtiene nueva información sobre patologías que ya están codificadas o incluir otras de nueva aparición. La modificación de las bases de datos no solo implica información explícita sobre la propia enfermedad sino también las condiciones físico-químicas que favorecen su aparición. Estas condiciones de aparición quedan codificadas como reglas (**A-A**) y permiten además la interacción con el módulo de simulación del programa de gestión de la planta (**CONTROLPLANT 1.0**).

Para poder llevar a cabo el proceso de diagnóstico y determinación del tratamiento a seguir, el programa consta de varios módulos desempeñando cada uno de los cuales una misión concreta. Sin duda, el más importante de ellos es el motor de inferencia, a través del cual el programa puede obtener conclusiones y respuestas sobre lo que está sucediendo en la planta a nivel patológico. Para ello la respuesta dada por el sistema se caracteriza por no ser única, es decir, en realidad ésta se compone de varias posibles opciones a las que el programa asigna distintos niveles de creencia. Por otra parte, el resto de los módulos facilitan al usuario información sobre lo que está haciendo el motor de inferencia al obtener sus conclusiones, guían a éste cuando aparentemente no es posible seguir o cuando las conclusiones obtenidas son poco concluyentes y aconsejan sobre los pasos a seguir para poder erradicar la enfermedad de la planta determinando los tratamientos, fármacos y dosis adecuadas.

#### **III.2.1. INICIO DEL PROGRAMA**

Para poder acceder al programa, el usuario tiene dos opciones: a) desde el programa de gestión **CONTROLPLANT 1.0** (figura AIII.1a) como una herramienta más o, b) desde un acceso directo del escritorio de Windows<sup>®</sup> (figura AIII.1b). Si se elige la primera opción tendremos que hacer 'click' en **SPI** (Sistema de Predicción Integrado) de la barra de herramientas de **CONTROLPLANT** y elegir **SEDPA**. Si optamos por acceder al programa

desde el escritorio de Windows® bastará con hacer doble 'click' sobre el icono de dicha aplicación. En ambos casos el resultado será la visualización de una pantalla de presentación (figura AIII.2) en la que el usuario debe introducir una clave de acceso. La existencia de una clave de acceso distinta para cada usuario permite tener un registro continuo de todo el personal de planta que tiene acceso al programa (día y hora de conexión-desconexión, etc.).



**Figura AIII.1a.** Acceso a SEDPA 1.0 desde la barra de herramientas de CONTROLPLANT 1.0

**Figura AIII.1b.** Inicio de SEDPA 1.0 desde el acceso directo en el escritorio de Windows®



**Figura AIII.2.** Pantalla de presentación de la aplicación 'Sistema Experto para el Diagnóstico Patológico de Anguilas' (SEDPA 1.0)

Si la clave de acceso es correcta se entra en la ventana principal del programa (figura AIII.3).

### III.2.2. VENTANA PRINCIPAL

La ventana principal de **SEDPA 1.0** se encuentra físicamente dividida en cuatro partes claramente diferenciadas: a) una barra de herramientas donde se pueden distinguir un total de cuatro opciones: *Archivo*, *Ver*, *Opciones* y *Ayuda*, b) un bloque con el título de propiedades donde se muestran los parámetros con los que trabaja el sistema junto con los botones *Aceptar síntoma*, *Sugerencia*, *Explicación* y *Siguiente paso* (bloque superior), c) un espacio central reservado para una caja de texto (donde el usuario puede escribir los síntomas/lesiones observadas) y para las mallas que muestran los detalles del proceso de razonamiento seguido por el motor de inferencia (bloque central) y, d) la parte inferior de la pantalla donde se muestran tres ventanas de hipertexto donde se van detallando los resultados finales obtenidos por el sistema (bloque inferior). A continuación se describe cada una de esas partes con mayor detalle.

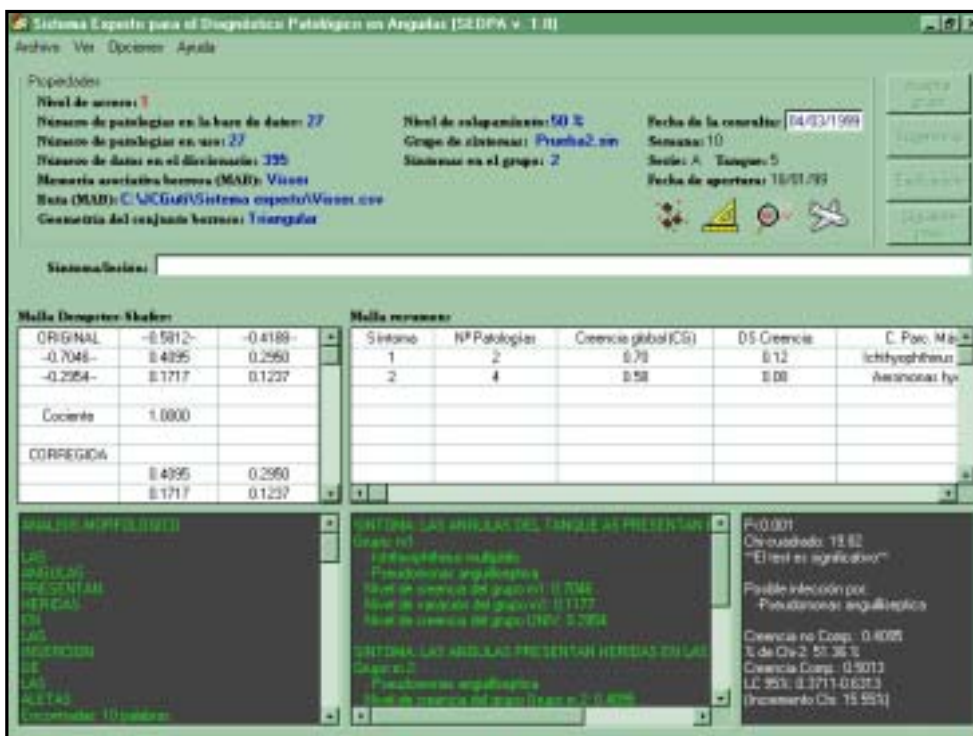
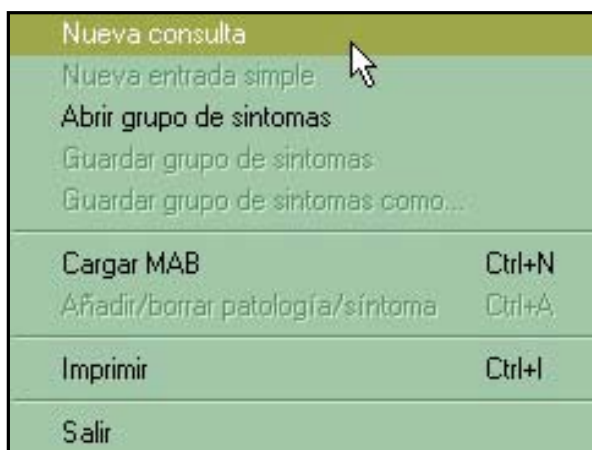


Figura AIII.3. Ventana principal de SEDPA 1.0



### III.2.2.1. Barra de herramientas

Como ya se ha mencionado en la sección anterior, la primera opción que encontramos en la barra de herramientas es *Archivo* (figura AIII.4). Cuando se hace ‘click’ en esta opción se despliega un menú donde se puede optar entre: *Nueva consulta*, *Nueva entrada simple*, *Abrir grupo de síntomas*, *Guardar grupo de síntomas*, *Guardar grupo de síntomas como...*, *Cargar MAB...Ctrl+N*, *Añadir/borrar patología/síntoma...Ctrl+A*, *Imprimir...Ctrl+I* y *Salir*.



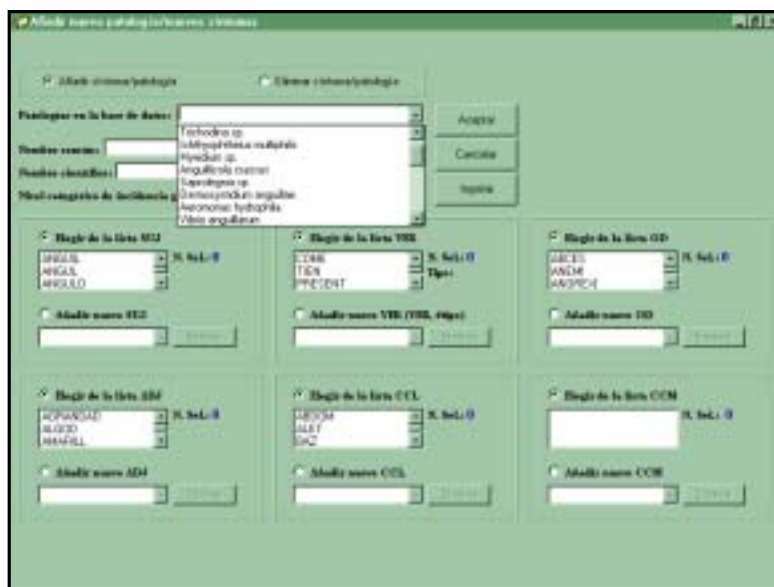
**Figura AIII.4.** Menú desplegable que aparece al hacer ‘click’ en la opción *Archivo* de la barra de herramientas de la ventana principal

La primera de las opciones (*Nueva consulta*) está por defecto siempre disponible. Haciendo ‘click’ en *Nueva consulta* se restablecen los valores iniciales de las variables y matrices utilizadas durante el proceso de diagnóstico, de tal forma que no sea necesario iniciar de nuevo el programa para hacer una nueva consulta. *Nueva entrada simple* inicialmente se encuentra inhabilitado. Esto es debido a que esta opción tan solo puede ser utilizada cuando se ha iniciado la sesión a partir de un grupo de síntomas a los que además se quiere añadir alguno nuevo detectado. *Abrir grupo de síntomas* permite recuperar un grupo de síntomas/lesiones probados en otra ocasión. Por ello, cuando se elige esta opción aparece un cuadro de diálogo donde el usuario debe seleccionar el archivo de grupo de síntomas (archivos con la extensión ‘.sin’). Cuando el archivo de grupo de síntomas es cargado se producen tres efectos significativos: a) la caja de texto de introducción de síntomas se inhabilita lo que impide momentáneamente la introducción de cualquier síntoma, b) el botón que inicia el proceso de diagnóstico cambia su aspecto mostrando el mensaje *Aceptar Grupo* en vez de *Aceptar síntoma* y, c) el icono representativo del tipo de síntoma utilizado (simple o grupo) cambia de una cruz simple a un grupo de cruces rojas. Las dos siguientes opciones (*Guardar grupo de síntomas* y *Guardar grupo de síntomas*

como...) son las típicas de cualquier aplicación para Windows®. Te permiten guardar los síntomas introducidos hasta el momento por primera vez o añadir los últimos a un archivo ya creado.

Como ya se comentó en la sección 3.4.3.2.2 la Memoria Asociativa Borrosa (**MAB**) con la que funciona el sistema puede cargarse independientemente no siendo, por tanto, una parte del código principal del programa. Esto se lleva a cabo desde la opción *Cargar MAB...Ctrl+A* del menú de *Archivo*. Al hacer 'click' en esta opción aparece un cuadro de diálogo donde el usuario debe seleccionar el archivo de memoria asociativa borrosa (archivos con extensión '.mab'). Por defecto, si el usuario decide no seleccionar ninguna, el programa carga la **MAB** denominada Visser. Hay que tener en cuenta que una vez que se comienza a trabajar con una **MAB** en concreto, la opción se inhabilita impidiendo la selección de otra distinta. Debido a que la selección de una u otra **MAB** tiene efectos importantes sobre el proceso de diagnóstico de forma temporal, esta opción está restringida a usuarios que tenga un nivel de acceso 2.

*Añadir/borrar patología/síntoma* es la opción más restrictiva del programa y, por tanto, por defecto se encuentra inhabilitada. A ella tan solo se puede entrar teniendo un nivel de acceso 3, por lo que está limitada a aquellos usuarios que conozcan la estructura de la Base de Conocimiento del Dominio Principal (**CDP**). De esta forma se impide que usuarios no experimentados en el manejo de la base de datos puedan llevar a cabo cambios en la **CDP** que altere gravemente el funcionamiento y las conclusiones del sistema. Si se está en posesión del nivel de acceso 3 y hacemos 'click' en esta opción nos aparecerá una nueva ventana (figura AIII.5) donde podemos elegir entre *Añadir síntoma/patología* o *Eliminar síntoma/patología*. Tanto en un caso como en otro tendremos que seleccionar en primer lugar el organismo/factor responsable de la aparición de la patología de la lista *Patologías en la base de datos*. Una vez seleccionada aparecerán en las cajas de texto de la parte inferior el nombre común, nombre científico y el nivel categórico de incidencia global (se corresponde con alguno de los conjuntos borrosos [FA]). En el caso de que lo que se pretenda sea introducir una nueva patología en la base de datos, se procederá directamente a escribir el nombre común, científico y el nivel categórico de incidencia global, las etiquetas sintácticas que se corresponden con la nueva patología que pueden ser elegidas de las listas de cada una de las categorías (**SUJ, VER, OD, ADJ, CCL** y **CCM**) o escritas en la caja texto *Añadir nuevo...* teniendo en cuenta que si no se encuentran recogidas en la **CDP**, el programa preguntará además si cada una de esas nuevas etiquetas sintácticas puede ser asociada a alguna de las patologías ya existentes.



**Figura AIII.5.** Ventana Añadir/Eliminar patologías o síntomas/lesiones. Para poder visualizar esta ventana se necesita un nivel de acceso 3

Las dos opciones siguientes son las típicas de cualquier aplicación para Windows®: *Imprimir* y *Salir*. Con la primera se obtiene un copia en papel de todas las acciones llevadas a cabo durante el proceso de diagnóstico así como el usuario, fecha, hora de conexión y desconexión, etc. Con la segunda se sale de la aplicación.

La segunda opción de la barra de herramientas de la ventana principal es *Ver*. Al desplegar este menú encontramos cuatro nuevas opciones: *Lista de patologías*, *Detalles de los resultados del último síntoma*, *Valoración del diagnóstico* y *Sugerir tratamiento* (figura AIII.6). Con la primera de ellas tan solo se muestra una lista de todos los organismos y factores causantes de enfermedad en la anguila. Si elegimos la segunda (figura AIII.7) se visualiza una nueva ventana dividida en tres partes: a) un bloque llamado *Propiedades del grupo/individuales* donde se muestran las características del grupo sintomático recién formado (número de patologías seleccionadas y nivel de creencia global) y de forma individual las características de cada patología (nombre común, nombre científico y nivel de creencia individual) gracias a la presencia de una barra de desplazamiento, b) una ventana hipertexto donde se proporciona al usuario información completa sobre la enfermedad seleccionada y, c) una ventana de imágenes donde al mismo tiempo que el texto se presentan fotografías (del organismo responsable de la enfermedad, sus efectos sobre las anguilas, etc.) que facilita la identificación del organismo/factor responsable.

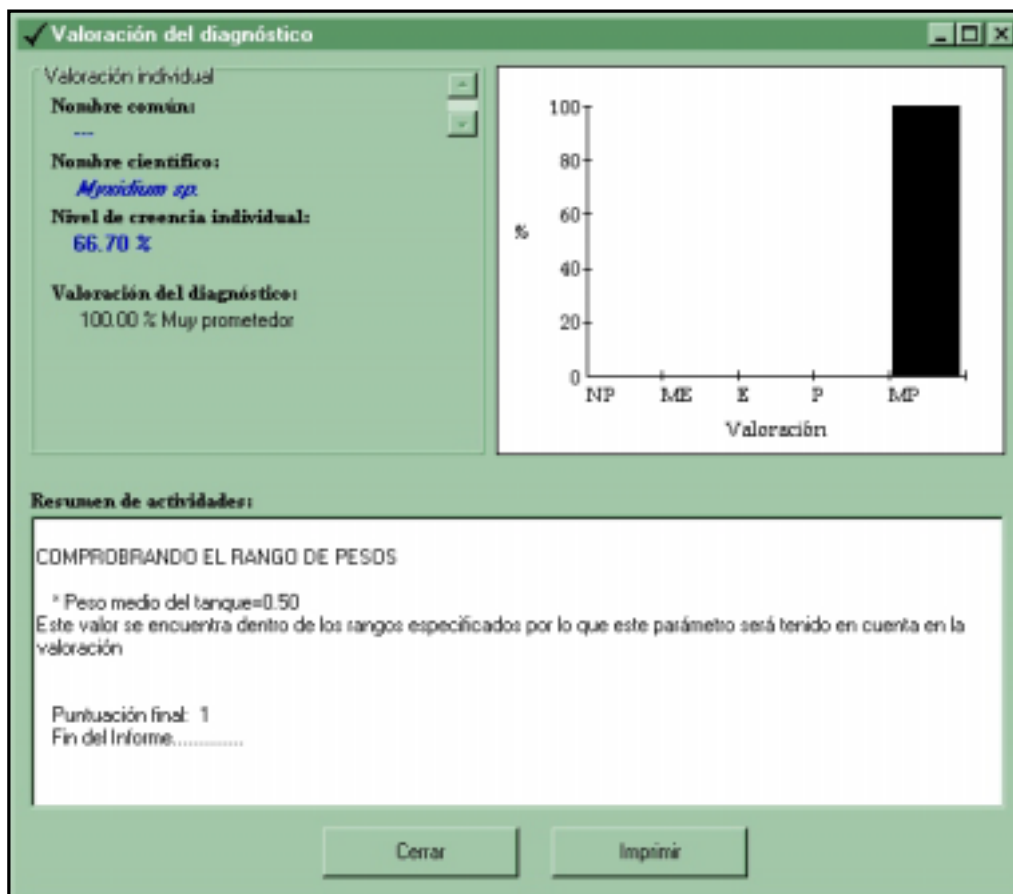


Figura AIII.6. Menú desplegable de la opción Ver de la barra de herramientas de la ventana principal



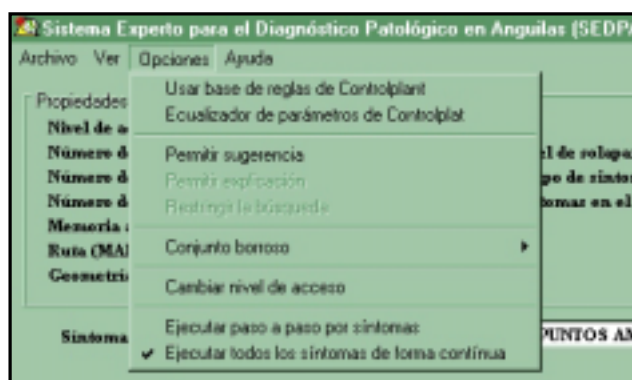
Figura AIII.7. Ventana Detalles donde se aporta al usuario información en forma de texto y fotografías que puede ayudar a diagnosticar de forma más precisa el agente responsable de la enfermedad

Asimismo, desde esta ventana podemos acceder a *Sugerir tratamiento*. Si elegimos la tercera opción del menú (*Valoración del diagnóstico*), **SEDPA** evalúa la posibilidad de que el agente patológico seleccionado sea el responsable de la enfermedad en función de las características físico-químicas y específicas de la planta, es decir, se computan los datos recogidos en la Base de Conocimiento del Dominio Secundario (**CDS**) a través de las reglas (**A-A**) de **CONTROLPLANT**. De esta forma aparece una ventana en la que se distinguen tres zonas: a) un bloque en el que se muestran las características del agente patológico evaluado, b) una gráfica que muestra los resultados de la evaluación y c) una ventana hipertexto donde se muestran los detalles del proceso seguido por el sistema para aceptar o rechazar condiciones de aparición. El último apartado, muestra en una ventana hipertexto el tratamiento a seguir en función del tanque (sus características biológicas) y el agente seleccionado por el sistema (figura AIII.8).



**Figura AIII.8.** Ventana Valoración del diagnóstico

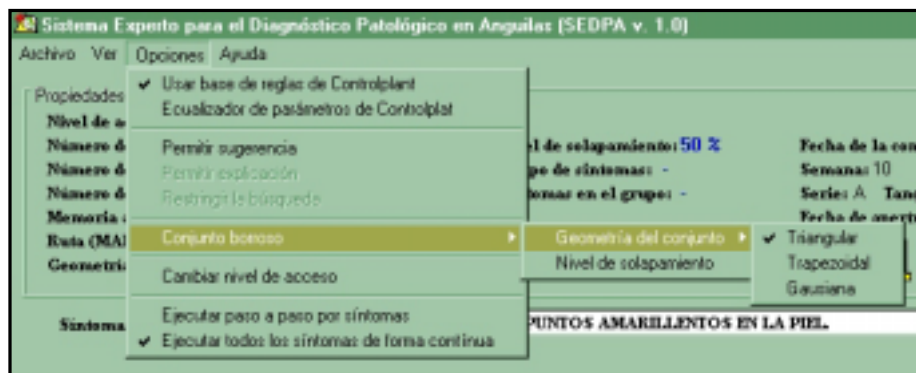
La tercera opción de la barra de herramientas de la ventana principal (*Opciones*) despliega un menú donde se puede elegir entre: *Usar base de reglas de CONTROLPLANT (A-A)*, *Ecualizador de parámetros de CONTROLPLANT*, *Permitir sugerencia*, *Permitir explicación*, *Restringir la búsqueda*, *Conjunto borroso*, *Cambiar nivel de acceso*, *Ejecutar paso a paso por síntomas* y *Ejecutar todos los síntomas de forma continua* (figura AIII.9). Si seleccionamos *Usar base de reglas de CONTROLPLANT (A-A)* estamos indicando al sistema que las conclusiones del motor de inferencia deben estar condicionadas por la información presente en la Base de Datos del Dominio Secundario (**CDS**), es decir, en el proceso de diagnóstico se tendrá en cuenta las características particulares de cada patología además de las condiciones físico-químicas que permiten su aparición (esto puede establecerse a través de la opción *Ecualizador de parámetros de CONTROLPLANT*). Si no se selecciona esta opción el diagnóstico sólo se lleva a cabo en función de los síntomas/lesiones observados. Además, su activación provoca la aparición del icono correspondiente en el bloque propiedades. Las dos siguientes (*Permitir sugerencia* y *Permitir explicación*) tan solo habilitan o inhabilitan los botones *Sugerencia* y *Explicación* (ver descripción del botón *Sugerencia* y *Explicación*) a través de los cuales se accede a esas opciones.



**Figura AIII.9.** Menú desplegable que aparece al hacer 'click' en *Opciones* de la barra de herramientas de la ventana principal

Mediante *Restringir la búsqueda* podemos limitar el rango de búsqueda durante el diagnóstico. Esta opción se incluye ya que en ocasiones puede suceder que la experiencia del experto que maneje el sistema le permite descartar directamente varias de los organismos/factores que se encuentran incluidos en las bases de datos. Al hacer 'click' en esta opción aparece una ventana donde tan solo se incluye una caja de texto desplegable donde el experto puede seleccionar los organismos/factores más probables según su experiencia.

A través de la opción *Conjunto borroso* se despliega un segundo menú (figura AIII.10) donde podemos seleccionar: *Geometría del conjunto* y *Nivel de solapamiento*.



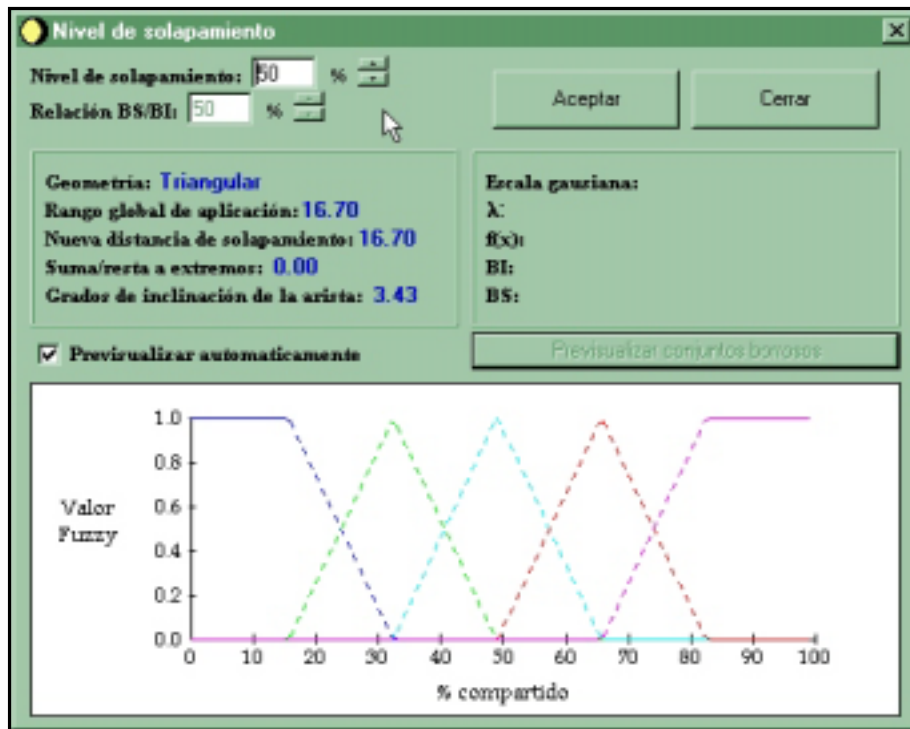
**Figura AIII.10.** Dentro de la opción *Conjunto borroso* podemos seleccionar el nivel de solapamiento de los conjuntos borrosos así como establecer su geometría (*Triangular, Trapezoidal y Gaussiana*)

Si seleccionamos la primera se despliega un tercer menú con tres opciones: *Triangular, Trapezoidal* y *Gaussiana* donde se encuentra marcada por defecto *Triangular*. La elección de una u otra se refleja mediante el icono correspondiente en el bloque propiedades. Sin embargo si elegimos *Nivel de solapamiento* se visualiza una nueva ventana (figura AIII.11) donde podemos distinguir los siguientes bloques: a) una parte superior donde se pueden encontrar dos cajas de texto asociadas a dos barras de desplazamiento a través de las cuales cambiamos los niveles de solapamiento de los conjuntos. En el caso de que se elija la geometría trapezoidal también podemos seleccionar la relación entre la base superior e inferior, b) una parte central donde se recogen las características que presentan los conjuntos borrosos en el momento de ejecución del sistema y, c) una parte inferior donde se presenta una gráfica que muestra la geometría final de los conjuntos borrosos. Como la configuración de los parámetros que aparecen en esta opción tienen un importante efecto temporal sobre las conclusiones obtenidas por el motor de inferencia, para poder visualizarla se necesita un nivel de acceso 2.

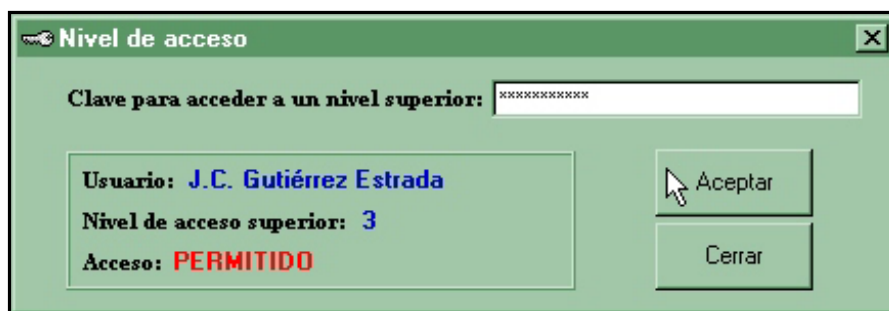
Al hacer ‘click’ en la opción *Cambiar nivel de acceso* nos aparece una ventana (figura AIII.12) en la que destaca una caja de texto donde debemos introducir la clave de acceso para poder pasar a un nivel superior. Si la clave es correcta, en la misma ventana se muestra el nombre del usuario, el nivel superior al que puede acceder y si el acceso está permitido. El nivel de acceso varía entre 1 (con acceso a la estructura básica del sistema) y el nivel 3 (con acceso a todos los subsistemas del programa).

Las dos últimas opciones de este menú actúan básicamente sobre el funcionamiento del sistema cuando se ha optado por escoger un grupo de síntomas. Si se elige la primera (*Paso a paso por síntomas*) se producen tantas interrupciones como síntomas incluye el grupo. De esta forma podemos observar las conclusiones intermedias que el sistema está obteniendo. Si por el contrario escogemos la segunda (*Ejecutar todos los síntomas de*

*forma) continua* el sistema muestra los resultados finales, pero no es posible comprobar que es lo que ha sucedido en los pasos intermedios. Al igual que sucede con otras opciones importantes del sistema descritas anteriormente, la elección de una u otra supone la visualización del icono correspondiente en el bloque de propiedades.



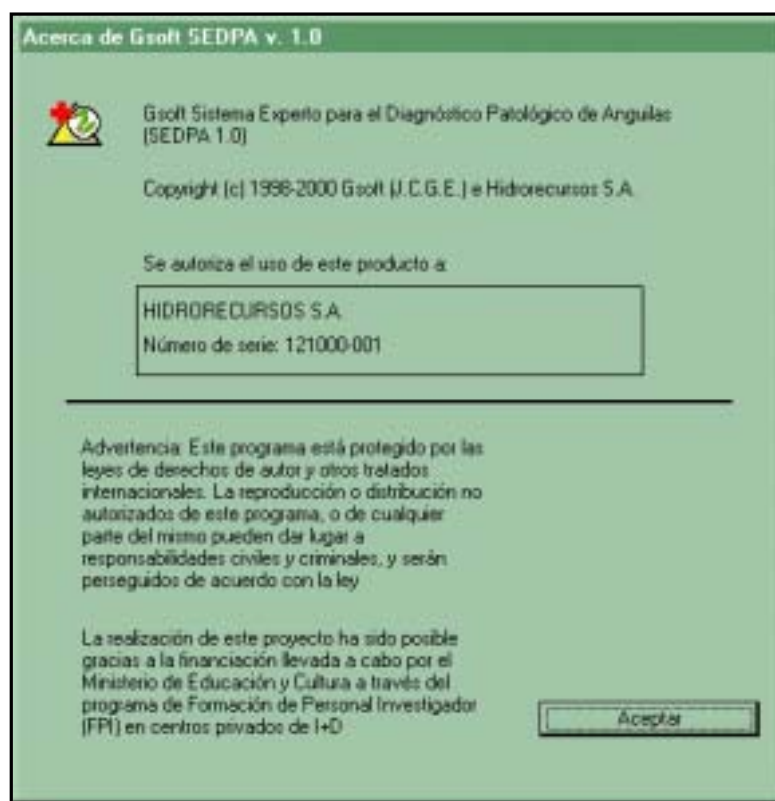
**Figura AIII.11.** A través de la ventana *Nivel de solapamiento* podemos establecer la superposición de los conjuntos borrosos. El gráfico de la parte inferior nos da una idea de la configuración de los conjuntos. En este caso se trabaja con unos conjuntos de geometría triangular y un nivel de solapamiento del 50 %



**Figura AIII.12.** Ventana a través de la cual es posible cambiar el nivel de acceso a los distintos subsistemas del programa



Finalmente, en la barra de herramientas encontramos la opción *Ayuda*. Haciendo ‘click’ se despliega un menú que contiene las opciones *Contenido* y *Acerca de...* La opción *Contenido* de este menú permite que el usuario obtenga información respecto a las bases teóricas de la metodología desarrollada en la aplicación y respecto a su manejo. La búsqueda de la información puede obtenerse de una ficha que describe el contenido general o de una ficha índice de palabras clave. La opción *Acerca de...* muestra una ventana con información general del programa: nombre, autor, organismo donde se ha desarrollado la aplicación, versión y año (figura AIII.13).

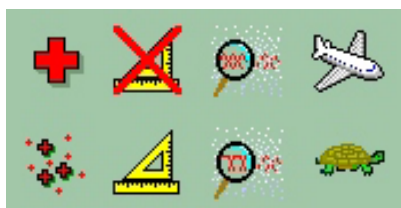


**Figura AIII.13.** Ventana *Acerca de...* que muestra información general de *SEDPA 1.0*

### III.2.2.2. Bloque superior

En el bloque superior hay que distinguir entre el cuadro titulado *Propiedades* y los botones *Aceptar*, *Sugerencia*, *Explicación* y *Siguiente paso*. En el cuadro *Propiedades* se muestran las características y propiedades de las que dispone el sistema y con las que va a trabajar para obtener sus conclusiones. En él se puede ver el nivel de acceso permitido al usuario en ese momento, el número de patologías codificadas en la **CDP**, el rango de búsqueda

(número de patologías en uso) cuyo valor variará si la búsqueda está restringida o no, el número de términos en el diccionario de sinónimos, la **MAB** con la que va a trabajar el sistema así como el lugar donde se encuentra guardado el archivo, la geometría de los conjuntos borrosos, el nivel de solapamiento, el nombre del grupo de síntomas si se ha escogido esta opción así como el número de síntomas en el grupo, características del tanque sobre el que estamos haciendo la consulta, y un total de cuatro iconos que nos recuerdan gráficamente las características más importantes del sistema (figura AIII.14).



**Figura AIII.14.** *Dos posibles configuraciones de los iconos que representan las características más importantes de funcionamiento del sistema*

El botón *Aceptar* desencadena el proceso de diagnóstico una vez que el usuario ha introducido el síntoma/lesión. Este proceso (como ya se comentó en el punto 3.4) consiste en primer lugar en un análisis morfológico-sintáctico con la formación de una **RTA** y en la transmisión de la incertidumbre procedente de varias fuentes de evidencia mediante la aplicación de la teoría de Dempster-Shafer. Los resultados de todos estos pasos se muestran en el bloque central e inferior.

Con el botón *Sugerencia* se visualiza una nueva ventana (figura AIII.15) a través de la cual el usuario puede obtener una guía que opcionalmente puede seguir cuando aparentemente el proceso de diagnóstico se detiene. Para ello, el sistema pregunta al usuario: ¿se aprecia algún síntoma/lesión localizado en...? cuando en el bloque *Matriz de resultados preliminares* se encuentra seleccionado *Muestra matriz de resultados preliminares CCL* o ¿Se aprecia/n...? cuando se encuentra seleccionado *Muestra matriz de resultados preliminares OD*. El mismo bloque tiene además una malla donde se muestran cada uno de los datos presentes tanto en la matriz **OD** como en la **CCL** y la relación existente con respecto a cada una de las patologías seleccionadas tras el último evento. Dicha relación indica con un SI o un NO, significando el SI que determinada etiqueta sintáctica solo se encuentra asociada a una de las patologías seleccionada y el NO que se encuentra asociada a más de una de las patologías seleccionadas o bien no se encuentra asociada a ninguna. En la misma malla se presentan además los porcentajes de comunidad de dichas etiquetas sintácticas entre todas las patologías reconocidas por el sistema, tan solo

en aquellos casos en los que se presenta un SI. El cálculo de este porcentaje es importante ya que va a determinar el factor de impacto de la respuesta dada por el usuario.

Todos los datos aportados en la malla se acompañan dentro del mismo bloque por información suplementaria: número de patologías seleccionadas por la sugerencia dentro del grupo, la patología seleccionada en un momento dado que depende tanto de la matriz de resultados preliminares como de la etiqueta sintáctica que escojamos, el número que la patología seleccionada ocupa dentro de la **CDP**, el número de síntomas afectados, es decir, el número de etiquetas sintácticas que en la malla aparecen asociados a un SI, y finalmente el factor de impacto de la respuesta recomendado por el sistema en función del porcentaje de comunidad. Este bloque se completa con un botón que transmite el factor de impacto de la respuesta a las cajas de texto del bloque siguiente y que son los que realmente se van a tener en cuenta durante el cálculo. En este otro bloque es donde se muestra al usuario el tipo de pregunta ya relacionada con la etiqueta sintáctica. Además el usuario puede variar la etiqueta sobre la que se pregunta mediante una barra de desplazamiento con tantos pasos como los indicados en *Numero de síntomas afectados*. Una vez que se ha seleccionado la etiqueta sintáctica se puede responder *Si*, *No* o *No es posible determinarlo*, apareciendo en una ventana de hipertexto las consecuencias de la respuesta sobre los niveles de creencia individuales así como sobre los globales cuando la respuesta es *Si* o *No*, y pasando a la siguiente etiqueta más restrictiva cuando se contesta *No es posible determinarlo*. Finalmente los resultados preliminares que se muestran en la ventana hipertexto se almacenan y se tienen en cuenta para el siguiente evento tras hacer 'click' en el botón *Aceptar*.

El botón *Explicación* muestra una nueva ventana donde se distingue una caja de hipertexto. En esta ventana el sistema registra que reglas de **CONTROLPLANT (A-A)** se cumplen y cuales no.

### **III.2.2.3. Bloque central**

En el bloque central (figura AIII.16) de la ventana principal se pueden distinguir tres estructuras principales: la caja de texto mediante la cual el usuario introduce los síntomas/lesiones que observa y dos mallas de datos: la situada más a la izquierda denominada *Dempster-Shafer* y la que se encuentra a la derecha (*Resumen*).

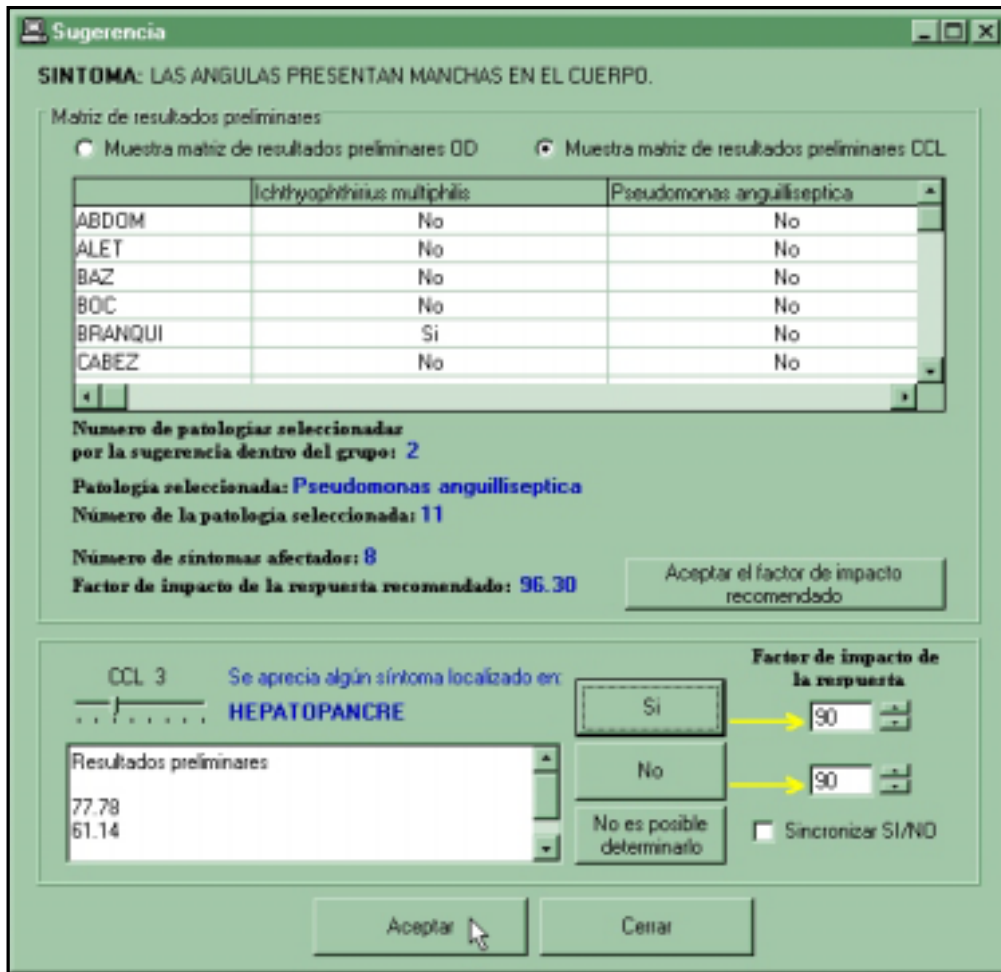


Figura AIII.15. Ventana que se visualiza cuando se hace 'click' en el botón Sugerencia

Sistema/lesión:

Malla Demerita-Shadiz:			Malla resumen:				
ORIGINAL	-0.6736-	-0.2254-	Sintoma	Nº Patologías	Gravidad global (GG)	D5 Deencia	C. Pac. Máx.
-0.3762-	0.2534	0.1228	1	4	0.57	0.00	Aeromonas hy.
-0.2790-	0.1679	0.0811	2	4	0.65	0.00	Vibrio angui
-0.1296-	0.0873	0.0423	3	3	0.66	0.00	Vibrio angui
-0.0861-	0.0647	0.0314	4	3	0.67	0.00	Vibrio angui
-0.0684-	0.0461	0.0223					
-0.0567-	0.0342	0.0166					
Cociente	1.0000						

Figura AIII.16. Bloque central de la ventana principal del programa SEDPA 1.0

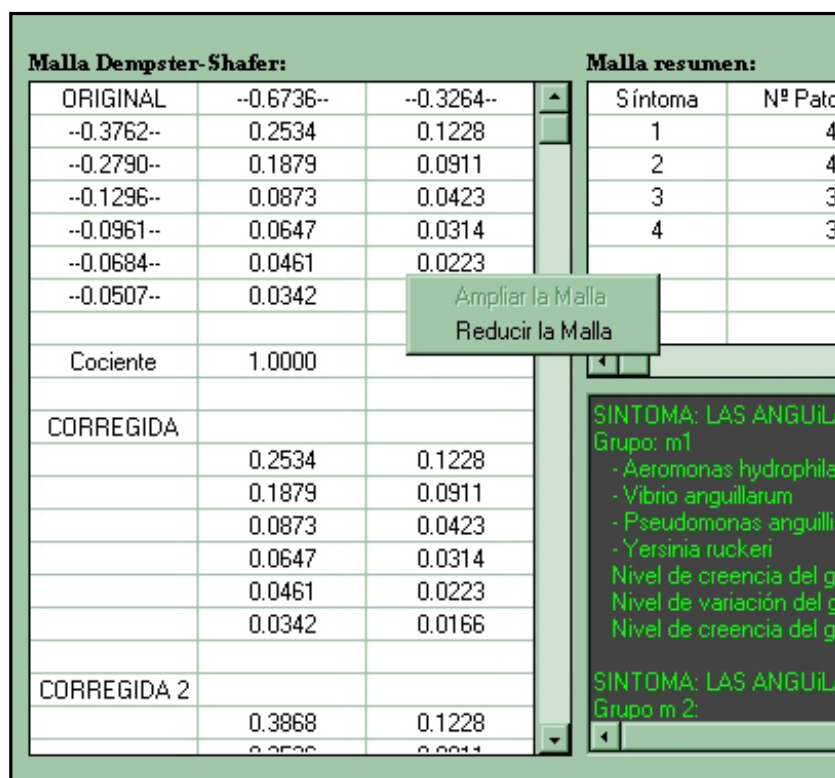
### III.2.2.3.1. Malla de Dempster-Shafer

La malla de Dempster-Shafer es la que muestra los cálculos que lleva a cabo el sistema cuando se presentan más de un foco de evidencia (dos o más síntomas/lesiones), es decir, cuando se aplica la teoría de transmisión de la evidencia de Dempster-Shafer. Los cálculos muestran la matriz inicial de combinación de evidencias identificada como ORIGINAL, un parámetro denominado *Cociente* que está indicando la cantidad de creencia que se asocia a conjuntos vacíos después de la aplicación de la regla de Dempster, una segunda matriz denominada CORREGIDA que es la matriz ORIGINAL transformada en función del cociente, y una tercera matriz identificada como CORREGIDA 2 en la que los grupos sintomáticos iguales se agrupan y se suman sus niveles de creencia.

Debido a la falta de espacio que existe en la ventana principal, tan solo es posible mostrar diez filas de la malla Dempster-Shafer aunque en todo momento, la presencia de las tres matrices supera la extensión máxima de la malla. Por ello, para poder ver todos los datos, la malla tiene una barra de desplazamiento que permite visualizar su parte inferior. También existe la posibilidad de ampliar la malla haciendo 'click' en el botón derecho del ratón sobre la malla. Cuando se hace esto, aparece un menú flotante donde se puede seleccionar *Ampliar la malla* con lo que se consigue visualizar el doble de filas. Para reducir la malla hasta su tamaño original, tendremos que volver a hacer 'click' sobre la malla con el botón derecho del ratón debiendo seleccionar en esta ocasión *Reducir la malla* (figura AIII.17).

### III.2.2.3.2. Malla Resumen

La malla *Resumen* muestra los resultados individualizados que el sistema obtiene tras cada síntoma/lesión introducido por el usuario. De esta forma, para cada síntoma/lesión (1, 2, ...,n) se especifica el número total de patologías asociadas (Nº Patologías), el nivel de creencia global del grupo sintomático (Creencia global –CG-), la desviación estándar del nivel de creencia del grupo sintomático (DS Creencia), el agente de los que se encuentran dentro del grupo sintomático que presenta una creencia parcial máxima (Creencia Parcial Máxima –CPM<sub>x</sub>-) así como el que presenta una creencia parcial mínima (Creencia Parcial Mínima –CPM<sub>m</sub>-). Cuando para un determinado síntoma/lesión coinciden CPM<sub>x</sub> y CPM<sub>m</sub>, la desviación estándar del nivel de creencia será cero.

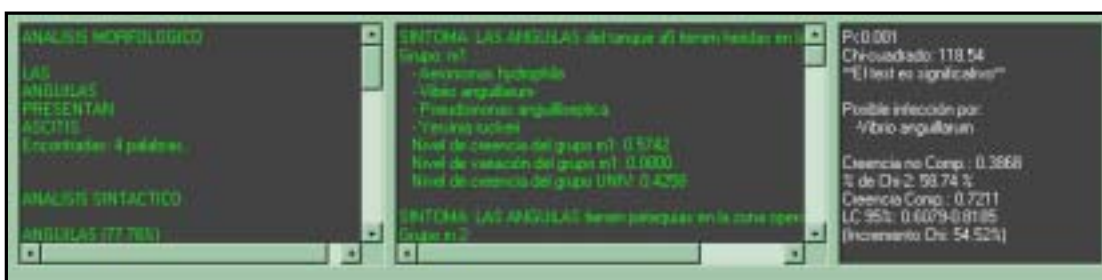


**Figura AIII.17.** Haciendo 'click' en el botón derecho del ratón sobre la malla Dempster-Shafer se visualiza un menú flotante que permite ampliar y reducir el tamaño de la malla

#### III.2.2.4. Bloque inferior

En el bloque inferior de la ventana principal (figura AIII.18) se pueden distinguir tres ventanas de hipertexto con el fondo de color negro. La situada más a la izquierda (con el texto de color verde) es la ventana que muestra los resultados del análisis morfológico-sintáctico indicándonos, por tanto, el número de palabras encontradas, la **RTA** formada así como el valor 'porcentaje compartido' asociado a cada estructura de la **RTA** y el valor del **[MPC]**. Como complemento a la información anterior también se pueden observar los conjuntos borrosos seleccionados en la **MAB** cargada durante el proceso de búsqueda.

En la ventana central (la de mayor tamaño, también con el texto de color verde) se pueden observar los síntomas/lesiones tal y como el usuario los escribió en la caja de texto que se encuentra en el bloque central. Debajo de cada síntoma/lesión se muestran los grupos formados, identificados por la letra **m** junto con su nivel de creencia, así como el nivel de creencia asignado al marco de discernimiento (identificado como **UNIV**) después de la entrada de nuevas fuentes de evidencia.



**Figura AIII.18.** Bloque inferior de la ventana principal de SEDPA 1.0

La tercera ventana, la situada más a la derecha (fondo negro con el texto en blanco) muestra los mejores resultados. Las tres primeras filas de la ventana están ocupadas en este orden por: el nivel de significación del test chi-cuadrado, el valor de chi-cuadrado y un mensaje que nos indica si el test es significativo o no. Un segundo bloque nos está indicando el grupo sintomático que presenta un mayor nivel de creencia, mientras que el tercer bloque nos muestra el nivel de creencia asociado al grupo sintomático anterior, el tanto por ciento de incremento del valor de chi-cuadrado, la creencia compensada asociada al grupo sintomático en función del incremento de chi-cuadrado, y los límites de confianza al 95 % para la creencia compensada cuando el número de síntomas introducidos es mayor a uno.

Para poder ver los detalles del test chi-cuadrado existe la opción de hacer 'click' con el botón derecho del ratón encima de la ventana. Aparecerá entonces un menú flotante con las opciones: *Mejor grupo*, *Chi-cuadrado* y *Valoración del diagnóstico*. Si seleccionamos *Chi-cuadrado* se visualiza una nueva malla que sustituye tanto a la ventana central como derecha, donde se pueden ver los cálculos detallados del test. Para volver a la situación anterior basta con pulsar nuevamente encima de la malla con el botón derecho del ratón y seleccionar *Mejor grupo* (figura AIII.19). La opción *Valoración del diagnóstico* nos presenta nuevamente la figura AIII.8.

Chi-cuadrado	118.54			
gl	8			
p<	0.001			
Grupo	Observado (O)	Esperado (E)	O-E	((O-E) <sup>2</sup> /E (% Contribución)
m10	38.7	11	27.7	69.63 (58.7%)
m11	25.3	11	14.3	18.49 (15.6%)
m12	3.4	11	-7.6	5.23 (4.4%)
m13	12.3			0.15 (0.1%)
m14	9.1			0.33 (0.3%)
m15	4.2			4.17 (3.5%)
m16	3.1			5.62 (4.7%)

Mejor grupo

✓ Chi-cuadrado

Valoración del diagnóstico

**Figura AIII.19.** Malla que presenta los detalles del test chi-cuadrado modificado y el menú flotante que permite su visualización

## CAPÍTULO 4

---

### CARACTERIZACIÓN Y PREDICCIÓN DE LOS PRINCIPALES PARÁMETROS FÍSICO-QUÍMICOS DEL AGUA

#### 4.1. Introducción

- 4.1.1. Descripción de los factores físico-químicos críticos
  - 4.1.1.1. Oxígeno disuelto
  - 4.1.1.2. Amoníaco, nitritos y nitratos
  - 4.1.1.3. Temperatura
  - 4.1.1.4. Carácter ácido-básico (pH)
- 4.1.2. Aprovechamiento de los efluentes de refrigeración de las centrales térmicas
- 4.1.3. El problema de la predicción
- 4.1.4. Desarrollo histórico de las Redes Neuronales Computacionales (**RNCs**)
- 4.1.5. Aplicaciones de las Redes Neuronales Computacionales
  - 4.1.5.1. Biología, producción vegetal, producción animal, química y medicina
  - 4.1.5.2. Hidrología, hidráulica e ingeniería
  - 4.1.5.3. Comparación con otras metodologías

#### 4.2. Material y métodos

- 4.2.1. Fuente de los datos
- 4.2.2. Métodos tradicionales de predicción: análisis multivariante (regresión múltiple) y análisis univariante (descomposición y **ARIMA**)
  - 4.2.2.1. Análisis de regresión múltiple
  - 4.2.2.2. Métodos de suavizado y modelos **ARIMA**
- 4.2.3. Predicción de los principales factores físico-químicos críticos mediante la aplicación de Redes Neuronales Computacionales
  - 4.2.3.1. Función de transferencia
  - 4.2.3.2. Entrenamiento de la Red Neuronal
  - 4.2.3.3. Arquitectura de las Redes Neuronales
  - 4.2.3.4. Patrones de entrenamiento durante la calibración
- 4.2.4. Generalización de los modelos
- 4.2.5. Pretratamiento de la información

#### 4.3. Resultados

- 4.3.1. Variables de la Central Térmica de Puente Nuevo. Análisis de la producción bruta
- 4.3.2. Variables de planta
  - 4.3.2.1. Temperatura en la serie 'A'



- 4.3.2.2. Temperatura en la serie 'B'
- 4.3.2.3. Temperatura en las series 'E' y 'F'
- 4.3.2.4. Amoniaco en la serie 'A'. Estimación a partir de las concentraciones de amoniaco de los días previos
- 4.3.2.5. Amoniaco en la serie 'A'. Estimación a partir de otras variables implicadas
- 4.3.2.6. Amoniaco en la serie 'A'. Estimación a partir de la serie filtrada
- 4.3.2.7. Amoniaco en las series 'B', 'E' y 'F'
- 4.3.2.8. Nitritos en la serie 'A'
- 4.3.2.9. Nitritos en las series 'B', 'E' y 'F'
- 4.3.2.10. Nitratos en la serie 'A'
- 4.3.2.11. Nitratos en las series 'B', 'E' y 'F'
- 4.3.2.12. pH en la serie 'A'

#### 4.4. Discusión

- 4.4.1. Temperatura
- 4.4.2. Amoniaco
- 4.4.3. Nitritos y nitratos
- 4.4.4. pH

#### 4.5. Referencias

#### 4.6. ANEXO IV.1. Glosario

#### 4.7. ANEXO IV.2. Manual del usuario del simulador de Redes Neuronales Computacionales (**REDGEN 1.0**)

## CAPÍTULO 4. CARACTERIZACIÓN Y PREDICCIÓN DE LOS PRINCIPALES PARÁMETROS FÍSICO-QUÍMICOS DEL AGUA

### 4.1. INTRODUCCIÓN

Uno de los factores más importantes que afecta al éxito o fracaso de una inversión en una piscifactoría es el del suministro de agua. La cantidad de agua disponible determina la cantidad de peces que se pueden producir, pero si la calidad de agua no es la adecuada, evidentemente la piscifactoría está abocada al fracaso. Cada especie de pez posee unos límites para los diversos parámetros de la calidad del agua, tales como la temperatura, oxígeno disuelto, sustancias nitrogenadas, pH, salinidad, etc., e idealmente la piscifactoría debería operar en los niveles óptimos de cada parámetro para conseguir así un rápido crecimiento y un comportamiento eficiente. Sin embargo al considerar la calidad del agua es necesario tener en cuenta que los diversos parámetros interaccionan entre sí y que el nivel preferido por el pez y las condiciones óptimas para dichos parámetros varían considerablemente a lo largo de su ciclo vital (Shepherd y Bromage, 1999). Por ello, el control y la predicción de la variación de los principales parámetros físico-químicos se constituye como aspectos fundamentales del buen funcionamiento de la explotación (Allan y Maguire, 1991; Bejda *et al.*, 1992; Alcaraz y Espina, 1995; Burel *et al.*, 1996; Culberson y Piedrahita, 1996).

Según Coll-Morales (1991) entre las variables físico-químicas de calidad del agua que pueden afectar a los cultivos acuáticos se encuentran: la temperatura, la concentración de amoníaco, el pH, la concentración de oxígeno, las concentraciones de nitritos, nitratos, fósforo, calcio, magnesio y cloro, la salinidad, las partículas sólidas en suspensión, los metales pesados y compuestos orgánicos. De todos estos factores, los más críticos sin duda son la temperatura (°C), la concentración de amoníaco (NH<sub>3</sub> mg/l), el oxígeno disuelto (O<sub>2</sub> mg/l), nitratos (NO<sub>3</sub><sup>-</sup> mg/l), nitritos (NO<sub>2</sub><sup>-</sup> mg/l) y la concentración de hidrogeniones (pH).

#### 4.1.1. Descripción de los factores físico-químicos críticos

##### 4.1.1.1. Oxígeno disuelto

El oxígeno disuelto es uno de los requerimientos más importantes en la calidad del agua en cualquier sistema de cultivo, ya que éste tiene una gran influencia sobre el consumo de alimento, tasa metabólica y gasto energético (Boyd *et al.*, 1978; Elliott, 1982). Existen tres factores principales que afectan a la solubilidad del oxígeno en el agua: la temperatura, la salinidad y la presión. Otros factores que pueden afectar a las concentraciones de oxígeno son: *blooms* fitoplanctónicos, elevadas cargas de materia orgánica y la propia respiración de los peces y otros organismos.

Los requerimientos básicos de los peces en cuanto al oxígeno varían entre especies, tamaño, temperatura, alimentación, estrés, pero se puede admitir que para peces en descanso los consumos se encuentran entre 80 y 500 mg de O<sub>2</sub> disuelto por kilogramo de peso húmedo y hora. Para peces en activo los consumos se disparan hasta 300-1500 mg de O<sub>2</sub> disuelto kg<sup>-1</sup> h<sup>-1</sup>. El oxígeno se demandará en mayores cantidades tras la alimentación, con el aumento de la temperatura o con el descenso del flujo de agua.

Doudoroff y Shumway (1970) y Brett y Blackburn (1981) han demostrado que los peces expuestos a condiciones de hipoxia muestran unas reducidas tasas de crecimiento y unas altas mortalidades. El mismo efecto se consigue cuando la concentración de oxígeno es excesivamente alta (Mellergaard y Dalsgaard, 1989; Krise, 1991). Así, el conocimiento de los requerimientos de oxígeno y la tolerancia a la hipoxia es fundamental para una optimización de las condiciones de cultivo. Wienbeck (1981) indica que cuando la concentración de oxígeno disuelto es de 2 mg l<sup>-1</sup>, se produce una disminución significativa en la actividad y en la ingestión de alimento en anguilas europeas, existiendo una tendencia a migrar hacia la zona de entrada de agua en el tanque. Cuando el contenido de oxígeno disminuye por debajo de 1.5 mg l<sup>-1</sup> las anguilas se dirigen hacia la superficie para iniciar un respiración aérea. Asimismo, Wienbeck (1981) determinó la existencia de una relación no lineal significativa entre el consumo máximo de oxígeno llevado a cabo por las anguilas y la temperatura del agua. Los estudios llevados a cabo por Egusa (1958), Tarr y Hill (1979) y Degani y Gallagher (1985) sobre la anguila europea y otras especies de anguilas, muestran de igual forma que el consumo de oxígeno es afectado por la temperatura y el tamaño del cuerpo. Cruz-Neto y Steffensen (1997) señalan que la anguila europea tiene un requerimiento energético mínimo de 88 mg de O<sub>2</sub> kg<sup>-1</sup> h<sup>-1</sup>, presentando una gran capacidad de regulación y compensación a niveles bajos de oxígeno cuando la concentración de CO<sub>2</sub> no es excesivamente alta. Por otra parte, Hermann *et al.* (1962), Adelman y Smith (1970), Buentello *et al.* (2000) y Drapcho y Brune (2000) entre otros, han encontrado asociaciones significativas entre las ganancias de peso y las concentraciones de oxígeno disuelto para distintas especies.

La necesidad de controlar y predecir el comportamiento de este importante parámetro ha llevado a varios autores a sugerir modelos que en la mayoría de los casos utilizan los datos de la declinación del oxígeno disuelto durante la noche, y están basados en la premisa de que la columna de agua es uniforme en función de la profundidad (Boyd *et al.*, 1978; Romaine y Boyd, 1979; Meyer y Brune, 1982; Klemetson y Rogers, 1985; Madenjian *et al.*, 1987). Losordo y Piedrahita (1991), Boyd *et al.* (1994) y Culberson y Piedrahita (1996) en cambio, incorporan modelos mecanicistas bastante precisos que caracterizan los procesos físicos y bioquímicos que implican de forma conjunta a los niveles de oxígeno disuelto y temperatura. Oliveira y Franca (1998) modelan los efectos de los perfiles de concentración de oxígeno tanto en la fase líquida como gaseosa y aplican los resultados de las simulaciones del modelo a tanques de acuicultura que incorporan sistemas de aireación. Forsberg (1994) indica que mediante la aplicación de regresiones múltiples es posible explicar el 70 % de la variación en el consumo de oxígeno en salmones atlánticos (*Salmo salar* L.).

Ya que como se ha expuesto anteriormente, la cantidad de oxígeno disponible es un factor crucial para la supervivencia y desarrollo de los animales confinados, en la mayor parte de las explotaciones se opta por un control activo en tiempo real de dicho parámetro. Así, en el caso de Hidrorecursos, la concentración de oxígeno disuelto está controlada por un automatismo que mantiene sus niveles dentro de los rangos óptimos (descrito en varios apartados del capítulo 2), por lo que la modelación de este parámetro no se hace necesaria.

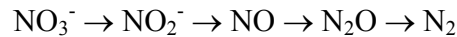
#### 4.1.1.2. Amoniac, nitritos y nitratos

Uno de los principales inconvenientes de los sistemas de acuicultura intensivos radica en la rápida acumulación en el agua de amoniac, nitritos y nitratos. El problema es causado principalmente por la presencia de compuestos nitrogenados (proteínas y ácidos nucleicos), alimento no consumido y los productos procedentes del desecho de los peces (Lin y Wu, 1996).

El catabolismo de estas moléculas biológicas resulta en la excreción de compuestos de nitrógeno reducido como el amoniac ( $\text{NH}_3$ ) (Hopkins *et al.*, 1993). En muchos animales el nitrógeno es excretado en forma de compuestos poco tóxicos como la urea o el ácido úrico. En animales acuáticos, el nitrógeno puede ser excretado directamente al medio a través de las branquias y piel en forma de amoniac, ya que éste se disuelve con facilidad en el agua (Hunnam *et al.*, 1982), encontrándose en un equilibrio dependiente de la temperatura y el pH con su forma ionizada o ion amonio ( $\text{NH}_4^+$ ) (Boyd, 1997). En ambientes acuáticos naturales, este nutriente es normalmente asimilado por las plantas pero en la mayor parte de los cuerpos de agua artificiales, como acuarios y sistemas de acuicultura, existe un número insuficiente de algas o macrófitas que sean capaces de eliminar el amoniac producido. En las explotaciones en las que no hay una recirculación del agua, la eliminación del amoniac se lleva a cabo mediante la renovación de ésta (Avnimelech *et al.*, 1994). En sistemas con recirculación, el amoniac y su forma ionizada son pasados a través de un filtro donde bacterias quimioautotróficas aeróbicas (*Nitrosomonas* y *Nitrobacter*) oxidan el amoniac en primer lugar hasta nitritos ( $\text{NO}_2^-$ ) y finalmente hasta nitratos ( $\text{NO}_3^-$ ).

Los procesos de nitrificación reducen la toxicidad del nitrógeno excretado, ya que la mayoría de los organismos acuáticos no toleran concentraciones de amoniac o nitritos superiores a  $10^{-3} \text{ g l}^{-1}$  (1.0 ppm) (Huey *et al.*, 1980; Colt *et al.*, 1981; Palachek y Tomaso, 1984; Lewis y Morris, 1986). En cambio, esos mismos organismos son capaces de soportar cantidades de nitratos superiores a los  $0.5 \text{ g l}^{-1}$  (500 ppm) (Hunnam *et al.*, 1982). La completa oxidación del amoniac es, por tanto, un proceso que consume oxígeno disuelto y aumenta la concentración de hidrogeniones ( $\text{H}^+$ ), por lo que tiende a acidificar el agua. Cuando las concentraciones de nitratos alcanzan niveles tóxicos se debe proceder a su eliminación. Este proceso en la mayoría de los casos se realiza mediante la sustitución de parte del agua del sistema o a través de una filtración anaerobia. En este último caso, el nitrato y el nitrito sirven como aceptores de electrones terminales en la respiración anaerobia

de algunas especies de bacterias. Esto permite a las bacterias en ausencia de oxígeno la eliminación de nitratos y nitritos directamente del agua siguiendo la secuencia (Boyd, 1997):



Los efectos de los productos nitrogenados de desecho de los peces y su relación con otros parámetros físico-químicos han sido comprobados por varios autores. Así, Huey *et al.*, (1984) sugieren que las altas temperaturas incrementan la captación de nitritos en *Ictalurus punctatus* aumentando por tanto su toxicidad. La relación entre el tamaño del pez y la toxicidad del amoníaco y nitritos es aun una cuestión abierta a debate. Almendras (1987) indica que la tolerancia a los nitritos es mayor en los peces más pequeños. Por contra, Wedemeyer y Yasutake (1978), Hilmy *et al.* (1987) y Alcaraz y Espina (1995) señalan que los peces de gran tamaño son menos sensibles a este tipo de contaminantes.

La aparición de estrés en los sistemas intensivos está asociado con una gran variedad de condiciones ambientales entre las que destacan los niveles altos de amoníaco y nitritos. Cuando los peces se encuentran estresados, se lleva a cabo una respuesta primaria consistente en la secreción de hormonas hepatopancreáticas (LDH, ADH, GDH, AST y ALT) que alteran la composición de la sangre y los tejidos y aumentan la frecuencia de ventilación y de latidos del corazón (Mazeaud *et al.*, 1977; Palackova *et al.*, 1990). Esto induce a una reducción en la efectividad del sistema inmune, aumentando la susceptibilidad a enfermedades y al ataque de parásitos, hace disminuir la tasa de crecimiento e incrementa la mortalidad. Todos los cambios producidos por la aparición de estrés llevan consigo la variación en el comportamiento y la fisiología del pez (Maule y Schreck, 1990; Israeli-Weinstein y Kimmel, 1998).

Dado el riesgo que la acumulación de este tipo de compuestos supone para la explotación, Kochba *et al.* (1994) y Avnimelech *et al.* (1994) entre otros han desarrollado modelos que predicen la concentración de moléculas nitrogenadas en los tanques. Por otra parte, un porcentaje importante de las investigaciones llevadas a cabo sobre el comportamiento de los compuestos nitrogenados que intervienen en el proceso de nitrificación se han centrado en sistemas de producción con tecnología de recirculación (Speece, 1973; Foster, 1974; Liao y Mayo, 1974). De esta forma, autores como Gujer y Boller (1986), Nijhof (1994a,b) y Kamstra *et al.* (1998) desarrollan modelos físicos que permiten describir el comportamiento de los reactores en los que se lleva a cabo la nitrificación. Estos modelos incorporan variables tales como la carga hidráulica, el tipo de sustrato del filtro, sus dimensiones y la estratificación bacteriana a lo largo de la columna de filtración, permitiendo el cálculo de la carga de amoníaco en el biofiltro y su tasa de eliminación.

#### 4.1.1.3. *Temperatura*

La temperatura corporal de los peces depende de las variaciones de la temperatura del medio acuático y se sitúa alrededor de 0.5 °C por encima del valor de esta última. Esto es debido a que los peces son seres poiquiloterms. De esta forma, la temperatura del agua afecta a la actividad, comportamiento, alimentación, crecimiento y reproducción de todas las especies de peces (Burns, 1976; Dosoretz y Degani, 1987; Degani *et al.*, 1988; Soderberg, 1990; Wootton, 1991; Hart y Purser, 1995; Hart *et al.*, 1996; Bevelhimer y Bennett, 2000). Después del oxígeno, la temperatura es posiblemente el factor más importante que interviene en el bienestar de los peces durante su ciclo de vida.

Las reacciones fisiológicas de los peces obedecen a la ley del 'Q10' según la cual una elevación térmica de 10 °C entraña la duplicación de la cinética de las reacciones enzimáticas, así como la de los fenómenos fisiológicos (Huner y Dupree, 1984; Kinkelin *et al.*, 1991). De este modo, mientras que un pez permanezca en los límites térmicos correspondientes a su especie, sus funciones se desarrollarán con normalidad aunque no forzosamente de forma óptima. Es decir, ciertas funciones se acelerarán constantemente hasta que se alcance la temperatura letal, mientras que otras sólo lo harán al llegar a la temperatura óptima fisiológica, y a partir de ahí, se ralentizarán.

Una consecuencia original de la vida a temperatura variable es que la edad de un individuo depende, no solo del número de días transcurridos desde la eclosión del huevo, sino también de la temperatura existente durante este período. Así, la edad de un pez puede ser evaluada en grados/día, lo que expresa mejor el valor de la edad fisiológica del animal. Esto es particularmente importante en explotaciones intensivas ya que la edad fisiológica condiciona las funciones de defensa y, por lo tanto, interviene en la receptividad a las infecciones (Dorson y Torchy, 1981).

Cairns *et al.* (1975) encontraron que la temperatura es un factor determinante en la toxicidad química de algunos compuestos. La interacción toxicidad-temperatura es una relación compleja, por lo que es difícil predecir su efecto. Temperaturas elevadas pueden incrementar la toxicidad de los nitritos directamente, ya que dicha elevación incrementa el flujo de agua y el intercambio iónico a través de las branquias. Como consecuencia la acumulación de nitritos en la sangre del pez se incrementa (Huey *et al.*, 1984). Alcaraz y Espina (1995) constatan este efecto sobre *Ctenopharyngodon idella* donde la mayor toxicidad de los nitritos es alcanzada a elevadas temperaturas. Sin embargo, temperaturas relativamente elevadas pueden proporcionar un efecto de protección frente a este contaminante. En este sentido, es importante considerar que la toxicidad de los nitritos no sólo depende de la cantidad captada por el organismo sino también del proceso de desintoxicación a través del sistema meteglobin-reductasa. Se sabe que los procesos bioquímicos y enzimáticos son sensibles a la temperatura, siendo éstos más eficientes en el intervalo termal correspondiente a la temperatura preferida por las especies.

Varios estudios también han relacionado las variaciones de temperatura con la asimilación de proteínas y ácidos grasos a partir de distintos tipos de alimentos, sugiriendo que la temperatura a la cual se consiguen unos crecimientos más rápidos en la anguila europea se encuentra alrededor de 25 °C (Degani, 1986; Dosoretz y Degani, 1987; Degani *et al.*, 1988).

Al igual que con otro tipo de parámetros, la evolución de la temperatura y partes del balance de energía del agua contenida en estanques ha sido modelada por varios autores (Losordo y Piedrahita, 1991; Culberson y Piedrahita, 1996). Bevelhimer y Bennett (2000) desarrollaron un modelo que evalúa aspectos temporales de la exposición termal y la dinámica del estrés termal acumulado, así como la recuperación de los peces expuestos a períodos de fluctuación de temperatura.

#### 4.1.1.4. *Carácter ácido-básico (pH)*

El valor del pH tiene un efecto significativo sobre la salud y el crecimiento de los peces tanto en poblaciones naturales como en sistemas de producción en medio acuático. El pH determina la concentración relativa de especies de sistemas ácido-base tal como amonico-ion amonio (Fivelstad, 1988; Colt y Orwicz, 1991; Boyd, 1997) o dióxido de carbono-ácido carbónico-bicarbonato-carbonato (Sanni y Forsberg, 1996; Boyd, 1997). En el caso del sistema amonico-ion amonio, la concentración relativa de las dos especies es particularmente importante dado la diferencia de toxicidad entre las dos formas. Por otra parte, el aumento de la concentración de CO<sub>2</sub> en el agua como consecuencia de la respiración de los peces se traduce en un proceso de acidificación (Colt y Orwicz, 1991). De la misma forma que con el resto de parámetros, el pH no es un factor fácil de predecir ni controlar, ya que su valor cambia como consecuencia de los procesos bioquímicos que ocurren en el agua y con los intercambios de gases que se producen entre el agua y la atmósfera (Piedrahita y Seland, 1995).

La obtención de modelos de comportamiento del pH ha sido llevado a cabo por varios autores. Piedrahita y Seland (1995) basan su modelo en la compleja resolución de la ecuación polinomial de sexto grado que relaciona la alcalinidad con el pH en función de la actividad del ion hidrógeno. En la valoración de un modelo de la capacidad de transporte llevado a cabo sobre agua dulce, Colt y Orwicz (1991) aplicaron un submodelo para el cálculo del pH tras la adición de CO<sub>2</sub>. Sanni y Forsberg (1996) presentaron una simplificación analítica de la obtención del valor del pH en un sistema de acuicultura de salmones, en el que las variables fueron el pH y alcalinidad a la entrada del sistema y el carbonato producido como consecuencia de la respiración de los peces.

#### 4.1.2. Aprovechamiento de los efluentes de refrigeración de las centrales térmicas

La refrigeración en las centrales térmicas productoras de electricidad es debida a la necesidad de condensar el vapor utilizado en la generación de energía. El vapor de agua a presión mueve una turbina que, a través de un alternador produce energía eléctrica con un rendimiento del 30-40 %. Para la condensación del vapor que permite el cierre del ciclo se utiliza agua procedente del ambiente como refrigerante. Este agua tras su paso por las torres de refrigeración arrastra, en función del rendimiento de la instalación, un 60-70 % de la energía producida en forma de calor que se vierte de nuevo al ambiente. Así, para evitar que la temperatura de vertido sea excesivamente alta, el volumen de agua bombeado en el proceso de refrigeración ha de ser elevado. Esta es la razón fundamental por la que las centrales se sitúan a orillas del mar, ríos o pantanos.

El vertido de los efluentes termales en el medio ambiente puede alterar la ecología de la zona (Bevelhimer y Bennett, 2000), si bien el incremento de la temperatura puede manipularse de manera que esa gran cantidad de energía que no tiene ningún otro uso sea aprovechada (Coll-Morales, 1991). Uno de los sectores que puede utilizar esa energía en su proceso de producción es la acuicultura, ya que la cantidad de agua templada descargada por una central térmica es considerablemente mayor que los requerimientos de cualquier piscifactoría. De esta forma, una central de una potencia de 2000 MW con cuatro turbinas generará una descarga de aproximadamente 65 m<sup>3</sup>/s. En estas condiciones, la cantidad de agua que una central genera con una sola turbina es de 16 m<sup>3</sup>/s que es aproximadamente 20 veces más de lo que necesita una piscifactoría que produzca 200 toneladas de anguilas al año (0.78 m<sup>3</sup>/s) (Aston, 1981). En este sentido, la central térmica de Puente Nuevo es capaz de generar una potencia de 318 MW producida por una única turbina, por lo que el volumen de agua bombeada para su refrigeración es de aproximadamente 12 m<sup>3</sup>/s.

Un factor importante en el diseño de posibles aplicaciones en acuicultura junto a centrales generadoras de energía son las medidas *antifouling* y anticorrosivas. Estas medidas se concretan en el uso de sulfatos y cloro. Ambos productos son tóxicos para muchas especies, aunque como indica Heath (1977) el problema no es grave si las especies cultivadas presentan una tolerancia relativamente alta a este tipo de sustancias (anguilas y carpas). Los metales pesados son otros tipos de compuestos que los efluentes de refrigeración pueden transportar como consecuencia del trasiego del agua a través de la instalación de la central. Si bien, las concentraciones de estos metales son por regla general muy pequeñas, algunos seres vivos son capaces de concentrarlos en sus tejidos, pudiendo ocasionar un grave problema sanitario. Romeril y Davis (1976) encontraron que las concentraciones de metales pesados en la piscifactoría de anguilas asociada a la central térmica de Ratcliffeon-Soar (Inglaterra) eran menores a las detectadas en otras explotaciones de características similares, cuya fuente de agua no procedía de centrales térmicas, así como a las encontradas en poblaciones naturales.

Lo más característico del efluente de refrigeración es, sin lugar a dudas, su temperatura. Normalmente el perfil de temperaturas del efluente es paralelo al perfil de



temperatura ambiental, pero con un incremento aproximado de entre 5 y 10 °C (Coll-Morales, 1991). Esta aproximación podría ser utilizada como una predicción relativamente fiable de la temperatura del agua en el interior de la piscifactoría, lo que permitiría la adecuación del régimen de bombeo óptimo de agua fría y caliente, y por tanto, el mantenimiento de la temperatura dentro de los rangos fisiológicos de la especie cultivada. Desgraciadamente, esta aproximación ideal en raras ocasiones es utilizada ya que el nivel de producción de las centrales generadoras de electricidad puede variar de forma significativa en 24 horas. Además estas instalaciones requieren un mantenimiento que en ocasiones supone la completa parada de las turbinas, y por tanto, la total interrupción del flujo de agua templada.

#### **4.1.3. El problema de la predicción**

Para resolver el problema de la predicción de la evolución de los principales parámetros de la calidad del agua, se pueden emplear varios tipos de técnicas. Una de ellas consiste en el cálculo de los factores a través de modelos matemáticos-físico (Boyd *et al.*, 1994; Giovannini y Piedrahita, 1994; Beg *et al.*, 1995; Culberson y Piedrahita, 1996; Piedrahita y Seland, 1995). Otras herramientas utilizadas para tales fines son las regresiones lineales simples y múltiples y los árboles de regresión y clasificación (**CART**) (Nerini *et al.*, 2000).

En este trabajo, junto con métodos tradicionales de predicción, como los análisis univariantes (métodos de descomposición y modelos **ARIMA**) y los análisis multivariantes (regresión múltiple) de series temporales, se propone el uso de Redes Neuronales Computacionales (**RNCs**), ya que han demostrado conseguir muy buenos resultados en el campo de las técnicas de control inteligente.

Las **RNCs** están siendo ampliamente utilizadas en el ajuste de funciones y en el modelado de sistemas altamente no lineales, ya que realizan una transformación no lineal entre los datos de entrada y salida, lo que supone importantes ventajas frente a las metodologías estadísticas convencionales tales como la regresión lineal múltiple o la regresión en componentes principales, en las que la relación entre los datos ha de ser lineal. Asimismo, no precisan de una ecuación para la relación empírica entre los datos como así ocurre en otras técnicas estadísticas paramétricas tales como la regresión no lineal o el filtro de Kalman. Por otro lado, se caracterizan por identificar la información relevante sobre una ingente cantidad de datos, discriminando la que es inconsistente, ambigua o incompleta. En siguientes apartados de esta introducción, se describe el desarrollo histórico de las **RNCs** así como algunas de sus diversas aplicaciones.

#### **4.1.4. Desarrollo histórico de las Redes Neuronales Computacionales (RNCs)**

A comienzos de la década de 1940, a medida que surgían las máquinas de calcular, unos pocos pensadores pioneros comenzaron a darse cuenta de que los ordenadores personales

podían ser más que masticadores de números. En ese momento, surgieron y lucharon por su reconocimiento dos visiones opuestas de lo que podían ser los ordenadores. Una facción los veía como sistemas para manipular símbolos por medio de reglas formales (rama simbólica), y la otra como un medio para modelar el cerebro (rama conexionista), buscando ambas la conducta inteligente. La primera usa los ordenadores para tener una representación formal del mundo; la segunda, para simular las interacciones de las neuronas. Estas dos tendencias, constituyen en 1956 el campo de la ciencia que se conoce como Inteligencia Artificial (Dreyfus y Dreyfus, 1999).

McCulloch y Pitts (1943) propusieron un modelo constituido por neuronas artificiales, en el que cada una de ellas se caracterizaba por estar encendida o apagada. El encendido se daba como respuesta a la estimulación producida por una cantidad suficiente de neuronas vecinas. Así, mostraron que cualquier función calculable podría determinarse mediante cierta red de neuronas interconectadas y que todos los conectores lógicos se podían lograr utilizando sencillas estructuras de red. Hebb (1949) demostró una sencilla regla de actualización para modificar las intensidades de las conexiones (pesos) entre neuronas, de manera que ocurriera el aprendizaje.

Minsky y Edmonds construyeron la primera computadora de red neuronal en 1951 denominada SNARC, que utilizaba un mecanismo de piloto automático obtenido de un avión bombardero B-24 para simular una red de 40 neuronas. Algunos de los investigadores del momento veían con escepticismo el que este tipo de trabajo pudiera considerarse como matemáticas (Russell y Norvig, 1996). En su disertación doctoral Minsky (1954) continuó con el estudio de las redes neuronales. Rosenblatt (1957) inventó el actual estilo perceptrón de las redes neuronales, constituido por unidades límite capacitables y, posteriormente, mostró la primera demostración del teorema de la convergencia del perceptrón (Rosenblatt, 1960). Caianello (1961) diseñó una teoría estadística del aprendizaje basándose en la mecánica estadística clásica.

La mayor parte del trabajo en el área anterior a 1970 fue enfocado a las máquinas tipo perceptrón de un nivel, aunque hubo algunas excepciones. Widrow y Hoff (1960) y Widrow (1962) desarrollaron dispositivos similares conocidos como sistemas autoorganizados de un nivel (adalines) o de varios niveles (madalines). Los libros *Principles of Neurodynamics* (Rosenblatt, 1962) y *Learning Machines* (Nilsson, 1965) resumen gran parte del trabajo realizado en esta época.

La publicación del libro *Perceptrons* (Minsky y Papert, 1969) señaló el fin de una era. Los autores criticaban severamente la experimentación sin directrices y la ausencia del rigor matemático que caracterizó buena parte de los primeros trabajos sobre perceptrones. Demostraron que si bien era posible lograr que un perceptrón aprendiera cualquier cosa que éste pudiera representar, su capacidad de representación era muy limitada. Un caso específico era la imposibilidad de capacitar a un perceptrón de dos entradas a discernir si éstas eran diferentes. Si bien los resultados que obtuvieron no eran aplicables a redes más

complejas, de substrato múltiple, los fondos para la investigación de redes neuronales se redujeron a prácticamente nada.

Puede considerarse que los artículos que aparecen en Hinton y Anderson (1981), señalaron el renacimiento del conexionismo. Algunos físicos compararon los grupos formados por neuronas con los conglomerados formados por átomos que constituyen los sólidos. Entre ellos destaca Hopfield (1982) que utilizó técnicas de mecánica estadística para analizar las propiedades de almacenamiento y optimización de las redes neuronales. La máquina de Boltzmann (Hinton y Sejnowski, 1983) y el análisis de las redes neuronales utilizando la teoría física de los lentes de spin magnético (Amit *et al.*, 1985) estrecharon los vínculos entre la mecánica estadística y la teoría de las redes neuronales. Les aportó no sólo soluciones matemáticas, sino también respetabilidad. El impulso más fuerte se produjo a mediados de la década de los 80, cuando algunos investigadores popularizan el algoritmo de aprendizaje de propagación posterior o retropropagación, mencionado por primera vez por Bryson y Ho (1969). Este algoritmo se aplicó a diversos problemas relacionados con el aprendizaje en computación y en psicología. La gran difusión que conocieron los resultados obtenidos, publicados por Rumelhart y McClelland (1986), suscitaron gran interés.

La capacidad de expresión de las redes de varios niveles fue objeto de investigación de Cybenko (1989), quien demostró que bastaban dos niveles ocultos o intermedios para representar cualquier función y que uno solo era suficiente para representar cualquier función continua. Asimismo, la complejidad del aprendizaje de las redes neuronales ha sido objeto del estudio de muchos investigadores. Algunos de los primeros resultados fueron obtenidos por Judd (1990) y Blum y Rivest (1992) que demostraron que en el espacio de pesos puede haber una cantidad exponencial de mínimos locales, y que por el contrario, un algoritmo de remontamiento de máximos con reinicio aleatorio era capaz de encontrar un óptimo global en un tiempo polinomial.

Todos estos estudios han convencido a un gran número de investigadores, provocando un creciente interés por estas técnicas, no sólo en el campo de la Inteligencia Artificial, sino en otras áreas de conocimiento en las que se hace uso de todos los recursos que esta herramienta puede proporcionar.

#### **4.1.5. Aplicaciones de las Redes Neuronales Computacionales**

Desde principios de la década de los 90, los modelos de **RNCs** se han aplicado en muchos campos de las ciencias y de la ingeniería, obteniéndose resultados bastante satisfactorios. En este apartado se presentan algunos ejemplos tomados de las muchas e importantes aplicaciones de las redes neuronales.

#### 4.1.5.1. *Biología, producción vegetal, producción animal, química y medicina*

En biología la mayor parte de las aplicaciones con redes neuronales se han desarrollado en medicina y biología molecular (Albiol *et al.*, 1995; Faraggi y Simon, 1995; Lo *et al.*, 1995). No obstante, se pueden encontrar algunas aplicaciones en sistemas ecológicos y en ciencias ambientales. Colasanti (1991) y Edwards y Morse (1995) encontraron similitudes entre las redes neuronales y los ecosistemas, recomendando la utilización de esta herramienta en la modelación de ecosistemas. Varios autores han empleado este tipo de técnicas en distintos campos de la ecología aplicada. De esta forma, Seginer *et al.* (1994) simula el comportamiento de las condiciones ambientales en el interior de un invernadero, Baran *et al.* (1996) y Lek *et al.* (1996a,b) predicen la variación de varios parámetros importantes en el manejo de la trucha común. Dagorn *et al.* (1997), Giske *et al.* (1998), Huse y Giske (1998) y Dagorn *et al.* (2000) modelan la dinámica espacial de varias especies de peces. Recknagel *et al.* (1997) predicen la producción del fitoplancton en lagos y embalses. Guegan *et al.* (1998) estiman la diversidad piscícola en ríos a través de parámetros ambientales y Gutiérrez-Estrada *et al.* (2000) predicen el índice gonadosomático en una población de perca sol introducida en la Península Ibérica.

En el área de producción vegetal las redes neuronales se están utilizando en temáticas muy diversas, como en el control de las condiciones ambientales de invernaderos (Seginer *et al.*, 1996; Sigrimis *et al.*, 2000), en la simulación de la temperatura de suelos cultivados (Yang *et al.*, 1997a), en la estimación de los niveles de contaminación de suelos agrarios (Parmar *et al.*, 1997; Yang *et al.*, 1997c), en la identificación de cultivos de olivo (Fiorino y Mancuso, 1998), y en la evaluación cuantitativa de la forma de la fruta (Morimoto *et al.*, 2000), entre otras. También en producción animal hay ejemplos de uso de esta herramienta, como en la clasificación cualitativa de jamones de cerdo ibérico (Hervás *et al.*, 1994), en la determinación del estado de confort de cerdos en naves de producción (Shao *et al.*, 1997), en la estimación del rendimiento de leche de ganado vacuno (Lacroix *et al.*, 1997) y en la predicción del nivel de aminoácidos de los ingredientes de los piensos para pollos (Roush y Cravener, 1997; Cravener y Roush, 1999)

En los últimos años la aparición de publicaciones utilizando las **RNCs** en química ha crecido enormemente como consecuencia de su capacidad de resolver problemas tales como el reconocimiento de patrones o la ordenación de datos (Zupan y Gasteiger, 1991; Burns y Whitesides, 1993; Ventura, 1996). El conjunto de aplicaciones abarca campos muy diversos, en los que se han utilizado las **RNCs** con varios fines tales como procesos de calibración, los cuales se han comparado con otras técnicas quimiométricas (Blank y Brown, 1993), métodos de modelación en la estimación de parámetros (Wessel y Jurs, 1994; Ventura *et al.*, 1995 y 1997a), reconocimiento o clasificación de patrones en análisis espectral (Hartnett *et al.*, 1993) y resolución de cinéticas altamente no lineales (Ventura *et al.*, 1997b; Hervás *et al.*, 1998).

En el ámbito de la medicina, las **RNCs** han establecido criterios de diagnosis sobre los síntomas de una enfermedad a partir de los registros médicos de pacientes que lo

padecen. Destacan los trabajos de diagnóstico del dolor abdominal y apendicitis agudos (Pesonen *et al.*, 1996, 1998a y 1998b), de enfermedades psiquiátricas como la neurosis, la esquizofrenia (Zou *et al.*, 1996) y del vértigo (Kentala *et al.*, 1997).

#### 4.1.5.2. Hidrología, hidráulica e ingeniería

La lluvia, fundamental en la mayoría de los procesos hidrológicos así como en la gestión de los sistemas hidráulicos, es difícil de predecir debido a su gran variabilidad espacial y temporal. Las **RNCs** se han utilizado con este fin, dando excelentes resultados en el pronóstico a corto plazo, lo cual las hace muy atractivas para la estimación en tiempo real (Zhang *et al.*, 1997; Palazón, 1998; Hsu *et al.*, 1997 y 1999). Asimismo, Shin y Salas (2000) se apoyan en esta técnica para el análisis regional de sequías basándose en datos anuales de precipitación.

Otros trabajos dignos de mención son los de Raman y Sunilkumar (1995), Raman y Chandramouli (1996) y Dibike *et al.* (1999) que analizan la gestión de embalses de agua con **RNCs**. Saad *et al.* (1994) las aplican en la operación y control conjuntos de embalses y centrales hidroeléctricas, Norreys y Cluckie (1997) y Sánchez *et al.* (1998) en el diseño y operación de sistemas de alcantarillado urbano y en el filtrado de información (precipitación, humedad del suelo, usos del suelo, vegetación, etc.) procedente de sensores remotos (Islam y Kothari, 2000).

Zhang *et al.* (1994), Pulido-Calvo (2001) y Pulido-Calvo *et al.* (2001) desarrollaron metodologías para predecir las demandas diarias de agua en núcleos urbanos y en zonas regables, respectivamente. Estos autores eligen como variables de entrada a la **RNC** las demandas de agua y factores ambientales de varios días anteriores al de estimación, obteniendo resultados bastante satisfactorios.

Son numerosas las aplicaciones de **RNCs** en ingeniería civil (Kartam *et al.*, 1997), como en la detección de daños causados por la vibración en estructuras (Wu *et al.*, 1992; Elkordy *et al.*, 1993), en la estimación y control de la productividad de la construcción (Murtaza y Fisher, 1994), en la ingeniería del transporte (Dougherty *et al.*, 1993) y en el modelado del comportamiento del suelo (Ellis *et al.*, 1995).

#### 4.1.5.3. Comparación con otras metodologías

En una gran cantidad de trabajos se compara el funcionamiento de las redes neuronales con técnicas estadísticas tradicionales, como las regresiones simples y múltiples (Karunanithi *et al.*, 1994; Wessel y Jurs, 1994; Tamari *et al.*, 1996; Muttiah *et al.*, 1997; Parmar *et al.*, 1997; Tokar y Johnson, 1999; Gutiérrez-Estrada *et al.*, 2000; Johnson y Rogers, 2000; Thirumalaiah y Deo, 2000; Pulido-Calvo, 2001), los modelos de Box-Jenkins de series

temporales (Hsu *et al.*, 1995; Tawfik *et al.*, 1997; Fernando y Jayawardena, 1998; Thirumalaiah y Deo, 2000; Pulido-Calvo, 2001), análisis discriminantes (Hervás *et al.*, 1994), y con modelos físicos conceptuales (Aziz y Wong, 1992; Hsu *et al.*, 1995; Álvarez y Bolado, 1996; Hjelmfelt y Wang, 1996; Tokar y Johnson, 1999; Tokar y Markus, 2000). En general, el comportamiento de las **RNCs** es superior al de los métodos clásicos, sobre todo cuando se pretende un ajuste de datos de gran dispersión numérica y/o que no se disponga de una longitud de datos estadísticamente consistente (Govindaraju, 2000).

## **4.2. MATERIAL Y MÉTODOS**

### **4.2.1. Fuente de los datos**

Los datos necesarios para la calibración y evaluación de los modelos han tenido tres orígenes fundamentales: a) las bases de datos contenidas en el ordenador central de Hidrorecursos; b) los análisis químicos diarios llevados a cabo por los operarios de la planta y c) la central térmica de Puente Nuevo.

De la primera fuente se ha extraído básicamente la temperatura así como el valor del pH (sólo para la serie 'A'), que eran medidos y almacenados de forma automática a intervalos regulares de 6 horas. Por otra parte, la concentración de amoníaco, nitratos y nitritos se obtuvo a través del test colorimétrico Microquant<sup>®</sup> de la casa comercial Merck<sup>®</sup>. Este test tiene la ventaja de que es utilizable particularmente para aguas turbias o coloreadas debido a su principio de luz transmitida más un blanco de muestra. El comparador de color está constituido de un plástico estable y un disco giratorio de 10 grados que se corresponden con la gradación entre los rangos máximos y mínimos de medición. Una ventaja adicional de este tipo de test es su fácil utilización por personal no especializado. La medición de estos tres parámetros se llevó a cabo dos veces al día (mañana y tarde) coincidiendo con el cambio de turno de los operarios. Finalmente, los datos de producción bruta generada en la central térmica de Puente Nuevo fue obtenida y cedida por el departamento de producción de la propia central.

### **4.2.2. Métodos tradicionales de predicción: análisis multivariante (regresión múltiple) y análisis univariante (descomposición y ARIMA)**

#### *4.2.2.1. Análisis de regresión múltiple*

El análisis de regresión múltiple pretende obtener una ecuación lineal que permita estimar la variable dependiente o criterio **e**, conocidos los valores que presentan las **q** variables independientes o predictivas, **x<sub>1</sub>,...,x<sub>q</sub>**:

$$e = \beta_0 + \beta_1 x_1 + \dots + \beta_q x_q + \varepsilon \quad (4.1)$$

donde el término  $\beta_0$  es una generalización del término bivariante de ordenada en el origen, los coeficientes  $\beta_1, \dots, \beta_q$  son los denominados coeficientes de regresión parcial y el valor  $\epsilon$  es el término de error que ajusta la predicción del modelo al valor exacto de  $e$ .

Se admite que el error  $\epsilon$ , ante una aplicación repetida del modelo, representa pequeños desajustes distribuidos normalmente alrededor de una media cero. Asimismo, los coeficientes  $\beta_0, \beta_1, \dots, \beta_q$  son los valores ideales que configurarían el modelo exacto para la población en la que el número de datos tiende a ser infinito. Ya que la muestra de datos es limitada, habrá que conformarse con una estimación de estos parámetros ( $b_0, b_1, \dots, b_q$ ) a partir de la ecuación:

$$e = b_0 + b_1 x_1 + \dots + b_q x_q \quad (4.2)$$

donde  $e$  representa el valor medio estimado para todos aquellos casos con un determinado perfil de  $x_1, \dots, x_q$ , y los parámetros  $b_0, b_1, \dots, b_q$  se calculan mediante un ajuste de mínimos cuadrados de forma que sea mínima la suma de cuadrados de los desajustes entre los valores reales de  $e$  en cada individuo de la muestra y los valores teóricos que proporciona la ecuación buscada. Para aceptar con una seguridad cuantificada la validez del modelo diseñado se analiza la significación estadística global de la relación de  $e$  con las variables predictivas  $x_1, \dots, x_q$ , mediante el análisis de la varianza (Hair *et al.*, 1999).

#### 4.2.2.2. Métodos de suavizado y modelos ARIMA

En el análisis univariante de series temporales se trata de hacer previsiones de valores futuros de una variable (parámetros físico-químicos), utilizando como información únicamente la contenida en los valores pasados de la serie temporal que, mide la evolución de la variable objeto de estudio. En estos análisis se pueden considerar dos grandes grupos: métodos de descomposición y modelos ARIMA. Los métodos de descomposición utilizados en este trabajo son el suavizado exponencial simple y el suavizado lineal exponencial de Holt (Uriel, 1995).

El suavizado exponencial simple consiste en determinar los valores del futuro de una variable  $d$  en función de una media ponderada de sus valores pretéritos. Si la variable  $d$  es sometida a un proceso de alisado exponencial simple se obtiene como resultado la variable alisada  $S_t$  (ecuación 4.3), siendo (4.4) la ecuación de predicción:

$$S_t = \alpha_d d_t + (1 - \alpha_d) S_{t-1} \quad (4.3)$$

$$e_{t+1} = S_t \quad (4.4)$$

La ponderación se efectúa mediante un parámetro  $\alpha_d$  ( $0 < \alpha_d < 1$ ). Si  $\alpha_d$  tiene un valor próximo a cero, se da mucho peso a las observaciones pasadas a través del término

$S_{t-1}$ ; por el contrario, si  $\alpha_d$  está próximo a la unidad, se da más importancia a la observación actual de la variable  $d$ . En el período de calibración (dependiendo del tipo de parámetro: temporadas de producción de 1997, 1998 y 1999), se prueba con distintos valores de  $\alpha_d$  y se elige el que de la mejor predicción. En la asignación del valor inicial de  $S_t$  ( $S_0$ ) se toma la media de los valores del parámetro de producción diario del período descriptivo (Uriel, 1995). Ambos parámetros serán los utilizados en la generalización (temporadas 2000 y 2001).

El método de Holt es también un método de alisado exponencial que utiliza dos parámetros de alisado  $\alpha_d$  y  $\beta_d$ , y asimismo, se calculan dos variables de alisado para cada instante de tiempo  $t$ ,  $S_t$  y  $T_t$ :

$$S_t = \alpha_d d_t + (1 - \alpha_d)(S_{t-1} + T_{t-1}) \quad (4.5)$$

$$T_t = \beta_d (S_t - S_{t-1}) + (1 - \beta_d) T_{t-1} \quad (4.6)$$

siendo la ecuación de predicción:

$$e_{t+1} = S_t + T_t \quad (4.7)$$

Los valores iniciales para comenzar la recursión ( $S_0$  y  $T_0$ ) se obtienen a partir de los coeficientes obtenidos en el ajuste de una recta por mínimos cuadrados con los datos disponibles en el período de calibración. Asimismo, se prueban distintos valores de  $\alpha_d$  y  $\beta_d$ , y se eligen los que den el mejor ajuste del modelo descriptivo (Uriel, 1995). Estos parámetros serán los utilizados en el período de generalización o validación externa.

El modelo **ARIMA(P,D,Q)** de Box-Jenkins (*AutoRegressive-Integrated-Moving Average*) puede representar la correlación de la estructura de la serie temporal con un mínimo de parámetros, proporcionando una alta eficiencia en la predicción. La formulación considerada es:

$$(1 - \Phi_1 B - \dots - \Phi_p B^p)(1 - B^D) d_t = (1 - \Theta_1 B - \dots - \Theta_q B^q) A_t \quad (4.8)$$

siendo  $d_t$  el valor observado del parámetro físico-químico en el instante  $t$ ;  $A_t$  la diferencia en valor absoluto entre el valor del parámetro observado y predicho en el instante  $t$ ;  $B$  el operador retroceso que asigna a una variable su valor en el instante anterior:  $Bd_t = d_{t-1}$  y  $B^m d_t = d_{t-m}$ ;  $\Phi$  y  $\Theta$  los parámetros del modelo y  $P$ ,  $D$  y  $Q$  los órdenes del modelo.

Los valores de  $P$ ,  $D$  y  $Q$  suelen variar entre 0 y 2 (Moreu, 1999), por lo que para la identificación de los órdenes óptimos se van aumentando cada uno en una unidad y se observa el comportamiento del nuevo modelo con respecto a los demás, eligiendo el que de una mejor aproximación. Una alternativa a la metodología propuesta por Moreu (1999) consiste en el análisis de las funciones de autocorrelación y autocorrelación parcial de la serie de datos originales (Holton-Wilson y Keating, 1996). Según esta última, el valor de los



parámetros  $\mathbf{P}$  y  $\mathbf{Q}$  viene determinado por el nivel de significación de las correlaciones seriales estimadas para los distintos retrasos y por la forma en que disminuyen ambas funciones a lo largo del tiempo retrasado.

La estimación de los parámetros  $\Phi$  y  $\Theta$  del modelo descriptivo se realiza mediante la teoría de los mínimos cuadrados, y se ha de comprobar el nivel de significación de éstos. Estos parámetros se utilizan para la estimación de los valores diarios de los parámetros físico-químicos en el período de generalización.

Estos tres modelos muestran su bondad de ajuste a través de un análisis de residuos donde se estudian las funciones de autocorrelación de los mismos (Uriel, 1995; Moreu, 1999).

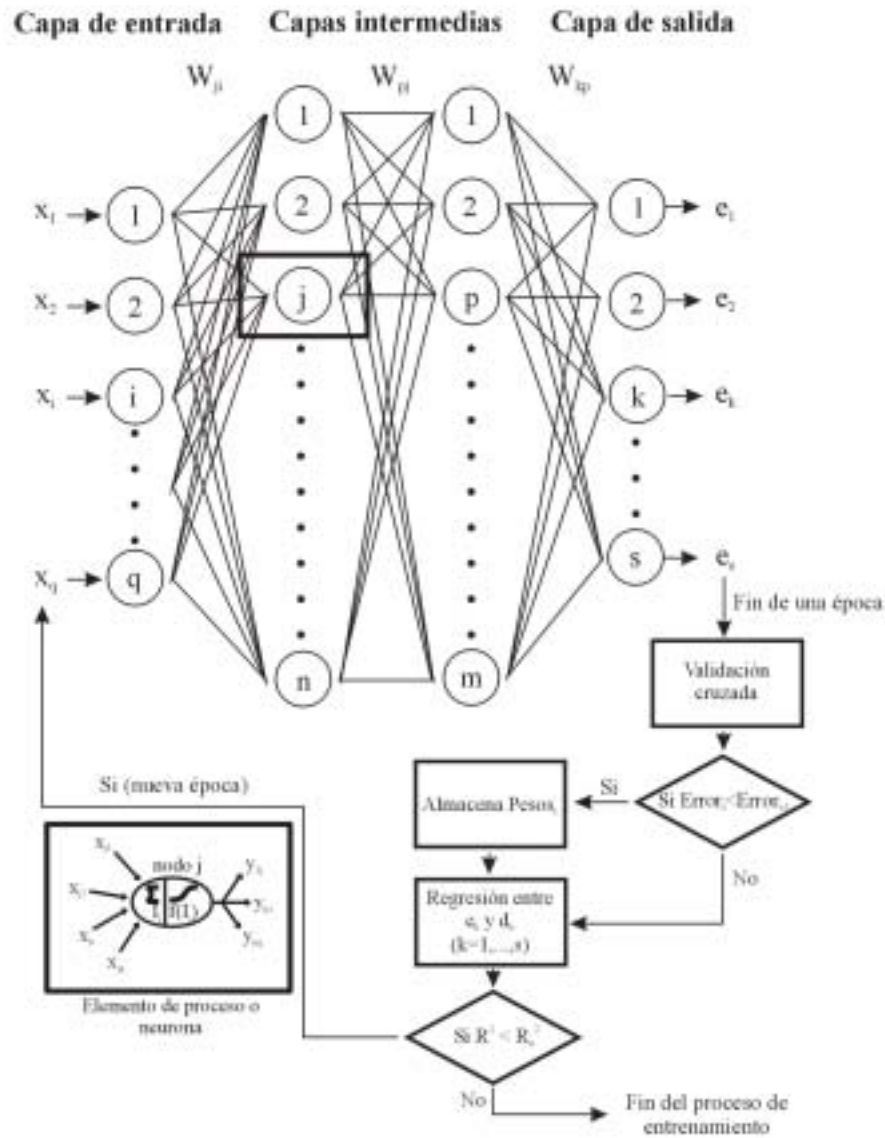
#### **4.2.3. Predicción de los principales factores físico-químicos críticos mediante la aplicación de Redes Neuronales Computacionales**

Las Redes Neuronales Computacionales (**RNCs**) son modelos matemáticos que copian la estructura de los sistemas neuronales biológicos con el fin de alcanzar una funcionalidad similar. El procesador elemental o neurona es un dispositivo simple de cálculo que, a partir de un vector de entrada procedente del exterior o de otras neuronas, proporciona una única respuesta o salida. En general, una red neuronal consiste en el establecimiento de un conjunto de nodos o neuronas agrupadas en varias capas e interconectadas entre sí. Las conexiones que unen a las neuronas tienen asociado un peso numérico, realizándose el aprendizaje mediante la actualización de tales pesos (figura IV.1).

De la multitud de modelos y variantes de **RNCs** que existen, en este trabajo se utilizan redes neuronales supervisadas y unidireccionales. En el aprendizaje supervisado se presenta a la red un conjunto de entradas y salidas deseadas u objetivos, e iterativamente ésta ajusta sus pesos hasta que su salida tiende a ser la deseada, utilizando para ello información detallada del error que comete en cada paso. En las redes unidireccionales la información circula en un único sentido, desde las neuronas de entrada hacia las de salida.

Las neuronas de entrada reciben los datos de partida, sin llevar a cabo ninguna operación matemática, solamente transfieren los valores a la siguiente capa. Cada neurona de las capas intermedias y de la capa de salida realiza la suma  $\mathbf{I}_j$  de los valores ponderados sobre los datos que recibe como entrada. De este modo, dado un conjunto de entradas  $\mathbf{x}_{ji}$  ( $i = 1, \dots, \mathbf{q}$ ) en la neurona  $\mathbf{j}$ , estos valores se multiplican por el conjunto de pesos  $\mathbf{W}_{ji}$  (conexión del nodo  $\mathbf{i}$  con el nodo  $\mathbf{j}$ ) (figura IV.1):

$$\mathbf{I}_j = \sum_{i=1}^{\mathbf{q}} \mathbf{x}_{ji} \mathbf{W}_{ji} \quad (4.9)$$



**Figura IV.1.** Red neuronal computacional con una capa de entrada de  $q$  neuronas, dos capas intermedias de  $n$  y  $m$  neuronas, respectivamente, y una capa de salida de  $s$  neuronas ( $q, n, m, s$ ). Variables de entrada:  $x_1, \dots, x_q$  Variables de salida de la red:  $e_1, \dots, e_s$  Variables deseadas de salida:  $d_1, \dots, d_s$

Posteriormente la neurona  $j$  calcula una respuesta  $y_{pj}$  ( $p = 1, \dots, m$ ) con una función de transferencia  $f(I_j)$ . Esta respuesta es transmitida a cada una de las neuronas  $p$  de la siguiente capa, que vuelve a efectuar la misma operación hasta alcanzar a las neuronas de salida (figura IV.1).

Las principales características de una **RNC** son la función de transferencia de las neuronas (respuesta a la suma ponderada de sus entradas), el algoritmo de aprendizaje (utilizado para el ajuste de los coeficientes ponderales de la red), y la arquitectura (números de capas y distribución de las neuronas en cada capa).

#### 4.2.3.1. Función de transferencia

La selección de la función de transferencia de las neuronas es arbitraria, mientras que cumpla la condición de ser una función acotada. En este trabajo, se han utilizado para las neuronas de salida la función lineal (l) y para cada neurona de las capas intermedias la función de transferencia logística, también conocida como sigmoide (s):

$$y_{pj} = f(I_j) = \frac{1}{1 + e^{-I_j}} \quad (4.10)$$

La ecuación (4.10) se ha seleccionado sobre la base de que, la respuesta de cada neurona de la red (incluida la salida), se encontrará en el intervalo  $[0,1]$  para cualquier entrada  $(-\infty, \infty)$ , lo que es coherente con el escalado de los datos experimentales del que se hace uso en este trabajo:

$$V_b^* = \frac{V_b - V_{\min,b}}{V_{\max,b} - V_{\min,b}} \quad (4.11)$$

donde  $V_b^*$  es el valor escalado de la variable  $V_b$ ;  $V_{\min,b}$  es el valor mínimo de la variable  $V_b$  minorado en un 15 % y  $V_{\max,b}$  es el valor máximo de la variable  $V_b$  aumentado en un 15 %. De esta forma se previene que valores mayores o menores a los utilizados en el entrenamiento o aprendizaje de la red, den respuestas inadecuadas en la validación del modelo.

#### 4.2.3.2. Entrenamiento de la Red Neuronal

La determinación de un conjunto de pesos que permita a la red realizar el tipo de procesamiento deseado se denomina aprendizaje. Este proceso consiste en un cálculo iterativo en el que un conjunto de pares de datos (entradas-salidas objetivo) es procesado, comparando la respuesta calculada por la red con la salida objetivo correspondiente y ajustando los coeficientes de peso de las neuronas para minimizar la diferencia entre la respuesta de la red y la salida objetivo. El algoritmo de aprendizaje utilizado en el presente trabajo es una variación del modelo de retropropagación estándar —introducido por Bryson y Ho (1969) y Werbos (1974), y estudiado posteriormente por otros autores tales como

Rumelhart y McClelland (1986)—, conocido como delta-barra delta extendida (Extended Delta-Bar Delta, **EDBD**) (Minai y Williams, 1990).

A cada coeficiente de peso  $W_{ji}$  le es asignado inicialmente un valor aleatorio que ha de ser ajustado durante el proceso de aprendizaje para minimizar la diferencia entre la respuesta de la red y la salida objetivo. El método más utilizado consiste en modificar progresivamente los coeficientes de peso desde la capa de salida hasta la entrada, propagando el error en sentido inverso al cálculo de la red. Este procedimiento se repite con todos los pares de datos iterativamente, hasta que se converge en un conjunto final de coeficientes de peso que pueden ser utilizados posteriormente en el cálculo de la respuesta de la red. La señal de error para cada neurona  $k$  de salida se estima a partir de la ecuación (4.12):

$$\delta_k = [d_k - f(I_k)] f'(I_k) \quad (4.12)$$

donde  $d_k$  es la salida deseada de la neurona  $k$ ,  $f(I_k)$  es la salida de la neurona  $k$  y  $f'(I_k)$  es la derivada de la función de transferencia de la neurona  $k$ . En el caso de que la neurona  $j$  sea de la capa intermedia, el error que se produce es la suma de los errores  $\delta_p$  en las neuronas a las que está conectada la salida de ésta, multiplicando cada uno de ellos por el peso de la conexión  $W_{pj}$ , y la señal de error se calcula como:

$$\delta_j = \left[ \sum_{p=1}^m \delta_p W_{pj} \right] f'(I_j) \quad (4.13)$$

donde el rango de  $p$  ( $p = 1, \dots, m$ ) cubre todas aquellas neuronas a las que está conectada la salida de la neurona  $j$ .

El cambio en el coeficiente de peso  $W_{ji}(t)$  de la conexión que une la neurona  $i$  con la neurona  $j$  se calcula iterativamente, a partir de (4.14) y (4.15):

$$\Delta W_{ji}(t+1) = \alpha(t) \delta_j(t) y_{ji}(t) + \mu(t) \Delta W_{ji}(t) \quad (4.14)$$

$$W_{ji}(t+1) = W_{ji}(t) + \Delta W_{ji}(t+1) \quad (4.15)$$

con  $\alpha(t)$  un coeficiente de proporcionalidad denominado tasa o coeficiente de aprendizaje que controla la velocidad con la que converge el algoritmo y  $\mu(t)$  el término de momento que tiende a mantener los cambios de los pesos en la misma dirección y obviar la permanencia en mínimos locales, para así alcanzar la convergencia rápidamente.

En el algoritmo estándar de retropropagación, el coeficiente de aprendizaje y el término de momento son los mismos para la corrección de todos los pesos de conexión, lo que ocasiona lentitud de convergencia. La evidencia empírica sugiere que cada dimensión del espacio de pesos puede ser bastante diferente en términos de la superficie de error total. Específicamente cada conexión de una red debería tener su propios coeficientes de

aprendizaje y de momento, ya que el tamaño de paso apropiado para una dimensión de peso puede no serlo para todas las otras dimensiones (Ventura, 1996). Por esto, será deseable que estos parámetros varíen con el tiempo para cada conexión, como así se lleva a cabo en el algoritmo **EDBD**.

Para implementar las reglas heurísticas de aumento o disminución del coeficiente de aprendizaje  $\alpha(\mathbf{t})$  y del término de momento  $\mu(\mathbf{t})$ , se usa una media ponderada del término de error de cada neurona  $\mathbf{j}$ :

$$\bar{\delta}_j(\mathbf{t}) = (1 - \theta) \delta_j(\mathbf{t}) + \theta \delta_j(\mathbf{t} - 1) \quad (4.16)$$

donde  $\theta$  es un factor de ponderación ( $0 < \theta < 1$ ).

La actualización del coeficiente de aprendizaje se realiza considerando los siguientes aspectos: a) si el componente del término de error actual y la media ponderada del componente del término de error anterior son del mismo signo, se aumenta la tasa de aprendizaje en función del valor absoluto de  $\delta_j(\mathbf{t})$ , b) si por el contrario sus signos son distintos, se disminuye el coeficiente de aprendizaje en una cantidad proporcional a su valor actual:  $-\varphi_\alpha \alpha(\mathbf{t})$  (ecuación 4.17). Para el término del momento se usa una regla semejante (ecuación 4.18):

$$\Delta\alpha(\mathbf{t}) = \left\{ \begin{array}{ll} \kappa_\alpha e^{(\gamma_\alpha |\bar{\delta}_j(\mathbf{t})|)} & \text{si } \bar{\delta}_j(\mathbf{t} - 1) \delta_j(\mathbf{t}) > 0 \\ -\varphi_\alpha \alpha(\mathbf{t}) & \text{si } \bar{\delta}_j(\mathbf{t} - 1) \delta_j(\mathbf{t}) < 0 \\ 0 & \text{en otro caso} \end{array} \right\} \quad (4.17)$$

$$\Delta\mu(\mathbf{t}) = \left\{ \begin{array}{ll} \kappa_\mu e^{(\gamma_\mu |\bar{\delta}_j(\mathbf{t})|)} & \text{si } \bar{\delta}_j(\mathbf{t} - 1) \delta_j(\mathbf{t}) > 0 \\ -\varphi_\mu \mu(\mathbf{t}) & \text{si } \bar{\delta}_j(\mathbf{t} - 1) \delta_j(\mathbf{t}) < 0 \\ 0 & \text{en otro caso} \end{array} \right\} \quad (4.18)$$

siendo  $\kappa_\alpha$  y  $\kappa_\mu$  factores de escalado para los coeficientes de aprendizaje y momento, respectivamente;  $\gamma_\alpha$  y  $\gamma_\mu$  factores exponenciales para el incremento de estos coeficientes;  $\varphi_\alpha$  y  $\varphi_\mu$  factores de decremento. Los valores usados de estos parámetros son:  $\theta = 0.7$ ;  $\kappa_\alpha = 0.095$ ;  $\gamma_\alpha = 0.1$ ;  $\varphi_\alpha = 0.1$ ;  $\kappa_\mu = 1$ ;  $\gamma_\mu = 0.05$  y  $\varphi_\mu = 0.01$ , los cuales aseguran un entrenamiento óptimo en un gran número de casos (Ventura, 1996). Asimismo, para evitar grandes oscilaciones en el espacio de pesos se limitan los valores máximos de  $\alpha(\mathbf{t})$  y  $\mu(\mathbf{t})$ , siendo los recomendados por Minai y Williams (1990) de 2 y 0.9, respectivamente.

La complejidad añadida al algoritmo estándar de retropropagación tanto desde el punto de vista computacional como de la cantidad de recursos necesarios, se ve ampliamente compensada por la aceleración que experimenta el aprendizaje de la red. Por estas razones, el algoritmo **EDBD** ha sido el empleado en el desarrollo de los modelos de **RNCs** presentados en este trabajo.

Se denomina época al período de tiempo que comprende todas las iteraciones realizadas al presentar todos los patrones una sola vez. Conforme más compleja sea la arquitectura de una **RNC** (mayor número de neuronas en las capas intermedias), mayor será el tiempo de procesamiento de una época. Sin embargo, el tiempo total del aprendizaje puede ser menor ya que pueden ser necesarias menos épocas. La actualización de los pesos se realiza tras la finalización de cada época.

#### 4.2.3.3. *Arquitectura de las Redes Neuronales*

Se denomina arquitectura a la topología o estructura de conexionado de una red neuronal. En la arquitectura estándar o totalmente conectada, éstas se organizan en unidades estructurales denominadas capas. Se distinguen tres tipos de capas: de entrada, de salida y ocultas. Una capa de entrada está compuesta por neuronas que reciben datos o señales procedentes del entorno. Una capa de salida es aquella cuyas neuronas proporcionan la respuesta de la red. Una capa oculta o intermedia es aquella que no tiene una conexión directa con el entorno (figura IV.1). Este tipo de capa proporciona a la red neuronal grados de libertad adicionales, gracias a los cuales puede encontrar representaciones internas correspondientes a determinados rasgos del entorno, proporcionando una mayor riqueza computacional.

De este modo, el número de neuronas en cada capa y el número de capas depende de la aplicación particular, es decir, del número de variables de entrada y del número y calidad de las respuestas buscadas en cada estudio. Aunque una red neuronal de una única capa intermedia puede representar cualquier función continua (Cybenko, 1989), en este trabajo se utilizan redes con dos capas ocultas (aproximador universal de cualquier función: Cybenko, 1989) ya que *a priori* no se conoce el tipo de relación entre las variables. Asimismo, puede ocurrir que con una sola capa intermedia el número de neuronas ocultas para alcanzar una cierta cota de error sea tan elevado que su aplicación resulte inabordable en la práctica (Martín del Brío y Sanz-Molina, 1997).

También aparece un fenómeno denominado sobreajuste cuando el número de parámetros (pesos) de la red resulta excesivo en relación al problema a tratar y al número de patrones de entrenamiento disponibles. Si el número de nodos ocultos es excesivo, el resultado que se obtiene en la generalización de la red se aparta de la realidad, pues no está capturando la tendencia del problema sino que está interpolando (memorizando) los datos, ajustando también el ruido presente en ellos. Por esto, se prueban distintas estructuras de la red neuronal y se elige la que de un menor error de generalización.

#### 4.2.3.4. Patrones de entrenamiento durante la calibración

A título de ejemplo, en la tabla IV.1 se muestran tres combinaciones de datos (patrones) de entrenamiento del modelo.

**Tabla IV.1.** Ejemplos de patrones de entrenamiento (entradas-salida objetivo) sin escalar

Nº patrón entrenamiento	Entrada Temperatura (°C)					Salida Temperatura (°C)
	Temp. (t-1)	Temp. (t-2)	Temp. (t-3)	Temp. (t-4)	Temp. (t-5)	Temp. (t)
1	25	25.4	25.5	26	25.7	23.9
2	23.9	25	25.4	25.5	26	24
3	24	23.9	25	25.4	25.5	24.8

Los patrones utilizados (dependiendo del tipo de parámetro) en el entrenamiento se corresponden con los datos de las temporadas 1997, 1998 y 1999, y en la generalización del modelo con los de la temporada 2000 y primeros meses del 2001.

Los modelos propuestos se han calibrado y validado con el programa de simulación de redes neuronales computacionales **REDGEN** 1.0 para MS Windows® desarrollado en lenguaje MS Visual Basic® y Borland C++® (Anexo IV.I).

#### 4.2.4. Generalización de los modelos

Un aspecto fundamental es la capacidad de un modelo de generalizar a partir de ejemplos. Por generalización se entiende la capacidad del modelo de aportar una respuesta correcta ante datos o patrones que no han sido empleados en su calibración o entrenamiento. En el caso de las redes neuronales es un hecho experimental fácilmente observable que si se entrena una red hasta alcanzar un error muy pequeño en el aprendizaje, la eficacia real del sistema se degrada (Martín del Brío y Sanz-Molina, 1997). Al principio, la red se adapta progresivamente al conjunto de aprendizaje, acomodándose al problema y mejorando la generalización. Sin embargo, en un momento dado el sistema se ajusta demasiado a las particularidades de los patrones empleados en el entrenamiento (sobreaprendizaje), por lo que crece el error que se comete ante patrones diferentes. De este modo, dada una arquitectura de red, ésta deberá entrenarse hasta alcanzar el punto óptimo en el que el error de generalización es mínimo.

En el contexto de predicción de variables existen métodos para la evaluación de modelos de **RNC** de entre los cuales el más utilizado es el método de validación cruzada

(Ventura, 1996) (figura IV.1). Para ello se selecciona una submuestra dentro del grupo de calibración que es sometida a una evaluación interna mediante la aplicación de los pesos obtenidos tras la última época. El error total cometido en el instante  $t$  es comparado con el error total obtenido en el instante  $t-1$ . Así, si el error en el instante  $t-1$  es menor que el error en el instante  $t$ , se almacenan los pesos para una posterior generalización o validación externa.

En este trabajo además, se propone un segundo sistema de control consistente en un aprendizaje controlado por un coeficiente de determinación umbral ( $R_u^2$ ) (Gutiérrez-Estrada *et al.*, 2000; Pulido-Calvo, 2001). Al final de cada época, se comparan los valores estimados por la **RNC** y los valores observados mediante una regresión lineal. Si la determinación del coeficiente de esta regresión ( $R^2$ ) es menor que  $R_u^2$ , se lleva a cabo una nueva época. El proceso de entrenamiento termina cuando  $R^2$  es mayor o igual que  $R_u^2$  (figura IV.1).

La magnitud más empleada en la literatura para medir el error cometido en la fase de generalización de un modelo es el llamado porcentaje de error estándar de predicción o error estándar de predicción (*percent standard error of prediction*, %SEP ó SEP), que es una generalización del error cuadrático medio (**RMS**) y viene dado por:

$$RMS = \sqrt{\frac{\sum_{g=1}^N (d_{kg} - e_{kg})^2}{N}} \quad \%SEP = \frac{100}{\bar{d}_{kg}} RMS \quad (4.19)$$

donde  $d_{kg}$  es la salida deseada  $k$  del patrón  $g$ ;  $e_{kg}$  es la salida estimada por el modelo para el patrón  $g$ ;  $N$  es el número total de patrones del conjunto de generalización y  $\bar{d}_{kg}$  es el valor medio de las salidas deseadas del conjunto de predicción. La principal ventaja de este coeficiente estriba en su adimensionalidad, que lo hace apto para realizar comparaciones entre modelos obtenidos para la resolución de problemas en los que se manejan datos de dimensiones muy distintas (Ventura, 1996).

Otra medida denominada coeficiente de eficiencia (*coefficient of efficiency*, **E**), también se usa en este trabajo para ver como el modelo explica la varianza total de los datos. La magnitud **E** representa la proporción de variación de los datos observados considerada por el modelo, y se expresa como (Kitanidis y Bras, 1980):

$$E = \frac{S_{obs} - S}{S_{obs}} \quad (4.20)$$

$$S_{obs} = \sum_{g=1}^N (d_{kg} - \bar{d}_{kg})^2; \quad S = \sum_{g=1}^N (d_{kg} - e_{kg})^2 \quad (4.21)$$

siendo  $S_{obs}$  la medida de la variabilidad de los valores observados frente a la media y **S** la medida de la asociación entre los valores observados y los estimados. El coeficiente de



eficiencia  $E$  varía entre 1 y  $-\infty$  indicando una buena predicción cuando se encuentre muy próximo a 1.

Los componentes periódicos de las series temporales, especialmente útiles en el análisis de los residuos durante la generalización de los modelos, pueden ser extraídos mediante el análisis espectral de la serie de residuos. Para ello se ha usado una técnica conocida como Transformada de Fourier Rápida o **FFT**. Esta transformación para series de tiempo discreto se define como:

$$F(k) = \sum_{n=0}^{N-1} f(t) \cdot e^{(-jv_k t)} \quad (4.22)$$

donde  $f(t)$  es serie temporal discreta,  $F(k)$  la transformada de Fourier de  $f(t)$ ,  $v$  es la frecuencia y  $N$  es el número de datos (Otnes y Enochson, 1978; Park, 1998). Al tratarse de series de tiempo discretas se ha establecido la frecuencia de barrido de la serie entre los límites 0 y 0.5 para evitar de esta forma efectos de rebote de frecuencias (*aliasing*) (Bloomfield, 1976). La **FFT** proporciona un par de valores para cada frecuencia de onda (coeficientes de Fourier  $-F(k)$ ), que se consideran como un número complejo con un componente cosenoidal (parte real) y otro senoidal (parte imaginaria). Ambos valores pueden combinarse para el cálculo del módulo (4.23) y el periodograma (4.24), aunque la detección de la periodicidad puede llevarse a cabo a través de la visualización de los coeficientes cosenoidales y senoidales en función de la frecuencia (Park, 1998).

$$\text{Módulo} = \sqrt{(\text{Parte real } F(k))^2 + (\text{Parte imaginaria } F(k))^2} \quad (4.23)$$

$$\text{Periodograma} = (\text{Parte real } F(k))^2 + (\text{Parte imaginaria } F(k))^2 \cdot \frac{N}{2} \quad (4.24)$$

Cuando para un determinado parámetro se usan modelos lineales y no lineales puede ser especialmente útil el empleo del coeficiente de desigualdad de Theil junto con los diagramas de predicción-realización para la evaluación de dichos modelos. El coeficiente de desigualdad de Theil ( $U$ ), se construye a partir de las tasas de cambio encontradas entre dos valores consecutivos dentro de una misma serie. Así, para una determinada serie de tasas de cambio observadas ( $y^*$ ) y su correspondiente serie de tasas de cambio estimadas ( $\hat{y}^*$ ), se puede definir el coeficiente de desigualdad  $U$  como:

$$U = \frac{\sqrt{\frac{1}{N} \sum (\hat{y}_i^* - y_i^*)^2}}{\sqrt{\frac{1}{N} \sum (\hat{y}_i^*)^2 + \frac{1}{N} \sum (y_i^*)^2}} \quad (4.25)$$

donde  $N$  es el número de datos contenidos en la serie. El valor de  $U$  está comprendido entre 0 y 1, de tal forma que un valor de  $U$  igual a 0 implica que los valores de predicción coinciden con los reales, mientras que la máxima desigualdad se corresponde con el valor  $U$  igual a 1. Para profundizar en el análisis del coeficiente de desigualdad, nos podemos basar en los resultados obtenidos por Theil (Theil, 1966; Stergiou *et al.*, 1997; Pulido-San-Román y Pérez-García, 2001), que descompone  $U$  en tres tipos de errores: a)  $U_s$  o error sistemático que indica como de parecidas en promedio son las tasas de cambio observadas y estimadas, b)  $U_d$  o error de dispersión que como su nombre indica, muestra la similitud entre las dispersiones en ambas series y c)  $U_c$  o error de correlación que indica la mayor o menor correlación existente entre los valores de predicción y los reales. Se debe cumplir que:

$$U^2 = U_s^2 + U_d^2 + U_c^2 \quad (4.26)$$

donde:

$$U_s = \frac{\bar{y}^* - \bar{y}}{D} \quad (4.27)$$

$$U_d = \frac{S_{\bar{y}^*} - S_{\bar{y}}}{D} \quad (4.28)$$

$$U_c = \frac{\sqrt{2 \cdot (1 - r^*) \cdot S_{\bar{y}^*} \cdot S_{\bar{y}}}}{D} \quad (4.29)$$

$$D = \sqrt{\frac{1}{N} \sum (\bar{y}_i^*)^2} + \sqrt{\frac{1}{N} \sum (\bar{y}_i)^2} \quad (4.30)$$

donde  $\bar{y}^*$  es la media de la serie de tasas de cambio estimadas,  $\bar{y}$  es la media de la serie de tasas de cambio observadas,  $S_{\bar{y}^*}$  es la desviación estándar de la serie de tasas de cambio estimadas,  $S_{\bar{y}}$  es la desviación estándar de la serie de tasas de cambio observadas y  $r^*$  es la correlación existente entre ambas series.

#### 4.2.5. Pretratamiento de la información

La fiabilidad de la predicción mediante las técnicas descritas en los apartados anteriores, dependen en gran medida de la calidad y cantidad de los datos disponibles. Es evidente que, si de un determinado parámetro se dispone de una gran cantidad de datos históricos distribuidos uniformemente a lo largo del tiempo y a lo largo de las unidades de muestreo, la calidad de la predicción será mucho mayor que en el caso de que tan solo se disponga de datos puntuales tomados aleatoriamente en el período de muestreo. Este problema se

agudiza en aquellos parámetros que pueden sufrir cambios drásticos de sus valores absolutos en un período de tiempo inferior al transcurrido entre dos muestreos consecutivos. De esta forma, la serie de datos temporales recoge las imprecisiones, el ruido y los posibles errores en la toma de datos, lo que puede provocar una predicción alejada de la tendencia media de la serie. Esto se reflejará en un bajo nivel de ajuste entre los valores estimados y observados y, por lo tanto, en unos errores de predicción alejados de los rangos de aceptación.

Muchos tipos de ruidos se muestran como cambios rápidos de amplitud dentro de la serie. En estos casos, es frecuente tratar de reducir el error de la serie utilizando un suavizado, de tal forma que los valores puntuales de la serie son modificados. Así los valores más altos que los inmediatamente adyacentes (supuestamente a causa del ruido) son reducidos, mientras que los valores más bajos son incrementados.

Para evitar dentro de lo posible este problema, se ha llevado a cabo un proceso de pretratamiento de la información contenida en la serie, en aquellos casos en los que los errores de predicción de todos los modelos empleados fue excesivamente alto. El pretratamiento ha consistido en un proceso de filtrado o alisamiento de tipo rectangular o media desplazante no ponderada empleado por otros autores (Coulbeck, 1988; Cabrera y Vela, 1994) que para un determinado incremento de tiempo viene determinado como:

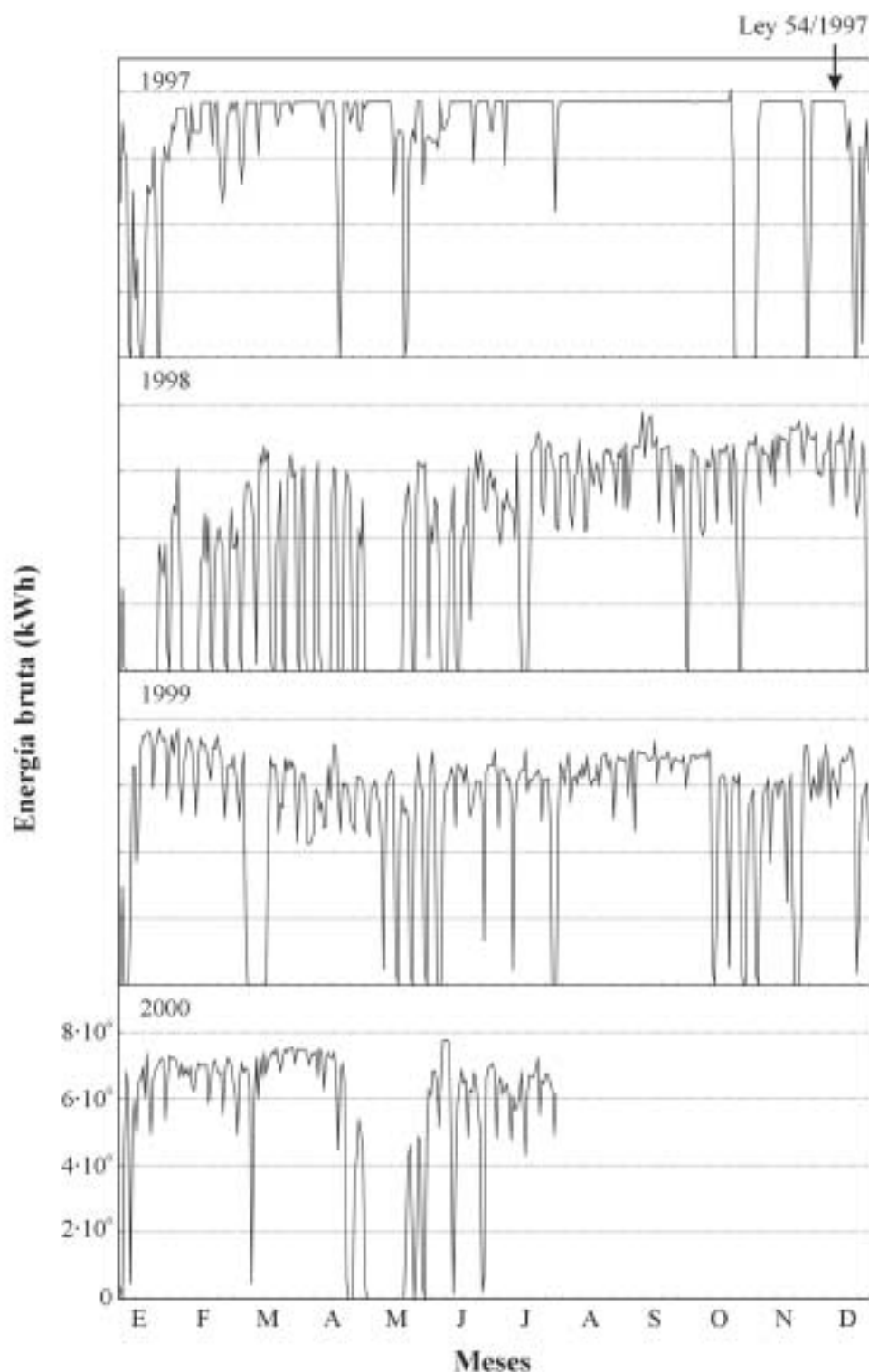
$$F_t = \frac{\sum_{q=a}^z V_q}{\Delta t + 1}; \quad z = a + \Delta t; \quad t = z - \frac{\Delta t}{2} \quad (4.31)$$

donde  $F_t$  es el valor de  $V_q$  filtrado para  $q=t$  y  $\Delta t$  es el rango de tiempo considerado.

## 4.3. RESULTADOS

### 4.3.1. Variables de la Central Térmica de Puente Nuevo. Análisis de la producción bruta

La variación temporal de la energía bruta generada por la Central Térmica de Puente Nuevo durante el período comprendido entre enero de 1997 y julio de 2000 se muestra en la figura IV.2. Se puede observar como la producción de energía durante el año 1997 fluctúa ligeramente en los primeros meses del año, aunque a partir de principios de febrero se estabiliza existiendo a partir de entonces tan solo tres paradas técnicas separadas por largos períodos en los que la energía producida se encuentra cercana a la máxima posible. A partir de la entrada en vigor de la Ley 54/1997 del Sector Eléctrico (BOE 28-11-1997) a finales de noviembre de 1997, la producción de energía desciende considerablemente. En los primeros meses de 1998 la producción media se estabiliza en los  $2.1 \cdot 10^6 \pm 2.5 \cdot 10^6$  kWh, produciéndose a partir de ese momento un incremento constante que lleva la producción hasta sus valores máximos anuales y, que continúa en los años 1999 y 2000.



**Figura IV.2.** Energía bruta producida por la Central Térmica de Puente Nuevo durante los años 1997, 1998, 1999 y 2000. Se indica el comienzo de la aplicación de la Ley 54/1997.

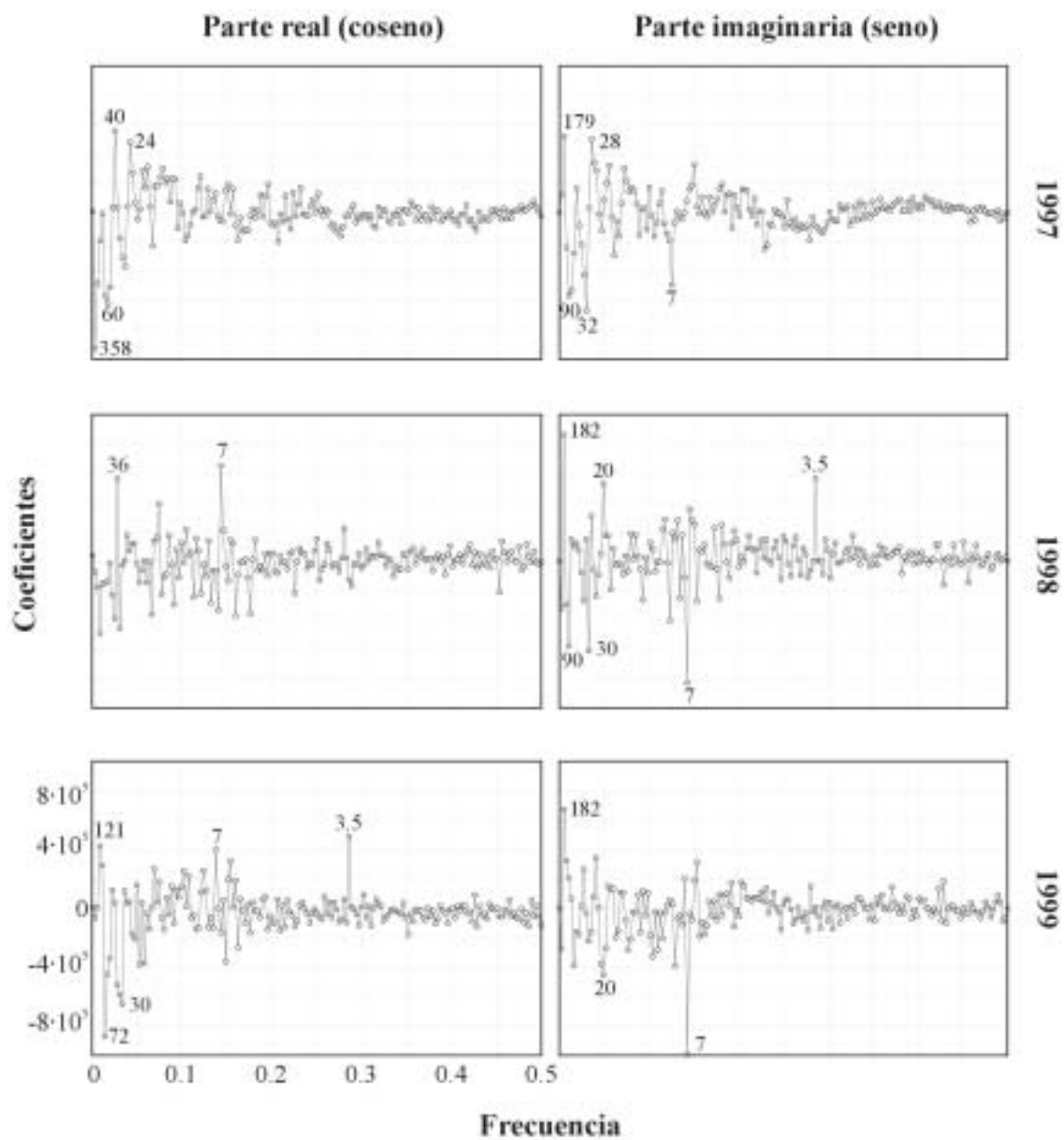
La transformada de Fourier de los datos muestra la presencia de coeficientes con valores absolutos muy alejados del valor cero, lo que indica la fuerte periodicidad de la producción. Se puede decir que en el año 1997, el ciclo de producción tiene un carácter mensual al encontrarse frecuencias con grandes coeficientes que se corresponden con períodos de 60 (2 meses), 40 y 24 (aproximadamente un mes) días en la parte real y con períodos de 28, 32 (aproximadamente un mes), 90 (tres meses) y 179 (6 meses) días en la parte imaginaria. En este año se encuentra un coeficiente importante relacionado con la frecuencia 0.143 o período de 7 días (semanal) en la parte imaginaria. Este período no se detecta en la parte real por lo que tiene menor importancia que los anteriormente citados. En cambio, en años sucesivos se detecta, además de una estacionalidad mensual, una importante periodicidad semanal al alcanzar la frecuencia 0.143 los coeficiente más grandes en la parte real e imaginaria del año 1998 y en la parte imaginaria de 1999 (figura IV.3).

Un análisis de la varianza ha mostrado la presencia de diferencias significativas entre el nivel de producción llevado a cabo en cada uno de los siete días de la semana para el período de producción dentro de la Ley 54/1997 (años 1998, 1999 y 2000) (ANOVA:  $F=15.451$ ,  $p_{\alpha}<0.001$ ) (figura IV.4a). Un test de Tukey posterior mostró que los dos días en los que los niveles de producción disminuían significativamente eran el sábado y el domingo (Tukey: en todos los casos nivel de significación entre lunes-martes-miércoles-jueves-viernes, sábado y domingo respectivamente:  $p_{\alpha\text{sábado}}<0.05$ ;  $p_{\alpha\text{domingo}}<0.001$ ). No se encontraron diferencias significativas entre ambos días. (Tukey:  $p_{\alpha\text{sábado-domingo}}=0.052$ ). Esta periodicidad semanal no fue detectada en el año 1997 antes de la implantación de la Ley 54/1997 (ANOVA:  $F=0.606$ ,  $p_{\alpha}=0.999$ ) (figura IV.4a). Un test  $\chi^2$  sobre el número de días a un determinado nivel de producción antes y después de la aplicación de la nueva Ley mostró la presencia de marcadas diferencias significativas ( $\chi^2=1772.181$ ,  $g.l.=7$ ,  $p_{\alpha}<0.001$ ) (figura IV.4b). Esto indica que la entrada en vigor de la nueva ley no sólo ha influido en la forma y momento de distribución de la energía si no que también ha sido responsable de la disminución de la cantidad de energía generada en la Central Térmica de Puente Nuevo.

## 4.3.2. Variables de planta

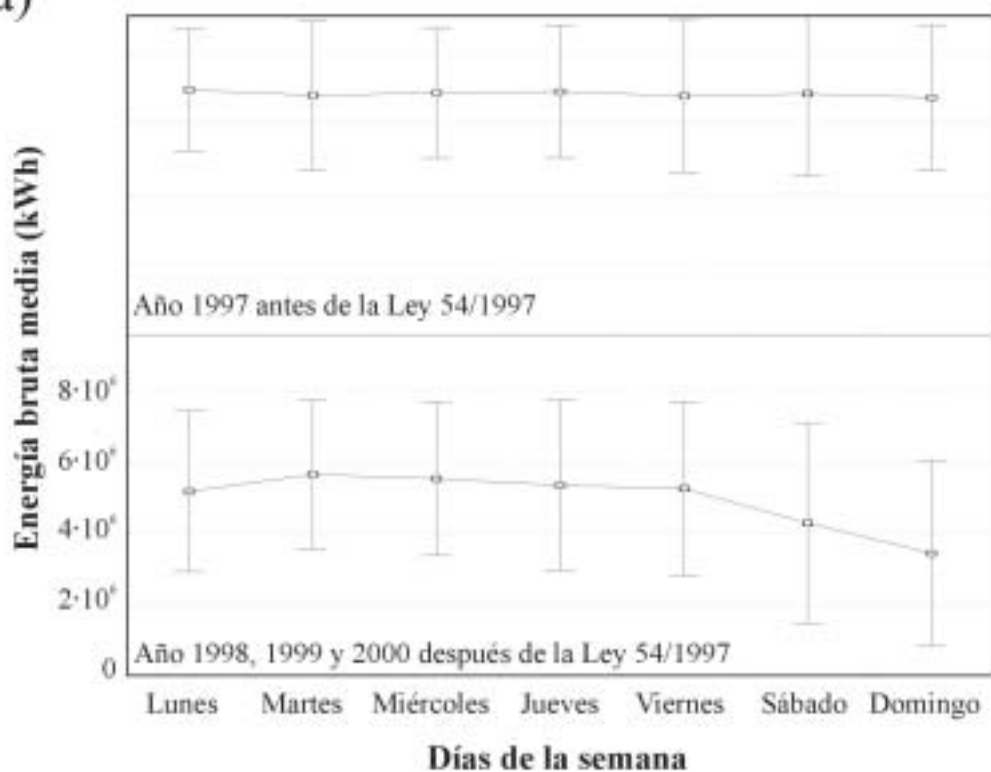
### 4.3.2.1. Temperatura en la serie 'A'

La temperatura media anual del agua en la serie 'A' ha variado significativamente a lo largo de los años en estudio (ANOVA:  $F=205.671$ ,  $p_{\alpha}<0.001$ ). De forma global, la temperatura alcanzó el valor medio más alto en el año 1997 ( $24.51 \pm 2.33$  °C) y el valor más bajo en el año 2001 ( $14.96 \pm 1.36$  °C). Asimismo, al considerar las series de datos existentes para los años 1997 y 2001 se han encontrado diferencias significativas cuando se comparan las temperaturas medias de los tres primeros meses del año (ANOVA:  $F=438.555$ ,  $p_{\alpha}<0.001$ ), y de los seis últimos meses del año (ANOVA:  $F=15.320$ ,  $p_{\alpha}<0.001$ ). En el primero de los casos se observa una fuerte disminución de la temperatura media desde los  $21.50 \pm 2.62$  °C

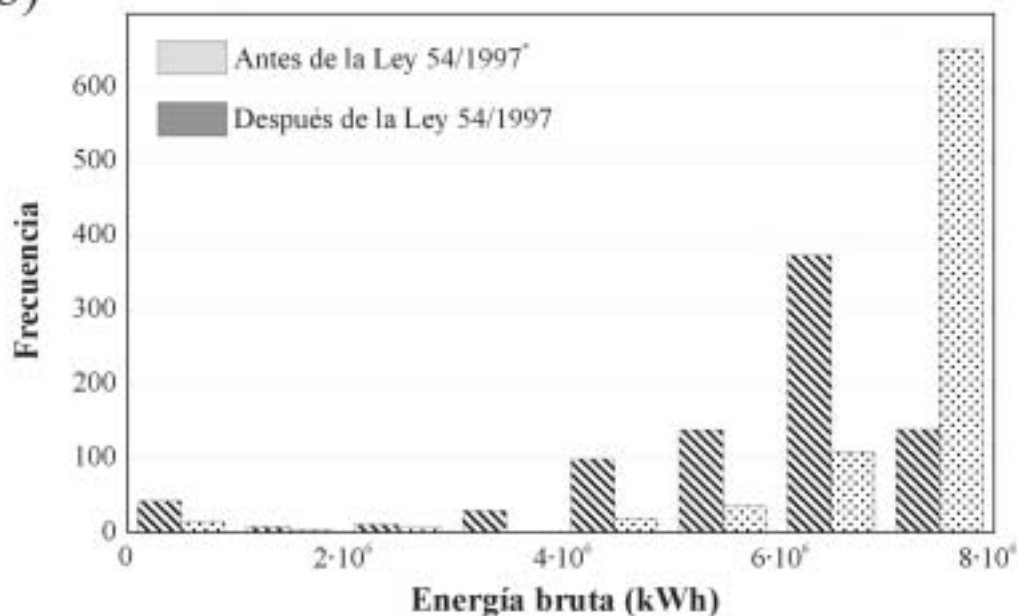


**Figura IV.3.** Frecuencia de las funciones coseno y seno frente a los coeficientes encontrados en la transformada de Fourier de las series de producción bruta en la Central Térmica de Puente Nuevo. Los números indican periodos (días) cuyos coeficientes tienen un gran peso dentro de la función transformada

a)

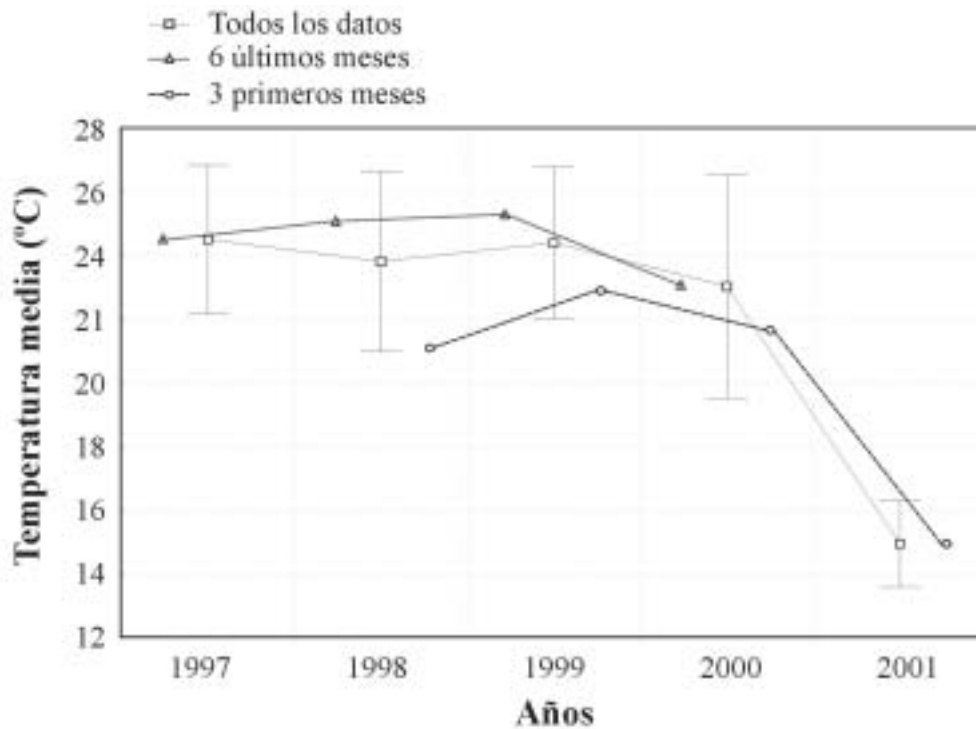


b)



**Figura IV.4.** a) Energía bruta media por día de la semana en el año 1997 antes de la implantación de la Ley 54/1997 y energía bruta media por día de la semana para los años 1998, 1999 y 2000. b) Distribución de frecuencias del número de días para distintos niveles de producción. Datos escalados para permitir la comparación

alcanzados en el 2000 hasta los  $14.96 \pm 1.36$  °C en el 2001. En el segundo caso se observa una tendencia a la baja en el año 2000 después de un incremento constante de la temperatura media desde el año 1997 hasta el año 1999 (figura IV.5).



**Figura IV.5.** *Temperatura media anual del agua durante el periodo de estudio considerando la totalidad de los datos disponibles (línea continua). La línea punteada indica la evolución de la temperatura media en los seis últimos meses de cada año, mientras que la línea rayada hace referencia a la evolución de la temperatura media en los tres primeros meses del año. En estos dos últimos casos se han omitido las desviaciones estándar por claridad.*

El análisis de las temperaturas medias del agua por meses muestra un patrón de evolución paralelo al experimentado por las temperaturas ambientales medias para el mismo período. Durante los meses más fríos (otoño-invierno y principio de la primavera) la temperatura media del agua en la serie 'A' se encuentra en todo momento por encima de las temperaturas medias ambientales. Estas diferencias se reducen paulatimamente a medida que las temperaturas ambientales aumentan (final de la primavera y verano), de tal forma que durante este período éstas tienden a coincidir con las temperaturas medias del agua (figura IV.6). Se han hallado grandes correlaciones significativas en los años 1997, 1999 y 2000, y correlaciones significativas inferiores en el año 1998 entre la temperatura del agua en la serie 'A' y la temperatura ambiental (1997:  $R=0.8780$ ;  $R^2=0.7709$ ;  $F(1;186)=625.950$ ;



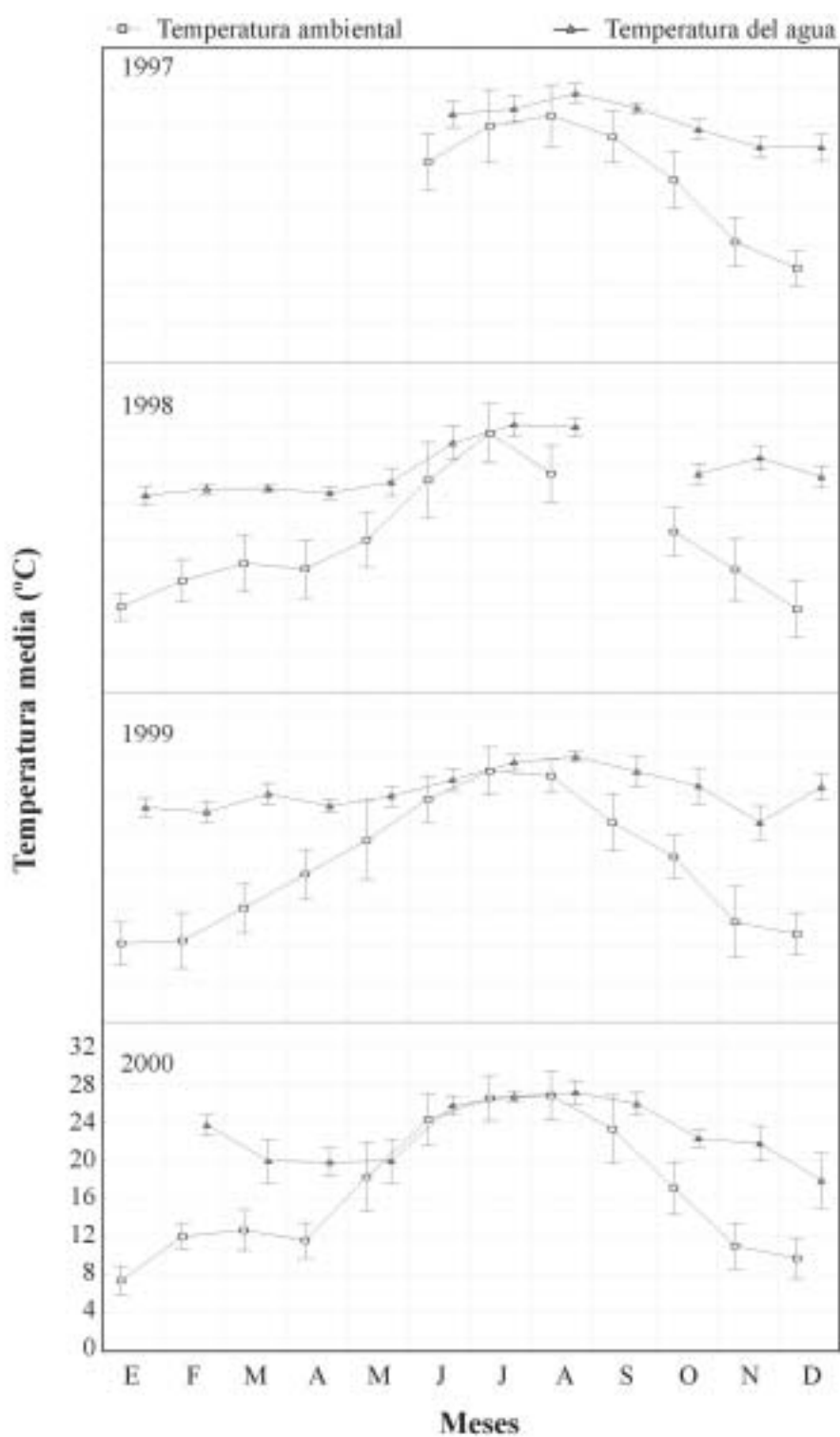


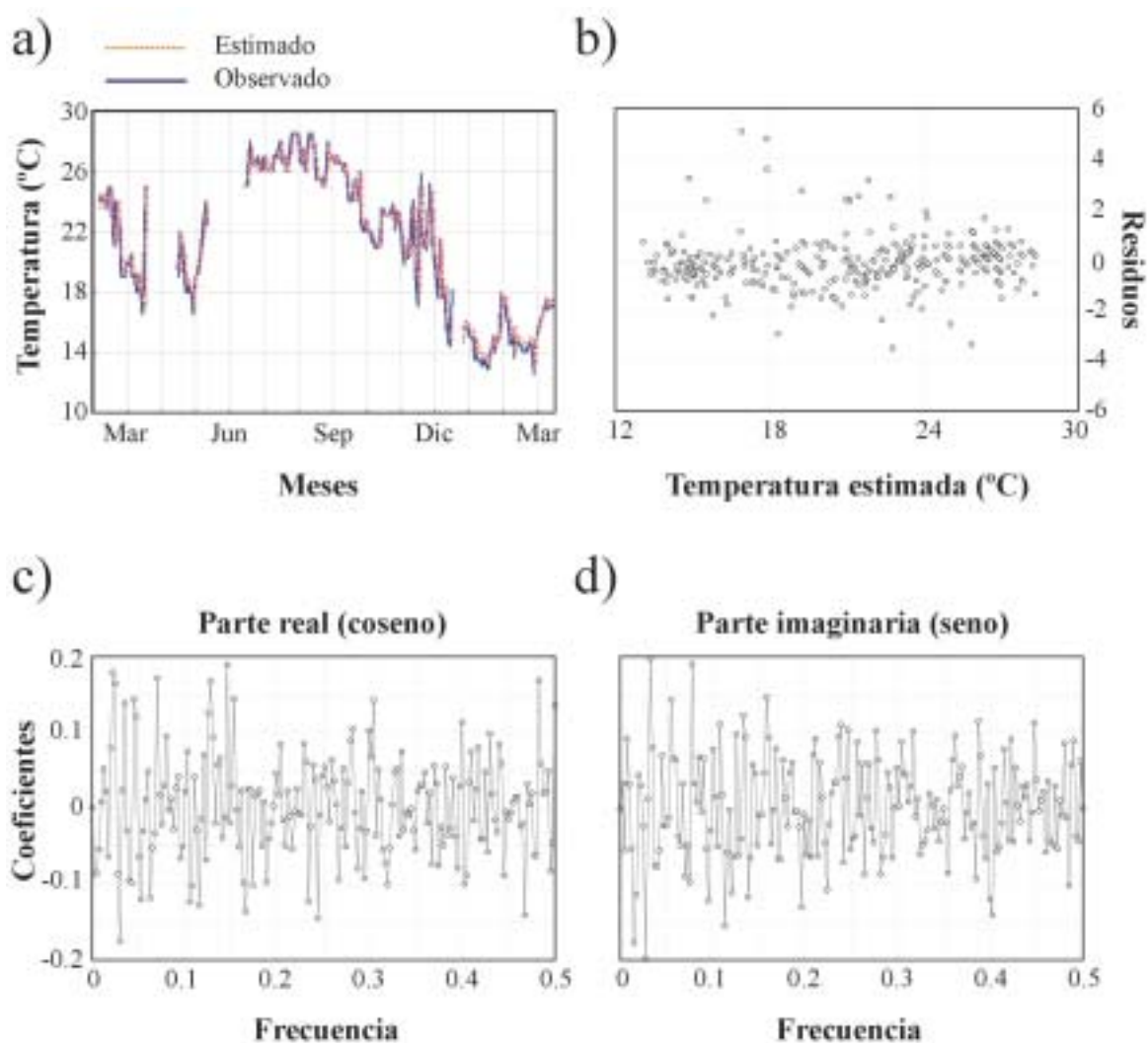
Figura IV.6. Temperaturas medias  $\pm$  desviaciones típicas del agua en la serie 'A' y en el ambiente durante el periodo de estudio

$p_{\alpha} < 0.001$ ;  $N=188$  - 1998:  $R=0.4438$ ;  $R^2=0.1970$ ;  $F(1;354)=86.820$ ;  $p_{\alpha} < 0.001$ ;  $N=356$  - 1999:  $R=0.7146$ ;  $R^2=0.5106$ ;  $F(1;344)=358.930$ ;  $p_{\alpha} < 0.001$ ;  $N=346$  - 2000:  $R=0.7845$ ;  $R^2=0.6154$ ;  $F(1;186)=436.830$ ;  $p_{\alpha} < 0.001$ ;  $N=275$ ). La menor correlación de la temperatura ambiental en el año 1998 parece estar relacionada con una mayor dependencia del nivel de producción de la central térmica. Así, la mayor correlación significativa entre la temperatura del agua en la serie 'A' y el nivel de producción de la central térmica ha sido encontrada en este año (1998:  $R=0.4888$ ;  $R^2=0.2389$ ;  $F(1;354)=111.120$ ;  $p_{\alpha} < 0.001$ ;  $N=356$ ), mientras que para los años 1997, 1999 y 2000 las correlaciones han sido inferiores aunque significativas (1997:  $R=0.3784$ ;  $R^2=0.1432$ ;  $F(1;186)=31.090$ ;  $p_{\alpha} < 0.001$ ;  $N=188$  - 1999:  $R=0.2166$ ;  $R^2=0.0469$ ;  $F(1;344)=16.940$ ;  $p_{\alpha} < 0.001$ ;  $N=346$  - 2000:  $R=0.3627$ ;  $R^2=0.1316$ ;  $F(1;124)=18.790$ ;  $p_{\alpha} < 0.001$ ;  $N=126$ ).

El análisis por días de la semana ha dado como resultado la ausencia total de diferencias significativas en todos los años (ANOVA 1997:  $F=0.123$ ;  $p_{\alpha}=0.9934$  - ANOVA 1998:  $F=0.028$ ;  $p_{\alpha}=0.9999$  - ANOVA 1999:  $F=0.201$ ;  $p_{\alpha}=0.9762$  - ANOVA 2000:  $F=0.078$ ;  $p_{\alpha}=0.9981$ ).

La regresión múltiple, con variable dependiente la temperatura del agua en el sistema 'A' y con variables independientes la temperatura del agua en el mismo sistema en los cinco días anteriores, ha proporcionado resultados estadísticamente satisfactorios en el proceso de calibración ( $R^2=0.9078$ ;  $F(5;869)=1712.200$ ;  $p_{\alpha} < 0.001$ ;  $N=875$ ). Durante el proceso de validación (temperatura del agua estimada frente a la temperatura del agua observada) también se obtienen buenos resultados al alcanzarse un coeficiente de correlación superior a 0.97 ( $R^2=0.9561$ ;  $F(1;131)=7204.900$ ;  $p_{\alpha} < 0.001$ ;  $N=333$ ) (figura IV.7). La validación del modelo se ve corroborada por un error estándar de predicción inferior al 5 % y por un coeficiente de error **E** de 0.9551 (tabla IV.2).

El análisis de los residuos existentes entre los valores observados y estimados frente a valores estimados durante la validación ha dado como resultado una nube de puntos compacta centrada en el cero, lo que indica la falta de autocorrelación entre residuos. Asimismo, el análisis espectral de la serie temporal de residuos muestra la ausencia de frecuencias significativamente importantes por lo que no se detecta periodicidad en la serie (figura IV.7).



**Figura IV.7.** a) Temperatura del agua observada y estimada del sistema 'A' en el periodo de validación (años 2000 y 2001) considerando en el análisis de regresión múltiple como variable dependiente la temperatura del agua y como variables independientes la temperatura del agua de los cinco días anteriores; b) Residuos (temperatura observada menos temperatura estimada) en el grupo de validación frente a la temperatura estimada; c) y d) Frecuencia de las funciones coseno y seno frente a los coeficientes encontrados en la transformada de Fourier

**Tabla IV.2.** Estimación de la temperatura del agua en la serie 'A' mediante el análisis de regresión múltiple con las variables independientes representadas por la temperatura del agua en los cinco días anteriores

<b>Resumen de la regresión múltiple en la calibración del modelo</b>			
R=0.9525; R <sup>2</sup> =0.9078; F(5;869)=1712.200; p <sub>α</sub> <0.001; N=875			
<b>Variable dependiente</b>	<b>Variable independiente</b>	<b>b<sub>i</sub>(i=0,1,...,q)</b>	<b>p<sub>α</sub></b>
Temperatura del agua (t)	Intersección (b <sub>0</sub> )	0.9083	0.0005**
	Temperatura del agua (t-1)	0.8964	0**
	Temperatura del agua (t-2)	0.0467	0.3049
	Temperatura del agua (t-3)	-0.0975	0.0322*
	Temperatura del agua (t-4)	0.0359	0.4300
	Temperatura del agua (t-5)	0.0810	0.0165*
<b>Resumen de la regresión en la validación del modelo</b>			
R=0.9778; R <sup>2</sup> =0.9561; F(1;331)=7204.900; p <sub>α</sub> <0.001; N=333			
Error estándar de predicción=4.6372 %; Coeficiente E=0.9551			
<b>Variable dependiente</b>	<b>Variable independiente</b>	<b>b<sub>i</sub>(i=0,1,...,q)</b>	<b>p<sub>α</sub></b>
Temperatura del agua observada (t)	Intersección (b <sub>0</sub> )	-0.5320	0.0434*
	Temperatura del agua estimada (t)	1.0193	0**

\* p<sub>α</sub><0.05; \*\* p<sub>α</sub><0.01

En la regresión anterior se puede observar como de las cinco variables independientes, tres son estadísticamente significativas (Temperatura del agua (t-1), (t-3) y (t-5)) (tabla IV.2). Éstas explican de forma significativa una mayor parte de la varianza de la variable dependiente, por lo que se repite el análisis de regresión utilizando como variables independientes las temperaturas del agua del primer, tercer y quinto días anteriores. Los resultados son similares al obtenerse el mismo coeficiente de correlación y magnitudes de error ligeramente mejores (R<sup>2</sup>=0.9561; F(1;131)=7212.100; p<sub>α</sub><0.001; N=333; SEP=4.6362 %; Coeficiente E=0.9551).

Los resultados también son estadísticamente aceptables cuando la estimación de la temperatura se lleva a cabo tan solo en función de una única variable independiente. De esta forma, cuando la variable independiente es la temperatura del día anterior se observa una ligera disminución del coeficiente de correlación para la serie de validación así como un aumento del error (R<sup>2</sup>=0.9542; F(1;131)=6892.400; p<sub>α</sub><0.001; N=333; SEP=4.7580 %; Coeficiente E=0.9527). En estas condiciones, la peor de las predicciones se alcanza cuando se utiliza la temperatura del agua de cinco días antes (R<sup>2</sup>=0.8361; F(1;131)=1688.500; p<sub>α</sub><0.001; N=333; SEP=9.5413 %; Coeficiente E=0.8098).

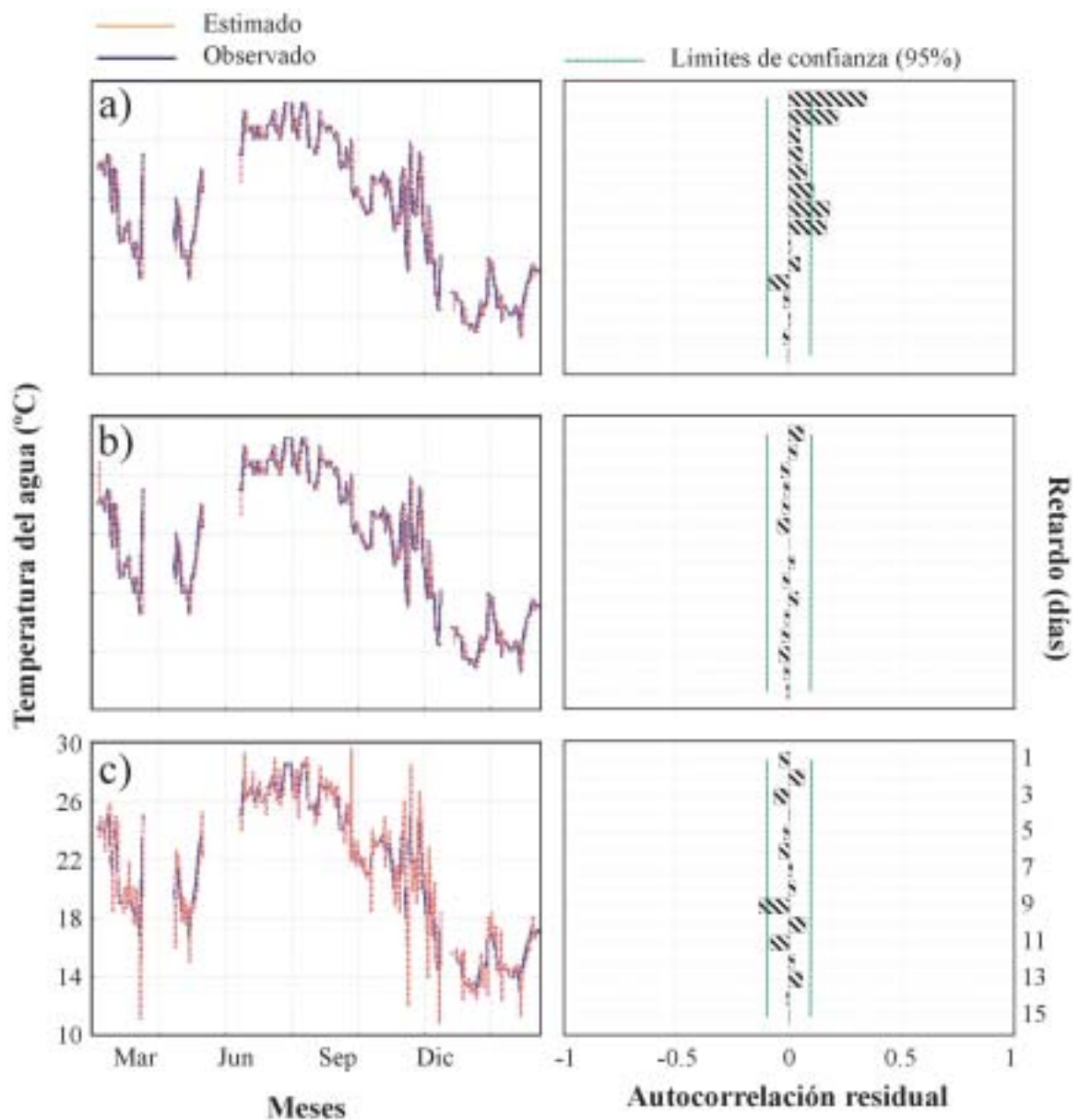
Los parámetros de estimación de la temperatura del agua a través del suavizado exponencial simple, suavizado lineal exponencial y modelos **ARIMA** se muestran en la tabla IV.3.

**Tabla IV.3.** *Parámetros de estimación y bondad de ajuste de los modelos de series temporales en la temperatura del agua de la serie 'A'*

Modelo	Parámetros	Validación			
		R <sup>2</sup>	F	%SEP	Coef. E
Suavizado exponencial simple	S <sub>0</sub> =24.1300; α <sub>d</sub> =0.9210	0.9528	(1;328)=6621.4	4.7658	0.9526
Suavizado lineal exponencial	S <sub>0</sub> =27; T <sub>0</sub> =-0.0040 α <sub>d</sub> =0.9290; β <sub>d</sub> =0	0.9520	(1;328)=6500.4	4.8039	0.9518
ARIMA (0,0,1)	Θ <sub>1</sub> =-0.9518**	0.4321	(1;330)=251.0	53.8141	-5.1696
ARIMA (1,0,0)	Φ <sub>1</sub> =0.9984**	0.9584	(1;330)=7599.5	4.6930	0.9531
ARIMA (0,1,1)	Θ <sub>1</sub> =-0.0486	0.9585	(1;330)=7620.1	4.6966	0.9530
ARIMA (1,0,1)	Φ <sub>1</sub> =0.9982**; Θ <sub>1</sub> =-0.0323	0.9584	(1;330)=7596.5	4.6930	0.9531
ARIMA (1,1,0)	Φ <sub>1</sub> =0.0488	0.9585	(1;330)=7621.4	4.6967	0.9530
ARIMA (1,1,1)	Φ <sub>1</sub> =0.0202; Θ <sub>1</sub> =-0.0289	0.9585	(1;330)=7620.9	4.6966	0.9530
ARIMA (1,1,2)	Φ <sub>1</sub> =0.7311**; Θ <sub>1</sub> =0.7287** Θ <sub>2</sub> =0.1126*	0.9589	(1;329)=7670.4	4.6139	0.9546
ARIMA (1,2,1)	Φ <sub>1</sub> =0.0999; Θ <sub>1</sub> =-0.9698**	0.9190	(1;326)=3699.4	6.8289	0.9006
ARIMA (2,2,1)	Φ <sub>1</sub> =0.1604**; Θ <sub>1</sub> =-0.9487** Φ <sub>2</sub> =0.4354**	0.9349	(1;327)=4695.0	6.0936	0.9209
ARIMA (2,1,2)	Φ <sub>1</sub> =1.2954**; Θ <sub>1</sub> =-1.2536** Φ <sub>2</sub> =-0.7736**; Θ <sub>1</sub> =-0.6386**	0.9604	(1;330)=8006.1	4.5407	0.9561
ARIMA (2,1,1)	Φ <sub>1</sub> =0.8285**; Θ <sub>1</sub> =-0.8001** Φ <sub>2</sub> =-0.1490**	0.9589	(1;330)=7702.7	4.6111	0.9547
ARIMA (1,2,2)	Φ <sub>1</sub> =0.3798; Θ <sub>1</sub> =0.0218 Θ <sub>2</sub> =0.9782	0.9327	(1;327)=4534.1	5.9960	0.9234
ARIMA (2,2,2)	Φ <sub>1</sub> =0.5634**; Θ <sub>1</sub> =0.0292 Φ <sub>2</sub> =-0.1388**; Θ <sub>2</sub> =0.9706**	0.9410	(1;327)=5219.9	5.6036	0.9331

\* p<sub>α</sub><0.05; \*\* p<sub>α</sub><0.01

En el modelo de suavizado exponencial simple, lineal exponencial y todos los modelos **ARIMA** excepto en el ARIMA (0,0,1), el porcentaje de varianza explicada supera el 91 % mientras que los términos de error se encuentran alrededor del 5 % (**SEP**) y 0.93 (coeficiente **E**). De todos ellos, el que presenta una mayor bondad de ajuste es el ARIMA (2,1,2) donde el coeficiente de determinación es de 0.9561 y el error es del 4.5407 %. De entre los dos modelos de suavizado y el ARIMA (2,1,2), tan solo el suavizado lineal exponencial presenta una ausencia total de significación en todos los coeficientes de correlación de los errores de predicción, aunque el término de error es ligeramente superior al modelo ARIMA (2,1,2). Esto puede ser debido al desfase en los valores observados y estimados perceptible tanto en el suavizado lineal exponencial como el exponencial simple. Este desfase no se observa en el modelo ARIMA (2,1,2). Sin embargo en este último existe significación estadística del noveno coeficiente de autocorrelación del error de predicción, lo que probablemente viene determinado por la sobreestimación de los valores extremos (figura IV.8).



**Figura IV.8.** Temperatura del agua observada y estimada en la serie 'A' y autocorrelación residual en validación para los modelos: a) suavizado exponencial simple; b) suavizado lineal exponencial (Holt) c) ARIMA (2,1,2)

#### 4.3.2.2. Temperatura en la serie 'B'

La evolución temporal de la temperatura en la serie 'B' y en la serie 'A' se muestran en la figura IV.9. Se puede observar un comportamiento similar de dicho parámetro en ambas series. De esta forma la regresión lineal simple en la que la temperatura de la serie 'B' es

estimada a través de la temperatura en la serie 'A' ha ofrecido muy buenos resultados tanto en el proceso de calibración como en el de validación (tabla IV.4).

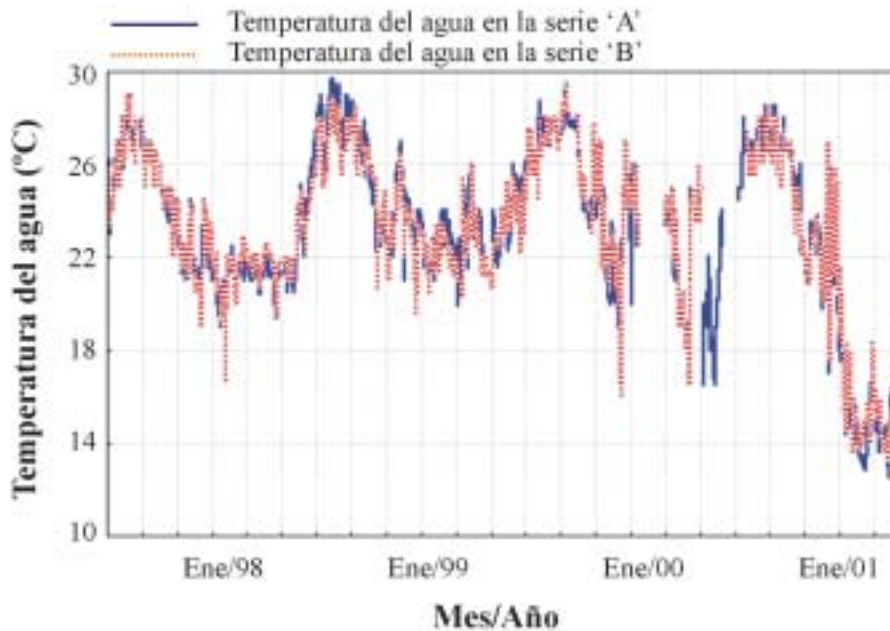


Figura IV.9. Evolución temporal de la temperatura del agua en las series 'A' y 'B'

**Tabla IV.4.** Estimación de la temperatura del agua en la serie 'B' a través de la temperatura del agua en la serie 'A'

<b>Resumen de la regresión en la calibración del modelo</b>			
R=0.9711; R <sup>2</sup> =0.9431; F(1;888)=14711.000; p <sub>α</sub> <0.001; N=890			
Variable dependiente	Variable independiente	b <sub>0</sub> , b <sub>1</sub>	p <sub>α</sub>
Temperatura del agua en 'B'	Intersección (b <sub>0</sub> )	1.7849	0**
	Temperatura del agua en 'A' (b <sub>1</sub> )	0.9249	0**
<b>Resumen de la regresión en la validación del modelo</b>			
R=0.9907; R <sup>2</sup> =0.9815; F(1;305)=16146.000; p <sub>α</sub> <0.001; N=307			
Error estándar de predicción=4.7095 %; Coeficiente E=0.9788			
Variable dependiente	Variable independiente	b <sub>0</sub> , b <sub>1</sub>	p <sub>α</sub>
Temperatura del agua en 'B' observada	Intersección (b <sub>0</sub> )	-1.0063	0**
	Temperatura del agua en 'B' estimada (b <sub>1</sub> )	1.0392	0**

\*\* p<sub>α</sub><0.01

Ya que existe una alta correlación significativa entre la temperatura del agua en las series 'A' y 'B', se han utilizado los mismos parámetros de regresión y de los modelos de series temporales en el proceso de predicción de la temperatura en la serie 'B' para el mismo período de validación. Así, en el caso de las regresiones múltiples se han obtenido en todos las ocasiones coeficientes de correlacion superiores a 0.9 y magnitudes de error por debajo del 9.9 % (tabla IV.5). De la misma forma que en el proceso de validación de la estimación de la temperatura del agua en la serie 'A', el modelo de regresión con el que se han obtenido los mejores resultados se corresponde con el que utiliza como variables independientes las temperaturas del agua del primer, tercero y quinto día anteriores ( $R^2=0.9531$ ;  $F(1;308)=6264.200$ ;  $p_\alpha<0.001$ ;  $N=310$ ;  $SEP=4.6886$  %; Coeficiente  $E=0.9519$ ). Asimismo, no se observan ni frecuencias significativas tras la aplicación de la transformada de Fourier a los residuos de validación ni ningún tipo de tendencia de los residuos frente a los valores estimados (figura IV.10).

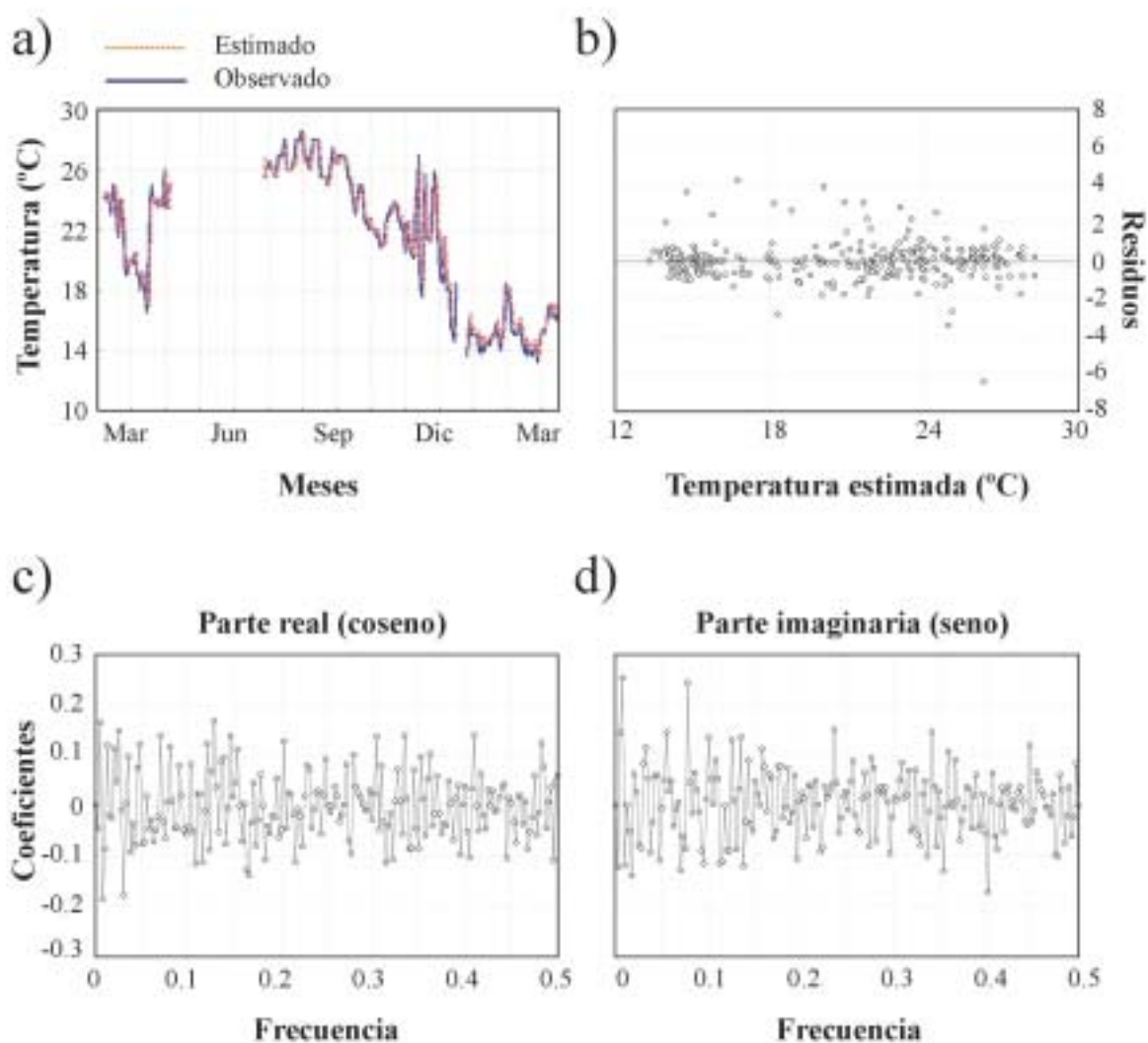
**Tabla IV.5.** *Parámetros de validación en la estimación de la temperatura del agua en la serie 'B' a través de los coeficientes de regresión obtenidos en el proceso de calibración con los datos de temperatura del agua en la serie 'A'*

Modelo en calibración	$R^2$ Validación	F	$p_\alpha$	% SEP	Coef. E
Temperatura de los 5 días anteriores	0.9526	(1;308)=6190.9	0**	4.7125	0.9514
Temperatura de 1 <sup>er</sup> , 3 <sup>er</sup> y 5 <sup>o</sup> días anteriores	0.9531	(1;308)=6264.2	0**	4.6886	0.9519
Temperatura del día anterior	0.9504	(1;308)=5901.3	0**	4.8372	0.9488
Temperatura del quinto día anterior	0.8164	(1;308)=1369.5	0**	9.8070	0.7897

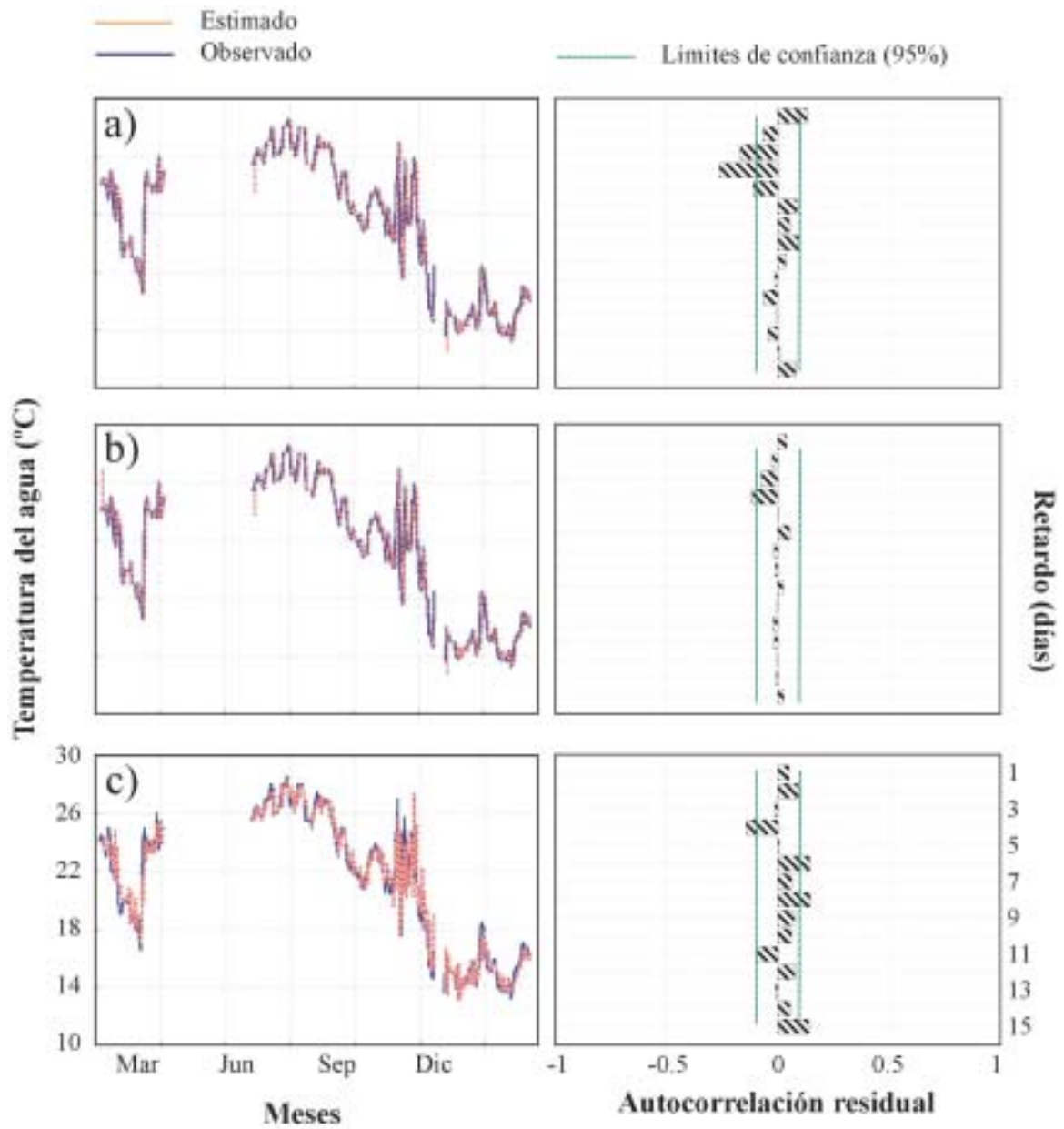
\*\*  $p_\alpha<0.01$

Las estimaciones de la temperatura del agua en la serie 'B' a partir de los parámetros obtenidos a través del suavizado exponencial simple y el suavizado lineal exponencial de las temperaturas de la serie 'A', también ofrecen ajustes estadísticamente aceptables y magnitudes de error pequeñas (Suavizado exponencial simple:  $R^2=0.9478$ ; % SEP=4.8939; Coeficiente  $E=0.9476$  – Suavizado lineal exponencial:  $R^2=0.9469$ ; % SEP=4.9321; Coeficiente  $E=0.9468$ ) (figura IV.11) (tabla IV.6). De forma general, los modelos **ARIMA** proporcionan coeficientes de determinación superiores a 0.87 y errores cercanos al 5 %, siendo en esta ocasion el modelo ARIMA (1,1,2) el que ofrece mejores resultados ( $R^2=0.9545$ ;  $F(1,305)=6391.600$ ;  $p_\alpha<0.001$ ;  $N=307$ ; %SEP=4.5644; Coef.  $E=0.9544$ ) (figura IV.11) (tabla IV.6). Tan solo en el caso del modelo ARIMA (0,0,1) se obtiene un error estándar de predicción superior al 53 % y la varianza explicada de la variable dependiente no supera el 39 %.





**Figura IV.10.** a) Temperatura del agua observada y estimada del sistema "B" en el periodo de validación, considerando en el análisis de regresión múltiple como variable dependiente la temperatura del agua y como variables independientes las temperaturas del agua del primer, tercer y quinto días anteriores; b) Residuos (temperatura observada menos temperatura estimada) en el grupo de validación frente a la temperatura estimada; c) y d) Frecuencia de las funciones coseno y seno frente a los coeficientes encontrados en la transformada de Fourier



**Figura IV.11.** Temperatura del agua observada y estimada en la serie 'B' y autocorrelación residual en validación (años 2000 y 2001) para los modelos: a) suavizado exponencial simple; b) suavizado lineal exponencial (Holt); c) ARIMA (1,1,2)

**Tabla IV.6.** *Parámetros de estimación y bondad de ajuste de los modelos univariantes de series temporales de la temperatura del agua de la serie 'B'*

Modelo	Parámetros	Validación			
		R <sup>2</sup>	F	%SEP	Coef. E
Suavizado exponencial simple	S <sub>0</sub> =24.1300; α <sub>d</sub> =0.9210	0.9478	(1;306)=5551.0	4.8939	0.9476
Suavizado lineal exponencial	S <sub>0</sub> =27; T <sub>0</sub> =-0.0040 α <sub>d</sub> =0.9290; β <sub>d</sub> =0	0.9469	(1;306)=5455.1	4.9322	0.9468
ARIMA (0,0,1)	Θ <sub>1</sub> =-0.9533**	0.3844	(1;308)=192.3	53.6644	-5.2974
ARIMA (1,0,0)	Φ <sub>1</sub> =0.9984**	0.8716	(1;308)=2090.4	4.7799	0.9500
ARIMA (0,1,1)	Θ <sub>1</sub> =-0.0669	0.9509	(1;307)=5951.5	4.7732	0.9502
ARIMA (1,0,1)	Φ <sub>1</sub> =0.9982**; Θ <sub>1</sub> =-0.0374	0.8720	(1;308)=2097.9	4.7708	0.9502
ARIMA (1,1,0)	Φ <sub>1</sub> =0.0594	0.9509	(1;307)=5947.7	4.7745	0.9501
ARIMA (1,1,1)	Φ <sub>1</sub> =-0.3169; Θ <sub>1</sub> =-0.3835	0.9527	(1;305)=6141.4	4.6632	0.9524
ARIMA (1,1,2)	Φ <sub>1</sub> =0.6244**; Θ <sub>1</sub> =0.6269** Θ <sub>2</sub> =0.1685**	0.9545	(1;305)=6391.6	4.5644	0.9544
ARIMA (1,2,1)	Φ <sub>1</sub> =0.0746; Θ <sub>1</sub> =-0.9710**	0.9013	(1;303)=2766.3	7.0709	0.8907
ARIMA (2,2,1)	Φ <sub>1</sub> =0.1196*; Θ <sub>1</sub> =-0.9535** Φ <sub>2</sub> =-0.4033**	0.9137	(1;303)=3206.7	6.5185	0.9071
ARIMA (2,1,2)	Φ <sub>1</sub> =1.2595**; Θ <sub>1</sub> =1.2394** Φ <sub>2</sub> =-0.6652**; Θ <sub>1</sub> =-0.5028**	0.9545	(1;306)=6416.0	4.5711	0.9543
ARIMA (2,1,1)	Φ <sub>1</sub> =0.8067*; Θ <sub>1</sub> =0.7826** Φ <sub>2</sub> =-0.1990**	0.9533	(1;306)=6246.2	4.6279	0.9532
ARIMA (1,2,2)	Φ <sub>1</sub> =0.4099**; Θ <sub>1</sub> =0.0250 Θ <sub>2</sub> =0.9749**	0.9194	(1;303)=3456.5	6.1542	0.9172
ARIMA (2,2,2)	Φ <sub>1</sub> =0.5691; Θ <sub>1</sub> =0.0246 Φ <sub>2</sub> =-0.2154; Θ <sub>2</sub> =0.9754	0.9306	(1;303)=4061.5	5.6817	0.9294

\* p<sub>α</sub><0.05; \*\* p<sub>α</sub><0.01

#### 4.3.2.3. Temperatura en las series 'E' y 'F'

Para la estimación de la temperatura del agua en las series correspondientes a la zona de preengorde (series 'E' y 'F'), se ha procedido de la misma forma que en la estimación de la temperatura del agua en la serie 'B'. Se ha llevado a cabo por tanto, la regresión simple entre la temperatura en la serie 'A' y preengorde y entre la serie 'B' y preengorde para el período de calibración. Los resultados obtenidos son estadísticamente aceptables aunque en el mejor de los casos el porcentaje de varianza explicada no supera el 62 % (tabla IV.7).

La validación de ambos modelos a proporcionado niveles de varianza explicada superiores a los encontrados durante el proceso de calibración (0.90 y 0.92, respectivamente), lo que puede deberse a que gran parte de la variabilidad de la temperatura

del agua en las series de preengorde haya sido recogida previamente por el grupo de calibración. Sin embargo un peor ajuste durante la calibración se refleja en los errores de predicción, que en esta ocasión se encuentran cercanos al 15 % (SEP) y alrededor de 0.65 (Coeficiente E) (tabla IV.7).

**Tabla IV.7.** Estimación de la temperatura del agua en las series 'E' y 'F' a través de la temperatura del agua en las series 'A' y 'B'

<b>Resumen de la regresión en la calibración del modelo</b>			
(1) $R=0.7516$ ; $R^2=0.5648$ ; $F(1;846)=1098.1$ ; $p_\alpha<0.001$ ; $N=848$			
(2) $R=0.7857$ ; $R^2=0.6173$ ; $F(1;862)=1390.4$ ; $p_\alpha<0.001$ ; $N=864$			
<b>Variable dependiente</b>	<b>Variable independiente</b>	<b><math>b_0, b_1</math></b>	<b><math>p_\alpha</math></b>
Temperatura del agua en 'E' y 'F' (1)	Intersección ( $b_0$ )	10.4502	0**
	Temperatura del agua en 'A' ( $b_1$ )	0.6325	0**
Temperatura del agua en 'E' y 'F' (2)	Intersección ( $b_0$ )	8.3035	0**
	Temperatura del agua en 'B' ( $b_1$ )	0.7193	0**
<b>Resumen de la regresión en la validación del modelo</b>			
(1) $R=0.9486$ ; $R^2=0.8999$ ; $F(1;351)=3155.6$ ; $p_\alpha<0.001$ ; $N=353$ ; SEP=14.6717; E=0.6278			
(2) $R=0.9587$ ; $R^2=0.9191$ ; $F(1;305)=3667.5$ ; $p_\alpha<0.001$ ; $N=325$ ; SEP=12.5581; E=0.6915			
<b>Variable dependiente</b>	<b>Variable independiente</b>	<b><math>b_0, b_1</math></b>	<b><math>p_\alpha</math></b>
Temperatura del agua en 'E' y 'F' observada (1)	Intersección ( $b_0$ )	-19.7341	0**
	Temperatura del agua en 'A' estimada ( $b_1$ )	1.7522	0**
Temperatura del agua en 'E' y 'F' observada (2)	Intersección ( $b_0$ )	-162247	0**
	Temperatura del agua en 'B' estimada ( $b_1$ )	1.6118	0**

\*\*  $p_\alpha<0.01$

Ya que los errores de validación son sensiblemente mayores a los encontrados en la estimación de la temperatura del agua en las series 'A' y 'B', se ha procedido a la calibración y validación de los modelos de estimación de la temperatura del agua en las series 'E' y 'F' de forma independiente a las dos series anteriores. De esta forma, el nivel de predicción obtenido es tan satisfactorio como los conseguidos para las series 'A' y 'B' (tabla IV.8 y IV.9).

En el caso de las regresiones múltiples encontramos pequeñas diferencias en cuanto a la significación de las variables independientes implicadas durante la calibración del modelo. Así, en una primera estimación de la temperatura del agua a partir de las temperaturas existentes en los cinco días previos, se puede observar como las variables que significativamente explican una mayor parte de la varianza de la variable dependiente son la temperaturas del primer, segundo y quintos días anteriores. De la misma forma que se hizo en la serie 'A', se ha ido disminuyendo el número de variables independientes implicadas en función de la significación encontrada en el modelo anterior. Como consecuencia de este proceso de reducción de complejidad del modelo se ha probado, además de los citados anteriormente, el que considera como variables independientes los dos días anteriores (tabla IV.8).

La validación de los modelos de regresión múltiple han proporcionado muy buenos resultados, ya que en todos los casos el coeficiente de determinación se encuentra por encima de 0.97 y los errores no superan el 3.8 % (SEP). De todos ellos, el que presenta un mayor ajuste con los datos reales es el que considera como variables independientes las temperaturas de los cinco días anteriores ( $R^2=0.9745$ ;  $F(1;371)=14195.6$ ;  $p_\alpha<0.001$ ;  $N=373$ ;  $SEP=3.7909$ ;  $Coef. E=0.9734$ ) (tabla IV.8; figura IV.12).

**Tabla IV.8.** *Parámetros de validación en la estimación de la temperatura del agua en las series 'E' y 'F' a través de modelos de regresión múltiple*

Modelo en calibración	$R^2$ Validación	F	$p_\alpha$	% SEP	Coef. E
<i>Temperaturas de los 5 días anteriores</i>	0.9745	(1;371)=14195.6	0**	3.7909	0.9734
<i>Temperatura de 1<sup>er</sup>, 2<sup>o</sup> y 5<sup>o</sup> días anteriores</i>	0.9744	(1;371)=14145.0	0**	3.7953	0.9733
<i>Temperatura de los dos días anteriores</i>	0.9745	(1;371)=14165.9	0**	3.7948	0.9733

\*\*  $p_\alpha<0.01$

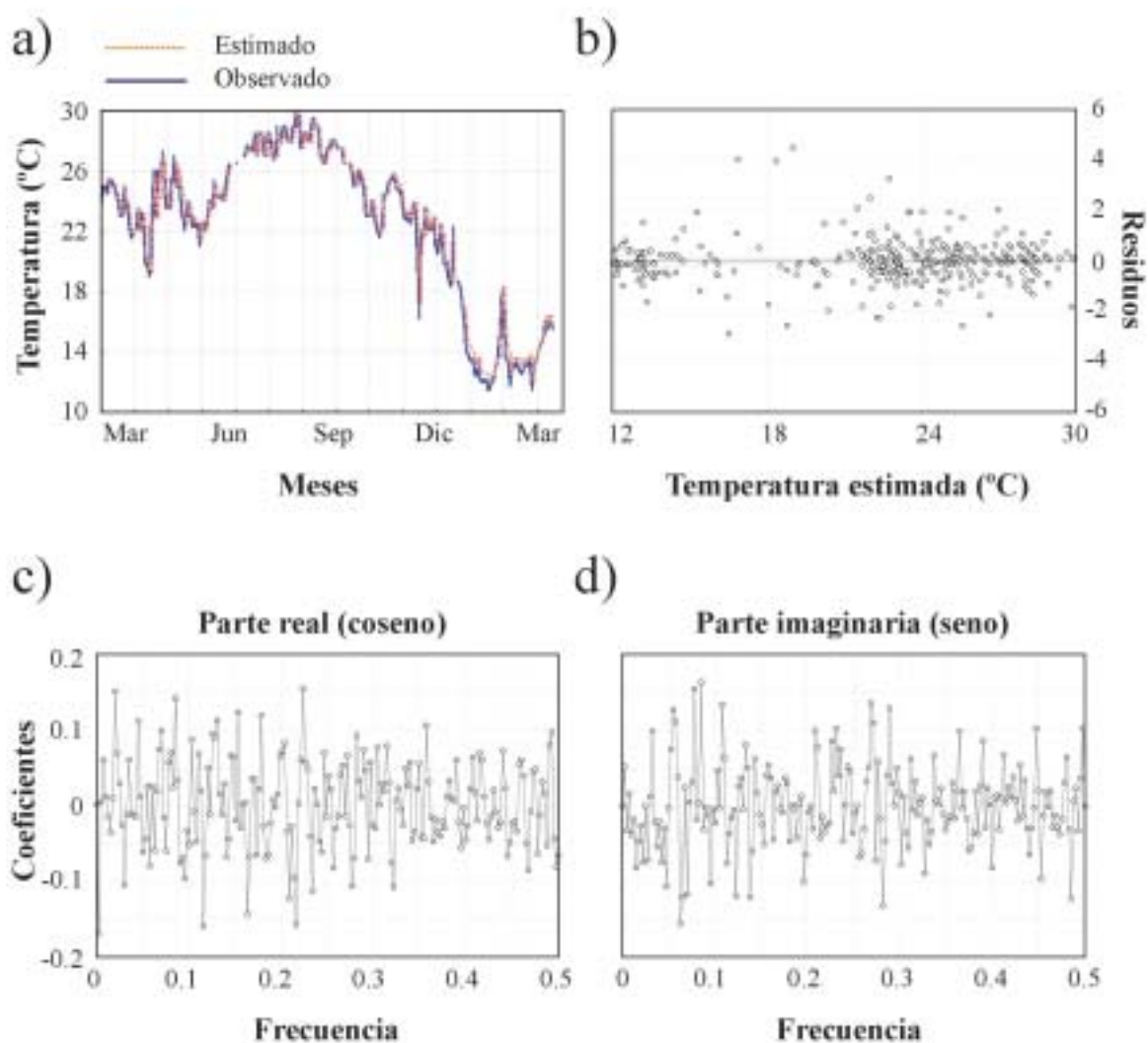
De la misma forma que las regresiones múltiples, los modelos de suavizado y **ARIMA** han dado como resultado muy buenos ajustes entre valores reales y estimados en validación. De entre todos los modelos probados destaca como el mejor el ARIMA (2,1,2) con el que se consigue explicar casi un 98 % de la varianza de la variable dependiente y un error de predicción del 3.6 % ( $R^2=0.9764$ ;  $F(1;385)=15941.4$ ;  $p_\alpha<0.001$ ;  $N=387$ ;  $SEP=3.5663$ ;  $Coef. E=0.9764$ ) (tabla IV.9; figura IV.13), coincidiendo de esta forma con el mejor modelo de serie temporal encontrado para la serie 'A'.

**Tabla IV.9.** *Parámetros de estimación y bondad de ajuste de los modelos de series temporales de la temperatura del agua de las series 'E' y 'F'*

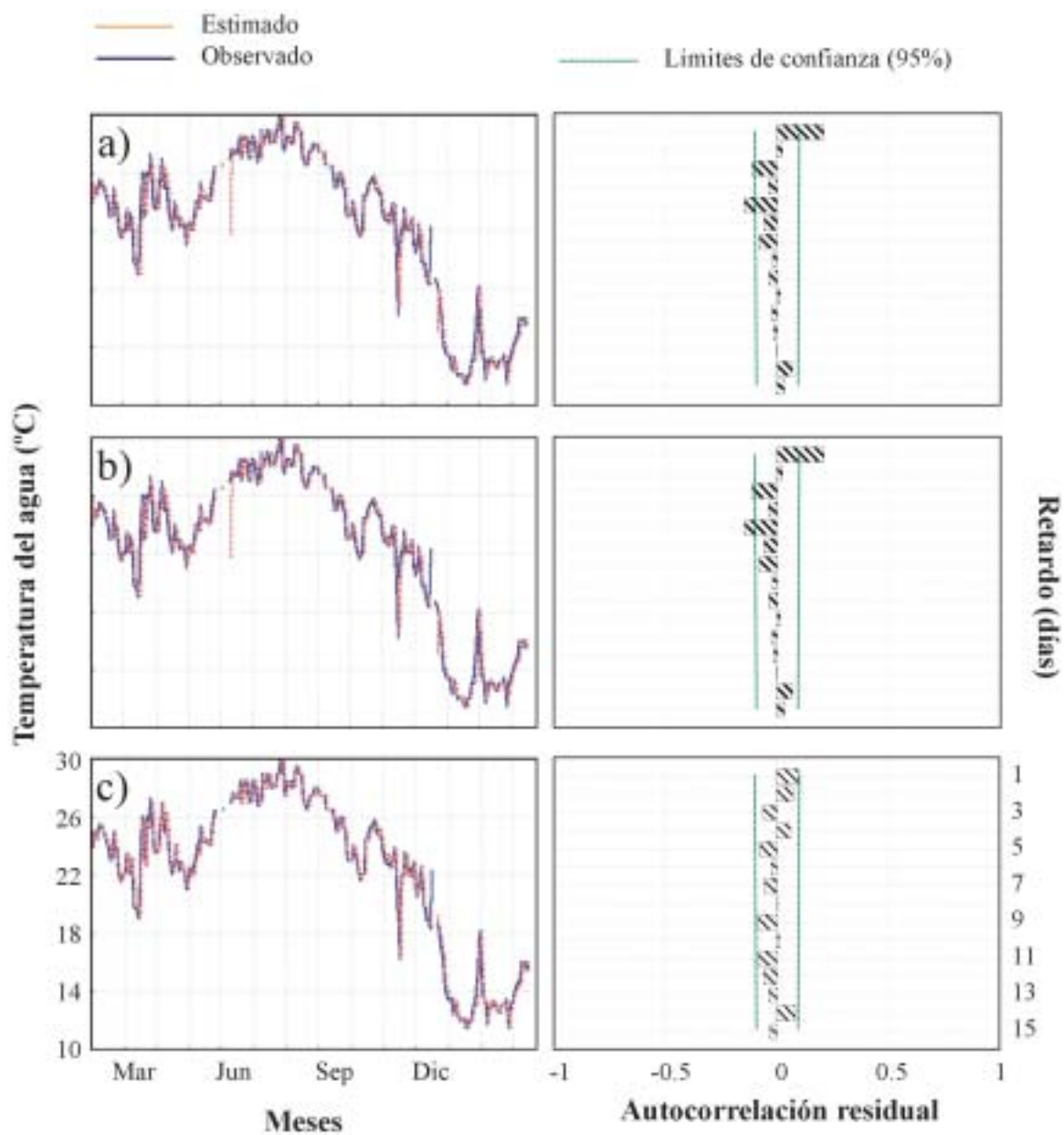
Modelo	Parámetros	Validación			
		R <sup>2</sup>	F	%SEP	Coef. E
Suavizado exponencial simple	S <sub>0</sub> =25.4300; α <sub>d</sub> =0.8040	0.8795	(1;386)=2817.3	3.9919	0.9705
Suavizado lineal exponencial	S <sub>0</sub> =26; T <sub>0</sub> =-0.0010 α <sub>d</sub> =0.8040; β <sub>d</sub> =0	0.9703	(1;384)=12527.1	4.0014	0.9703
ARIMA (0,0,1)	Θ <sub>1</sub> =-0.9498**	0.5285	(1;385)=431.5	53.3984	-4.2841
ARIMA (1,0,0)	Φ <sub>1</sub> =0.9986**	0.9758	(1;385)=15544.4	3.6123	0.9758
ARIMA (0,1,1)	Θ <sub>1</sub> =-0.0353	0.9759	(1;385)=15577.9	3.6112	0.9758
ARIMA (1,0,1)	Φ <sub>1</sub> =0.9985** ; Θ <sub>1</sub> =-0.0441	0.9760	(1;385)=15669.9	3.6000	0.9760
ARIMA (1,1,0)	Φ <sub>1</sub> =0.0363	0.9759	(1;385)=15579.2	3.6111	0.9758
ARIMA (1,1,1)	Φ <sub>1</sub> =0.0595 ; Θ <sub>1</sub> =0.0232	0.9759	(1;385)=15579.7	3.6111	0.9758
ARIMA (1,1,2)	Φ <sub>1</sub> =0.8012** ; Θ <sub>1</sub> =0.7895** Θ <sub>2</sub> =0.0801	0.9761	(1;385)=15753.4	3.5864	0.9762
ARIMA (1,2,1)	Φ <sub>1</sub> =0.0557 ; Θ <sub>1</sub> =-0.9703**	0.9537	(1;382)=7872.1	5.1131	0.9516
ARIMA (2,2,1)	Φ <sub>1</sub> =0.1021* ; Θ <sub>1</sub> =-0.9466** Φ <sub>2</sub> =-0.4916	0.9638	(1;382)=10171.9	4.4940	0.9626
ARIMA (2,1,2)	Φ <sub>1</sub> =1.6656** ; Θ <sub>1</sub> =1.6650** Φ <sub>2</sub> =-0.7986** ; Θ <sub>1</sub> =-0.7607**	0.9764	(1;385)=15941.4	3.5663	0.9764
ARIMA (2,1,1)	Φ <sub>1</sub> =-0.8333** ; Θ <sub>1</sub> =-0.8811** Φ <sub>2</sub> =0.0923	0.9763	(1;385)=15844.7	3.5800	0.9762
ARIMA (1,2,2)	Φ <sub>1</sub> =0.5190** ; Θ <sub>1</sub> =0.0154 Θ <sub>2</sub> =0.9671**	0.9634	(1;385)=10042.8	4.5143	0.9622
ARIMA (2,2,2)	Φ <sub>1</sub> =0.7241** ; Θ <sub>1</sub> =0.0228 Φ <sub>2</sub> =-0.3991** ; Θ <sub>2</sub> =0.9481**	0.9685	(1;382)=11728.3	4.1736	0.9677

\* p<sub>α</sub><0.05; \*\* p<sub>α</sub><0.01

De forma global, se puede observar una disminución significativa del error de validación en las series 'E' y 'F' (ANOVA: F=14.533, p<sub>α</sub><0.001 – No se considera el error de predicción del modelo ARIMA (0,0,1)). Un test de Tukey posterior confirmó la presencia de diferencias significativas entre los errores de predicción cometidos en la serie 'A' y las series 'E' y 'F', y la serie 'B' y las series 'E' y 'F'. Asimismo, no se encontraron diferencias significativas entre los errores de validación de las series 'A' y 'B'.



**Figura IV.12.** a) Temperatura del agua observada y estimada de los sistemas 'E' y 'F' en el periodo de validación (años 2000 y 2001), considerando en el análisis de regresión múltiple como variable dependiente la temperatura del agua y como variables independientes la temperatura del agua de los cinco días anteriores; b) Residuos (temperatura observada menos temperatura estimada) en el grupo de validación frente a la temperatura estimada; c) y d) Frecuencia de las funciones coseno y seno frente a los coeficientes encontrados en la transformada de Fourier

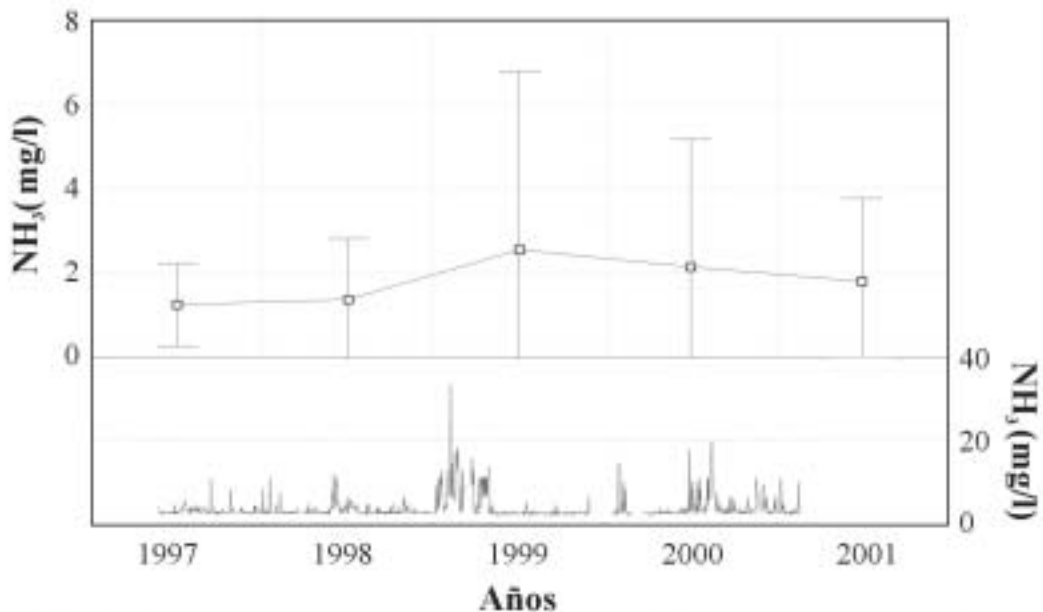


**Figura IV.13.** Temperatura del agua observada y estimada en las series 'E' y 'F', y autocorrelación residual en validación para los modelos: a) suavizado exponencial simple; b) suavizado lineal exponencial (Holt); c) ARIMA (2,1,2)



#### 4.3.2.4. Amoniac en la serie 'A'. Estimación a partir de las concentraciones de amoniac de los días previos

La concentración media anual de  $\text{NH}_3$  en la serie 'A' así como la evolución temporal de los valores absolutos de  $\text{NH}_3$  se muestra en la figura IV.14. Se puede observar un aumento significativo de las concentraciones de amoniac total a partir del año 1999 que se mantienen durante el 2000 para finalmente decaer ligeramente en el 2001 (ANOVA:  $F=10.755$ ,  $p_\alpha < 0.001$ ). Un test de Tukey posterior detectó la presencia de diferencias significativas entre los valores medios de los años 1997, 1998 y 2001 con respecto a los valores medios de los años 1999 y 2000. Los altos valores promedios encontrados tanto para el año 1999 y 2000 ( $2.55 \pm 4.24$  y  $2.14 \pm 3.03$ , respectivamente) se deben a la presencia de períodos con concentraciones altas de este parámetro, que en ocasiones puntuales alcanza los 35 mg/l.



**Figura IV.14.** Valores medios anuales de la concentración de  $\text{NH}_3$  presente en el agua de la serie 'A' (gráfica superior). Evolución temporal a lo largo del periodo de estudio (valores absolutos) en la serie 'A' (gráfica inferior)

De la misma forma que en el proceso de estimación de la temperatura del agua, se ha procedido al cálculo de las ecuaciones de regresión que relaciona la concentración de amoniac en un día concreto con la concentración de dicho parámetro en los cinco días anteriores. En esta ocasión, al contrario de los que sucedía con la temperatura, los resultados obtenidos tanto en el proceso de calibración como en el de validación, son estadísticamente pobres. Así con este primer modelo, la cantidad de varianza de la variable dependiente explicada por las variables independientes durante la calibración es del 58 % ( $R^2=0.5808$ ). Este porcentaje disminuye hasta el 41 % ( $R^2=0.4088$ ) en la generalización. De este modo, el

error de predicción (**SEP**) se eleva hasta el 107.3569 % y el coeficiente de eficiencia (**E**) disminuye hasta 0.4162 (tabla IV.10). En el análisis de los residuos se puede observar una agrupación desigual de éstos y la presencia de frecuencias con componentes significativamente distintos de cero en el análisis de Fourier. El estadístico de Durbin-Watson ( $D=1.9361$ ), muy próximo a 2, indica la ausencia de correlación serial en los residuos durante la calibración aunque a través del análisis de Fourier se puede intuir la presencia de periodicidad en esta serie (figura IV.15).

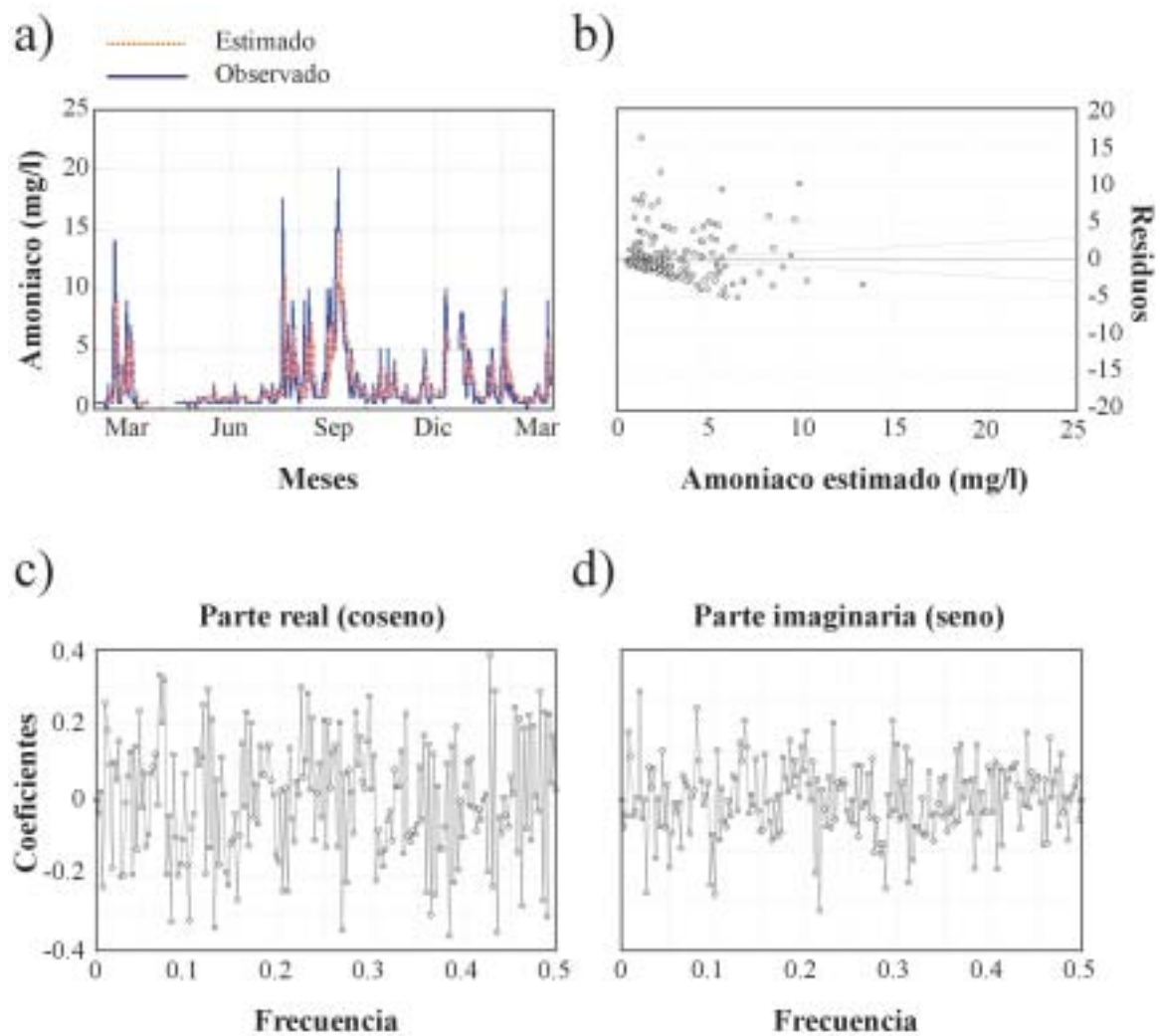
**Tabla IV.10.** Estimación del  $\text{NH}_3$  del agua en la serie 'A' mediante el análisis de regresión múltiple con las variables independientes representadas por el  $\text{NH}_3$  en los cinco días anteriores

<b>Resumen de la regresión múltiple en la calibración del modelo</b>			
R=0.7621; $R^2=0.5808$ ; $F(5;870)=241.7$ ; $p_\alpha < 0.001$ ; N=876			
Variable dependiente	Variable independiente	$b_i(i=0,1,\dots,q)$	$p_\alpha$
$\text{NH}_3(t)$	Intersección ( $b_0$ )	0.2766	0.0004**
	$\text{NH}_3(t-1)$	0.6051	0**
	$\text{NH}_3(t-2)$	0.0028	0.9429
	$\text{NH}_3(t-3)$	0.0821	0.0353*
	$\text{NH}_3(t-4)$	0.0612	0.1198
	$\text{NH}_3(t-5)$	0.0863	0.0112*
<b>Resumen de la regresión en la validación del modelo</b>			
R=0.6394; $R^2=0.4088$ ; $F(1;361)=249.6$ ; $p_\alpha < 0.001$ ; N=363			
Error estándar de predicción=107.3569 %; Coeficiente E=0.4162			
Variable dependiente	Variable independiente	$b_i(i=0,1,\dots,q)$	$p_\alpha$
$\text{NH}_3$ observado (t)	Intersección ( $b_0$ )	0.2818	0.0850
	$\text{NH}_3$ estimado (t)	0.8942	0**

\*  $p_\alpha < 0.05$ ; \*\*  $p_\alpha < 0.01$

En la regresión anterior, se puede observar como tan solo tres de las cinco variables independientes contribuyen significativamente en la varianza de la variable dependiente. Se ha procedido por tanto a la estimación de la concentración de  $\text{NH}_3$  en función tan solo de estas variables significativas ( $\text{NH}_3$  en el primer, tercer y quinto días anteriores). De esta forma, los resultados obtenidos tanto en calibración como en validación son estadísticamente similares (Calibración:  $R=0.7613$ ;  $R^2=0.5796$ ;  $F(3;872)=400.8$ ;  $p_\alpha < 0.001$ ; N=876. Validación:  $R=0.6382$ ;  $R^2=0.4073$ ;  $F(1;361)=248.0$ ;  $p_\alpha < 0.001$ ; N=363;  $\text{SEP}=107.5335$ ; Coeficiente E=0.4143).

Tampoco han proporcionado buenas estimaciones los modelos de suavizado, al encontrarse durante la validación magnitudes de error incluso mayores a las detectadas en las regresiones múltiples (Suavizado Exponencial Simple:  $\text{SEP}=118.4885$  %; Coeficiente E=0.2888. Suavizado de Holt:  $\text{SEP}=114.0999$  %; Coeficiente E=0.3405) (figura IV.16). Los modelos **ARIMA** en los que se han encontrado todos sus parámetros significativos son el  $\text{ARIMA}(0,0,1)$ ,  $\text{ARIMA}(1,0,0)$  y  $\text{ARIMA}(0,1,1)$ . Sin embargo, los resultados en validación



**Figura IV.15.** a) Concentración de amoniaco observada y estimada del sistema 'A' en el periodo de validación, considerando en el análisis de regresión múltiple como variable dependiente la concentración de amoniaco y como variables independientes las concentraciones de amoniaco de los cinco días anteriores; b) Residuos (concentración de amoniaco observada menos concentración de amoniaco estimada) en el grupo de validación frente a la concentración de amoniaco estimada; c) y d) Frecuencia de las funciones coseno y seno frente a los coeficientes encontrados en la transformada de Fourier

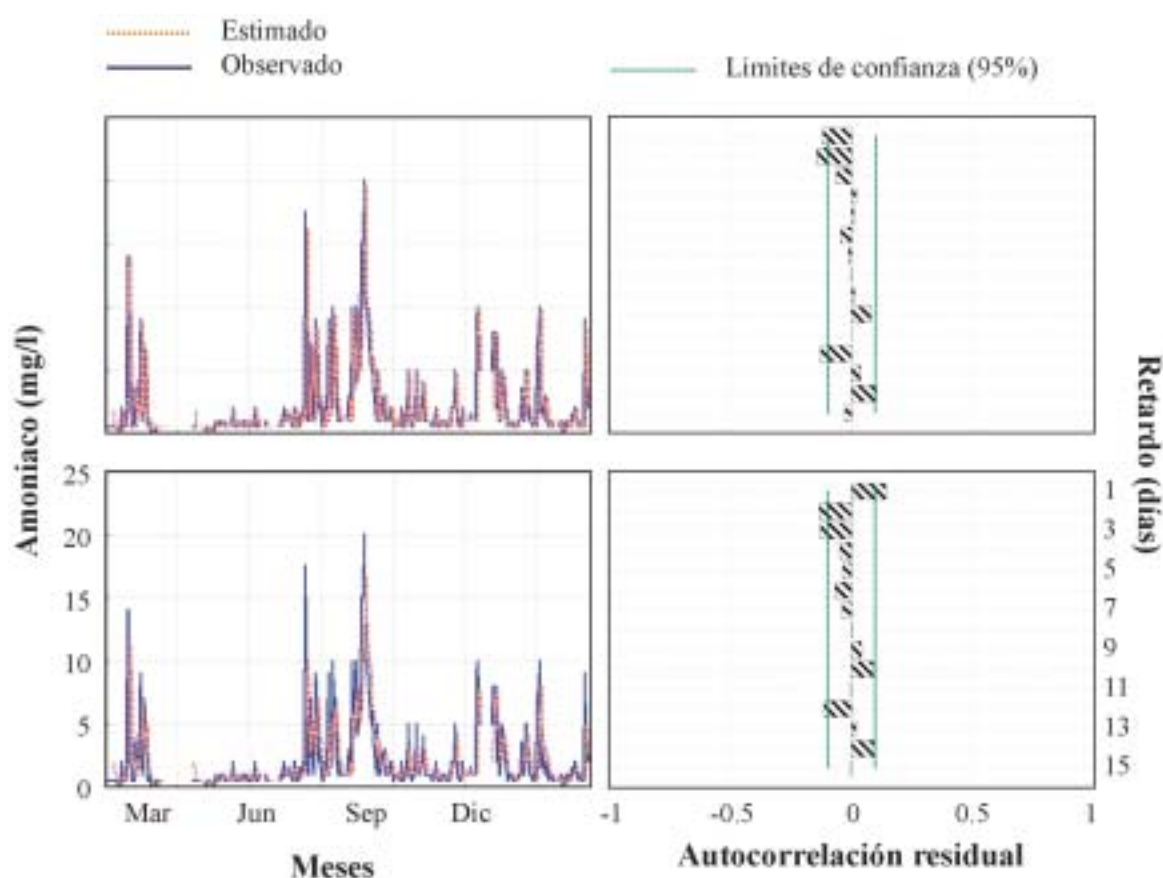


Figura IV.16. Concentración de amoniaco en el agua observada y estimada en la serie 'A' y autocorrelación residual en validación para los modelos: suavizado exponencial simple (arriba) y suavizado lineal exponencial (Holt) (abajo)

reflejan un bajo nivel de varianza explicada (17 %, 40 % y 35 %, respectivamente) y magnitudes de error que en todo momento son mayores al 110 % (**SEP**) e inferiores a 0.36 (Coeficiente **E**). El mejor resultado en cuanto al término de error lo encontramos en el modelo ARIMA(1,1,1), donde el porcentaje de error de predicción es del 109 % y el coeficiente de eficiencia es 0.39 (tabla IV.11).

La falta de bondad de ajuste encontrada entre los valores observados y estimados mediante el empleo de modelos de regresión múltiple y modelos univariantes de series temporales durante la validación, puede estar condicionada por el reparto de información relevante sobre los sucesivos retrasos implicados en la estimación. De esta forma, un examen de las funciones de autocorrelación y autocorrelación parcial de la serie ha mostrado la presencia de correlación serial para al menos nueve y un retrasos respectivamente (figura IV.17a). Así, el modelo ARIMA(1,0,9) incrementó el porcentaje de varianza explicada con respecto al modelo ARIMA(1,1,1) en 19 puntos, alcanzando ésta un valor del 57 % (tabla IV.11). A pesar de que durante la validación del modelo la función de autocorrelación residual no mostró diferencias significativas con respecto al cero en ninguno de sus términos

(figura IV.17b), las magnitudes de error mejoraron de forma poco importante, situándose el porcentaje de error de predicción alrededor del 105 % mientras que el coeficiente de eficiencia fue de 0.43. El aumento de la cantidad de varianza explicada junto con un estancamiento de los términos de error puede estar relacionado con la sobreestimación de los valores máximos y mínimos que se producen en la validación de este modelo, sobre todo cuando existen importantes variaciones de las concentraciones de amoniaco entre dos días consecutivos (figura IV.17). Por otra parte, la regresión múltiple en la que las variables independientes se corresponden con la concentración de amoniaco de los nueve días anteriores tampoco mejoró la predicción (Calibración:  $R=0.7675$ ;  $R^2=0.5892$ ;  $F(9;854)=136.1$ ;  $p_\alpha<0.001$ ;  $N=864$ . Validación:  $R=0.6270$ ;  $R^2=0.3931$ ;  $F(1;349)=226.1$ ;  $p_\alpha<0.001$ ;  $N=351$ ;  $SEP=107.3417$ ; Coeficiente  $E=0.4163$ ).

**Tabla IV.11.** *Parámetros de estimación y bondad de ajuste de los modelos de series temporales del  $NH_3$  de la serie 'A'*

Modelo	Parámetros	Validación			
		$R^2$	F	%SEP	Coef. E
Suavizado exponencial simple	$S_0=1.8820$ ; $\alpha_d=0.5270$	0.3971	(1;263)=239.1	118.4885	0.2888
Suavizado lineal exponencial	$S_0=2.0010$ ; $T_0=-0.0020$ $\alpha_d=0.5260$ ; $\beta_d=0$	0.3797	(1;265)=223.4	114.1000	0.3405
ARIMA (0,0,1)	$\Theta_1=-0.8372^{**}$	0.1711	(1;376)=77.6	147.0969	-0.1293
ARIMA (1,0,0)	$\Phi_1=0.9283^{**}$	0.4040	(1;376)=254.9	111.9265	0.3462
ARIMA (0,1,1)	$\Theta_1=0.6340^*$	0.3521	(1;375)=203.8	111.7047	0.3505
ARIMA (1,0,1)	$\Phi_1=0.9999$ ; $\Theta_1=0.3811$	0.3934	(1;375)=243.8	110.4236	0.3636
ARIMA (1,1,0)	$\Phi_1=-0.2936$	0.3904	(1;375)=240.2	113.2426	0.3325
ARIMA (1,1,1)	$\Phi_1=0.1545$ ; $\Theta_1=0.6866^*$	0.3821	(1;375)=231.9	108.5963	0.3862
ARIMA (1,1,2)	$\Phi_1=-0.6921^{**}$ ; $\Theta_1=-0.3006$ $\Theta_2=0.6590^*$	0.2906	(1;375)=153.6	126.2684	0.1701
ARIMA (1,2,1)	$\Phi_1=0.2702$ ; $\Theta_1=-0.9999$	0.2441	(1;372)=120.1	185.4264	-0.7752
ARIMA (2,2,1)	$\Phi_1=0.5754$ ; $\Theta_1=0.3280$ $\Phi_2=-0.9999$	0.1434	(1;372)=62.3	195.9579	-0.9826
ARIMA (2,1,2)	$\Phi_1=-0.7571$ ; $\Theta_1=-0.3221$ $\Phi_2=-0.0670$ ; $\Theta_1=0.6374^*$	0.2787	(1;375)=144.9	127.8792	0.1487
ARIMA (2,1,1)	$\Phi_1=0.1371$ ; $\Theta_1=0.6098$ $\Phi_2=-0.2325$	0.3786	(1;375)=228.5	108.9728	0.3819
ARIMA (1,2,2)	$\Phi_1=0.3209$ ; $\Theta_1=0.4188$ $\Theta_2=0.5811$	0.1363	(1;372)=58.7	168.9216	-0.4733
ARIMA (2,2,2)	$\Phi_1=0.8160$ ; $\Theta_1=0.7140$ $\Phi_2=-0.9999$ ; $\Theta_2=-0.0579$	0.1190	(1;372)=50.2	196.7923	-0.9995
ARIMA (1,0,9)	-----	0.5726	(1;373)=499.75	104.8325	0.4310

\*  $p_\alpha<0.05$ ; \*\*  $p_\alpha<0.01$

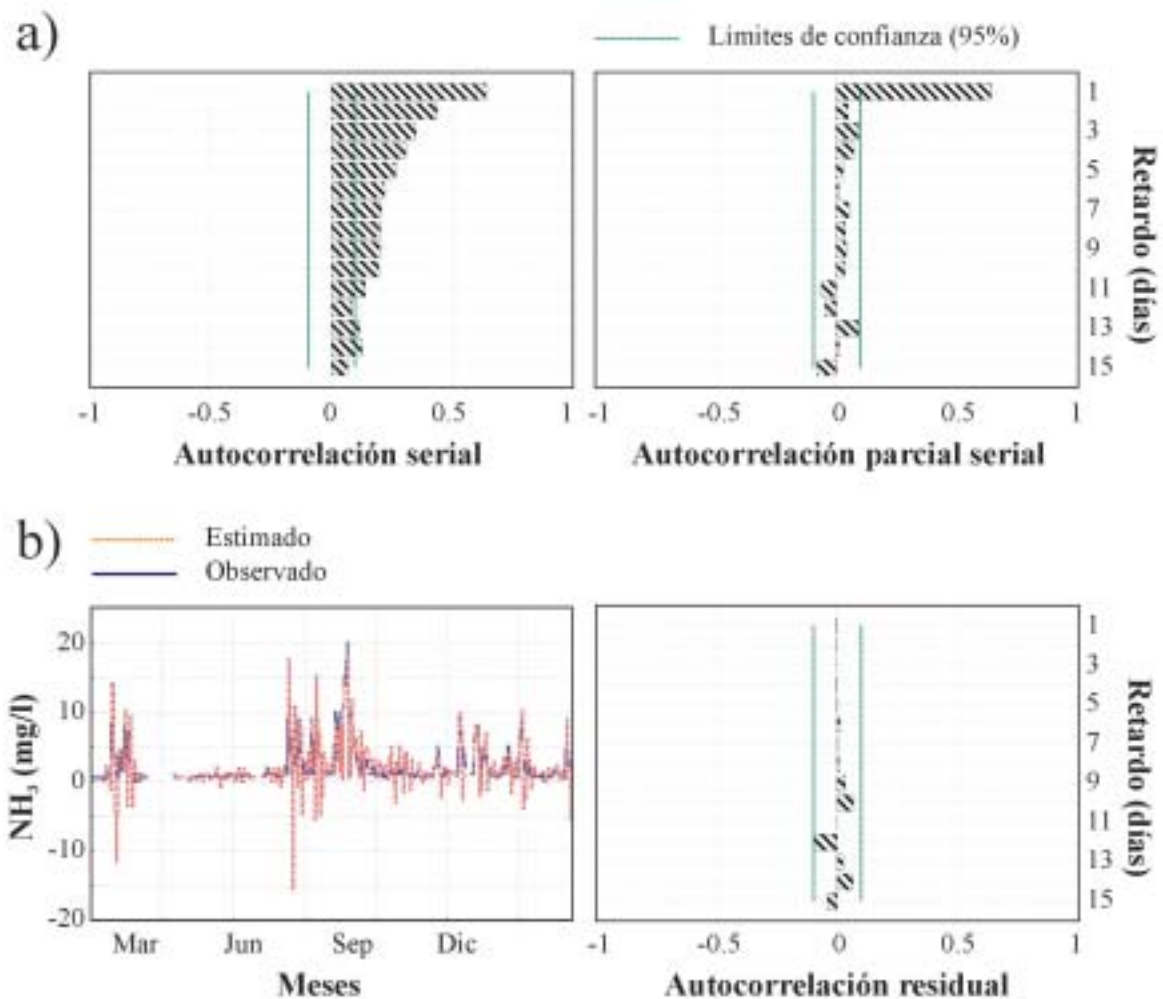


Figura IV.17. a) Autocorrelación serial y autocorrelación parcial serial de la concentración de amoníaco en el sistema 'A'. b) Concentración de amoníaco observado y estimado y, función de autocorrelación residual para el modelo ARIMA(1,0,9)

Ya que ni los modelos de regresión múltiple ni los modelos univariantes de series temporales han proporcionado resultados estadísticamente satisfactorios y, el error estándar de predicción se ha encontrado en todo momento por encima del 5 %, se ha procedido a la estimación del parámetro NH<sub>3</sub> en función de sus concentraciones en los días previos mediante el empleo de Redes Neuronales Computacionales (RNCs). Para ello se han entrenado de un total de 10 RNCs con distintas arquitecturas en las que el nivel de complejidad de la red varía en función del número de neuronas en las capas ocultas. Así, partiendo de una red relativamente simple como la 5:5s:5s:11, se ha llegado hasta la arquitectura 5:20s:20s:11, añadiendo 5 neuronas en cada una de las capas ocultas en los sucesivos entrenamientos.

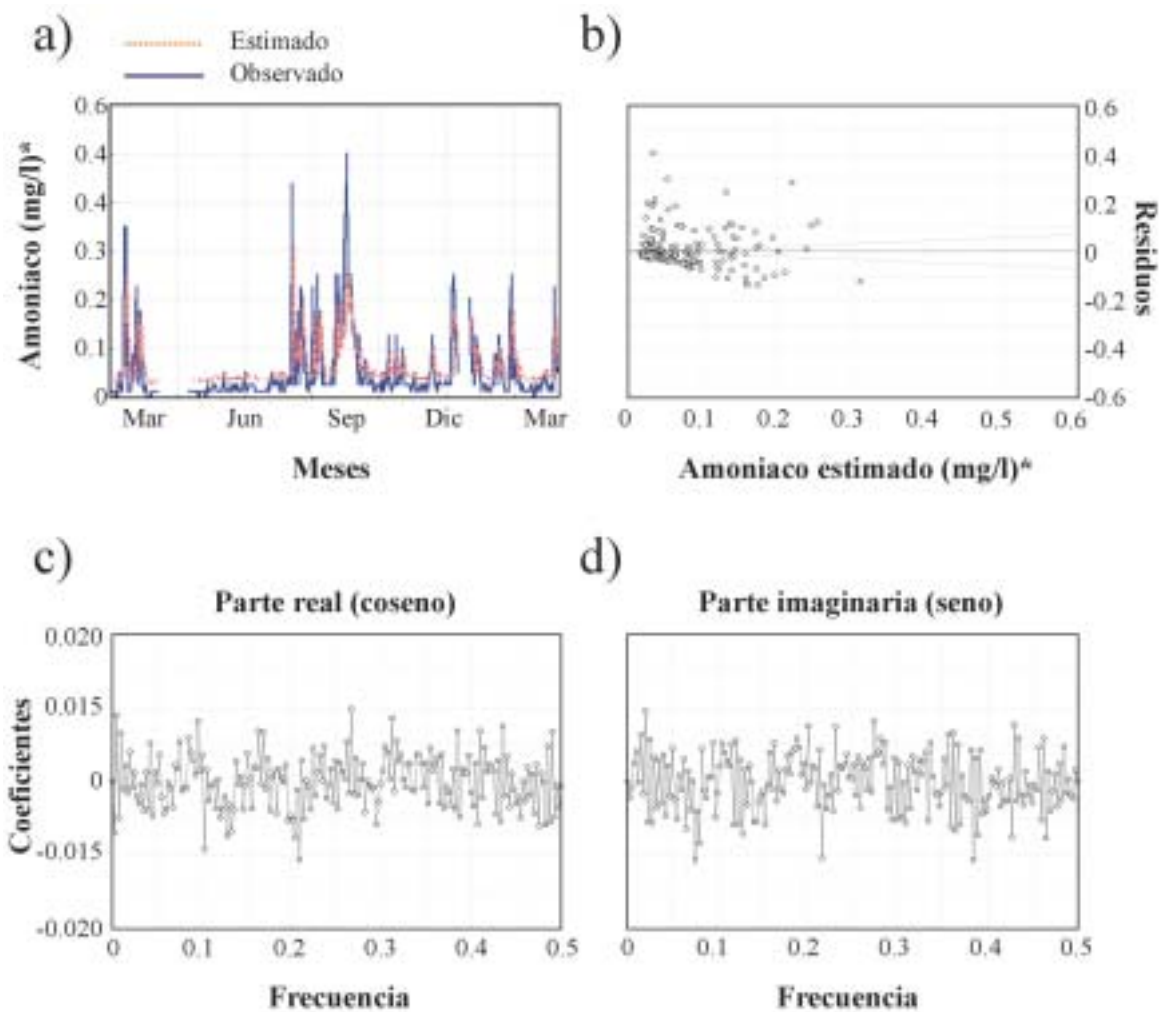
A pesar de la reconocida potencia de los modelos neuronales, los resultados obtenidos son tan estadísticamente pobres como los conseguidos en los casos anteriores (tabla IV.12).

**Tabla IV.12.** Resultados de calibración y generalización de las RNCs en las que las variables de entrada son las concentraciones de amoniaco en el sistema 'A' en los cinco días anteriores. Parada del entrenamiento mediante validación cruzada (50 patrones) y un coeficiente de determinación umbral igual a 0.8 ( $\epsilon$ =número de pesos/número de patrones)

Red	Pesos	Calibración		Generalización				
		$\epsilon$	Épocas	R	R <sup>2</sup>	RMS <sup>1</sup>	%SEP	E
5:5s:5s:11	55	0.1884	771	0.5353	0.2865	0.0623	119.0040	0.2476
5:5s:10s:11	85	0.2911	243	0.4228	0.1788	0.0653	124.9249	0.1709
5:10s:5s:11	105	0.3596	49	0.6237	0.3890	0.0568	108.6231	0.3732
5:10s:10s:11	160	0.5479	81	0.6259	0.3918	0.0565	108.2092	0.3779
5:10s:15s:11	215	0.7363	44	0.6312	0.3984	0.0565	108.0680	0.3796
5:15s:10s:11	235	0.8048	569	0.4763	0.2269	0.0664	136.8623	0.1450
5:15s:15s:11	315	1.0788	149	0.6101	0.3722	0.0575	109.9402	0.3579
5:15s:20s:11	395	1.3527	683	0.5917	0.3501	0.0593	113.3078	0.3179
5:20s:15s:11	415	1.4212	12	0.5157	0.2659	0.0656	125.4776	0.1635
5:20s:20s:11	520	1.7808	67	0.6031	0.3637	0.0578	110.5627	0.3506

<sup>1</sup>Datos escalados en el rango [0,1]

En este caso, el mejor resultado se obtiene con la red 5:10s:15s:11, en la que la cantidad de varianza explicada es del 63.12 % y los términos de error son 108 % y 0.38 (**SEP** y coeficiente **E**, respectivamente). El error cometido también se refleja en una ligera anormalidad en la distribución de los residuos frente a las estimaciones de las concentraciones de amoniaco, aunque el valor del estadístico de Durbin-Watson indica la ausencia de correlación serial (Estadístico D=1.9734) (figura IV.18). La regresión simple entre los términos de error (**SEP** y **E**) y el número de pesos o parámetros de la red ha dado como resultado una clara independencia de ambos términos de error con respecto a la complejidad de la red (**SEP**: R=-0.0784;  $p_{\alpha}$ =0.8295; N=10. Coeficiente **E**: R=0.0799;  $p_{\alpha}$ =0.8264; N=10) (figura IV.19). Por otra parte se han encontrado marcadas diferencias significativas entre el porcentaje de error de predicción en función del número de parámetros y el error estimado durante la calibración ( $\epsilon$ ) (test de comparación de dos rectas de regresión:  $F_{exp}$ =47.5549;  $t_{exp}$ =101;  $t_{(0.001)}(23545)$ <3.2910; ANCOVA,  $p_{\alpha}$ <0.001).



**Figura IV.18.** a) Concentración de amoníaco observada y estimada del sistema 'A' en el período de validación (años 2000 y 2001) usando la red neuronal 5:10s:15s:11, donde las variables de entrada son las concentraciones de amoníaco en los cinco días previos; b) Residuos (concentración de amoníaco observada menos concentración de amoníaco estimada) en el grupo de validación frente a la concentración de amoníaco estimada; c) y d) Frecuencia de las funciones coseno y seno frente a los coeficientes encontrados en la transformada de Fourier. \*Valores escalados en el rango [0,1]



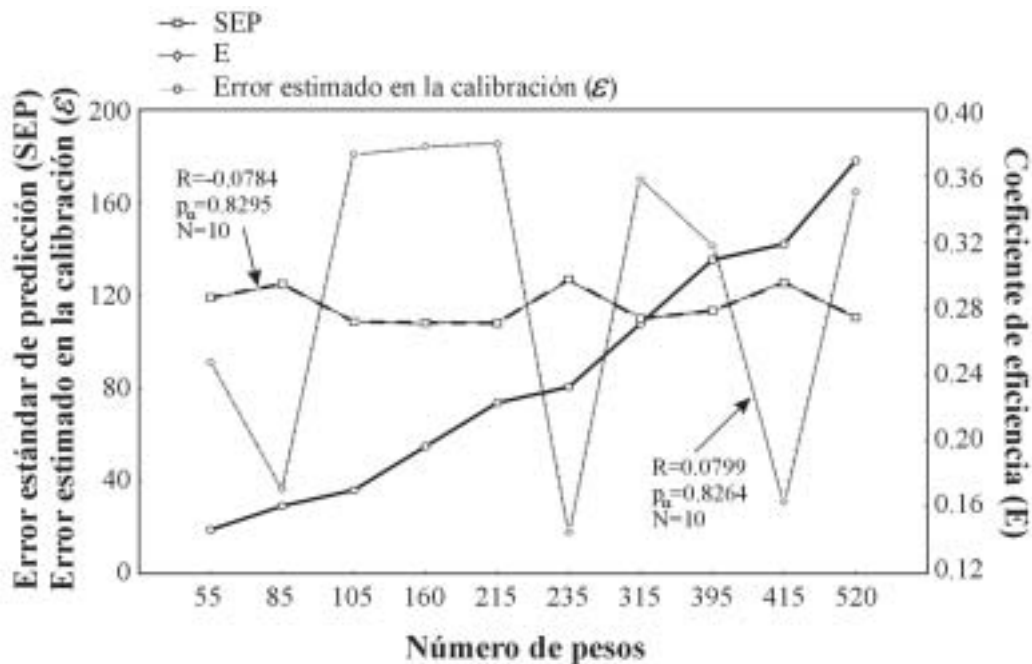


Figura IV.19. Error de predicción (SEP), coeficiente de eficiencia (E) y error estimado durante la calibración (e = número de pesos/número de patrones) de las Redes Neuronales Computacionales en las que las variables de entrada fueron las concentraciones de amoníaco en los cinco días previos en la serie 'A'. Se indica además la correlación y el nivel de significación encontrado para las regresiones entre SEP y número de pesos, y E y el número de pesos

#### 4.3.2.5. Amoníaco en la serie 'A'. Estimación a partir de otras variables implicadas

Como se comentó en el apartado 4.1.1.2 la concentración de  $\text{NH}_3$  en el agua depende de varios factores, por lo que se ha procedido a la calibración de ecuaciones de regresión y modelos de **RNCs** en los que dichos factores pasan a ser las variables independientes o variables de entrada. Desde un punto de vista físico-químico, las variables que tienen una mayor influencia en la cantidad de  $\text{NH}_3$  en el agua son la temperatura y el pH. Por otra parte, es indudable el papel fundamental que juegan otros parámetros como la cantidad de alimento suministrado, la proporción de este alimento que no es consumido, el índice de conversión del alimento, la carga de amoníaco en el biofiltro, así como la tasa de eliminación de amoníaco en el biofiltro.

De esta forma, se ha calibrado una primera regresión entre la concentración de  $\text{NH}_3$ , la temperatura y el pH del agua de los cinco días previos. Los resultados obtenidos se muestran en la tabla IV.13.

**Tabla IV.13.** Estimación del  $\text{NH}_3$  del agua en la serie 'A' mediante el análisis de regresión múltiple con las variables independientes representadas por la temperatura y el pH en los cinco días anteriores

<b>Resumen de la regresión múltiple en la calibración del modelo</b>			
R=0.2323; $R^2=0.0540$ ; $F(10;865)=4.9$ ; $p_\alpha < 0.001$ ; N=876			
Variable dependiente	Variable independiente	$b_i(i=0,1,\dots,q)$	$p_\alpha$
NH <sub>3</sub> (t)	Intersección ( $b_0$ )	14.6739	0**
	Temperatura del agua (t-1)	-0.0026	0.9810
	Temperatura del agua (t-2)	-0.0100	0.9506
	Temperatura del agua (t-3)	-0.0831	0.6057
	Temperatura del agua (t-4)	0.0450	0.7799
	Temperatura del agua (t-5)	0.1113	0.3520
	pH (t-1)	-0.7491	0.0278*
	pH (t-2)	-0.8694	0.0290*
	pH (t-3)	-0.0296	0.9409
	pH (t-4)	-0.0758	0.8493
	pH (t-5)	-0.3139	0.3538
<b>Resumen de la regresión en la validación del modelo</b>			
R=0.2519; $R^2=0.0635$ ; $F(1;331)=22.4$ ; $p_\alpha < 0.001$ ; N=333			
Error estándar de predicción=139.1316 %; Coeficiente E=0.1005			
Variable dependiente	Variable independiente	$b_i(i=0,1,\dots,q)$	$p_\alpha$
NH <sub>3</sub> observado (t)	Intersección ( $b_0$ )	-0.2363	0.6618
	NH <sub>3</sub> estimado (t)	1.1569	0**

\*  $p_\alpha < 0.05$ ; \*\*  $p_\alpha < 0.01$

Igual que cuando se utilizan como variables independientes los valores de las concentraciones de  $\text{NH}_3$  en los días previos, los resultados obtenidos tanto en calibración como en validación son estadísticamente muy poco satisfactorios. Como se puede observar, tan solo el pH de los dos días anteriores contribuye significativamente sobre la varianza de la variable dependiente. Por otra parte, el coeficiente de determinación de la regresión es tan solo ligeramente mayor a 0.05, indicando el bajo nivel de ajuste entre las variables independientes y dependiente. Como es esperable a partir de estos resultados, la validación del modelo también proporciona resultados pobres. El porcentaje de varianza explicada es del 6.35 % mientras que los términos de error equivalen a 139.1316 % (**SEP**) y 0.1005 (**E**). Resultados similares se obtienen cuando las variables independientes se reducen a las significativas de la regresión anterior (pH de los dos días anteriores, Calibración:  $R=0.1978$ ;  $R^2=0.0391$ ;  $F(2;882)=18.0$ ;  $p_\alpha < 0.001$ ; N=885. Validación:  $R=0.2115$ ;  $R^2=0.0447$ ;  $F(1;370)=17.3$ ;  $p_\alpha < 0.001$ ; N=372; SEP=143.3410; Coeficiente E=0.0452).

El modelo de regresión múltiple que considera como variables independientes la carga de amoníaco en el biofiltro, la cantidad de pienso suministrado y la tasa de eliminación de amoníaco en los cinco días anteriores muestra unos resultados estadísticamente correctos en el proceso de calibración tal y como se muestra en la tabla IV.14.

**Tabla IV.14.** Estimación del  $\text{NH}_3$  del agua en la serie 'A' mediante el análisis de regresión múltiple con las variables independientes representadas por la carga de amoníaco en el biofiltro, la cantidad de pienso suministrado y la tasa de eliminación de amoníaco en los cinco días anteriores

<b>Resumen de la regresión múltiple en la calibración del modelo</b>			
R=0.7961; $R^2=0.6338$ ; $F(15;299)=34.5$ ; $p_\alpha < 0.001$ ; N=315			
Variable dependiente	Variable independiente	$b_i(i=0,1,\dots,q)$	$p_\alpha$
NH <sub>3</sub> (t)	Intersección ( $b_0$ )	-0.7265	0.1679
	Carga de amoníaco (t-1)	0.0189	0.0008**
	Carga de amoníaco (t-2)	-0.0087	0.1549
	Carga de amoníaco (t-3)	0.0064	0.2885
	Carga de amoníaco (t-4)	-0.0007	0.9151
	Carga de amoníaco (t-5)	0.0055	0.2150
	Kilogramos de pienso (t-1)	0.0045	0.7072
	Kilogramos de pienso (t-2)	0.0066	0.6422
	Kilogramos de pienso (t-3)	-0.0056	0.6809
	Kilogramos de pienso (t-4)	-0.0036	0.7973
	Kilogramos de pienso (t-5)	0.0097	0.4225
	$R_i$ (t-1)	1.7157	0.1616
	$R_i$ (t-2)	1.9018	0.1553
	$R_i$ (t-3)	-1.1819	0.3734
	$R_i$ (t-4)	0.8695	0.5228
$R_i$ (t-5)	-1.3846	0.2652	
<b>Resumen de la regresión en la validación del modelo</b>			
R=0.6819; $R^2=0.4650$ ; $F(1;283)=246.0$ ; $p_\alpha < 0.001$ ; N=285			
Error estándar de predicción=113.1647 %; Coeficiente E=0.4409			
Variable dependiente	Variable independiente	$b_i(i=0,1,\dots,q)$	$p_\alpha$
NH <sub>3</sub> observado (t)	Intersección ( $b_0$ )	0.4609	0.0060**
	NH <sub>3</sub> estimado (t)	0.8182	0**

\*  $p_\alpha < 0.05$ ; \*\*  $p_\alpha < 0.01$

Durante la validación se observa como el porcentaje de varianza explicada casi alcanza el 50 % ( $R^2=0.465$ ). Sin embargo, los términos de error vuelven a encontrarse otra vez excesivamente alejados de los rangos de aceptación (SEP=113.1647, Coeficiente E=0.4409). De todas las variables incluidas en la calibración tan solo la carga de amoníaco en el biofiltro del día anterior muestra significación estadística con la concentración de amoníaco. Los resultados en calibración y validación para la regresión simple, donde esta variable se considera como la única independiente, son estadísticamente similares a los obtenidos con la regresión anterior aunque los valores alcanzados por el porcentaje de error de predicción y el coeficiente de eficiencia son peores. (Carga de amoníaco en el biofiltro del día anterior, Calibración:  $R=0.7769$ ;  $R^2=0.6035$ ;  $F(1;313)=476.5$ ;  $p_\alpha < 0.001$ ; N=315. Validación:  $R=0.6723$ ;  $R^2=0.4521$ ;  $F(1;283)=233.5$ ;  $p_\alpha < 0.001$ ; N=285; SEP=131.1528; Coeficiente E=0.2491).

Los modelos de redes neuronales en los cuales los grupos de variables de entrada son los mismos que los especificados para las regresiones múltiples anteriores tampoco han mejorado significativamente la predicción (tablas IV.15 y IV.16).

**Tabla IV.15.** Resultados de calibración y generalización de las RNCs en las que las variables de entrada son las temperaturas y pH en el sistema 'A' en los cinco días anteriores. Parada del entrenamiento mediante validación cruzada (50 patrones) y un coeficiente de determinación umbral igual a 0.8

		Calibración		Generalización				
Red	Pesos	$\epsilon$	Épocas	R	R <sup>2</sup>	RMS <sup>1</sup>	%SEP	E
10:5s:5s:1l	80	0.0914	1320	0.1483	0.0220	0.0736	133.7301	0.0110
10:5s:10s:1l	110	0.1257	49	-0.1460	0.0213	0.0747	135.7102	-0.0185
10:10s:5s:1l	155	0.1771	1	0.1624	0.0264	0.0745	135.3355	-0.0129
10:10s:10s:1l	210	0.2400	1305	0.1436	0.0206	0.0735	133.6532	0.0121
10:10s:15s:1l	265	0.3029	4	-0.1476	0.0218	0.0747	135.8713	-0.0209
10:15s:10s:1l	310	0.3543	6	-0.1583	0.0251	0.0755	137.2066	-0.0411
10:15s:15s:1l	390	0.4457	11	-0.1324	0.0175	0.0750	136.3416	-0.0280
10:15s:20s:1l	470	0.5371	6	0.2601	0.0676	0.0744	135.2211	-0.0112
10:20s:15s:1l	515	0.5886	54	-0.0371	0.0014	0.0743	135.0417	-0.0085
10:20s:20s:1l	620	0.7086	1	-0.0985	0.0097	0.0743	134.9801	-0.0076

<sup>1</sup>Datos escalados en el rango [0,1]

**Tabla IV.16.** Resultados de calibración y generalización de las RNCs en las que las variables de entrada son las cargas de amoníaco en el biofiltro, las cantidades de pienso suminitrado y la tasa de eliminación de amoníaco en el sistema 'A' en los cinco días anteriores. Parada del entrenamiento mediante validación cruzada (50 patrones) y un coeficiente de determinación umbral igual a 0.8

		Calibración		Generalización				
Red	Pesos	$\epsilon$	Épocas	R	R <sup>2</sup>	RMS <sup>1</sup>	%SEP	E
15:5s:5s:1l	105	0.3333	9786	0.5455	0.2976	0.0492	124.7286	0.2620
15:5s:10s:1l	135	0.4286	702	0.3808	0.1450	0.0535	135.3617	0.1308
15:10s:5s:1l	205	0.6508	40	0.6146	0.3777	0.0490	124.2195	0.2680
15:10s:10s:1l	260	0.8254	372	0.6067	0.3680	0.0479	121.2873	0.3021
15:10s:15s:1l	315	1.0000	513	0.5904	0.3485	0.0487	123.3003	0.2788
15:15s:10s:1l	385	1.2222	319	0.5672	0.3217	0.0494	124.9926	0.2588
15:15s:15s:1l	465	1.4762	285	0.6162	0.3797	0.0488	123.7047	0.2740
15:15s:20s:1l	545	1.7302	38	-0.1187	0.0141	0.0582	147.4216	-0.0310
15:20s:15s:1l	615	1.9524	438	0.5502	0.3027	0.0507	128.3547	0.2184
15:20s:20s:1l	720	2.2857	60	0.6454	0.4165	0.0455	115.1301	0.3712

<sup>1</sup>Datos escalados en el rango [0,1]

En el caso de las variables de tipo químico se observa una ligera mejoría no significativa tanto en el coeficiente de correlación como en el error estándar de predicción

para la red neuronal 10:15s:20s:11. A pesar de ello, la cantidad de varianza explicada durante el proceso de validación es tan solo del 6.76 %. El bajo nivel de ajuste entre los valores estimados y esperados hace que los términos de error nuevamente se encuentren por encima de lo estadísticamente aceptable. Así, el error estándar de predicción se encuentra en todo momento alrededor del 135 % mientras que el coeficiente de eficiencia se centra en torno al cero, alcanzándose en algunas ocasiones valores negativos.

Algo parecido sucede con la estimación de la concentración de amoníaco a partir de la cantidad de pienso suministrado, la carga de amoníaco en el biofiltro y la tasa de eliminación de amoníaco. En esta ocasión, ninguna de las redes neuronales supera en ninguno de sus términos a la regresión múltiple, aunque las diferencias entre regresión múltiple y red neuronal no fueron significativas (test R,  $p_{\alpha}=0.2615$ ). De esta forma, el mejor resultado con las redes neuronales es alcanzado con la red 15:20s:20s:11 donde el coeficiente de determinación fue del 0.4165, mientras que con la regresión múltiple fue de 0.4650. De igual forma, los términos de error son muy parecidos aunque ligeramente mejores en la regresión múltiple. En ambos casos los errores se centran alrededor del 120 % y los niveles de eficiencia no superan en ningún caso el 0.45.

#### 4.3.2.6. *Amoníaco en la serie 'A'. Estimación a partir de la serie filtrada*

Como se comentó en el apartado 4.2.5, la presencia de ruido en la serie de datos originales, puede originar un bajo nivel de ajuste entre los datos observados y los generados mediante los modelos de predicción. Este fenómeno se ha podido observar con las calibraciones y estimaciones llevadas a cabo con el parámetro  $\text{NH}_3$  en la serie 'A'. Para intentar corregir ese defecto, se ha procedido a filtrar previamente los datos originales antes de ser utilizados en la calibración del modelo.

De este modo utilizando los mismos criterios de calibración y validación, los resultados obtenidos son significativamente mejores en comparación con los conseguidos con los datos brutos y las predicciones son estadísticamente aceptables para todos los modelos.

Para la regresión múltiple con las variables independientes las concentraciones de amoníaco en los cinco días previos (tabla IV.17), la cantidad de varianza explicada tanto en calibración como en validación se encuentra por encima del 95 %, es decir, unos 40 puntos en calibración y 55 en validación mejor que cuando los datos utilizados eran los datos brutos. Sin embargo, el aspecto más significativo se obtiene en los términos de error alcanzados en la validación. Mientras que en el caso de los datos brutos los errores para todos los modelos estaban muy por encima del 100 %, y los niveles de eficiencia eran inferiores a 0.5, con los datos filtrados encontramos que el error estándar de predicción disminuye hasta el 21.3 % mientras que la eficiencia se sitúa en un 0.95.

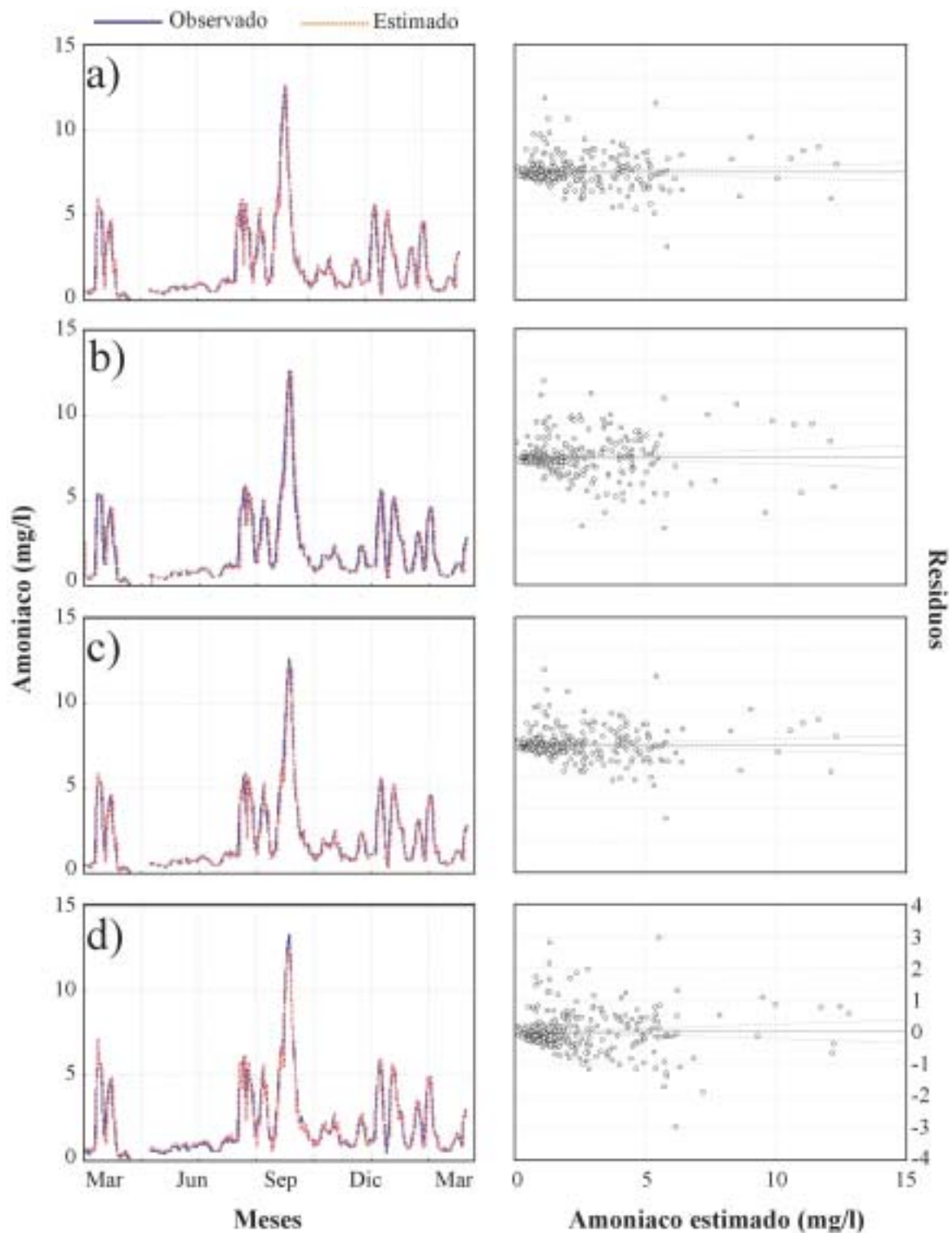
**Tabla IV.17.** Estimación del  $\text{NH}_3$  del agua en la serie 'A' mediante el análisis de regresión múltiple con las variables independientes representadas por el  $\text{NH}_3$  filtrado en los cinco días anteriores

<b>Resumen de la regresión múltiple en la calibración del modelo</b>			
R=0.9891; $R^2=9783$ ; F(5;866)=3031.6; $p_\alpha < 0.001$ ; N=872			
Variable dependiente	Variable independiente	$b_i(i=0,1,\dots,q)$	$p_\alpha$
NH <sub>3</sub> (t)	Intersección ( $b_0$ )	0.0639	0.0709
	NH <sub>3</sub> (t-1)	1.5842	0**
	NH <sub>3</sub> (t-2)	-0.6600	0**
	NH <sub>3</sub> (t-3)	0.0863	0.4207
	NH <sub>3</sub> (t-4)	-0.0344	0.7356
	NH <sub>3</sub> (t-5)	-0.0069	0.8983
<b>Resumen de la regresión en la validación del modelo</b>			
R=0.9782; $R^2=0.9568$ ; F(1;376)=8335.1; $p_\alpha < 0.001$ ; N=378			
Error estándar de predicción= 21.3099 %; Coeficiente E=0.9545			
Variable dependiente	Variable independiente	$b_i(i=0,1,\dots,q)$	$p_\alpha$
NH <sub>3</sub> observado (t)	Intersección ( $b_0$ )	0.5531	0.0725
	NH <sub>3</sub> estimado (t)	1.0186	0**

\*  $p_\alpha < 0.05$ ; \*\*  $p_\alpha < 0.01$

Cuando de la regresión anterior sólo se seleccionan las variables significativas, es decir, la concentración de amoníaco en los dos días previos, y se procede nuevamente a la calibración y validación del modelo, los resultados obtenidos mejoran ligeramente, aunque desde un punto de vista estadístico los resultados son similares (Calibración: R=0.9890;  $R^2=0.9782$ ; F(2;880)=7626.6;  $p_\alpha < 0.001$ ; N=882. Validación: R=0.9785;  $R^2=0.9574$ ; F(1;376)=8450.9;  $p_\alpha < 0.001$ ; N=378; SEP=20.7449; Coeficiente E=0.9569 (Figura IV.20).

Los modelos de suavizado exponencial simple y suavizado lineal exponencial o de Holt, también han proporcionado resultados bastante satisfactorios. De entre estos dos modelos, el que ha obtenido mejores resultados tanto en calibración como en validación ha sido el suavizado lineal exponencial con un total de varianza explicada del 95.12 %, un error estándar de predicción del 24.13 % y un coeficiente de eficiencia por encima de 0.94. (Tabla IV.18) (Figura IV.20). Asimismo, todos los modelos **ARIMA** calibrados excepto uno, han proporcionado resultados similares a los conseguidos con las regresiones múltiples y los modelos de suavizado en la validación. En esta ocasión, los modelos con todos sus términos significativos que presentan los mejores comportamientos durante el proceso de validación son el ARIMA(1,1,0) y ARIMA(2,1,1) (Figura IV.20). El primero de ellos es el modelo que explica una mayor cantidad de varianza de la variable dependiente ( $R^2=0.9572$ ) y, por lo tanto, presenta una mayor correlación entre valores observados y estimados. El ARIMA(2,1,1) en cambio, ha presentado los mejores términos de error para el conjunto de todos los modelos (SEP=20.7257 y E=0.9569) (Tabla IV.18).



**Figura IV.20.** Concentración de amoniaco observado y esperado, y residuos para los modelos: a) Regresión múltiple; b) Suavizado lineal exponencial; c) ARIMA(2,1,1) y d) Red neuronal 5:10s:15s:1l. En todos los casos los datos utilizados para la calibración del modelo han sido previamente filtrados

**Tabla IV.18.** Parámetros de estimación y bondad de ajuste de los modelos de series temporales en el NH<sub>3</sub> filtrado de la serie 'A'

Modelo	Parámetros	Validación			
		R <sup>2</sup>	F	%SEP	Coef. E
Suavizado exponencial simple	S <sub>0</sub> =2.5940; α <sub>d</sub> =1	0.9328	(1;376)=5218.9	26.1196	0.9316
Suavizado lineal exponencial	S <sub>0</sub> =0.6437; T <sub>0</sub> =-0.0020 α <sub>d</sub> =1; β <sub>d</sub> =0.3800	0.9512	(1;375)=7311.5	24.1336	0.9416
ARIMA (0,0,1)	Θ <sub>1</sub> =-0.9254**	0.8363	(1;376)=1920.3	76.5036	0.4134
ARIMA (1,0,0)	Φ <sub>1</sub> =0.9849**	0.9358	(1;376)=5482.0	25.4364	0.9352
ARIMA (0,1,1)	Θ <sub>1</sub> =-0.4725**	0.9527	(1;375)=7552.0	21.9946	0.9515
ARIMA (1,0,1)	Φ <sub>1</sub> =0.9751**; Θ <sub>1</sub> =-0.4766**	0.9527	(1;376)=7569.9	21.8918	0.9520
ARIMA (1,1,0)	Φ <sub>1</sub> =0.5539**	0.9572	(1;375)=8379.6	21.2352	0.9548
ARIMA (1,1,1)	Φ <sub>1</sub> =0.4906**; Θ <sub>1</sub> =-0.0918	0.9570	(1;375)=8362.1	21.2060	0.9549
ARIMA (1,1,2)	Φ <sub>1</sub> =0.4174; Θ <sub>1</sub> =-0.1673 Θ <sub>2</sub> =-0.0504	0.9570	(1;375)=8344.4	21.2020	0.9549
ARIMA (1,2,1)	Φ <sub>1</sub> =0.2906**; Θ <sub>1</sub> =-0.9946**	0.7986	(1;372)=1474.7	48.7636	0.7617
ARIMA (2,2,1)	Φ <sub>1</sub> =0.4357**; Θ <sub>1</sub> =-0.9904** Φ <sub>2</sub> =-0.4927	0.9444	(1;372)=6312.7	25.8011	0.9333
ARIMA (2,1,2)	Φ <sub>1</sub> =1.5318**; Θ <sub>1</sub> =0.9956** Φ <sub>2</sub> =-0.5820**; Θ <sub>1</sub> =-0.0053	0.9571	(1;375)=8363.4	20.7257	0.9570
ARIMA (2,1,1)	Φ <sub>1</sub> =1.5282**; Θ <sub>1</sub> =0.9903** Φ <sub>2</sub> =-0.5785**	0.9571	(1;375)=8363.4	20.7257	0.9569
ARIMA (1,2,2)	Φ <sub>1</sub> =-0.1454**; Θ <sub>1</sub> =-1.8701** Θ <sub>2</sub> =-0.8769**	0.9469	(1;372)=6630.8	24.8852	0.9379
ARIMA (2,2,2)	Φ <sub>1</sub> =-0.1084**; Θ <sub>1</sub> =-1.7963** Φ <sub>2</sub> =-0.2925**; Θ <sub>2</sub> =-0.8047**	0.9510	(1;372)=7217.2	24.0028	0.9423

\* p<sub>α</sub><0.05; \*\* p<sub>α</sub><0.01

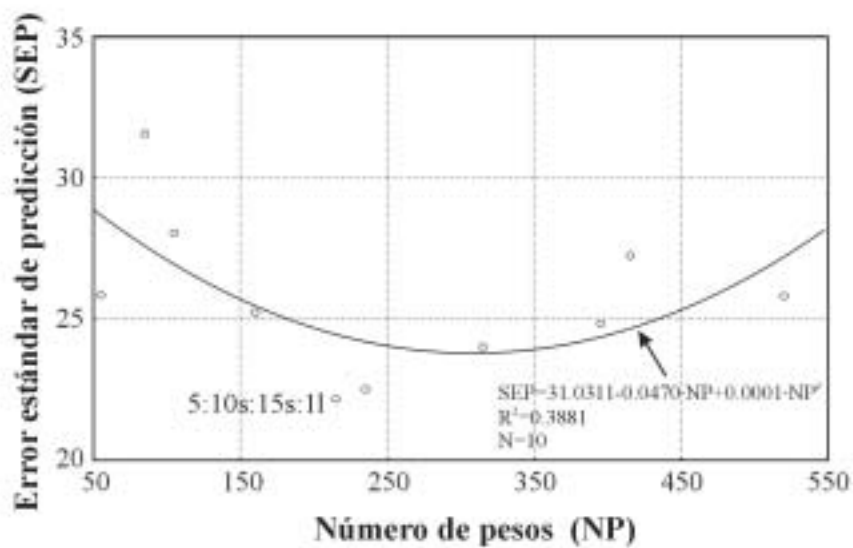
Con los datos filtrados, las redes neuronales también estiman correctamente la concentración de amoníaco aunque, tanto la cantidad de varianza explicada como los términos de error en la mejor red neuronal son ligeramente inferiores (Red 5:10s:15s:11: R<sup>2</sup>=0.9508; SEP=22.1665; E=0.9508) (Tabla IV.19) (Figura IV.20). Para el conjunto de todas las redes neuronales entrenadas, la regresión simple entre los términos de error y el número de pesos de la red indica la no existencia de correlación significativa (SEP: R=0.2752; p<sub>α</sub>=0.4415; N=10. Coeficiente E: R=0.2953; p<sub>α</sub>=0.4075; N=10). Sin embargo, el porcentaje de varianza explicada aumenta hasta un 38.8 % cuando se relacionan ambas variables a través de una regresión polinómica de segundo grado (Figura IV.21). También se han encontrado marcadas diferencias significativas entre el error estándar de predicción en función del número de pesos y el error mínimo teórico estimado durante la calibración (test de comparación de dos rectas de regresión: F<sub>exp</sub>=41.8180; t<sub>exp</sub>=559.4084; t<sub>(0.001)</sub>(2611)<3.2910; ANCOVA, p<sub>α</sub><0.001).



**Tabla IV.19.** Resultados de calibración y generalización de las RNCs en las que las variables de entrada son las concentraciones de amoníaco filtrado en el sistema 'A' en los cinco días anteriores. Parada del entrenamiento mediante validación cruzada (50 patrones) y un coeficiente de determinación umbral igual a 0.99

Red	Pesos	Calibración		Generalización				
		$\epsilon$	Épocas	R	R <sup>2</sup>	RMS <sup>1</sup>	%SEP	E
5:5s:5s:11	55	0.0626	55	0.9696	0.9402	0.0333	25.8434	0.9331
5:5s:10s:11	85	0.0968	789	0.9680	0.9369	0.0407	31.5544	0.9002
5:10s:5s:11	105	0.1196	7930	0.9671	0.9354	0.0361	28.0585	0.9211
5:10s:10s:11	160	0.1822	6836	0.9700	0.9409	0.0325	25.2448	0.9361
5:10s:15s:11	215	0.2449	1975	0.9751	0.9508	0.0286	22.1664	0.9508
5:15s:10s:11	235	0.2677	976	0.9753	0.9511	0.0290	22.5176	0.9492
5:15s:15s:11	315	0.3588	1286	0.9741	0.9488	0.0309	24.0225	0.9422
5:15s:20s:11	395	0.4499	756	0.9729	0.9466	0.0320	24.8347	0.9382
5:20s:15s:11	415	0.4727	148	0.9689	0.9389	0.0351	27.2398	0.9256
5:20s:20s:11	520	0.5923	713	0.9688	0.9385	0.0333	25.8179	0.9332

<sup>1</sup>Datos escalados en el rango [0,1]



**Figura IV.21.** Regresión polinómica de segundo grado entre el error estándar de predicción (SEP) alcanzado durante la validación y el número de pesos de la red neuronal. Se indica la ecuación de la regresión, su coeficiente de determinación, el número de datos así como la arquitectura neuronal con la que se obtienen los mejores resultados

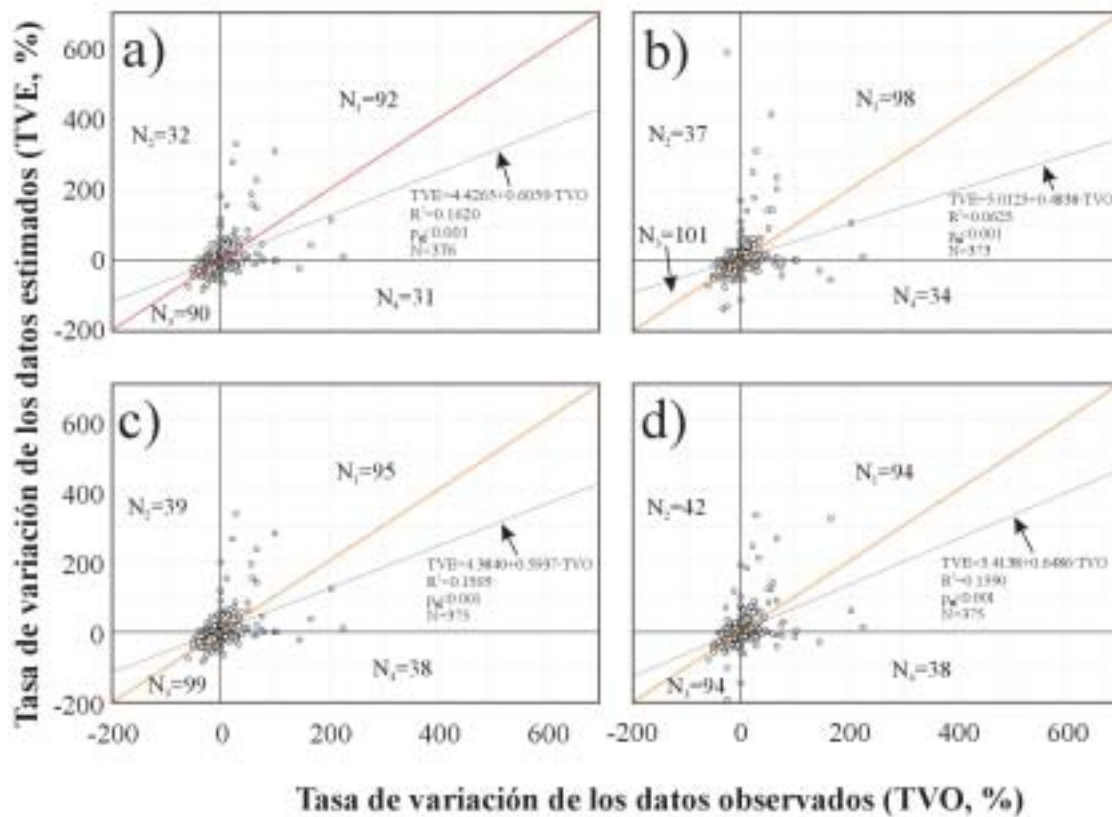
La semejanza en la predicción por parte de los cuatro mejores modelos se pone de manifiesto a través del análisis del coeficiente de desigualdad de Theil y del análisis de los diagramas de predicción-realización (Tabla IV.20) (Figura IV.22). El valor más cercano a cero del coeficiente  $U$  o coeficiente de desigualdad de Theil lo encontramos en el modelo **ARIMA** ( $U_{\text{ARIMA}}=0.5690$ ). Por el contrario, el peor resultado es obtenido con el modelo de suavizado de Holt ( $U_{\text{HOLT}}=0.8462$ ), de tal forma que se encontraron diferencias significativas entre el coeficiente  $U$  de este modelo y el resto.

**Tabla IV.20.** Valores de cada componente de error del coeficiente de desigualdad de Theil

	Coeficiente de Theil ( $U$ )	Error sistemático ( $U_s$ )	Error de dispersión ( $U_d$ )	Error de correlación ( $U_c$ )
<b>Regresión múltiple</b>	0.5692	0.0447	0.2009	0.5331
<b>Suavizado de Holt</b>	0.8462	-0.0175	0.6383	0.5620
<b>ARIMA(2,1,1)</b>	0.5690	0.0436	0.1954	0.5349
<b>RNC 5:10s:15s:11</b>	0.5776	0.0291	0.2317	0.5328

Por otra parte, el mejor error sistemático o componente  $U_s$  del coeficiente de Theil fue encontrado en la red neuronal ( $U_s=0.0291$ ), denotando este modelo una mayor tendencia a la estimación de valores promedio que el resto. En este caso, tanto la regresión múltiple como el modelo **ARIMA** tienden a proporcionar una supervaloración de las tasas de cambio en la serie, mientras que el suavizado de Holt proporciona una infravaloración de dichas tasas. En cambio, los valores más correctos del error de dispersión o componente  $U_d$  del coeficiente de Theil es encontrado en el modelo **ARIMA** ( $U_d=0.1954$ ). De esta forma, el modelo **ARIMA** es el que presenta una menor dispersión de las tasas de cambio alrededor de sus valores promedio. A pesar de ello, no se encontraron diferencias significativas con los errores de dispersión de la regresión múltiple y la red neuronal. El peor valor del error de dispersión corresponde al modelo de Holt ( $U_s=0.6383$ ). Nuevamente, el mejor valor para el error de correlación o componente  $U_c$  del coeficiente de Theil es el de la red neuronal ( $U_c=0.5328$ ), seguido de la regresión múltiple, el modelo **ARIMA** y finalmente el suavizado de Holt. Este error de correlación indica una mayor correlación significativa entre las tasas de cambio en la serie original y la serie estimada por la red neuronal.

Asimismo, el análisis de los diagramas de predicción-realización muestran la ausencia de diferencias significativas entre las distribuciones de frecuencias absolutas por cuadrantes para los cuatro modelos (Tabla IV.21).



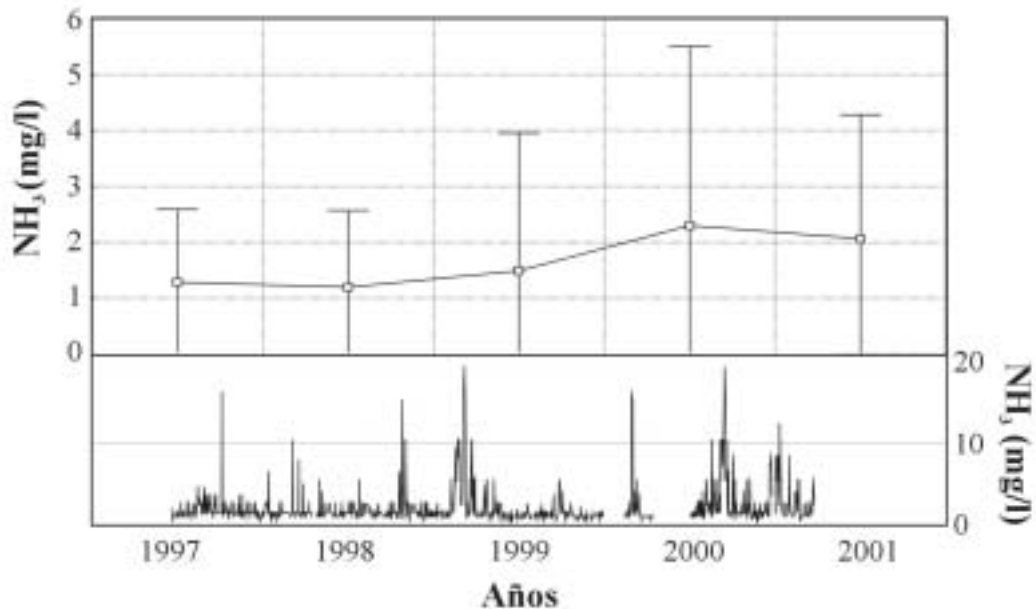
**Figura IV.22.** a) Tasa de variación de los datos observados frente a la tasa de variación de los datos estimados por: a) la regresión múltiple con variables independientes la concentración de amoníaco filtrada en los dos días previos; b) suavizado de Holt; c) ARIMA(2,1,1) y d) red neuronal 5:10s:15s:1l que considera como variables de entrada la concentración de amoníaco filtrada en los cinco días previos. Se indica el número de casos por cuadrante ( $N_1$ ,  $N_2$ ,  $N_3$  y  $N_4$ ) así como la ecuación de regresión, el coeficiente de determinación, el nivel de significación y el número de datos. La línea roja punteada indica la recta de correlación 1 que pasa por el origen de coordenadas

**Tabla IV.21.** Resultados de los test  $\chi^2$  entre las frecuencias por cuadrantes en los diagramas de Theil para los cuatro mejores modelos seleccionados. Se indica el valor del estadístico  $\chi^2$  así como el nivel del significación del test entre paréntesis

	Suavizado de Holt	ARIMA(2,1,1)	Red neuronal 5:10s:15s:1l
Regresión múltiple	0.2327 (0.9721)	1.1799 (0.7578)	2.3102 (0.5106)
Suavizado de Holt	----	0.6599 (0.8826)	1.6865 (0.6399)
ARIMA(2,1,1)		----	0.4522 (0.9293)

#### 4.3.2.7. Amoniac en las series 'B', 'E' y 'F'

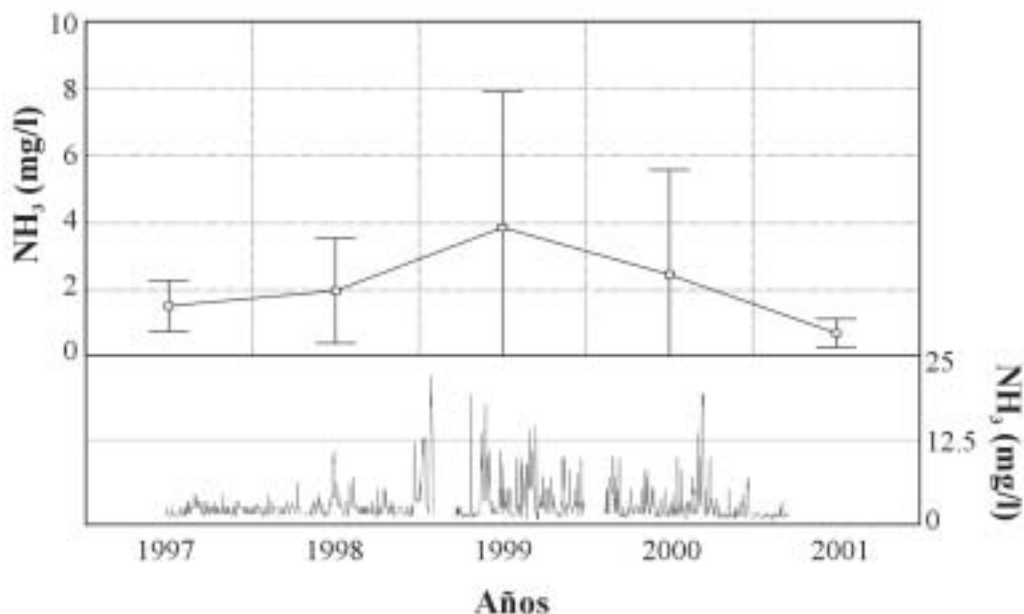
La evolución temporal de la concentración de amoniac en la serie 'B' se muestra en la figura IV.23. Se puede observar como existe un patrón de comportamiento distinto al encontrado en la serie 'A'. Mientras en esta última se producía un incremento significativo de las concentraciones de amoniac a partir del año 1999, en la serie 'B' el incremento significativo se produce en el año 2000 (ANOVA:  $F=10.954$ ;  $p_{\alpha}<0.001$ ), manteniéndose el mismo nivel durante los primeros meses del 2001 (test de Tukey:  $p_{\alpha}=0.9262$ ). En cambio, en las series 'E' y 'F' (figura IV.24), se vuelve a repetir la situación encontrada en la serie 'A' con importantes incrementos de las concentraciones de amoniac en el año 1999 (ANOVA:  $F=36.954$ ;  $p_{\alpha}<0.001$ ), aunque con un descenso significativo en el año 2000 y primeros meses del 2001 (test Tukey:  $p_{\alpha(1999-2000)}<0.001$ ;  $p_{\alpha(1999-2001)}<0.001$ ).



**Figura IV.23.** Valores medios anuales de la concentración de  $NH_3$  presente en el agua de la serie 'B' (gráfica superior). Evolución temporal a lo largo del periodo de estudio (valores absolutos) en la serie 'B' (gráfica inferior)

Hay correlaciones significativas entre las concentraciones de amoniac en la serie 'A' y las series 'B', 'E' y 'F' (Serie 'A'-Serie 'B':  $R=0.5429$ ;  $p_{\alpha}<0.05$ . Serie 'A'-Serie 'E' y 'F':  $R=0.2649$ ;  $p_{\alpha}<0.05$ ), por lo que se ha procedido a la evaluación de los modelos seleccionados en la serie 'A' con el mismo periodo de validación en las series 'B', 'E' y 'F'.

De esta forma, los resultados obtenidos son tan satisfactorios como los encontrados durante la validación con la serie 'A' (tabla IV.22).



**Figura IV.24.** Valores medios anuales de la concentración de NH<sub>3</sub> presente en el agua de las series 'E' y 'F' (gráfica superior). Evolución temporal a lo largo del periodo de estudio (valores absolutos) en las series 'E' y 'F' (gráfica inferior)

**Tabla IV.22.** Estimación de la concentración de amoníaco en el agua en las series 'B', 'E' y 'F' a partir de los mejores modelos en validación calibrados en la serie 'A'

	Serie 'B'				
	R	R <sup>2</sup>	RMS	SEP	E
Regresión múltiple	0.9840	0.9682	0.4137	18.1548	0.9681
Suavizado de Holt	0.9771	0.9547	0.5424	23.9283	0.9446
ARIMA(2,1,1)	0.9857	0.9716	0.4252	18.6563	0.9663
RNC 5:10s:15s:11	0.9831	0.9664	0.4228	18.5569	0.9667
	Series 'E' y 'F'				
Regresión múltiple	0.9803	0.9610	0.4277	18.7714	0.9609
Suavizado de Holt	0.9775	0.9555	0.5038	22.1086	0.9458
ARIMA(2,1,1)	0.9806	0.9618	0.4220	18.5212	0.9620
RNC 5:10s:15s:11	0.9788	0.9580	0.3987	17.4983	0.9660

A pesar de que en ningún caso se han encontrado diferencias significativas en cuanto a la cantidad de varianza explicada (test R:  $p_{\alpha} > 0.05$  en todos los casos) tanto para la validación en la serie 'B' como en las series 'E' y 'F', hay pequeñas variaciones en cuanto al comportamiento de los distintos modelos con respecto al parámetro de evaluación. Así, con respecto al coeficiente de determinación, los mejores resultados los encontramos en ambos casos en el modelo ARIMA(2,1,1). Por otra parte cuando el parámetro de evaluación es el error estándar de predicción, los mejores resultados son los conseguidos por la regresión múltiple en la serie 'B' y por la red neuronal en las series 'E' y 'F'. El mismo resultado se consigue cuando se tiene en cuenta el coeficiente de eficiencia (E).

El análisis del coeficiente de Theil para los valores estimados en la serie 'B' (tabla IV.23) muestra una concentración de sus mejores componentes en los modelos de series temporales. Así, el coeficiente U con el valor más cercano a cero y la menor dispersión de las tasas de variación es el calculado para el suavizado de Holt. Por otra parte, los valores medios de las tasas de variación más próximas a las reales así como el menor error de correlación es proporcionado por el modelo ARIMA.

**Tabla IV.23.** Valores de cada componente de error del coeficiente de desigualdad de Theil en la serie 'B'

	<b>Coefficiente de Theil (U)</b>	<b>Error sistemático (U<sub>s</sub>)</b>	<b>Error de dispersión (U<sub>d</sub>)</b>	<b>Error de correlación (U<sub>c</sub>)</b>
<b>Regresión múltiple</b>	0.9141	-0.0420	0.8269	0.4017
<b>Suavizado de Holt</b>	0.6158	0.0469	0.3280	0.5774
<b>ARIMA(2,1,1)</b>	0.9345	0.0761	0.9038	0.2560
<b>RNC 5:10s:15s:11</b>	0.9552	-0.055	0.9379	0.3036

En el caso del análisis del coeficiente de Theil en las series 'E' y 'F' los resultados difieren ligeramente (tabla IV.24). El coeficiente U más cercano a cero y el componente de dispersión más bajo se corresponde con la red neuronal ( $U=0.6134$ ;  $U_d=0.3254$ ). Por otra parte, el mejor error sistemático y el mejor error de correlación son los del suavizado de Holt.

**Tabla IV.24.** Valores de cada componente de error del coeficiente de desigualdad de Theil en las series 'E' y 'F'

	<b>Coefficiente de Theil (U)</b>	<b>Error sistemático (U<sub>s</sub>)</b>	<b>Error de dispersión (U<sub>d</sub>)</b>	<b>Error de correlación (U<sub>c</sub>)</b>
<b>Regresión múltiple</b>	0.8026	0.0688	0.6001	0.5333
<b>Suavizado de Holt</b>	0.9286	-0.0357	0.8585	0.3647
<b>ARIMA(2,1,1)</b>	0.7674	0.0634	0.5262	0.5605
<b>RNC 5:10s:15s:11</b>	0.6134	0.0653	0.3254	0.6277

En el análisis de los diagramas de predicción-realización de la serie 'B' se han encontrado diferencias significativas entre el número de casos por cuadrante en el modelo **ARIMA** y el resto (tabla IV.25). En todos los casos, la mayor contribución al valor del estadístico  $\chi^2$  se debe al número de casos contenido en el segundo cuadrante que es significativamente menor en el modelo **ARIMA**, indicando que éste tiende a proporcionar menores errores de signos en las tasas de cambio. En las series 'E' y 'F', el test  $\chi^2$  realizado sobre las frecuencias por cuadrante, no reflejó en ningún caso la presencia de diferencias significativas (tabla IV.26).

**Tabla IV.25.** Resultados de los test  $\chi^2$  entre las frecuencias por cuadrantes en los diagramas de Theil de la serie 'B' para los cuatro modelos seleccionados. Se indica el valor del estadístico  $\chi^2$  así como el nivel de significación del test entre paréntesis. \* Porcentaje de contribución al estadístico  $\chi^2$  del segundo cuadrante

	<b>Suavizado de Holt</b>	<b>ARIMA(2,1,1)</b>	<b>Red neuronal 5:10s:15s:11</b>
<b>Regresión múltiple</b>	1.4474 (0.8359)	23.7972 (<0.001) 72.6686 %*	1.6038 (0.8081)
<b>Suavizado de Holt</b>	----	14.4674 (0.0060) 82.7640 %*	4.9438 (0.2931)
<b>ARIMA(2,1,1)</b>		----	21.6204 (0.0002) 50.8335 %*

**Tabla IV.26.** Resultados de los test  $\chi^2$  entre las frecuencias por cuadrantes en los diagramas de Theil de las series 'E' y 'F' para los cuatro modelos seleccionados. Se indica el valor del estadístico  $\chi^2$  así como el nivel de significación del test entre paréntesis

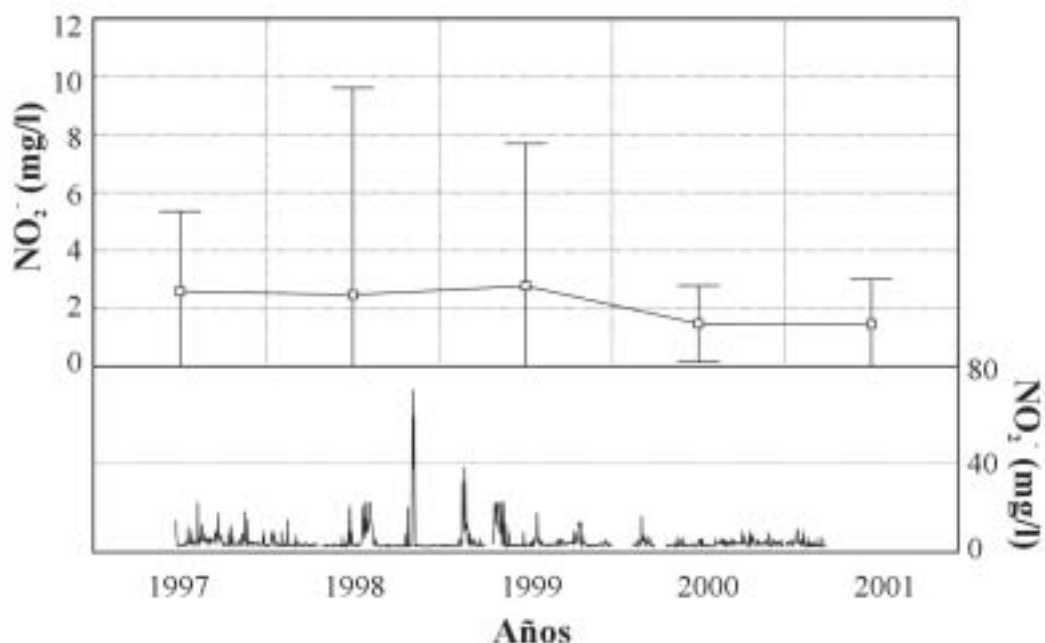
	Suavizado de Holt	ARIMA(2,1,1)	Red neuronal 5:10s:15s:11
<b>Regresión múltiple</b>	1.5358 (0.6740)	3.5389 (0.3157)	1.7164 (0.6333)
<b>Suavizado de Holt</b>	----	1.8605 (0.6019)	1.8789 (0.5979)
<b>ARIMA(2,1,1)</b>		----	0.5878 (0.8992)

#### 4.3.2.8. Nitritos en la serie 'A'

La evolución temporal de la concentración de nitritos en la serie 'A' se muestra en la figura IV.25. Los valores medios anuales de la concentración de  $\text{NO}_2^-$  en la serie 'A' se mantienen relativamente constantes a lo largo del período de estudio. A pesar de ello, un análisis de la varianza mostró la presencia de diferencias significativas (ANOVA:  $F=4.0341$ ;  $p_\alpha < 0.05$ ). Un test de Tukey posterior indicó que estas diferencias se deben fundamentalmente a un descenso significativo de la concentración de  $\text{NO}_2^-$  a partir del año 1999 (test de Tukey:  $p_\alpha (1999-2000) < 0.05$ ).

La calibración de la ecuación de regresión que relaciona la concentración de nitritos en un día determinado con la presente en los cinco días anteriores a dado como resultado un buen ajuste tanto en el proceso de calibración como en validación (tabla IV.27). Así, la cantidad de varianza de la variable dependiente explicada por las variables independientes supera el 98 % en la calibración y el 91 % en la validación. También los términos de error en la validación externa alcanzan niveles estadísticamente aceptables, superando en esta ocasión incluso a los alcanzados por los mejores modelos calibrados para el parámetro amoniaco en esta serie. De esta forma, el error estándar de predicción está ligeramente por encima del 18 % mientras que el coeficiente de eficiencia es superior a 0.9. Sin embargo en esta ocasión el análisis de los residuos ha dado como resultado que el estadístico de Durvin-Watson ( $D=1.7603$ ) se encuentre en la región 'D', indicando un desplazamiento de la autocorrelación hacia la zona de correlación serial positiva (Región 'E'). La descomposición de la serie de residuos en sus frecuencias fundamentales muestra la presencia de dos módulos con valores superiores a 1 para frecuencias alrededor de 0.33, lo que se corresponde con un período de aproximadamente 3 días.





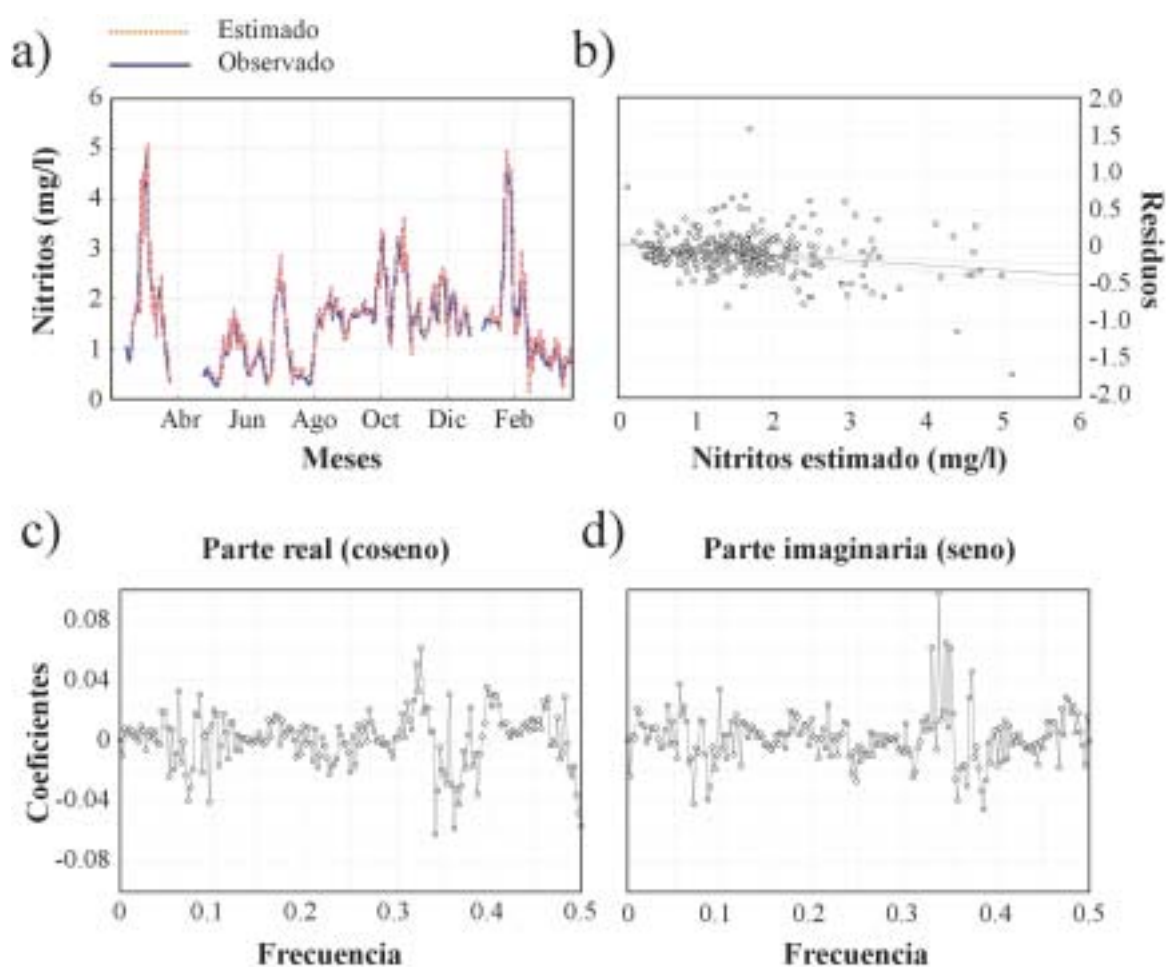
**Figura IV.25.** Valores medios anuales de la concentración de  $\text{NO}_2^-$  presente en el agua de la serie 'A' (gráfica superior). Evolución temporal a lo largo del periodo de estudio (valores absolutos) en la serie 'A' (gráfica inferior)

**Tabla IV.27.** Estimación del  $\text{NO}_2^-$  del agua en la serie 'A' mediante el análisis de regresión múltiple con las variables independientes representadas por el  $\text{NO}_2^-$  filtrado en los cinco días anteriores

<b>Resumen de la regresión múltiple en la calibración del modelo</b>			
R=0.9911; R <sup>2</sup> =0.9822; F(5;344)=3792.3; p <sub>α</sub> <0.001; N=350			
Variable dependiente	Variable independiente	b <sub>i</sub> (i=0,1,...,q)	p <sub>α</sub>
NO <sub>2</sub> <sup>-</sup> (t)	Intersección (b <sub>0</sub> )	0.1421	0**
	NO <sub>2</sub> <sup>-</sup> (t-1)	1.6189	0**
	NO <sub>2</sub> <sup>-</sup> (t-2)	-0.7929	0**
	NO <sub>2</sub> <sup>-</sup> (t-3)	-0.0174	0.8664
	NO <sub>2</sub> <sup>-</sup> (t-4)	0.6109	0**
	NO <sub>2</sub> <sup>-</sup> (t-5)	-0.4716	0**
<b>Resumen de la regresión en la validación del modelo</b>			
R=0.9550; R <sup>2</sup> =0.9120; F(1;343)=3560.5; p <sub>α</sub> <0.001; N=345			
Error estándar de predicción=18.1561 %; Coeficiente E=0.9018			
Variable dependiente	Variable independiente	b <sub>i</sub> (i=0,1,...,q)	p <sub>α</sub>
NO <sub>2</sub> <sup>-</sup> observado (t)	Intersección (b <sub>0</sub> )	0.0438	0.1235
	NO <sub>2</sub> <sup>-</sup> estimado (t)	0.9319	0**

\* p<sub>α</sub><0.05; \*\* p<sub>α</sub><0.01

La regresión múltiple en la que se elimina la variable independiente  $\text{NO}_2^-$  (t-3) también ofrece buenos resultados en ambas fases. La cantidad de varianza explicada en validación es la misma que en el caso anterior aunque el error estándar de predicción disminuye muy ligeramente. El mismo efecto lo encontramos con el coeficiente de eficiencia (Calibración:  $R=0.9910$ ;  $R^2=0.9822$ ;  $F(4;345)=4753.7$ ;  $p_\alpha < 0.001$ ;  $N=350$ . Validación:  $R=0.9550$ ;  $R^2=0.9121$ ;  $F(1;343)=3559.2$ ;  $p_\alpha < 0.001$ ;  $N=345$ ;  $\text{SEP}=18.1405$ ; Coeficiente  $E=0.9020$ ). También se detecta un aumento del estadístico de Durvin-Watson ( $D=1.7698$ ), a pesar de lo cual la autocorrelación residual se mantiene en la zona 'D' indeterminada y cercana a la correlación serial positiva (figura IV.26).



**Figura IV.26.** a) Concentración de nitritos observada y estimada del sistema 'A' en el periodo de validación, considerando en el análisis de regresión múltiple como variable dependiente la concentración de nitritos y como variables independientes la concentración de nitritos del primer, segundo, cuarto y quinto días anteriores; b) Residuos en el grupo de validación frente a la concentración de nitritos estimada; c) y d) Frecuencia de las funciones coseno y seno frente a los coeficientes encontrados en la transformada de Fourier

Los modelos univariantes de series temporales también han mostrado un comportamiento estadísticamente correcto (tabla IV.28). Tanto el suavizado exponencial simple como el suavizado de Holt explican más del 90 % de la varianza de la variable dependiente, aunque de la misma forma que ocurría con el parámetro amoniaco, el suavizado de Holt aporta unos términos de error estadísticamente mejores. Así, el error estándar de predicción en éste último se encuentra por debajo del 18 % mientras que la eficiencia en la predicción está por encima de 0.9 (figura IV.27). Entre los modelos **ARIMA** con todos sus términos significativos destaca el ARIMA(2,1,2) que rebaja el nivel de error hasta valores cercanos al 15 % con una eficiencia de 0.9278. A pesar de ello, la función de autocorrelación residual muestra correlación significativa para al menos 7 y 12 retrasos (figura IV.28). Este mismo efecto se muestra tanto en el suavizado lineal como en el suavizado de Holt para 1, 2, 7 y 8 retrasos.

**Tabla IV.28.** Parámetros de estimación y bondad de ajuste de los modelos de series temporales en el  $\text{NO}_2^-$  filtrado de la serie 'A'

Modelo	Parámetros	Validación			
		R <sup>2</sup>	F	%SEP	Coef. E
Suavizado exponencial simple	$S_0=2.6990; \alpha_d=1$	0.9001	(1;358)=3225.3	18.4692	0.8967
Suavizado lineal exponencial	$S_0=0.3995; T_0=0.0100$ $\alpha_d=1; \beta_d=0.3270$	0.9190	(1;357)=4048.4	17.7984	0.9040
ARIMA (0,0,1)	$\Theta_1=-0.8847^{**}$	0.6984	(1;358)=828.9	63.9181	-0.2377
ARIMA (1,0,0)	$\Phi_1=0.9894^{**}$	0.9131	(1;358)=3762.9	17.1017	0.9114
ARIMA (0,1,1)	$\Theta_1=-0.4725^{**}$	0.9256	(1;357)=4441.2	15.9047	0.9234
ARIMA (1,0,1)	$\Phi_1=0.9842^{**}; \Theta_1=-0.2805^{**}$	0.9219	(1;358)=4225.8	16.2779	0.9197
ARIMA (1,1,0)	$\Phi_1=0.3502^{**}$	0.9283	(1;357)=4625.3	15.7009	0.9253
ARIMA (1,1,1)	$\Phi_1=0.5073^{**}; \Theta_1=0.1793$	0.9292	(1;357)=4681.9	15.6664	0.9256
ARIMA (1,1,2)	$\Phi_1=0.4802^*; \Theta_1=0.1538$ $\Theta_2=-0.0154$	0.9291	(1;357)=4680.1	15.6654	0.9257
ARIMA (1,2,1)	$\Phi_1=0.1536^*; \Theta_1=-0.9883^{**}$	0.8881	(1;354)=2810.2	21.3527	0.8619
ARIMA (2,2,1)	$\Phi_1=0.2299^{**}; \Theta_1=-0.9802^{**}$ $\Phi_2=-0.4571^{**}$	0.9104	(1;354)=3598.1	18.9945	0.8907
ARIMA (2,1,2)	$\Phi_1=1.6063^{**}; \Theta_1=1.3293^{**}$ $\Phi_2=-0.6482^{**}; \Theta_1=-0.3335^{**}$	0.9296	(1;357)=4716.5	15.5716	0.9278
ARIMA (2,1,1)	$\Phi_1=0.3512; \Theta_1=0.0247$ $\Phi_2=0.0587$	0.9291	(1;357)=4678.8	15.6654	0.9257
ARIMA (1,2,2)	$\Phi_1=-0.2990^{**}; \Theta_1=-1.8559^{**}$ $\Theta_2=-0.8743^{**}$	0.9135	(1;354)=3736.5	18.5042	0.8963
ARIMA (2,2,2)	$\Phi_1=-0.3081^{**}; \Theta_1=-1.7713^{**}$ $\Phi_2=-0.3644^{**}; \Theta_2=-0.7964^{**}$	0.9235	(1;354)=4273.9	17.4522	0.9077

\*  $p_\alpha < 0.05$ ; \*\*  $p_\alpha < 0.01$

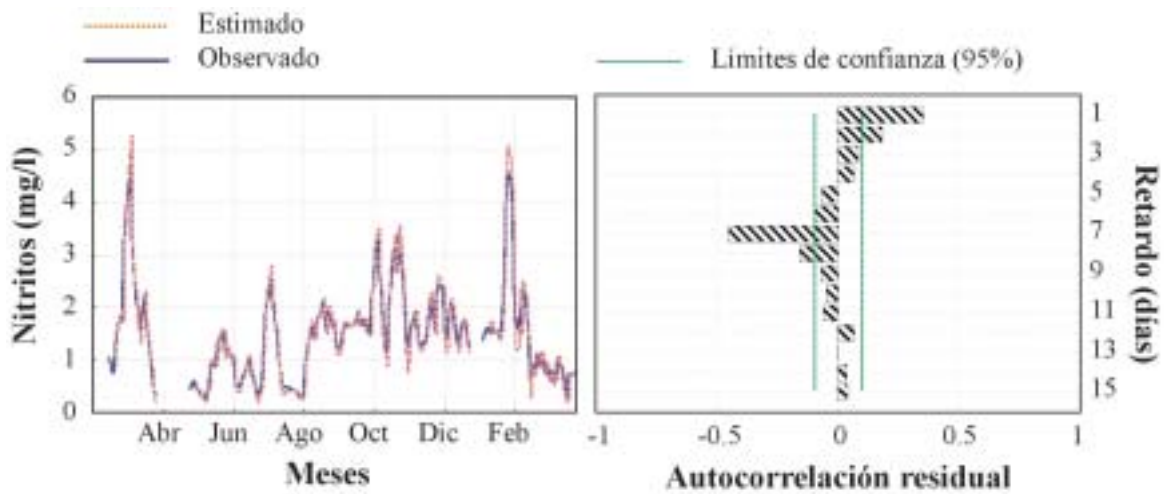


Figura IV.27. Concentración de nitritos observada y estimada en la serie 'A' y autocorrelación residual en validación para el modelo suavizado lineal exponencial (Holt)

Aunque no se han encontrado diferencias significativas entre las correlaciones obtenidas para cada uno de los mejores tipos de modelos (test R:  $p_{\alpha} > 0.05$  en todos los casos), la cantidad de varianza explicada por los modelos de redes neuronales computacionales ha sido inferior a la calculada para los modelos univariantes de series temporales, aunque superior a la encontrada en las regresiones múltiples (tabla IV.29). La red neuronal que ha mostrado un mejor comportamiento durante el proceso de validación ha sido la 5:10s:5s:11, con la que se ha cometido un error del 19.8297 % (figura IV.29). En este caso, la eficiencia de la predicción no llegó a 0.9 ( $E=0.8828$ ), quedándose este valor por debajo del calculado incluso para las regresiones múltiples. Al igual que en los modelos anteriores, tampoco se ha detectado periodicidad en la serie de residuos. Por otra parte, la regresión simple entre el número de parámetros de la red y el error alcanzado durante la validación mostró un nivel bajo de ajuste ( $R=0.2013$ ;  $R^2=0.0405$ ;  $F(1;8)=0.3$ ;  $p_{\alpha}=0.5770$ ;  $N=10$ ). El porcentaje de varianza explicada disminuye hasta un 18.2650 % cuando ambas variables se relacionan a través de una regresión polinómica de segundo grado, aunque con la concavidad contraria a la encontrada para el parámetro amoníaco (figura IV.30). También hay marcadas diferencias significativas entre el error teórico mínimo que se debería cometer durante la calibración y el error estándar de predicción en función del número de pesos (test de comparación de dos rectas de regresión:  $F_{exp}=65.7752$ ;  $t_{exp}=767.2115$ ;  $t_{(0.001)}(16770) < 3.2910$ ; ANCOVA,  $p_{\alpha} < 0.001$ ).

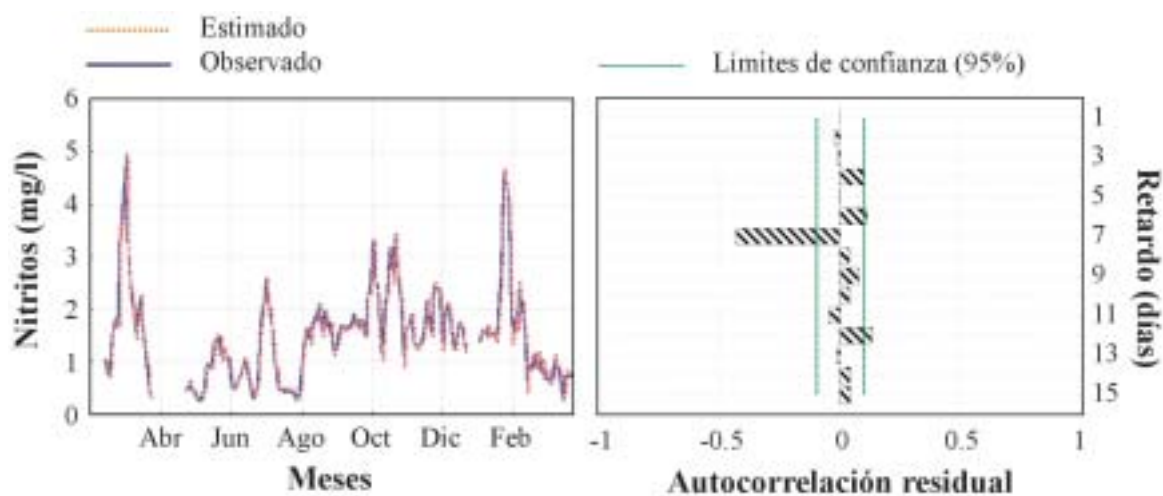
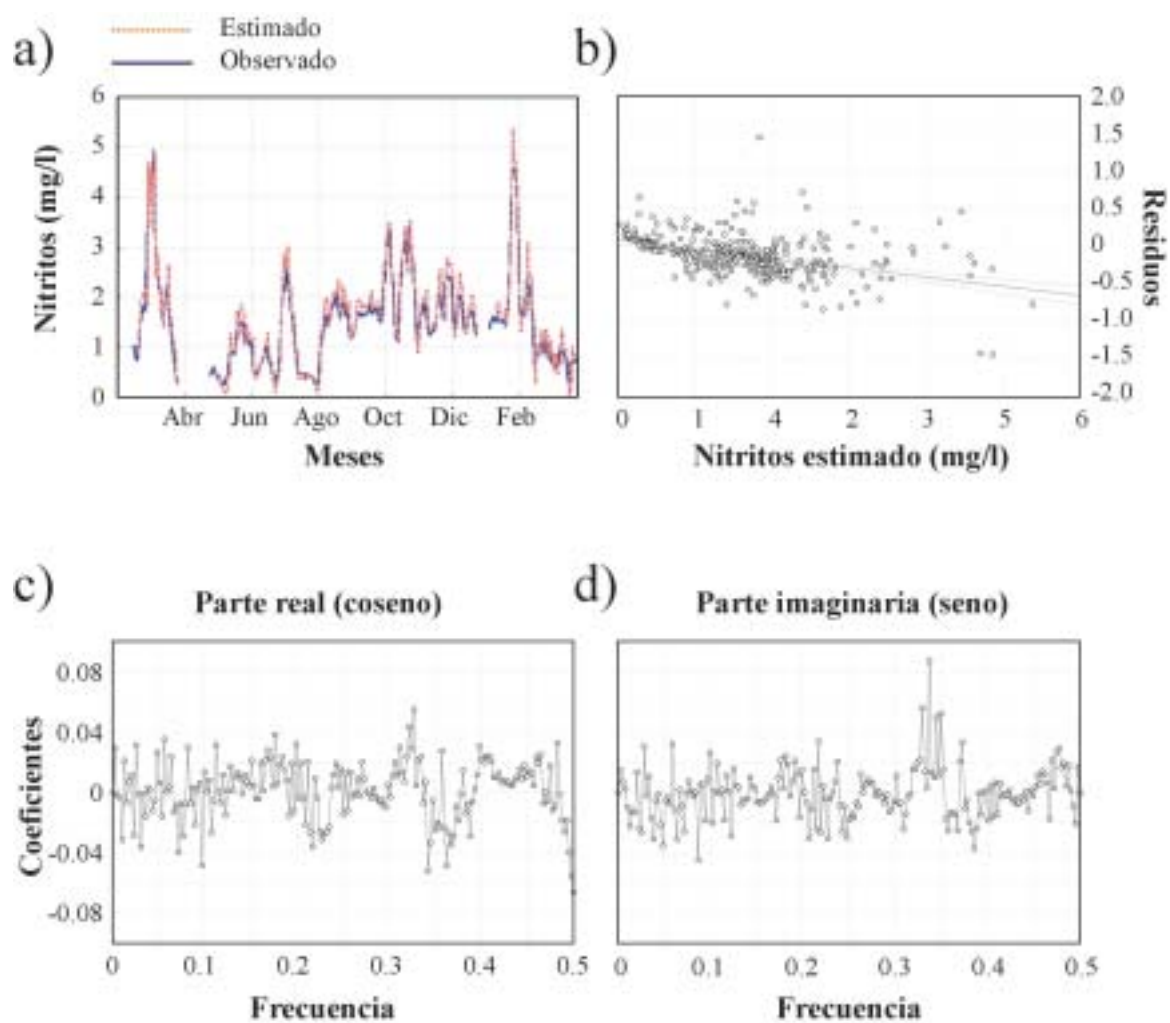


Figura IV.28. Concentración de nitritos observada y estimada en la serie 'A' y autocorrelación residual en validación para el modelo ARIMA(2,1,2)

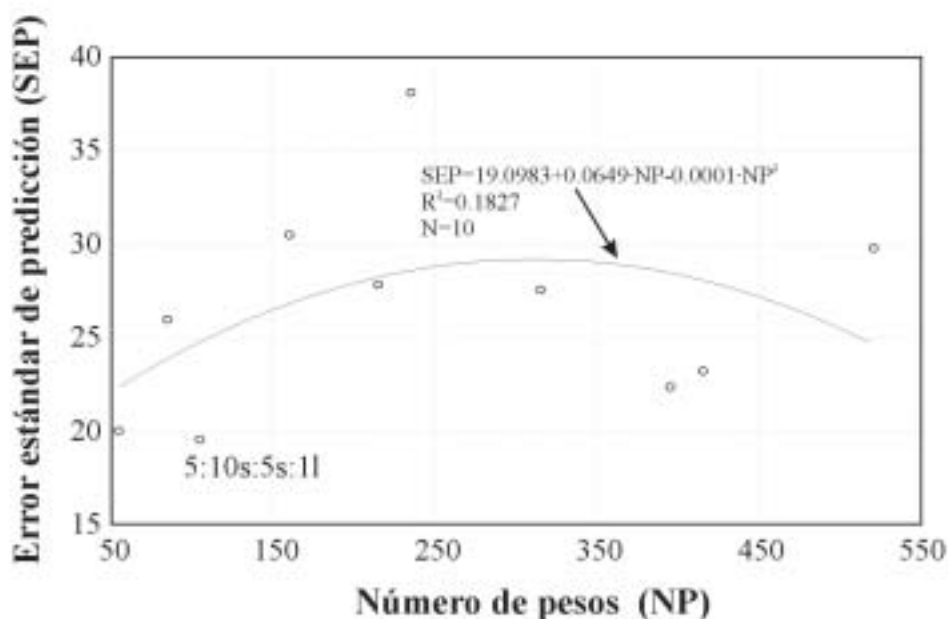
Tabla IV.29. Resultados de calibración y generalización de las RNCs con las variables de entrada las concentraciones de nitritos filtradas en el sistema 'A' en los cinco días anteriores. Parada del entrenamiento mediante validación cruzada (50 patrones) y un coeficiente de determinación umbral igual a 0.99

Red	Pesos	Calibración		Generalización				
		$\epsilon$	Épocas	R	R <sup>2</sup>	RMS <sup>1</sup>	%SEP	E
5:5s:5s:11	55	0.1632	177	0.9510	0.9043	0.0139	20.0297	0.8768
5:5s:10s:11	85	0.2522	9089	0.9423	0.8880	0.0180	25.9744	0.7929
5:10s:5s:11	105	0.3116	1937	0.9583	0.9184	0.0135	19.8297	0.8829
5:10s:10s:11	160	0.4748	956	0.9361	0.8763	0.0211	30.4990	0.7144
5:10s:15s:11	215	0.6380	1913	0.9435	0.8902	0.0193	27.8394	0.7620
5:15s:10s:11	235	0.6973	1837	0.9587	0.9191	0.0264	38.1045	0.5542
5:15s:15s:11	315	0.9347	2195	0.9438	0.8908	0.0191	27.5622	0.7668
5:15s:20s:11	395	1.1721	1251	0.9536	0.9093	0.0155	22.3637	0.8464
5:20s:15s:11	415	1.2315	1436	0.9509	0.9041	0.0161	23.2299	0.8343
5:20s:20s:11	520	1.5430	1969	0.9470	0.8968	0.0206	29.7864	0.7276

<sup>1</sup>Datos escalados en el rango [0,1]



**Figura IV.29.** a) Concentración de nitritos observada y estimada del sistema 'A' en el período de validación, considerando en la red neuronal 5:10s:5s:11 como variables de entrada la concentración de nitritos de los cinco días anteriores; b) Residuos en el grupo de validación frente a la concentración de nitritos estimado; c) y d) Frecuencia de las funciones coseno y seno frente a los coeficientes encontrados en la transformada de Fourier



**Figura IV.30.** Regresión polinómica de segundo grado entre el error estándar de predicción (SEP) alcanzado durante la validación (concentración de nitritos en la serie 'A') y el número de pesos de la red neuronal. Se indica la ecuación de la regresión, su coeficiente de determinación, el número de datos así como la arquitectura neuronal con la que se obtienen los mejores resultados

Los resultados del análisis del coeficiente de Theil se muestran en la tabla IV.30. El mejor valor del coeficiente  $U$  se corresponde con el calculado para el modelo ARIMA(2,1,2). También es en este modelo donde los componentes sistemático y de dispersión alcanzan sus valores más cercanos al cero ( $U_s=0.0209$ ;  $U_d=0.1108$ ). Sin embargo, el menor error de correlación ha sido el calculado para el modelo de red neuronal 5:10s:5s:11 ( $U_c=0.5145$ ). El análisis de las distribuciones de frecuencias absolutas por cuadrantes en los diagramas de predicción-realización muestra a través del test  $\chi^2$  la ausencia de diferencias significativas entre los cuatro modelos considerados, excepto cuando se comparan las frecuencias encontradas en la regresión múltiple y el suavizado de Holt (tabla IV.31). Estas diferencias se deben a un mayor número de casos contenidos en el segundo cuadrante del diagrama de predicción-realización de la regresión múltiple, de tal forma que la contribución porcentual del segundo cuadrante al valor del estadístico  $\chi^2$  es de casi el 50 %.

**Tabla IV.30.** Valores de cada componente de error del coeficiente de desigualdad de Theil en la serie 'A'

	Coeficiente de Theil (U)	Error sistemático ( $U_s$ )	Error de dispersión ( $U_d$ )	Error de correlación ( $U_c$ )
<b>Regresión múltiple</b>	0.7736	0.0573	0.5168	0.5782
<b>Suavizado de Holt</b>	0.6219	0.0319	0.1940	0.5938
<b>ARIMA(2,1,2)</b>	0.6123	0.0209	0.1108	0.6055
<b>RNC 5:10s:5s:11</b>	0.7887	0.0783	0.5988	0.5145

**Tabla IV.31.** Resultados de los test  $\chi^2$  entre las frecuencias por cuadrantes en los diagramas de Theil de la serie 'A' para los cuatro modelos seleccionados. Se indica el valor del estadístico  $\chi^2$  así como el nivel de significación del test entre paréntesis. \*Contribución porcentual del segundo cuadrante en el valor del estadístico  $\chi^2$

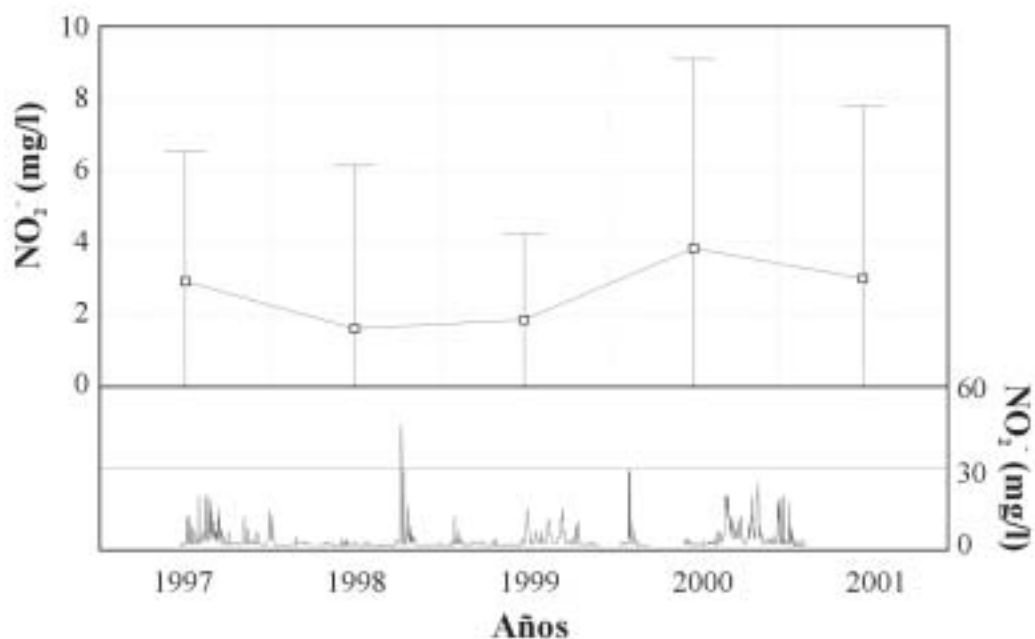
	Suavizado de Holt	ARIMA(2,1,2)	Red neuronal 5:10s:5s:11
<b>Regresión múltiple</b>	8.2478 (0.0412) 48.7863 %*	4.7350 (0.1923)	1.6392 (0.6505)
<b>Suavizado de Holt</b>	----	0.4125 (0.9376)	2.6460 (0.4495)
<b>ARIMA(2,1,2)</b>		----	1.0781 (0.7824)

#### 4.3.2.9. Nitritos en las series 'B', 'E' y 'F'

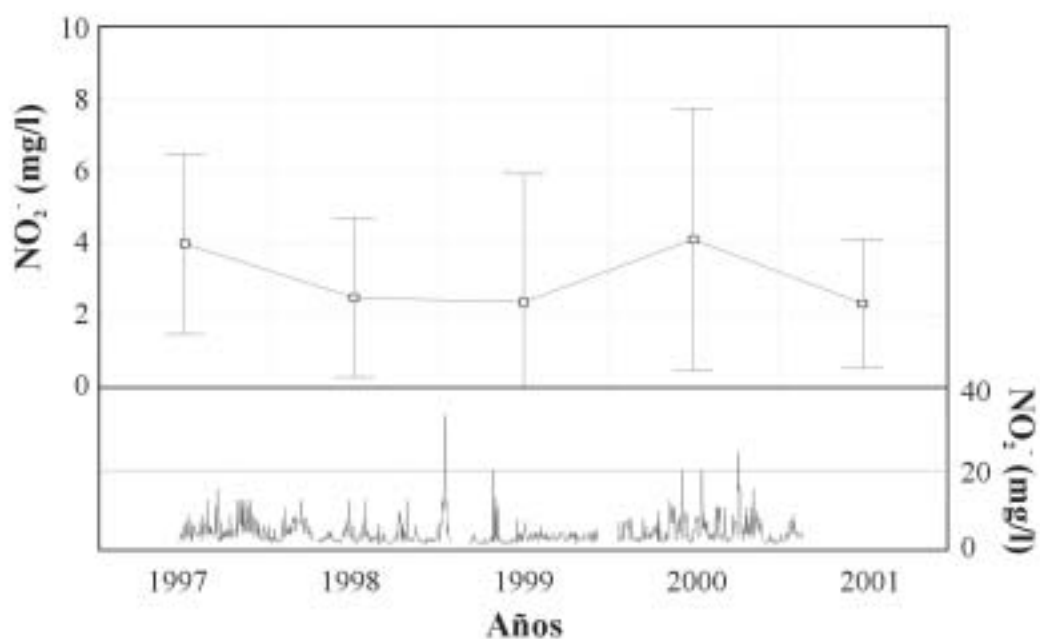
De la misma forma que en la serie 'A', el análisis de la varianza de la concentración de nitritos tanto en la serie 'B' como en las series de la zona de preengorde (series 'E' y 'F') han mostrado marcadas diferencias significativas (ANOVA Serie 'B':  $F=13.8009$ ;  $p_{\alpha}<0.05$ ; ANOVA Serie 'E' y 'F':  $F=15.7904$ ;  $p_{\alpha}<0.05$ ). En el caso de la serie 'B', las concentraciones medias por años siguen un patrón semejante al encontrado para la concentración de amoníaco en la misma serie. Se puede observar por tanto, un claro aumento de la concentración media de nitritos a partir del año 2000, que se mantiene prácticamente constante en el año siguiente (test de Tukey:  $p_{\alpha(1999-2000)}<0.05$ ;  $p_{\alpha(2000-2001)}=0.5039$ ). En cambio, la repetición del patrón no se ha detectado para la concentración de nitritos en las series 'E' y 'F' donde los menores y mayores promedios anuales se localizan en los años 1999 y 2000, respectivamente (test de Tukey:  $p_{\alpha(1999-2000)}<0.05$ ) (Figuras IV.31 y IV.32).

El análisis de correlaciones parciales entre las concentraciones de nitritos para las series 'A', 'B', 'E' y 'F' indica la presencia de una baja correlación significativa entre las series 'A' y 'B', así como entre las series 'B', 'E' y 'F'. Dicha correlación significativa no ha sido detectada entre la serie 'A' y las dos series de la zona de preengorde ('E' y 'F') (Serie 'A'-Serie 'B':  $R=0.1506$ ;  $p_{\alpha}<0.05$ ; Serie 'B'-Series 'E' y 'F':  $R=0.1278$ ;  $p_{\alpha}<0.05$ ; Serie 'A'-Series 'E' y 'F':  $R=0.0431$ ;  $p_{\alpha}>0.05$ ). Esta distribución de la correlación ha permitido la comparación del comportamiento de los modelos lineales con los no lineales en la ampliación de la generalización de los modelos calibrados para la serie 'A' en el resto de las series de circuito cerrado de la planta. Los resultados de la aplicación en las series 'B', 'E' y 'F' de los modelos calibrados para la serie 'A' se resumen en la tabla IV.32.





**Figura IV.31.** Valores medios anuales de la concentración de  $\text{NO}_2^-$  presente en el agua de la serie 'B' (gráfica superior). Evolución temporal a lo largo del periodo de estudio (valores absolutos) en la serie 'B' (gráfica inferior)



**Figura IV.32.** Valores medios anuales de la concentración de  $\text{NO}_2^-$  presente en el agua de las series 'E' y 'F' (gráfica superior). Evolución temporal a lo largo del periodo de estudio (valores absolutos) en las series 'E' y 'F' (gráfica inferior)

**Tabla IV.32.** Estimación de la concentración de nitritos en el agua en las series ‘B’, ‘E’ y ‘F’ a partir de los mejores modelos en validación calibrados en la serie ‘A’

	Serie ‘B’				
	R	R <sup>2</sup>	RMS	SEP	E
Regresión múltiple	0.9644	0.9302	0.9345	29.3750	0.9270
Suavizado de Holt	0.9730	0.9468	0.8760	27.5354	0.9359
ARIMA(2,1,2)	0.9809	0.9622	0.6896	21.6769	0.9603
RNC 5:10s:5s:1l	0.9568	0.9154	1.0568	33.2191	0.9067
	Serie ‘E’ y ‘F’				
Regresión múltiple	0.9681	0.9371	0.5770	17.7910	0.9335
Suavizado de Holt	0.9698	0.9405	0.5999	18.4976	0.9281
ARIMA(2,1,2)	0.9735	0.9477	0.5205	16.0495	0.9459
RNC 5:10s:5s:1l	0.9581	0.9180	0.6526	20.1213	0.9150

Tanto en la serie ‘B’ como en las series ‘E’ y ‘F’ los mejores resultados son los calculados para el modelo ARIMA (2,1,2). En la serie ‘B’ la cantidad de varianza explicada por el modelo ARIMA es superior al 96 %, mientras que el error estándar de predicción está cercano al 22 %. También el mejor coeficiente de eficiencia se encuentra en este modelo. Por el contrario, en esta ocasión los peores resultados son los calculados para la red neuronal. Aunque un test de comparación de coeficientes de correlación no mostró la presencia de diferencias significativas (test R:  $p_{\alpha} > 0.05$ ) entre las correlaciones en validación para la red neuronal y el modelo ARIMA(2,1,2), las diferencias entre los errores estándar de predicción fueron de casi 12 puntos. Estas diferencias se acortaron al comparar los coeficientes de eficiencia, donde el modelo ARIMA(2,1,2) superó en 6 puntos a la red neuronal.

En las series ‘E’ y ‘F’ tampoco se presentaron diferencias significativas entre las correlaciones de validación del mejor y el peor modelo [ARIMA(2,1,2) y red neuronal, respectivamente] (test R:  $p_{\alpha} > 0.05$ ), pero en esta ocasión las diferencias entre los errores estándar de predicción se reducen a 4 puntos. Comparando la serie ‘B’ con ambas series de la zona de preengorde, se puede observar como el modelo que cuantitativamente mejora más su predicción es la red neuronal. Esto también se refleja en el coeficiente de eficiencia. A pesar de que el error estándar de predicción del modelo ARIMA(2,1,2) en las serie ‘E’ y ‘F’ es inferior al encontrado para la serie ‘B’, el coeficiente de eficiencia de este modelo para las series ‘E’ y ‘F’ es peor al calculado para la serie ‘B’. Nuevamente la mayor mejora cuantitativa en el coeficiente de eficiencia en las serie ‘E’ y ‘F’ en relación a la serie ‘B’ se obtiene en la red neuronal. Por otra parte, la mejora del modelo de red neuronal también se manifiesta en el análisis de los residuos ya que es el único modelo que no presenta una correlación significativa entre la serie estimada y los residuos.

El análisis del coeficiente de error de Theil en la serie 'B' muestra que el modelo que obtiene una peor puntuación es el ARIMA(2,1,2), mientras que los mejores modelos son el suavizado de Holt y la red neuronal (tabla IV.33). Al dividir el coeficiente  $U$  en sus tres componentes, el mayor error sistemático lo presenta la red neuronal ( $U_s=0.0592$ ) aunque la contribución de este componente al error total en esta ocasión es la menor en todos los casos. En la componente  $U_d$  hay importantes diferencias, fundamentalmente en el modelo ARIMA(2,1,2) y en la regresión múltiple. Por el contrario, el menor error de correlación  $U_c$  es el calculado para el modelo ARIMA(2,1,2).

**Tabla IV.33.** Valores de cada componente de error del coeficiente de desigualdad de Theil en la serie 'B'

	Coeficiente de Theil ( $U$ )	Error sistemático ( $U_s$ )	Error de dispersión ( $U_d$ )	Error de correlación ( $U_c$ )
<b>Regresión múltiple</b>	0.8358	0.0577	0.6565	0.5207
<b>Suavizado de Holt</b>	0.6773	0.0426	0.2757	0.6208
<b>ARIMA(2,1,2)</b>	0.9978	0.0575	0.9988	0.0686
<b>RNC 5:10s:5s:11</b>	0.6846	0.0592	0.3037	0.6144

En las series 'E' y 'F' el análisis del coeficiente de desigualdad de Theil muestra que de forma global el mejor modelo es el suavizado de Holt, aunque seguido muy de cerca por el modelo ARIMA y la red neuronal (tabla IV.34). Sin embargo, el modelo ARIMA obtiene mejores valores tanto para el error sistemático como para el error de dispersión. En el error de correlación, el suavizado de Holt nuevamente obtiene el mejor valor.

**Tabla IV.34.** Valores de cada componente de error del coeficiente de desigualdad de Theil en las series 'E' y 'F'

	Coeficiente de Theil ( $U$ )	Error sistemático ( $U_s$ )	Error de dispersión ( $U_d$ )	Error de correlación ( $U_c$ )
<b>Regresión múltiple</b>	0.7671	0.0806	0.5516	0.5300
<b>Suavizado de Holt</b>	0.5488	0.0429	0.1772	0.5184
<b>ARIMA(2,1,2)</b>	0.5575	0.0353	0.1626	0.5336
<b>RNC 5:10s:5s:11</b>	0.5896	0.0471	0.2399	0.5389

El estadístico  $\chi^2$  calculado para las frecuencias absolutas por cuadrantes en los diagramas de predicción-realización de la serie 'B' no ha mostrado en ninguno de los casos la presencia de diferencias significativas (tabla IV.35). En el caso de las series 'E' y 'F' tan solo cuando se comparan las frecuencias de la regresión múltiple con el suavizado de Holt, se encuentran diferencias significativas como consecuencia de un alto número de casos en el segundo cuadrante de la regresión múltiple (tabla IV.36).

**Tabla IV.35.** Resultados de los test  $\chi^2$  entre las frecuencias por cuadrantes en los diagramas de Theil de la serie 'B' para los cuatro modelos seleccionados. Se indica el valor del estadístico  $\chi^2$  así como el nivel de significación del test entre paréntesis

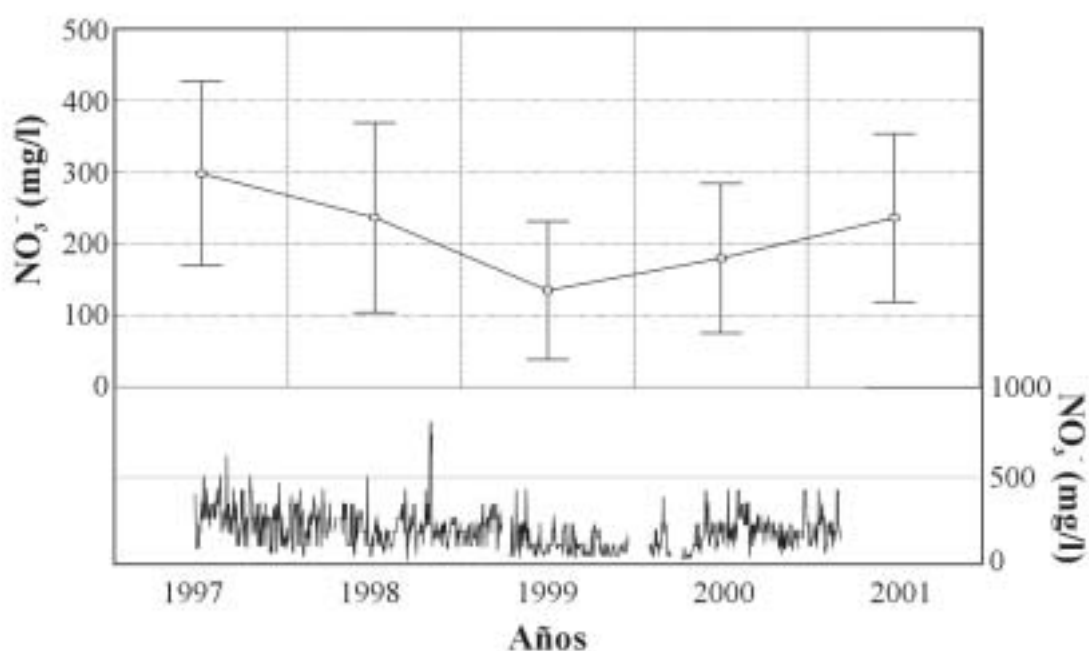
	Suavizado de Holt	ARIMA(2,1,2)	Red neuronal 5:10s:5s:11
<b>Regresión múltiple</b>	4.4903 (0.2132)	3.2220 (0.3587)	5.4376 (0.1424)
<b>Suavizado de Holt</b>	----	4.3618 (0.2250)	0.0403 (0.9979)
<b>ARIMA(2,1,2)</b>		----	5.6969 (0.1273)

**Tabla IV.36.** Resultados de los test  $\chi^2$  entre las frecuencias por cuadrantes en los diagramas de Theil de las series 'E' y 'F' para los cuatro modelos seleccionados. Se indica el valor del estadístico  $\chi^2$  así como el nivel de significación del test entre paréntesis. \* Contribución porcentual del segundo cuadrante en el valor del estadístico  $\chi^2$

	Suavizado de Holt	ARIMA(2,1,2)	Red neuronal 5:10s:5s:11
<b>Regresión múltiple</b>	11.6293 (0.0088) 72.6321 %*	7.7662 (0.0511)	1.7893 (0.6173)
<b>Suavizado de Holt</b>	----	0.5235 (0.9137)	4.3106 (0.2298)
<b>ARIMA(2,1,2)</b>		----	2.0598 (0.5601)

#### 4.3.2.10. Nitratos en la serie 'A'

El análisis de la varianza de la concentración de nitrato en la serie 'A' por años muestra la presencia de marcadas diferencias significativas entre los valores medios para el período de estudio considerado (ANOVA:  $F=74.1891$ ;  $p_{\alpha}<0.05$ ) (figura IV.33). En la evolución de las concentraciones medias por años de este parámetro se puede observar un claro descenso significativo desde el año 1997 hasta 1999, así como un aumento significativo a partir de ese año hasta el 2001 (test de Tukey: en todos los casos  $p_{\alpha}<0.05$ ).



**Figura IV.33.** Valores medios anuales de la concentración de  $\text{NO}_3^-$  presente en el agua de la serie 'A' (gráfica superior). Evolución temporal a lo largo del periodo de estudio (valores absolutos) en la serie 'A' (gráfica inferior)

La ecuación de regresión múltiple que relaciona la concentración de nitratos en un día determinado con la concentración de nitratos en los cinco días anteriores muestra unos altos niveles de correlación tanto en el proceso de calibración como en validación. Así, la correlación durante la calibración está cercana al 99 %, y por tanto, la cantidad de varianza explicada durante esta fase es prácticamente del 98 %. Durante la validación, los valores de correlación y de varianza explicada se encuentran muy cercanos a los calculados durante la calibración ( $R=0.9865$ ;  $R^2=0.9732$ ). Por otra parte los términos de error son los mejores de los conseguidos para todas las sustancias nitrogenadas, de tal forma que el error estándar de predicción en validación se encuentra por debajo del 7 % ( $\text{SEP}=6.9664$ ) y el nivel de

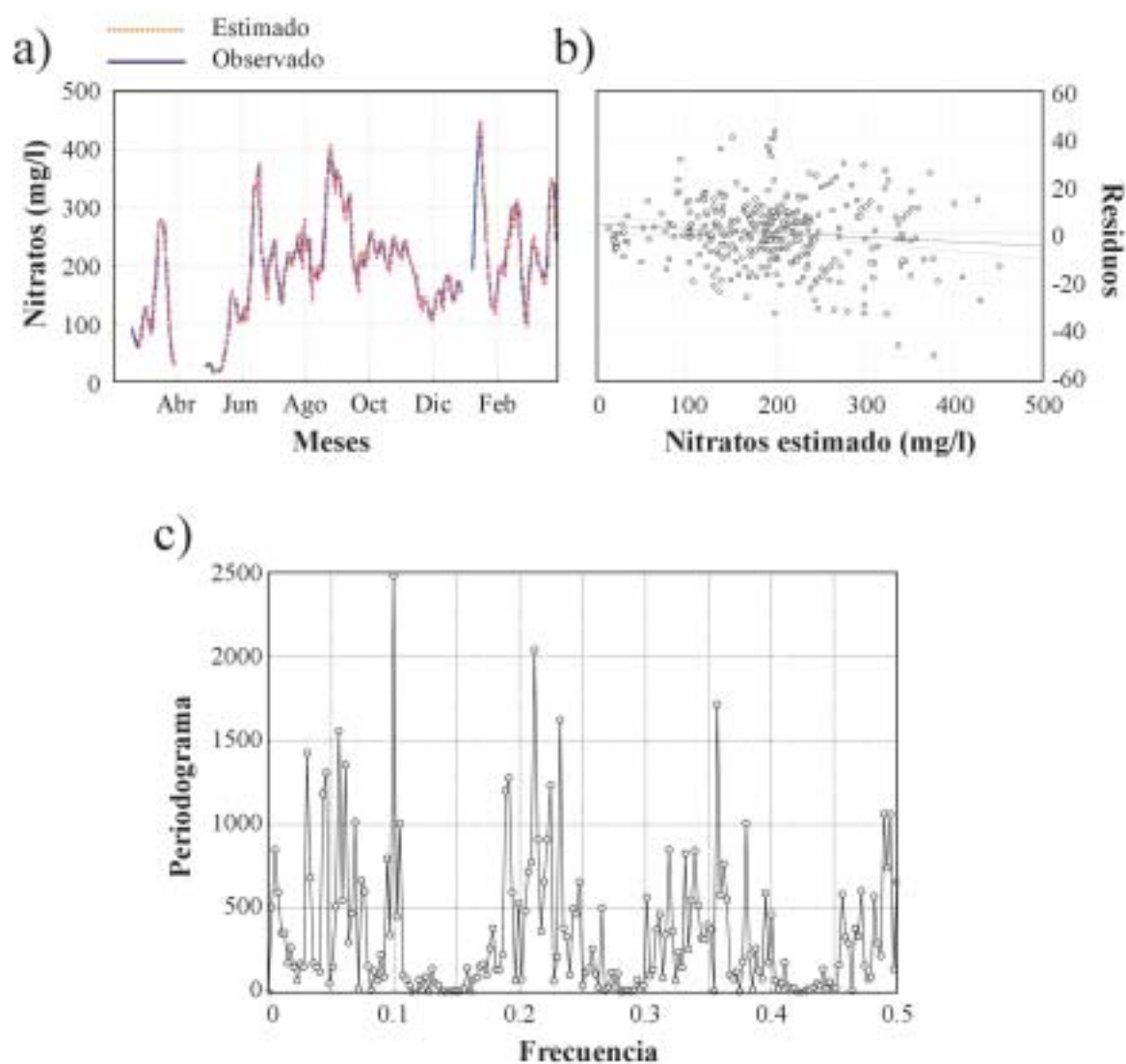
eficiencia de la validación es de 0.9727 (tabla IV.37). El análisis de la correlación serial en el proceso de calibración indica la ausencia de autocorrelación, al encontrarse el estadístico de Durvin-Watson muy próximo a 2. Sin embargo, el análisis espectral de los residuos en la validación muestra la presencia de tres frecuencias con valores en el periodograma significativamente importantes que se corresponden con períodos de 3, 5 y 10 días.

**Tabla IV.37. Estimación del  $\text{NO}_3^-$  del agua en la serie 'A' mediante el análisis de regresión múltiple con las variables independientes representadas por el  $\text{NO}_3^-$  filtrado en los cinco días anteriores**

<b>Resumen de la regresión múltiple en la calibración del modelo</b>			
R=0.9890; $R^2=0.9781$ ; F(5;319)=2849.3; $p_\alpha < 0.001$ ; N=325			
Variable dependiente	Variable independiente	$b_i(i=0,1,\dots,q)$	$p_\alpha$
$\text{NO}_3^- (t)$	Intersección ( $b_0$ )	3.2827	0.0150*
	$\text{NO}_3^- (t-1)$	1.5830	0**
	$\text{NO}_3^- (t-2)$	-0.7200	0**
	$\text{NO}_3^- (t-3)$	0.2183	0.0513
	$\text{NO}_3^- (t-4)$	-0.1343	0.1992
	$\text{NO}_3^- (t-5)$	0.0291	0.6030
<b>Resumen de la regresión en la validación del modelo</b>			
R=0.9865; $R^2=0.9732$ ; F(1;343)=12462.0; $p_\alpha < 0.001$ ; N=345			
Error estándar de predicción=6.9664 %; Coeficiente E=0.9727			
Variable dependiente	Variable independiente	$b_i(i=0,1,\dots,q)$	$p_\alpha$
$\text{NO}_3^-$ observado (t)	Intersección ( $b_0$ )	4.7727	0.0118**
	$\text{NO}_3^-$ estimado (t)	0.9822	0**

\*  $p_\alpha < 0.05$ ; \*\*  $p_\alpha < 0.01$

En la regresión múltiple anterior tan solo los dos días anteriores contribuyen significativamente sobre la varianza de la variable independiente, mientras que la concentración de nitratos de tres días antes presenta un nivel de significación marginal ( $p_\alpha=0.0513$ ). Se ha procedido, por tanto, nuevamente a la calibración y validación de la regresión múltiple en la que las variables independientes son las concentraciones de nitratos en los dos días anteriores. En ambas fases los resultados son tan satisfactorios como los conseguidos con la regresión anterior aunque con una mejoría no significativa en el error estándar de predicción (Calibración: R=0.9889;  $R^2=0.9778$ ; F(2;328)=7239.6;  $p_\alpha < 0.001$ ; N=331. Validación: R=0.9865;  $R^2=0.9731$ ; F(1;343)=12431.0;  $p_\alpha < 0.001$ ; N=345; SEP=6.9638; Coeficiente E=0.9727). De la misma forma que ocurría con la regresión anterior, no se ha detectado autocorrelación entre residuos durante la calibración (estadístico de Durvin-Watson=1.9221). Nuevamente el análisis espectral de los residuos de validación muestra tres frecuencias significativas en el periodograma ( $v=0.0995$ ,  $v=0.2117$  y  $v=0.3571$ ) que se corresponden con períodos de aproximadamente 3, 5 y 10 días (figura IV.34).



**Figura IV.34.** a) Concentración de nitratos observada y estimada del sistema 'A' en el periodo de validación, considerando en el análisis de regresión múltiple como variable dependiente la concentración de nitratos y como variables independientes la concentración de nitratos de los dos días anteriores; b) Residuos en el grupo de validación frente a la concentración de nitratos estimada y c) Periodograma de la serie de residuos obtenida en la fase de validación

Los modelos de suavizado también proporcionan resultados aceptables. En esta ocasión, de la misma forma que ha ocurrido con el resto de parámetros, el suavizado de Holt ha proporcionado mejores términos de error que el suavizado lineal simple (tabla IV.38) (figura IV.35). Por otra parte, de entre los modelos **ARIMA** con todos sus términos significativos destaca el modelo ARIMA(1,1,0). En éste la cantidad de varianza explicada es superior al 97 % y los términos de error son de 7.1509 % (**SEP**) y 0.9712 (**E**) (figura IV.35). A pesar de ello, la función de autocorrelación de los residuos presenta correlación significativa al menos para 7 retrasos. Asimismo, como en el resto de los parámetros, el modelo **ARIMA** que presenta una peor bondad de ajuste es el modelo ARIMA(0,0,1), denotando el carácter autoregresivo de la serie.

**Tabla IV.38.** Parámetros de estimación y bondad de ajuste de los modelos de series temporales en el  $NO_3^-$  filtrado de la serie 'A'

Modelo	Parámetros	Validación			
		R <sup>2</sup>	F	%SEP	Coef. E
Suavizado exponencial simple	$S_0=136.6000; \alpha_d=1$	0.9400	(1;358)=5605.3	10.8340	0.9383
Suavizado lineal exponencial	$S_0=178.7000; T_0=-0.2720$ $\alpha_d=0.9000; \beta_d=0.9000$	0.9674	(1;355)=10545.4	8.7704	0.9596
ARIMA (0,0,1)	$\Theta_1=-0.9785^{**}$	0.8028	(1;358)=1457.6	56.2509	-0.6635
ARIMA (1,0,0)	$\Phi_1=0.9949^{**}$	0.9516	(1;358)=7035.6	9.7000	0.9505
ARIMA (0,1,1)	$\Theta_1=-0.5718^{**}$	0.9697	(1;357)=11428.9	7.6534	0.9692
ARIMA (1,0,1)	$\Phi_1=0.9916^{**}; \Theta_1=-0.5262^{**}$	0.9657	(1;358)=10069.1	8.1895	0.9647
ARIMA (1,1,0)	$\Phi_1=0.7289^{**}$	0.9750	(1;357)=13905.6	7.1509	0.9712
ARIMA (1,1,1)	$\Phi_1=0.7194^{**}; \Theta_1=-0.0182$	0.9750	(1;357)=13904.6	7.1205	0.9733
ARIMA (1,1,2)	$\Phi_1=0.7477^{**}; \Theta_1=0.0028$ $\Theta_2=-0.0462$	0.9750	(1;357)=13944.6	7.1210	0.9733
ARIMA (1,2,1)	$\Phi_1=0.3289^{**}; \Theta_1=-0.9811^{**}$	0.9561	(1;354)=7706.2	9.7928	0.9496
ARIMA (2,2,1)	$\Phi_1=0.4606^{**}; \Theta_1=-0.9708^{**}$ $\Phi_2=-0.3781^{**}$	0.9644	(1;354)=9603.3	8.8489	0.9588
ARIMA (2,1,2)	$\Phi_1=1.7758^{**}; \Theta_1=1.1362^{**}$ $\Phi_2=-0.8485^{**}; \Theta_1=-0.1917$	0.9750	(1;357)=13921.2	6.9504	0.9746
ARIMA (2,1,1)	$\Phi_1=0.2779; \Theta_1=0.4725$ $\Phi_2=0.3122$	0.9750	(1;357)=13900.5	7.1178	0.9734
ARIMA (1,2,2)	$\Phi_1=-0.1522^*; \Theta_1=-1.8427^{**}$ $\Theta_2=-0.8670^{**}$	0.9679	(1;354)=10671.7	8.3052	0.9637
ARIMA (2,2,2)	$\Phi_1=-0.1050; \Theta_1=-1.7745^{**}$ $\Phi_2=-0.2134^{**}; \Theta_2=-0.8036^{**}$	0.9697	(1;354)=11314.0	8.1193	0.9653

\*  $p_\alpha < 0.05$ ; \*\*  $p_\alpha < 0.01$



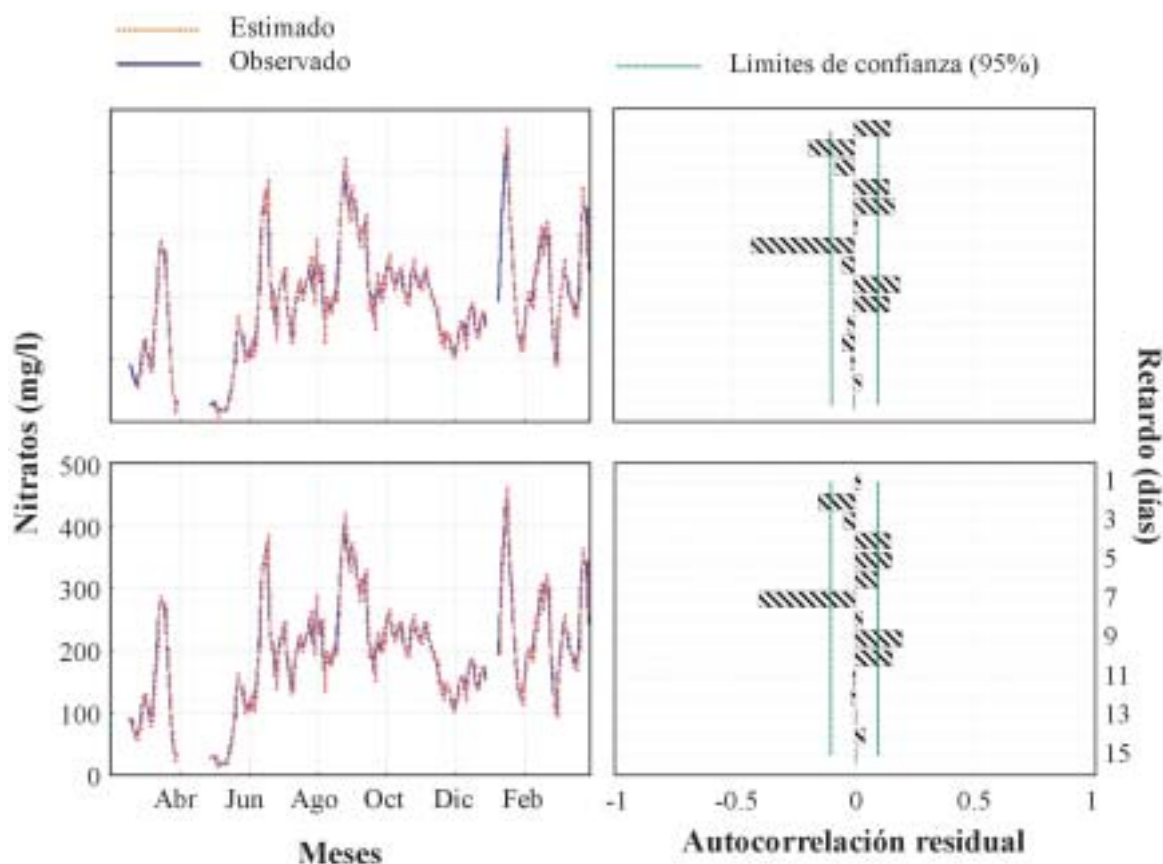


Figura IV.35. Concentración de nitratos observada y estimada en la serie 'A' y autocorrelación residual en validación para los modelos: suavizado lineal exponencial (Holt) (arriba) y ARIMA(1,1,0) (abajo)

En función del coeficiente de determinación, del error estándar de predicción y del coeficiente de eficiencia, las redes neuronales en esta ocasión proporcionan las peores estimaciones. El error estándar de predicción varía entre el 25.7597 % para la red 5:5s:10s:11 y el 11.5963 % para la arquitectura 5:15s:10s:11, mientras que el coeficiente de eficiencia varía entre 0.6527 y 0.9243, respectivamente (tabla IV.39). Sin embargo, al contrario de lo que ocurre con los modelos de regresión múltiple y univariantes de series temporales, el análisis espectral de los residuos en validación para la red 5:15s:10s:11 tan solo mostró valores de periodograma significativamente grandes para frecuencias muy pequeñas, lo que se corresponde con períodos mayores a 150 días (figura IV.36). La regresión lineal entre el número de pesos de la red neuronal y el error estándar de predicción ha dado una correlación relativamente alta ( $R=0.6296$ ) aunque con un nivel de significación marginal ( $p_{\alpha}=0.0511$ ). Una regresión polinómica de segundo grado incrementó el porcentaje de varianza explicada en más de 11 puntos, estableciendo el coeficiente de determinación en el 0.5108 (figura IV.37). Asimismo, se han encontrado importantes diferencias significativas entre el error

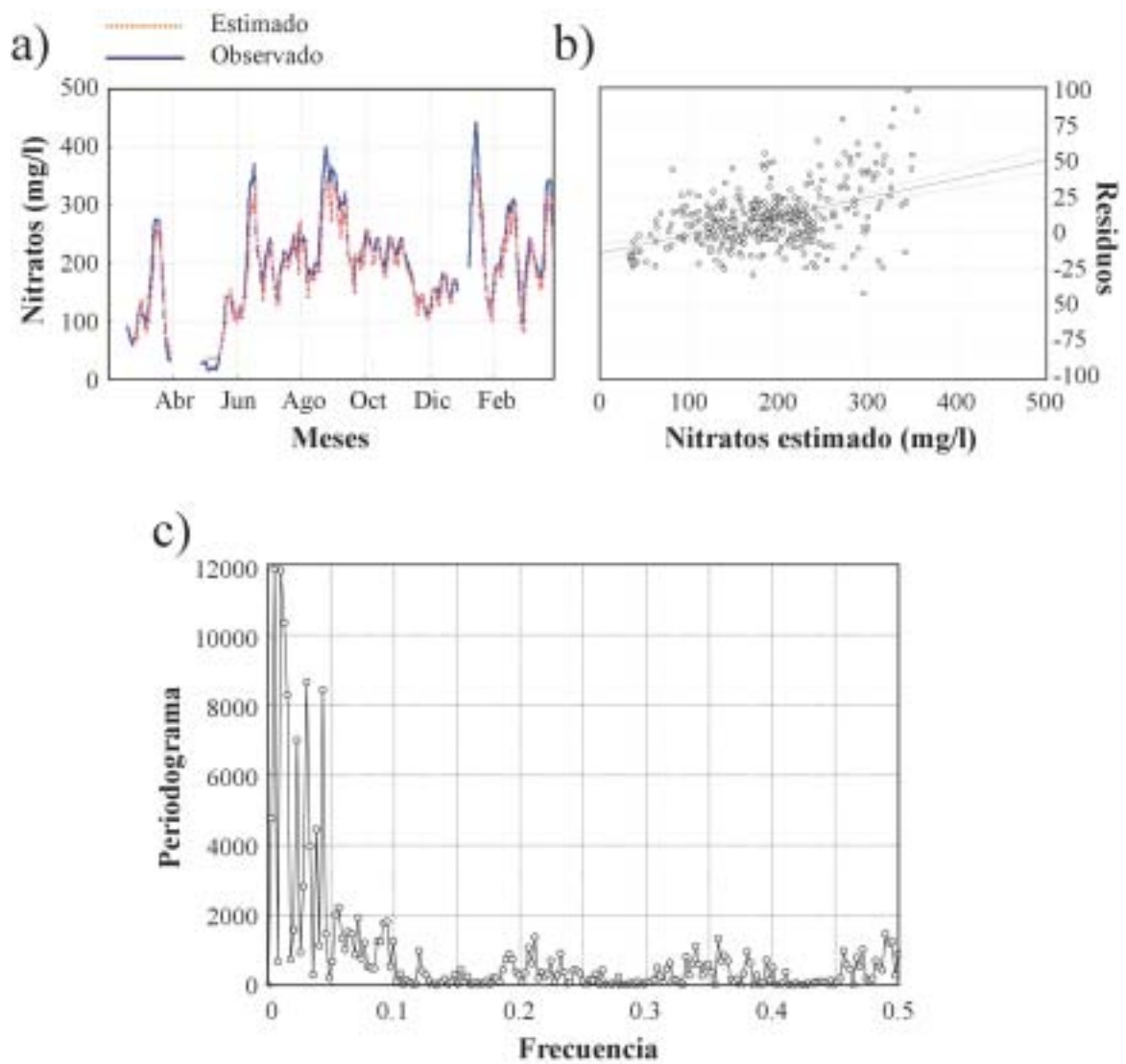
teórico mínimo alcanzable durante la calibración y el error estándar de predicción en función del número de pesos (test de comparación de dos rectas de regresión:  $F_{exp}=12.7073$ ;  $t_{exp}=185.4212$ ;  $t_{(0.001)}(17819.4)< 3.2910$  ; ANCOVA,  $p_{\alpha}<0.001$ ).

**Tabla IV.39.** Resultados de calibración y generalización de las RNCs en las que las variables de entrada son las concentraciones de nitratos filtradas en el sistema 'A' en los cinco días anteriores. Parada del entrenamiento mediante validación cruzada (50 patrones) y un coeficiente de determinación umbral igual a 0.99

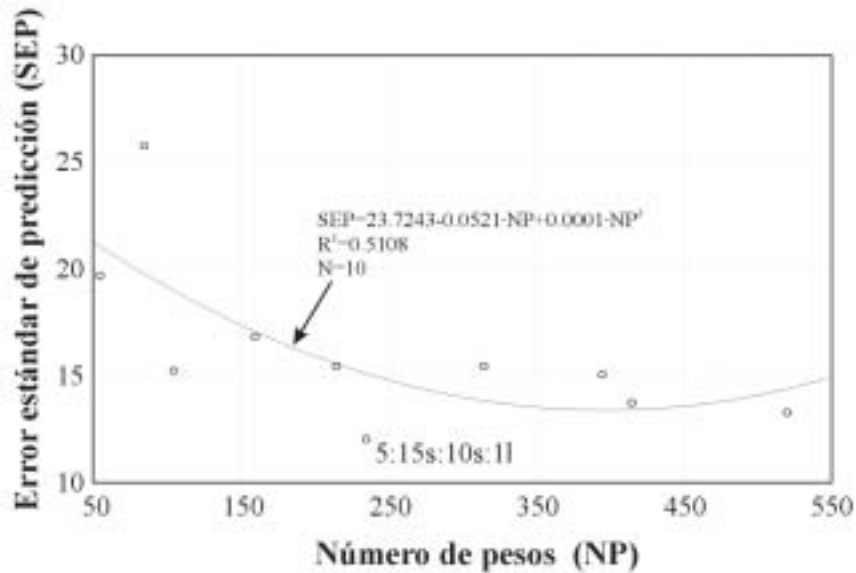
Red	Pesos	Calibración		Generalización				
		$\epsilon$	Épocas	R	R <sup>2</sup>	RMS <sup>1</sup>	%SEP	E
5:5s:5s:11	55	0.1594	8092	0.9473	0.8974	0.0844	19.6753	0.7974
5:5s:10s:11	85	0.2464	9984	0.9288	0.8627	0.1105	25.7597	0.6527
5:10s:5s:11	105	0.3043	8162	0.9628	0.9269	0.0653	15.2328	0.8785
5:10s:10s:11	160	0.4638	9836	0.9463	0.8955	0.0722	16.8323	0.8517
5:10s:15s:11	215	0.6232	7921	0.9557	0.9133	0.0663	15.4519	0.8750
5:15s:10s:11	235	0.6812	9620	0.9760	0.9526	0.0516	11.5963	0.9243
5:15s:15s:11	315	0.9130	9742	0.9592	0.9201	0.0663	15.4509	0.8750
5:15s:20s:11	395	1.1449	9314	0.9616	0.9247	0.0646	15.0634	0.8812
5:20s:15s:11	415	1.2029	9139	0.9655	0.9322	0.0589	13.7298	0.9013
5:20s:20s:11	520	1.5072	9910	0.9691	0.9391	0.0569	13.2733	0.9078

<sup>1</sup>Datos escalados en el rango [0,1]

El análisis del coeficiente de Theil proporciona nuevamente unos resultados que aportan una ligera contradicción en cuanto a la validación de los modelos. Así, en función del valor calculado del coeficiente U, el modelo que presenta un mejor comportamiento en validación es la regresión múltiple seguido de la red neuronal, el modelo ARIMA(1,1,0) y el suavizado de Holt (tabla IV.40) Estos resultados cambian cuando el coeficiente U se descompone en su componente sistemático, de dispersión y de correlación. De esta forma, los mejores valores para los dos primeros componentes han sido los calculados para la red neuronal ( $U_s=0.0085$ ;  $U_d=0.1084$ ). El peor componente de correlación, sin embargo, es el obtenido para la red neuronal aunque las diferencias con el calculado para la regresión múltiple son mínimas. Por otra parte, la comparación de las frecuencias absolutas por cuadrantes en el diagrama de realización-predicción (figura IV.38) para los cuatro mejores modelos ha dado como resultado la ausencia total de diferencias significativas (tabla IV.41).



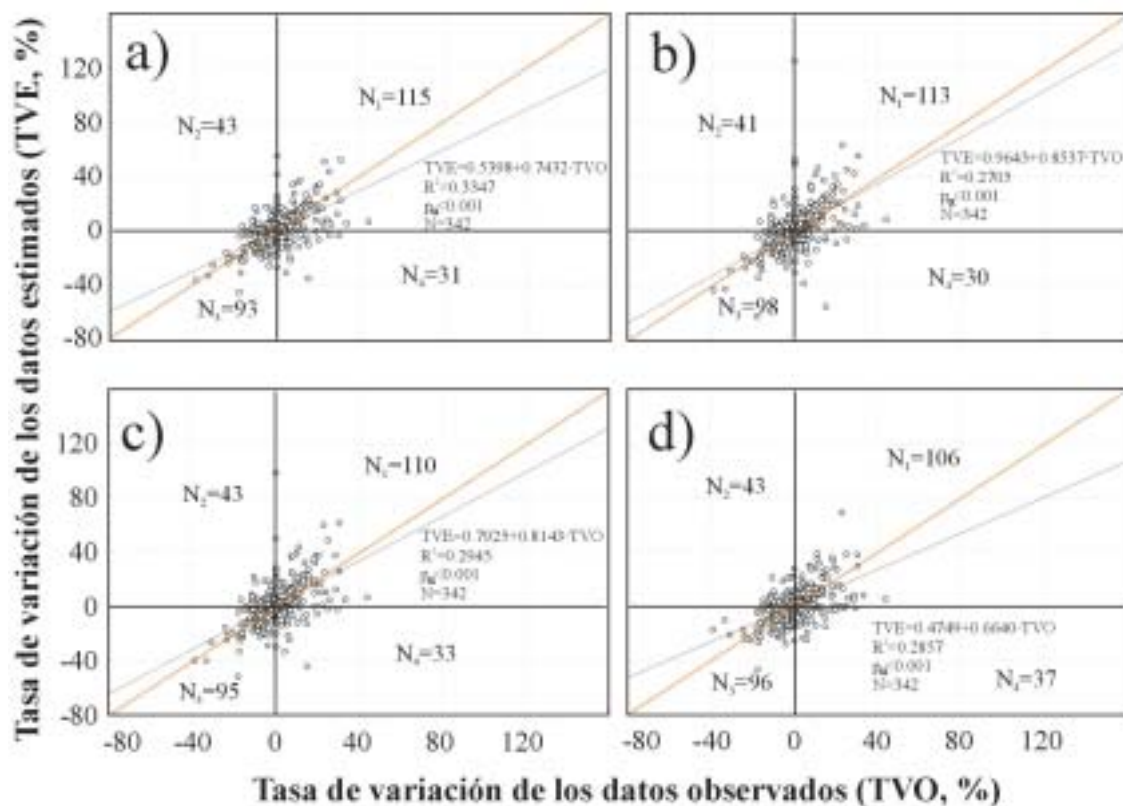
**Figura IV.36.** a) Concentración de nitratos observada y estimada del sistema 'A' en el periodo de validación, considerando en la red neuronal 5:15s:10s:11 como variables de entrada las concentraciones de nitratos en los cinco días anteriores ; b) Residuos en el grupo de validación frente a la concentración de nitratos estimada y c) Periodograma de la serie de residuos obtenida en la fase de validación



**Figura IV.37.** Regresión polinómica de segundo grado entre el error estándar de predicción (SEP) alcanzado durante la validación (concentración de nitratos en la serie 'A') y el número de pesos de la red neuronal. Se indica la ecuación de la regresión, su coeficiente de determinación, el número de datos así como la arquitectura neuronal con la que se obtienen los mejores resultados

**Tabla IV.40.** Valores de cada componente de error del coeficiente de desigualdad de Theil en la serie 'A'

	<b>Coefficiente de Theil (U)</b>	<b>Error sistemático (U<sub>s</sub>)</b>	<b>Error de dispersión (U<sub>d</sub>)</b>	<b>Error de correlación (U<sub>c</sub>)</b>
<b>Regresión múltiple</b>	0.4709	0.0136	0.1249	0.4567
<b>Suavizado de Holt</b>	0.5324	0.0301	0.2433	0.4761
<b>ARIMA(1,1,0)</b>	0.5081	0.0241	0.2005	0.4695
<b>RNC 5:15s:10s:11</b>	0.4903	0.0085	0.1084	0.4810



**Figura IV.38.** Tasa de variación de los datos observados frente a la tasa de variación de los datos estimados por: a) la regresión múltiple que considera como variables independientes la concentración de nitratos filtrada en uno y dos días antes; b) suavizado de Holt; c) ARIMA(1,1,0) y d) red neuronal 5:15s:10s:11 que considera como variables de entrada la concentración de nitratos filtrada en los cinco días previos. Se indica el número de casos por cuadrante ( $N_1$ ,  $N_2$ ,  $N_3$  y  $N_4$ ) así como la ecuación de regresión, el coeficiente de determinación, el nivel de significación y el número de datos. La línea roja punteada indica la recta de correlación 1 que pasa por el origen de coordenadas

**Tabla IV.41.** Resultados de los test  $\chi^2$  entre las frecuencias por cuadrantes en los diagramas de Theil de la serie 'A' para los cuatro modelos seleccionados. Se indica el valor del estadístico  $\chi^2$  así como el nivel del significación del test entre paréntesis

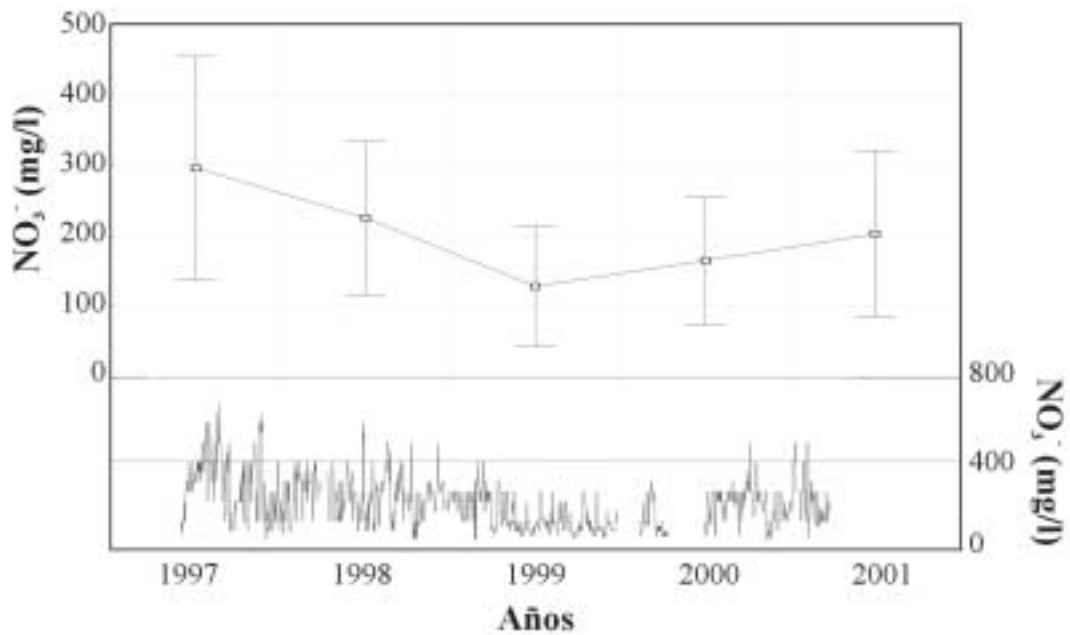
	Suavizado de Holt	ARIMA(1,1,0)	Red neuronal 5:15s:10s:11
Regresión múltiple	0.4541 (0.9288)	0.3620 (0.9480)	1.7832 (0.6183)
Suavizado de Holt	----	0.5568 (0.9107)	1.9213 (0.5889)
ARIMA(1,1,0)		----	0.5945 (0.8977)

#### 4.3.2.11. Nitratos en las series 'B', 'E' y 'F'

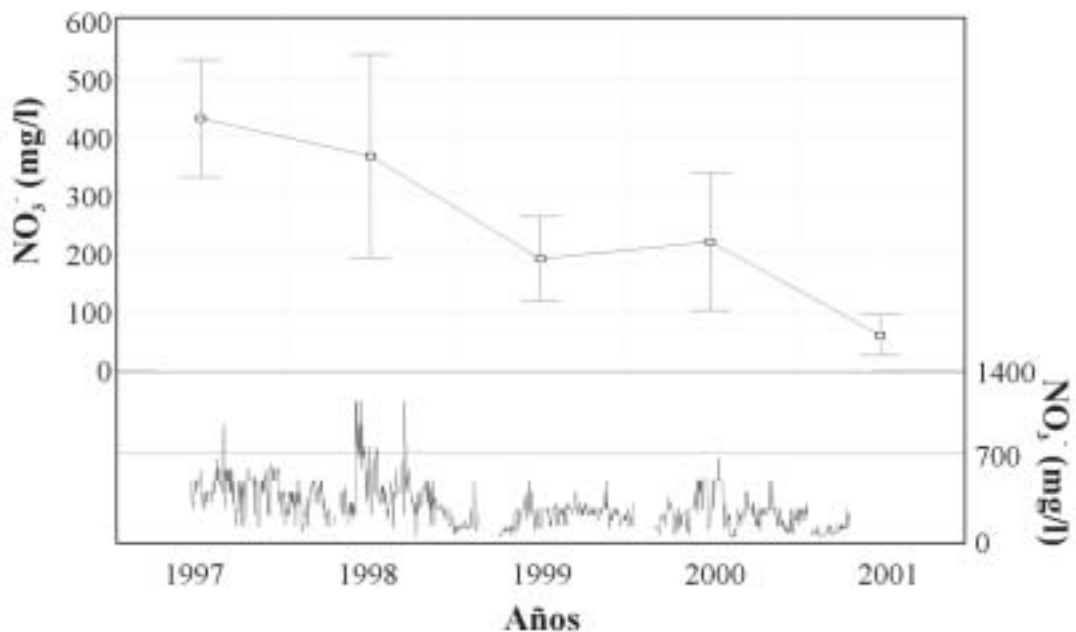
La evolución temporal de la concentración de nitratos en las series 'B', 'E' y 'F' se muestra en las figuras IV.39 y IV.40. En ambos casos (serie 'B' y las dos series de preengorde 'E' y 'F') se han encontrado marcadas diferencias significativas entre los valores medios anuales (Serie 'B': ANOVA:  $F=85.6681$ ;  $p_{\alpha}<0.001$ ; Serie 'E' y 'F': ANOVA:  $F=156.6907$ ;  $p_{\alpha}<0.001$ ). En el caso de la serie 'B' se muestra un patrón de evolución similar al encontrado para la serie 'A', es decir, una disminución desde el año 1997 que se prolonga hasta 1999, para posteriormente volver a aumentar hasta llegar a los primeros meses del año 2001 (test de Tukey:  $p_{\alpha}<0.05$  en todos los casos). En las series de la zona de preengorde se observa el mismo descenso desde el año 1997 hasta 1999, pero a partir de ese momento los niveles de nitratos se mantienen prácticamente constantes durante el año siguiente (test de Tukey:  $p_{\alpha}=1058$ ) para continuar en descenso significativo (test de Tukey:  $p_{\alpha}<0.001$ ) en los primeros meses del año 2001.

El análisis de correlaciones entre las series originales de la concentración de nitratos en las series 'A', 'B', 'E' y 'F' muestra la existencia de correlación significativa en todos los casos (Serie 'A'-Serie 'B':  $R=0.5561$ ,  $p_{\alpha}<0.05$ ; Serie 'A'-Series 'E' y 'F':  $R=0.3648$ ,  $p_{\alpha}<0.05$ ; Serie 'B'-Series 'E' y 'F':  $R=0.3552$ ,  $p_{\alpha}<0.05$ ). Así, aplicando los mejores modelos en validación calibrados en la serie 'A' los resultados son en todos los casos muy satisfactorios (tabla IV.42).

De la misma forma que ocurría con la extensión hacia la serie 'B' y las series de preengorde de los mejores modelos calibrados en la serie 'A' para el parámetro nitrato, los mejores resultados en validación se obtienen para el modelo **ARIMA**. En el caso de la serie 'B' la cantidad de varianza explicada por el modelo ARIMA(1,1,0) se encuentra al mismo nivel que en la serie 'A' (Serie 'A':  $R^2=0.9750$ ; Serie 'B':  $R^2=0.9717$ ). Lo mismo sucede con el error estándar de predicción y el coeficiente de eficiencia, aunque en el caso de la serie 'B' son ligeramente peores. Los peores resultados para esta misma serie y este parámetro químico lo encontramos en la regresión múltiple, que pasa de un error del 6.9638 % en la serie 'A' a un 10.4100 % en la serie 'B'.



**Figura IV.39.** Valores medios anuales de la concentración de  $\text{NO}_3^-$  presente en el agua de la serie 'B' (gráfica superior). Evolución temporal a lo largo del periodo de estudio (valores absolutos) en la serie 'B' (gráfica inferior)



**Figura IV.40.** Valores medios anuales de la concentración de  $\text{NO}_3^-$  presente en el agua de las series 'E' y 'F' (gráfica superior). Evolución temporal a lo largo del periodo de estudio (valores absolutos) en las series 'E' y 'F' (gráfica inferior)

**Tabla IV.42.** *Estimación de la concentración de nitratos en el agua en las series ‘B’, ‘E’ y ‘F’ a partir de los mejores modelos en validación calibrados en la serie ‘A’*

	<b>Serie ‘B’</b>				
	<b>R</b>	<b>R<sup>2</sup></b>	<b>RMS</b>	<b>SEP</b>	<b>E</b>
<b>Regresión múltiple</b>	0.9705	0.9418	18.4995	10.4100	0.9394
<b>Suavizado de Holt</b>	0.9787	0.9579	16.5225	9.2975	0.9517
<b>ARIMA(1,1,0)</b>	0.9858	0.9717	13.0367	7.3360	0.9699
<b>RNC 5:15s:10s:1l</b>	0.9794	0.9592	18.1641	10.2212	0.9416
	<b>Series ‘E’ y ‘F’</b>				
<b>Regresión múltiple</b>	0.9941	0.9882	13.4986	7.1282	0.9880
<b>Suavizado de Holt</b>	0.9935	0.9870	14.2384	7.5189	0.9866
<b>ARIMA(1,1,0)</b>	0.9943	0.9886	13.4088	7.0808	0.9881
<b>RNC 5:15s:10s:1l</b>	0.9883	0.9768	25.9442	13.7004	0.9556

En las serie ‘E’ y ‘F’ es también el modelo ARIMA(1,1,0) el que aporta las mejores tasas de error de entre todos los modelos seleccionados. Asimismo este modelo es capaz de explicar una mayor cantidad de la varianza de la variable dependiente ( $R^2=0.9886$ ), aunque seguido muy de cerca por el resto de modelos. En esta ocasión, el modelo para el que se han obtenido los peores resultados es la red neuronal, que incrementa ligeramente los términos de error alcanzados en la series ‘A’ y ‘B’. Sin embargo es precisamente este modelo el único en el que, a pesar de que el error estándar de predicción aumenta, el coeficiente de eficiencia también aumenta.

La evaluación y calificación del mejor modelo en función del coeficiente de desigualdad de Theil tampoco coincide en esta ocasión ni para la serie ‘B’ ni para las dos series de la zona de preengorde (tablas IV.43 y IV.44). Así, el valor más cercano a cero en ambas ocasiones es calculado para la red neuronal (Serie ‘B’:  $U=0.5136$ ; Serie ‘E’ y ‘F’:  $U=0.5229$ ). A ello contribuye fundamentalmente los bajos valores en los errores sistemáticos y de dispersión, donde nuevamente la red neuronal alcanza los mejores resultados. En cuanto al error de correlación en ambas series el mejor valor es alcanzado por la regresión múltiple, aunque los elevados errores sistemáticos y de dispersión hacen que para el conjunto de todas las series sea el peor modelo.



**Tabla IV.43.** Valores de cada componente de error del coeficiente de desigualdad de Theil en la serie 'B'

	<b>Coefficiente de Theil (U)</b>	<b>Error sistemático (U<sub>s</sub>)</b>	<b>Error de dispersión (U<sub>d</sub>)</b>	<b>Error de correlación (U<sub>c</sub>)</b>
<b>Regresión múltiple</b>	0.9807	0.0763	0.9700	0.1672
<b>Suavizado de Holt</b>	0.6915	0.0370	0.4998	0.4859
<b>ARIMA(1,1,0)</b>	0.5377	0.0271	0.2454	0.4840
<b>RNC 5:15s:10s:11</b>	0.5136	-0.0005	0.0789	0.5226

**Tabla IV.44.** Valores de cada componente de error del coeficiente de desigualdad de Theil en las series 'E' y 'F'

	<b>Coefficiente de Theil (U)</b>	<b>Error sistemático (U<sub>s</sub>)</b>	<b>Error de dispersión (U<sub>d</sub>)</b>	<b>Error de correlación (U<sub>c</sub>)</b>
<b>Regresión múltiple</b>	0.9855	-0.0470	0.9752	0.1601
<b>Suavizado de Holt</b>	0.6176	-0.0125	0.3008	0.5433
<b>ARIMA(1,1,0)</b>	0.7025	-0.0257	0.4466	0.5453
<b>RNC 5:15s:10s:11</b>	0.5229	-0.0101	-0.0112	0.5280

Por otra parte, el análisis de las frecuencias absolutas por cuadrantes del diagrama de realización-predicción, muestra la presencia de diferencias significativas entre las frecuencias en la regresión múltiple y la red neuronal, y entre el suavizado de Holt y la red neuronal para la serie 'B'. Así en el primer caso, las diferencias estadísticas se deben fundamentalmente a que el número de casos en el cuadrante uno es mayor en la regresión múltiple. En el segundo caso, las diferencias se alcanzan al ser la frecuencia mayor en el cuarto cuadrante de la red neuronal. A pesar de ello, la menor contribución en ambos casos de menores frecuencias en el segundo cuadrante de la red neuronal, hace que el coeficiente de desigualdad de Theil sea mejor para este último modelo (tabla IV.45).

Al contrario de lo sucedido para la serie 'B', en las series de preengorde no se detectan en ningún momento diferencias significativas debidas a una mayor concentración de casos en alguno de los cuatro cuadrantes de los diagramas de realización-predicción (tabla IV.46).

**Tabla IV.45.** Resultados de los test  $\chi^2$  entre las frecuencias por cuadrantes en los diagramas de Theil de la serie 'B' para los cuatro modelos seleccionados. Se indica el valor del estadístico  $\chi^2$  así como el nivel de significación del test entre paréntesis. <sup>1</sup>Contribución porcentual del primer cuadrante al valor del estadístico  $\chi^2$ . <sup>2</sup>Contribución porcentual del cuarto cuadrante al valor del estadístico  $\chi^2$

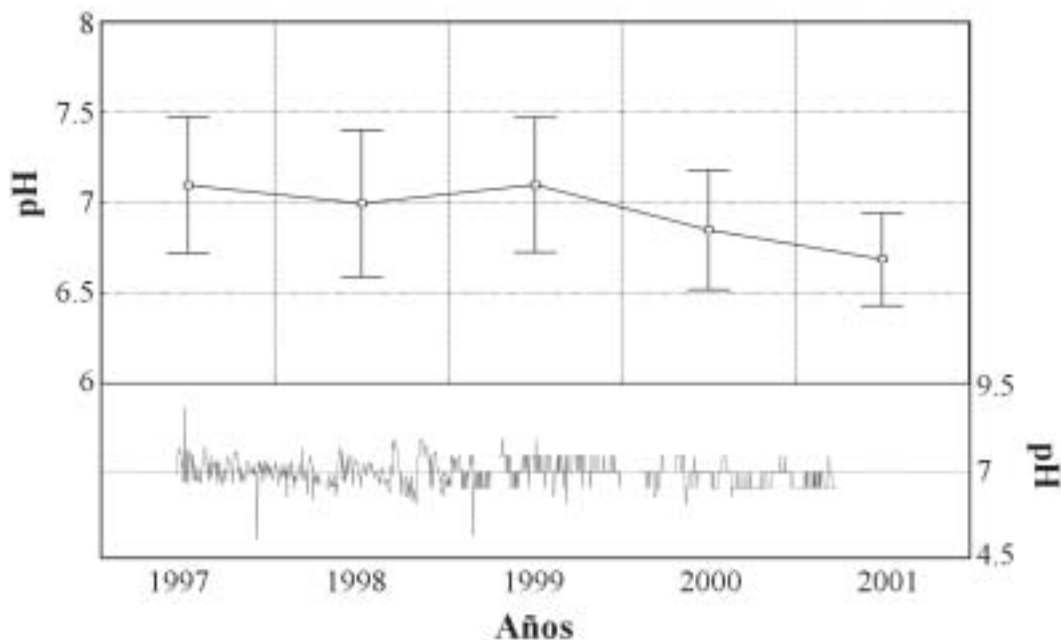
	Suavizado de Holt	ARIMA(1,1,0)	Red neuronal 5:15s:10s:11
<b>Regresión múltiple</b>	0.0322 (0.9985)	0.5040 (0.9180)	10.4568 (0.0151) 42.6756 % <sup>1</sup>
<b>Suavizado de Holt</b>	----	0.7512 (0.8611)	11.4994 (0.0093) 43.5127 % <sup>2</sup>
<b>ARIMA(1,1,0)</b>		----	6.7834 (0.0792)

**Tabla IV.46.** Resultados de los test  $\chi^2$  entre las frecuencias por cuadrantes en los diagramas de Theil de las series 'E' y 'F' para los cuatro modelos seleccionados. Se indica el valor del estadístico  $\chi^2$  así como el nivel de significación del test entre paréntesis

	Suavizado de Holt	ARIMA(1,1,0)	Red neuronal 5:15s:10s:11
<b>Regresión múltiple</b>	0.2562 (0.9680)	0.7629 (0.8583)	1.6509 (0.6479)
<b>Suavizado de Holt</b>	----	0.2577 (0.9678)	1.8163 (0.6114)
<b>ARIMA(1,1,0)</b>		----	1.2628 (0.7380)

#### 4.3.2.12. pH en la serie 'A'

El valor medio anual del pH en la serie 'A' ha variado significativamente a lo largo del período de estudio (ANOVA:  $F=35.7964$ ;  $p_{\alpha}<0.05$ ) (figura IV.41). A pesar de que los valores medios anuales durante los años 1997, 1998 y 1999 se mantienen muy cerca del valor 7, durante el año 2000 y primeros meses del 2001, se observa un descenso significativo del pH (test de Tukey:  $p_{\alpha}<0.05$  en todos los casos excepto cuando se comparan los años 1997 y 1999) que en ocasiones puntuales es de 6.



**Figura IV.41.** Valores medios anuales del pH de la serie 'A' (gráfica superior). Evolución temporal a lo largo del período de estudio (valores absolutos) en la serie 'A' (gráfica inferior)

La regresión múltiple en la que la variable dependiente es el valor de pH y las variables independientes el valor de pH en los cinco días anteriores, ha proporcionado resultados muy satisfactorios tanto en el proceso de calibración como en validación. Así, durante el proceso de calibración la cantidad de varianza explicada fue superior al 95 %. Durante el proceso de validación se observa como el porcentaje de varianza explicada es ligeramente superior al 97 %, lo que puede indicar que en este caso el grupo de validación puede constituirse como un subgrupo de los datos de calibración. De esta forma, el error estándar de predicción y el coeficiente de eficiencia alcanzan niveles más que aceptables (SEP=0.6897 %; E=0.9646). Además no se detectó correlación serial durante la calibración (estadístico de Durvin-Watson=2.0167). El análisis espectral de los residuos obtenidos durante la validación no muestra la presencia de frecuencias con valores de periodograma significativamente grandes. El mayor valor de periodograma (VP=0.02856) se corresponde con la última frecuencia de muestreo seleccionada ( $v=0.5$ ) que equivale a un período de 196 días.

En la regresión anterior, tan solo el valor de pH de los dos días anteriores contribuyeron significativamente a la varianza de la variable dependiente, de tal forma que, se ha procecido a la calibración de una nueva regresión múltiple con tan solo esas dos variables independientes. Los resultados tanto en la fase de calibración como en la de validación son semejantes a los obtenidos anteriormente aunque con ligeras mejorías en los términos de error. Así en este caso, el error estándar de predicción fue de 0.6611 % y el

coeficiente de eficiencia quedó establecido en 0.9675 (tabla IV.47). Del mismo modo que en la regresión anterior, no se detectaron indicios de autocorrelación durante la calibración (Durvin-Watson=2.0359) ni perioricidad en la serie de residuos de validación (figura IV.42).

**Tabla IV.47.** Estimación del pH del agua en la serie 'A' mediante el análisis de regresión múltiple con las variables independientes representadas por el pH filtrado en los dos días anteriores

<b>Resumen de la regresión múltiple en la calibración del modelo</b>			
R=0.9764; R <sup>2</sup> =0.9534; F(2;331)=3386.3; p <sub>α</sub> <0.001; N=334			
Variable dependiente	Variable independiente	b <sub>i</sub> (i=0,1,...,q)	p <sub>α</sub>
pH (t)	Intersección (b <sub>0</sub> )	0.3691	0**
	pH (t-1)	1.4704	0**
	pH (t-2)	-0.5224	0**
<b>Resumen de la regresión en la validación del modelo</b>			
R=0.9857; R <sup>2</sup> =0.9717; F(1;343)=11771.0; p <sub>α</sub> <0.001; N=345			
Error estándar de predicción= 0.6611 %; Coeficiente E=0.9675			
Variable dependiente	Variable independiente	b <sub>i</sub> (i=0,1,...,q)	p <sub>α</sub>
pH observado (t)	Intersección (b <sub>0</sub> )	-0.1333	0.0380*
	pH estimado (t)	1.0172	0**

\* p<sub>α</sub><0.05; \*\* p<sub>α</sub><0.01

Los modelos de suavizado y **ARIMA** también han proporcionado resultados estadísticamente muy aceptables, aunque se han encontrado diferencias en cuanto a la bondad de ajuste entre los distintos tipos de modelos. Así, de forma global, los modelos de suavizado han mostrado un error estándar de predicción que en todo momento ha oscilado alrededor del 7.7 %. En cambio, en los modelos **ARIMA** si exceptuamos el modelo de media móvil de orden 1, el error se ha encontrado al mismo nivel calculado para la regresión múltiple, es decir, alrededor del 0.7 %. También se han presentado importantes diferencias entre el coeficiente de eficiencia del mejor modelo de suavizado (en esta ocasión es el suavizado exponencial simple) y el mejor modelo **ARIMA** con todos sus términos significativos [ARIMA(2,1,2)]. Mientras que el coeficiente de eficiencia en el mejor suavizado ha alcanzado valores negativos (E=-3.0605), en el modelo ARIMA(2,1,2) fue de 0.9752.

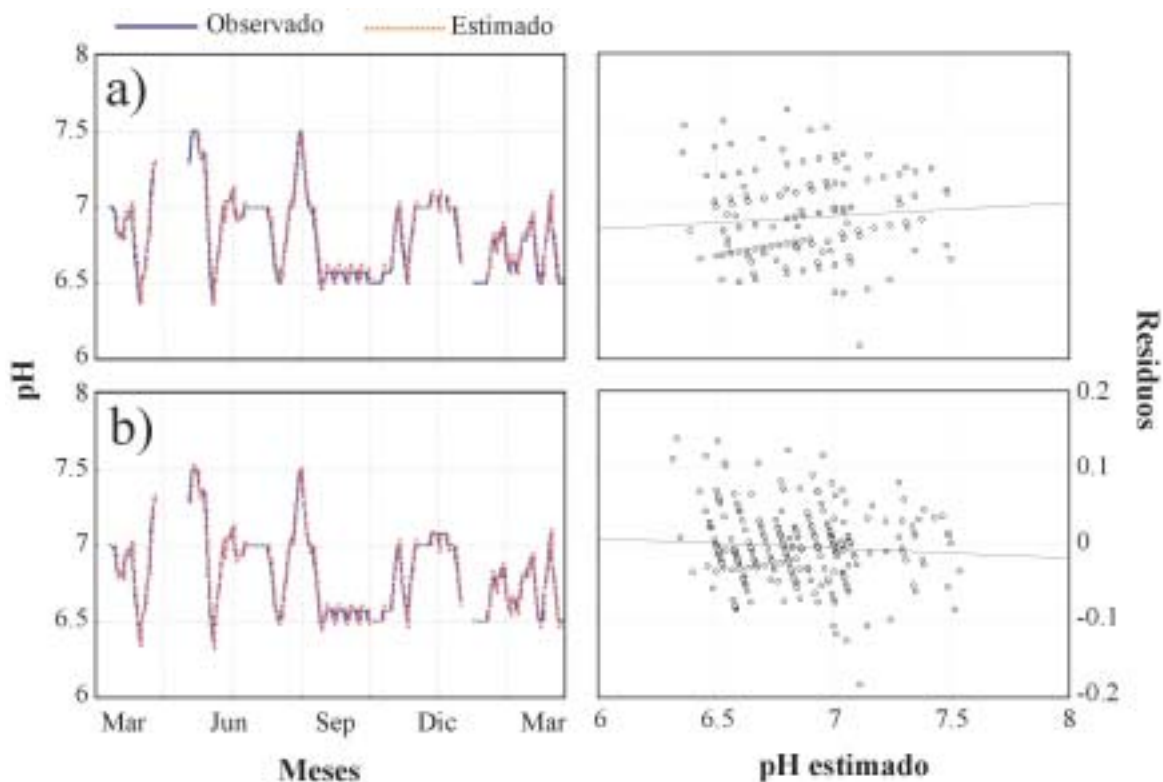


Figura IV.42. pH observado, estimado y residuos para los modelos: a) Regresión múltiple y b) ARIMA(2,1,2)

## 4.4. DISCUSIÓN

### 4.4.1. Temperatura

Como se ha mostrado en el apartado 4.3.2.1, tanto los modelos de regresiones múltiples como los modelos univariantes de series temporales han presentado una significativa capacidad predictiva, al encontrarse en todo momento errores estándar de predicción por debajo del 5 % y coeficientes de eficiencia superiores a 0.95 durante la generalización. Estos resultados se han alcanzado, a diferencia del resto de parámetros físico-químicos considerados, sin la necesidad de llevar a cabo el proceso de filtrado detallado en la sección 4.2.5. A pesar de ello hay que tener en cuenta que, el valor diario de temperatura utilizado es el resultado de hallar el valor promedio para todos los valores de temperatura medidos en un mismo día, lo que en cierta forma constituye un tipo especial de filtro.

De forma global se puede afirmar que para el parámetro temperatura los mejores resultados se alcanzan a través de los modelos **ARIMA**, aunque todos los modelos probados representan correctamente la tendencia general de la serie. Este mismo resultado es encontrado por otros autores como Stergiou *et al.* (1997), quienes comparan la capacidad

predictiva de las capturas mensuales de 16 especies de peces a través de los modelos **ARIMA** y las regresiones múltiples. A pesar de ello, los modelos **ARIMA** tienden a producir valores extremos (tanto positivos como negativos) cuando la variación entre dos valores consecutivos de la serie es brusca, lo que puede estar determinado por la determinación incorrecta del modelo (esto puede ser detectado a través del análisis de la función de autocorrelación residual. Véanse las diferencias entre el modelo **ARIMA**(2,1,2) en la serie 'A' y en las series 'E' y 'F'). Así estos valores contribuyen con más del 90 % al error total cometido durante la generalización.

Tanto en las regresiones múltiples como en los modelos de suavizado no se produce sobreestimación o subestimación de valores cuando en la serie original hay cambios abruptos. A pesar de ello, los errores cometidos durante la generalización se encuentran en todo momento por encima del calculado para el mejor modelo **ARIMA**. En este caso, el error cometido para cada valor es pequeño aunque repartido equitativamente para todos los valores de la serie. Esto es consecuencia de un fenómeno de retraso temporal entre la serie real y la estimada durante la generalización. El efecto de retraso viene determinado por la formulación matemática de este tipo de modelos. Así, al calibrar regresiones múltiples en las que las variables independientes son las mismas que la variable dependiente pero retrasadas en unidades de tiempo, se obtiene en todos los casos que la variable independiente que contribuye en un mayor porcentaje a la varianza de la variable dependiente es la que se corresponde con el valor del parámetro que deseamos predecir retrasado en una unidad de tiempo. Es decir, si tratamos de predecir el valor de la temperatura para el momento  $t$  utilizando una regresión múltiple, la variable independiente que va a contar con un mayor peso en el modelo es la temperatura del momento  $t-1$ . De esta forma el coeficiente  $b_1$  de esa variable independiente siempre se encontrará muy próximo a uno y siempre será significativo. Lo mismo sucede en el suavizado exponencial simple y en el suavizado de Holt. En el primero de estos dos modelos, el efecto de retraso temporal se produce cuando el parámetro  $\alpha_d$  se encuentra muy próximo a uno mientras que en el segundo caso se tiene que cumplir que  $\alpha_d$  sea prácticamente uno y  $\beta_d$  se encuentre muy cerca de cero.

El efecto de retraso temporal ha sido claramente observado en los trabajos de otros autores que trataban de predecir parámetros que evolucionan en función del tiempo mediante distintas metodologías univariantes. De esta forma Thirumalaiah y Deo (2000) estimaron tanto la escorrentía superficial horaria y el estado de los niveles diarios de ríos en la zona central de la India a través de redes neuronales computacionales, observándose el retraso en las estimaciones y provocando que en la mayor parte de los modelos calibrados los coeficientes de eficiencia no superasen el nivel 0.8. También Yang *et al.* (1997b) muestran el retraso en alguno de sus modelos de redes neuronales para la predicción del drenaje superficial en tierras de cultivo. Los modelos de regresión de ventana móvil y **ARIMAs** calibrados por Junttila (2001) sobre la predicción de la inflación, sufren retraso en la generalización en función de la anticipación en la predicción aplicada durante la calibración, lo que implica valores del coeficiente de desigualdad de Theil que en todo momento se encuentran por encima de 0.8. Por otra parte, Prybotuk *et al.* (2000) tratando de predecir la

concentración de ozono en la ciudad de Houston, usaron modelos **ARIMA** y regresiones múltiples obteniendo errores estándar de predicción durante la generalización superiores al 90 %. Este elevado error está íntimamente relacionado con el fenómeno de retraso temporal. Asimismo, Pulido-Calvo (2001) y Pulido-Calvo *et al.* (2002) detectan el fenómeno de retraso al predecir las demandas diarias de agua en una comunidad de regantes de la provincia de Córdoba. Las causas del retraso entre la serie a estimar y la estimada expuestas anteriormente contrastan con las publicadas por otros autores. Así, Park (1998) explica el retraso entre las series de capturas de *Theragra chalcogramma* real y estimada a través de un modelo **ARIMA** como consecuencia de la influencia de factores bióticos (zooplankton, etc.) y abióticos (temperatura, etc.).

Tal y como se comenta en el apartado 2.2.2, la principal razón por la que Hidrorecursos se construyó junto a la Central Térmica de Puente Nuevo fue el aprovechamiento del agua caliente procedente del efluente de refrigeración de la Central Térmica. El análisis de las temperaturas ambientales y del agua muestra como durante todo el período de estudio la temperatura del agua se encuentra siempre por encima de las temperaturas ambientales, gracias al aprovechamiento del efluente de agua caliente. La recirculación del agua en algunos sistemas ha favorecido además la conservación del calor dentro de la planta. Esto ha permitido mantener la mayor parte del tiempo la temperatura media dentro de los rangos de tolerancia y crecimiento rápido de la especie.

El cambio en la forma de generación de energía a partir del año 1998 como consecuencia de la entrada en vigor de la Ley 54/1997 del Sector Eléctrico (BOE 28-11-1997) ha introducido potenciales desventajas. La aplicación de la nueva Ley implica que la energía generada en cada período de programación será retribuida al coste marginal del sistema eléctrico basado en la oferta realizada por el último grupo de generación cuya puesta en marcha haya sido necesaria para atender la demanda. De esta forma, aquellas centrales generadoras que lleven a cabo una producción de energía más barata (centrales nucleares y grandes centrales térmicas o de ciclo combinado) proporcionarán un volumen de agua a una temperatura más constante. La Central Térmica de Puente Nuevo puede llegar a generar una potencia máxima de 318 MW, por lo que se puede considerar una planta generadora de tamaño pequeño. Además la generación se lleva a cabo a partir de la quema de carbón. Estos dos factores intervienen de forma definitiva para que la producción disminuya de forma drástica cuando los requerimientos energéticos son menores (normalmente los fines de semana). De esta forma se favorece una aclimatación de los individuos de las series en circuito abierto (series 'D', 'C' y 'G') a temperaturas más bajas durante el fin de semana y se provoca estrés térmico durante los primeros días de la semana cuando la central genera más energía. Este efecto de aclimatación-estrés térmico no se detecta en las series de circuito cerrado (series 'A', 'B', 'E' y 'F') como consecuencia de la amortiguación de los cambios de temperatura llevada a cabo por parte del biofiltro.

#### 4.4.2. Amoniaco

La estimación del parámetro amoniaco a partir de los datos originales ha supuesto niveles de error muy alejados de los rangos estadísticos de aceptación para todos los modelos empleados (regresiones múltiples, suavizados, **ARIMAs** y **RNCs**). Esta falta de bondad de ajuste puede deberse a que la concentración de amoniaco en el agua está fuertemente influenciada por el manejo que se hace del filtro biológico. De esta forma, debido a que continuamente se llevan a cabo distintas manipulaciones del biofiltro de forma directa o indirecta (cambio en los caudales circulantes, limpieza, inyección de aire, aplicación de tratamientos, etc), es probable que la variable amoniaco se comporte para esta frecuencia de muestreo como una variable de ruido blanco. Es decir, en esencia, la serie original de la concentración de amoniaco sería una serie puramente aleatoria de números y por lo tanto no correlacionados. Esto podría explicar el hecho de que con independencia del modelo empleado el error estándar de predicción se encontrase siempre alrededor del 100 %.

Sin embargo, el análisis de las funciones de autocorrelación serial y autocorrelación parcial serial de la serie original descarta en cierta forma la posibilidad anterior. Así, en función del nivel de significación de los distintos retardos en ambas funciones se puede establecer, o al menos intuir, la existencia de un proceso autorregresivo de orden 1 (una caída suave de la función de autocorrelación a lo largo de nueve retardos y una única aguja significativa en la función de autocorrelación parcial) tal y como establece Holton Wilson y Keating (1996). La aplicación del modelo [ARIMA(1,0,9)] incrementa significativamente el porcentaje de varianza explicada y produce una función de autocorrelación residual donde ninguno de sus términos son significativos. A pesar de ello, el error estándar de predicción sigue encontrándose por encima del 100 %. Es decir, la falta de correlación serial en los residuos de la generalización se consigue a base de producir valores extremos (tanto positivos como negativos) en la función estimada. Este mismo efecto es encontrado por Hsu *et al.* (1995) en la simulación de la escorrentía mediante la aplicación de modelos **ARMAX**. Asimismo, Pulido-Calvo *et al.* (2002) detectan este fenómeno al intentar predecir las demandas diarias de agua en una comunidad de regantes.

La presencia de correlación serial también podría explicar en parte la falta de bondad de ajuste de los modelos lineales. Si se da correlación serial, pueden producirse problemas en el uso e interpretación de los modelos lineales y sobre todo de la función de regresión por mínimos cuadrados ordinarios. Esto es debido a que a pesar de que la correlación serial no induce sesgos en los coeficientes estimados, hace que las estimaciones de los errores estándar sean más pequeñas que los errores estándar verdaderos. Esto significa que las relaciones **t** calculadas para cada coeficiente sean sobreestimadas, lo que a su vez puede conducir al rechazo de hipótesis nulas que no deberían rechazarse. Es decir, los coeficientes de regresión pueden considerarse estadísticamente significativos cuando en realidad no lo son o al contrario, pueden considerarse no significativos cuando si lo son. Además, la existencia de correlación serial hace que el coeficiente de determinación y los estadísticos **F** no sean fiables al evaluar la significación global de la función de regresión.



El efecto de la correlación serial puede ser relativamente importante en los modelos lineales y especialmente en las regresiones múltiples, ya que de esta forma se incumple uno de los requisitos fundamentales para la aplicación de este tipo de modelos (Hair *et al.*, 1999). Esta condición no es imprescindible en modelos como las redes neuronales computacionales, por lo que han sido utilizadas en la resolución de problemas complejos con gran éxito (García-Villares *et al.*, 1995). A pesar de ello, los resultados obtenidos al aplicar las **RNCs** a las series de amoníaco originales son desde un punto de vista estadístico tan poco aceptables como en el resto de modelos, al alcanzarse errores estándar de predicción superiores al 100 %. La fuente de error en la fase de generalización, tal y como se explicó en la sección 4.4.1, se encuentra en el retraso temporal observado entre la serie original y la estimada. Esto puede apoyar nuevamente la hipótesis de que la serie original tendiese a comportarse como una serie de ruido blanco. De esta forma también se podría explicar la ausencia de correlación significativa entre los términos de error (**SEP** y coeficiente **E**) y la complejidad de la arquitectura neuronal.

La estimación de la concentración de amoníaco a partir de otras variables implicadas como la temperatura del agua, el pH, la carga de amoníaco en el biofiltro, los kilogramos de pienso suministrados o la tasa de eliminación de amoníaco en el biofiltro, tampoco ha resultado estadísticamente satisfactoria. En este sentido, los resultados obtenidos son semejantes a los alcanzados por Kamstra *et al.* (1998). Esto puede ser debido a que los valores observados para esos parámetros presentan rangos de variación relativamente pequeños. De entre todas las variables utilizadas destaca por su poca influencia la cantidad de alimento suministrado. La relación entre los kilogramos de pienso suministrado y la concentración del amoníaco ha sido apuntada por otros autores como Heinsbroek y Kamstra (1990) y Kamstra *et al.* (1998). Estos señalan que la razón de la baja correlación entre cantidad de alimento y concentración de amoníaco puede ser debida a la ausencia de un patrón de suministro-utilización del alimento y la producción de desechos a lo largo del tiempo. Esto puede verse favorecido por el hecho de que en la mayor parte de las piscifactorías de anguilas, la alimentación se lleva a cabo mediante comederos de autodemanda (también en las series de preengorde de Hidrorecursos) lo que implica que la tasa instantánea de consumo de alimento no puede ser calculada de forma precisa.

La estimación de la concentración de amoníaco a partir de la serie original filtrada produce estimaciones correctas y términos de error que se pueden considerar aceptables. Con el proceso de filtrado definido en la sección 4.2.5 se lleva a cabo un cambio de escala de la serie original, aumentando amplitudes y disminuyendo frecuencias. De esta forma, se consigue espaciar temporalmente eventos de elevación o disminución brusca del parámetro a estimar, que es el origen del aumento del error. En definitiva, se consigue un efecto de suavizado gracias al cual se elimina el ruido e imprecisiones presentes en la serie original. Otra razón por la que se obtienen buenas estimaciones es el cambio en la naturaleza de la serie. Se pasa de una serie no lineal a una serie filtrada lineal (Bloomfield, 1976). Esto explica el buen funcionamiento de los modelos lineales (regresiones múltiples, suavizados y modelos **ARIMA**) y del correcto funcionamiento de las redes neuronales computacionales.

En este sentido, hay que tener en cuenta que debido al método de calibración de las **RNCs** no se garantiza el alcanzar el mínimo absoluto de la función de error a minimizar, por lo que frente a una función lineal las **RNCs** pueden proporcionar estimaciones menos correctas que las llevadas a cabo mediante las regresiones múltiples o los modelos **ARIMA**. A pesar de ello, la mayor capacidad de generalización de las **RNCs** queda patente cuando las condiciones de estimación se hacen más cambiantes, como por ejemplo la estimación de la concentración de amoníaco en las series de preengorde a partir de los modelos calibrados en la serie 'A' (SEP=17.4983 y U=0.6134).

#### 4.4.3. Nitritos y nitratos

Partiendo de los datos previamente suavizados, los resultados alcanzados tanto en la estimación de las concentraciones de nitritos como en las de nitratos, son estadísticamente tan correctos como en el caso anterior. En el parámetro nitrito nuevamente existe una tendencia de los modelos lineales a proporcionar mejores estimaciones que las **RNCs**, aunque tan solo de forma absoluta. Se puede observar que cuando las condiciones de manejo son muy distintas a las existentes en la serie en la que originalmente se han calibrado los modelos (por ejemplo: la extensión de los modelos calibrados en la serie 'A' hacia las series 'E' y 'F'), porcentualmente las mayores disminuciones en los términos de error se consiguen con las **RNCs**, aunque de forma absoluta los modelos lineales siguen siendo mejores. Esto es debido a que al partir de la serie filtrada los modelos lineales (regresiones múltiples, suavizados y modelos **ARIMA**) cuentan con cierta ventaja en su calibración. La falta de afinidad de las **RNCs** en el caso de los nitritos puede estar indicando una deficiencia en la calibración de estos modelos. De esta forma, se podría explicar la relación encontrada entre el número de pesos y el error estándar de predicción en las **RNCs** calibradas para este parámetro.

La mejora relativa de las **RNCs** cuantificada en el parámetro nitrito no se observa en el parámetro nitrato. En este caso, la extensión de este modelo calibrado en la serie 'A' hacia las series de preengorde produce peores resultados tanto de forma absoluta como relativa. De la misma forma que en el caso anterior, parece intuirse una falta de adecuación del modelo a los datos, aunque en esta ocasión, puede estar condicionada por un proceso de sobreajuste del modelo durante la calibración. Así, puede observarse como de entre todos los modelos calibrados para la serie 'A', la mayor pendiente significativa para la regresión simple entre los nitratos estimados y los residuos de la estimación es encontrada en el mejor modelo de red neuronal. Esta correlación significativa entre valores estimados y residuos también se manifiesta en los diagramas de predicción-realización, donde la regresión simple entre la tasa de variación de los datos observados y la tasa de variación de los datos estimados que presenta una pendiente más alejada de uno es calculada para la red neuronal. Sin embargo, a pesar de lo comentado anteriormente, los mejores valores tanto para la serie 'B' como en ambas series de preengorde del coeficiente de desigualdad Theil se consiguen con el mejor modelo de red neuronal. Esta aparente contradicción puede tener su respuesta

en el hecho de que es la red neuronal el único modelo que no reduce el error estándar de predicción cuando éste es recalculado al retrasar artificialmente la serie original con respecto a la estimada.

#### 4.4.4. pH

Los modelos calibrados para la estimación del pH son los que han obtenido los mejores resultados durante el proceso de la generalización, ya que los errores se han encontrado en todo momento por debajo del 1 %. Estos buenos resultados parecen estar influenciados por la escala y el rango de los valores a estimar. Este fenómeno ya ha sido apuntado por otros autores que emplean medidas de agudeza relativas como el error estándar de predicción (**SEP**) o el porcentaje de error medio absoluto (**MAPE**) (Noakes *et al.*, 1985).

#### 4.5. REFERENCIAS

- Adelman, I.R. y L.L. Smith. 1970. Effect of oxygen on growth and food conversion efficiency of northern pike. *Progressive Fish-Culturist*, 32: 93-96.
- Albiol, J., C. Campmajo, C. Casas y M. Poch. 1995. Biomass estimation in plant cell cultures: a neural network approach. *Biotechnology Progress*, 11: 88-92.
- Alcaraz, G. y S. Espina. 1995. Acute toxicity of nitrite in juvenile grass carp modified by weight and temperature. *Bulletin of Environmental Contamination and Toxicology*, 55: 473-478.
- Allan, G.L. y G.B. Maguire. 1991. Lethal levels of low dissolved oxygen and effects of short-term oxygen stress on subsequent growth of juvenile *Panaeus monodon*. *Aquaculture*, 94: 27-37.
- Almendras, J.M.E. 1987. Acute toxicity and methemoglobinemia in juvenile milkfish (*Chanos chanos* Forsskae). *Aquaculture*, 61: 33-40.
- Álvarez, J. y S. Bolado. 1996. Descripción de los procesos de infiltración mediante redes neurales artificiales. *Ingeniería del Agua*, 3(2): 39-46.
- Amit, D., H. Gutfreund y H. Sompolinsky. 1985. Spin-glass models of neural networks. *Physical Review*, 32A: 1007-1018.
- Aston, R.J. 1981. The availability and quality of power station cooling water for aquaculture. En *Proceedings of the World Symposium on Aquaculture 'Heated Effluents and Recirculation System'*. Berlín (Alemania).

- Avnimelech, Y., M. Kochva y S. Diab. 1994. Development of controlled intensive aquaculture system with a limited water exchange and adjusted carbon to nitrogen ratio. *The Israeli Journal of Aquaculture*, 46(3): 119-131.
- Aziz, A.R.A. y K.F.V. Wong. 1992. Neural network approach to the determination of aquifer parameters. *Ground Water*, 30: 164-166.
- Baran, P., S. Lek, M. Delacoste y A. Belaud. 1996. Stochastic models that predict trouts population densities or biomass on macrohabitat scale. *Hydrobiologia*, 337: 1-9.
- Bevelhimer, M. y W. Bennett. 2000. Assessing cumulative thermal stress in fish during chronic intermittent exposure to high temperatures. *Environmental Science and Policy*, 3: 211-216.
- Beg, S.A., M.M. Hassan y M.A.S. Chaudhry. 1995. Multi-substrate analysis of carbon oxidation and nitrification in an upflow packed-bed biofilm reactor. *Journal of Chemical and Technical in Biotechnology*, 64: 367-378.
- Bejda, A.J., B.A. Phelan y A.L. Studholme. 1992. The effect of dissolved oxygen on the growth of young-of-the-year winter flounder, *Pseudopleuronectes americanus*. *Environmental Biology of Fishes*, 34: 321-327.
- Blank, T.B. y S.D. Brown. 1993. Non-linear multivariate mapping of chemical data using feed-forward neural networks. *Analytical Chemistry*, 65: 3081-3089.
- Bloomfield. 1976. *Fourier analysis of time series: an introduction*. Edita: John Wiley & Sons. New York.
- Blum, A.L. y R.L. Rivest. 1992. Training a 3-node neural network is NP-complete. *Neural Networks*, 5: 117-127.
- BOE 28-11-1997. *Ley 54/1997, de 27 de noviembre, del Sector Eléctrico*, núm. 285: 35097-35126.
- Boyd, C.E. 1997. Practical aspects of chemistry in pond aquaculture. *The Progressive Fish-Culturist*, 59: 85-93.
- Boyd, C.E., B.J. Watten, V. Goubier y R. Wu. 1994. Gas supersaturation in surfaces waters of aquaculture ponds. *Aquacultural Engineering*, 13: 31-39.
- Boyd, C.E., R.P. Romaine y E. Johnston. 1978. Predicting early orning dissolved oxygen concentrations in chanel catfish ponds. *Transactions of the American Fisheries Society*, 107: 484-492.

- Brett, J.R. y J.M. Blackburn. 1981. Oxygen requirement for growth of young coho (*Oncorhynchus kisutch*) and sockeye (*O. nerka*) salmon at 15 °C. *Canadian Journal of Fisheries and Aquatic Sciences*, 38: 399-404.
- Bryson, A.E. y Y.C. Ho. 1969. *Applied optimal control*. Edita: Blaisdell. New York.
- Buentello, J.A., D.M. Gatlin III y W.H. Neill. 2000. Effects of water temperature and dissolved oxygen on daily feed consumption, feed utilization and growth of channel catfish (*Ictalurus punctatus*). *Aquaculture*, 182: 339-352.
- Burel, C., J. Person-Le Ruyet, F. Gaumet, A. Le Roux, A. Severe y G. Boeuf. 1996. Effects of temperature on growth and metabolism in juvenile turbot. *Journal of Fish Biology*, 49: 678-692.
- Burns, J.A. y G.M. Whitesides. 1993. Feed-forward neural networks in chemistry: mathematical systems for classification and pattern recognition. *Chemical Review*, 93: 2583-2601.
- Burns, J.R. 1976. The reproductive cycle and its environmental control in the pumpkinseed, *Lepomis gibbosus* (Pisces Centrarchidae). *Copeia*, 3: 449-455.
- Cabrera, E. y A.F. Vela. 1994. *Mejora del rendimiento y de la fiabilidad en sistemas de distribución de agua*. Edita: Cabrera y Vela, Universidad Politécnica de Valencia. Valencia.
- Caianello, E.R. 1961. Outline of a theory of thought and thinking machines. *Journal of Theoretical Biology*, 1: 204-235.
- Cairns, J., A.G. Heath y B.C. Parker. 1975. The effects of temperature upon the toxicity of chemical to aquatic organisms. *Hydrobiology*, 47: 135-171
- Colasanti, R.L. 1991. Discussions of the possible use of neural network algorithms in ecological modelling. *Binary*, 3: 13-15.
- Coll-Morales, J. 1991. *Acuicultura marina animal*. Edita: Mundi-Prensa. Madrid.
- Colt, J. y K. Orwicz. 1991. Modeling production capacity of aquatic culture system under freshwater conditions. *Aquacultural Engineering*, 10: 1-29.
- Colt, J., R. Ludwig, G. Tchobanoglous y J.J. Cech. 1981. The effects of nitrite on the short term growth and survival of channel catfish, *Ictalurus punctatus*. *Aquaculture*, 24: 111-122.
- Coulbeck, B. 1988. *Computer control of water supply*. Edita: Leicester Polytechnic. Leicester.

- Cravener, T.L. y W.B. Roush. 1999. Improving neural network prediction of amino acid levels in feed ingredients. *Poultry Science*, 78: 983-991.
- Cruz-Neto, A.P. y J.F. Steffensen. 1997. The effects of acute hypoxia and hypercapnia on oxygen consumption of the freshwater european eel. *Journal of Fish Biology*, 50: 759-769.
- Culberson, S.D. y R.H. Piedrahita. 1996. Aquaculture pond ecosystem model: temperature and dissolved oxygen prediction – mechanism and application. *Ecological Modelling*, 89: 231-258.
- Cybenco, G. 1989. Approximation by superpositions of a sigmoidal function. *Mathematics of Controls, Signals, and Systems*, 2: 303-314.
- Dagorn, L., F. Menezes, P. Bach y R.J. Olson. 2000. Co-evolution of movement behaviours by tropical pelagic predatory fishes in response to prey environment: a simulation model. *Ecological Modelling*, 134: 325-341.
- Dagorn, L., M. Petit y J.M. Stretta. 1997. Simulation of large-scale tropical tuna movements in relation with daily remote sensing data: the artificial life approach. *Byosystem*, 44: 167-180.
- Degani, G. 1986. Dietary effect of lipids source, lipid level, temperature and fatty acid growth of glass eel (*Anguilla anguilla*). *Aquaculture*, 56: 207-214.
- Degani, G. y M. Gallagher. 1985. The relationship between growth, food conversion and oxygen consumption in developed and underdeveloped American eels, *Anguilla rostrata* L. *Journal of Fish Biology*, 27: 635-641.
- Degani, G., D. Levanon y M.L. Gallagher. 1988. Relationship between growth, food conversion, body size, body composition and temperature in the European eel, *Anguilla anguilla* L. *Aquaculture and Fisheries Management*, 19: 139-143.
- Dibike, Y.B., D. Solomatine y M.B. Abbott. 1999. On the encapsulation of numerical-hydraulic models in artificial neural network. *Journal of Hydraulic Research*, 37: 147-161.
- Dorson, M. y C. Torchy. 1981. The influence of fish age and water temperature on mortalities of rainbow trout, *Salmo gairdneri* Richardson, caused by an European strain of infectious pancreatic necrosis virus. *Journal of Fish Diseases*, 4: 213-221.
- Dosoretz, C. y G. Degani. 1987. Effect of fat rich diet and temperature on growth and body composition of european eels (*Anguilla anguilla*). *Comparative Biochemistry and Physiology*, 87A(3): 733-736.

- Doudoroff, P. y D.L. Shumway. 1970. Dissolved oxygen requirements of freshwater fishes. *FAO Fisheries Technical Paper*, 86.
- Dougherty, M.S., H.R. Kirby y R.D. Boyle. 1993. Use of neural networks to recognize and predict traffic congestion. *Traffic Engineering and Control*, 34: 311-314.
- Drapcho, C.M. y D.E. Brune. 2000. The partitioned aquaculture system: impact of design and environmental parameters on algal productivity and photosynthetic oxygen production. *Aquacultural Engineering*, 21: 151-168.
- Dreyfus, H.L. y S.E. Dreyfus. 1999. Fabricar una mente versus modelar el cerebro: la inteligencia artificial se divide de nuevo. En: *El nuevo debate sobre la inteligencia artificial. Sistemas simbólicos y redes neuronales*. S.R. Graubard. Edita: Gedisa. Barcelona.
- Edwards, M. y D.R. Morse. 1995. The potential for computer-aided identification in biodiversity research. *Trends in Ecology and Evolution*, 10: 153-158.
- Egusa, S. 1958. On the oxygen consumption rate of the pond cultured eel, *Anguilla japonica*. *Japanese Journal of Ichthyology*, 7: 24-56.
- Elkordy, M.F., K.C. Chang y G.C. Lee. 1993. Neural networks trained by analytically simulated damage states. *Journal of Computing in Civil Engineering*, 7: 130-145.
- Elliott, J.M. 1982. The effects of temperature and ration size on the growth and energetics of salmonids in captivity. *Comparative Biochemistry and Physiology*, 73B: 81-91.
- Ellis, G.W., C. Yao, R. Zhao y D. Penumadu. 1995. Stress-strain modeling of sands using artificial neural networks. *Journal of Geotechnical Engineering*, 121: 429-435.
- Faraggi, D. y R. Simon. 1995. A neural network model for survival data. *Statistics in Medicine*, 14: 73-82.
- Fernando, D.A.K. y A.W. Jayawardena. 1998. Runoff forecasting using RBF networks with OLS algorithm. *Journal of Hydrologic Engineering*, 3: 203-209.
- Fiorino, P. y S. Mancuso. 1998. Aplicación de una red neuronal artificial para la identificación de 24 cultivares de olivo. *Olivae*, 73: 37-43.
- Fivelstad, S. 1988. Waterflow requirements for salmonids in single pass and semi-closed land-based seawater and freshwater system. *Aquacultural Engineering*, 7: 183-200.
- Forsberg, O.I. 1994. Modelling oxygen consumption rates of post-smolt atlantic salmon in commercial-scale, land-based farms. *Aquaculture International*, 2: 180-196.

- Foster, J.R.M. 1974. Studies on nitrification in marine biological filters. *Aquaculture*, 4: 387-397.
- García-Villares, J.L., J.A. Blasco-Alberto, B. Martín del Brío, I.J. Ramírez-Rosado, J. Barquillas-Pueyo, J.A. Domínguez-Navarro, N. Medrano-Marqués. 1995. Short-term electric power load-forecasting using artificial neural networks. Part II. Multilayer perceptron networks for hourly electric-demand forecasting. *Proceedings of the Fourteenth IASTED International Conference Modeling, Identification and Control*, 223-227. Austria.
- Giovannini, P. y R.H. Piedrahita. 1994. Modeling photosynthetic production optimization for aquaculture ponds. *Aquacultural Engineering*, 13: 83-100.
- Giske, J., G. Huse y O. Fiksen. 1998. Modelling spatial dynamics of fish. *Review on Fish Biology and Fisheries*, 8: 57-91.
- Govindaraju, R.S. 2000. Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering*, 5: 124-137.
- Guegan, J.F., S. Lek y T. Oberdorff. 1998. Energy availability and habitat heterogeneity predict global riverine fish diversity. *Nature*, 391: 382-384.
- Gujer, W. y M. Boller. 1986. Design of a nitrifying tertiary trickling filter based on theoretical concepts. *Water Resources*, 20(11): 1353-1362.
- Gutiérrez-Estrada, J.C., I. Pulido-Calvo y J. Prenda. 2000. Gonadosomatic index estimates of an introduced pumpkinseed (*Lepomis gibbosus*) population in a Mediterranean stream, using computational neural networks. *Aquatic Science*, 62: 350-363.
- Hair, J.F., R.E. Anderson, R.L. Tatham y W.C. Black. 1999. *Análisis multivariante*. Edita: Prentice Hall. Madrid.
- Hart, P.R. y G.J. Purser. 1995. Effects of salinity and temperature on eggs and yolk sac larvae of the greenback flounder (*Rhombosolea tapirina* Günther, 1862). *Aquaculture*, 136: 221-230.
- Hart, P.R., W.G. Hutchinson y G.J. Purser. 1996. Effects fo photoperiod, temperature and salinity on hatchery-reared larvae of the greenback flounder (*Rhombosolea tapirina* Günther, 1862). *Aquaculture*, 144: 303-311.
- Hartnett, M., D. Diamond y P.G. Barker. 1993. Neural networks based recognition of flow injection patterns. *Analyst*, 118: 347-354.
- Heath, A.G. 1977. Toxicity of intermittent chlorination of freshwater fish: influence of temperature and chlorine form. *Hydrobiologia*, 56: 39-47.



- Hebb, D.O. 1949. *The organization of behavior*. Edita: Wiley. New York.
- Heinsbroek, L.T.N. y A. Kamstra. 1990. Design and performance of water recirculation systems for eel culture. *Aquaculture Engineering*, 9: 187-207.
- Hermann, R.B., C.E. Warren y P. Doudoroff. 1962. Influence of oxygen concentration on the growth of juvenile coho salmon. *Transactions of the American Fisheries Society*, 91: 155-167.
- Hervás, C., A. Garrido, B. Lucena, N. García y E. De Pedro. 1994. Near infrared spectroscopy for classification of Iberian pig carcasses using an artificial neural network. *Journal of Near Infrared Spectroscopy*, 2: 177-184.
- Hervás, C., S. Ventura, M. Silva y D. Pérez-Bendito. 1998. Computational neural networks for resolving nonlinear multicomponent systems based on chemiluminescence methods. *Journal of Chemical Informatic and Computer Science*, 38: 1119-1124.
- Hilmy, A.M., N.A. El-Domiaty y K. Wershana. 1987. Acute and chronic toxicity of nitrite to *Clarias lazera*. *Comparative Biochemistry and Physiology*, 86: 247-253.
- Hinton, G.E. y J.A. Anderson. 1981. *Parallel models of associative memory*. Edita: Lawrence Erlbaum Associates, Potomac. Maryland.
- Hinton, G.E. y T. Sejnowski. 1983. Optimal perceptual inference. *IEEE Proceedings of the Computer Society. Conference on Computer Vision and Pattern Recognition*, Washington, D.C., IEEE Computer Society Press, 448-453.
- Hjelmfelt, A.T. y M. Wang. 1996. Predicting runoff using artificial neural networks. *Surface-Water Hydrology*, 233-244.
- Holton-Wilson, J. y B. Keating. 1996. *Previsiones en los negocios*. Edita: IRWIN. Illinois.
- Hopfield, J.J. 1982. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Science*, 79: 2554-2558.
- Hopkins, J.S., R.D. Hamilton, P.S. Sandifer, C.L. Browdy y A.D. Stokes. 1993. Effect of water exchange rate on production, water quality, effluent characteristics and nitrogen budgets of intensive shrimp ponds. *Journal of World Aquacultural Society*, 24(3): 304-320.
- Hsu, K., H.V. Gupta y S. Sorooshian. 1995. Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research*, 31: 2517-2530.

- Hsu, K., H.V. Gupta, X. Gao y S. Sorooshian. 1999. Estimation of physical variables from multichannel remotely sensed imagery using a neural network: application to rainfall estimation. *Water Resource Research*, 35: 1605-1618.
- Hsu, K., X. Gao, S. Sorooshian y H.V. Gupta. 1997. Precipitation estimation from remotely sensed information using artificial neural networks. *Journal of Applied Meteorology*, 36: 1176-1190.
- Huey, D.W., B.A. Simco y D.W. Criswell. 1980. Nitrite-induced methemoglobin formation in channel carfish. *Transactions of the American Fisheries Society*, 109: 558-562.
- Huey, D.W., T.L. Beitinger y M.C. Wooten. 1984. Nitrite-induced methemoglobin formation and recovery in channel catfish (*Ictalurus punctatus*) at three acclimation temperatures. *Bulletin of Environmental Contamination and Toxicology*, 32: 674-681.
- Huner, J.V. y H.K. Dupree. 1984. *Pond culture systems*. En: *Third report to the fish farmers*. Edita: U.S. Department of the Interior Fish and Wildlife Service. Washington.
- Hunnam, P., A. Milne y P. Stebbing. 1982. *El acuario vivo. Agua dulce y salada*. Edita: Libros del Carballo. Madrid.
- Huse, G. y J. Giske. 1998. Ecology in mare pentium: an individual-based spatio-temporal model for fish with adapted behaviour. *Fisheries Research*, 37: 163-178.
- Islam, S. y R. Kothari. 2000. Artificial neural networks in remote sensing of hydrologic processes. *Journal of Hydrologic Engineering*, 5: 138-144.
- Israeli-Weinstein, D. y E. Kimmel. 1998. Behavioral response of carp (*Cyprinus carpio*) to ammonia stress. *Aquaculture*, 165: 81-93.
- Johnson, V.M. y L.L. Rogers. 2000. Accuracy of neural network approximators in simulation-optimization. *Journal of Water Resources Planning and Management*, 126: 48-56.
- Judd, J.S. 1990. *Neural network design and the complexity of learning*. Edita: MIT Press, Cambridge. Massachusetts.
- Junttila, J. 2001. Structural breaks, ARIMA model and finnish inflation forecasts. *International Journal of Forecasting*, 17: 203-230.
- Kamstra, A., J.W. Van der Heul y M. Nijhof. 1998. Performance and optimisation of trickling filters on eel farms. *Aquacultural Engineering*, 17: 175-192.

- Kartam, N., I. Flood y J.H. Garrett Jr. 1997. *Artificial neural networks for civil engineers: fundamentals and applications*. Edita: American Society of Civil Engineers, ASCE. New York.
- Karunanithi, N., W.J. Grenney, D. Whitley y K. Bovee. 1994. Neural networks for river flow prediction. *Journal of Computing in Civil Engineering*, 8: 201-220.
- Kentala, E., I. Pyykkö, Y. Auramo y M. Juhola. 1997. Neural networks in neurotologic expert systems. *Acta Otolaryngol*, 529: 127-129.
- Kinkelin, P, Ch. Michel y P. Ghittino. 1991. *Tratado de las enfermedades de los peces*. Edita: ACRIBIA, S.A. Zaragoza.
- Kitanidis, P.K. y R.L. Bras. 1980. Real time forecasting with a conceptual hydrological model. 2. Applications and results. *Water Resources Research*, 16: 1034-1044.
- Klemetson, S.L. y G.L. Rogers. 1985. Aquaculture pond temperature modeling. *Aquaculture Engineering*, 4: 191-208.
- Kochba, M., S. Diab y Y. Avnimelech. 1994. Modeling of nitrogen transformation in intensive aerated fish ponds. *Aquaculture*, 120: 95-104.
- Krise, W.F. 1991. Hatchery management of lake trout exposed to chronic dissolved gas supersaturation. *American Fisheries Society Symposium*, 10: 368-371.
- Lacroix, R., F. Salehi, X.Z. Yang y K.M. Wade. 1997. Effects of data preprocessing on the performance of artificial neural networks for dairy yield prediction and cow culling classification. *Transactions of the ASAE*, 40: 839-846.
- Lek, S., A. Belaud, P. Baran, I. Dimopoulos y M. Delacoste. 1996a. Role of some environmental variables in trout abundance models using neural networks. *Aquatic Living Research*, 9: 23-29.
- Lek, S., M. Delacoste, P. Baran, I. Dimopoulos, J. Lauga y S. Aulagnier. 1996b. Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling*, 90: 39-52.
- Lewis, W.M. y D.P. Morris. 1986. Toxicity of nitrite to fish. A review. *Transactions of the American Fisheries Society*, 115: 183-195.
- Liao, P.B. y R.D. Mayo. 1974. Intensified fish culture combining water reconditioning with pollution abatement. *Aquaculture*, 3: 61-85.
- Lin, S.H. y C.L. Wu. 1996. Electrochemical removal of nitrite and ammonia for aquaculture. *Water Research*, 30: 715-721.

- Lo, J.Y., J.A. Baker, P.J. Kornguth y C.E. Floyd. 1995. Application of artificial neural networks to interpretation of mammograms on the basis of the radiologists impression and optimized image features. *Radiology*, 197: 242-252.
- Losordo, T.M. y R.H. Piedrahita. 1991. Modeling temperature variation and thermal stratification in shallow aquaculture ponds. *Ecological Modelling*, 54: 189-226.
- Madenjian, C.P., G.L. Rogers y A.W. Fast. 1987. Predicting night time dissolved oxygen loss in prawn ponds of Hawaii: Part II. A new method. *Aquaculture Engineering*, 6: 209-225.
- Martín del Brío, B. y A. Sanz-Molina. 1997. *Redes neuronales y sistemas borrosos*. Edita: Ra-Ma. Madrid.
- Maule, A.G. y C.B. Schreck. 1990. Changes in numbers of leukocytes in immune organs of juvenile Coho salmon after acute stress or cortisol treatment. *Journal of Aquatic Animal and Health*, 2: 298-304.
- Mazeaud, M.M., F. Mazeaud y E.M. Donaldson. 1977. Primary and secondary effects of stress in fish: some new data with a general review. *Transactions of the American Fisheries Society*, 106(3): 201-212.
- McCulloch, W.S y W. Pitts. 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5: 115-137.
- Møllergaard, S. y I. Dalsgaard. 1989. Handbook of eel diseases. *Damns Fisk og Havunders Rapport*, 293: 1-47.
- Meyer, D.I. y D.E. Brune. 1982. Computer simulation of diurnal oxygen levels in a stillwater aquaculture pond. *Aquacultural Engineering*, 1: 245-261.
- Minai, A.A. y R.D. Williams. 1990. Acceleration of back-propagation through learning rate and momentum adaptation. *International Joint Conference of Neural Networks*, 1: 676-679.
- Minsky, M.L. 1954. *Neural nets and the brain-model problem*. Tesis Doctoral. Universidad de Princeton. EE.UU.
- Minsky, M.L. y S. Papert. 1969. *Perceptrons: an introduction to computational geometry*. Edita: MIT Press, Cambridge. Massachusetts.
- Moreu, P. 1999. *Estadística Informatizada*. Edita: Paraninfo. Madrid.

- Morimoto, T., T. Takeuchi, H. Miyata y Y. Hashimoto. 2000. Pattern recognition of fruit shape based on the concept of chaos and neural networks. *Computers and Electronics in Agriculture*, 26: 171-186.
- Murtaza, M.B. y D.J. Fisher. 1994. NEUROMODEX-Neural network system for modular construction decision making. *Journal of Computing in Civil Engineering*, 8: 221-233.
- Muttiah, R.S., R. Srinivasan y P.M. Allen. 1997. Prediction of two-year peak stream discharges using neural networks. *Journal of America Water Resources Association*, 33: 625-630.
- Nerini, D., J.P. Durbec y C. Manté. 2000. Analysis of oxygen rate time series in a strongly polluted lagoon using a regression tree method. *Ecological Modelling*, 133: 95-105.
- Nijhof, M. 1994a. Bacterial stratification and hydraulic loading effects in a plug-flow model for nitrifying trickling filters applied in recirculating fish culture system. *Aquaculture*, 134(1-2): 49-64.
- Nijhof, M. 1994b. Diffusional transport mechanisms and biofilm nitrification characteristics influencing nitrite levels in a nitrifying trickling filter effluents. *Water Resources*, 29(10): 2287-2292.
- Nilsson, N.J. 1965. *Learning machines: foundations of trainable pattern-classifying systems*. Edita: McGraw-Hill. New York.
- Noakes, D.J., A.I. MacLeod y K.W. Hipel. 1985. Forecasting monthly riverflow time series. *International Journal of Forecasting*, 1: 179-190.
- Norreys, R. e I. Cluckie. 1997. A novel approach to the real-time modelling of large urban drainage systems. *Water Science Technology*, 36: 19-24.
- Oliveira, M.E.C. y A.S. Franca. 1998. Simulation of oxygen mass transfer in aeration systems. *International Communication in Heat and Mass Transfer*, 25(6): 853-862.
- Otnes, R.K. y L. Enochson. 1978. *Applied time series analysis*. Edita: Wiley Interscience. New York.
- Palachek, R.M. y J.R. Tomaso. 1984. Toxicity of nitrite to channel catfish, tilapia and largemouth bass. *Canadian Journal of Fisheries and Aquatic Science*, 41: 1739-1744.
- Palackova, J., S. Gajdusek, J. Jirasek y K. Fasaic. 1990. Effect of sublethal concentration of ammonia in water on changes in and correlations of some biochemical indices in carp fry (*Cyprinus carpio* L.). *Ichthyologia*, 22(1): 57-67.

- Palazón, J. 1998. *Modelización de la intensidad de la lluvia: métodos estadísticos y uso de redes neuronales artificiales*. Trabajo Profesional Fin de Carrera. Universidad de Córdoba.
- Park, H.H. 1998. Analysis and prediction of walleye pollock (*Theragra chalcogramma*) landings in Korea by time series analysis. *Fisheries Research*, 38: 1-7.
- Parmar, R.S., R.W. McClendon, G. Hoogenboom, P.D. Blankenship, R.J. Cole y J.W. Dorner. 1997. Estimation of aflatoxin contamination in preharvest peanuts using neural networks. *Transactions of the ASAE*, 40: 809-813.
- Pesonen, E., M. Eskelinen y M. Juhola. 1996. Comparison of different neural network algorithms in the diagnosis of acute appendicitis. *International Journal of Bio-Medical Computing*, 40: 227-233.
- Pesonen, E., M. Eskelinen y M. Juhola. 1998a. Treatment of missing data values in a neural network based decision support system for acute abdominal pain. *Artificial Intelligence in Medicine*, 13: 139-146.
- Pesonen, E., C. Ohmann, M. Eskelinen y M. Juhola. 1998b. Diagnosis of acute appendicitis in two databases. Evaluation of different neighborhoods with an LVQ neural network. *Methods of Information in Medicine*, 37: 59-63.
- Piedrahita, R.H. y A. Seland. 1995. Calculation of pH in fresh and sea water aquaculture system. *Aquacultural Engineering*, 14: 331-346.
- Prybotuk, V.R., Y. Junsun y D. Mitchell. 2000. Comparison of neural models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. *European Journal of Operational Research*, 122: 31-40.
- Pulido-Calvo, I. 2001. *Diseño y gestión óptimos de sistemas de impulsión y de almacenamiento de agua para riego*. Tesis Doctoral. Universidad de Córdoba.
- Pulido-Calvo, I., J. Roldán, R. López-Luque y J.C. Gutiérrez-Estrada. 2001. Redes neuronales computacionales en la gestión de sistemas de distribución de agua para riego. *V Simposio sobre el Agua en Andalucía*, 1: 351-362.
- Pulido-Calvo, I., J. Roldán, R. López-Luque y J.C. Gutiérrez-Estrada. 2002. Técnicas de predicción a corto plazo de la demanda de agua. Aplicación al uso agrícola. *Ingeniería del Agua*, 9(3): 319-331.
- Pulido-San-Román, A. y J. Pérez-García. 2001. *Modelos econométricos*. Edita: Pirámide. Madrid.

- Raman, H. y N. Sunilkumar. 1995. Multi-variate modeling of water resources time series using artificial neural networks. *Hydrological Science*, 40: 145-163.
- Raman, H. y V. Chandramouli. 1996. Deriving a general operating policy for reservoirs using neural networks. *Journal of Water Resources Planning and Management*, 122: 342-347.
- Recknagel, F., T. Petzoldt, O. Jaeke y F. Krusche. 1997. Hybrid expert-system delacqua-A toolkit for water-quality control of lakes and reservoirs. *Ecological Modelling*, 71: 17-36.
- Romairé, R.P. y C.E. Boyd. 1979. Effects of solar radiation on the dynamic of dissolved oxygen in channel catfish ponds. *Transactions of the American Fisheries Society*, 108: 472-478.
- Romeril, M.G. y M.H. Davis. 1976. Trace metals in eels grown in power station cooling water. *Aquaculture*, 8: 139-150.
- Rosenblatt, F. 1957. *The perceptron: a perceiving and recognizing automaton*. Report 85-460-1, Project PARA, Cornell Aeronautical Laboratory, Ithaca, New York.
- Rosenblatt, F. 1960. *On the convergence of reinforcement procedures in simple perceptrons*. Report VG-1196-G-4, Cornell Aeronautical Laboratory, Ithaca, New York.
- Rosenblatt, F. 1962. *Principles of neurodynamics*. Edita: Spartan. Chicago.
- Roush, W.B. y T.L. Cravener. 1997. Artificial neural network prediction of amino acid levels in feed ingredients. *Poultry Science*, 76: 721-727.
- Rumelhart, D.E. y J.L. McClelland. 1986. *Parallel Distributed Processing*. Edita: MIT Press, Cambridge, Massachusetts.
- Russell, S. y P. Norvig. 1996. *El aprendizaje en las redes neuronales y de creencia. Inteligencia artificial: un enfoque moderno*. Edita: Prentice Hall Hispanoamericana, Mexico.
- Saad, M., A. Turgeon, P. Bigs y R. Duquete. 1994. Learning disaggregation technique for the operation of long-term hydro-electric power systems. *Water Resources Research*, 30: 3195-3202.
- Sánchez, L., V. Arroyo, J. García, K. Koev y J. Revilla. 1998. Use of neural networks in design of coastal sewage systems. *Journal of Hydraulic Engineering*, 124: 457-464.
- Sanni, S. y O.I. Forsberg. 1996. Modelling pH and carbon dioxide in single-pass seawater aquaculture systems. *Aquacultural Engineering*, 15: 91-110.

- Seginer, I., T. Buolard y B.J. Bailey. 1994. Neural network models of the greenhouse climate. *Journal of Agricultural Engineering Research*, 59: 203-216.
- Seginer, I., Y. Hwang, T. Boulard y J.W. Jones. 1996. Mimicking an expert greenhouse grower with a neural-net policy. *Transactions of the ASAE*, 39: 299-306.
- Shao, J., H. Xin y J.D. Harmon. 1997. Neural network analysis of postural behavior of young swine to determine the IR thermal comfort state. *Transactions of the ASAE*, 40: 755-760.
- Shepherd, J. y N. Bromage. 1999. *Piscicultura intensiva*. Edita: ACRIBIA, S.A. Zaragoza.
- Shin, H. y J.D. Salas. 2000. Regional drought analysis based on neural networks. *Journal of Hydrology Engineering*, 5: 145-155.
- Sigrimis, N.A., K.G. Arvanitis y R.S. Gates. 2000. A learning technique for a general purpose optimizer. *Computers and Electronics in Agriculture*, 26: 83-103.
- Soderberg, R.W. 1990. Temperature effects on the growth of blue tilapia in intensive aquaculture. *The progressive Fish-Culturist*, 52: 155-157.
- Speece, R.E. 1973. Trout metabolism characteristics and the rational design of nitrification facilities for water re-use in hatcheries. *Transactions of American Fisheries Society*, 102: 323-332.
- Stergiou, K.I., E.D. Christou y G. Petrakis. 1997. Modelling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods. *Fisheries Research*, 29: 55-95.
- Tamari, S., J.H.M. Wösten y J.C. Ruiz-Suárez. 1996. Testing an artificial neural network for predicting soil hydraulic conductivity. *Soil Science Society of America Journal*, 60: 1732-1741.
- Tarr, R.J. y B.J. Hill. 1979. Oxygen consumption, food assimilation and energy content of Southern African elvers (*Anguilla*). *Aquaculture*, 15: 141-150.
- Tawfik, M., A. Ibrahim y H. Fahmy. 1997. Hysteresis sensitive neural network for modeling rating curves. *Journal of Computing in Civil Engineering*, 11: 206-211.
- Theil, H. 1966. *Applied economic forecasting*. Edita: North-Holland. Amsterdam.
- Thirumalaiah, K. y M.C. Deo. 2000. Hydrological forecasting using neural networks. *Journal of Hydrologic Engineering*, 5: 180-189.
- Tokar, A.S. y M. Markus. 2000. Precipitation-runoff modeling using artificial neural networks and conceptual models. *Journal of Hydrologic Engineering*, 5: 156-161.



- Tokar, A.S. y P.A. Johnson. 1999. Rainfall-runoff modeling using artificial neural networks. *Journal of Hydrologic Engineering*, 4: 232-239.
- Uriel, E. 1995. *Análisis de datos. Series temporales y análisis multivariante*. Edita: AC. Madrid.
- Ventura, S. 1996. *Diseño de redes neuronales en métodos cinéticos*. Tesis Doctoral. Universidad de Córdoba.
- Ventura, S., M. Silva, D. Pérez-Bendito y C. Hervás. 1995. Artificial neural networks for estimation of kinetic analytical parameters. *Analytical Chemistry*, 67: 1521-1525.
- Ventura, S., M. Silva, D. Pérez-Bendito y C. Hervás. 1997a. Estimation of parameters of kinetic compartmental models by use of computational neural networks. *Journal of Chemical Informatic and Computer Science*, 37: 517-521.
- Ventura, S., M. Silva, D. Pérez-Bendito y C. Hervás. 1997b. Computational neural networks in conjunction with principal component analysis for resolving highly nonlinear kinetics. *Journal of Chemical Informatic and Computer Science*, 37: 287-291.
- Wedemeyer, G.A. y W.T. Yasutake. 1978. Prevention and treatment of nitrite toxicity in juvenile steelhead trout (*Salmo gairdneri*). *Journal of the Fisheries Research Board of Canada*, 35: 822-827.
- Wessel, M.D. y P. Jurs. 1994. Prediction of reduced ion mobility constants from structural information using multiple lineal regression analysis and computational neural networks. *Analytical Chemistry*, 66: 2480-2487.
- Werbos, P.J. 1974. *Beyond regression: New tools for prediction and analysis in the behavioural sciences*. Tesis Doctoral. Universidad de Harvard.
- Widrow, B. 1962. Generalization and information storage in networks of adaline 'neurons'. En: *Self-organizing systems*. Yovits, M.C., G.T. Jacobi y G.D. Goldstein. Edita: Spartan. Chicago.
- Widrow, B. y M.E. Hoff. 1960. Adaptive switching circuits. *1960 IRE WESCON Convention Record*, New York, 96-104.
- Wienbeck, H. 1981. On the oxygen balance of an experimental eel farm operated in thermal effluents of a conventional power station. En *Proceedings of the World Symposium on Aquaculture 'Heated Effluents and Recirculation System'*. Berlín (Alemania).
- Wootton, R.J. 1991. *Ecology of teleost fishes*. Edita: Chapman and Hall. Londres.

- Wu, X., J. Ghaboussi y J.H. Garret. 1992. Use of neural networks in detection of structural damage. *Computers and Structures*, 42: 649-659.
- Yang, C.C., S.O. Prasher, G.R. Mehuys y N.K. Patni. 1997a. Application of artificial neural networks for simulating of soil temperature. *Transactions of the ASAE*, 40: 649-656.
- Yang, C.C., S.O. Prasher, R. Lacroix, S. Sreekanth, N.K. Patni y L. Masse. 1997b. Artificial neural network model for subsurface-drained farmlands. *Journal of Irrigation and Drainage Engineering*, 123(4): 285-292.
- Yang, C.C., S.O. Prasher, S. Sreekanth, N.K. Patni y L. Masse. 1997c. An artificial neural network model for simulating pesticide concentrations in soil. *Transactions of the ASAE*, 40: 1285-1294.
- Zhang, M., J. Fulcher y R.A. Scofield. 1997. Rainfall estimation using artificial neural network group. *Neurocomputing*, 16: 97-115.
- Zhang, S.P., H. Watanabe y R. Yamada. 1994. Prediction of daily water demands by neural networks. En: *Stochastic and statistical method in hydrology and environmental engineering*. Edita: Kluwer. Dordrecht.
- Zou, Y., Y. Shen, L. Shu, Y. Wang, F. Feng, K. Xu, Y. Qu, Y. Song, Y. Zhong, M. Wang y W. Liu. 1996. Artificial neural network to assist psychiatric diagnosis. *British Journal of Psychiatry*, 169: 64-67.
- Zupan, J. y J. Gasteiger. 1991. Neural networks: a new method for solving chemical problems or just a passing phase?. *Analytica Chimica Acta*, 248: 1-30.



## 4.6. ANEXO IV.1. GLOSARIO

$\alpha_d$	Factor de alisado
$\alpha(t)$	Tasa de aprendizaje
$\beta$	Coefficiente de regresión poblacional
$\beta_d$	Factor de alisado en el método de Holt
$\gamma_\alpha$	Factor exponencial para el incremento de la tasa de aprendizaje
$\gamma_\mu$	Factor exponencial para el incremento del momento
$\delta_j$	Error en las neuronas de salida
$\Phi$	Parámetro estimado en los modelos <b>ARIMA</b>
$\varphi_\alpha$	Factor de decremento de la tasa de aprendizaje
$\varphi_\mu$	Factor de decremento del momento
$\kappa_\alpha$	Factor de escalado para la tasa de aprendizaje
$\kappa_\mu$	Factor de escalado para el momento
$\mu(t)$	Término de momento
$\nu$	Frecuencia
$\Theta$	Parámetro estimado en los modelos <b>ARIMA</b>
$\theta$	Factor de ponderación
$A_t$	Diferencia en valor absoluto entre el valor del parámetro observado y predicho en el instante <b>t</b>
<b>B</b>	Operador de retroceso en los modelos <b>ARIMA</b>
<b>b</b>	Coefficiente de regresión muestral
<b>D</b>	Factor de estacionalidad en los modelos <b>ARIMA</b>
<b>d</b> ó <b>d<sub>t</sub></b>	Variable temporal
<b>d<sub>k</sub></b>	Salida deseada de la neurona <b>k</b>
<b>d<sub>kg</sub></b>	Salida deseada <b>k</b> del patrón <b>g</b>

E	Coeficiente de eficiencia
e	Variable dependiente o criterio
$e_{kg}$	Salida estimada <b>k</b> por el modelo para el patrón <b>g</b>
F(k)	Transformada de Fourier de <b>f(t)</b>
$F_t$	Valor filtrado de la variable original en el momento <b>t</b>
$f(I_k)$	Salida de la neurona <b>k</b>
$f'(I_k)$	Derivada de la función de transferencia de la neurona <b>k</b>
$f(t)$	Serie temporal discreta
$I_j$	Suma ponderada en cada una de las capas intermedias y de salida de una red neuronal
N	Número de datos en la muestra
P	Orden autoregresivo en los modelos <b>ARIMA</b>
Q	Orden de media móvil en los modelos <b>ARIMA</b>
$R^2$	Coeficiente de determinación
$R_u^2$	Coeficiente de determinación umbral
S	Medida de asociación entre los valores observados y estimados
$S_{obs}$	Variabilidad de los valores observados frente a la media
$S_t$	Variable <b>d</b> alisada
$T_t$	Segunda variable de alisamiento en el método de Holt
U	Coeficiente de desigualdad de Theil
$U_c$	Componente de correlación del coeficiente de desigualdad de Theil
$U_d$	Componente de dispersión del coeficiente de desigualdad de Theil
$U_s$	Componente sistemático del coeficiente de desigualdad de Theil
$V_b$	Variable a ser estimada
$V_b^*$	Valor escalado de la variable <b>V<sub>b</sub></b>

$V_{\max,b}$	Máximo de la variable $V_b$
$V_{\min,b}$	Mínimo de la variable $V_b$
$W_{ji}$	Peso que conecta la neurona $i$ con la neurona $j$
$x$	Variable independiente
$x_{ji}$	Conjunto de entrada en la neurona $j$
$y^*$	Tasa de cambio en la serie observada
$\dot{y}^*$	Tasa de cambio en la serie estimada
$y_{pj}$	Respuesta de la neurona $j$ tras la aplicación de la función de transferencia



## 4.7. ANEXO IV.2. MANUAL DEL USUARIO DEL SIMULADOR DE REDES NEURONALES COMPUTACIONALES (REDGEN 1.0)

La aplicación de ordenador desarrollada tiene como objetivo la simulación de Redes Neuronales Computacionales (RNCs). En función del tipo de aprendizaje y de la arquitectura, los modelos de redes neuronales que se pueden considerar en esta aplicación son: supervisados unidireccionales (retropropagación estándar, **DBD**, **EDBD** y **LVQ**), híbridos (**RBF**), no supervisados realimentados (**ART 1**, **ART 2** y Hopfield) y supervisados bidireccionales (**EDBDM**).

En el aprendizaje supervisado se presenta a la red un conjunto de patrones (entrada-salida objetivo) e iterativamente ésta ajusta sus pesos hasta que su salida tienda a ser la deseada, mientras que en el no supervisado se presentan a la red los datos de entrada sin adjuntar la respuesta que deseamos. En el aprendizaje híbrido coexisten en la red los dos tipos de aprendizaje anteriores, normalmente en distintas capas de neuronas. Asimismo, atendiendo al flujo de datos en la red neuronal, podemos hablar de redes unidireccionales, donde la información circula en un único sentido desde las neuronas de entrada hacia las de salida, y de redes realimentadas en las que la información puede circular en cualquier sentido.

En este caso se describe con amplitud el manejo del módulo de la aplicación correspondiente a la red neuronal supervisada unidireccional **EDBD**. Para iniciar el programa basta con hacer doble 'click' en el icono de dicha aplicación que se encuentra en el escritorio de Windows® (figura AIV.1). De este modo aparece la ventana principal, que está compuesta de distintas opciones en la barra de herramientas y en el entorno de la ventana (figura AIV.2). A continuación se describe cada una estas opciones.



**Figura AIV.1.** Inicio de la aplicación de ordenador REDGEN 1.0 desde el acceso directo en el escritorio de Windows®

### IV.2.1. VENTANA PRINCIPAL

La ventana principal de **REDGEN 1.0** se encuentra físicamente dividida en tres partes: a) una barra de herramientas con seis opciones *Archivo*, *Red*, *Arquitectura de la red*, *Aprendizaje*, *Generalización* y *Ayuda*, que sirve principalmente para introducir los datos de entrenamiento y de generalización, así como para seleccionar la arquitectura y el tipo de aprendizaje de la **RNC**; b) dos bloques en la parte izquierda del entorno de la ventana, *Arquitectura* y *Parámetros de aprendizaje*, donde se muestran las características fundamentales de la arquitectura y del tipo de aprendizaje de la **RNC** elegida; c) los bloques *Entrenamiento* y *Parámetros estadísticos* (figura AIV.2).



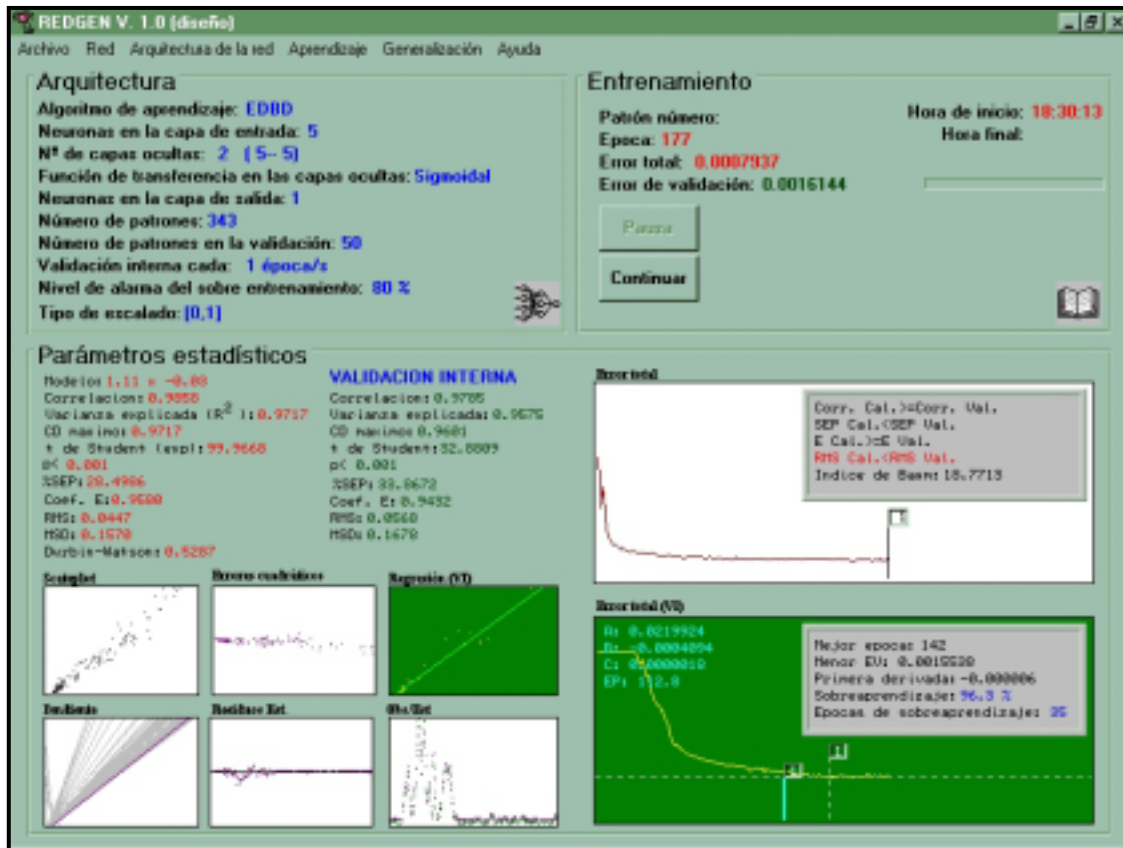


Figura AIV2. Ventana principal de la aplicación de ordenador REDGEN 1.0

#### IV.2.1.1. Opción de barra Archivo

La opción de barra *Archivo* presenta un menú desplegable con seis opciones: *Nuevo*, *Entrenamiento*, *Guardar resultados como*, *Guardar pesos como*, *Imprimir* y *Salir* (figura AIV.3). La opción *Nuevo* permite la introducción de nuevos datos para la simulación de todas las redes neuronales como se desee sin necesidad de iniciar la aplicación.

La opción *Entrenamiento* es el soporte principal para la lectura de los datos relativos al conjunto de entrenamiento (entradas-salidas objetivo). De este modo, al marcarla aparecen dos alternativas: *Calibración* y *Validación interna*. En la primera de ellas podemos seleccionar: *Cargar entradas de entrenamiento* y *Cargar salidas de entrenamiento*, las cuales requieren los archivos con los valores de las variables de entrada y de salida de la **RNC**, respectivamente. Estos ficheros pueden ser de tipo texto o Excel<sup>®</sup>, y han de presentar los valores de cada variable en columnas. Al hacer ‘click’ en la segunda opción podremos elegir entre cuatro nuevas opciones: a) requiere tanto las entradas como las salidas en un archivo externo para la validación interna (*Cargar entradas validación* y *Cargar salidas validación*), b) *Seleccionar un grupo fijo aleatorio*, que indica al programa que del grupo de entrenamiento seleccione un subgrupo aleatorio a utilizar en la validación interna, c)

*Validación cruzada tipo ISI* opción mediante la cual el grupo de validación interna queda constituido por los primeros elementos del grupo de calibración y d) *Validación cruzada completa (SAS)*, opción que permite realizar validación cruzada de la misma forma que la lleva a cabo el paquete estadístico SAS, en la que todos los patrones de entrenamiento pasan a ser en algún momento parte del subgrupo de validación.



**Figura AIV.3.** Menú desplegable de la opción de barra Archivo

La opción *Guardar resultados como*, tal y como su nombre indica, pide el nombre del fichero donde quedan almacenados los resultados obtenidos en el entrenamiento de la **RNC**. La extensión del archivo es '.res' y presenta en columnas los valores observados frente a los estimados por la **RNC**, así como todos los parámetros estadísticos calculados durante la calibración.

La opción *Guardar pesos como* pide el nombre del archivo donde se guardan los valores de los pesos o conexiones de cada una de las neuronas de la red. Su extensión es '.pes' y tiene un número de columnas igual al número de capas de la **RNC** menos 1. De este modo, si tenemos una red neuronal con una capa de entrada, dos capas ocultas y una capa de salida, la primera columna del archivo presenta las conexiones de las neuronas de la capa de entrada con las neuronas de la primera capa oculta, la segunda columna los pesos de las conexiones de las neuronas de las dos capas intermedias, y la tercera columna los pesos de las conexiones de los nodos de la segunda capa oculta con la capa de salida.

La opción *Imprimir* tiene a su vez dos alternativas: *Configuración de impresora y Datos*. La primera permite seleccionar la impresora que se va a utilizar así como la configuración de ésta: orientación, tamaño y origen del papel, y tipo y calidad de gráficos. Asimismo se puede indicar el intervalo de impresión (todo, selección, páginas), la calidad de la impresión y el número de copias a realizar. La segunda alternativa solicita la dirección del fichero de datos o resultados que se quiere imprimir. Por último, la opción *Salir* implica la terminación de la aplicación.

#### IV.2.1.2. Opción de barra *Red*

El menú desplegable que resulta al marcar esta opción permite seleccionar entre redes neuronales supervisadas, no supervisadas e híbridas. Del primer tipo podemos seleccionar como algoritmos de aprendizaje: retropropagación estándar (Back-propagation), **DBD** (Delta-Bar-Delta), **EDBD** (Extended Delta-Bar-Delta), **LVQ** (Learning Vector Quantization) y **EDBDM** (Extended Delta-Bar-Delta Modificado). De la segunda clase, los algoritmos de aprendizaje disponibles son: **ART 1**, **ART 2** (Adaptative Resonance Theory) y Hopfield, y del tercer tipo el algoritmo **RBF** (Radial Basis Function).

#### IV.2.1.3. Opción de barra *Arquitectura de la red*

Esta opción es la que permite definir el tipo de estructura de la red neuronal. *Nodos* es la primera alternativa que aparece en el menú que se despliega al marcar esta opción, y asimismo, cuenta con tres posibilidades: *Entradas*, *Capas ocultas* y *Salidas*. Al seleccionar *Entradas* aparece una ventana en la que hay que indicar el número de neuronas que va a tener la capa de entrada. Asimismo, al marcar *Capas ocultas* y *Salidas* aparecen nuevas ventanas donde se ha de señalar el número de capas ocultas y de neuronas en cada una de estas capas (figura AIV.4), y el número de neuronas de la capa de salida, respectivamente. Una vez introducida la arquitectura deseada, aparece una nueva ventana donde hay que escribir el nombre del archivo donde se guarda esta estructura. La extensión de este fichero es '.top'.

La segunda alternativa de esta opción de la barra de herramientas se denomina *Sesgos* y consiste en añadir una neurona adicional a la **RNC** que se encuentra conectada a todas las demás neuronas de una capa de la red y que tiene permanentemente como valor de salida la unidad. Cada una de las neuronas a las que se conecta el sesgo consiguen que su entrada total pertenezca al dominio donde puede ser usada por la función de transferencia con un mejor rendimiento total de la red. Al seleccionar esta opción aparece una ventana donde se ha de indicar las capas de la red neuronal afectadas por una neurona de sesgo.



**Figura AIV.4.** Ventanas donde se indica el número de capas ocultas y el número de neuronas en estas capas de la RNC

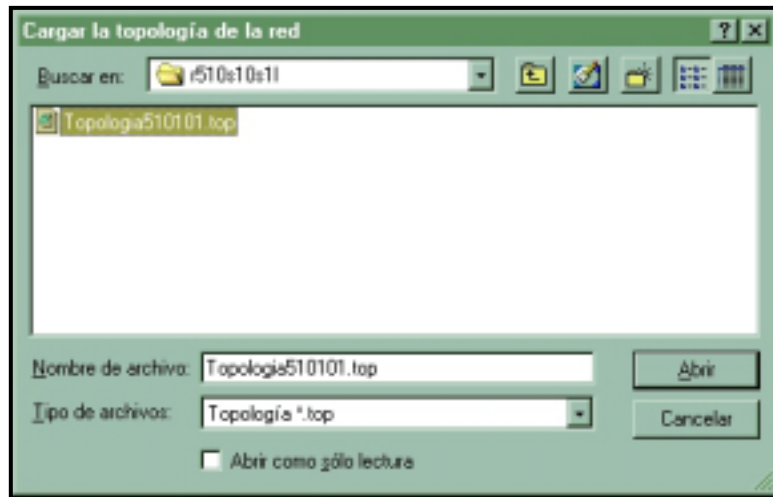
#### IV.2.1.4. Opción de barra *Aprendizaje*

En función del tipo de aprendizaje seleccionado se tendrán operativas unas u otras alternativas del menú desplegable que aparece al seleccionar la opción *Aprendizaje*. Las alternativas operativas se subdividen en cuatro bloques: los dos primeros contienen parámetros necesarios en el aprendizaje de la **RNC**, que en el caso de elegir el algoritmo **EDBD** son: factor de convergencia o ponderación (*Factor de convergencia*), valores máximos de la tasa de aprendizaje (*Tasa de aprendizaje máxima*) y del término de momento (*Momento máximo*), factores de escalado y de decremento para los coeficientes de aprendizaje y momento (*Factor de escalado de alfa*, *Factor de escalado del momento*, *Factor de decremento de alfa*, *Factor de decremento del momento*) y factores exponenciales para el incremento de estos coeficientes (*Factor exponencial de alfa*, *Factor exponencial del momento*); en el tercero de ellos podemos seleccionar el *Factor de convergencia (Teta)*, el valor máximo aceptable del término de error de cada patrón (*Término de error*), *Opciones* y *Coefficiente de determinación umbral* (habrá que seleccionar uno de ellos). En el último bloque encontramos la opción para el inicio del proceso de entrenamiento: *Comenzar entrenamiento*. En la sección *Opciones* del tercer bloque encontramos: *Número de términos en la validación interna* (donde podremos seleccionar: *Número fijo*, *Porcentaje* y *Épocas para validación interna*) y *Nivel de alarma del sobreentrenamiento*.

#### IV.2.1.5. Opción de barra *Generalización*

Esta opción permite cargar los resultados obtenidos en el entrenamiento para proceder a la validación del modelo con los datos del conjunto de generalización correspondiente. De este modo, las ventanas que se muestran con esta alternativa son: *Cargar topología de la red*, *Cargar datos de aplicación (entradas)*, *Cargar datos de aplicación (salidas)*, *Cargar fichero de pesos* y *Guardar resultados*.

En la alternativa *Cargar topología de la red* se abre el fichero con la extensión '.top' en el que se recoge la arquitectura de la red neuronal (número de capas y de neuronas en cada capa) previamente guardada en la opción de barra *Arquitectura de la red* (figura AIV.5). Las opciones *Cargar datos de aplicación (entradas)* y *Cargar datos de aplicación (salidas)* son el soporte principal para la lectura de los datos relativos al conjunto de generalización (entradas-salidas objetivo). Al marcarlas aparecen las ventanas que requieren los archivos con los valores de las variables de entrada y de salida para la validación de la **RNC**, respectivamente. Estos ficheros pueden ser de tipo texto o Excel<sup>®</sup>, y han de presentar los valores de cada variable en columnas. La opción *Cargar fichero de pesos* pide el nombre del archivo donde se han guardado los valores de las conexiones de cada una de las neuronas de la red obtenidos en el entrenamiento, cuyo nombre se indicó en la opción de barra *Archivo*. La opción *Guardar resultados*, tal y como su nombre indica, pide el nombre del fichero donde quedan almacenados los resultados obtenidos en la generalización de la **RNC**.



**Figura AIV.5.** Ventana donde se indica el nombre del archivo que contiene la estructura del conexionado de las neuronas de la red

#### IV.2.1.6. Opción de barra *Ayuda*

La opción *Contenido* de este menú permite que el usuario obtenga información respecto a las bases teóricas de la metodología desarrollada en esta aplicación de ordenador y respecto a su manejo. La búsqueda de la información puede obtenerse de una ficha que describe el contenido general o de una ficha índice de palabras clave. La opción *Acerca de...* muestra una caja de diálogo con la información general del programa: nombre, autor, organismo donde se ha desarrollado la aplicación, versión y año (figura AIV.6).

#### IV.2.1.7. Bloques *Arquitectura* y *Parámetros de aprendizaje*

En estos dos bloques, situados en la parte izquierda del entorno de la ventana principal, se muestran las características fundamentales de la arquitectura y del tipo de aprendizaje de la **RNC**. En el bloque *Arquitectura* se visualiza el tipo de algoritmo de aprendizaje seleccionado, el número de capas y de neuronas en cada capa (detrás del número de capas ocultas se muestra entre paréntesis el número de neuronas de estas capas), el número de patrones (entradas-salidas objetivo) utilizados en el entrenamiento de la red neuronal, el tipo de función de transferencia utilizado, el número de patrones utilizados en la validación interna, el intervalo de épocas para una validación interna, el nivel de alarma del sobreentrenamiento, y el tipo de escalado de los datos. Toda esta información ha sido introducida con anterioridad en las opciones de barra *Archivo*, *Red* y *Arquitectura de la red* (figura AIV.2).

En el bloque *Parámetros de aprendizaje* se visualizan los valores de los parámetros necesarios para el aprendizaje (según el tipo de algoritmo elegido en la opción de barra *Red*).



**Figura AIV.6.** Caja de diálogo con la información general del programa REDGEN 1.0

#### IV.2.1.8. Bloques *Entrenamiento* y *Parámetros estadísticos*

El bloque *Entrenamiento* se encuentra en la parte derecha del entorno de la ventana principal y permite controlar algunos parámetros del entrenamiento tales como el error de calibración y validación interna, el patrón introducido y la época. Asimismo, cuenta con dos botones (*Pausa* y *Continuar*) que permiten parar o continuar un entrenamiento a voluntad (figura AIV.2). También hay un indicador de la hora de comienzo y final del entrenamiento.

En el bloque *Parámetros estadísticos* se ofrece al usuario información que le permite seguir en tiempo real el proceso de entrenamiento. El bloque ocupa la mitad inferior de la ventana principal y aparece una vez que comienza el entrenamiento. En el cuadrante superior izquierdo se proporcionan varios parámetros tanto de la calibración como de la validación interna: el modelo de regresión simple entre los valores observados y estimados por la red una vez finalizada una época, la correlación, la varianza explicada, el máximo coeficiente de determinación alcanzado por el modelo, la *t* de Student y su nivel de significación, el porcentaje de error estándar de predicción (**SEP**), el coeficiente de eficiencia (**E**), el error cuadrático medio (**RMS**), la media del grupo de calibración (**MSD**) y el estadístico de Durbin-Watson. En la parte derecha de este primer cuadrante encontramos los mismos términos (excepto la ecuación del modelo y el estadístico de Durbin-Watson) pero para el subgrupo de validación interna.

En el cuadrante inferior izquierdo nos encontramos con 6 pequeños gráficos. De izquierda a derecha y de arriba hacia abajo muestran: una representación punto a punto (*Scatterplot*) de los valores observados frente a los estimados para el grupo de calibración; una representación punto a punto de los errores cuadráticos; una representación punto a punto y la recta de regresión entre valores observados y estimados en el grupo de validación interna; un gráfico de pendientes correspondiente a la recta de regresión entre los valores observados y estimados en el grupo de calibración; una representación punto a punto de los residuos del grupo de calibración (observados menos estimados), y finalmente una representación punto a punto de los valores observados y estimados frente al número de orden de la serie original.

El cuadrante superior derecho está ocupado por un gráfico de mayor tamaño donde se representa el error total de salida de la red neuronal para el grupo de calibración. El último punto dibujado se acompaña de una recta vertical a la cual se une un pequeño cuadro en cuyo interior se indica el número de ciclos transcurridos (1 ciclo=260 épocas). En el interior de la ventana del gráfico se encuentra un cuadro que nos indica si el entrenamiento se está llevado a cabo de forma correcta mediante la comparación de la correlación, el error estándar de predicción, el coeficiente de eficiencia y el error cuadrático medio del grupo de calibración con el de validación. Asimismo se indica el valor del índice de Baum.

El cuadrante inferior derecho lo ocupa el gráfico del error total de validación interna. De la misma forma que en el caso anterior, el gráfico dibujado se acompaña de una recta vertical punteada en cuyo extremo hay un pequeño cuadro sobre el que se indica el ciclo en el que se ha conseguido el menor error de validación. En la esquina superior izquierda de la ventana del gráfico se encuentran cuatro números precedidos por las letras **A**, **B**, **C** y **EP**. Las tres primeras indican los coeficientes de la regresión polinómica de segundo orden entre el número de épocas y el error total en validación. **EP** indica el mínimo de la función calculada mediante mínimos cuadrados, de tal forma que se puede estimar el final del entrenamiento cuando la recta vertical punteada y la recta vertical de color azul claro coinciden. También dentro de la ventana del gráfico encontramos otro cuadro donde se indica la época en la que se ha obtenido el menor error de validación, su valor, el gradiente de cambio entre los tres últimos puntos dibujados (primera derivada o primera diferenciación), el nivel de sobreaprendizaje calculado como una suma ponderada entre las veces que la función crece y las veces que la función decrece, y el número de épocas transcurridas desde la última mejora en la función de error.

#### **IV.2.2. REQUERIMIENTOS MÍNIMOS DE HARDWARE E INSTALACIÓN**

Para el funcionamiento adecuado del programa se requiere un ordenador con un procesador Pentium® 75 o superior, con un mínimo de 32 Megabytes de memoria RAM. El sistema operativo Windows 95® o superior, y un mínimo de 1 Megabyte de espacio libre en el disco duro.

El programa está codificado en lenguaje MS Visual Basic® versión 4.0 y Borland C++® versión 4.0. El procedimiento de instalación es el típico de los programas diseñados para este

sistema operativo y se hace a partir de 1 disco de instalación. La desinstalación debe hacerse desde el icono del panel de control.





## CAPÍTULO 5

---

### CARACTERIZACIÓN Y PREDICCIÓN DE PARÁMETROS BIOLÓGICOS. ESTRATEGIA DE PRODUCCIÓN

#### 5.1. Introducción

5.1.1. Modelación del crecimiento y las estrategias de producción

5.1.2. Crecimiento de la anguila

#### 5.2. Material y métodos

5.2.1. Caracterización de la biomasa y mortalidad real

5.2.2. Estimación de la desviación típica o estándar. Derivación de los modelos de varianza

5.2.3. La tasa de crecimiento

5.2.4. La tasa de mortalidad

5.2.5. Modelación de la cosecha

5.2.6. Proceso de simulación dinámica. Valoración económica inicial

#### 5.3. Resultados

5.3.1. Error de predicción del gestor

5.3.2. Análisis de las muestras del tanque 'A1'. Relación longitud-peso

5.3.3. Estimación de los rangos de variación de la desviación típica. Aplicación de los modelos de varianza

5.3.4. Caracterización de la biomasa

5.3.5. Caracterización de la mortalidad

5.3.6. Análisis de los resultados obtenidos con el Programa Operativo de Simulación (**POS**)

5.3.6.1. Simulaciones de referencia

5.3.6.2. Simulación de referencia 1

5.3.6.3. Simulación de referencia 2

5.3.6.4. Simulación 3. Incorporación de una nueva cohorte

5.3.6.5. Simulación 4. Tres cohortes, tres años e incorporación del factor de crecimiento debido a la temperatura

5.3.6.6. Simulación 5. Incorporación del factor de crecimiento debido a la concentración de amoníaco

5.3.6.7. Simulación 6. Incorporación del factor de crecimiento debido a la concentración de nitritos

5.3.6.8. Simulación 7. Incorporación del factor de crecimiento debido a la concentración de nitratos

5.3.6.9. Simulación 8. Incorporación del factor de crecimiento debido al pH

5.3.6.10. Simulación 9. Factores de crecimiento debidos a la temperatura y amoníaco. Influencia de los porcentajes iniciales en cada subgrupo de crecimiento

5.3.6.11. Simulación 10. Influencia del lote de cosecha

- 5.3.6.12. Simulación 11. Influencia del cambio del lote de cosecha durante la simulación
- 5.3.6.13. Simulación 12. Influencia del tamaño de cosecha
- 5.3.6.14. Simulación 13. Aplicación de los modelos de predicción de las variables temperatura y amoniaco
- 5.3.6.15. Simulación 14. Conexión con **SEDPA** 1.0 a través de las reglas **A-A** de **CONTROLPLANT** 1.0

#### 5.4. Discusión

- 5.4.1. Variación de la biomasa y mortalidad
- 5.4.2. Comportamiento y robustez del modelo
- 5.4.3. Inclusión de factores de crecimiento debidos a parámetros físico-químicos
- 5.4.4. Conexión con **SEDPA** y reglas **A-A** de **CONTROLPLANT**

#### 5.5. Referencias

#### 5.6. ANEXO V.1. Glosario

#### 5.7. ANEXO V.2. Manual del usuario del Programa Operativo de Simulación (**POS**)

## CAPÍTULO 5. CARACTERIZACIÓN Y PREDICCIÓN DE PARÁMETROS BIOLÓGICOS. ESTRATEGIA DE PRODUCCIÓN

### 5.1. INTRODUCCIÓN

Uno de los principales problemas con los que a medio plazo se enfrenta el gestor de cualquier piscifactoría es la definición de la estrategia de producción de la planta. Se entiende como tal, el manejo y las acciones encaminadas a alcanzar la cosecha y tasas de crecimiento y mortalidad para una capacidad de producción deseada. En gran parte de las ocasiones, la estrategia de producción óptima vendrá impuesta por las características constructivas de las instalaciones o como indica Colt (1991) estará en función de criterios de calidad del agua. Sin embargo, a pesar de las limitaciones propias del sistema, la estrategia de producción estará sujeta a variaciones marcadas por factores internos como el *stock* (cantidad de biomasa almacenada en el sistema) y por factores externos como las condiciones ambientales o el estado del mercado. En base a estos factores, el gestor tendrá que decidir el número de ejemplares en el/los lote/s de reclutamiento y venta así como el tipo de cosecha.

Básicamente, el sistema de producción de cualquier piscifactoría puede englobarse en alguna de las tres estrategias siguientes: cultivo de lotes (**BC**, *Batch Culture*), almacenamiento y cosecha de lotes coexistentes (**CBSH**, *Current Batch Stocking and Harvesting*), y almacenamiento de lotes mezclados y cosecha por clasificación (**CMSGH**, *Concurrent Mixed-Stocking and Graded Harvesting*) (Summerfelt *et al.*, 1993).

El cultivo de lotes (**BC**) es la estrategia más simple dentro del cultivo de peces. Una piscifactoría que basase su producción siguiendo esta estrategia operaría de la siguiente forma: a) almacenaría un grupo de peces, b) los engordaría hasta que alcanzasen la talla de mercado, c) cosecharía el lote completo, y d) repetiría el proceso con el siguiente lote de peces. La estrategia **BC** cuenta como ventaja frente al resto de estrategias de cultivo la facilidad en la predicción de la producción final, ya que ésta puede ser llevada a cabo mediante la aplicación de modelos simples de crecimiento en los que se tienen en cuenta la temperatura, la alimentación y el tamaño de los individuos (Bjorndal, 1988; Wootton, 1991). Sin embargo, desde un punto de vista económico es la menos eficiente ya que infrautiliza las instalaciones al existir una única cosecha por período de crecimiento y presenta la menor relación producción/tiempo.

El almacenamiento y cosecha de lotes coexistentes (**CBSH**) es una estrategia de producción más avanzada y compleja que la anterior. De esta forma, el procedimiento de operación de una planta con una estrategia de tipo **CBSH** sería: a) almacenamiento en el sistema de grupos individuales de peces a determinados intervalos de tiempo, b) engorde de los grupos individuales de forma aislada de otros grupos hasta que la longitud o peso medio de los peces alcanza el tamaño comercial, y c) cosecha completa de los grupos individuales. Así, se puede decir que la estrategia **CBSH** es la repetición en paralelo de la **BC** a distintos intervalos de tiempo. Este sistema de producción ha

sido estudiado por varios autores (Paessun y Allison, 1984; Watten, 1992), coincidiendo todos en destacar que la principal ventaja radica en la operación continua y en la *cuasi* producción continua. A pesar de ello, el sistema **CBSH** arrastra algunos de los inconvenientes de la estrategia **BC** al incrementarse fácilmente los costes de inversión, el tiempo, el seguimiento y las labores de mantenimiento de los grupos de peces aislados en función de la producción final requerida.

Finalmente, el almacenamiento de lotes mezclados y cosecha por clasificación (**CMSGH**) es la estrategia de producción más avanzada y compleja de las tres descritas. El procedimiento de operación con este sistema de producción sería: a) almacenamiento de peces en el sistema a distintos intervalos de tiempo, b) engorde de los peces, y c) cosecha selectiva de todos los individuos de talla comercial a intervalos de tiempo determinados e independientemente de su localización en el sistema. Sobre este planteamiento básico se pueden llevar a cabo ligeras modificaciones que proporcionan elasticidad y complejidad al sistema. Así, por ejemplo, se puede presentar la posibilidad de realizar cosechas selectivas en función de lotes de venta predeterminados o que tanto la cosecha como los lotes de ventas queden establecidos por requerimientos de un mercado a la demanda. Este sistema presenta importantes ventajas frente a los descritos anteriormente, ya que permite una operación continua que facilita el alcanzar la capacidad plena de la planta y, al igual que en la estrategia **CBSH**, la producción es *cuasi* continua, aunque en este caso, los lotes de cosecha están compuestos por individuos de la misma talla. Sin embargo, una mayor complejidad en el sistema de operación conlleva la presencia de importantes inconvenientes. De esta forma, una clasificación continua de los individuos presentes en la planta incrementa significativamente el tiempo destinado a labores de clasificación y aumenta el riesgo de la introducción de enfermedades y de la aparición de estrés.

### **5.1.1. Modelación del crecimiento y las estrategias de producción**

Debido a implicaciones económicas, la modelación del crecimiento de especies piscícolas destinadas a consumo ha suscitado un gran interés a partir de principios de los años 80 (From y Rasmussen, 1984; Hewett y Johnson, 1987; Olsen y Balchen, 1992; Cho y Bureau, 1998). La aplicación de estos modelos se ha llevado a cabo con gran éxito en multitud de instalaciones y se han implementado como herramientas muy útiles que proporcionan básicamente: a) predicciones sobre factores secundarios determinantes en el crecimiento como la densidad o la calidad del agua, o b) reducciones de la cantidad de información, lo que permite comparaciones entre crecimientos bajo diferentes grupos de condiciones (temperatura, tipo de alimento, tasa de alimentación, estirpes genéticas o especies, densidad, etc.) (Arnason *et al.*, 1992).

Existen dos formas básicas de modelos de crecimiento: a) modelos que asumen que el crecimiento es una función dependiente del tamaño alcanzado (por ejemplo: modelo Logístico, Gompertz, Monomolecular o de Richards), y b) modelos que basan su formulación en principios bioenergéticos (por ejemplo: modelo de Von Bertalanffy) (Wootton, 1991). El primer tipo es más adecuado cuando los datos disponibles son

muestrales y lo que se persigue es conocer la biomasa de una cohorte a un tamaño determinado (Van Nes *et al.*, 2002). El segundo tipo es utilizado cuando se dispone de medidas biométricas individuales (básicamente longitud, peso y edad) y tradicionalmente han sido aplicados en la descripción del crecimiento de poblaciones naturales (Crivelli y Mestre, 1988; Herrera y Fernández-Delgado, 1992; Martínez-Garmendia, 1998; Gutiérrez-Estrada *et al.*, 2000).

De los tres tipos de estrategias de producción expuestas anteriormente, sin duda la que supone una modelación más compleja es la **CMSGH**. Esto se debe fundamentalmente a que la cosecha se compone de individuos de aproximadamente la misma talla obtenidos mediante clasificación, lo que obliga no sólo a definir y modelar el crecimiento de las distintas cohortes, sino que requiere la incorporación al modelo de ecuaciones que describan la variación del tamaño de los peces en una cohorte determinada en función del tiempo. Generalmente, cuando los modelos han sido utilizados con la finalidad de reducir información, la variación sobre el modelo es sólo usada como una estimación de la variación del error, lo que facilita la comparación entre distintos grupos de condiciones. Sin embargo, es difícil relacionar la variación del error con la variación del tamaño de los peces, ya que la mayoría de los datos usados para la generación de los modelos de crecimiento han sido tradicionalmente lotes de pesos medios y/o transformaciones de esas medias. Por otra parte, no es normal obtener datos de varianza muestral debido a que esto requiere medidas de pesos individuales para cada lote.

La incorporación de ecuaciones que describen la variación del tamaño de los peces proporciona al modelo la posibilidad de predecir la distribución por tallas (Sparre, 1976; Arnason y Scuse, 1984) así como el análisis de estrategias de manejo realistas que incluyen la necesidad y el efecto de las clasificaciones. De esta forma, un modelo determinista de crecimiento puede pronosticar cuando un lote de peces alcanzará un tamaño medio objetivo y que proporción puede ser cosechada (Arnason *et al.*, 1992).

### **5.1.2. Crecimiento de la anguila**

A pesar de que la anguila es una especie explotada comercialmente desde hace bastantes años, pocos son los estudios llevados a cabo sobre su crecimiento. La mayoría de los trabajos se han concentrado en la caracterización del crecimiento de poblaciones naturales (Rossi *et al.*, 1988; Fernández-Delgado *et al.*, 1989; Andersson *et al.*, 1991; Mann y Blackburn, 1991; Aprahamian, 2000), reduciéndose considerablemente el número de publicaciones que se centran en individuos confinados para consumo (Deelder, 1978; Kuhlmann, 1979; Deelder, 1981; Wickins, 1985; Kamstra, 1993; Pedersen, 1998). Una posible explicación a este fenómeno podemos encontrarla en la dificultad en el cálculo exacto de la edad y el crecimiento de esta especie (Deelder, 1981), lo que provoca una gran divergencia de los parámetros incluidos en los modelos de crecimiento. De esta forma, la longitud asintótica ( $L_{\infty}$ ) que junto con la constante  $K_v$  determinan la forma de la curva de crecimiento en el modelo de Von Bertalanffy,

pueden variar para distintas poblaciones naturales entre 37.5-135 cm y 0.978-0.058 años<sup>-1</sup> respectivamente (Rossi *et al.*, 1988).

Las anguilas confinadas en tanques de cultivo también presentan una pronunciada variabilidad en el crecimiento después de su captura. Se puede decir que, cuatro son los factores principales que contribuyen a tal variabilidad: a) el origen y estado físico de la población reclutada, b) la respuesta de los individuos al estrés de captura, transporte y condiciones ambientales en el tanque de cultivo, c) la proporción de anguilas que reanuda la alimentación, la de peces que se alimenta rápidamente y la adaptación a los piensos comerciales, y d) la densidad en el sistema así como la variación inicial del tamaño en la población. En este sentido, algunos trabajos indican que sólo un 6 % de cada lote inicial de angulas presenta un crecimiento rápido (con una tasa instantánea de crecimiento  $-G-$  de 1.75), un 10 % presenta un crecimiento normal ( $G=1.40$ ), mientras que un 44 % crecen de forma lenta ( $G=0.10$ ) (Jover-Cerdá, 1993). El 26 % restante presenta pérdidas de peso o se mantienen en un estado de crecimiento latente denominado 'estado de cola'. En este sentido, Degani y Levanon (1983) y Degani *et al.* (1984) observaron que un 3 % presentaba un crecimiento rápido, el 20 % presentó un crecimiento moderado, el 25 % un crecimiento lento y el 52 % un crecimiento nulo o negativo.

Una pronunciada variabilidad en el crecimiento también ha sido comprobada en otras especies comerciales como la carpa (*Cyprinus carpio*) (Wohlfarth, 1977) o la platija (*Pleuronectes platessa*) (Jobling, 1982). Sin embargo, en estas especies la proporción de la población que incrementa su talla de forma rápida es aproximadamente equivalente a la fracción de la población que incrementa su talla de forma lenta. Así, la mayor parte de la población se encuentra en todo momento con una talla que oscila alrededor de la talla media poblacional. En el caso de la anguila, la simetría en la distribución por tallas sólo es encontrada en las primeras fases del proceso de engorde o dentro de los subgrupos de crecimiento dentro de un mismo tanque, de ahí la constante necesidad de las reclasificaciones. De esta forma se consiguen establecer pequeños lotes homogéneos por tanques, evitando así el canibalismo y el dominio de los individuos de mayor porte. A pesar de ello, para el conjunto de la población en el sistema, la distribución del número de individuos tenderá a estar sesgada hacia las tallas superiores aproximándose por tanto a una distribución de tipo Poisson (Kuhlmann, 1979). Este fenómeno complica aun más la determinación del modelo de crecimiento al no cumplirse las condiciones de normalidad.

## **5.2. MATERIAL Y MÉTODOS**

### **5.2.1. Caracterización de la biomasa y mortalidad real**

La caracterización de la biomasa de la planta se ha llevado a cabo a partir de los datos registrados semanalmente en los 'partes de alimentación'. En estos partes de alimentación (utilizados por los operarios para repartir las raciones diarias por tanques), además del tipo y cantidad de pienso, se suministra información sobre el peso medio de

los individuos así como la biomasa total por tanque. Ambos términos son estimados semanalmente por el gestor de la planta a partir de la cantidad de pienso consumido en la semana anterior, las tablas de conversión del alimento suministrado por el fabricante así como la mortalidad registrada. A estos datos se suma la experiencia acumulada del gestor que, en función de la existencia de incidencias, focos patológicos, aplicación de tratamientos, clasificaciones, etc., aumenta o disminuye el índice de conversión del alimento, aumentando o disminuyendo, por tanto, la ganancia en biomasa y el peso medio de la población. De esta forma, para cada tanque tenemos un dato de biomasa total y peso medio por semana (tabla V.1).

**Tabla V.1.** *Ejemplo de parte de alimentación para dos tanques de la serie ‘A’ y los dos primeros días de la semana. Para cada día se indica la cantidad de alimento suministrado por la mañana y por la tarde*

Tanque	Fecha de inicio	Biomasa (kg)	Peso medio (g)	Tipo pienso	kg toma	Lunes	Martes
A1	02/11/99	255	8.0	X3	1.9	2.4 – 1.9	1.9 – 1.9
A8	28/10/99	216	4.7	X2	2.2	2.5 - 2.2	2.2 – 2.2

Teniendo en cuenta lo comentado anteriormente, la validación del modelo estará condicionada por la agudeza de la predicción del gestor, por lo que es interesante conocer su error de predicción. Para ello, se llevó a cabo una regresión lineal entre la biomasa real obtenida tras la clasificación y la estimada por el gestor para un total de 40 tanques. El error de predicción se estimó mediante la ecuación (4.19) (Capítulo 4, apartado 4.2.4).

De igual forma, en el parte de mortalidad de la planta se registran diariamente el número de individuos muertos por tanque.

Por lo tanto, tal y como se comento en el apartado 5.1.1, se cuenta tan solo con valores estimados de biomasa y peso medio para los cuales no se conoce la desviación alrededor de la media. De esta forma, para un tanque y momento determinado se conoce la biomasa en el tanque pero es imposible conocer la proporción de ejemplares que se encuentran por encima o por debajo de una talla en concreto. Teniendo en cuenta que cada tanque es reclasificado aproximadamente cada 30 ó 40 días, podemos suponer que la distribución de frecuencias por tallas en un tanque se aproxima a una distribución normal. Así, si una variable aleatoria  $x$  sigue una distribución normal, podemos decir que su función de densidad viene definida por:

$$f(x)_k = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-0.5\left(\frac{x-\mu}{\sigma}\right)^2\right]} \quad (5.1)$$



donde  $\mu$  es el peso medio de los ejemplares en el tanque  $k$  y  $\sigma$  es la desviación típica o estándar. La representación gráfica de la densidad de la distribución normal se muestra en la figura V.1 y se caracteriza porque: a) tiene forma de campana, b) es simétrica con respecto a la vertical de  $\mu$ , c) alcanza un máximo en  $\mu$ , d) es asintótica al eje de abscisas y, e) la distancia entre el eje vertical trazado en  $\mu$  y el punto de inflexión es  $\sigma$ .

A partir de la ecuación (5.1), es fácil obtener la proporción de individuos o la cantidad de biomasa por encima o por debajo de una determinada talla ya que, la probabilidad de que un individuo presente una talla  $x$  comprendida entre  $a$  y  $b$  queda definida como:

$$P_k(a \leq \text{Talla} \leq b) = \int_a^b f(x)_k dx \quad (5.2)$$

Asimismo, también es posible la obtención de un perfil de distribución semanal por tamaños, ya que se cumplirá que la función de distribución de la biomasa total [ $f(x)_B$ ] en la planta será:

$$f(x)_B = \sum_{T=1}^k f(x)_T \quad (5.3)$$

Sin embargo, como se puede desprender del examen de la ecuación (5.1), la función de densidad está determinada por la media y su desviación típica. Por tanto, es necesario calcular o estimar la desviación antes de poder aplicar la función de densidad normal por tanque.

### 5.2.2. Estimación de la desviación típica o estándar. Derivación de los modelos de varianza

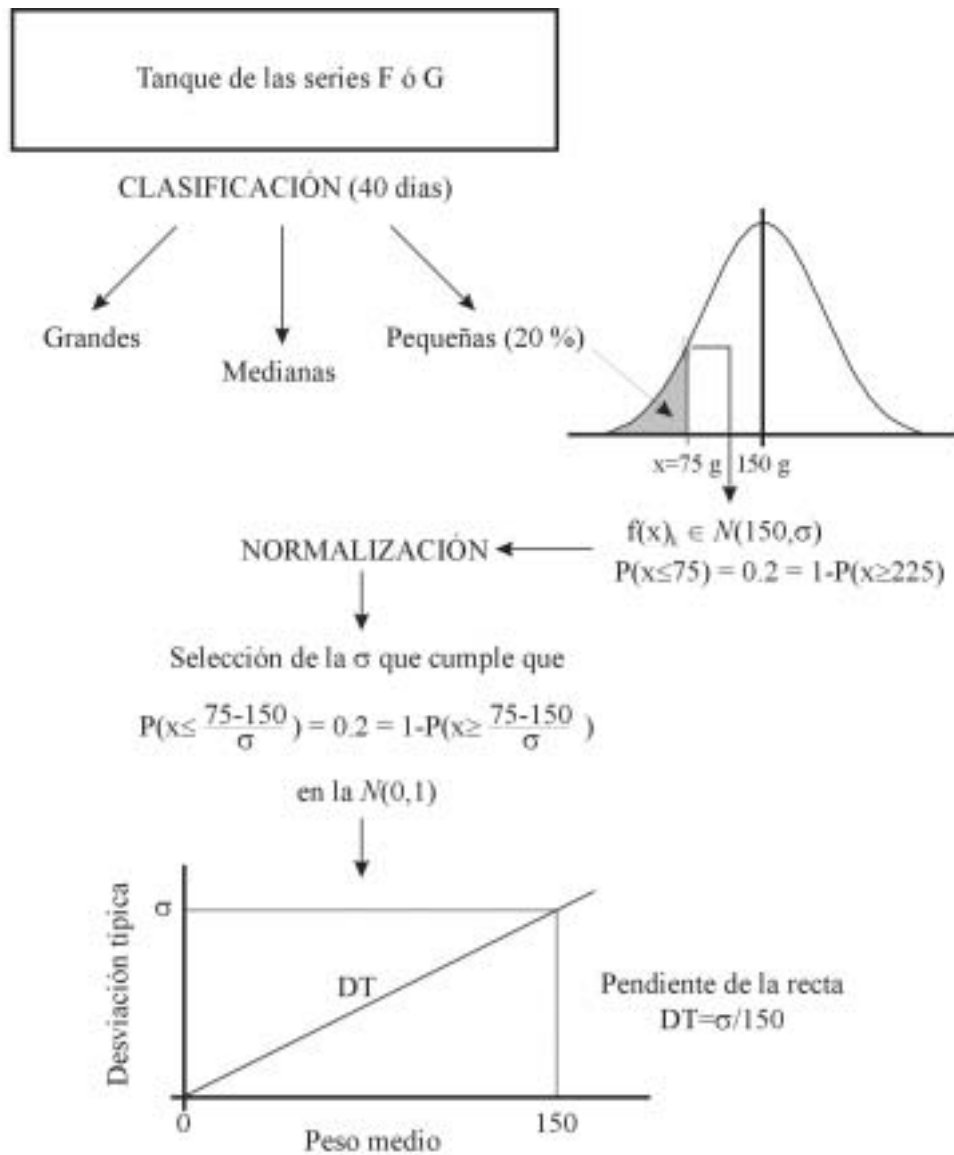
Matemáticamente, se puede definir la desviación típica como:

$$\sigma = \sqrt{\sum_{vi} \frac{(x_i - \mu)^2}{n - 1}} \quad (5.4)$$

donde  $x_i$  es la talla correspondiente al ejemplar  $i$ , y  $n$  es el número de ejemplares dentro de un tanque. De esta forma, el cálculo de la desviación típica para un tanque en un momento  $t$ , implica la medición de todos los ejemplares del tanque, o al menos de una muestra, para cada momento  $t$ , lo cual es inabordable. Por tanto, es necesario llevar a cabo una estimación de la desviación y establecer unos criterios de extensión de la desviación a todos los tanques.

Una primera estimación de la desviación típica puede realizarse a partir de los resultados de las clasificaciones llevadas cabo durante el proceso normal de manejo de la piscifactoría. Así, un examen de los resultados de las clasificaciones de aquellos tanques en los que el peso medio por ejemplar era de 150 gramos y que habían sido reclasificados en un período inferior a 40 días, reveló que aproximadamente el 20 % de la biomasa no superaba los 75 gramos de peso medio. A partir de esta información y

suponiendo, como ya se comentó en el apartado 5.2.1, que la distribución de esta subpoblación se aproxima a una normal, podemos calcular su desviación típica haciendo uso de las tablas de probabilidades de la distribución normal con media igual a cero y desviación típica igual a uno  $[N(0,1)]$ . A partir de esta primera estimación, podemos conocer la pendiente de la recta de variación de la desviación típica en función del peso medio, forzando el corte con el centro del eje de coordenadas, es decir, suponiendo que la desviación típica para un peso medio de cero gramos es cero (figura V.1).



**Figura V.1.** Esquema de la estimación de la desviación típica a partir de los resultados de las clasificaciones en los tanques de las series F y G

Se asume, por tanto, que la desviación típica es función del peso medio de los ejemplares y que se incrementa de forma lineal, aunque proporcionando una varianza distinta de cero para el momento  $t=0$ .

Una segunda estimación se puede llevar a cabo a partir del modelo propuesto por Sparre (1976) y modificado posteriormente por Arnason *et al.* (1992). Sparre desarrolló un modelo de crecimiento estocástico de un grupo de peces, asumiendo que todos tenían un mismo peso medio inicial y que crecían durante un período de tiempo  $t$  bajo condiciones de alimentación y temperatura constante. En este modelo, el peso de los ejemplares en el momento  $t$  está normalmente distribuido, con una media  $\mu_t$  y una varianza  $\sigma_t$ , donde  $\mu_t$  es función del tiempo ( $t$ ), del peso medio inicial ( $\mu_0$ ), de la temperatura ( $T$ ) y de la menor tasa de alimentación que a la temperatura  $T$  produce el máximo crecimiento ( $TAT$ ). Por otra parte, la varianza viene definida como:

$$\sigma_t^2 = \alpha^2 t^2 \mu_t^2 \quad (5.5)$$

donde el parámetro  $\alpha$  especifica la tasa de crecimiento en el coeficiente de variación ( $CV$ ) por unidad de tiempo. Como se puede observar en la ecuación (5.5), este modelo de varianza no está directamente relacionado con las condiciones de crecimiento de la población ( $T$  y  $TAT$ ) ni con el tamaño inicial ( $\mu_0$ ), pero predice la variación en peso final en función del peso medio ( $\mu_t$ ) y del tiempo ( $t$ ). Una consecuencia directa de la formulación del modelo es que su aplicación a dos subpoblaciones que crecen bajo condiciones diferentes (por ejemplo: alta temperatura y baja tasa de alimentación, y baja temperatura y alta tasa de alimentación) en un intervalo de tiempo dado, producirán las mismas varianzas en sus crecimientos.

Desde un punto de vista biológico, el modelo de varianza de Sparre (ecuación 5.5) incorpora algunas propiedades realistas ya que, como consecuencia de la inclusión del parámetro tiempo en la ecuación, el modelo predice que un lote de peces mantenido a un peso medio constante mediante manipulación de la tasa de alimentación y la temperatura, continuará incrementándose en varianza. Por otra parte, el modelo permite también incrementar o disminuir la varianza a tasas de alimentación por debajo de la de mantenimiento, en función de como de rápido disminuye  $\mu_t$  con respecto al tiempo. Sin embargo, el modelo también incorpora propiedades no realistas ya que asume una varianza cero en el momento  $t=0$ , y supone un incremento lineal del coeficiente de variación en función del tiempo ( $CV_t = \alpha t$ ).

Arnason *et al.* (1992) modifican este modelo simple (ecuación 5.5) mediante la incorporación de tres parámetros ( $\alpha$ ,  $\beta$  y  $\gamma$ , reemplazando  $\alpha^2$ , y los dos exponentes de  $t$  y  $\mu_t$  respectivamente):

$$\sigma_t^2 = \alpha t^\beta \mu_t^\gamma \quad (5.6)$$

Sin embargo, el modelo propuesto por Arnason *et al.* (1992) arrastra alguno de los inconvenientes del modelo del cual se origina ya que sigue considerando que la varianza en el momento  $t=0$  es cero. De esta forma, Arnason *et al.* (1992) incorporan un nuevo término al modelo cuando  $t \neq 0$ , de tal forma que:

$$\sigma_t^2 - \sigma_0^2 = \alpha t^\beta \mu_t^\gamma \quad (5.7)$$

donde  $\sigma_0^2$  es la varianza de la población en el momento  $t=0$ .

Un inconveniente que plantea la ecuación (5.7) es que los parámetros  $\alpha$ ,  $\beta$  y  $\gamma$  (independientes de las condiciones de crecimiento) deben ser calculados experimentalmente para cada especie y cada población. Este problema puede ser superado si se tiene en cuenta que la ecuación (5.7) puede ser transformada a:

$$\ln(\sigma_t^2 - \sigma_0^2) = \ln(\alpha) + \beta \ln(t) + \gamma \ln(\mu_t) \quad (5.8)$$

Esta ecuación (5.8) presenta la misma estructura que el modelo general de regresión múltiple:

$$y_i = b_0 + b_1 L t_i + b_2 L w_i + \varepsilon \quad (5.9)$$

donde  $y_i = \ln(\sigma_i^2 - \sigma_0^2)$ ,  $L t_i = \ln(t_i)$ ,  $L w_i = \ln(w_i)$  y  $b_0 = \ln(\alpha)$ ,  $b_1 = \beta$ ,  $b_2 = \gamma$  y  $\varepsilon =$  **error residual**, son los estimadores de la regresión múltiple.

El cálculo de los parámetros mencionados en la ecuación (5.7) se llevó a cabo aprovechando los datos tomados de un total de 700 ejemplares del tanque 'A1' durante el período comprendido desde el fin de la primera clasificación (20/08/1999) hasta la segunda clasificación del tanque (16/09/1999). Los 700 ejemplares, repartidos en lotes de 50 anguilas permitieron obtener una estimación de la desviación estándar en el momento  $t=0$ , a través de un análisis de regresión simple entre el peso medio de cada lote y su desviación. Asimismo, se hicieron mediciones de longitud total (**LT**, mm) y de peso fresco (**PF**, g) en un total de 84 anguilas para la obtención de la relación longitud-peso. A través de la ecuación de esta regresión se estimó la longitud total para los 700 ejemplares pesados anteriormente.

### 5.2.3. La tasa de crecimiento

A una temperatura constante, se puede asumir que el peso medio de una cohorte de peces se incrementa de forma exponencial durante su primer año y medio de vida. Esto puede ser definido matemáticamente como (Wootton, 1991):

$$\mu_t = e^{[G(t-t_0) + \ln(\mu_0)]} \quad (5.10)$$

donde  $\mu_t$  es el peso medio en el momento  $t$ ,  $G$  es la tasa instantánea de crecimiento y  $\mu_0$  es el peso medio inicial.

Este tipo de crecimiento es fácilmente observable en cualquier especie cultivada, aunque como es lógico, varía en función de la especie, temperatura del agua y tasa de alimentación. Como se puede ver en la ecuación (5.10), el modelo asume que el crecimiento es independiente de la densidad ( $\text{kg/m}^3$ ) cuando ésta se encuentra por debajo de la capacidad de carga de la instalación. Esto es válido siempre y cuando no se den comportamientos inhibitorios del crecimiento a altas densidades y las condiciones

de calidad del agua se mantengan relativamente constantes (Brett, 1979). En el caso de la anguila, varios estudios indican que la tasa de crecimiento alcanzada por grupos cultivados a altas densidades es mayor a la encontrada para grupos cultivados a bajas densidades (Wickins, 1985). Por otra parte, Kebus *et al.* (1992) demostraron que cuando la calidad del agua es adecuada, incluso especies muy exigentes como la trucha, pueden ser cultivadas a densidades cercanas a los 270 kg/m<sup>3</sup> sin interrupción del crecimiento.

Ya que como se expuso en el apartado 5.1.2., el crecimiento de la anguila está sometido a una fuerte variación a lo largo del tiempo, se ha dividido cada cohorte en tres subgrupos de crecimiento siguiendo los resultados encontrados por Jover-Cerdá (1993) entre otros. Así, se han definido los subgrupos de crecimiento como ‘Lento’, ‘Normal’ y ‘Rápido’, asignándole a cada uno de ellos una tasa de conversión del alimento (**FCR**). Se cumple por tanto que  $\mathbf{FCR}_{\text{Lento}} > \mathbf{FCR}_{\text{Normal}} > \mathbf{FCR}_{\text{Rápido}}$ , siendo:

$$\mathbf{FCR} = \frac{\text{Alimento aportado (kg)}}{\Delta \text{Biomasa (kg)}} \quad (5.11)$$

Así se asume, que el alimento no es un factor limitante, o dicho de otra forma, se aporta la cantidad de alimento que sea necesaria para mantener **FCR** constante. La ecuación (5.11) está íntimamente relacionada con la (5.10) ya que, al mantener un **FCR** constante también se consigue mantener constante la tasa de crecimiento (**G**) de forma independiente para cada subgrupo.

Dado que el crecimiento de la anguila (como la de cualquier especie piscícola) depende de forma considerable de la temperatura, se hace necesario incluir este parámetro físico en el modelo. De esta forma, según Brown (2000) el rango de crecimiento de la anguila se encuentra entre 12.7 y 29.9 °C. Otros autores como Más-Álvarez y Barrera-Orozco (1989) incrementan el límite inferior hasta 15 °C. Por otra parte, Degani *et al.* (1988) establecen la temperatura óptima de crecimiento entre 23 y 25 °C. A partir de estos datos podemos construir una curva que relaciona el grado o factor de crecimiento (**FC**) con la temperatura simplemente asignando un factor de crecimiento igual a uno cuando la temperatura es de 24 °C, y cero cuando la temperatura es inferior a 15 °C o superior a 29.9 °C. Una regresión polinómica de tercer grado nos aportará la expresión matemática que relaciona ambos términos (figura V.2):

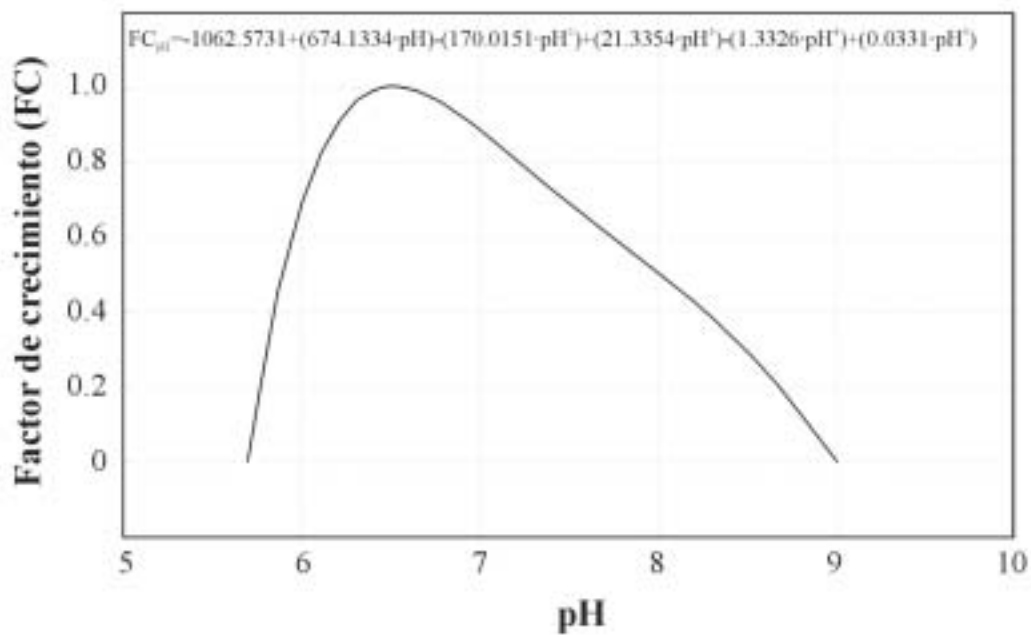
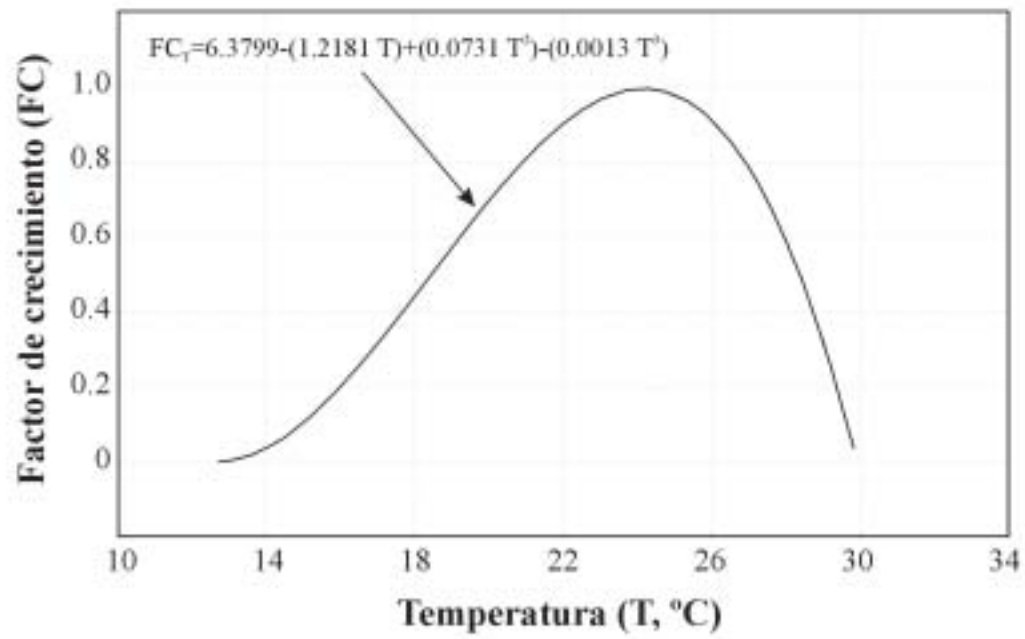
$$\mathbf{FC}_T = 6.3799 - (1.2181 \cdot T) + (0.0731 \cdot T^2) - (0.0013 \cdot T^3) \quad (5.12)$$

Relacionando esta nueva expresión con la ecuación (5.10) se tiene que:

$$\mathbf{G}_t^T = \mathbf{G}_t \cdot \mathbf{FC}_T \quad (5.13)$$

donde  $\mathbf{G}_t^T$  es la tasa de crecimiento de la cohorte en el momento **t** para la temperatura **T** y  $\mathbf{FC}_T$  es el factor de crecimiento para la temperatura **T**.

Con los parámetros amoníaco, nitrito y nitratos se procede de la misma forma que con la temperatura. Sin embargo en este caso al no conocerse con exactitud la relación entre dichos parámetros y el factor de crecimiento, se supone que la relación existente es de tipo lineal. Así en el caso del amoníaco, el rango de crecimiento se



**Figura V.2.** Funciones de los factores de crecimiento (FC) debidos a la temperatura del agua y al valor del pH

encuentra entre 0 mg/l y 5.8 mg/l (suponiendo un pH=8.5 y una T=25 °C) (Brown, 2000). Matemáticamente podemos expresarlo como:

$$FC_{Am} = 1-(0.1725 \cdot [NH_3]) \quad (5.14)$$

donde  $FC_{Am}$  es el factor de crecimiento para la concentración de amoníaco  $[NH_3]$ .

Para los nitritos, el rango de crecimiento se encuentra aproximadamente entre 0 mg/l y 15.5 mg/l, mientras que para los nitratos el rango de crecimiento se extiende desde 0 mg/l hasta aproximadamente 500 mg/l. De esta forma, para estos dos parámetros:

$$FC_{Ni} = 1-(0.0645 \cdot [NO_2^-]) \quad (5.15)$$

$$FC_{Na} = 1-(0.0020 \cdot [NO_3^-]) \quad (5.16)$$

donde  $FC_{Ni}$  es el factor de crecimiento para la concentración de nitritos  $[NO_2^-]$ , y  $FC_{Na}$  es el factor de crecimiento para la concentración de nitratos  $[NO_3^-]$ .

En el caso del pH, los rangos de crecimiento se corresponden con pHs entre 5.7 y 9.0, aunque se tiende a encontrar mejores crecimientos en valores de pH ligeramente ácidos, ya que de esta forma el porcentaje de amoníaco total disuelto no ionizado ( $NH_3$ ) es menor (Shepherd y Bromage, 1999). Matemáticamente podemos expresarlo como una función polinómica de quinto grado (figura V.2):

$$FC_{pH} = -1062.5731+(674.1334 \cdot pH)-(170.0151 \cdot pH^2)+(21.3354 \cdot pH^3)-(1.3326 \cdot pH^4)+(0.0331 \cdot pH^5) \quad (5.17)$$

donde  $FC_{pH}$  es el factor de crecimiento para el valor **pH**.

Por tanto, relacionando la ecuación (5.10) con las ecuaciones (5.12) y (5.14) a (5.17), se tiene que:

$$G_t^{FC} = G_t \cdot FC_T \cdot FC_{Am} \cdot FC_{Ni} \cdot FC_{Na} \cdot FC_{pH} \quad (5.18)$$

donde  $G_t^{FC}$  es la tasa de crecimiento de la cohorte en el momento **t** teniendo en cuenta la temperatura, concentración de amoníaco, concentración de nitritos, concentración de nitratos y el valor del pH.

El cálculo de los factores de crecimiento se ha llevado a cabo a partir de las series históricas de cada parámetro físico-químico promediado con su correspondiente predicción, es decir, haciendo uso de los modelos calibrados en el capítulo 4.

#### 5.2.4. La tasa de mortalidad

Por regla general, en un sistema real la tasa de mortalidad (**Z**) es muy elevada durante las primeras fases de aclimatación de los peces a la planta. Después de la aclimatación, la tasa de mortalidad disminuye y se hace proporcional al número de ejemplares en una

cohorte en un momento determinado (McNown y Seireg, 1983), asumiendo la ausencia de mortalidad aguda por la presencia de enfermedades, fallos en el manejo, etc. Para poder incluir la mortalidad en el modelo es necesario suponer que: a) la tasa de mortalidad es constante a lo largo del tiempo, y b) la alta mortalidad inicial es contabilizada el primer día de estancia del lote en la planta. De este modo, podemos definir matemáticamente la tasa de mortalidad como:

$$Z = -\frac{\ln(N_{t+1}) - \ln(N_t)}{\Delta t} \quad (5.19)$$

donde  $N_{t+1}$  es el número de ejemplares en la cohorte en el momento  $t+1$ , y  $N_t$  es el número de ejemplares en el momento  $t$ .

Una vez calculada la tasa de mortalidad diaria media, los datos son agrupados semanalmente para poder llevar a cabo las estimaciones de mortalidad relativa con respecto a la biomasa total. De esta forma, se puede proceder de la misma manera que con la biomasa. Es decir, aplicando la función de distribución normal por tanque para sumarlas posteriormente. Así podemos obtener un patrón de mortalidad semanal similar al patrón de biomasa.

### 5.2.5. Modelación de la cosecha

En un sistema de cosecha a la demanda (como en el caso de Hidrorecursos), la constitución de los lotes de venta está supeditada al proceso de clasificación. De esta forma, tras la solicitud de pedido, se reclasifican aquellos tanques que son susceptibles de contener ejemplares con un tamaño igual o superior al demandado, separando aquellos que cumplen las condiciones solicitadas. Este proceso se llevará a cabo tantas veces como sea necesario hasta llegar al peso establecido en el lote de venta. Una vez clasificados, los ejemplares que no cumplen las condiciones se reincorporan a aquellos tanques cuyas características biológicas (básicamente el peso medio del tanque y biomasa) se aproximan más a las calculadas tras la clasificación. Por tanto, las características biológicas del tanque se modifican en función de la nueva incorporación.

Suponiendo un subgrupo que se distribuye normalmente en el cual, parte de los ejemplares que lo componen han superado el tamaño de cosecha ( $W_c$ ), podemos expresar la biomasa de cosecha ( $B_c$ ) o un lote de venta como:

$$B_c = \int_{W_{BC}}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-0.5\left(\frac{x-\mu}{\sigma}\right)^2\right]} dx \quad (5.20)$$

siendo  $W_{BC} \geq W_c$ .

Si se consideran dos subgrupos de crecimiento en los que alguno de sus ejemplares supera la talla de cosecha, la biomasa de cosecha se calculará siguiendo la ecuación (5.20) pero teniendo en cuenta de forma prioritaria al primer subgrupo en



entrar en el sistema, hasta que los tamaños máximos dentro de ambos subgrupos se igualan. Si tenemos más de dos subgrupos se procede de igual forma. El efecto global de la constitución de lotes de venta sobre el total de la población se puede observar en la figura V.3.

Una vez cosechado el lote de venta hay que proceder al reajuste de la media y desviación del subgrupo modificado en función del número de individuos restantes en el subgrupo. Este reajuste produce aparentemente un efecto irreal al disminuir la biomasa correspondiente a la cola derecha (individuos de mayor porte), pero simula aproximadamente el efecto que de forma global se produce como consecuencia del estrés en la población tras la clasificación.

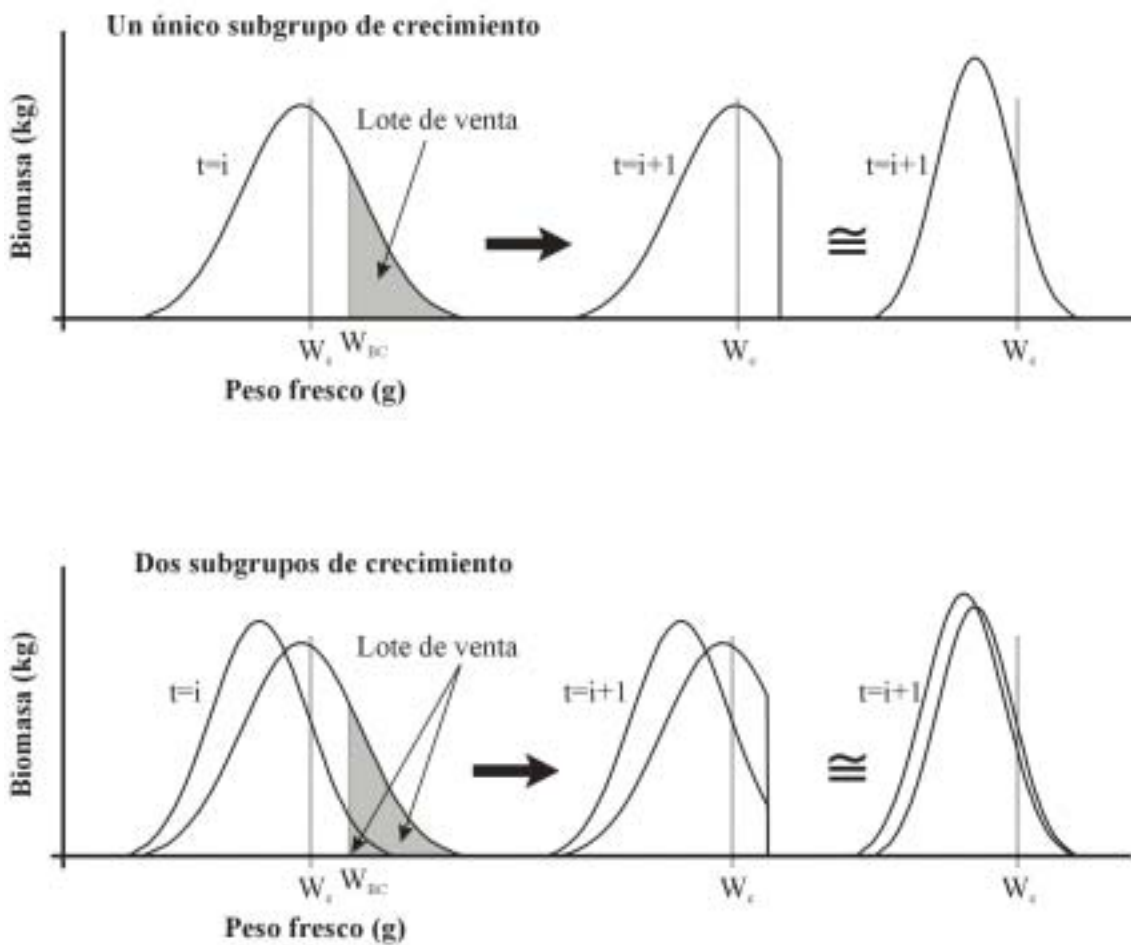


Figura V.3. Esquema de la formación de los lotes de ventas y del reescalado y ajuste del nuevo peso medio para uno y dos subgrupos de crecimiento

### 5.2.6. Proceso de simulación dinámica. Valoración económica inicial

Establecidos de forma básica los rangos de las condiciones iniciales de producción (tipo de estrategia, crecimiento y subgrupos de crecimiento, desviaciones iniciales asociadas a los crecimientos, modelos de varianza y mortalidad) es posible simular la evolución del *stock*, producción y mortalidad en función del tiempo. Este proceso de simulación se ha llevado a cabo mediante lo que en el capítulo 2 (apartado 2.3.2.1) se denominó **POS** o ‘Programa Operativo de Simulación’ (manual del usuario descrito en el anexo V.2). Éste permite no solo simular multitud de escenarios de variación de la biomasa en la planta, sino que también puede conectarse fácilmente con **CONTROLPLANT 1.0** y **SEDPA 1.0** a través del conjunto de reglas **A-A** de **CONTROLPLANT**. En su código fuente se han incorporado los modelos de predicción de parámetros físico-químicos que en el capítulo 4 mostraron un mejor comportamiento durante la validación. De esta forma, además de predecir la variación en crecimiento, **POS** puede prever la aparición de focos patológicos. Un esquema del funcionamiento de **POS** se muestra en la figura V.4.

**POS** también lleva a cabo una valoración económica básica en la que se tienen en cuenta costes de personal, la inversión en biomasa (compra de angulas), costes por consumo de pienso, costes energéticos, la amortización de la inversión inicial en la instalación y la venta del producto. Ya que la simulación se lleva a cabo día a día, todos los costes considerados (fijos y variables) se van a cuantificar en esa unidad temporal. En los costes de personal se han tenido en cuenta el número de empleados técnicos y operarios de planta.

A pesar de que la compra de angulas suele realizarse repartida en dos o tres partidas procedentes de la misma zona a lo largo de un período tiempo relativamente corto (un mes o mes y medio), se ha contabilizado en un único día para toda la temporada ya que esto facilita la simulación y los resultados no son significativamente distintos.

Como se comentó en el apartado 5.2.3, la cantidad de pienso consumido se obtiene a partir de la ecuación (5.11). Una vez calculada, tan solo es necesario multiplicar por los precios marcados por el fabricante. A lo largo de todo el período de estudio siempre se han utilizado las mismas granulometrías de pienso extruído proporcionadas por la marca Dibaq. En la tabla V.2 se muestran las características y precios del pienso utilizado.

Para contabilizar los costes energéticos, se ha utilizado el tipo de contrato que Hidrorecursos S.A. tiene con la compañía suministradora, en este caso Endesa Energía S.A. Este contrato, enmarcado dentro del mercado liberalizado de la electricidad, se caracteriza por la inclusión de un término de energía en el cual el precio por kilovatio-hora (kWh) es de 0.05 € (8.47 pta). La revisión de las 5 últimas facturas generadas indican que el consumo medio mensual de Hidrorecursos durante el período de estudio fue de  $71489 \pm 3843$  kWh. De esta forma, se ha tenido en cuenta un consumo fijo de 71489 kWh/mes.

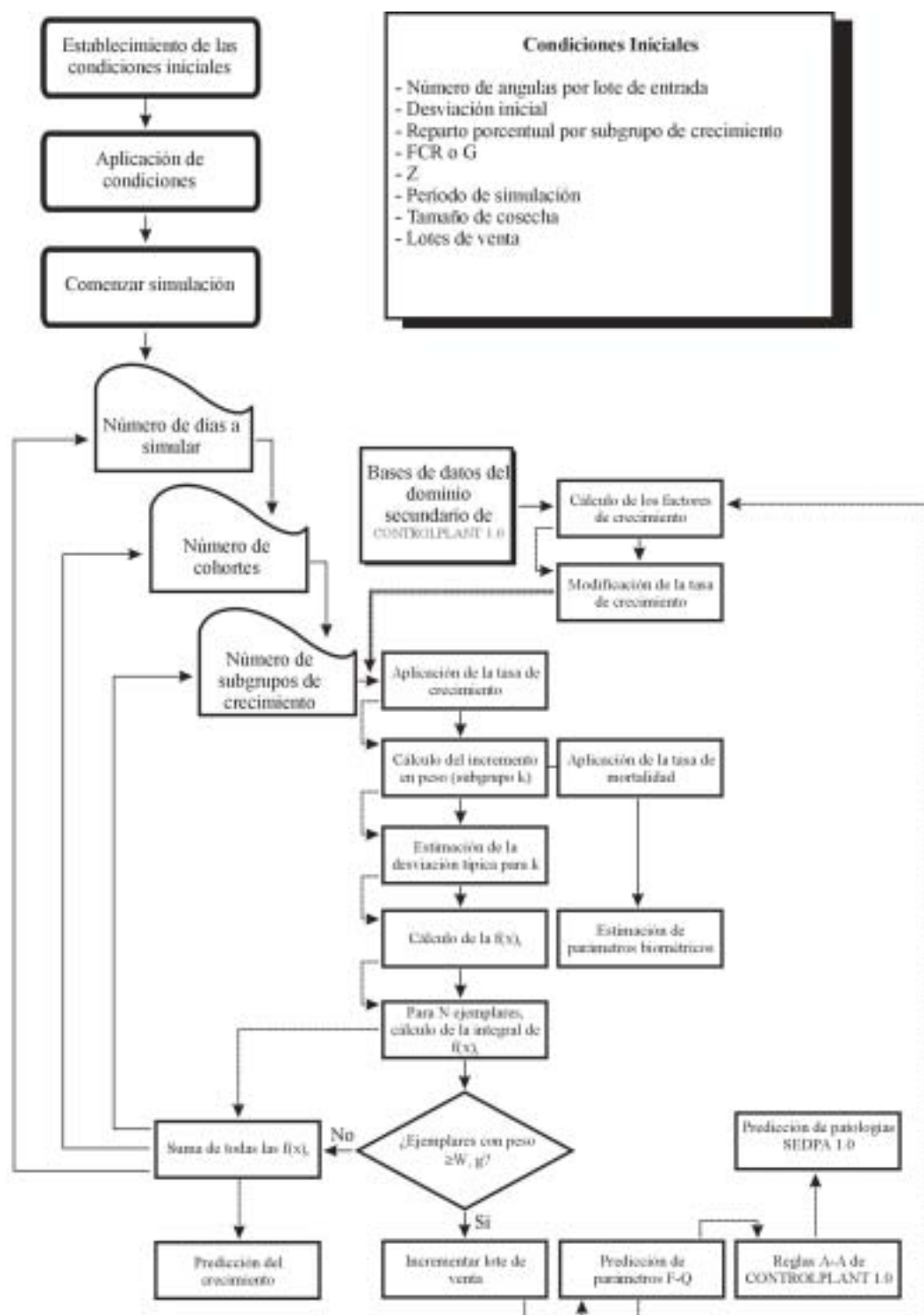


Figura V.A. Esquema del funcionamiento del Programa Operativo de Simulación (POS). Las flechas continuas indican simulación en tiempo real mientras que las discontinuas indican simulación en predicción para al menos un día de antelación

**Tabla V.2.** Características del pienso extruído utilizado durante el período de estudio

Denominación	Tamaño del grano (mm)	Precio (€/kg)	Peso del pez recomendado (g)
Microbaq-0 ó X0	-	1.36	<1
Microbaq-1 ó X1	-	1.36	<1
Microbaq-8 ó X2	-	1.30	1-3
Microbaq-10 ó X3	-	1.26	3-10
Microbaq-15 ó X4	-	1.24	10-20
Ext-Anguila 2mm	2	1.08	20-100
Ext-Anguila 3mm	3	1.06	>100

Los costes de inversión inicial en instalaciones constituyen por regla general un pago único que es necesario realizar para acometer la construcción y puesta en servicio de la planta. Para poder comparar el coste de inversión con el resto de costes es necesario referirlo en primer lugar a un término anual. De este modo se utiliza comúnmente el factor de amortización  $a_t$ , el cual representa el coste anual de amortización de una inversión producida en el año inicial y que se amortiza a lo largo de  $A$  años a una tasa de interés  $r$ . De esta forma, la inversión de un capital inicial  $C_I$  representa unas cargas anuales de amortización  $C_I \cdot a_t$  unidades monetarias, donde:

$$a_t = \frac{r(1+r)^A}{(1+r)^A - 1} \quad (5.21)$$

Una vez calculada la carga  $C_I \cdot a_t$  tan solo es necesario dividir por 365 para obtener el coste de amortización por día.

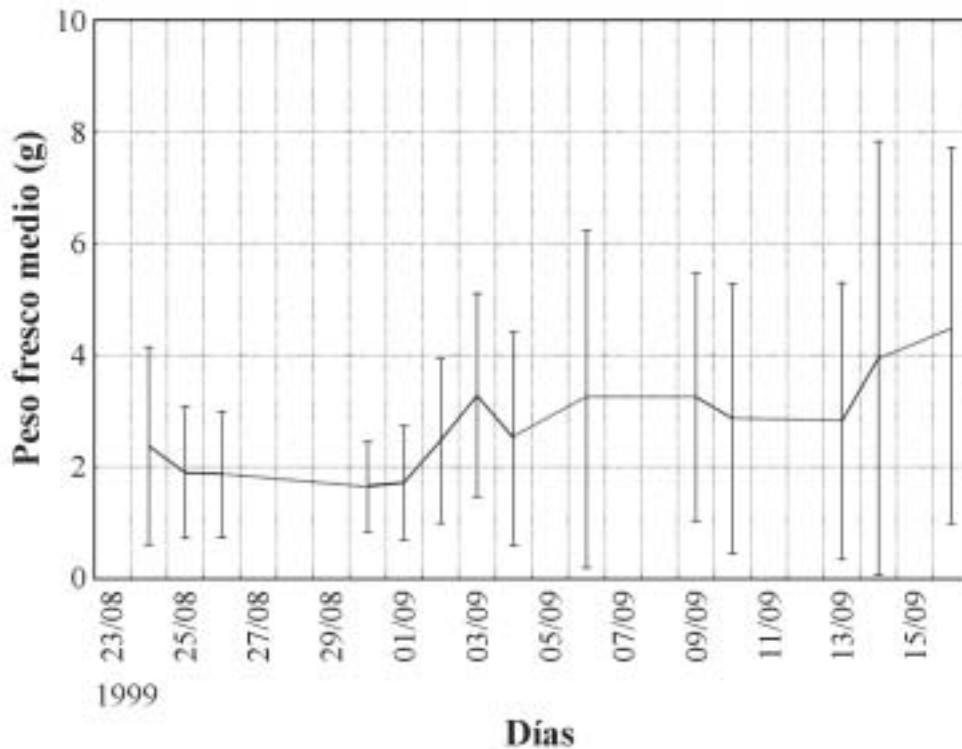
La suma de todos los costes (fijos y variables) tiene que se contrarrestada en base a la venta del producto final. Como ya se comentó en el capítulo de introducción, el precio del kilogramo de anguila se ha mantenido prácticamente constante desde 1986 oscilando entre 5.39-13.08 €/kg (897-2177 pta/kg). Así, se ha considerado un precio medio estimado en 6.61 €/kg (1100 pta/kg).

## 5.3. RESULTADOS

### 5.3.1. Error de predicción del gestor

La regresión lineal entre la biomasa estimada por el gestor de la planta y la biomasa real contenida en un total de 40 tanques se muestra en la figura V.5. Como se puede observar la regresión produce resultados muy satisfactorios ya que se consigue explicar más del 97 % de la variable dependiente. Asimismo, la pendiente y la intersección de la recta con el eje de abscisas indica que el gestor tiende a infraestimar ligeramente en tanques donde el contenido de biomasa es pequeño y a sobreestimar cuando la biomasa

almacenada es grande. Por otra parte, se encontró que el error estándar de predicción fue del 18.5 %.

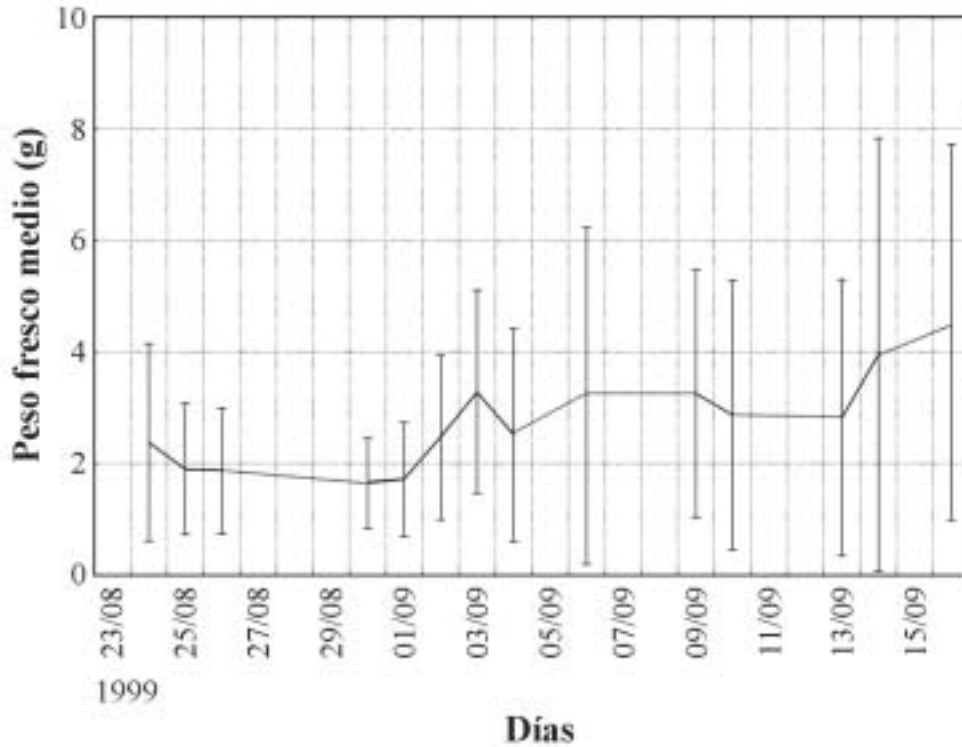


**Figura V.6.** Variación del peso fresco medio en el tanque 'A1' desde el 24 de agosto de 1999 hasta el 16 de septiembre de 1999. Para cada valor medio se pesaron un total de 50 ejemplares. Las zonas punteadas indican ausencia de datos.

### 5.3.2. Análisis de las muestras del tanque 'A1'. Relación longitud-peso

El análisis de las muestras obtenidas del tanque 'A1' desde el 24 de agosto de 1999 hasta el 16 de agosto del mismo año, mostró un incremento significativo en el peso medio fresco a partir del séptimo muestreo (03/09/1999) como así mostró un análisis de la varianza y un posterior test de Tukey (ANOVA:  $F=7.092$ ;  $p_{\alpha}<0.001$ ; Tukey:  $p_{\alpha}<0.05$ ) (figura V.6).

Por otra parte, en la tabla V.3 se muestra el resultado de la regresión no lineal entre el peso fresco medio de cada lote y su desviación típica correspondiente.

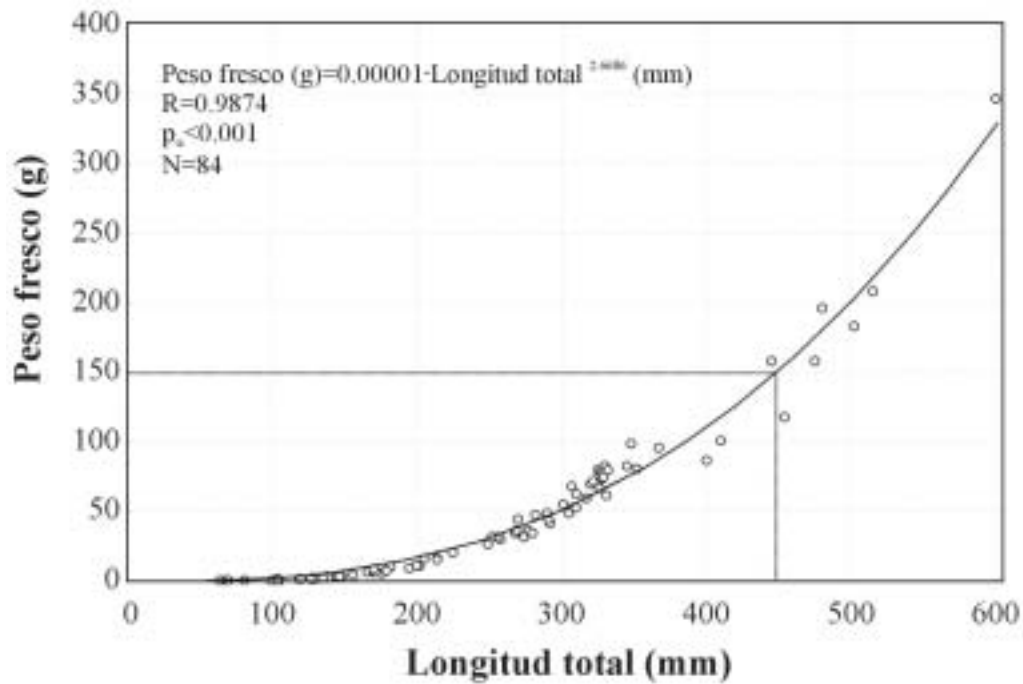


**Figura V.6.** Variación del peso fresco medio en el tanque 'A1' desde el 24 de agosto de 1999 hasta el 16 de septiembre de 1999. Para cada valor medio se pesaron un total de 50 ejemplares. Las zonas punteadas indican ausencia de datos

<b>Tabla V.3.</b> Parámetros estimados de la regresión no lineal entre la desviación típica y el peso medio de 14 lotes de 50 anguilas en el tanque 'A1'	
Modelo: $\sigma = a \cdot \mu^b$	
R=0.9306; Varianza explicada=86.5950 %; $p_\alpha < 0.05$ ; N=14	
<b>a</b>	<b>b</b>
1.5295	0.6904

Asimismo, los resultados de la relación longitud-peso (figura V.7) obtenida mediante regresión no lineal de tipo multiplicativo se muestra en la tabla V.4.

<b>Tabla V.4.</b> Parámetros estimados de la regresión no lineal entre la longitud y el peso fresco para un total de 84 anguilas	
Modelo: $PF = a \cdot LT^b$	
R=0.9874; Varianza explicada=97.4850 %; $p_\alpha < 0.001$ ; N=84	
<b>a</b>	<b>b</b>
0.00001	2.6686



**Figura V.7.** Relación longitud-peso. Se indica el modelo, el coeficiente de correlación, el nivel de significación y el número de datos. La zona punteada relaciona la longitud y el peso para una anguila de tamaño comercial

### 5.3.3. Estimación de los rangos de variación de la desviación típica. Aplicación de los modelos de varianza

A partir de la metodología expuesta en el apartado 5.2.2 (figura V.1), se ha llevado a cabo una primera estimación de la desviación típica para aquellos tanques con peso medio de 150 gramos y clasificados antes de 40 días. De esta forma, la desviación típica calculada que cumple que:

$$P\left(x \leq \frac{75-150}{\sigma}\right) = 0.2 = 1 - P\left(x \geq \frac{75-150}{\sigma}\right) \text{ en la } N(0,1) \quad (5.22)$$

es  $\sigma=89$ . Por tanto, la pendiente de la recta **DT** (figura V.1) es  $89/150=0.59$ .

Esto significa que, suponiendo una relación lineal entre el tamaño medio y la variación en tamaño, la desviación típica inicial para una cohorte de entrada a la planta con un peso medio inicial de 0.30 gramos es 0.18 gramos (figura V.1).

El valor de 0.18 gramos de desviación suponiendo una relación lineal entre el peso medio y su variación en el tiempo, es significativamente inferior al calculado a partir de las mediciones llevadas a cabo sobre anguilas del tanque 'A1'. Utilizando el modelo de la tabla V.3, la desviación típica para una cohorte de entrada de 0.3 gramos de peso medio es 0.67 gramos.

En la tabla V.5 se muestran los parámetros de la regresión múltiple (veánse ecuaciones 5.6 a 5.9) estimados a partir de las medidas realizadas en el tanque 'A1'. A partir de esta table se puede deducir que el parámetro  $\alpha=2.0522$ .

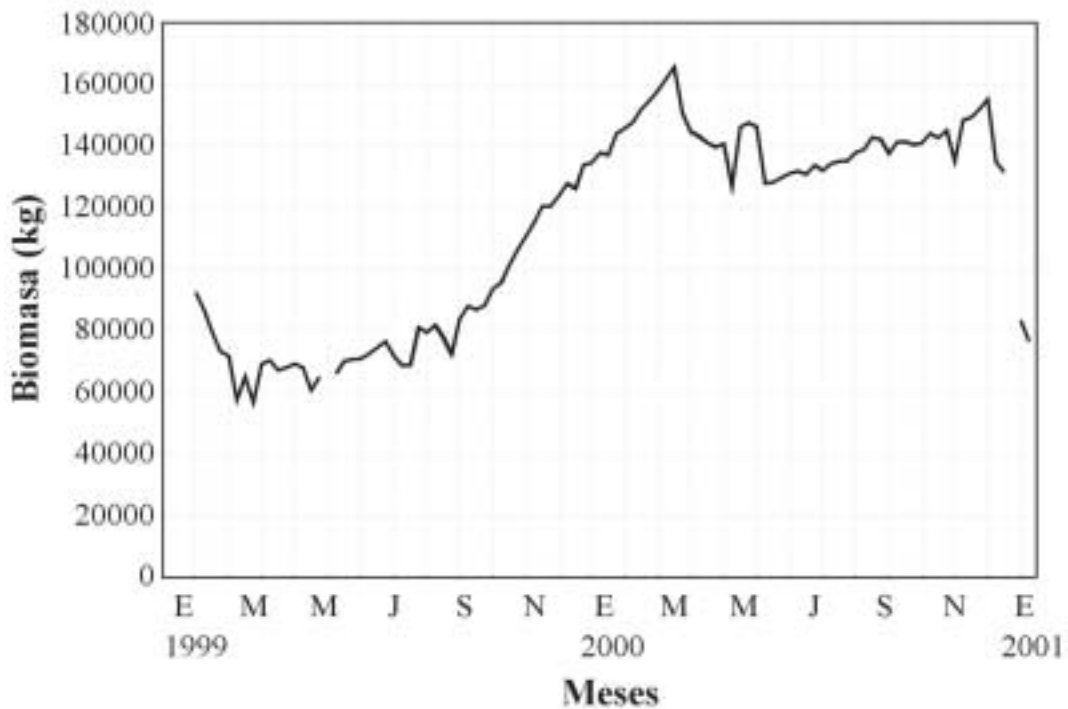
**Tabla V.5.** *Parámetros de estimación de la regresión múltiple en el modelo de varianza de Arnason-Sparre*

R=0.9503; R <sup>2</sup> =0.9031; F(2;11)=51.276; p <sub>α</sub> <0.001; N=14		
<b>b<sub>0</sub>=ln(α)</b>	<b>b<sub>1</sub>=β</b>	<b>b<sub>2</sub>=γ</b>
0.7189*	1.3223*	0.0710 <sup>†</sup>

\* p<sub>α</sub><0.001; <sup>†</sup> No significativo

### 5.3.4. Caracterización de la biomasa

La evolución de la biomasa total en la planta desde enero del año 1999 hasta enero del año 2001 se muestran en la figura V.8.



**Figura V.8.** *Biomasa total durante el periodo comprendido entre el 11/02/1999 y el 02/01/2001*



Se puede observar como durante ese período, la biomasa oscila alrededor de un valor promedio como consecuencia de fases de crecimiento rápido y lento, incorporación de nuevos ejemplares, mortalidades y ventas. Los límites mínimos y máximos se encuentran separados por aproximadamente un año, desde un nivel por debajo de los 60000 kg en febrero de 1999 hasta cerca de los 170000 kg en febrero-marzo del año 2000, lo que supone un incremento porcentual anual de 142 puntos. A partir de este nivel máximo, la biomasa se mantiene estable alrededor de las 140 toneladas, para disminuir drásticamente a principios del año 2001.

Un análisis de la varianza mostró la presencia de marcadas diferencias significativas entre la biomasa almacenada por meses (ANOVA:  $F=73.192$ ;  $p_{\alpha}<0.001$ ) a lo largo de todo el período de estudio. Un test de Tukey posterior indicó un aumento significativo de la biomasa desde septiembre de 1999 hasta febrero de 2000 ( $p_{\alpha}<0.001$  en todos los casos) así como un descenso significativo entre noviembre de 2000 y principios de 2001 ( $p_{\alpha}<0.001$ ).

Desde un punto de vista anual también se encontraron importantes diferencias significativas tal y como mostró un test t de Student de comparación de medias ( $p_{\alpha}<0.001$ ), siendo la biomasa almacenada en el año 2000 significativamente mayor que en el año 1999.

En las figuras V.9a a V.9e se muestra la evolución por semanas de la biomasa total almacenada en la planta tras la aplicación de funciones normales por tanques y asumiendo una desviación típica inicial de 0.2 gramos. Tal y como mostró Kuhlmann (1979) la distribución semanal por tallas para todo el período de estudio se caracteriza por un claro sesgo hacia las tallas de mayor porte, intuyéndose una aproximación hacia una función de tipo Poisson si ésta es discretizada (figura V.10). Por otra parte también se identifica en todas las semanas un tipo de distribución bimodal, trimodal o con un mayor número de modas, lo que coincide con los dos, tres o más subgrupos de crecimiento observados por Jover-Cerdá (1993) entre otros (figuras V.11 y V.12). Asimismo, en todas las semanas es posible observar una moda que dependiendo de la época del año varía entre 0 y 25 gramos, que se corresponde con una parte importante de la población que se mantiene en un estado de crecimiento extremadamente lento o nulo (figura V.13).

Cuando se representa la biomasa a lo largo del período de estudio para una talla concreta se puede observar la incorporación y disminución de ejemplares o biomasa a dicha talla. Así, en la figura V.14 se muestra la variación en biomasa de las tallas menores o igual a 0.5 gramos. En éstas se muestran con claridad los momentos de incorporación de nuevos ejemplares a la planta así como los períodos de aclimatación caracterizados por la pérdida de peso temporal de los ejemplares de mayor tamaño. Por otra parte, en la figura V.15 se representa conjuntamente la evolución temporal de la biomasa correspondiente a los tamaños 30, 40, 50 y 60 gramos. Se puede apreciar un ligero desplazamiento hacia la derecha a medida que aumenta el tamaño como consecuencia de la incorporación de individuos desde las tallas inferiores a las superiores. Asimismo, la suma de la biomasa del rango de tamaños comerciales comprendidos entre 130 y 170 gramos, pone de manifiesto algunos procesos de venta (figura V.16).

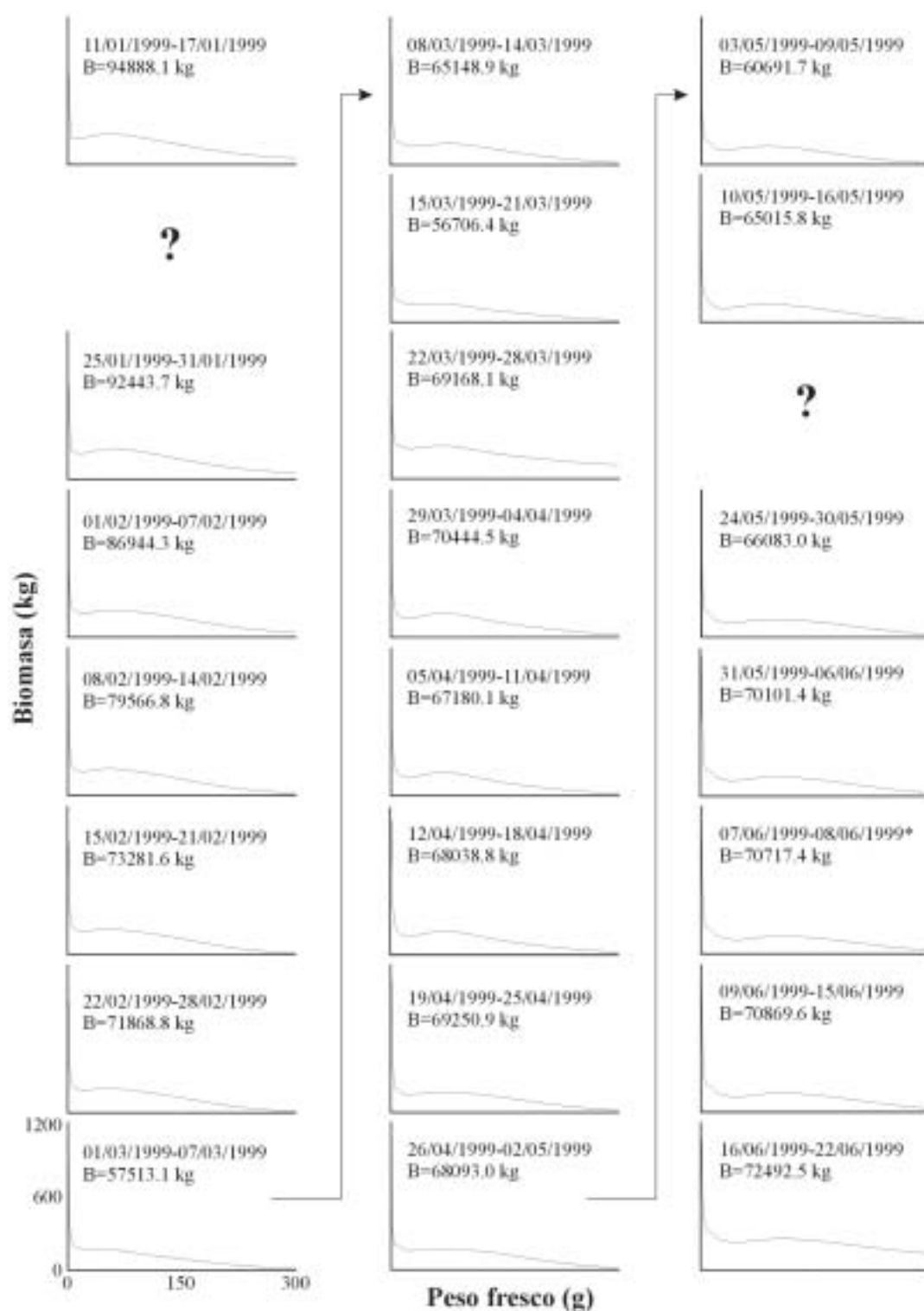


Figura V.9a. Evolución semanal de la biomasa (B) almacenada en la planta en función del rango de variación del peso fresco de un ejemplar, en el periodo comprendido entre el 11/01/1999 hasta el 02/01/2001. Extensión de la varianza mediante el modelo de Arnason-Sparre.  $\sigma_v=0.2$ . \*Cambio de inicio de semana de lunes a miércoles



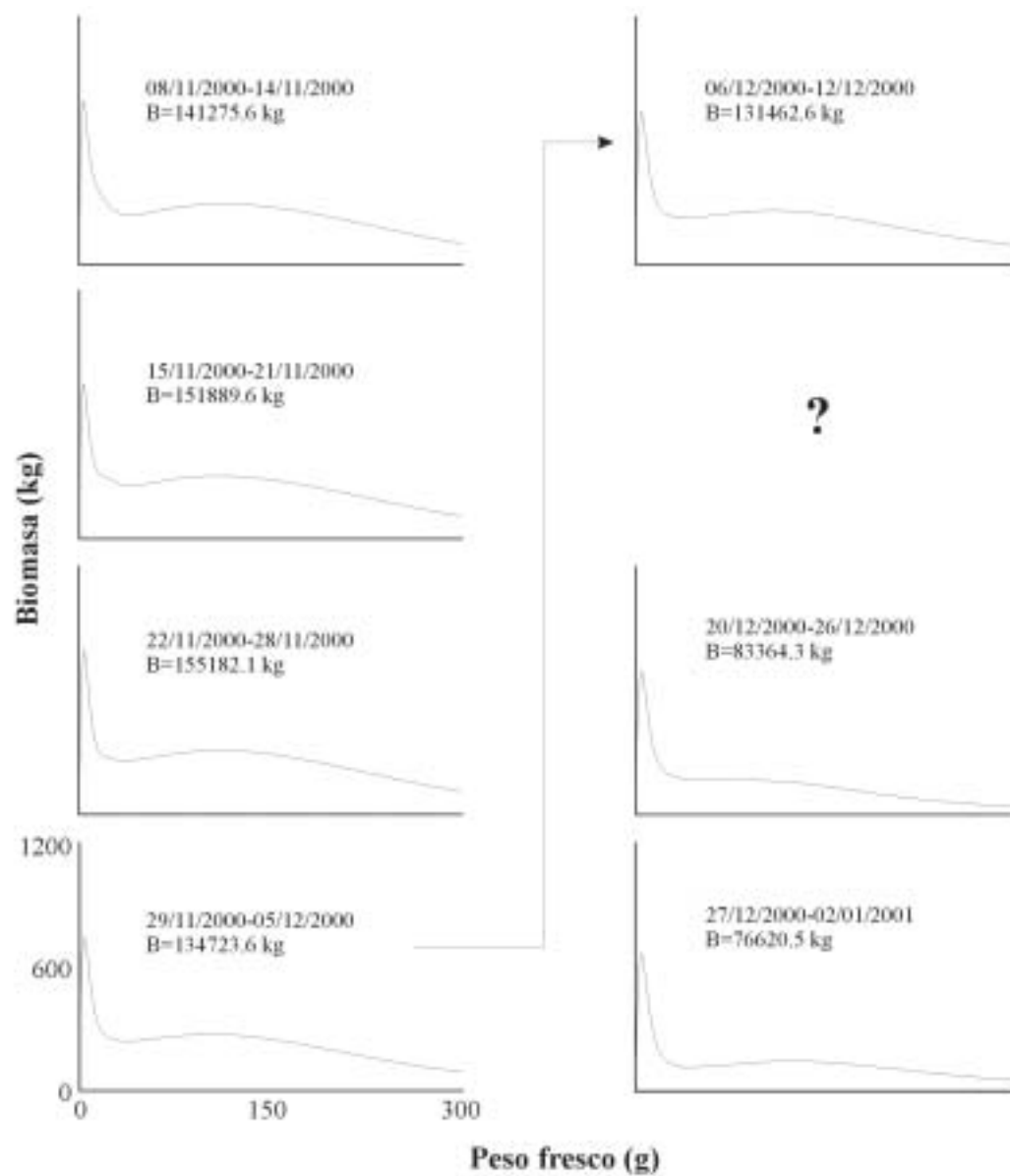
**Figura V.9b.** Evolución semanal de la biomasa (B) almacenada en la planta en función del rango de variación del peso fresco de un ejemplar, en el período comprendido entre el 11/01/1999 hasta el 02/01/2001. Extensión de la varianza mediante el modelo de Arrason-Sparre.  $\sigma_v=0.2$



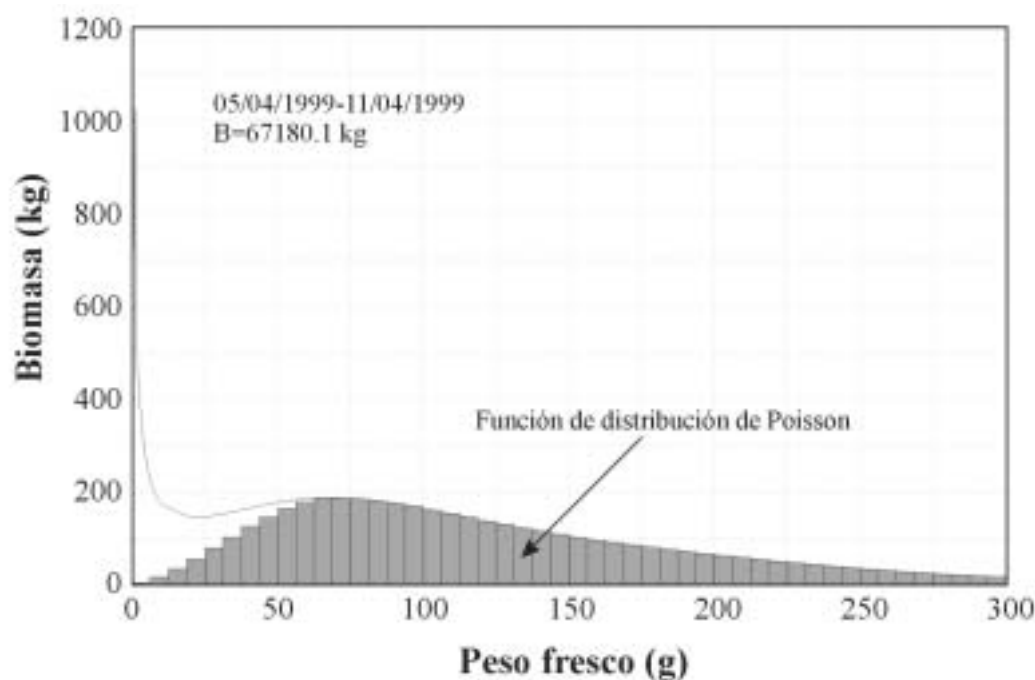
**Figura V.9c.** Evolución semanal de la biomasa (B) almacenada en la planta en función del rango de variación del peso fresco de un ejemplar, en el período comprendido entre el 11/01/1999 hasta el 02/01/2001. Extensión de la varianza mediante el modelo de Arnason-Sparre.  $\sigma_a=0.2$



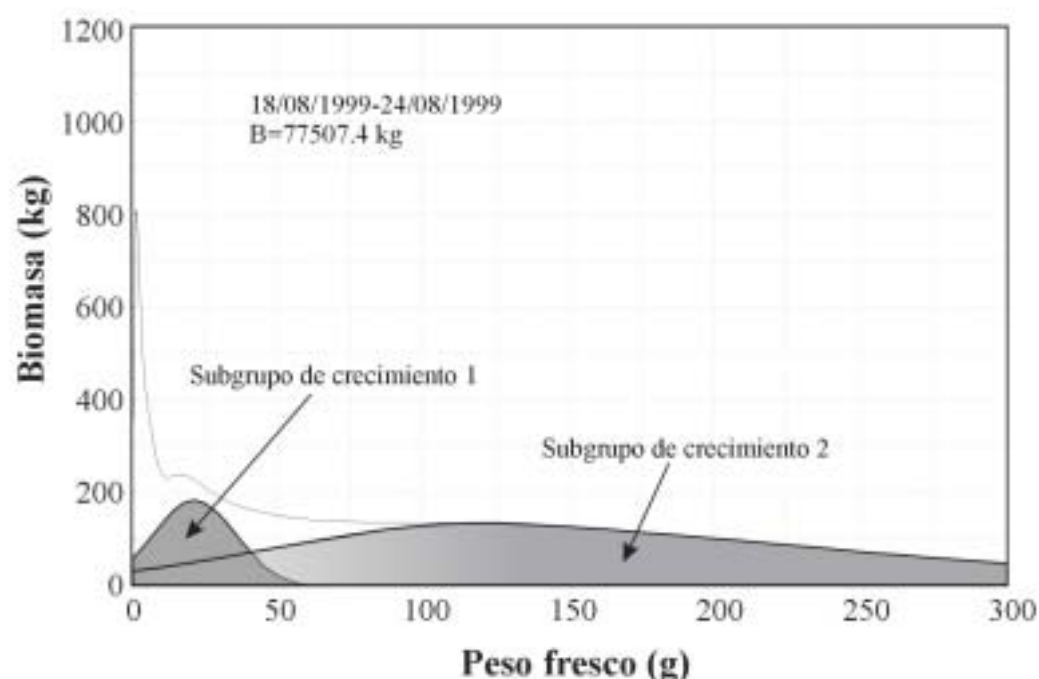
Figura V.9d. Evolución semanal de la biomasa (B) almacenada en la planta en función del rango de variación del peso fresco de un ejemplar, en el periodo comprendido entre el 11/01/1999 hasta el 02/01/2001. Extensión de la varianza mediante el modelo de Arnason-Sparre.  $\sigma_s=0.2$



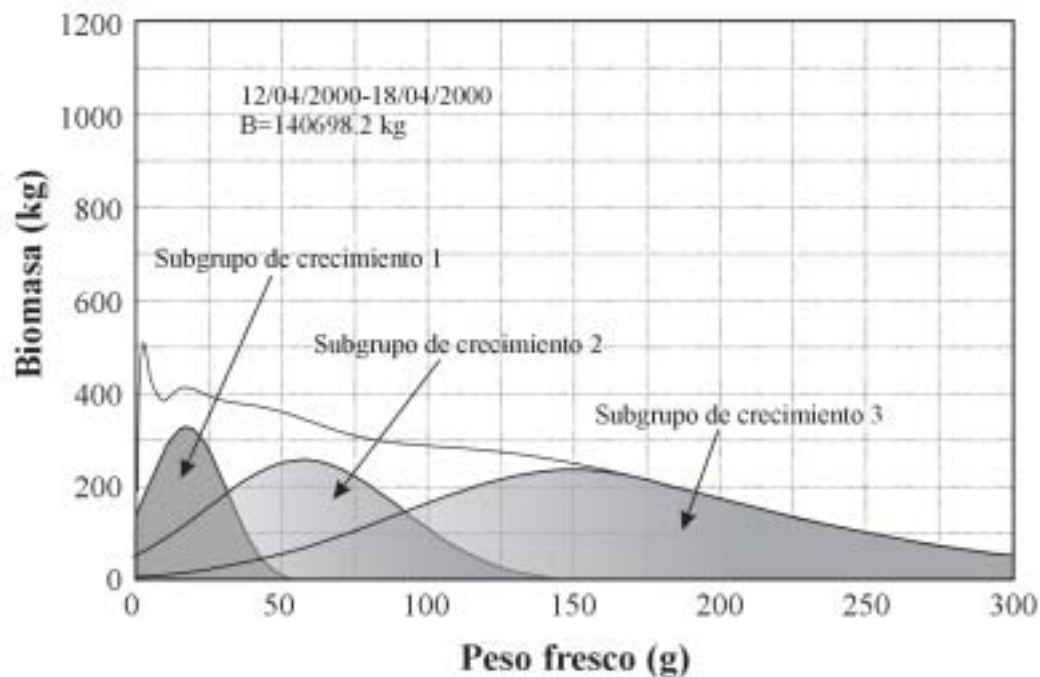
**Figura V.9e.** Evolución semanal de la biomasa (B) almacenada en la planta en función del rango de variación del peso fresco de un ejemplar, en el periodo comprendido entre el 11/01/1999 hasta el 02/01/2001. Extensión de la varianza mediante el modelo de Arnason-Sparre.  $\sigma_s=0.2$



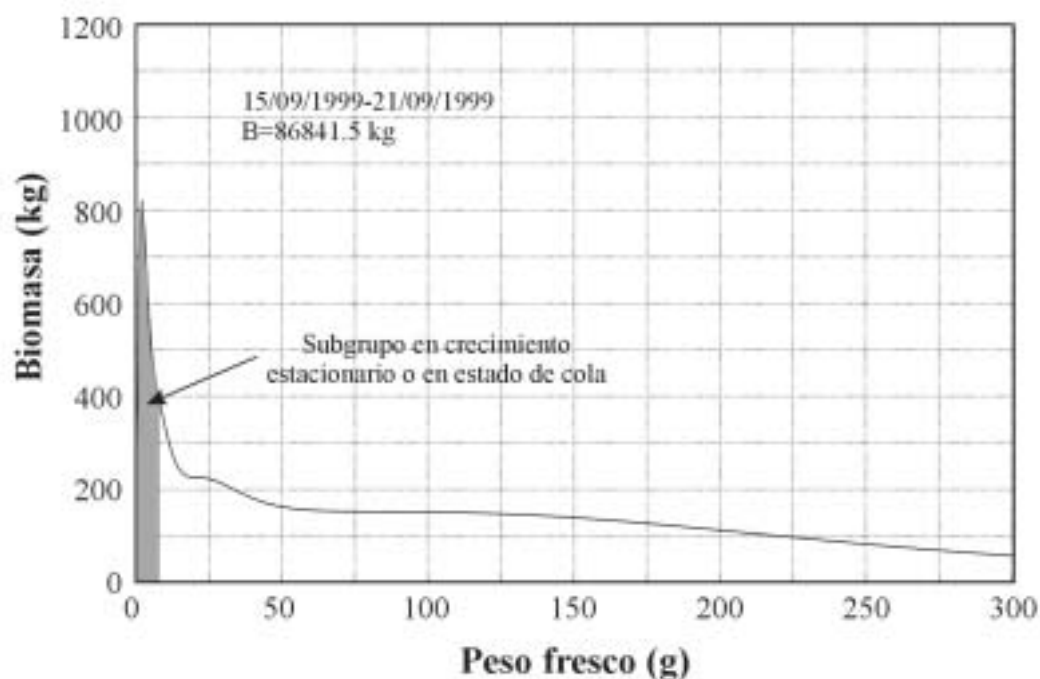
**Figura V.10.** Distribución de la biomasa total ( $B$ ) en la semana del 5 de abril de 1999 al 11 de abril de 1999. El histograma representan la función de distribución de Poisson asociada a la mayor parte de la población.



**Figura V.11.** Distribución de la biomasa total ( $B$ ) de tipo bimodal en la que se pueden identificar al menos dos grupos de crecimiento distintos. Se ha utilizado como referencia la distribución de la biomasa de la semana del 18 de agosto de 1999 al 24 de agosto de 1999.

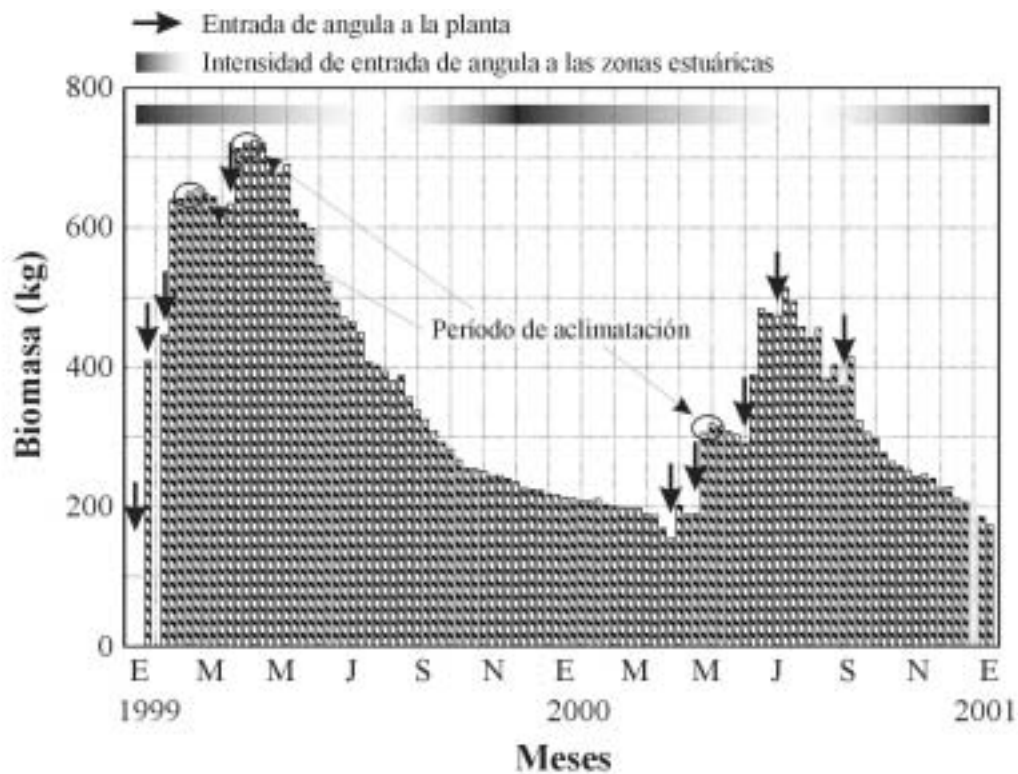


**Figura V.12.** Distribución de la biomasa total (B) de tipo trimodal en la que se pueden identificar al menos tres grupos de crecimiento distintos. Se ha utilizado como referencia la distribución de la biomasa de la semana del 12 de abril de 2000 al 18 de abril de 2000

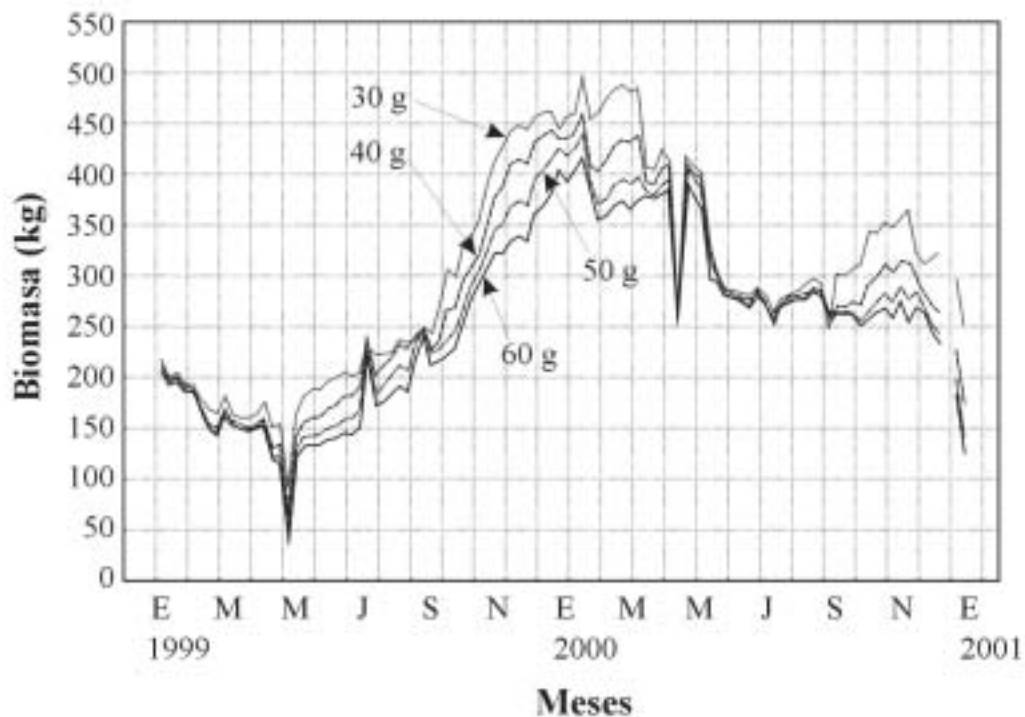


**Figura V.13.** Distribución de la parte de la población que permanece en crecimiento estacionario o estado de cola. Se ha utilizado como referencia la distribución de la biomasa de la semana del 15 de septiembre de 1999 al 21 de septiembre de 1999

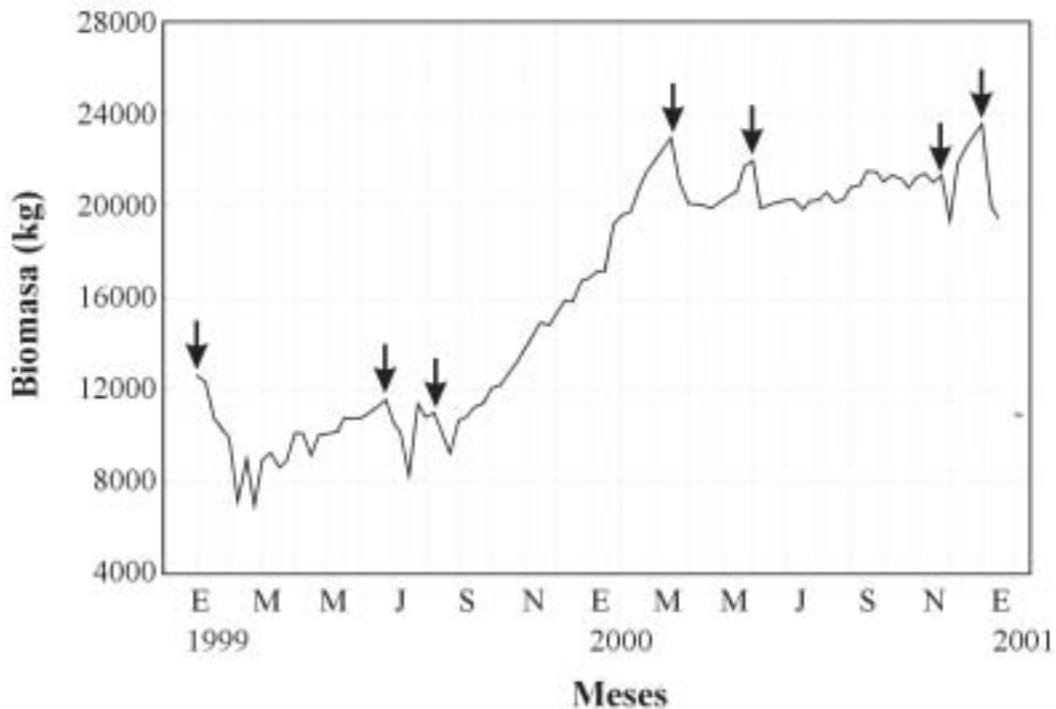




**Figura V.14.** Evolución a lo largo del periodo de estudio de la biomasa para los tamaños menores o igual a 0.5 gramos. Las flechas verticales indican la entrada de angulas a la planta mientras que el degradado de color negro superior indica la intensidad de entrada de angula a las zonas estuáricas y, por tanto, la intensidad de su captura



**Figura V.15.** Evolución temporal de la biomasa correspondiente a los tamaños 30, 40, 50 y 60 gramos



**Figura V.16.** Evolución temporal de la suma de biomasa del rango de tamaños comprendido entre 130 y 170 gramos. Las flechas verticales indican algunos periodos de ventas

### 5.3.5. Caracterización de la mortalidad

La figura V.17 muestra la mortalidad semanal contabilizada a lo largo del período de estudio. Como se puede observar, existen períodos en los que la mortalidad máxima se mantiene por debajo de 1000 kg (prácticamente todo el año 1999) y otros donde las bajas alcanzan cotas cercanas a los 5000 kg (año 2000). Un análisis de la varianza mostró marcadas diferencias significativas entre las mortalidades medias mensuales (ANOVA:  $F=17.359$ ;  $p_{\alpha}<0.001$ ), haciéndose patente un incremento significativo desde el mes de enero de 2000 hasta junio del mismo año, para a partir de ese momento descender hasta niveles semejantes a los encontrados durante todos los meses del año 1999 (test de Tukey:  $p_{\alpha}<0.001$  en todos los casos) (figura V.18). Anualmente, se produjo un incremento significativo de la mortalidad en el año 2000 (test t de Student:  $t(103)=-7.010$ ;  $p_{\alpha}<0.001$ ). Así, de forma general, la tasa de mortalidad diaria media estimada haciendo uso de la ecuación (5.9) fue de un 0.03 % aunque puntualmente se encontraron cercanos al 0.5 %.

Igual que ocurría con la biomasa, en la distribución semanal de la mortalidad por tallas se pueden identificar modas de mortalidad correspondientes a los distintos subgrupos de crecimiento, tal y como predicen McNown y Seireg (1983). Sin embargo, en esta ocasión los patrones encontrados presentan una configuración más variada que en el caso de la biomasa viva, lo que puede deberse a la distinta naturaleza de los datos.

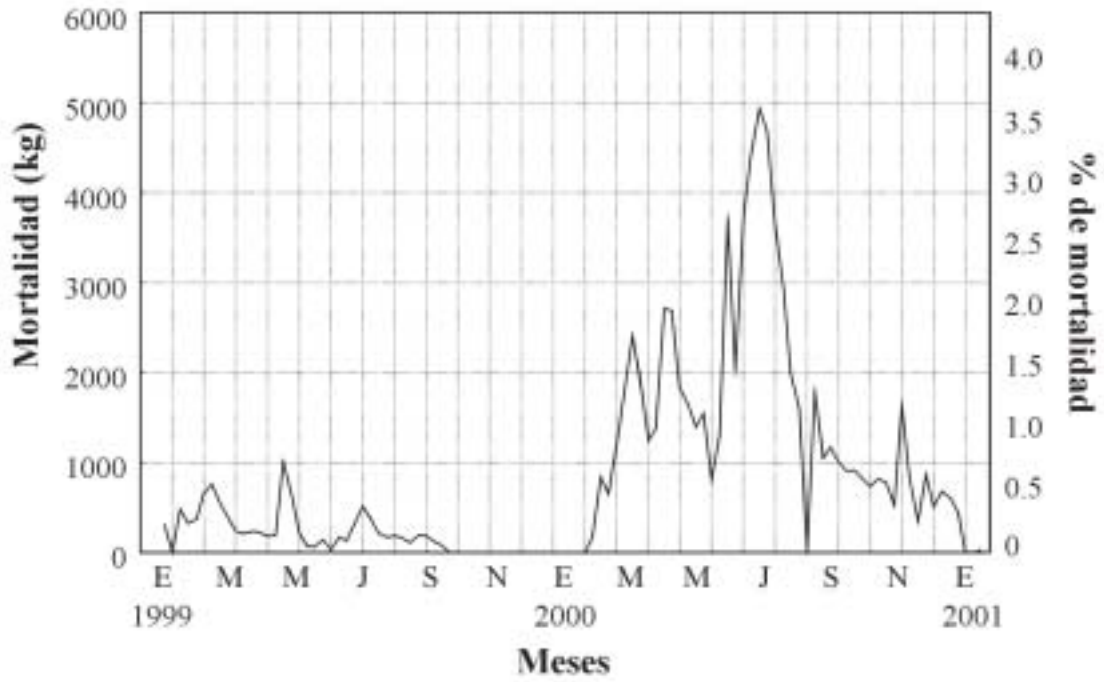


Figura V.17. Mortalidad semanal absoluta y relativa durante el periodo comprendido entre el 11/02/1999 y el 02/01/2001

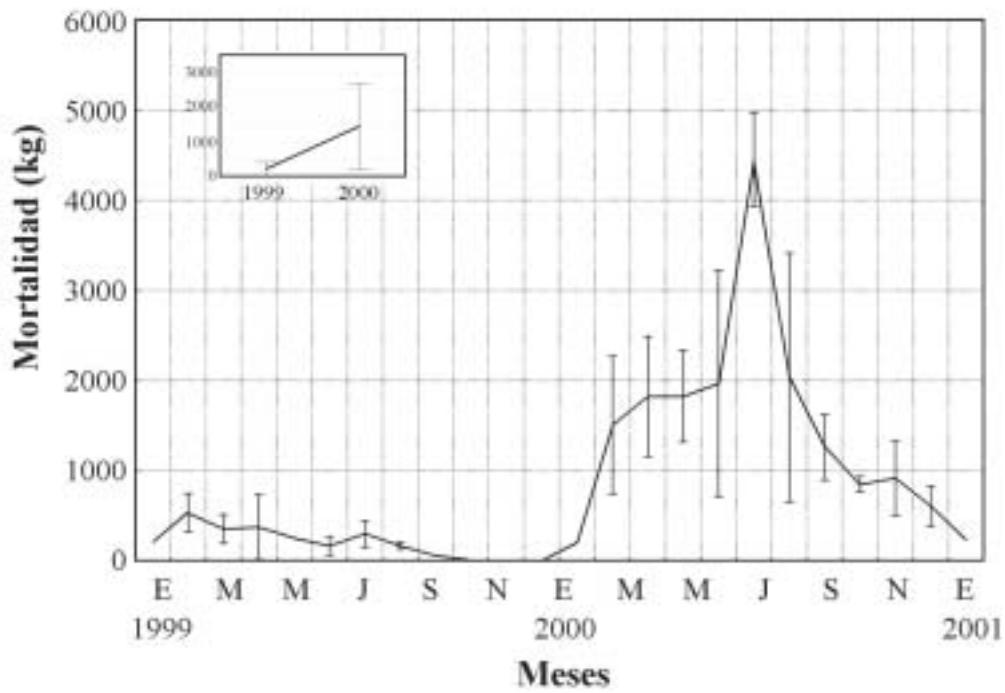
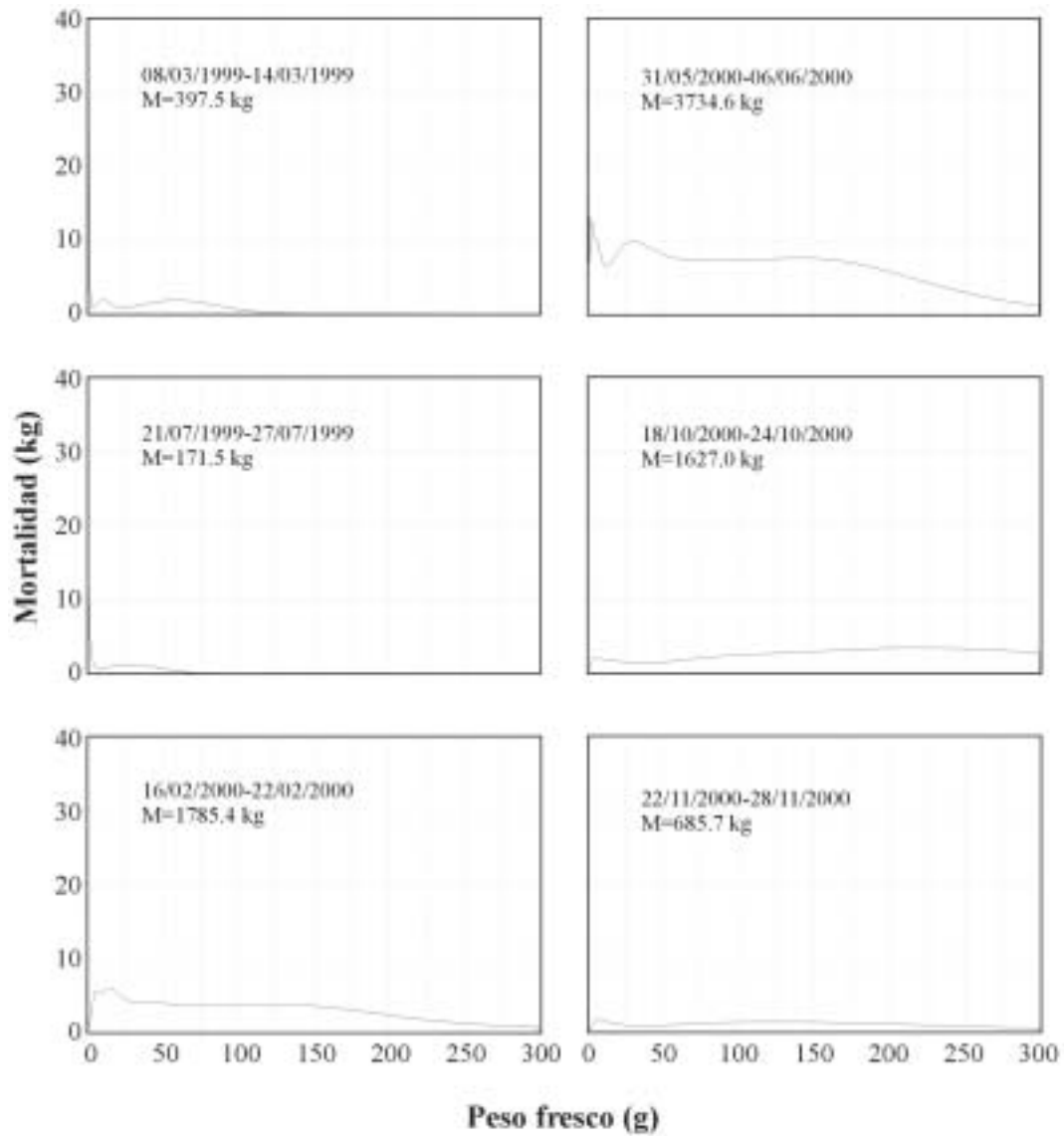


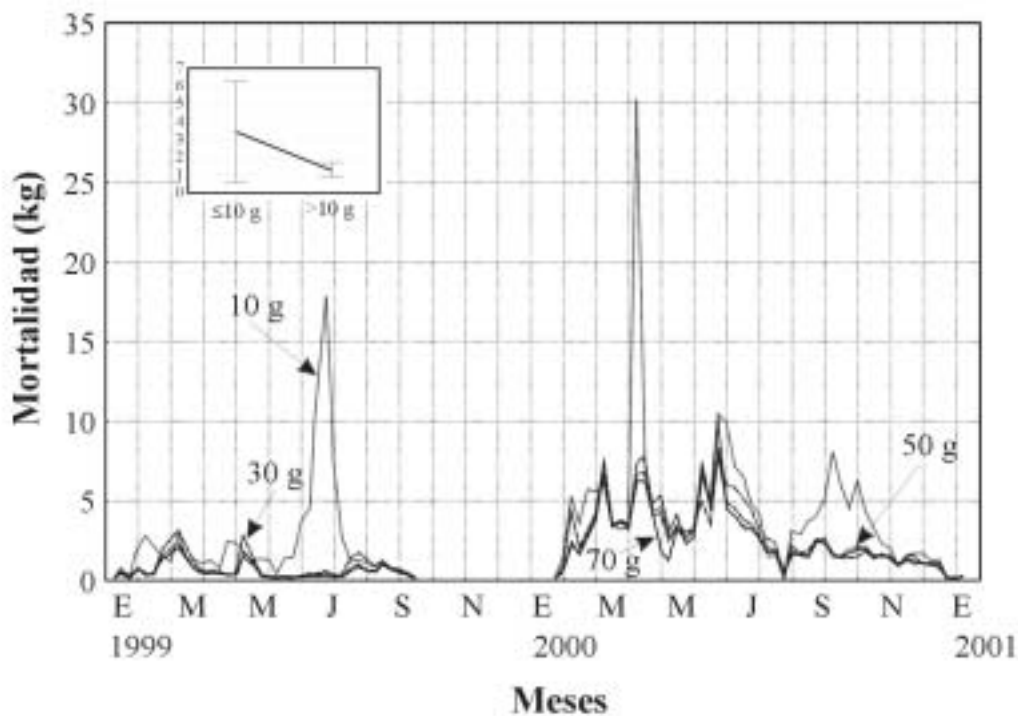
Figura V.18. Mortalidad mensual y anual media durante el periodo comprendido entre el 11/02/1999 y el 02/01/2001

Así, al ser los datos de mortalidad de tipo diario, puede que reflejen con mayor exactitud el número de subgrupos de crecimiento coexistentes en la planta. A título de ejemplo, en la figura V.19 se muestran algunos de los patrones encontrados.



**Figura V.19.** Seis patrones distintos de mortalidad semanal (M) en la planta en función del rango de variación del peso de un ejemplar. Extensión de la varianza mediante el modelo de Arnason-Sparre.  $\sigma_e=0.2$

Al considerar la evolución de la mortalidad por tamaños, encontramos que aproximadamente un 98 % de los rangos de tallas consideradas presentan un patrón similar de distribución, mientras que el resto muestra un patrón distinto o con mortalidades puntualmente acusadas. Así por ejemplo, en la figura V.20 se puede observar la evolución de las mortalidades para los tamaños 10, 30, 50 y 70 gramos. En los individuos de 10 gramos encontramos tres picos de mortalidad acusada a finales de junio de 1999, finales de marzo-principios de abril de 2000 y durante el mes de septiembre de 2000. En los tamaños 30, 50 y 70 gramos no se aprecian dichos picos de mortalidad o al menos son menos importantes. La presencia de momentos puntuales en los que la mortalidad es alta es muy común en individuos por debajo de 10 gramos lo que incrementa significativamente la media de mortalidad a lo largo de todo el período de estudio (test t de Student:  $t(598)=14.449$ ;  $p_{\alpha}<0.001$ ). Por el contrario, a partir de 10 gramos en adelante, todas las tallas presentan un comportamiento similar con la ausencia de mortalidades acusadas. Esto puede interpretarse como una mayor sensibilidad de los individuos menores a 10 gramos a situaciones estresantes tales como las clasificaciones, cambios en la dieta, etc.



**Figura V.20.** Evolución temporal de la mortalidad correspondiente a los tamaños 10, 30, 50 y 70 gramos. En la gráfica superior izquierda se muestra la mortalidad media para el conjunto de individuos menores o iguales a 10 gramos y mayores a 10 gramos

### 5.3.6. Análisis de los resultados obtenidos con el Programa Operativo de simulación (POS)

Los resultados de las simulaciones llevadas a cabo con **POS** fueron comparadas entre sí y con los datos reales recogidos durante las campañas de 1999 y 2000. Para ello se parte

de una serie de simulaciones de referencia en la que tan solo se consideran las condiciones básicas para que el programa funcione. Las condiciones iniciales de simulación son variadas y los parámetros físico-químicos son incorporados, produciéndose por tanto simulaciones más complejas.

### 5.3.6.1. Simulaciones de referencia

Las condiciones iniciales de las dos simulaciones de referencia llevadas a cabo se muestran en la tabla V.6 . Las únicas diferencias entre las dos simulaciones radica en el número inicial de angulas (1000000 ó 3000000). El resto de las condiciones se mantiene constante.

**Tabla V.6.** Valores de las condiciones iniciales de las dos simulaciones de referencia

Parámetros	Simulación 1	Simulación 2
Número inicial de angulas	1000000	3000000
Peso medio inicial (g)	0.3	0.3
Desviación estándar inicial (g)	0.2	0.2
% inicial en crecimiento lento	44	44
% inicial en crecimiento normal	10	10
% inicial en crecimiento rápido	6	6
% restante (considerado como..)	30 (Mort.)	30 (Mort.)
FCR lento	2.0	2.0
FCR normal	1.6	1.6
FCR rápido	1.2	1.2
Tasa de mortalidad diaria	0.0003	0.0003
Tiempo de simulación (días)	730	730
Mes de inicio	Octubre	Octubre
Parámetros F-Q	NO	NO
Peso medio de cosecha (g)	150	150
Lote mínimo de cosecha (kg)	2000	2000
Número de cohortes	1	1
Zona de alevinaje hasta (g)	40	40
Zona de preengorde hasta (g)	120	120

Por otra parte, en la tabla V.7 se muestran los valores de los parámetros económicos. En este caso, todos los parámetros se han mantenido constantes para las dos simulaciones, con la única particularidad de que no se han contabilizado los costes de amortización.

**Tabla V.7. Valores de los parámetros económicos para las dos simulaciones de referencia**

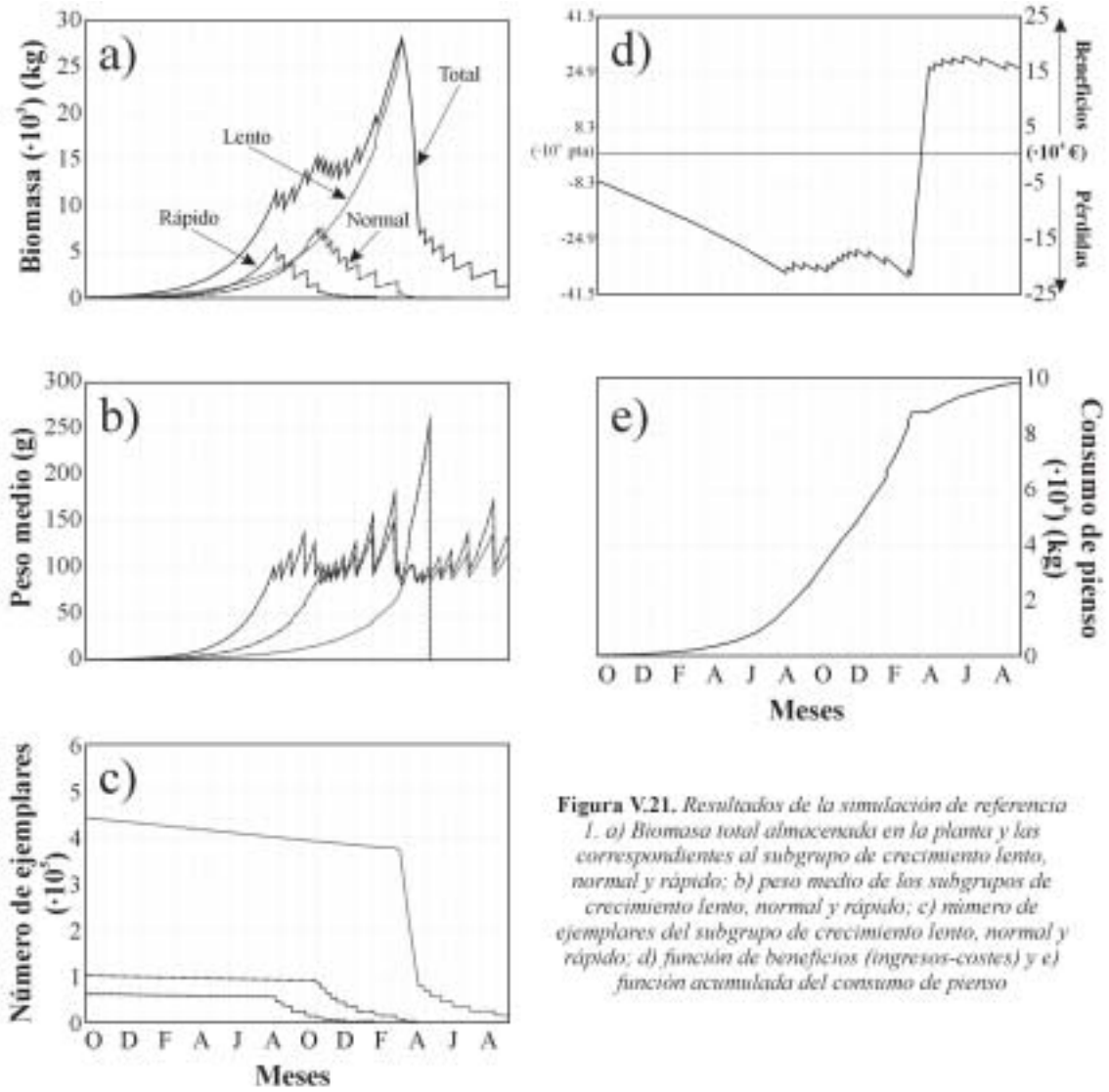
Parámetros	Para las dos simulaciones de referencia
Número de técnicos	2
Salario bruto por técnico (€)	2200
Número de operarios	2
Salario bruto por operario (€)	1085
€/kg de angula	150.6
kWh/mes de consumo	71489
€/kWh	0.051
Considerar amortización	NO

### 5.3.6.2. Simulación de referencia 1

Los resultados de la simulación de referencia 1 se presentan en la figura V.21. En la figura V.21a se muestra la evolución temporal de la biomasa total y las correspondientes de los subgrupos de crecimiento lento, normal y rápido, de una única cohorte a lo largo de un período de dos años. Observando la función de biomasa total, el primer lote de venta se obtiene a mediados del mes de agosto del primer año. Es decir, si se considera que la cohorte se introduce a principios de mes de octubre, tan solo transcurren 11 meses hasta que se lleva a cabo la primera venta de 2000 kg. Esto supone que, en condiciones ideales el tamaño comercial se alcanza unos 7 meses antes de lo que se alcanza en condiciones de explotación. A partir de ese momento, se producen continuas oscilaciones en la biomasa total almacenada, aunque con un aumento prácticamente constante, hasta llegar a marzo del segundo año donde la biomasa total alcanza un máximo algo superior a las 28 toneladas. Desde ahí hasta mediados del mes de agosto se produce un descenso continuo de la biomasa, que a partir de ese momento oscila nuevamente para hacerse prácticamente cero a finales de septiembre del segundo año.

Una consecuencia importante de esta primera simulación la encontramos en la máxima biomasa almacenada (28 tn). Este nivel se encuentra muy por debajo del máximo registrado en condiciones de explotación ( $\approx 160$  tn), lo que puede estar indicando que el número de ejemplares del reclutamiento inicial debe ser mayor.

También de forma global, se puede observar como la mayor contribución a la biomasa total corresponde al subgrupo de crecimiento lento, mientras que la menor aportación es llevada a cabo por el subgrupo de crecimiento rápido. Es decir, éste último es el que en primera instancia aporta individuos para la venta (a los 11 meses), pero al mismo tiempo es el subgrupo que presenta un menor potencial de aportación (aproximadamente unas 5 tn). Teniendo en cuenta tan solo la biomasa correspondiente al subgrupo de crecimiento lento, se puede observar como las primeras aportaciones para la venta se producen durante el mes de marzo del segundo año (17 meses), lo que prácticamente coincide con lo indicado por varios autores como el período mínimo que tiene que transcurrir para que la angula alcance el tamaño comercial de 150 gramos.



**Figura V.21.** Resultados de la simulación de referencia 1. a) Biomasa total almacenada en la planta y las correspondientes al subgrupo de crecimiento lento, normal y rápido; b) peso medio de los subgrupos de crecimiento lento, normal y rápido; c) número de ejemplares del subgrupo de crecimiento lento, normal y rápido; d) función de beneficios (ingresos-costes) y e) función acumulada del consumo de pienso



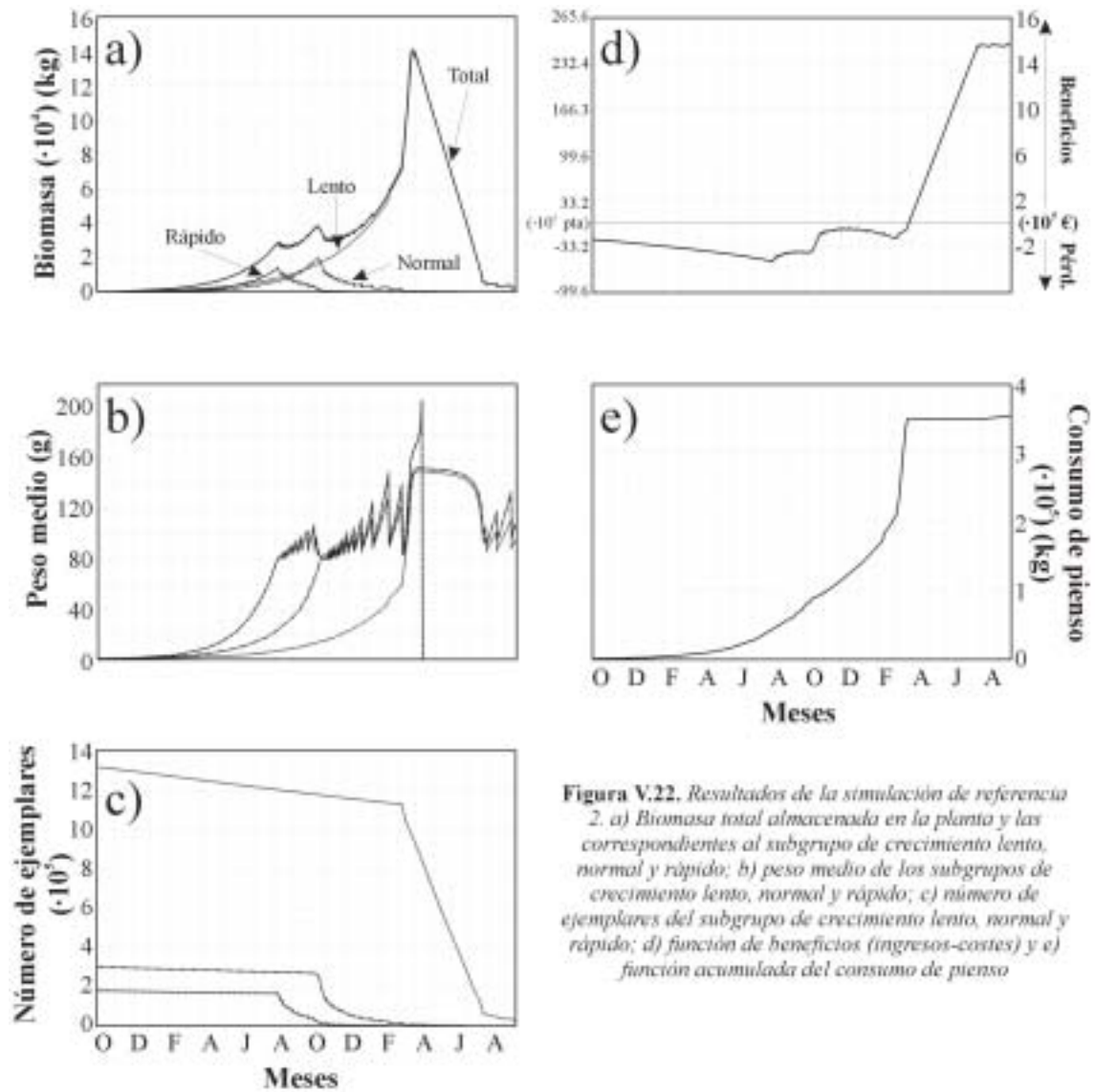
La disminución del número de individuos como consecuencia del proceso de venta, se refleja tanto en las curvas de evolución del peso medio de cada subgrupo de crecimiento como en las funciones de evolución del número de ejemplares almacenados en la planta. Así en el primer caso (figura V.21b) se puede ver como el peso medio de cada subgrupo de crecimiento oscila al mismo tiempo que se producen las ventas. Esto es debido al reajuste de la media del subgrupo en función del número de ejemplares restantes. También es interesante comprobar como a pesar de que se producen continuas ventas, el peso medio del subgrupo de crecimiento no rebasa el peso de cosecha (150 g) excepto en cuatro ocasiones. Éstas coinciden con el agotamiento de los subgrupos de crecimiento rápido y normal. Por otra parte, como es lógico, las ventas hacen disminuir el número de ejemplares almacenados, lo que se refleja en la transformación de la disminución lineal en una disminución en forma de escalón (figura V.21c).

La función de beneficios (figura V.21d) también muestra oscilaciones como consecuencia de los ingresos producidos por las ventas. Estas oscilaciones no se presentan hasta el mes de agosto del primer año que es cuando se produce la primera venta. Hasta ese momento y partiendo de unas pérdidas iniciales de unos 50000 € (8.3 millones de pesetas), las pérdidas se incrementan de forma no lineal hasta un nivel de 212000 € (35192000 pta). A partir de este momento se produce una pequeña recuperación como consecuencia de la venta de la totalidad del subgrupo de crecimiento rápido y la mayor parte del subgrupo de crecimiento normal. A pesar de estas ventas, las pérdidas se acumulan hasta alcanzar 213800 € (35490800 pta) en marzo del segundo año. Sin embargo, las ventas correspondientes al subgrupo de crecimiento lento disparan en poco tiempo los ingresos para de esta forma obtener beneficios a partir de principios de abril del segundo año. A partir de ese instante, el volumen de beneficios oscila alrededor de 150000 € (24900000 pta).

La función acumulada del consumo de pienso se muestra en la figura V.21e. Ésta se identifica como una función sigmoideal que presenta su punto de inflexión acusado a mediados del mes de febrero del segundo año y que se hace casi asintótica a nivel de 100000 kg de pienso consumido.

### 5.3.6.3. Simulación de referencia 2

La simulación de referencia 2 se muestra en la figura V.22. A pesar de que los resultados son muy parecidos a los obtenidos en la simulación de referencia 1, se presentan algunas diferencias significativas. En la evolución temporal de la biomasa total se puede comprobar como, de igual forma que en la simulación 1, se produce un aumento casi constante hasta marzo del segundo año unido a las oscilaciones correspondientes a los procesos de venta. Sin embargo, como en esta simulación se parte de un número mayor de angulas (3 millones), la biomasa máxima almacenada alcanza una cota ligeramente superior a las 140 toneladas (figura V.22a). Esta cifra se encuentra muy próxima a la producción máxima alcanzada por Hidrorecusos en todos sus años de funcionamiento (165.5 tn). Por otra parte, la disminución brusca de biomasa a partir de marzo del segundo año se lleva a cabo de forma más gradual. Esto es debido



**Figura V.22.** Resultados de la simulación de referencia 2. a) Biomasa total almacenada en la planta y las correspondientes al subgrupo de crecimiento lento, normal y rápido; b) peso medio de los subgrupos de crecimiento lento, normal y rápido; c) número de ejemplares del subgrupo de crecimiento lento, normal y rápido; d) función de beneficios (ingresos-costes) y e) función acumulada del consumo de pienso

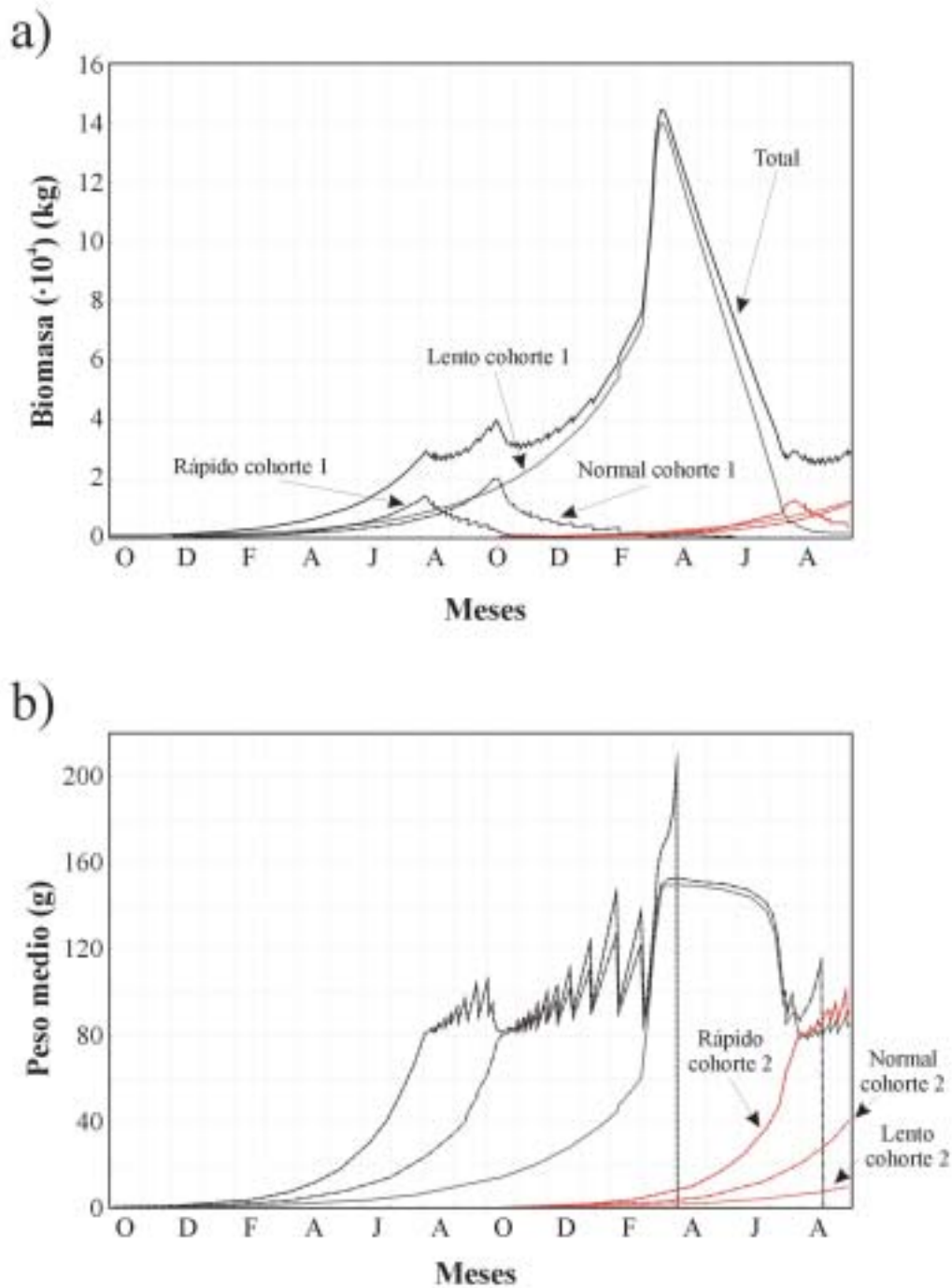
a que la cola derecha de la función de distribución de biomasa del subgrupo de crecimiento lento presenta una cantidad de biomasa muy superior al lote de cosecha. De esta forma, el reajuste de la media de este subgrupo de crecimiento se ve amortiguado hasta que su biomasa disminuye por debajo de las 10 toneladas. Este efecto se puede observar con claridad en las figuras V.22b y V.22c donde se muestra la evolución de los pesos medios y del número de ejemplares de cada subgrupo de crecimiento.

El efecto de la presencia de una mayor cantidad de biomasa se refleja tanto en la función de beneficios como el consumo de pienso. En el primer caso, también se obtienen beneficios a partir de abril del segundo año, pero en esta ocasión estos se elevan hasta 1500000 €(249 millones de pta) (figura V.22d). En cuanto al consumo de pienso, la forma de la curva es muy parecida al caso anterior, aunque cuantitativamente muy superior (figura V.22e). Asimismo, a mediados de marzo el consumo de pienso disminuye drásticamente. Esto es debido al efecto de las ventas y las clasificaciones conjuntamente.

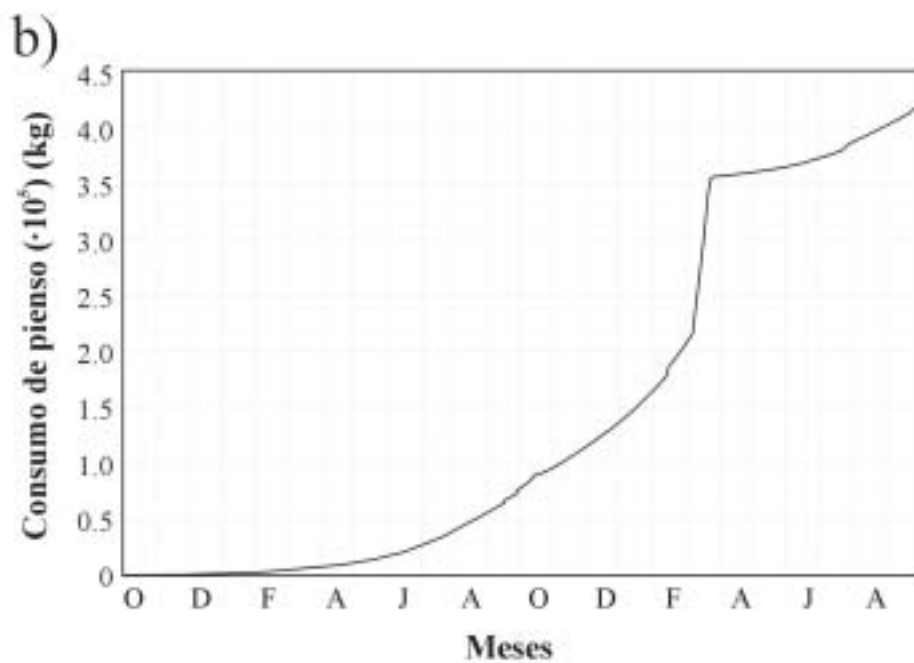
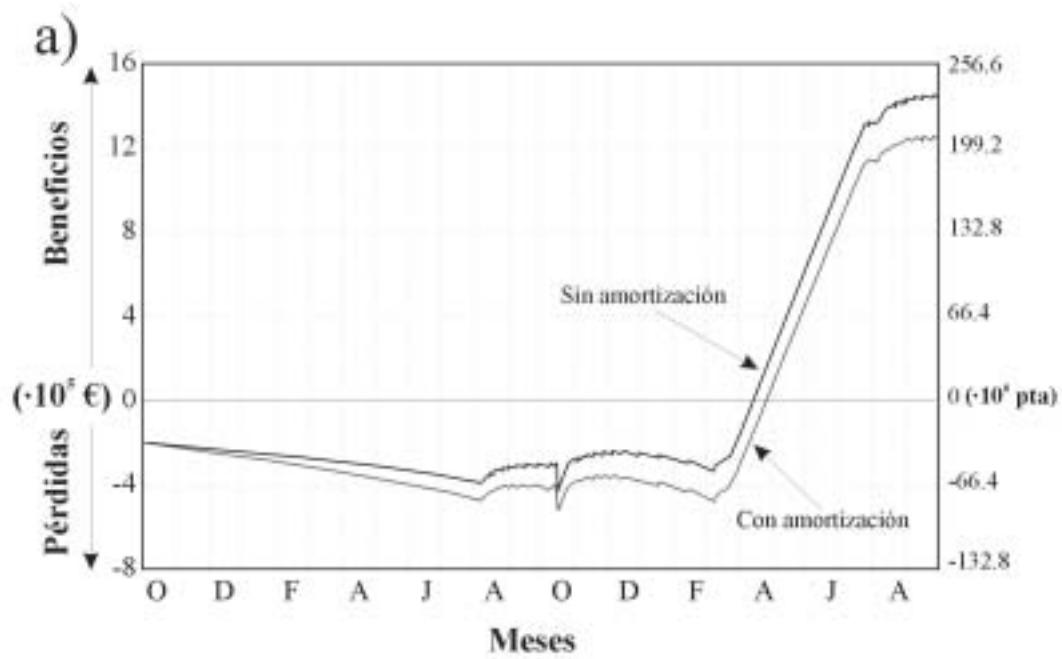
#### *5.3.6.4. Simulación 3. Incorporación de una nueva cohorte*

El efecto de introducir una nueva cohorte en la simulación de referencia 2 a los 15 días del comienzo de la nueva temporada (15 de octubre) sobre la biomasa total se muestra en la figura V.23a. Se puede observar como prácticamente hasta finales de marzo del segundo año la simulación es la misma. Sin embargo a partir de ese momento la caída de la biomasa total se suaviza ligeramente como consecuencia del incremento en peso de los ejemplares correspondientes a la cohorte 2. De esta forma, la biomasa total disminuye hasta una cota cercana a las 30 toneladas a partir de la cual, se inicia su recuperación. Esto es consecuencia de la coincidencia de las dos cohortes en el espacio y en el tiempo, como se puede también comprobar en la evolución de los pesos medios de cada subgrupo (figura V.23b).

También la función de beneficios (figura V.24a) y el consumo de pienso (figura V.24b) se ven modificados al incorporarse una nueva cohorte. En el primer caso, se puede observar con claridad el desembolso debido a la compra de los nuevos ejemplares a mediados del mes de octubre del segundo año, lo que provoca que las pérdidas sean acusadas hasta comienzos del mes de abril del segundo año. Si se considera el coste de amortización, la obtención de beneficios se retrasa aproximadamente un mes. Las diferencias entre los beneficios sin considerar la amortización y considerando la última son más importantes a finales de septiembre del segundo año, debido al agotamiento del subgrupo de crecimiento lento de la primera cohorte. Por otra parte, la función acumulada de consumo de pienso se ve modificada básicamente a partir de mediados de marzo del segundo año, ya que en ese momento el incremento en biomasa de los subgrupos de la cohorte 2 empieza a ser significativo.



**Figura V.23.** Resultados de la simulación número 3. a) Biomasa total almacenada en la planta y las correspondientes al subgrupo de crecimiento lento, normal y rápido de las cohortes 1 y 2; b) peso medio de los subgrupos de crecimiento lento, normal y rápido de las cohortes 1 y 2



**Figura V.24.** Resultados de la simulación número 3. a) Función de beneficios (ingresos-costes) excluyendo e incluyendo la amortización de la inversión inicial; b) función acumulada del consumo de pienso

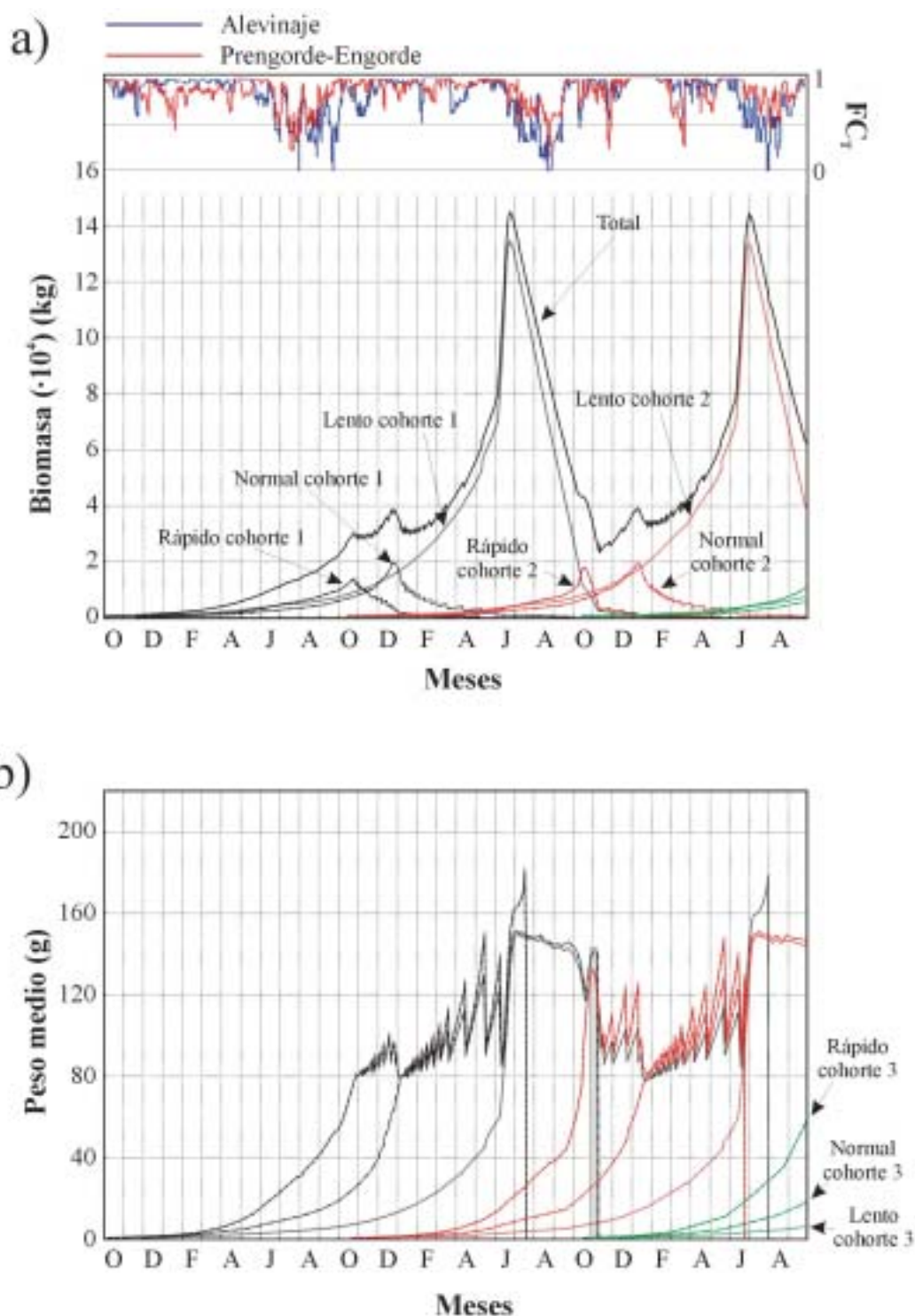
#### 5.3.6.5. *Simulación 4. Tres cohortes, tres años e incorporación del factor de crecimiento debido a la temperatura*

Los resultados de la simulación número 4 se muestran en las figuras V.25 y V.26. En esta nueva simulación se ha aumentado el parámetro tiempo (un año más que en el caso anterior) y se ha incorporado una nueva cohorte (cohorte 3). Asimismo, se ha aplicado el factor de crecimiento debido a la temperatura del agua. Como se puede comprobar existen diferencias importantes con respecto a la simulación anterior. La más significativa es el retraso de los crecimientos en todos los subgrupos. Así, en las biomazas máximas para cada subgrupo, se han desplazado en todos los casos hacia la derecha entre mes y medio y dos meses (figura V.25a). El máximo del subgrupo de crecimiento rápido de la cohorte 1 que se encuentra en la simulación 3 a principios del mes de agosto, se ha desplazado en la simulación 4 hasta mediados del mes de octubre. También se encuentra el mismo desplazamiento en el subgrupo de crecimiento normal y lento.

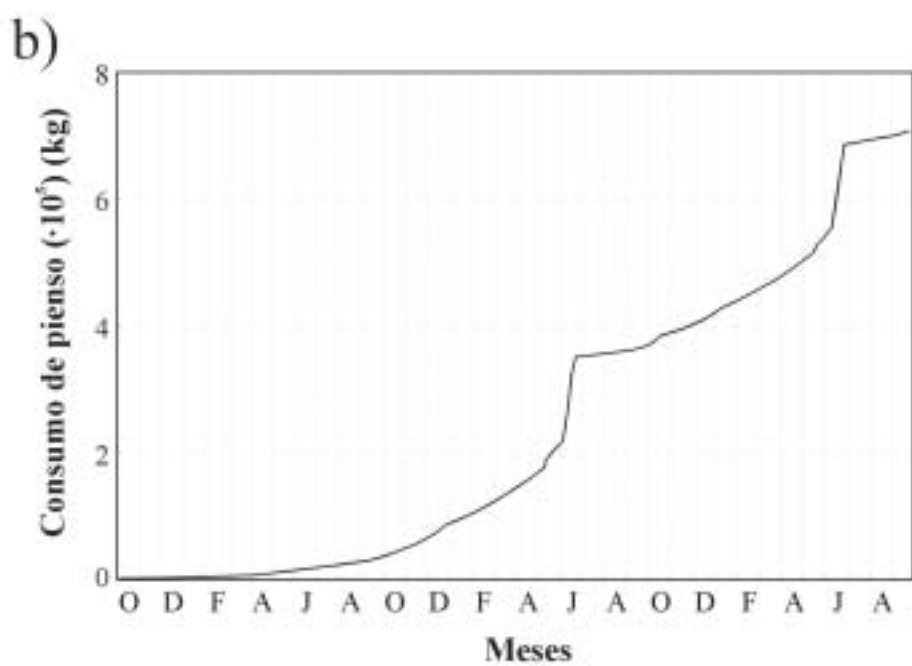
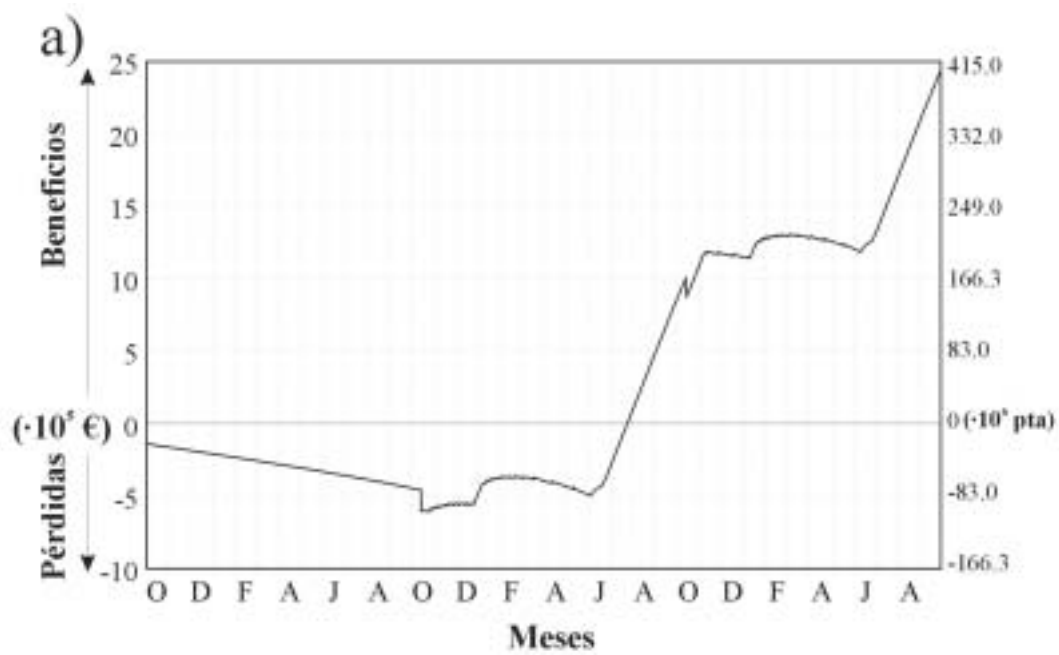
Una consecuencia de este desplazamiento es una mayor coincidencia de subgrupos correspondientes a distintas cohortes. En este sentido, se puede ver con claridad que cuando se incorpora la cohorte 2, el subgrupo de crecimiento rápido de la cohorte 1 aun no ha completado el primer lote de venta de 2000 kg. De esta forma, a mediados de octubre del tercer año llegan a coincidir ejemplares pertenecientes a 8 subgrupos de crecimiento distintos (figura V.25b). Esto es debido a que en ese momento aun podemos encontrar en la planta individuos incluidos en los subgrupos de crecimiento normal y lento de la cohorte 1. De este subgrupo encontramos los últimos ejemplares a principios de agosto del tercer año.

El desplazamiento observado en todos los subgrupos de crecimiento está condicionado por la inclusión del factor de crecimiento debido a la temperatura. Como se puede ver en la parte superior de la figura V.25a, el período de menor crecimiento se corresponde con los meses de junio, julio y agosto, cuando la temperatura del agua es excesivamente alta. Sin embargo, este fenómeno no afecta de igual forma a las distintas partes de la planta. Durante los meses de verano la zona más afectada de forma global es el alevinaje, ya que es donde se registran las mayores temperaturas. En el otoño e invierno ocurre todo lo contrario, es decir, las temperaturas más bajas se registran en preengorde-engorde, con lo que el crecimiento se reduce en esta zona. Por otra parte la inclusión del factor de crecimiento de la temperatura no es suficiente para adaptar la curva de crecimiento simulada a la real, ya que se obtienen ejemplares mayores a 150 gramos aproximadamente entre 6 y 10 meses antes de lo observado por autores como Más-Álvarez y Barrera-Orozco (1989).

En relación a la función de beneficios también encontramos importantes diferencias. Al disminuir los crecimientos para un mismo período de tiempo, en la simulación número 4 no se obtienen beneficios hasta el mes de julio del segundo año, mientras que en la simulación 3 éstos se obtenían tres meses antes. También de forma absoluta los beneficios son menores, ya que a finales de septiembre del segundo año en la simulación 3 se situaban en una cota alrededor de 1500000 €, mientras que en la simulación 4 éstos se encuentran algo por encima de los 820000 €(figura V.26a).



**Figura V.25.** Resultados de la simulación número 4. a) Biomasa total almacenada en la planta y las correspondientes al subgrupo de crecimiento lento, normal y rápido de las cohortes 1, 2 y 3. En la parte superior de la gráfica se muestra asimismo el factor de crecimiento correspondiente a la temperatura del agua en las zonas de alevinaje y preengorde-engorde ( $FC_t$ ); b) peso medio de los subgrupos de crecimiento lento, normal y rápido de las cohortes 1, 2 y 3. La barra de color gris indica el periodo en el cual coinciden un mayor número de subgrupos de crecimiento distintos



**Figura V.26.** Resultados de la simulación número 4. a) Función de beneficios (ingresos-costes) incluyendo la amortización de la inversión inicial; b) función acumulada del consumo de pienso



Por otra parte, la curva de consumo acumulado de pienso (figura V.26b) muestra dos escalones coincidentes con la época de mayor venta (junio del segundo y tercer año), que de igual forma que en los casos anteriores se encuentran desplazados con respecto al observado en la simulación 3 (mes de marzo del segundo año).

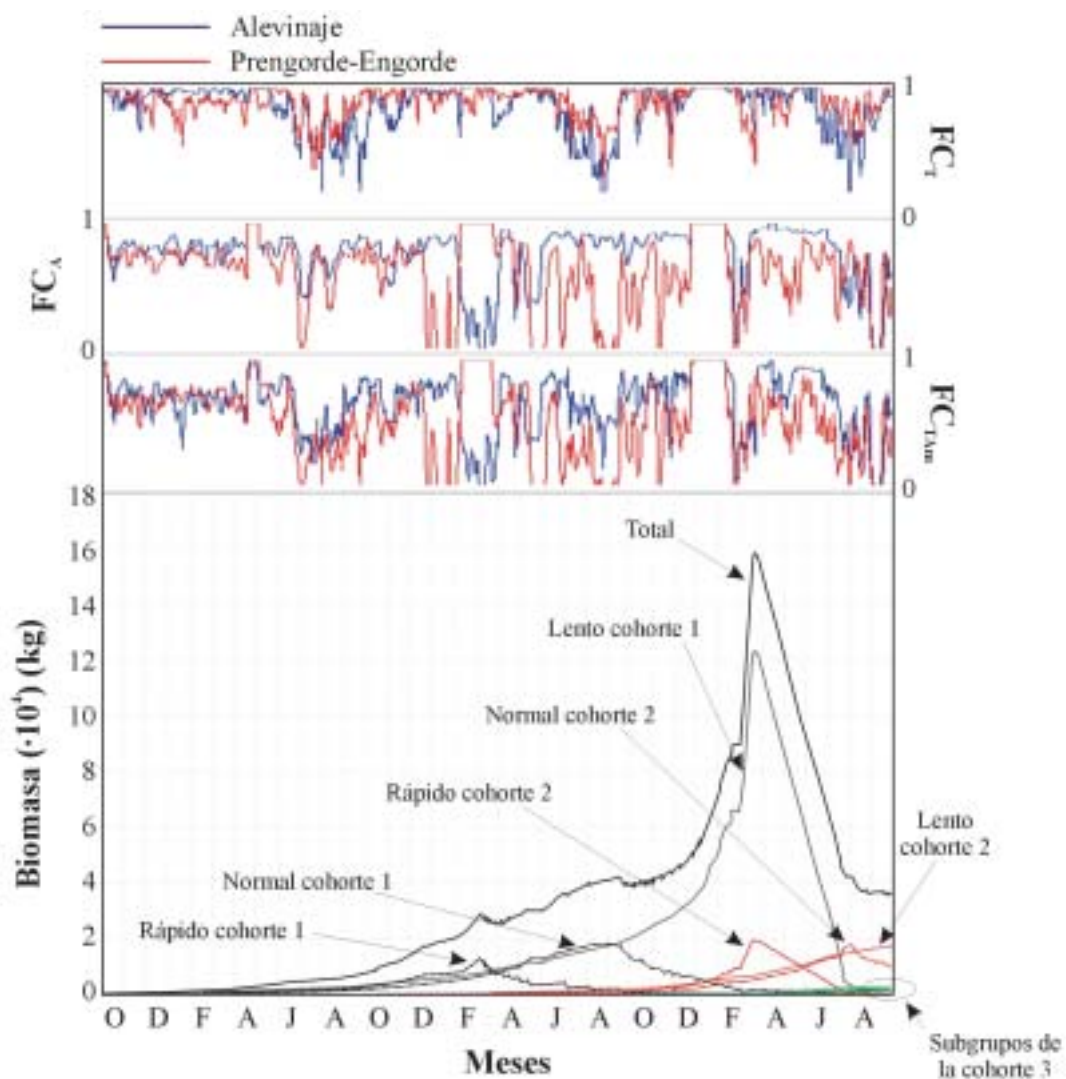
#### 5.3.6.6. *Simulación 5. Incorporación del factor de crecimiento debido a la concentración de amoniaco*

La incorporación del factor de crecimiento debido a la concentración de amoniaco aporta a la simulación un carácter más realista (figura V.27). Por una parte los crecimientos se muestran irregulares a lo largo de las tres temporadas, con fases de mayor y menor crecimiento, y se produce un retraso generalizado de las biomásas máximas con respecto a la simulación en la que tan solo se tenía en cuenta el factor temperatura. De esta forma, las primeras ventas del subgrupo de crecimiento rápido de la cohorte 1 se producen 18 meses después de su introducción en la planta. Por otra parte, la biomasa total almacenada prácticamente alcanza las 160 toneladas. Este se debe a la coincidencia del máximo de biomasa del subgrupo de crecimiento lento de la cohorte 1 con el subgrupo de crecimiento rápido de la cohorte 2. Asimismo, también se produce la coincidencia del máximo de biomasa total almacenada en la simulación con la real en el mes de marzo (figura V.8).

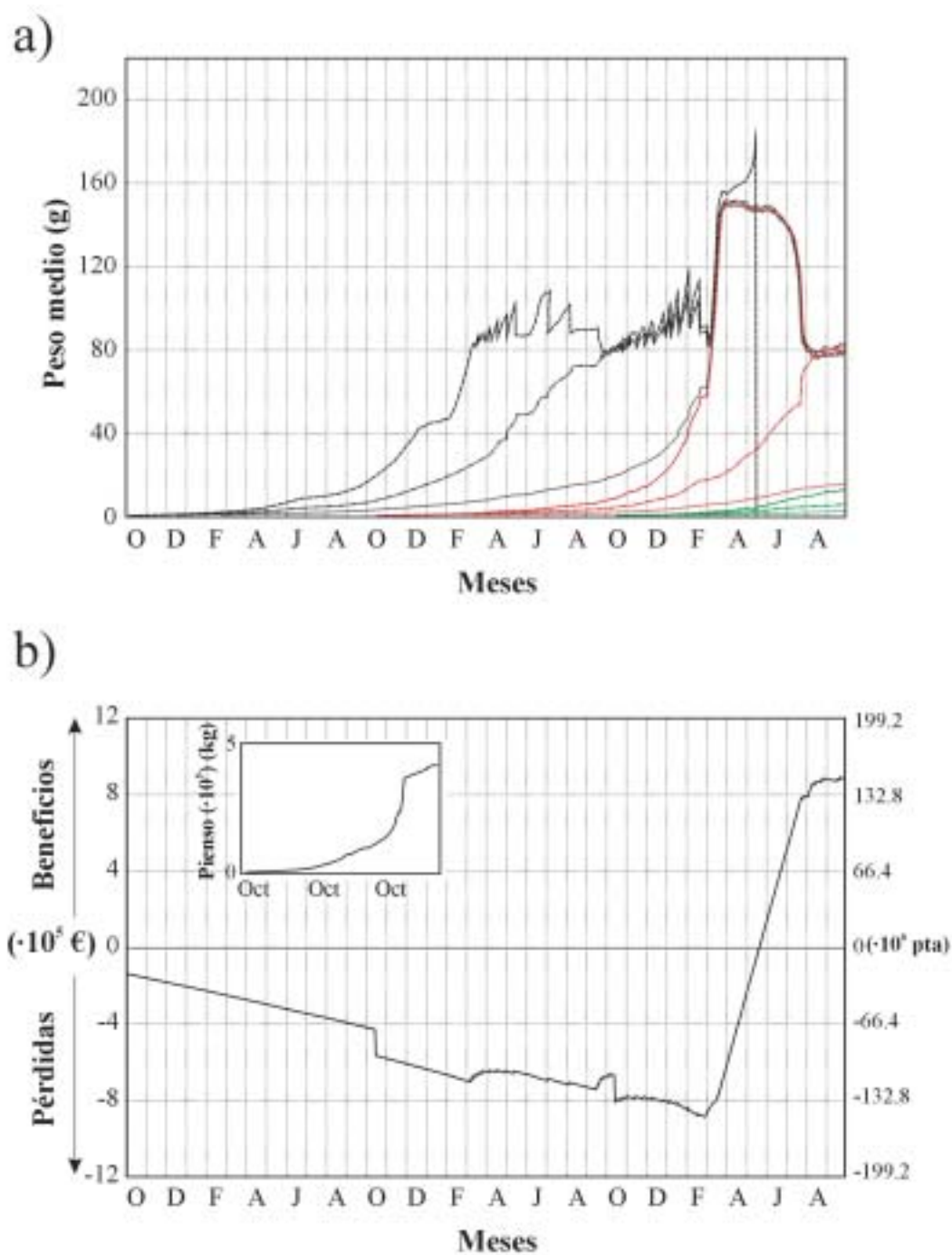
La evolución temporal de los pesos medios de cada subgrupo de crecimiento también reflejan con claridad el irregular crecimiento de cada cohorte (figura V.28a). El desplazamiento en el crecimiento provocado por la inclusión del parámetro amoniaco permite la coincidencia de los 9 subgrupos de crecimiento bien entrado el tercer año (mes de abril), donde la cohorte 3 presenta un crecimiento de al menos 7 meses. Por otra parte la disminución del crecimiento incide directamente sobre la función de beneficios, de tal forma que en la simulación 5 se registran pérdidas hasta el mes de mayo del tercer año por un valor cercano a los 900000 €. Sin embargo, una venta continuada desde marzo hasta julio del tercer año de un total de 120 toneladas proporciona un saldo positivo a mediados del mes de mayo del tercer año (figura V.28b). Así a finales de la tercera temporada los beneficios se elevan hasta cerca de 900000 €. El consumo de pienso también se ve afectado (figura V.28b). A diferencia de lo que se podía observar en la simulación 4, tan solo se identifica un escalón en función del consumo de pienso acumulado, que coincide como en el caso anterior con la venta masiva del tercer año.

#### 5.3.6.7. *Simulación 6. Incorporación del factor de crecimiento debido a la concentración de nitritos*

En esta nueva simulación se vuelve a repetir la pauta encontrada en las simulaciones 4 y 5. Es decir, la incorporación de un nuevo parámetro químico (en esta ocasión la concentración de nitritos) incide significativamente sobre la velocidad en los crecimientos de las distintas cohortes implicadas (figura V.29). De esta forma, el máximo de biomasa total así como los máximos de todos los subgrupos de crecimiento



**Figura V.27.** Resultados de la simulación número 5. Biomasa total almacenada en la planta y las correspondientes al subgrupo de crecimiento lento, normal y rápido de las cohortes 1, 2 y 3. En la parte superior de la gráfica se muestran los factores de crecimiento correspondientes a la temperatura del agua, a la concentración de amoníaco y la combinación de ambos en las zonas de alevinaje y prengorde-engorde ( $FC$ ,  $FC_{sum}$  y  $FC_{tot}$ )



**Figura V.28.** Resultados de la simulación número 5. a) Peso medio de los subgrupos de crecimiento lento, normal y rápido de las cohortes 1, 2 y 3; b) función de beneficios (ingresos-costes) incluyendo la amortización de la inversión inicial. En esta misma gráfica se incluye la función acumulada del consumo de pienso

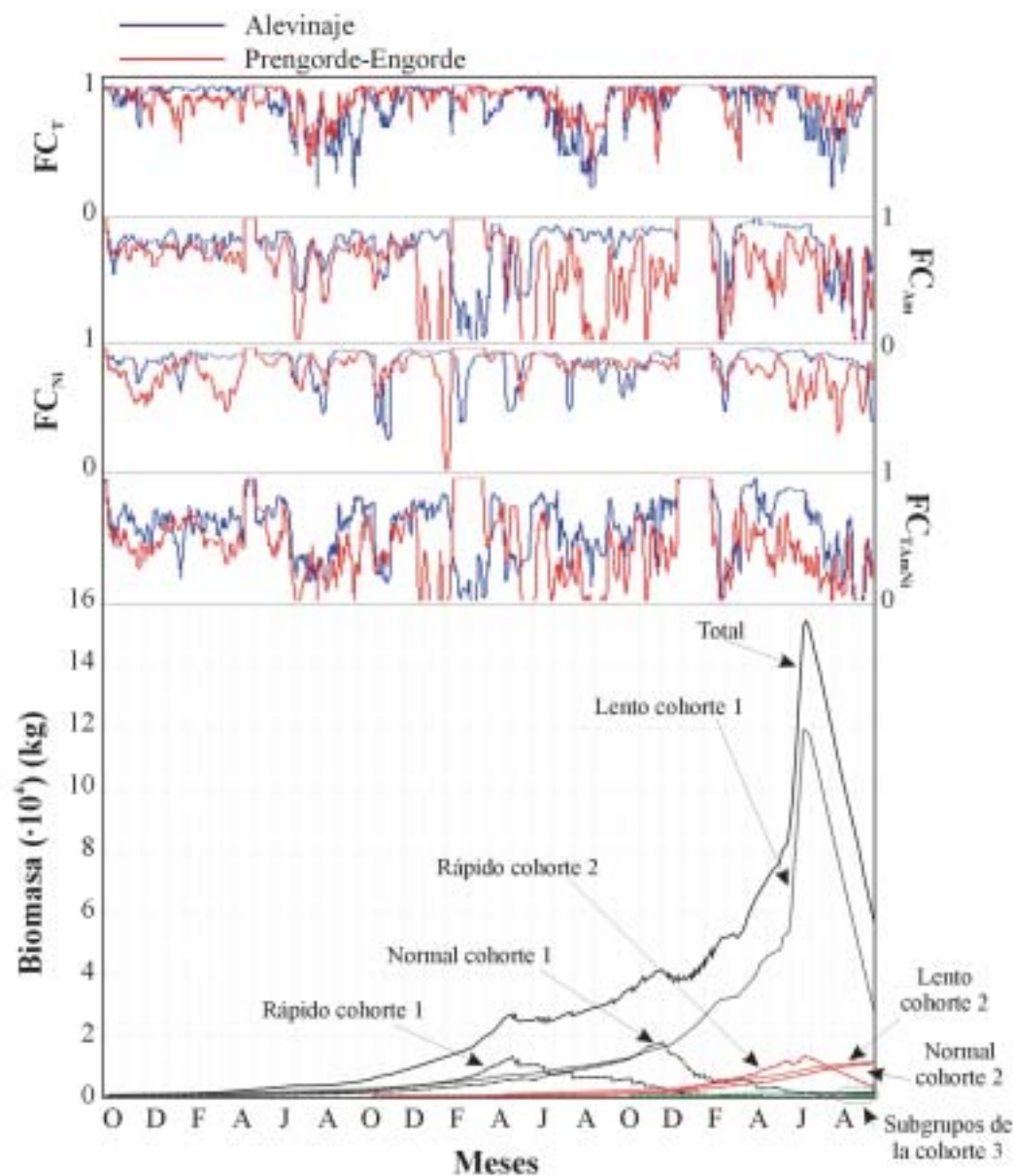
se vuelven a retrasar entre 1 y 3 meses. Por otra parte, aunque la biomasa máxima alcanzada durante la simulación se encuentra cercana a las 160 toneladas, en ningún momento supera dicha cota. La causa de este hecho se debe a que el máximo del subgrupo rápido de la cohorte 2 en esta ocasión se encuentra cercano a los 10000 kg. Esto es debido a que el factor de crecimiento compuesto por la temperatura, amoníaco y nitritos a principios de junio del tercer año es mucho más desfavorable que durante el final del mes de febrero y principios del mes de marzo. De igual forma, el retraso en el crecimiento, implica que la primera venta se lleve a cabo a principios de junio del segundo año, es decir, aproximadamente 20 meses después de la incorporación.

El análisis de la evolución de los pesos medios de cada cohorte, muestra que la extinción del subgrupo de crecimiento rápido de la cohorte 1 se produce prácticamente 3 años después de su incorporación a la planta. Por otra parte se produce un fenómeno no claramente observado en las simulaciones anteriores, que es el mayor incremento del peso medio del subgrupo de crecimiento rápido de la cohorte 2 con respecto al subgrupo de crecimiento lento de la cohorte 1 (figura V.30a). Asimismo, como era de esperar, la cota de pérdidas aumenta en relación a la simulación anterior y la obtención de beneficios se retrasa hasta finales de agosto del tercer año (figura V.30b). También, la incorporación del factor nitrito influye en un menor consumo de pienso así como en la presencia de un único escalón que se inicia a principios del mes de junio del tercer año (figura V.30b).

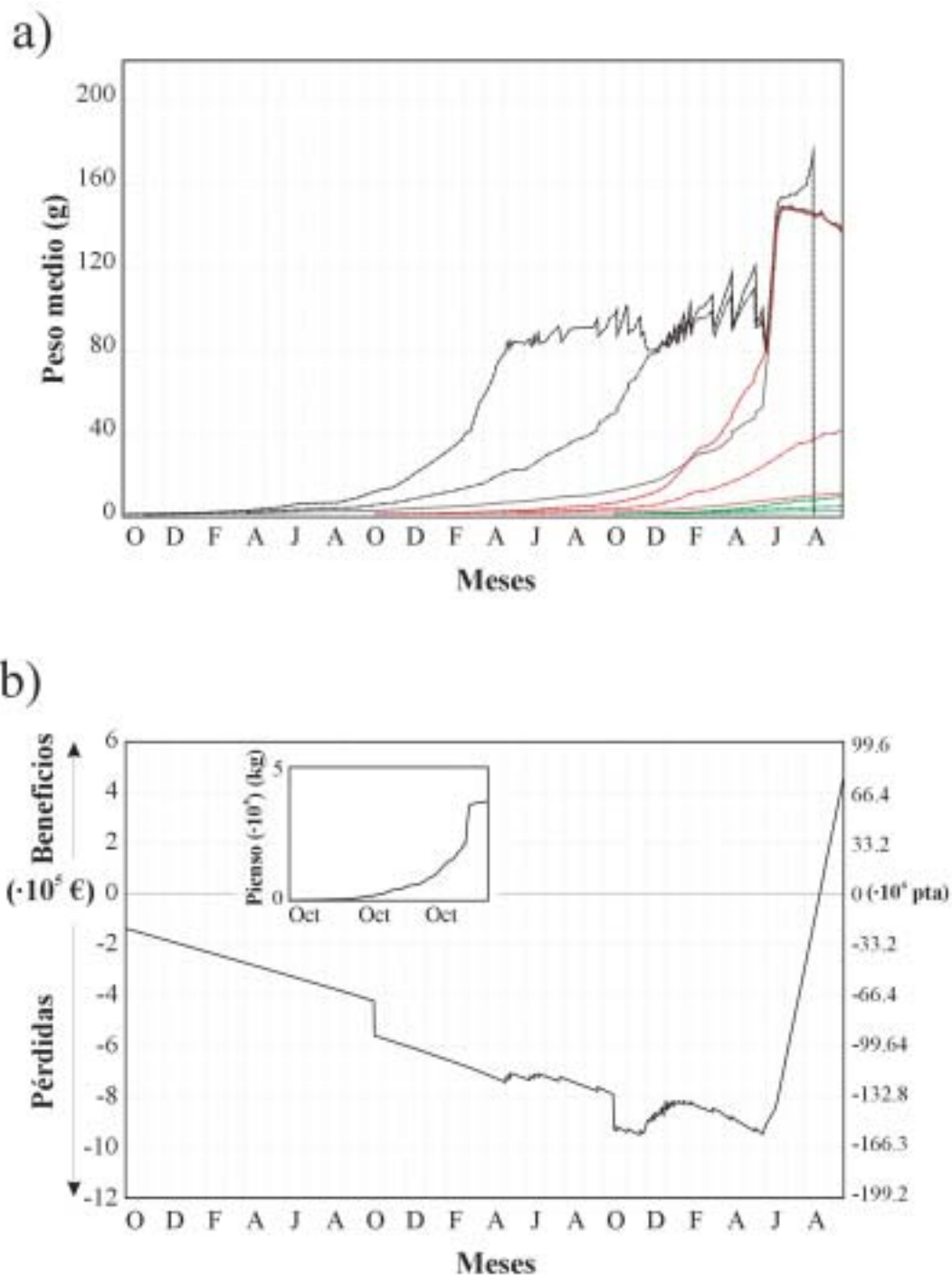
#### 5.3.6.8. *Simulación 7. Incorporación del factor de crecimiento debido a la concentración de nitratos*

La incorporación del factor de crecimiento debido a la concentración de nitratos en el agua desplaza de forma significativa la fecha de la primera venta con respecto a la simulación 6, en la que este factor no era tenido en cuenta (figura V.31). Así, la primera venta del subgrupo de crecimiento rápido de la cohorte 1 se produce en abril del tercer año de simulación, lo que supone aproximadamente 31 meses después de la entrada en la planta. Esto a su vez, provoca que el máximo de biomasa total almacenada en la planta se alcance el último día de simulación, indicando que el máximo real para esta simulación se encuentra, desde un punto de vista temporal, más avanzado. Esto puede explicarse porque en el último día de simulación tan solo un subgrupo de crecimiento ha alcanzado su máximo potencial.

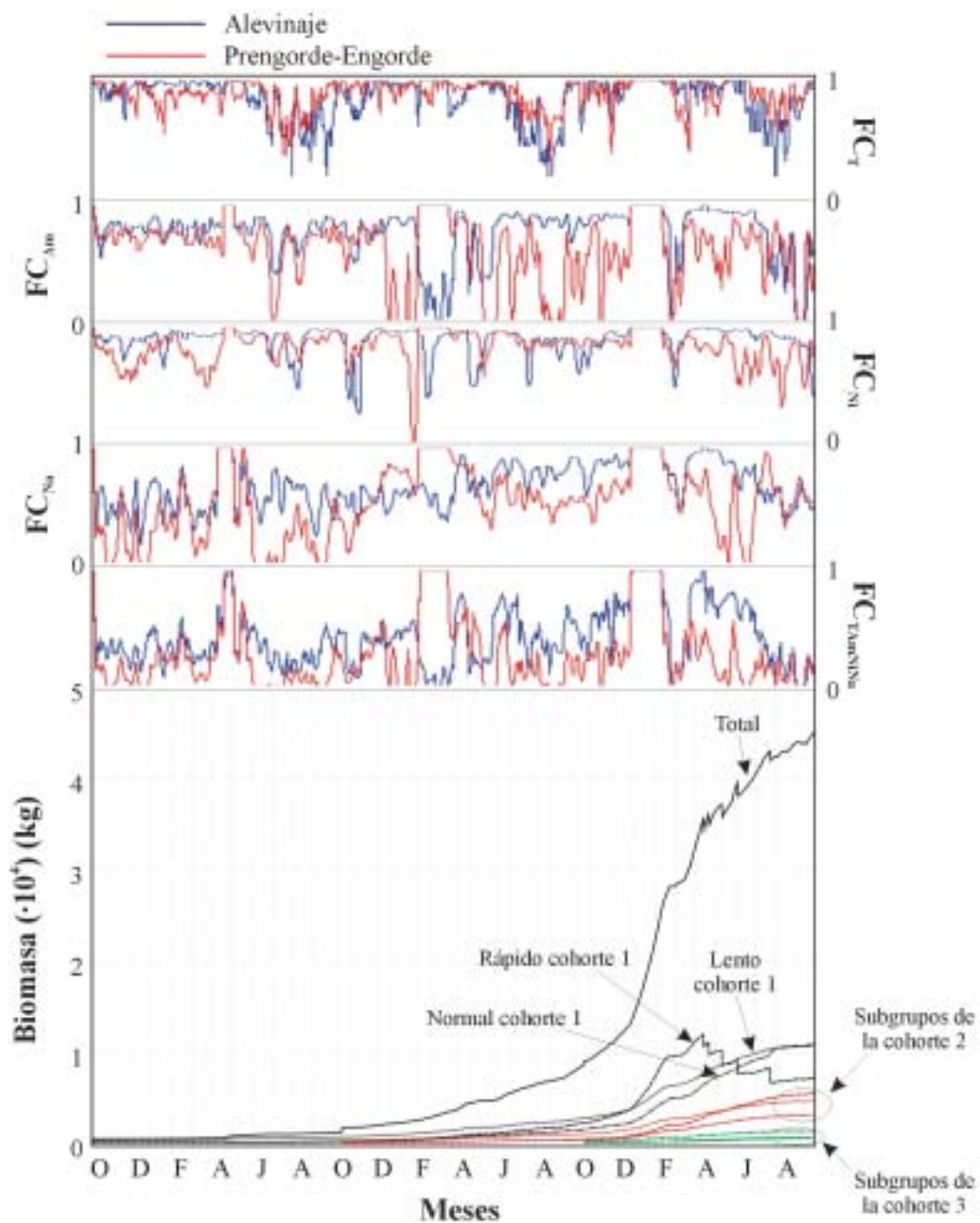
También en esta simulación los crecimientos de algunos subgrupos se adelantan a otros introducidos con anterioridad. De este modo, en la figura V.32a se puede observar como en la parte final de la simulación, los pesos medios de los subgrupos de crecimiento rápido y normal de la cohorte 2 son superiores al peso medio del subgrupo de crecimiento lento de la cohorte 1. Asimismo, en esta figura también se puede comprobar como ninguno de los subgrupos que intervienen en la simulación acaban por extinguirse.



**Figura V.29.** Resultados de la simulación número 6. Biomasa total almacenada en la planta y las correspondientes al subgrupo de crecimiento lento, normal y rápido de las cohortes 1, 2 y 3. En la parte superior de la gráfica se muestran los factores de crecimiento correspondientes a la temperatura del agua, a la concentración de amoníaco, la concentración de nitritos y la combinación de los tres factores en las zonas de alevinaje y preengorde-engorde ( $FC_T$ ,  $FC_{Am}$ ,  $FC_N$  y  $FC_{TAmN}$ )



**Figura V.30.** Resultados de la simulación número 6. a) Peso medio de los subgrupos de crecimiento lento, normal y rápido de las cohortes 1, 2 y 3; b) función de beneficios (ingresos-costes) incluyendo la amortización de la inversión inicial. En esta misma gráfica se incluye la función acumulada del consumo de pienso



**Figura V.31.** Resultados de la simulación número 7. Biomasa total almacenada en la planta y las correspondientes al subgrupo de crecimiento lento, normal y rápido de las cohortes 1, 2 y 3. En la parte superior de la gráfica se muestran los factores de crecimiento correspondientes a la temperatura del agua, concentración de amoníaco, concentración de nitritos, concentración de nitratos y la combinación de los cuatro factores en las zonas de alevinaje y prengorde-engorde ( $FC_T$ ,  $FC_{Am}$ ,  $FC_N$ ,  $FC_{Ta}$  y  $FC_{Tambien}$ ).

Por otra parte, en la figura V.32b se muestra tanto la función de beneficios como la cantidad de pienso consumido. La introducción del factor nitratos impide la recogida de beneficios durante toda la simulación. Ésta se caracteriza por un descenso casi continuo con dos escalones que se corresponden con el desembolso por la compra de angulas. De igual forma, al reducirse considerablemente la biomasa almacenada, también se reduce el consumo de pienso (figura V.32b).

#### 5.3.6.9. Simulación 8. Incorporación del factor de crecimiento debido al pH

La adición del factor de crecimiento debido al pH en la zona de alevinaje supone un retraso acusado del crecimiento. La biomasa total al final de la tercera temporada no supera las 11 toneladas (figura V.33) y el peso medio máximo ni siquiera alcanza los 20 gramos (figura V.34a). Por otra parte, la función de beneficios es en todo momento descendente, de tal forma que en ningún momento se obtiene un saldo positivo (figura V.34b).

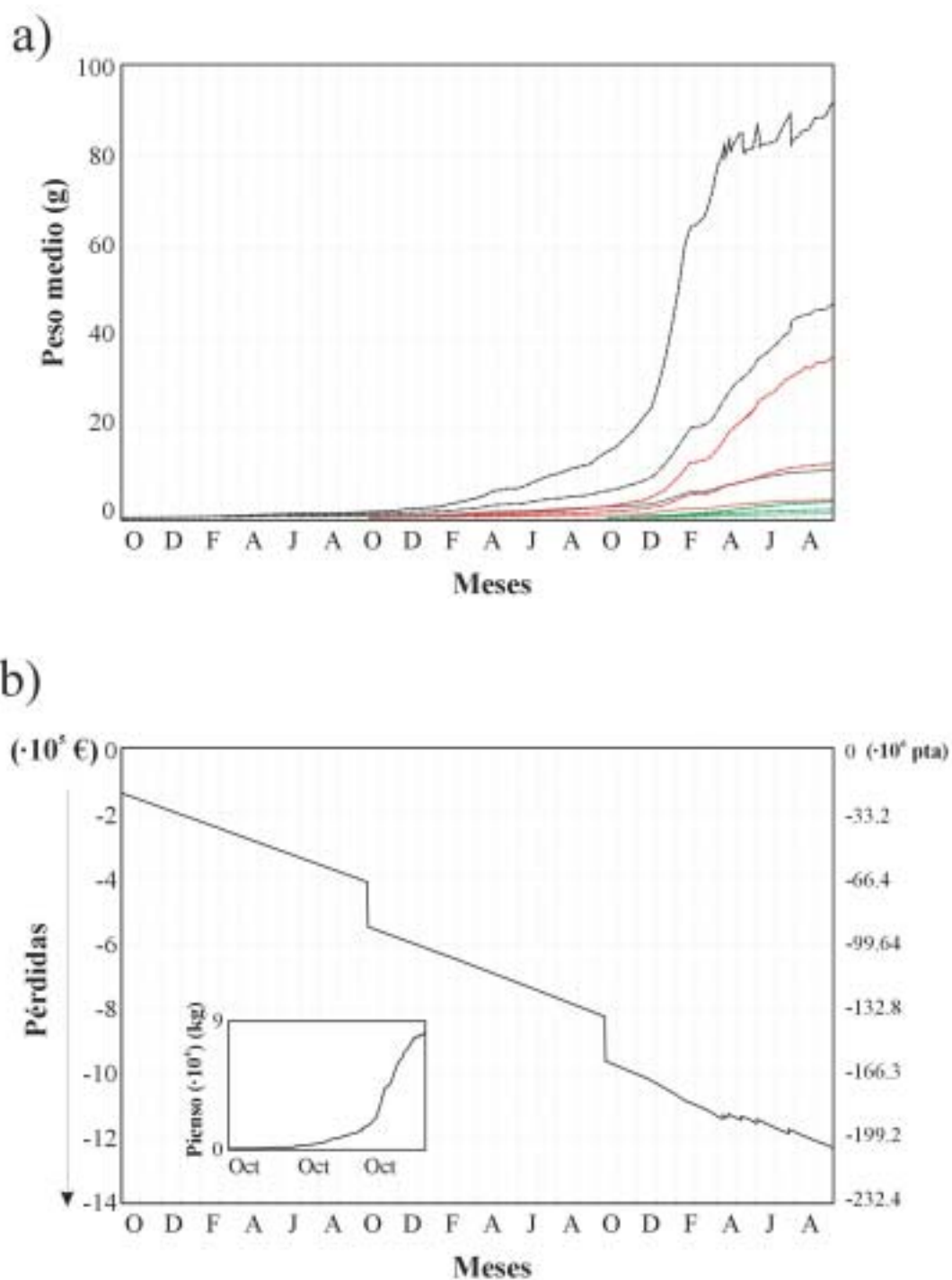
Como se ha podido comprobar en las simulaciones anteriores, la incorporación o eliminación de algunos factores de crecimiento provoca la aparición de escenarios de producción más o menos realistas. De esta forma, la simulación 5 en la que tan solo se tuvo en cuenta la temperatura y la concentración de amoníaco, ha proporcionado los resultados más próximos a los observados en la realidad.

#### 5.3.6.10. Simulación 9. Factores de crecimiento debidos a la temperatura y amoníaco. Influencia de los porcentajes iniciales en cada subgrupo de crecimiento

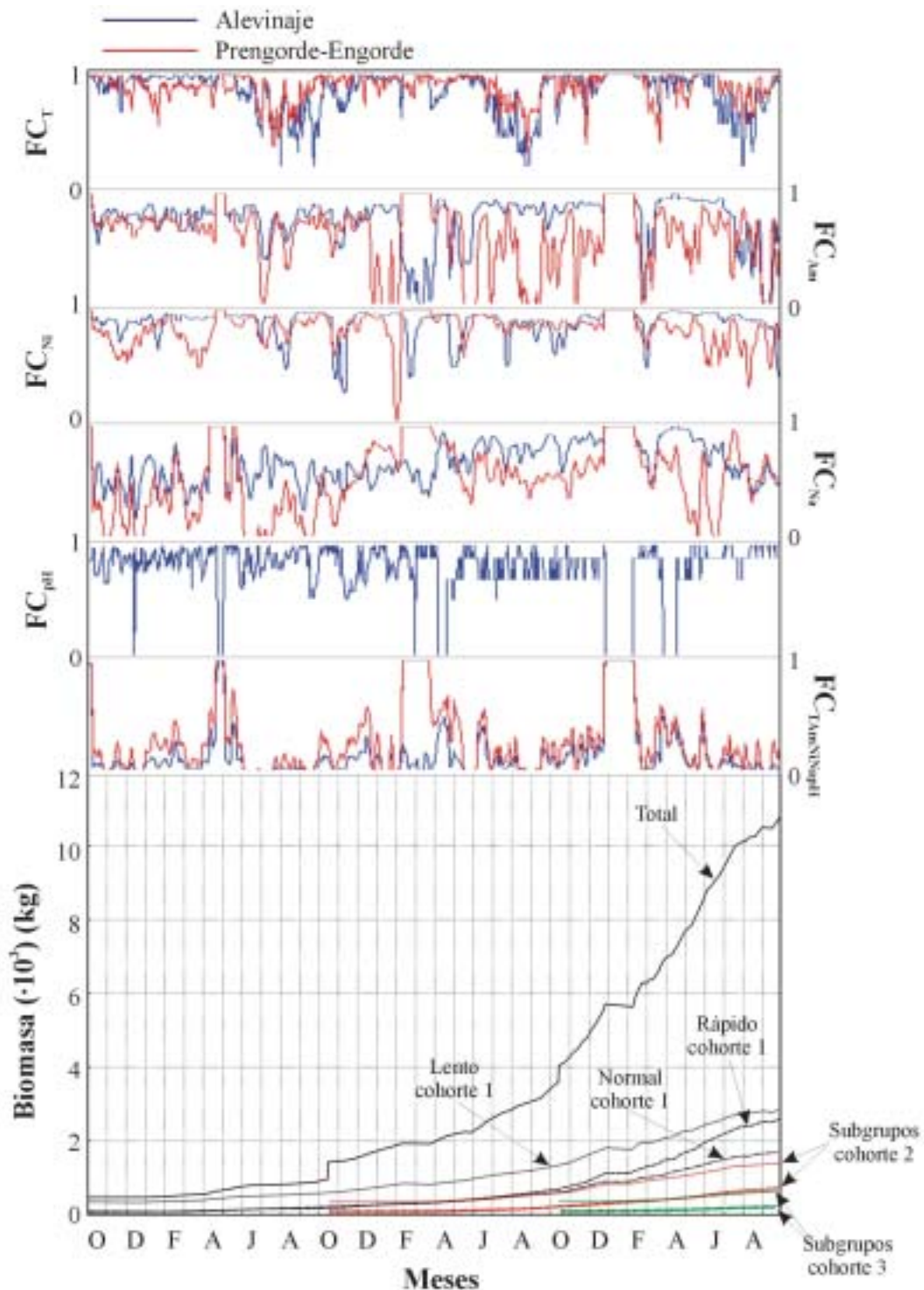
En la figura V.35a se muestra la biomasa total almacenada en la planta a lo largo de tres temporadas para dos simulaciones distintas. En la primera de ellas se ha considerado que las tres temporadas son exactamente iguales, con una variación en los porcentajes de los subgrupos de crecimiento normal y rápido con respecto al resto de simulaciones. La fracción correspondiente con el subgrupo de crecimiento normal es del 20 % mientras que del subgrupo de crecimiento rápido es del 15 %. El máximo correspondiente con el subgrupo de crecimiento rápido de la primera cohorte, encontrado a principios del mes de marzo del segundo año, está en una cota próxima a las 60 toneladas. De igual forma, el máximo del subgrupo de crecimiento normal también está cercano a ese valor 5 meses más tarde. Sin embargo, la coincidencia del máximo desarrollo del subgrupo de crecimiento lento de la cohorte 1 con el subgrupo de crecimiento rápido de la cohorte 2, hace que el *stock* máximo para las tres temporadas se sitúe por encima de las 200 toneladas. La variación de los porcentajes de los subgrupos de crecimiento normal y rápido de las cohortes 2 y 3 al 10 % y 6 %, respectivamente, restablece el *stock* máximo alrededor de 160 toneladas.

El reajuste de las fracciones dentro de cada cohorte implica pocas variaciones desde un punto de vista económico, ya que tan solo se encuentran pequeñas diferencias desde enero hasta abril del tercer año, y en el último mes de simulación. Estas

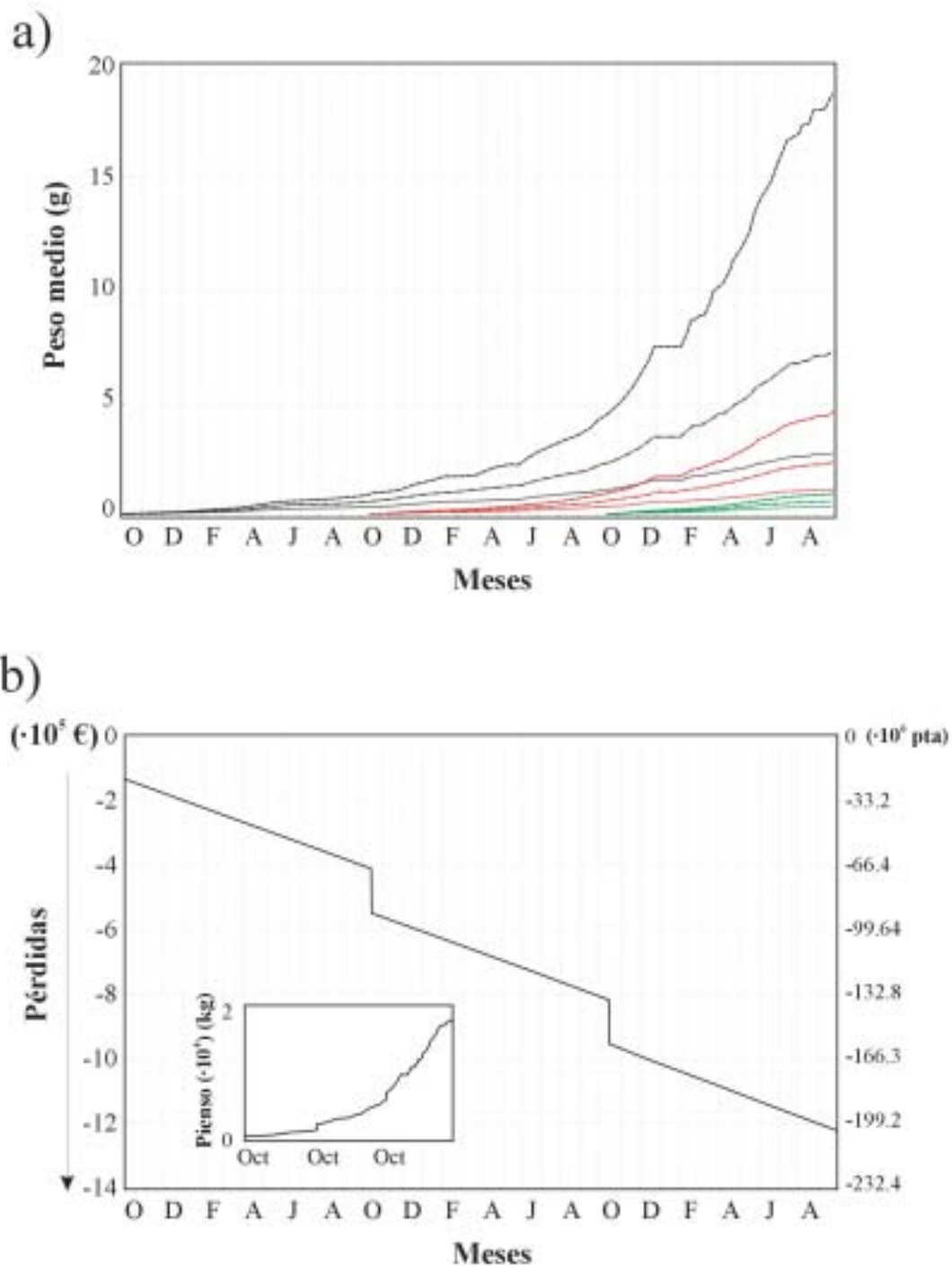




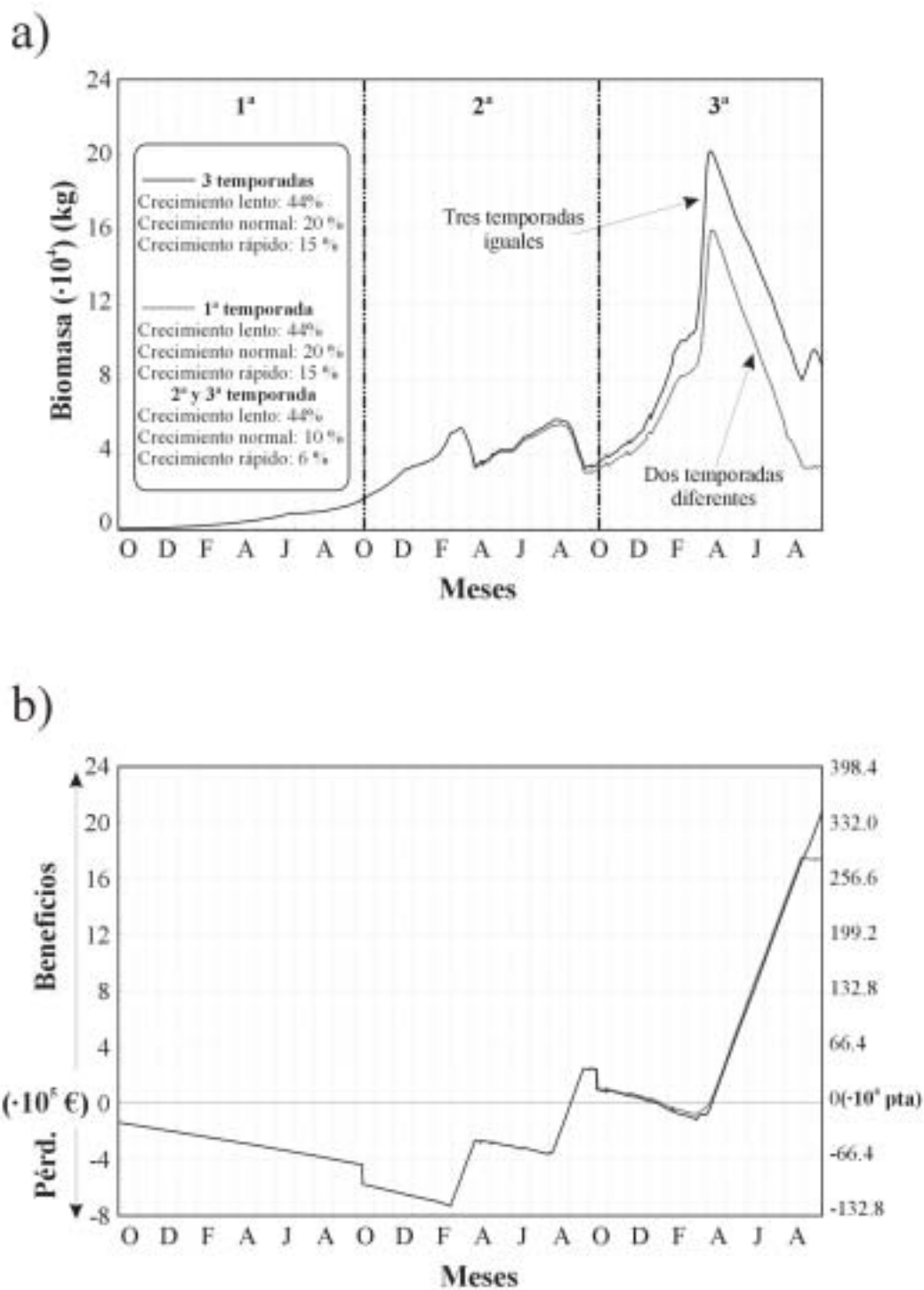
**Figura V.32.** Resultados de la simulación número 7. a) Peso medio de los subgrupos de crecimiento lento, normal y rápido de las cohortes 1, 2 y 3; b) función de beneficios (ingresos-costes) incluyendo la amortización de la inversión inicial. En esta misma gráfica se incluye la función acumulada del consumo de pienso



**Figura V.33.** Resultados de la simulación número 8. Biomasa total almacenada en la planta y las correspondientes al subgrupo de crecimiento lento, normal y rápido de las cohortes 1, 2 y 3. En la parte superior de la gráfica se muestran los factores de crecimiento correspondientes a la temperatura del agua, concentración de amoníaco, concentración de nitritos, concentración de nitratos, pH y la combinación de los cinco factores en las zonas de alevinaje y preengorde-engorde ( $FC_T$ ,  $FC_{Am}$ ,  $FC_{Na}$ ,  $FC_N$ ,  $FC_{pH}$  y  $FC_{Tem/pH}$ )



**Figura V.34.** Resultados de la simulación número 8. a) Peso medio de los subgrupos de crecimiento lento, normal y rápido de las cohortes 1, 2 y 3; b) función de beneficios (ingresos-costes) incluyendo la amortización de la inversión inicial. En esta misma gráfica se incluye la función acumulada del consumo de pienso



**Figura V.35.** Resultados de la simulación número 9. a) Biomasa total considerando tres temporadas idénticas en cuanto a la distribución en porcentaje de los distintos subgrupos en cada cohorte y considerando las dos últimas temporadas distintas a la primera; b) función de beneficios (ingresos-costes) incluyendo la amortización de la inversión inicial

diferencias son debidas a una disminución de costes de pienso cuando los porcentajes son reajustados durante la simulación (figura V.35b).

#### 5.3.6.11. Simulación 10. Influencia del lote de cosecha

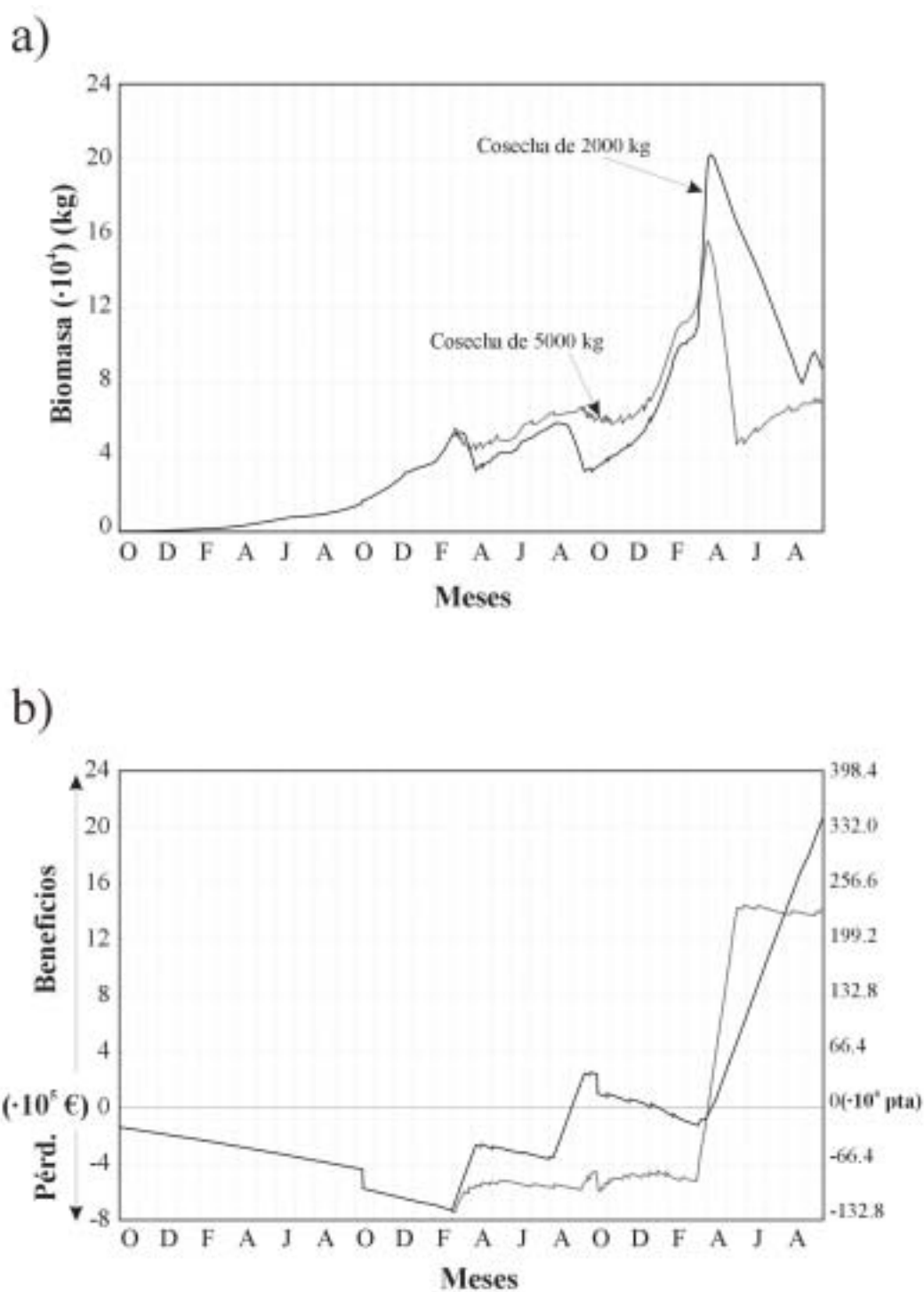
Partiendo de la simulación anterior (simulación 9) en la que se consideran tres temporadas idénticas en cuanto al reparto porcentual de individuos por subgrupos de crecimiento (44 % subgrupo de crecimiento lento, 20 % subgrupo de crecimiento normal y 15 % subgrupo de crecimiento rápido), se ha llevado a cabo una nueva simulación en la que el único factor a variar ha sido el tamaño de los lotes de cosecha, incrementándose éste desde 2000 kg hasta 5000 kg. Se han tenido en cuenta asimismo, los factores de crecimiento debidos a la temperatura y a la concentración de amoníaco. Los resultados se muestran en las figuras V.36a y V.36b.

La biomasa total almacenada en la planta es la misma en ambos casos, hasta que se produce la formación del primer lote de venta aproximadamente a los 18 meses de entrada en la planta (figura V.36a). A partir de ese momento y hasta marzo del tercer año de simulación, el *stock* de la planta cuando el lote de cosecha es de 5000 kg (*stock 1*) siempre es mayor que en el caso de que el lote de cosecha sea de 2000 kg (*stock 2*). Posteriormente, a principios de marzo del tercer año el *stock 1* alcanza su máximo cercano a las 160 toneladas para seguidamente descender bruscamente hasta mediados de mayo. En cambio, el *stock 2* (tal y como se comprobó en la simulación 9) sitúa su máximo ligeramente por encima de las 200 toneladas a principios de abril del tercer año, descendiendo de forma más suave que en el caso anterior hasta agosto. La variación del tamaño de cosecha tiene una gran influencia sobre la biomasa almacenada en la planta, y se consiguen resultados parecidos a los obtenidos mediante la variación del reparto del número de individuos por subgrupos durante la simulación.

Estos resultados influyen directamente sobre las funciones de beneficios. En la figura V.36b se puede ver como en el *stock 1* una mayor biomasa almacenada durante el período transcurrido desde principios de marzo del segundo año hasta mediados del mismo mes del tercer año, impide la consecución de un saldo positivo. Asimismo, en la parte final de la simulación se muestra que, a pesar de que en el *stock 1* se produce un incremento brusco de los ingresos como consecuencia de un volumen alto de ventas, los obtenidos finalmente por el *stock 2* son mayores.

#### 5.3.6.12. Simulación 11. Influencia del cambio del lote de cosecha durante la simulación

Los resultados de la simulación 11 se muestran en la figura V.37. Con esta simulación se ha pretendido observar la evolución de la biomasa total cuando se producen cambios en los lotes de venta durante la simulación. Así, estas condiciones se han mantenido constantes hasta mediados de abril del tercer año de simulación (lotes de 5000 kg). A partir de ese momento, en el resto de simulaciones, se producía un descenso brusco del



**Figura V.36.** Resultados de la simulación número 10. a) Biomasa total considerando tres temporadas idénticas en cuanto a la distribución en porcentaje de los distintos subgrupos en cada cohorte y con distintos lotes de cosecha (2000 y 5000 kg); b) función de beneficios (ingresos-costes) incluyendo la amortización de la inversión inicial



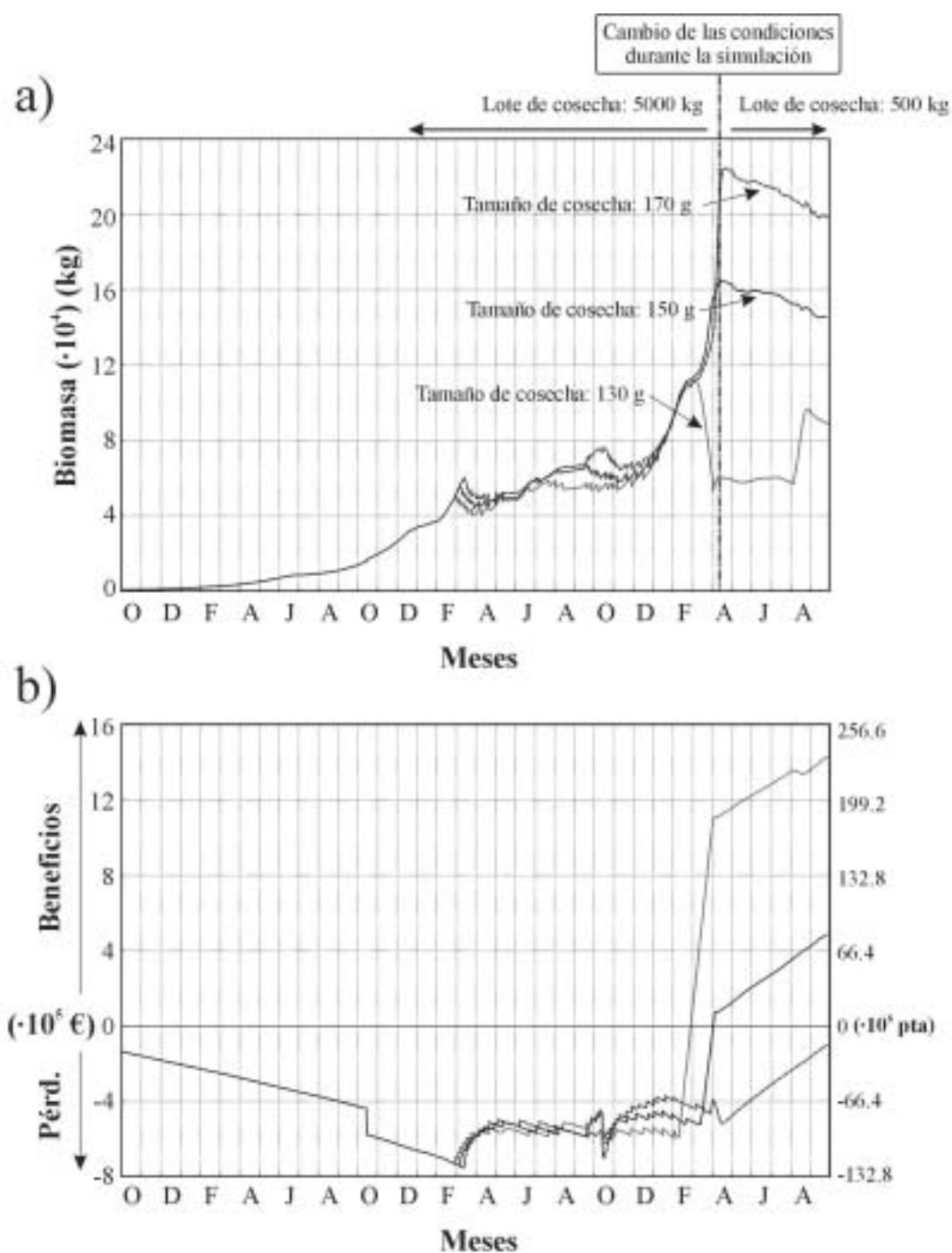
**Figura V.37.** Resultados de la simulación número 11. Biomasa total almacenada en la planta con lotes de venta de 5000 kg hasta mediados de abril. A partir de ese momento se contemplan dos posibilidades: a) lotes de venta de 500 kg y b) lotes de venta de 1000 kg

stock debido a que la biomasa por encima de la talla de cosecha era muy abundante. Al mismo tiempo tan solo se había contemplado la posibilidad de lotes de venta de 2000 y 5000 kg. Esto provocaba un período de ventas intenso que llevaba la biomasa almacenada hasta niveles alrededor de las 40 toneladas. Sin embargo, considerando lotes de venta menores (500 y 1000 kg) se consigue un descenso suave que deja a la planta con un nivel de almacenamiento de entre 103 y 145 toneladas.

Como es de esperar, un cambio en los lotes de venta introduce cambios en la función de beneficios de tal forma que, cuando se consideran lotes de venta menores (500 kg) los beneficios aumentan, aunque no de forma tan acusada como cuando el producto se vende en lotes de 1000 kg.

### 5.3.6.13. Simulación 12. Influencia del tamaño de cosecha

El efecto del tamaño de cosecha sobre la cantidad de biomasa almacenada en la planta se muestra en la figura V.38a. El efecto es muy significativo básicamente cuando el subgrupo de crecimiento lento de la primera cohorte alcanza su máximo potencial. De esta forma, si el tamaño de cosecha se adelanta hasta los 130 gramos se produce un descenso de la biomasa desde las 110 toneladas, que alcanza a finales de febrero del tercer año, hasta algo menos de 60 toneladas a principios de abril de ese mismo año.



**Figura V.38.** Resultados de la simulación número 12. a) Biomasa total considerando tres temporadas idénticas en cuanto a la distribución en porcentaje de los distintos subgrupos en cada cohorte y con distintos tamaños de cosecha (130, 150 y 170 g). También se considera un cambio del tamaño del lote de cosecha durante la simulación, pasando de 5000 a 500 kg; b) funciones de beneficios (ingresos-costes) incluyendo la amortización de la inversión inicial



Esto está propiciado por varias ventas consecutivas de 5000 kg y prácticamente coincide con el cambio de las condiciones de simulación. Este cambio, al igual que en la simulación 11, consiste en la reducción de los lotes de cosecha desde 5000 kg hasta 500 kg. A partir de ese momento la biomasa se estabiliza para volver a aumentar a principios de agosto.

El descenso en la biomasa detectado cuando se cosecha a partir de 130 gramos no es apreciable cuando se cosecha a 150 ó 170 gramos. Esto favorece el aumento de la biomasa hasta cotas por encima de 160 y 220 toneladas, respectivamente. Estos máximos coinciden con el cambio de las condiciones de simulación por lo que no es posible una venta masiva de lotes de 5000 kg. La ausencia de dichas ventas se refleja con claridad en la función de beneficios (figura V.38b). Mientras que la cosecha a 130 gramos produce los mayores beneficios al final de la simulación, el retraso del tamaño de cosecha hasta 170 gramos ni siquiera genera beneficios. Esto se produce a pesar de que en este último caso se encuentra la mayor cantidad de biomasa acumulada, lo que sin duda contribuye a incrementar los costes.

#### 5.3.6.14. *Simulación 13. Aplicación de los modelos de predicción de las variables temperatura y amoniaco*

Partiendo de la simulación número 11, en la que se considera un cambio en el lote de cosecha a partir de mediados de abril del tercer año, se ha llevado a cabo una nueva simulación en la que los factores de crecimiento son calculados a partir de las estimaciones con uno y dos días de antelación de la temperatura y la concentración de amoniaco. Ambos parámetros son obtenidos mediante un proceso de realimentación, es decir, el valor predicho en un instante  $t$  pasa en el momento  $t+1$  a convertirse en el valor de entrada  $t-1$  ó en el  $t-2$  en el modelo de predicción.

En la figura V.39 se muestra como existe un retraso de los picos de biomasa máxima, ligero en el caso de la predicción con un día de antelación y acusado cuando se tienen en cuenta dos días de antelación. Así, cuando se realizan las regresiones simples entre la biomasa simulada con factores de crecimiento estimados para uno y dos días de antelación, y la biomasa simulada con factores de crecimiento reales para uno, dos y tres años de simulación (figura V.40), se puede comprobar como en los tres casos se han obtenido mejores resultados con los factores de crecimiento estimados con un día de antelación. Asimismo, a pesar de que con dos años de simulación el porcentaje estándar de predicción es menor al calculado para un único año de simulación, se observa una tendencia al aumento del error en función del número años de simulación, como se puede desprender del análisis del coeficiente de determinación, **RMS** y coeficiente de eficiencia (**E**). De este modo, se puede decir que para el primer año de simulación y con un día de antelación el error cometido sobre la biomasa total almacenada es de aproximadamente un 17 %, en el segundo un 11 % y en el tercero está cercano al 40 %.

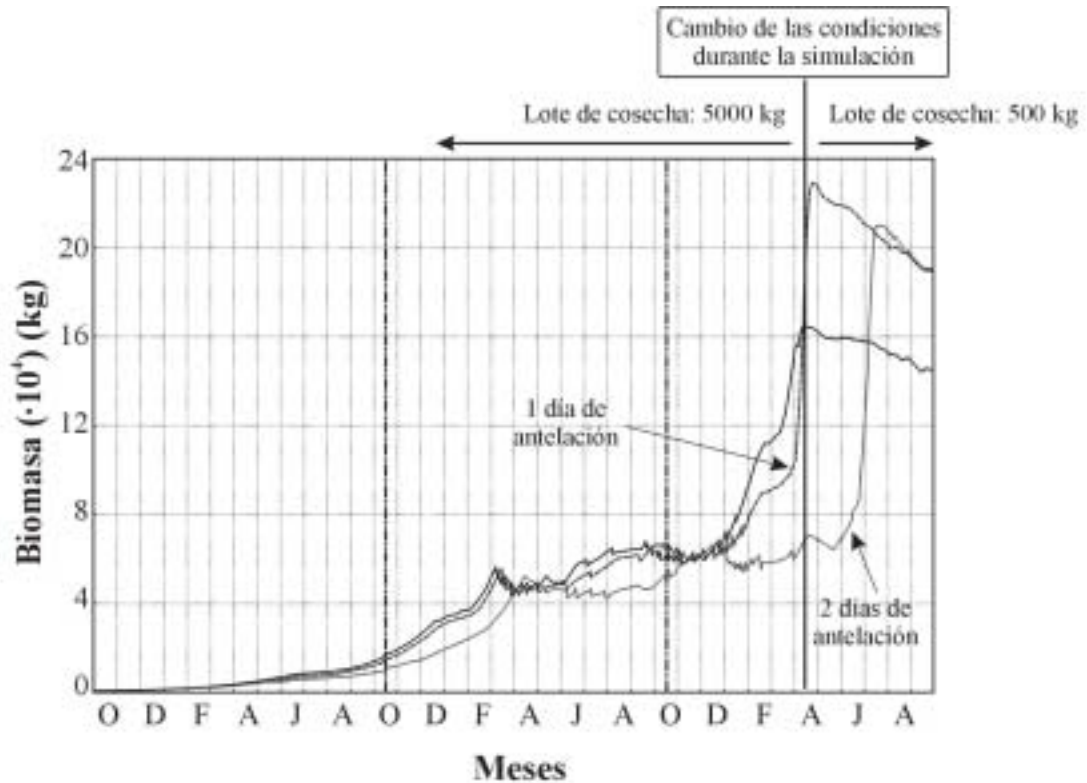
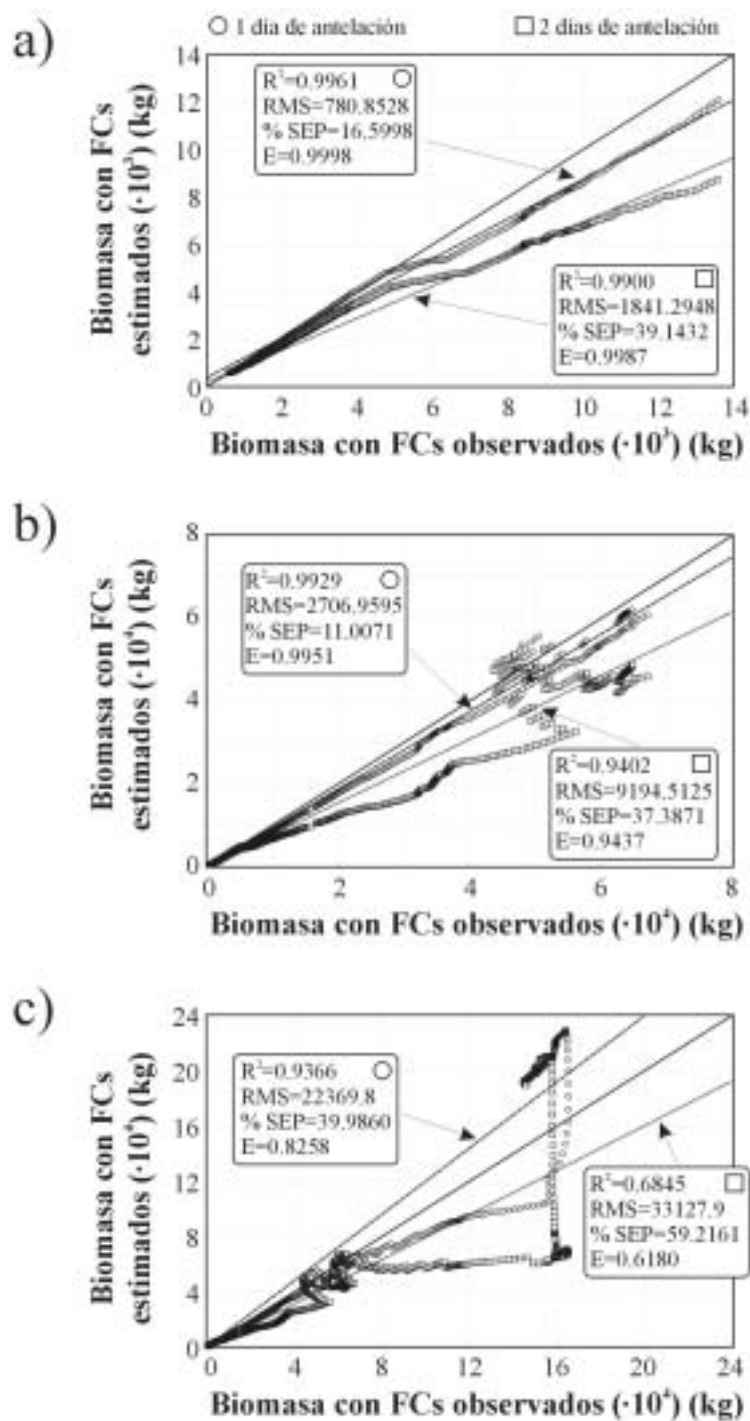


Figura V.39. Simulación número 13. Evolución de la biomasa total observada y estimada en función de los factores de crecimiento debidos a la temperatura y concentración de amoniaco predichos con uno y dos días de antelación

#### 5.3.6.15. Simulación 14. Conexión con SEDPA 1.0 a través de las reglas A-A de CONTROLPLANT 1.0

Nuevamente, partiendo de la simulación número 11, se ha llevado a cabo la conexión del Programa Operativo de Simulación con el sistema experto **SEDPA 1.0** a través de las reglas **A-A** de **CONTROLPLANT 1.0**. **POS** ha detectado la posibilidad de aparición de un total de hasta 12 agentes patológicos, de los cuales 8 (66.7 %) fueron identificados en el período transcurrido desde enero de 1999 hasta enero de 2001 (tabla V.8).

De todos los agentes patológicos detectados por **POS** con presencia real en la planta tan solo en dos ocasiones presentaron una categoría de presencia 'Rara' (*Saprolegnia* sp. y *Flexibacter columnaris*), mientras que el resto se encontraron en la categoría 'Común' o 'Muy común'. Por otra parte, aquellos que fueron detectados por **POS** pero no se identificaron en la planta se encontraron dentro de las categorías 'Rara' o 'Muy rara'.



**Figura V.40.** Regresiones lineales entre la biomasa simulada con factores de crecimiento (FCs) reales y la biomasa simulada con factores de crecimiento estimados debido a la temperatura y concentración de amoníaco predichos con uno y dos días de antelación. a) regresión del primer año de simulación; b) regresión de los dos primeros años de simulación y c) regresión de los tres años de simulación

**Tabla V.8.** Agentes patológicos detectados por el POS así como su presencia real en la planta durante el período de estudio. Se indica asimismo la intensidad de la presencia de cada agente patológico en cultivos intensivos de anguilas

Agente patológico	Presencia real	Detección por POS	Presencia en cultivos intensivos de anguilas
<i>Dactylogyrus anguillae</i>	Si	Si	Muy común
<i>Gyrodactylus</i> sp.	Si	Si	Muy común
<i>Trichodina</i> sp.	Si	Si	Común
<i>Ichthyophthirius multifiliis</i>	Si	Si	Muy común
<i>Saprolegnia</i> sp.	Si	Si	Rara
<i>Dermocystidium anguillae</i>	No	Si	Rara
<i>Aeromonas hydrophila</i>	Si	Si	Común
<i>Pseudomonas anguilliseptica</i>	Si	Si	Común
Papilomatosis	No	Si	Muy rara
<i>Flexibacter columnaris</i>	Si	Si	Rara
<i>Edwardsiella tarda</i>	No	Si	Rara
<i>Yersinia ruckeri</i>	No	Si	Muy rara

## 5.4. DISCUSIÓN

Como muestran los resultados, la simulación de la estrategia de producción mediante la utilización de modelos sencillos de crecimiento y extensión de modelos de varianza, puede proporcionar al gestor de la planta una herramienta mediante la cual ajustar los parámetros de producción a corto y medio plazo. Resultados similares a los obtenidos fueron alcanzados por Summerfelt *et al.* (1993) en la previsión del crecimiento y cosecha en una piscifactoría de pez gato (*Ameiurus melas*). Asimismo, la simulación permite establecer la estructura básica de la población, gracias a la cual es posible una previsión de ventas realista y facilita una cuenta de resultados inicial.

Mediante técnicas de simulación, autores como McDonald *et al.* (1996) y Cho y Bureau (1998) han conseguido resultados satisfactorios en la modelación de los sistemas de cultivo de peces para consumo. En la mayoría de las ocasiones ésta se basa en modelos bioenergéticos en los que el consumo de alimento, la asimilación de proteínas, la excreción de productos de desecho y la respiración son estimados a partir de submodelos o son obtenidos experimentalmente a partir de individuos confinados en condiciones controladas. Sin embargo, de igual forma que concluyen estos mismos autores, la inclusión de hipótesis relativamente subjetivas, el desconocimiento de muchos aspectos fisiológicos, ecológicos y biológicos, y la dificultad añadida de estudiar animales que se desenvuelven en un ambiente totalmente ajeno al nuestro, provoca que los resultados puedan ser considerados subóptimos.

#### **5.4.1. Variación de la biomasa y mortalidad**

A lo largo del período de estudio se han observado variaciones de la cantidad de biomasa real almacenada en la planta, en parte debido a cambios intrínsecos del sistema (crecimiento, ventas, mortalidad y entrada de nuevos ejemplares, fundamentalmente) y en parte debido a pequeños desajustes tras la aplicación de las funciones normales a las medias estimadas por el gestor. Esto puede ser una consecuencia de que la estimación parta de las biomazas individuales de cada tanque y de la suposición de que los individuos distribuyen sus tallas de forma normal. Así, cuando se aplican las funciones de distribución normal, la distribución de frecuencias por tallas es obviamente normal, pero su transformación a biomasa produce un sesgo de la función hacia la izquierda, y por tanto, un cambio en la ubicación del peso medio de la población. La razón de este desplazamiento tenemos que buscarla en el tipo de relación existente entre el peso y la longitud en los peces. Esta viene determinada por el coeficiente de forma o de condición (**K**), que es la relación entre el peso y el cubo de la longitud (García de Jalón-Lastra *et al.*, 1991; Bone *et al.*, 1995). De esta forma, si la distribución por longitudes es normal, la distribución por peso no lo es, puesto que el peso de un individuo cuya longitud pertenezca a un intervalo de la cola izquierda de la función es menor al de un individuo que se encuentre en un intervalo de la cola derecha. Análogamente ocurre lo mismo con la biomasa debida a la mortalidad. Asimismo, esto también explica la anormal reducción del número de individuos detectada de forma recurrente tras cada clasificación.

Este mismo fenómeno se detecta cuando durante la simulación se llevan a cabo procesos de venta. Cuando se produce una venta el reajuste del peso medio de la población parte de la cantidad de biomasa restante tras eliminar una porción del sistema, lo que provoca momentáneamente (en tiempo real durante la simulación) un cambio de distribución irreal que se corrige en el instante de tiempo siguiente.

#### **5.4.2. Comportamiento y robustez del modelo**

La comparación de los resultados de la simulación con los parámetros poblacionales reales ha resultado más compleja y quizás más difícilmente evaluable que en capítulos anteriores. Esto es debido a que en este caso la escala temporal es más amplia, por lo que probablemente es más interesante la precisión que la exactitud del modelo. De esta forma, las coincidencias físicas y temporales, en lo que podemos denominar puntos críticos de la producción (biomasa máxima, nivel de crecimiento de una temporada a otra, volumen de ventas, etc.), proporcionan un test de si las suposiciones del modelo están justificadas y si las ecuaciones utilizadas las representan adecuadamente, aportando una idea intuitiva de la robustez del modelo. Asimismo, la bondad de ajuste del modelo está condicionada por la calidad de los datos reales tomados en la planta. En este sentido, hay que tener en cuenta que las funciones de biomasa total han sido obtenidas a partir de datos semanales que a su vez han sido estimados por el gestor con un error de partida cercano al 20 %. Es por ello, por lo que quizás es más correcto hablar del comportamiento del modelo en vez de su validación (Hay y Walker, 1989).

La aplicación del modelo de extensión de la varianza de Arnason-Sparre conjuntamente con funciones de distribución normal, ha proporcionado la posibilidad de observar la variación de la biomasa a lo largo del tiempo para todo el rango de tamaños comerciales de la anguila. Como se comentó en el apartado 5.2.2, el principal inconveniente del modelo de Arnason-Sparre radica en el establecimiento de una desviación estándar inicial y el cálculo de los coeficientes  $\alpha$ ,  $\beta$  y  $\gamma$ . En este sentido, los datos procedentes de las clasificaciones de aquellos tanques en la que los ejemplares permanecieron menos de 40 días, han proporcionado una estimación de la desviación estándar inicial más próxima a la real que el estimado mediante el análisis de las mediciones de los ejemplares del tanque 'A1'. Dicha estimación (0.18) es muy próxima a la calculada por otros autores como Wickins (1985), quien para varios lotes de angulas de tamaños comprendidos entre 0.26 y 0.45 gramos obtuvo desviaciones típicas de entre 0.05 y 0.20 gramos. Resultados parecidos también son obtenidos por Kuhlmann (1979) para angulas del mar del Norte y del mar Tirreno ( $0.48 \pm 0.09$  y  $0.25 \pm 0.04$ , respectivamente). Estos valores quedan más alejados de los estimados a partir de los datos del tanque 'A1', probablemente debido a que estos últimos no fueron tomados específicamente para esa tarea.

Por otra parte, la estimación de los coeficientes  $\alpha$ ,  $\beta$  y  $\gamma$  mediante el análisis de regresión múltiple ha resultado relativamente satisfactoria. Así, a pesar de que el porcentaje de varianza explicada fue ligeramente superior al 90 % y de que el modelo fue significativo a un nivel inferior al 0.1 %, ésta fue obtenida a costa de los estimadores  $b_0 = \ln(\alpha)$  y  $b_1 = \beta$ , siendo el estimador  $b_2 = \gamma$  no significativo. En base a la construcción del modelo, podemos encontrar básicamente dos razones que justifican este comportamiento. En primer lugar, tal y como se ha comentado en el párrafo anterior, los datos extraídos del tanque 'A1' no fueron obtenidos específicamente para la estimación de los coeficientes del modelo de Arnason-Sparre. La base experimental de este modelo se fundamenta en lotes de ejemplares aislados y sometidos a distintos tratamientos. En este sentido, los ejemplares del tanque 'A1' estuvieron sometidos a temperaturas que oscilaron en un rango de entre 26.5 y 29.5 °C. Esto, tal y como señala Arnason *et al.* (1992) se refleja a escala del parámetro  $\alpha$  y no a escala temporal o sobre el coeficiente de peso ( $\beta$  y  $\gamma$ , respectivamente). Por otra parte, estos mismos autores someten a distintos tratamientos diferentes estirpes genéticas de la trucha arco iris (*Oncorhynchus mykiss*), encontrando que cuando comparan estirpes genéticas semejantes las diferencias entre los parámetros  $\beta$  y  $\gamma$  no son significativas. Sin embargo, si se encuentran importantes diferencias significativas cuando los ejemplares pertenecen a estirpes genéticas diferentes. Esto indica la influencia que la procedencia de las capturas del medio natural tiene sobre la variación del crecimiento. De hecho, el grupo de ejemplares del tanque 'A1' fue conformado mediante la mezcla de distintos lotes de angulas capturados en el bajo Guadalquivir por distintos pescadores.

#### **5.4.3. Inclusión de factores de crecimiento debidos a parámetros físico-químicos**

La inclusión en la simulación de factores de crecimiento (FC) ha proporcionado resultados más realistas que las simulaciones de referencia, en la que estos factores no

eran considerados. En el caso de las simulaciones llevadas a cabo con **POS**, los factores que modifican el proceso de crecimiento de forma más acorde con lo observado en la realidad son la temperatura del agua y la concentración del amoníaco. La introducción de estos factores es un procedimiento utilizado por otros autores. Así, Cho y Bureau (1998) introducen el efecto de la temperatura por medio del coeficiente de crecimiento por unidad termal que depende del peso y del número de días de crecimiento. En nuestro caso, dada la naturaleza de la simulación, **FC** modifica la tasa de crecimiento inicial siempre de forma restrictiva, lo que favorece un cambio acusado de la tasa de crecimiento cuando el número de factores tenidos en cuenta es alto. Esto puede ser fácilmente observable en la simulación 8.

Lo observado en la simulación 8 y de forma menos acusada en las simulaciones 6 y 7 puede tener varias interpretaciones. Por una parte, se puede pensar que la metodología propuesta (multiplicación de **FCs** debido a la temperatura, concentración de amoníaco, nitritos y nitratos y valor del pH) no se acopla correctamente con el resto del modelo. De esta forma, la inclusión de otros **FCs** incidirían negativamente sobre el crecimiento hasta que éste fuese nulo. Esto puede constituir una situación real en ausencia de condiciones de habitabilidad. Sin embargo, la serie de propiedades físico-químicas utilizadas en los cálculos de los **FCs** son las reales registradas en la planta. Teniendo en cuenta estas condiciones, no se han producido crecimientos tan lentos como los registrados en las simulaciones 6, 7 y 8. Es más probable, por tanto, que la calibración de los **FCs** no sea correcta al menos en aquellos casos en los que la incidencia sobre el crecimiento del parámetro físico-químico del que depende esté menos estudiada, es decir, todos excepto la temperatura. En el caso de la temperatura, Jauncey (1982) encontró relaciones polinómicas de tercer grado con el crecimiento específico para tres niveles de alimentación distintos en la carpa común (*Cyprinus carpio*), con la particularidad de que el máximo de crecimiento para los tres tratamientos se producía a la misma temperatura. A pesar de ello, hay que tener en cuenta que, ni siquiera en el caso de la temperatura, se comprende del todo la vía fisiológica por la cual influye en el crecimiento.

Por otra parte, existe una falta de información sobre la combinación de parámetros físico-químicos y su impacto sobre el crecimiento de la anguila, y en general, sobre cualquier especie piscícola cultivada. Es posible que el impacto sobre el crecimiento de un determinado parámetro físico-químico dependa de otro o varios parámetros, de igual forma que la toxicidad del amoníaco depende de la temperatura y del pH.

La incorporación de los **FCs** calculados a partir de parámetros físico-químicos (temperatura del agua y concentración de amoníaco) estimados mediante los modelos calibrados en el capítulo anterior han proporcionado resultados bastante satisfactorios. Las estimaciones realizadas para un único día de antelación indican que tan solo en el conjunto de los tres años de simulación el error cometido es excesivamente alto. Sin embargo, cuando se tienen en cuenta el primer año o los dos primeros años de simulación, se puede considerar que el error cometido es admisible. Esto indica la robustez de los modelos de predicción calibrados en el capítulo anterior. Sin embargo, cuando estos modelos son forzados a estimar la biomasa con dos días de antelación, se

produce un aumento del error para todo el período de estudio considerado. De este modo, se pone de manifiesto la especificidad y dependencia de los modelos de la escala temporal. A pesar de ello, cuando se consideran dos días de antelación, la mayor contribución a la variación de la estimación se debe al factor de crecimiento de la temperatura (dependiente en todos los sistemas de modelos **ARIMA**), mientras que el factor de crecimiento debido a la concentración de amoníaco (dependiente en función del sistema de un modelo **ARIMA**, una regresión múltiple y una red neuronal) contribuye menos a dicha variación. Esto puede ser debido, tal y como indican Pulido-Calvo *et al.* (2002) a que las redes neuronales tienden a estimar valores menos extremos, es decir, más centrados alrededor de un valor medio, lo que supone principalmente una infraestimación de las concentraciones de amoníaco. Por otra parte, los modelos **ARIMA** suelen presentar una estimación de los valores extremos máximos y mínimos alejada de los valores observados, como también les ocurre a Hsu *et al.* (1995), mientras que las regresiones múltiples tienden a producir un desplazamiento entre la curvas estimadas y observadas. Así, la combinación de los tres tipos de modelos produce una amortiguación de la respuesta en el caso de la concentración de amoníaco.

#### **5.4.4. Conexión con SEDPA y reglas A-A de CONTROLPLANT**

La conexión entre el **POS** y **SEDPA** a través de las reglas **A-A** de **CONTROLPLANT** ha producido resultados satisfactorios. Así, para las condiciones reales, el **POS** fue capaz de predecir el 100 % de las patologías detectadas durante el período de simulación. Asimismo, se encontró una alta coincidencia de los períodos de aparición de cada patología con los predichos por el **POS**. En este sentido, el conjunto de programas **CONTROLPLANT**, **SEDPA** y **POS** se han mostrado superiores a otros programas de diagnóstico de patologías de peces como **HAMES** o **FISHVET**, ya que estos no son capaces de situar temporalmente la aparición de una u otra patología. Sin embargo, una propiedad de la que carece tanto el conjunto **CONTROLPLANT**, **SEDPA** y **POS** así como otros programas de diagnóstico, es la de cuantificar o correlacionar el grado incidencia de cada patología en función del tiempo y de las condiciones ambientales. Esto es debido básicamente a la prácticamente total ausencia de estudios llevados a cabo en este sentido.

## **5.5. REFERENCIAS**

- Andersson, J., O. Sandstrom y H.J.M. Hansen. 1991. Elver (*Anguilla anguilla* L.) stocking in a Swedinh thermal effluent: Recaptures, growth and body condition. *Journal of Applied Ichthyology*, 7: 78-89.
- Aprahamian, M.W. 2000. The growth rate of eel in a tributaries of the lower River Severn, England, and its relationship with stock size. *Journal of Fish Biology*, 56: 223-227.



- Arnason, A.N. y D.H. Scuse. 1984. Design and implementation of a computerized decision support system for hatchery management. *International Symposium on Automation and Data Processing in Aquaculture*: 231-241.
- Arnason, A.N., M.H. Papst y G.E. Hopky. 1992. Modelling the increase in variance of fish weight. *Canadian Journal of Fisheries and Aquatic Science*, 49:2-16.
- Bjorndal, T. 1988. Optimal harvesting of farmed fish. *Marine Resources Economics*, 5: 139-159.
- Bone, Q., N.B. Marshall y J.H.S. Blaxter. 1995. *Biology of fishes*. Edita: Blackie Academic and Professional. Londres.
- Brett, J.R. 1979. *Fish physiology*. Edita: Academic Press. New York.
- Brown, L. 2000. *Acuicultura para veterinarios. Producción y clínica de peces*. Edita: ACRIBIA S.A. Zaragoza.
- Cho, C.Y. y D.P. Bureau. 1998. Development of bioenergetic models and the Fish-PrFEQ software to estimate production, feeding ration and waste output in aquaculture. *Aquatic Living Resources*, 11(4): 199-210.
- Colt, J. 1991. Modelling production capacity of aquatic culture system under freshwater conditions. *Aquacultural Engineering*, 10: 1-29.
- Crivelli, A.J. y D. Mestre. 1988. Life history traits of pumpkinseed, *Lepomis gibbosus*, introduced into the Camargue, a mediterranean wetland. *Archive of Hydrobiology*, 111(3): 449-466.
- Deelder, C.L. 1978. A short note on the intensive culture of eel (*Anguilla anguilla* L.). *Aquaculture*, 13: 289-290.
- Deelder, C.L. 1981. On the age and growth of cultured eels, *Anguilla anguilla* (Linnaeus, 1758). *Aquaculture*, 26: 13-22.
- Degani, G. y D. Levanon. 1983. The influence of low density on food adaptation and cannibalism on growth of eels (*Anguilla anguilla*). *Bamidgeh*, 35: 53-60.
- Degani, G., D. Levanon y G. Trieger. 1984. Preliminary study on the influence of different feeds on mortality and growth of eels (*Anguilla anguilla* L.) reared on a expanded granulated diet. *Aquaculture*, 30: 155-172.
- Degani, G., D. Levanon y M.L. Gallagher. 1988. Relationship between growth, food conversion, body size, body composition and temperature in the European eel, *Anguilla anguilla* L. *Aquaculture and Fisheries Management*, 19: 139-143.
- Fernández-Delgado, C., J.A. Hernando, M. Herrera y M. Bellido. 1989. Age and growth of yellow eels, *Anguilla anguilla*, in the estuary of the Guadalquivir River (southwest Spain). *Journal of Fish Biology*, 34: 561-570.

- From, J. y G. Rasmuseen. 1984. A growth model, gastric evacuation and body composition in rainbow trout, *Salmo gairdneri* Richardson, 1836. *Dana*, 3: 61-139.
- García de Jalón-Lastra, D., M. Mayo-Rustarazo, F. Hervella-Rodriguez, E. Barceló-Culebras y T. Fernández-Couto. 1991. *Principios y técnicas de gestión de la pesca en aguas continentales*. Edita: Mundi-Prensa. Madrid.
- Gutiérrez-Estrada, J.C., I. Pulido-Calvo y C. Fernández-Delgado. 2000. Age-structure, growth and reproduction of the introduced pumpkinseed (*Lepomis gibbosa*, L. 1758) in a tributary of the Guadalquivir river (southern Spain). *Limnetica*, 19: 21-29.
- Hay, R.K.M. y A.J. Walker. 1989. *An introduction to the physiology of crop yield*. Edita: Longman Scientific and Technical. New York.
- Herrera, M. y C. Fernández-Delgado. 1992. The life-history patterns of *Barbus bocagei sclateri* (Günther, 1868) in a tributary stream of the Guadalquivir River basin, souther Spain. *Ecology of Freshwater Fish*, 1: 42-51.
- Hewett, S.W. y B.L. Johnson. 1987. *A generalized bioenergetics model of fish growth for microcomputers*. Edita: University of Wisconsin Sea Grant Institute. Wisconsin.
- Hsu, K., H.V. Gupta y S. Sorooshian. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research*, 31: 2517-2530.
- Jauncey, K. 1982. *Carp (Cyprinus carpio L.) nutrition, a review*. En: Muir, J.F. y R.J. Roberts. *Recent advances in aquaculture*. Edita: Croom Helm. Londres.
- Jobling, M. 1982. Some observations on the effects of feeding frequency of the food intake and growth of plaice, *Pleuronectes platessa* L. *Journal of Fish Biology*, 20: 431-444.
- Jover-Cerdá, M. 1993. Producción de anguilas. *Productos del Mar*, Junio: 78-84.
- Kamstra, A. 1993. The effect of size grading on individual growth in eel, *Anguilla anguilla*, measured by individual marking. *Aquaculture*, 112: 67-77.
- Kebus, M.J., M.T. Collins, M.S. Brownfield, C.H. Amundson, T.B. Kayes y J.A. Mallison. 1992. Effects of rearing density on the stress response and growth of rainbow trout. *Journal of Aquatic Animal Health*, 4(1): 1-6.
- Kuhlmann, H. 1979. The influence of temperature, food, initial size, and origin on the growth of elvers (*Anguilla anguilla* L.). *Conseil International pour l'Exploration de la Mer*, 174: 59-63.

- Mann, R.H.K. y J.H. Blackburn. 1991. The biology of the eel *Anguilla anguilla* (L.) in an English chalk stream and interactions with juvenile trout *Salmo trutta* L. and *Salmo salar* L. *Hydrobiologia*, 218(1): 65-76.
- Martínez-Garmendia, J. 1998. Simulation analysis of evolutionary response of fish populations to size-selective harvesting with the use of an individual-based model. *Ecological Modelling*, 111: 37-60.
- Más-Álvarez, B. y R. Barrera-Orozco. 1989. *Producción de anguilas*. Hojas divulgadoras (Nº: 14/89 HD) del Ministerio de Agricultura Pesca y Alimentación. Madrid.
- McDonald, M.E., C.A. Tikkanen, R.P. Axler, C.P. Larsen y G. Host. 1996. Fish simulation culture model (FIS-C): a bioenergetics based model for aquacultural wasteload application. *Aquacultural Engineering*, 15(4): 243-259.
- McNown, W. y A. Seireg. 1983. Computer aided optimum design and control of staged aquaculture system. *Journal of World Mariculture Society*, 14: 417-433.
- Olsen, O.A. y J.G. Balchen. 1992. Structured modeling of fish physiology. *Mathematical Biosciences*, 112: 81-113.
- Paessun, M.A. y R. Allison. 1984. Maximising tilapia production in recirculating system by sequential rearing. *Aquaculture*, 42: 185-189.
- Pedersen, M.I. 1998. Recapture rate, growth and sex of stocked cultured eels *Anguilla anguilla* L. *Bulletin Francais de la Peche et de la Pisciculture*, 349: 153-162.
- Pulido-Calvo, I., J. Roldán, R. López-Luque y J.C. Gutiérrez-Estrada. 2002. Técnicas de predicción a corto plazo de la demanda de agua. Aplicación al uso agrícola. *Ingeniería del Agua*, 9(3): 319-331..
- Rossi, R., A. Carrieri, P. Franzoi, G. Cavallini y A. Gnes. 1988. A study of eel (*Anguilla anguilla* L.) population dynamics in the Comacchio lagoons (Italy) by mark-recapture method. *Oebalia*, 14: 1-14.
- Shepherd, J. y N. Bromage. 1999. *Piscicultura intensiva*. Edita: ACRIBIA, S.A. Zaragoza.
- Sparre, P. 1976. A markovian decision process applied to optimization of production planning in fish farming. *Meddel Danmarks fisk Havunders*, 7: 275-317.
- Summerfelt, S.T., J.A. Hankins, S.R. Summerfelt y J.M. Heinen. 1993. Modeling continuous culture with periodic stocking and selective harvesting to measure the effect on productivity and biomass capacity of fish culture system. *Proceedings of an Aquacultural Engineering Conference, Washington*: 581-595.

- Van Nes, E.H., E.H.R.R. Lammens y M. Scheffer. 2002. PISCATOR, an individual-based model to analyze the dynamics of lake fish communities. *Ecological Modelling*, 152: 261-278.
- Watten, B.J. 1992. Modelling the effects of sequential rearing on the potential production of controlled environment fish-culture systems. *Aquacultural Engineering*, 11: 33-46.
- Wickins, J.F. 1985. Growth variability in individually confined elvers, *Anguilla anguilla* (L.). *Journal of Fish Biology*, 27: 469-478.
- Wohlfarth, G.W. 1977. Shoot carp. *Bamidgeh*, 29: 35-56.
- Wootton, R.J. 1991. *Ecology of teleost fishes*. Edita: Chapman and Hall. Londres.



## 5.6. ANEXO V.1. GLOSARIO

$\alpha$	Parámetro de variación del modelo de Arnason-Sparre
$\beta$	Parámetro temporal del modelo de Arnason-Sparre
$\gamma$	Parámetro de peso del modelo de Arnason-Sparre
$\mu$	Peso medio
$\sigma$	Desviación típica o estándar
$a_t$	Factor de amortización
$B_c$	Biomasa de cosecha
$C_I$	Inversión inicial
CV	Coefficiente de variación
DT	Pendiente de la recta de regresión entre el peso medio y su desviación típica
$f(x)_B$	Función de distribución de biomasa total
$f(x)_k$	Función de densidad del tanque o subgrupo <b>k</b>
$FC_{Am}$	Factor de crecimiento debido a la concentración de amoníaco
$FC_{Na}$	Factor de crecimiento debido a la concentración de nitratos
$FC_{Ni}$	Factor de crecimiento debido a la concentración de nitritos
$FC_{pH}$	Factor de crecimiento debido al valor del pH
FCR	Factor de conversión del alimento
$FC_T$	Factor de crecimiento debido a la temperatura del agua
$G_t^{FC}$	Tasa de crecimiento en el momento <b>t</b> teniendo en cuenta todos los <b>FCs</b>
<b>K</b>	Factor de condición
$K_v$	Constante del modelo de Von Bertalanffy
$L_\infty$	Longitud asintótica en el modelo de Von Bertalanffy
<b>N</b>	Número de ejemplares de una cohorte

n	Número de ejemplares de un tanque
PF	Peso fresco
r	Tasa de interés anual
T	Temperatura
TAT	Menor tasa de alimentación que a la temperatura <b>T</b> produce el máximo crecimiento
$W_c$	Talla de cosecha
x	Variable aleatoria. Talla individual
Z	Tasa de mortalidad

## 5.7. ANEXO V.2. MANUAL DEL USUARIO DEL PROGRAMA OPERATIVO DE SIMULACIÓN (POS)

Como se ha comentado en los apartados anteriores del capítulo 5, el Programa Operativo de Simulación (POS) permite la simulación de multitud de escenarios y estrategias de producción.

### V.2.1. INICIO DEL PROGRAMA

El inicio de POS puede llevarse a cabo de dos formas distintas. Así, si nos encontramos dentro de CONTROLPLANT, se puede iniciar a partir de la barra de herramientas, opción SPI, donde podremos hacer 'click' en POS. De esta forma, al igual que en SEDPA o el propio CONTROLPLANT, aparecerá una ventana donde se requiere al usuario una clave de entrada. Esta misma petición se producirá si se inicia el programa desde el icono presente en el escritorio de Windows (figura AV.1).



Figura AV.1. Icono del Programa Operativo de Simulación en el escritorio de Windows®

Una vez introducida la clave de acceso correcta nos aparecerá la ventana principal de POS (figura AV.2).

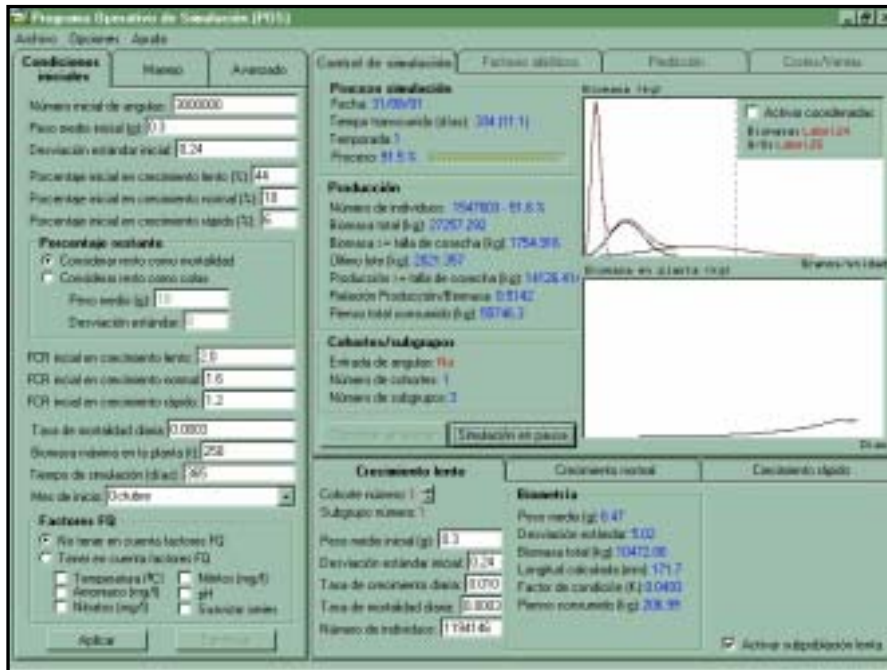


Figura AV.2. Ventana principal de POS



## V.2.2. VENTANA PRINCIPAL

La ventana principal de **POS** se encuentra dividida en tres tableros o paneles múltiples, presentando de esta forma varios cuadros de diálogo o pantallas de información en un único formulario y usando la misma interfaz que aparece en otras muchas aplicaciones de Microsoft Windows®. Un panel múltiple proporciona un grupo de fichas, cada una de las cuales actúa a modo de contenedor para otros controles. El primer panel múltiple (situado en el margen izquierdo de la ventana principal) contiene un total de tres fichas denominadas: *Condiciones iniciales*, *Manejo* y *Avanzado*. El segundo panel múltiple (cuadro superior derecho) contiene cuatro fichas: *Control de simulación*, *Factores abióticos*, *Predicción* y *Costes/Ventas*. El tercero nuevamente contiene tres fichas: *Crecimiento lento*, *Crecimiento normal* y *Crecimiento rápido*.

### V.2.2.1. Condiciones iniciales

En esta ficha el usuario puede cambiar las condiciones iniciales en las que se va a desarrollar la simulación (figura AV.3). Éstas quedan definidas mediante el número

Figura AV.3. Ficha de Condiciones iniciales

inicial de angulas, su peso medio y desviación estándar inicial del lote de entrada, la distribución porcentual del número de individuos dentro de cada subgrupo de crecimiento (lento, normal y rápido), el porcentaje de ejemplares que muere inicialmente o que permanece en estado de crecimiento latente, el factor de conversión del alimento medio para cada subgrupo de crecimiento, la tasa de mortalidad diaria, la capacidad de carga de la planta, el número de días de simulación, el mes de inicio de la simulación y la aplicación o no de factores físico-químicos. En este último apartado, es posible seleccionar uno o varios factores físico-químicos entre la temperatura del agua, la concentración de amoníaco, la concentración de nitritos, la concentración de nitratos y el pH. Además siguiendo el procedimiento adoptado en el capítulo 4, existe la posibilidad de filtrar o suavizar previamente las series de datos antes de ser utilizadas en la simulación.

Finalmente, en esta misma ficha se han incluido los botones *Aplicar* y *Continuar*. El primero de ellos indica a **POS** que las condiciones de simulación han sido verificadas por parte del usuario. El segundo, que aparece

no activo por defecto, sólo queda habilitado cuando las condiciones iniciales son cambiadas durante el proceso de simulación, permitiendo la finalización del proceso tras el cambio.

### V.2.2.2. Manejo

La extensión de las condiciones de simulación se completan por una parte con la segunda ficha del primer panel múltiple o ficha de *Manejo* (figura AV.4), y por otra con la cuarta ficha del segundo panel múltiple (situado en el cuadro superior derecho de la ventana principal) o ficha *Costes/Ventas*.

Condiciones iniciales	<b>Manejo</b>	Avanzado
Peso medio de cosecha (g): 150		
<b>Tipo de cosecha</b>		
<input checked="" type="radio"/> Lote mínimo de cosecha (kg): 2000		
<input type="radio"/> Intervalo de cosecha (días): 20		
<b>Cohortes</b>		
Número de cohortes: 1		
Tiempo transcurrido en incorporar una nueva cohorte desde el comienzo de la temporada (días): 15		
<b>Distribución de talla por zonas</b>		
Zona de alevinaje hasta (g): 40		
Zona de preengorde hasta (g): 120		

Figura AV.4. Ficha de Manejo

En la ficha de manejo se definen características de la simulación relacionadas con el manejo de los ejemplares. De esta forma, es posible establecer el peso medio de cosecha, el tipo de cosecha, en el que se puede escoger entre un lote mínimo de cosecha o un intervalo de días que transcurre entre cosecha y cosecha, el número de cohortes que van a coexistir en la planta así como el número de días transcurridos desde el comienzo de la temporada hasta que la nueva cohorte es introducida, y la distribución de tallas por zonas. Como se comentó anteriormente, estas características quedan fijadas cuando el usuario hace 'click' sobre el botón *Aplicar* de la ficha de *Condiciones iniciales*.

### V.2.2.3. Avanzado

Esta ficha se encuentra actualmente vacía a la espera de ser completada con información complementaria que pueda ser incluida en la simulación.

#### V.2.2.4. Control de simulación

La función básica de la ficha de *Control de simulación* es la de iniciar o parar el proceso de simulación y la de supervisar cuantitativamente y gráficamente la evolución de la simulación en transcurso. Para ello, en esta ficha se pueden identificar en su margen izquierdo tres cuadros identificados como: *Proceso de simulación*, *Producción* y *Cohortes/subgrupos*. En la zona derecha hay dos gráficos en los que se puede observar la biomasa de cada subgrupo de crecimiento en relación al peso fresco de los ejemplares (gráfico superior), y la biomasa total almacenada en la planta en función del tiempo (gráfico inferior). Finalmente, en la parte inferior izquierda tenemos el botón *Comenzar simulación* que inicia el proceso (figura AV.2).

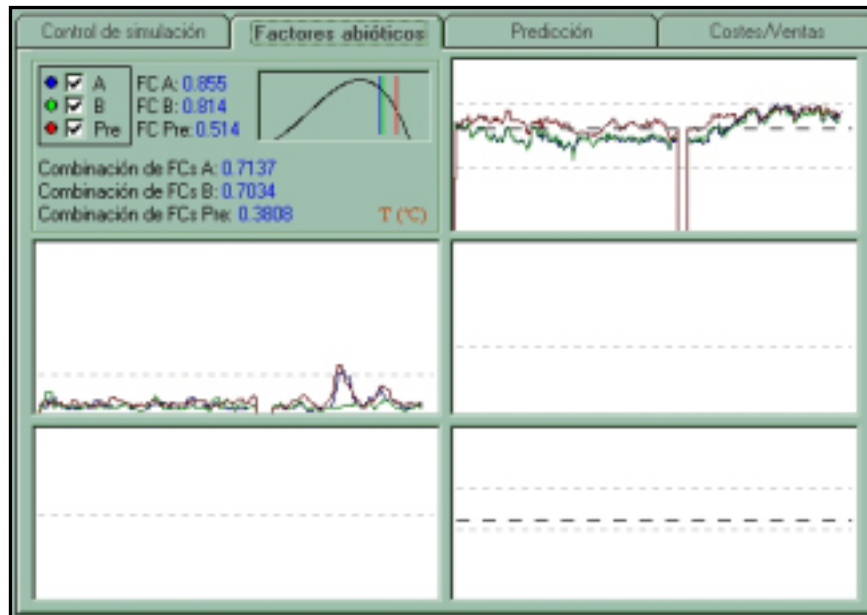
Dentro del primero de los cuadros (*Proceso de simulación*) hay una serie de etiquetas que indican como evoluciona la simulación. También se incluye una barra de proceso mediante la cual es fácil saber con rapidez el porcentaje de simulación transcurrido así como lo que resta del proceso.

El cuadro de *Producción* aporta al usuario información cuantitativa sobre el número de individuos presentes, biomasa total, biomasa almacenada con una talla igual o superior a la talla de cosecha especificada en las condiciones de manejo, biomasa en el último lote de cosecha, producción total, relación producción/biomasa y la cantidad total de pienso consumido.

Finalmente, el cuadro de *Cohortes/subgrupos* muestra la posibilidad o no de entrada de angulas a la planta, el número de cohortes presentes así como el número de subgrupos totales.

#### V.2.2.5. Factores abióticos

En la ficha de *Factores abióticos* (figura AV.5) se puede identificar un cuadro de diálogo no titulado cuyo interior muestra varias etiquetas relacionadas con el cálculo de los factores de crecimiento. El cálculo de los factores de crecimiento se puede controlar de dos formas distintas. Por una parte por la inclusión o no del parámetro físico-químico seleccionado en el cuadro de *Factores FQ* de la ficha de *Condiciones iniciales*, y por otra por la selección o no de la serie a la que queremos que afecte (serie A, B, E o F). Esto se realiza mediante la selección de dichas series en la parte superior izquierda del cuadro de diálogo incluido en esta ficha. El resto de la ficha está ocupado por un total de cinco gráficos cada uno de los cuales se corresponde con cada uno de los cinco posibles factores físico-químicos tenidos en cuenta en la simulación. Cada gráfica muestra el valor real o suavizado de cada parámetro físico-químico. Asimismo en el cuadro de diálogo se incluye una pequeña gráfica que indica al usuario el tipo de función utilizada en el cálculo de los factores de crecimiento.



**Figura AV.5.** Ficha de Factores abióticos. La gráfica superior derecha es de temperaturas mientras que la de la izquierda representa amoniaco

### V.2.2.6. Predicción

En la ficha *Predicción* (figura AV.6) tan solo es posible identificar cuatro cuadros de diálogo que se corresponden con las predicciones de las condiciones físico-químicas de las series A, B, E y F a través de los mejores modelos validados en el capítulo 4, así como con la conexión de **POS** con las reglas **A-A** de **CONTROLPLANT**. De esta forma, en cada cuadro de dialogo se puede identificar el mejor modelo para cada serie y para cada parámetro.



**Figura AV.6.** Ficha de Predicción

En el último cuadro de diálogo de esta ficha (*SEDPA 1.0/Reglas A-A de CONTROLPLANT*) se muestra la conexión de **POS** con el resto de aplicaciones desarrolladas y se indica la posibilidad de aparición de las distintas patologías.

### V.2.2.7 Costes/Ventas

En esta ficha (figura AV.7) **POS** muestra al usuario todas las características de la simulación relacionadas con los parámetros económicos y que permiten llevar a cabo el balance inicial. Se pueden identificar un total seis cuadros de diálogo: *Personal*, *Biomasa*, *Alimentación*, *Energía (mercado liberalizado)*, *Amortizaciones* y *Ventas*. En el primero de ellos el usuario puede modificar el número de empleados (operarios y técnicos) así como los salarios brutos correspondientes. En el segundo, se define el precio en el mercado del kilogramo de angulas (0.3 gramos). En el tercero es posible seleccionar la marca comercial de pienso extruído así como editar los precios para las distintas granulometrías (figura AV.8). En el cuadro de *Energía (mercado liberalizado)* el usuario puede establecer la cantidad media de energía consumida por mes así como el precio del kWh. En el penúltimo cuadro de diálogo (*Amortización*), **POS** pide al usuario la introducción de los parámetros de amortización, es decir, la vida útil de la instalación, la tasa de interés anual y la inversión inicial. Finalmente, en el último cuadro de diálogo (*Ventas*) se puede establecer el precio de venta del kilogramo de angulas con un tamaño igual o superior al establecido como tamaño comercial.

Control de simulación	Factores abióticos	Predicción	Costes/Ventas
<b>Personal</b> Número de técnicos: <input type="text" value="2"/> Salario bruto por técnico (€/mes): <input type="text" value="2200"/> Número de operarios: <input type="text" value="5"/> Salario bruto por operario (€/mes): <input type="text" value="1085"/>		<b>Amortizaciones</b> Vida útil de la instalación (años): <input type="text" value="30"/> Tasa de interés anual (%): <input type="text" value="5"/> Inversión inicial (€): <input type="text" value="1500000"/> Factor de amortización: 0.065	
<b>Biomasa</b> €/kg angula 0.3 gramos: <input type="text" value="150.6"/>		<b>Ventas</b> €/kg angula >= 150 gramos: <input type="text" value="6.61"/>	
<b>Alimentación</b> Pienso extruído marca: <input type="text" value="Dibaq-Diprotec"/> <input checked="" type="checkbox"/> Aplicar tabla de conversión del fabricante <input type="button" value="Editar precios"/>		<input checked="" type="checkbox"/> Considerar amortización <input type="checkbox"/> Mostrar balance anual Personal (€/día): 327.5 Biomasa (€/día): 0.0 Alimentación (€/día): 221.2 Energía (€/día): 127.5 Amortización (€/día): 267.3 Costes totales (€/día): 943.6 Ventas (€): 66452.1	
<b>Energía (mercado liberalizado)</b> Energía consumida media (kWh/mes): <input type="text" value="75000"/> Precio del kWh (€/kWh): <input type="text" value="0.051"/>		<input type="text" value="-255201.5 €"/> <small>-42 461 949</small>	

Figura AV.7. Ficha de Costes/Ventas

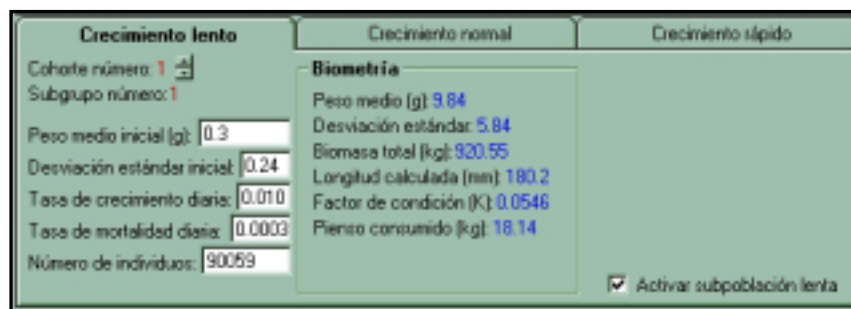


**Figura AV.8.** Ventana de edición del tipo y precios del pienso extruído

Independientemente de los cuadros de diálogo anteriores, en la parte inferior derecha de esta ficha se muestran una serie de etiquetas mediante las cuales el usuario puede observar la evolución durante la simulación de cada uno de los costes, así como los costes totales y los ingresos debidos a las ventas.

### V.2.2.8. Crecimiento lento, normal y rápido

En el último panel múltiple se pueden identificar las fichas *Crecimiento lento*, *Crecimiento normal* y *Crecimiento rápido* (figura AV.9).



**Figura AV.9.** Fichas de Crecimiento lento, normal y rápido

Las tres son exactamente iguales excepto en que en la primera hay un selector de cohortes que permite ver de forma individual una u otra cohorte. Por lo demás, en las tres fichas se indican las características inherentes a cada subgrupo de crecimiento correspondiente a cada cohorte. Así, en la parte izquierda de cada ficha se indica el peso medio y su desviación inicial, la tasa de crecimiento y mortalidad diaria y el número de individuos. Por otra parte, en la zona central de cada ficha se puede observar un cuadro

de diálogo denominado *Biometría*. En este cuadro de diálogo el usuario puede comprobar la variación durante la simulación del peso medio y su desviación típica, la biomasa total correspondiente, la longitud calculada para el peso medio del subgrupo y el factor de condición (**K**).

#### **V.2.2.9. Barra de herramientas**

En la barra de herramientas de **POS** tan solo pueden identificarse tres opciones: *Archivo*, *Opciones* y *Ayuda* (figura AV.2). En la primera de ellas (*Archivo*) el usuario puede elegir entre *Imprimir* y *Salir*. La opción *Imprimir* genera un archivo impreso en el que se muestran todas las características y propiedades de la simulación en curso. La opción *Salir* cierra la aplicación y devuelve el control del sistema a **CONTROLPLANT** en el caso de que se haya accedido a **POS** desde esta otra aplicación.

La opción *Opciones* muestra la posibilidad *Cambiar condiciones al comienzo de la temporada*, lo que permite a **POS** la detención automática de la simulación al inicio de la siguiente temporada para que el usuario pueda cambiar algunas condiciones de la simulación. La reanudación de la simulación se lleva a cabo mediante el botón *Continuar* situado en la parte inferior de la ficha *Condiciones iniciales* del primer panel múltiple.

Finalmente, la opción *Ayuda* es la típica de cualquier otra aplicación. Al hacer ‘click’ sobre ella aparecen dos nuevas opciones: *Ayuda* y *Acerca de...*

## **CAPÍTULO 6**

---

### **CONCLUSIONES**





## CAPÍTULO 6. CONCLUSIONES

1. El sistema experto **SEDPA** es una aplicación potencialmente útil en el diagnóstico de las distintas enfermedades y agentes patológicos que afectan a las explotaciones intensivas de anguilas europeas.
2. La inclusión en el motor de inferencia de **SEDPA** de un método de asignación de creencias individuales mediante lógica borrosa y la facilidad de combinación de eventos (síntomas/lesiones) altamente difusos a través de la teoría de Dempster-Shafer, ha proporcionado al sistema una capacidad de discriminación de patologías significativamente mayor a la mostrada por los expertos humanos.
3. El test  $\chi^2$  modificado ha funcionado como elemento de selección de las creencias significativas asignadas a patologías en el marco de la teoría de Dempster-Shafer.
4. El test basado en la generalización bivariante del test no paramétrico de Cramér Von Mises aplicado sobre las memorias asociativas borrosas (**MABs**), ha permitido la detección *a priori* de la existencia de diferencias significativas entre los diagnósticos asociados a dichas **MABs**.
5. Las conclusiones del subsistema sugerencia han sido relativamente satisfactorias, intuyéndose una relación entre la respuesta aportada y el volumen de información contenida en las bases de Conocimiento del Dominio Principal.
6. La obtención de conclusiones en ausencia de reglas establecidas como código en la corriente de control del sistema, facilita su adaptación a distintas condiciones de explotación. Sin embargo, esta misma propiedad impide la aportación al usuario de explicaciones robustas sobre el proceso de razonamiento.
7. La utilización de modelos estadísticos clásicos junto con las Redes Neuronales Computacionales, han proporcionado una capacidad aceptable de predicción a muy corto plazo de parámetros físico-químicos (temperatura del agua, amoníaco, nitritos, nitratos y pH), al encontrarse errores estándar de predicción (**SEP**) y valores de eficiencia (**E**) que en todo momento están alrededor del 7 % y 0.9, respectivamente.
8. Excepto para la temperatura del agua, las series originales de los parámetros físico-químicos tienden a comportarse como series de ruido blanco para la frecuencia de muestreo considerada.
9. El filtrado previo de la serie original (independientemente del parámetro tratado), normaliza la serie favoreciendo, en general, un mejor ajuste de los modelos de regresión múltiple y modelos univariantes de series temporales (suavizados y **ARIMA**).
10. En general, los modelos **ARIMA** proporcionan niveles de error (**SEP** y **E**) inferiores en el proceso de validación externa. Sin embargo, porcentualmente las Redes

Neuronales tienden a reducir más el error cometido que los modelos **ARIMA** cuando se lleva a cabo una extensión de los modelos calibrados para una serie y un parámetro, hasta otra serie, siendo la correlación entre las series originales baja y no significativa.

11. El Programa Operativo de Simulación (**POS**) es una herramienta potencialmente útil para la selección de la estrategia de producción.

12. La utilización de modelos de crecimiento conjuntamente con el método de extensión de la varianza propuesto por Arnason y Sparre, permite llevar a cabo simulaciones de la variación del *stock* en la planta.

13. La incorporación de los factores de crecimiento debidos a las condiciones físico-químicas proporcionan a la simulación un carácter más realista, aproximando los resultados simulados en los puntos críticos a la realidad.

14. La integración de **CONTROLPLANT**, **SEDPA** y **POS**, es una aplicación que permite el control y la predicción de los principales parámetros físico-químicos, biológicos y biométricos en una explotación intensiva de anguilas.