

teledetección: teoría y aplicaciones



carlos pinilla

Índice

Presentación	iii		
I PARTE: TEORÍA			
1. Panorámica de la teledetección	01-01		
Introducción	01-01		
Bosquejo histórico de la teledetección	01-03		
Campos de aplicación	01-06		
2. La radiación electromagnética	02-01		
Introducción	02-01		
El campo electromagnético	02-01		
El movimiento ondulatorio	02-04		
La onda electromagnética	02-05		
La teoría cuántica	02-06		
El espectro electromagnético	02-06		
Generación de la radiación	02-09		
Leyes de la radiación electromagnética	02-10		
<i>El cuerpo negro</i>	02-10		
<i>Leyes de la radiación</i>	02-11		
Terminología radiométrica	02-13		
3. Interacción de la radiación con la materia	03-01		
Introducción	03-01		
Interacción macroscópica	03-01		
Emisividad	03-04		
Interacción atómica y molecular	03-06		
4. Interacción atmosférica	04-01		
Introducción	04-01		
Componentes de la atmósfera	04-01		
Dispersión atmosférica	04-04		
Absorción atmosférica	04-06		
Óptica atmosférica	04-08		
Transferencia radiativa	04-10		
Efecto atmosférico	04-12		
Nubes	04-16		
5. Respuesta espectral de la superficie	05-01		
Introducción	05-01		
El suelo	05-02		
Características reflexivas del suelo desnudo	05-09		
La vegetación	05-15		
Respuesta espectral del sistema suelo-vegetación	05-19		
6. La imagen digital	06-01		
Introducción	06-01		
Sistemas	06-01		
La imagen digital	06-03		
La función de dispersión puntual	06-05		
Grabación	06-09		
<i>Ruido</i>	06-10		
<i>Formatos</i>	06-10		
Calibración	06-11		
7. Sensores y plataformas	07-01		
Introducción	07-01		
Resolución	07-01		
<i>Resolución espacial</i>	07-02		
<i>Resolución espectral</i>	07-02		
<i>Resolución radiométrica</i>	07-03		
<i>Resolución temporal</i>	07-04		
<i>Efemérides astronómicas del satélite</i>	07-06		
<i>Relación entre las distintas resoluciones</i>	07-07		
Tipos de sensores	07-08		
		Misiones y programas de teledetección	07-10
8. Tratamiento digital de imágenes	08-01		
Introducción	08-01		
Distorsiones	08-01		
<i>Distorsiones geométricas</i>	08-01		
<i>Distorsiones radiométricas</i>	08-02		
Operadores	08-03		
Estadísticos básicos de la imagen	08-04		
Transformaciones radiométricas	08-05		
Correcciones radiométricas	08-07		
Correcciones atmosféricas	08-10		
Realce de imágenes	08-13		
<i>Operaciones con el contraste</i>	08-13		
9. Empleo del color	09-01		
Introducción	09-01		
El color	09-01		
Síntesis del color	09-02		
Paletas de color	09-03		
Imágenes en color compuesto	09-03		
Imágenes en pseudocolor	09-05		
Otros sistemas de especificación del color	09-05		
10. Realce geométrico	10-01		
Introducción	10-01		
Cambios de escala	10-01		
El filtrado digital	10-02		
Filtrado en el dominio de la frecuencia	10-09		
Restauración de imágenes	10-12		
11. Georreferenciación	11-01		
Introducción	11-01		
Métodos polinómicos de corrección	11-01		
Otros métodos de corrección	11-06		
Control de calidad del ajuste	11-07		
Ortorrectificación	11-08		
Remuestreo	11-09		
Modelización matemática	11-11		
12. Álgebra de imágenes	12-01		
Introducción	12-01		
Análisis de componentes principales	12-01		
Transformación <i>tasseled cap</i>	12-05		
Índices de vegetación	12-07		
13. Clasificación digital	13-01		
Introducción	13-01		
La imagen como variable aleatoria multivariante	13-01		
Entrenamiento del clasificador	13-03		
Análisis de separabilidad	13-06		
Fase de asignación	13-08		
Control de calidad temático	13-14		
14. Teledetección multitemporal	14-01		
Introducción	14-01		
Métodos de normalización radiométrica	14-02		
Técnicas de detección de cambios	14-06		
15. Metodología en teledetección	15-01		
Introducción	15-01		
Decisiones iniciales	15-01		
Fases del proyecto de teledetección	15-02		
Apéndice 1. Glosario	A1-01		

Presentación

Los archivos que el lector encontrará en esta carpeta son parte del resultado de la experiencia docente acumulada durante 28 años en la Escuela Politécnica Superior de Jaén. Durante este tiempo se han impartido casos de las titulaciones de Ingeniería en Geodesia y Cartografía, Ingeniería Técnica en Topografía y, más recientemente en Ingeniería Geomática y Topográfica, bajo distintas denominaciones. Sin embargo, el esquema seguido siempre ha sido el de establecer tres niveles en el aprendizaje de las herramientas propias de la teledetección: los fundamentos, dentro de una asignatura de *teledetección elemental*, conocimientos más elaborados, ampliados o profundos, dentro de una *teledetección avanzada*, y, finalmente, el mundo de aplicaciones, tanto en el ámbito de la investigación como el profesional, bajo la denominación de *teledetección aplicada*.

Los siguientes capítulos abarcan los tres niveles, reunidos en dos partes, *teoría* y *aplicaciones*, que se ofrecen para su libre distribución y estarán en continuo proceso de revisión y actualización en contenidos e ilustraciones.

El formato de edición pretende adaptarse tanto a su impresión en formato A4 como a la lectura directa en la mayoría de los monitores.

Marzo de 2020
Carlos Pinilla.

1

Panorámica de la teledetección

Introducción

El término *teledetección* fue el mayoritariamente adoptado en los países de habla hispana como traducción de la expresión inglesa “*remote sensing*”, utilizándose a partir del año 1960 para describir las observaciones de un objeto efectuadas sin mediar contacto físico con él. Otras traducciones se refieren a la disciplina como *percepción remota* o *sensores remotos*.

Este tipo de adquisición de información solamente puede ser posible mediante la detección y medida de los cambios que el objeto observado induce en su entorno, bien sean en forma de perturbaciones electromagnéticas, bien como ondas acústicas reflejadas o alteradas, o como perturbaciones del campo gravitatorio debidas a la presencia del objeto. Con carácter más restrictivo, el término teledetección se reserva a las técnicas electromagnéticas de adquisición de información en las zonas del espectro electromagnético comprendidas entre las ondas de radio de baja frecuencia y los rayos gamma.

Los sistemas de teledetección se caracterizan por la concurrencia de los siguientes procesos:

- La emisión de radiación electromagnética desde una fuente natural o artificial.
- La interacción de la radiación con la superficie observada.
- La interacción de la radiación con la atmósfera, de haberla, como en el caso de la Tierra.
- La recepción de las ondas electromagnéticas en instrumentos a bordo de una plataforma y su transformación en señal electrónica.

- La transmisión de la señal a tierra, su almacenamiento y su distribución.
- El tratamiento digital de señal, su transformación en información útil y la interpretación final de los datos.

El uso de plataformas ha sufrido un cambio esencial con la incorporación de los satélites artificiales. Esto ha permitido una concepción más general de los temas de estudio, pudiendo abordarse proyectos a escala regional, continental o incluso global.

En el proceso de teledetección intervienen numerosas ramas de la ciencia y de la tecnología. Corresponde a la ingeniería electrónica el diseño de los sensores, en colaboración con los futuros usuarios, que serán quienes impongan los requerimientos del producto final, la imagen. El diseño del satélite a bordo del cual se instalarán los instrumentos de adquisición y su puesta en órbita es el ámbito propio de la ingeniería aeroespacial. La ingeniería de telecomunicaciones se encarga de recolectar, codificar, comprimir y transmitir los datos. En el segmento de tierra, es misión de la ingeniería geomática el procesamiento, manipulación, análisis e interpretación de la señal.

Diversas ingenierías y ciencias son usuarios finales de los datos. Cada vez más ámbitos científicos y técnicos hacen uso de las imágenes proporcionadas por los satélites para la investigación, el desarrollo, la explotación y la toma de decisiones. Entre ellos se encuentran la ingeniería agronómica y forestal, la biología, la física, la geografía, la geología, las ciencias ambientales, etc.

Hasta hace relativamente pocos años existía un debate acerca de cuáles eran los límites entre la teledetección y disciplinas como la fotogrametría y la fotointerpretación. Con el desarrollo en las últimas décadas de los sistemas de adquisición y de las técnicas de tratamiento, los campos se fueron definiendo hasta configurarse disciplinas independientes y bien consolidadas. Dejando al margen la fotointerpretación, cuyos principios básicos siguen siendo vigentes e incluso aplicables a la teledetección, la fotogrametría incluía los sistemas y técnicas de adquisición de imágenes, en sus inicios básicamente fotográficas, desde aeronaves, dejando a la teledetección las imágenes orbitales tomadas desde satélite. La tipología de los productos propios de cada técnica era bien diferenciada. La fotogrametría entonces trabajaba sobre emulsiones fotográficas y papel, en tanto que las imágenes de satélite, exceptuadas algunas misiones puramente fotográficas, eran mayoritariamente digitales.

Sin embargo, existían campos en los que la distinción era confusa. Sobre aeronaves se instalaban sistemas de radar lateral cuyo estudio y explotación recaía dentro del ámbito de la teledetección, mientras que se comenzaban a utilizar cámaras digitales aerotransportadas que emulaban los sensores satelitales o se empleaban pares estereoscópicos de imágenes de satélite para obtener modelos digitales de elevación, con lo cual se desdibujaban las fronteras trazadas inicialmente.

En la actualidad, el debate se ha ventilado no insistiendo en delimitar los campos de actuación de forma nítida. Los instrumentos, las metodologías y los fines son cada vez más compartidos entre la fotogrametría y la teledetección, con independencia de la plataforma desde la que se adquieran los datos. A fin de cuentas, *todo es teledetección*. Pero la práctica ha

ido consolidando un cuerpo de doctrina según el cual, la fotogrametría se ocupa de fines eminentemente geométricos, aunque no descarta los temáticos, y la teledetección persigue caracterizar la naturaleza de la superficie observada y sus propiedades físicas y químicas, sin olvidar obviamente el rigor geométrico de los productos.

La transformación de los datos brutos en información útil requiere de una serie de operaciones de tratamiento de la imagen más o menos complejas en función del fin perseguido y de la naturaleza de los datos originales. Así pues, la utilización de los datos brutos capturados mediante radar requiere de operaciones muy elaboradas para hacer de ellos una imagen interpretable. Los datos procedentes de la observación en el espectro visible, sin embargo, revisten menos complicación y su explotación necesita solamente de la corrección previa de las imágenes y de un tratamiento digital más ligero.

Precisamente, el factor distintivo entre la teledetección y la fotointerpretación clásica es el análisis digital de la imagen como forma de extraer información a partir del brillo de cada celda del terreno. La investigación en este campo está posibilitando exprimir hasta extremos, en muchas ocasiones sorprendentes, los datos brutos en busca de nuevas utilidades.

Por otro lado, la teledetección necesita también del contraste en campo, modelizando los datos adquiridos desde el satélite, a fin de establecer las funciones de conversión entre los niveles digitales medidos y el fenómeno a detectar. Por ello, es de esencial importancia conocer el comportamiento espectral de cada cubierta observable, lo cual requiere de un sólido bagaje físico que permita comprender la reacción de las superficies a la incidencia de la radiación electromag-

nética. La complejidad de los estudios de campo necesarios para alcanzar ese conocimiento también varía con el objetivo de la actividad de teledetección propuesta, existiendo un nivel apropiado para cada fin. Por ejemplo, la clasificación de cubiertas precisa de simples estudios de campo orientados a la separación en clases temáticas discretas. En cambio, proyectos de más envergadura técnica requieren abarcar toda la gama de estudios, desde el nivel físico hasta el temático.

En resumen, la teledetección reúne bajo un mismo punto de vista campos tan heterogéneos como los

fundamentos físicos de la respuesta espectral de los cuerpos, el proceso de corrección de las imágenes y el tratamiento digital de la información, bajo el cual se engloban las operaciones que realmente permiten la identificación de las superficies y la cuantificación de los fenómenos naturales o antrópicos observados.

La irrupción de los satélites artificiales a finales de la década de 1950 ha permitido abordar de forma más general los temas de estudio, pudiendo acometerse proyectos a escala regional, continental o incluso global.

Bosquejo histórico de la teledetección



Cartel promocional de Nadar.

El desarrollo temprano de la teledetección como campo científico está estrechamente ligado al desarrollo de la fotografía. Las primeras imágenes indelebles producidas por acción de la luz directa fueron las llamadas *heliografías*, efectuadas por el francés Niépce en 1822. Atendiendo a la formación de imágenes, pueden ser consideradas como las precursoras de cualquier sistema de percepción remota. El también francés Daguerre mejoró el procedimiento en 1839 que, a pesar de todo, seguía sin ser suficientemente operativo. Por esta razón, durante la siguiente década fue superado por el de Talbot, cuyo proceso negativo-positivo estableció las bases de la moderna fotografía.

Si se toma en consideración el antecedente de la fotografía aérea como precedente directo de la teledetección, podrá establecerse el inicio de esta técnica en el año 1858, en el que, desde la canasta de un globo aerostático, Gaspar F. Tournachon («Nadar») tomó las primeras fotografías de París.

El año siguiente el coronel francés Laussedat comenzó a utilizar fotografías para la confección de planos topográficos, aunque desde 1846 se venían empleando perspectivas dibujadas por medio de una cámara clara con ese propósito. En 1859 Laussedat consiguió reemplazar la cámara clara por la cámara oscura hasta construir un prototipo de fototeodolito. El procedimiento se conoce como *fotogrametría de intersección*.

Por su parte, el desarrollo de la aviación fue rápidamente aprovechado por la fotografía como eficaz medio de extensión de la experiencia que suponía volar en esa época. La primera fotografía desde un avión fue tomada por W. Wright en 1909, si bien la primera cámara aérea como tal no se desarrolló hasta la Primera Guerra Mundial, en 1915. La mejora de las emulsiones empleadas, el perfeccionamiento de las cámaras y la investigación sobre las propias plataformas de vuelo permitieron durante la Segunda Guerra Mundial el asentamiento de la fotografía aérea y el nacimiento de la fotointerpretación.

Durante la década de los 50 se desarrollan los primeros satélites de observación como medio de salvaguarda de la paz en pleno período de *guerra fría* entre los Estados Unidos de Norteamérica (EEUU) y la antigua Unión de Repúblicas Socialistas Soviéticas (URSS). El lanzamiento del Sputnik en 1957 por parte de la URSS marcó el hito. De forma casi inmediata los sectores civiles reaccionaron apostando por la utilización pacífica de la información proporcionada por los sensores orbitales. Fruto de ello fue la primera generación de satélites de observación meteorológica TIROS (*Television and InfraRed Observation Satellite*), de carácter experimental, cuya primera unidad fue lanzada por los Estados Unidos el 1 de abril de 1960. A partir del décimo satélite de la serie en 1965, el programa fue asumido por la ESSA (*Environmental Science Services Administration*), poniéndose en órbita los primeros satélites meteorológicos auténticamente operativos, desde el ESSA-1, en 1966, hasta el ESSA-9, en 1969. A esta serie le siguió en 1970 la muy conocida NOAA (*National Oceanic and Atmospheric Administration*), el primero de cuyos satélites tomó el nombre de ITOS-1 (*Improved TIROS Operational System*). Este programa de observación meteorológica sigue vigente con la puesta en servicio del NOAA-20, lanzado en 2017 y redominado JPSS-1 (*Joint Polar Satellite System*). De esta serie está previsto en 2022 el lanzamiento del segundo satélite. Paralelamente a la serie NOAA, la URSS puso en marcha su proyecto METEOR de satélites heliosincrónicos con características similares a los citados.

Simultáneamente a estos programas, la NASA desarrolló otros de índole científica, como el proyecto NIMBUS, que llegó a poner en órbita siete satélites entre 1964 y 1978, o el geoestacionario ATS, que generó seis satélites entre 1966 y 1974.

La necesidad de establecer un sistema de observación meteorológica mundial, hoy denominado WWW (*World Weather Watch*), fomentó el desarrollo de los satélites geoestacionarios civiles a los que pertenecen el norteamericano GOES, el ruso GOMS y el europeo Meteosat, del que ya se han puesto en órbita una decena.

Si bien son evidentes los beneficios obtenidos con la explotación de la información proporcionada por los satélites meteorológicos, la teledetección ha alcanzado su mayor popularidad a raíz de la puesta en marcha de lo que se conoce como Programas de Observación de la Tierra y de evaluación de los recursos naturales. Entre ellos, quizá el más destacado ha sido —y no ha perdido vigencia— el programa Landsat, que puso en órbita su primer satélite en 1972 con el nombre ERTS (*Resource Technology Satellite*), siendo continuado con el Landsat-2, y así sucesivamente hasta el Landsat-8, operativo en la actualidad. De los instrumentos que llevaba a bordo, los antiguos sensores TM (*Thematic Mapper*) y ETM+ (*Enhanced Thematic Mapper*) y los actuales OLI (*Operational Land Imager*) y TIRS (*Thermal Infrared Sensor*) proporcionan unos datos de gran resolución espectral y radiométrica, por cuya razón, unido a que sus imágenes han pasado de a ser de dominio público, ha alcanzado gran demanda.

Con fines más cartográficos, el consorcio franco-belga-sueco SPOT desarrolló una familia de satélites de muy alta resolución espacial que, desde 1986, ha puesto en órbita tres satélites que están gozando de una gran aceptación, especialmente cuando los datos de su sensor HRV son explotados en combinación con los proporcionados por Landsat. En esta misma línea, la aparición en el ámbito civil de los datos procedentes del satélite Helios, con resoluciones espaciales de 1 y 2 metros, vino por esas fechas a satisfacer



El satélite Sputnik.

las demandas de los profesionales sobre aplicaciones de gran precisión geométrica, hasta entonces pertenecientes en exclusiva a la esfera militar.

La escena de la teledetección ha empezado a completarse con la explotación de datos en regiones del espectro electromagnético hasta entonces no exploradas desde satélites. La extensión al dominio de las microondas mediante la incorporación de sistemas activos como el radar en sus múltiples versiones, como sistemas formadores de imágenes, como dispersómetros o como altímetros, ha proporcionado un nuevo auge a la teledetección.

La utilización del radar se remonta a la Segunda Guerra Mundial, con la finalidad casi exclusiva de detectar aviones enemigos en vuelo. Sin embargo, posteriormente en la esfera civil, su eficaz aprovechamiento en modo lateral sobre aviones estimuló la investigación de nuevos procedimientos, con el fin de lograr la resolución espacial requerida para utilizarlo a bordo de satélites. De este modo, se desarrollaron los sistemas de radar de apertura sintética SAR (*Synthetic Aperture Radar*), de los cuales constituyó un exponente muy importante el SAR del satélite ERS (*European Remote Sensing*) de la Agencia Espacial Europea con dos satélites en órbita, el primero lanzado en 1991 y el segundo en 1995, de cuya operación en tándem se obtuvieron resultados interferométricos muy interesantes. Ambos fueron sustituidos por el programa ENVISAT, en cuyo programa —ya en estado no operativo— participaba España. Otro satélite expresamente diseñado para el radar es el canadiense Radarsat, cuyos datos han sido explotados durante los últimos años y se ha convertido en referencia entre los satélites radar.

En los últimos años se está asistiendo al incremento de la resolución espacial hasta límites que hacen dudar de si el producto es una imagen de satélite o una

fotografía aérea. Satélites como Ikonos, QuickBird, GeoEye o WorldView superaron la barrera de 1 m de resolución espacial y algunos de ellos proporcionan imágenes con celdas de 25 cm.

Por su parte, la Agencia Espacial Europea ha puesto en marcha el ambicioso programa Sentinel, con una nutrida familia de satélites cada una de ellos destinado a un objetivo concreto, radar, observación de la Tierra, altimetría, oceanografía y estudios atmosféricos, cuyos datos son también de dominio público.

Simultáneamente a los programas especiales, altamente costosos, de gran complejidad organizativa y que requieren años para su diseño, fabricación, puesta a punto y lanzamiento, se están desarrollando a un ritmo rapidísimo las aplicaciones basadas en vehículos aéreos no tripulados (*Unmanned Aerial Vehicle*, UAV) que utilizan sensores simplificados, pero especializados en aquellas regiones del espectro electromagnético que permiten caracterizar el fenómeno estudiado, prescindiendo de objetivos multipropósito. El abaratamiento de los costes está fomentando su utilización cada vez más habitual en las disciplinas más diversas.

Junto con el incremento de la resolución espacial y de la utilización de los datos radar, el desarrollo de las técnicas de interpretación hiperespectral es otro de los campos de investigación actual en teledetección. Es de suponer que una de las vías naturales de evolución de las plataformas espaciales será la incorporación de sensores de mayor resolución espectral, como de hecho ya se utilizan en teledetección aérea. A partir de ahí se comenzará a sustituir el tradicional análisis multiespectral, basado en los datos de un puñado de bandas, por el hiperespectral, que habrá de manejar datos de varios centenares de bandas simultáneamente, o el ultraspectral, que lo hará con miles de ellas.



Imagen actual de satélite, indistinguible de una fotografía aérea.

La explotación de los datos que proporcionan los satélites no habría sido posible sin el desarrollo simultáneo de los sistemas de análisis de imagen. La incorporación de nuevas posibilidades de tratamiento y el robustecimiento de los procesos que ya estaban vigentes, unido todo ello a la simplificación de los procedimientos, la amigabilidad de las aplicaciones informáticas y la cada vez mayor potencia y rapidez de las plataformas de proceso están permitiendo abordar problemas más complejos. Por otro lado, los datos espaciales procedentes de imágenes de satélite constituyen volúmenes de información muy considerables, particularmente cuando los estudios se desarrollan sobre regiones geográficas extensas. Las herramientas capaces de gestionar eficazmente ese tipo de

información son los sistemas de información geográfica (SIG) de carácter teselar o *raster*. El desarrollo de los SIG y el adiestramiento de analistas en este campo ha posibilitado la gestión útil de la teledetección en muchos ámbitos hasta entonces inexplorados.

La investigación aplicada a los SIG, los rotundos y espectaculares éxitos obtenidos a partir de las imágenes radar, las posibilidades de mejorar su interpretación y su integración con la información procedente de sensores de gran resolución espacial en el espectro solar, el análisis hiperespectral y el uso extendido de vehículos no tripulados para proyectos de detalle están indicando con toda seguridad los derroteros por los que se encaminará la teledetección en los próximos años.

Campos de aplicación

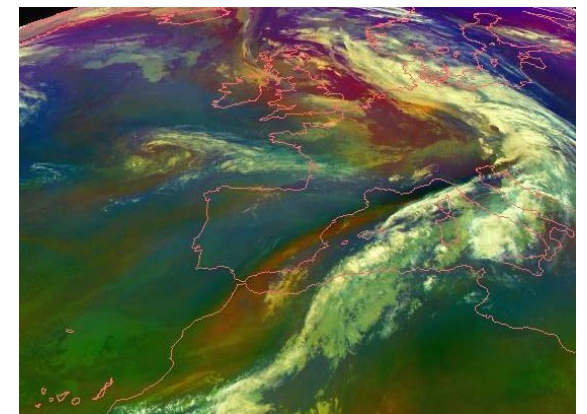
Sin entrar en detalles que excederían de la extensión de este capítulo, se ofrece a continuación una panorámica de las aplicaciones en las que la teledetección está demostrando ser una herramienta útil, tanto en la vertiente de investigación como en la de producción.

Tradicionalmente se han venido clasificando las aplicaciones de la teledetección de acuerdo con el lugar en el que acontece el fenómeno estudiado. Así hay aplicaciones centradas en la atmósfera, la litosfera, la hidrosfera y la biosfera. Profundizando un poco más, habría que añadir la criosfera, la gestión de emergencias, la cartografía o las infografías, por citar sólo unas cuantas.

Atmósfera

La atmósfera fue uno de los objetivos de la teledetección desde sus inicios, especialmente en cuanto se refiere a la meteorología. Las aplicaciones sobre la atmósfera siguen manteniendo como estrella a la predicción meteorológica, pero también se realizan estudio de perfiles atmosféricos, se determina el contenido en ozono y de otros gases, especialmente los llamados *de efecto invernadero*, se realiza con cada vez mayor precisión la predicción y seguimiento de los huracanes y se trabaja incesantemente en el establecimiento de modelos climatológicos locales y globales.

Fenómenos como el de *el Niño*, no han sido bien comprendidos hasta hace un par de décadas, con el



Composición de masas de aire de diferentes temperaturas sobre una imagen de la nubosidad en el espectro visible con fines de predicción meteorológica.

uso masivo de datos meteorológicos y climáticos a escala global, proporcionados por los satélites meteorológicos. La virulencia de los huracanes está siendo evaluada por misiones orbitales que miden la cantidad de lluvia, la altura de las nubes y la velocidad del viento en el interior de la masa convectiva. La medida exhaustiva de la temperatura de la superficie desde satélite está permitiendo alimentar objetivamente los modelos climáticos con los cuales se vigilan las anomalías térmicas y pluviométricas y se establecen las proyecciones futuras del clima. Dentro de este ámbito, la medida de la radiación solar a nivel de la superficie terrestre proporcionada por los satélites es otra entrada básica en dichos modelos.



Estudio de eutrofización de masas de agua.

Litosfera

El otro campo de aplicación de mayor empuje, sobre todo en los inicios de la teledetección, fue el relacionado con la geología y, por extensión, todos aquellos fenómenos que suceden en la superficie sólida de la Tierra. La teledetección es una herramienta muy apreciada en la elaboración de la cartografía edafológica y geológica, en el análisis geomorfológico, en los estudios de movimientos tectónicos o en la cartografía de riesgos de erosión. Desde un enfoque altimétrico está siendo utilizada en el seguimiento y análisis de las consecuencias de los sismos, en el seguimiento de deformaciones de la corteza terrestre en general y de subsidencias del terreno en particular, y desde el punto de vista hiperespectral tiene su aplicación en las prospecciones geomíneras y en la búsqueda de yacimientos de hidrocarburos fuera de los mares.

Hidrosfera

Como componente mayoritario en superficie del planeta, el agua reúne numerosas aplicaciones de la teledetección, distinguiendo aguas marinas y continen-

tales. Atendiendo a las aguas marinas, la teledetección oceanográfica se ocupa de la medida de la temperatura superficial del mar, básica en la comprensión de la redistribución del calor por todo el globo, de la determinación y el estudio de las corrientes de agua, de la altura de las olas, de la salinidad, de la caracterización de los contaminantes costeros, del vertido de hidrocarburos, del análisis del contenido en fitoplancton o de la localización de pesquerías para el guiado de las flotas pesqueras. Las ondas internas producidas por el movimiento de las masas de agua en las proximidades de accidentes geográficos es posible medirlas estudiando en detalle la altimetría de la superficie marina desde satélite, descontada la acción de las mareas y el oleaje.

En el ámbito de las aguas continentales, la teledetección se utiliza para medir la turbidez y el contenido en contaminantes y la eutrofización de las aguas confinadas como las de los embalses y lagunas. También es posible evaluar la innivación, midiendo el espesor de la capa de nieve invernal y anticipar el deshielo primaveral con el que alimentar los modelos hidrológicos de las cuencas y establecer los balances hídricos.

Biosfera

El estudio de las variadas manifestaciones de la vida también es objeto de atención por parte de la teledetección. Así pues, es posible y recurrente realizar controles e inventarios agrícolas, monitorizar el estado fenológico de la vegetación, estimar la producción agrícola, hace años sólo a nivel regional y ahora ya a nivel de parcela agraria, detectar las plagas y enfermedades de las plantas, evaluar la cantidad de biomasa forestal o agrícola, detectar la deforestación, hacer el seguimiento casi en tiempo real de los incendios forestales o medir la extensión y severidad de los daños ocasionados.

Una de las esferas de mayor éxito de la teledetección, especialmente la hiperespectral, es la agricultura de precisión, conjunto de sistemas y procedimientos destinados a evaluar el estado de la vegetación cultivada con gran resolución espacial, determinar sus necesidades nutricionales, contrastarlas con la capacidad del suelo, aplicar punto a punto las dosis adecuadas de fertilizantes y de productos fitosanitarios y, finalmente, medir la cosecha georreferenciando la producción con la misma resolución.

Cartografía

Dada la resolución espacial que en la actualidad ofrecen las imágenes de satélite, es posible ofrecer nuevas aplicaciones en el ámbito de la cartografía, y ya no solamente de pequeña escala, sino de casi cualquiera de ellas. Hasta hace relativamente pocos años, la teledetección tenía vetado el contexto urbano, fundamentalmente por dos razones. De una parte, la resolución de las imágenes sólo permitía la producción de ortoimágenes de escala inferior a 1:10 000 y en ciertos casos puntuales, 1:5 000. Por encima de ellas, el tamaño del píxel se hacía demasiado evidente y sobrepasaba el límite de la percepción visual, ampliamente establecido en 0,2 mm. Además, dentro del tejido urbano son abundantes las sombras, las que arrojan las edificaciones sobre el suelo y las que proyectan ellos entre sí. Una gran proporción de la información útil quedaba dentro de estas zonas sombreadas y la escasa resolución radiométrica de los sensores utilizados reservaban sólo unos pocos niveles digitales a esas áreas tan extensas. En definitiva, era difícil extraer información de valor de las imágenes urbanas. El desarrollo reciente de sensores orbitales de muy alta resolución espacial y radiométrica permite adquirir imágenes con píxeles decimétricos y, lo que también es muy importante, destinan gran parte del margen dinámico a las zonas oscuras, de tal modo

que ahora es posible estudiarlas en detalle, distinguiendo y caracterizando los elementos que quedan englobados en ellas.

Por ello, actualmente la teledetección está proporcionando cartografías a todas las escalas y permite realizar levantamientos regionales a escalas 1: 100 000 y menores, y actualizaciones de mapas 1:50 000 y 1:25 000. Así mismo se están llevando a cabo actualizaciones de la cartografía urbana a escalas 1:10 000, 1:5 000 e incluso 1:2 000. La disciplina está siendo, pues, utilizada en la ordenación del territorio, estudiando la situación inicial y la evolución de las áreas urbanas como base para la planificación. Es posible mediante teledetección realizar con los nuevos sensores el seguimiento del movimiento vehículos, aunque de forma restringida al ámbito militar.

Otras aplicaciones

Aunque el elenco de aplicaciones es extensísimo, cabe destacar que la ingeniería medioambiental y del paisaje han encontrado en la teledetección una fuente de datos para la elaboración de vistas realistas en construcciones denominadas 3D, fusionando la información procedente de una imagen de satélite de alta resolución con un modelo digital de elevaciones, que se sombrea con una fuente solar arbitrariamente colocada y a cuyo conjunto se le puede añadir un fondo celeste adornado por nubes virtuales. La superficie puede incluso enriquecerse con la introducción de texturas diversas.

Este mismo procedimiento es el utilizado en la creación de escenarios para videojuegos de aviación y simuladores de vuelos, tratados por equipos informáticos potentes dotados de tarjetas gráficas específicas que permiten simular el movimiento, recalculando y reconfigurando la vista de la escena de acuerdo a la



Realismo logrado en videojuegos a partir de imágenes de satélite.

geometría de cada instante.

2

La radiación electromagnética

Introducción

La finalidad de este manual es establecer los elementos básicos de la teledetección espacial de la superficie terrestre. La forma de adquirir información desde un satélite es capturar y medir la radiación electromagnética emitida o reflejada por la superficie.

El objetivo de este capítulo es introducir los fundamentos físicos de la teledetección, comenzando por la propia radiación electromagnética. No se trata de desarrollar un tratado sobre esta materia sino simplemente presentar sus principios básicos, los cuales serán empleados a lo largo de los siguientes capítulos.

El campo electromagnético

Uno de los tipos de interacción entre las partículas elementales que componen la materia es la electromagnética, en virtud de una propiedad de ésta denominada carga eléctrica. La influencia ejercida en el espacio por la presencia de un cuerpo cargado viene establecida por la generación de un campo electromagnético.

Un campo es una región del espacio en la que puede definirse una función uniforme cuyo valor depende de la posición, de tal modo que a cada punto de la región le corresponda una magnitud física determinada. Si la magnitud asociada a la posición es una cantidad, el campo se denomina escalar. Si, por el contrario, a cada punto del espacio se le puede hacer corresponder un vector, esto es, una magnitud, una dirección y un sentido, el campo recibe el nombre de vectorial. Por extensión, también se denomina campo a la propia función.

Pues bien, el campo electromagnético es el ejercido por una partícula cargada eléctricamente que se mueve por el espacio y está definido mediante dos

vectores: el campo eléctrico \vec{E} y el campo magnético \vec{H} . La fuerza ejercida sobre una carga eléctrica q que se mueve en el seno de un campo electromagnético a una velocidad \vec{v} es:

$$\vec{F} = q(\vec{E} + \vec{v} \wedge \vec{B}) \quad (1)$$

siendo

$$\vec{B} = \mu_0 \vec{H} \quad (2)$$

el vector inducción magnética y μ_0 la permeabilidad magnética del medio, supuesto éste isótropo. Los campos eléctrico y magnético dependen de las posiciones y del movimiento de las cargas en el espacio.

Maxwell (1831-1879), reuniendo los resultados experimentales de Coulomb, Gauss, Ampère y Faraday, puso de manifiesto la relación que hay entre la electricidad y el magnetismo y demostró que la luz es una manifestación del campo electromagnético. Las interacciones electromagnéticas están definidas por las

denominadas *leyes de Maxwell*, también llamadas ecuaciones de campo:

Ley de Gauss para el campo eléctrico

El flujo eléctrico es una magnitud escalar que expresa la magnitud del campo eléctrico que atraviesa una determinada superficie. La Ley de Gauss para el campo eléctrico establece que el flujo de ese campo eléctrico a través de una superficie cerrada es proporcional a la carga que hay en su interior:

$$\operatorname{div} \vec{E} = \frac{1}{\epsilon_0} \rho \quad (3)$$

donde *div* es la divergencia, que se define como el siguiente producto escalar:

$$\operatorname{div} \vec{E} = \vec{\nabla} \cdot \vec{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \quad (4)$$

siendo:

ρ = densidad de carga eléctrica.
 ϵ_0 = constante dieléctrica del medio.

$\vec{\nabla} = \frac{\partial}{\partial x} \vec{i} + \frac{\partial}{\partial y} \vec{j} + \frac{\partial}{\partial z} \vec{k}$ es un operador dispuesto a ser aplicado a un vector como en campo eléctrico.

El campo eléctrico siempre diverge a partir de una carga puntual.

Ley de Gauss para el campo magnético

Las líneas del campo magnético son cerradas. Si se

confina un dipolo magnético en el interior de una superficie cerrada, el flujo saliente es igual al entrante, es decir, la cantidad de líneas de fuerza que salen del polo norte magnético es igual a las que llegan al polo sur. En consecuencia, el flujo magnético neto en el exterior de dicha superficie es nulo. El campo magnético no diverge y, por lo tanto, no existen monopolos magnéticos.

$$\operatorname{div} \vec{B} = 0 \quad (5)$$

Ley de la inducción de Faraday-Henry

La variación del campo magnético con el tiempo origina un campo eléctrico:

$$\operatorname{rot} \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (6)$$

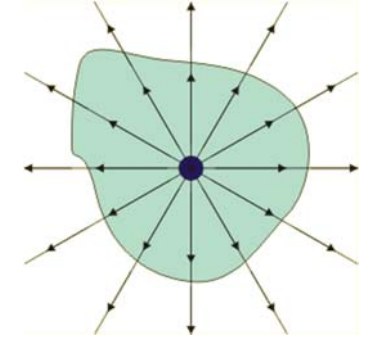
siendo *rot* el rotacional, que se define como siguiente producto vectorial:

$$\operatorname{rot} \vec{E} = \vec{\nabla} \wedge \vec{E} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix} \quad (7)$$

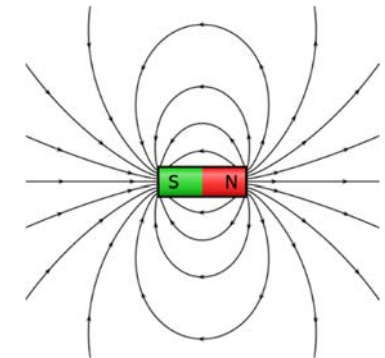
y

$$\frac{\partial \vec{B}}{\partial t} = \frac{\partial B_x}{\partial t} \vec{i} + \frac{\partial B_y}{\partial t} \vec{j} + \frac{\partial B_z}{\partial t} \vec{k} \quad (8)$$

Igualando ambas expresiones:



El campo eléctrico diverge desde una carga puntual.



El campo magnético no diverge desde un dipolo.

$$\begin{aligned}
 \text{rot } \vec{E} &= \\
 &= \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \right) \vec{i} - \left(\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z} \right) \vec{j} + \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right) \vec{k} = \\
 &= -\frac{\partial B_x}{\partial t} \vec{i} - \frac{\partial B_y}{\partial t} \vec{j} - \frac{\partial B_z}{\partial t} \vec{k} = -\frac{\partial \vec{B}}{\partial t}
 \end{aligned}$$

Es decir:

$$\begin{aligned}
 \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\frac{\partial B_x}{\partial t} \\
 \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\frac{\partial B_y}{\partial t} \\
 \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\frac{\partial B_z}{\partial t}
 \end{aligned} \tag{9}$$

Los signos negativos de la derecha indican que la corriente inducida se opone a la causa que la produce, contrarrestando así la variación de flujo magnético.

Ley de Ampère-Maxwell

La corriente que circula por un conductor induce la aparición de un campo magnético rotacional alrededor:

$$\text{rot } \vec{B} = \mu_0 \varepsilon_0 \frac{\partial \vec{E}}{\partial t} + \mu_0 \vec{J} \tag{10}$$

donde el primer sumando del segundo miembro es

proporcional a lo que Maxwell llamó corriente de desplazamiento, que se define así:

$$\vec{J}_d = \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \tag{11}$$

Esta expresión indica que la corriente de desplazamiento es originada por la variación temporal del campo eléctrico. Por su parte, el término \vec{J} es la densidad de corriente, que depende directamente del campo eléctrico:

$$\vec{J} = \sigma \cdot \vec{E} \tag{12}$$

donde σ es la conductividad eléctrica.

Ley de Ampère-Maxwell, pues, pone de manifiesto que el campo magnético depende de la densidad de *corriente de desplazamiento*, o velocidad de variación del campo eléctrico, y de la densidad de la *corriente de conducción*, o variación de la carga con el tiempo.

En el vacío, las ecuaciones de Maxwell se simplifican al ser $\rho=0$ y $\vec{J}=0$, quedando:

$$\begin{aligned}
 \text{div } \vec{E} &= 0; & \text{rot } \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\
 \text{div } \vec{B} &= 0; & \text{rot } \vec{B} &= \mu_0 \varepsilon_0 \frac{\partial \vec{E}}{\partial t}
 \end{aligned} \tag{13}$$

El movimiento ondulatorio

La onda es una perturbación que se propaga en el espacio. Considérese un campo que describa alguna propiedad física de la materia. Si las condiciones en cualquier lugar dependen del tiempo, la perturbación originada en un cierto punto se propagará a través del espacio. Sea una función $\xi = \xi(x)$ que describe el estado de los puntos de una recta. Pues bien, la función $\xi = \xi(x-a)$ tiene la misma forma que la anterior, diferenciándose solamente en un cierto desplazamiento a . Si $a = vt$, siendo v una cierta velocidad, entonces la función $\xi = \xi(x-vt)$ representará una curva que se desplaza sin deformación a lo largo del eje x a la velocidad v . Si la función de estado es del tipo sinusoidal, como:

$$y = \xi(x-vt) = \xi_0 \text{senk}(x-vt) \quad (14)$$

donde ξ_0 representa la amplitud de dicha función, puede comprobarse que su valor en el punto x es el mismo que en el punto $x + 2\pi/k$, ya que:

$$\begin{aligned} y_{x+2\pi/k} &= \xi_0 \text{sen} k(x + 2\pi/k - vt) = \\ &= \xi_0 \text{sen} [k(x - vt) + 2\pi] = \\ &= \xi_0 \text{sen} k(x - vt) = \\ &= y_x \end{aligned} \quad (15)$$

La constante $\lambda = 2\pi/k$ que representa el espacio tras el cual la forma de la curva se repite nuevamente se denomina *longitud de onda*. Se suele expresar en metros o en submúltiplos.

Si, en cambio, se estudia la perturbación periódica de un solo punto a lo largo del tiempo, el mínimo plazo medido en segundos en el que la función vuelve a alcanzar el mismo valor se denomina período T . La frecuencia es adimensional y expresa el número de veces que oscila una onda completa durante un segundo, es decir, el número de ciclo por segundo y se expresa en *hercios* (Hz) o s^{-1} .

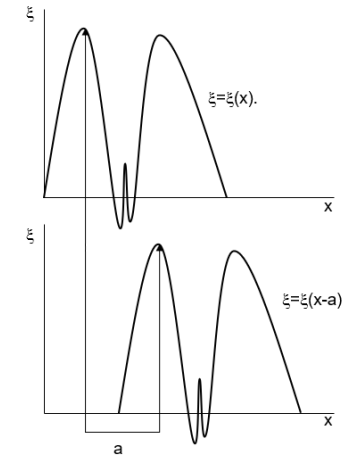
Las relaciones entre el período, la frecuencia f de la perturbación y la longitud de onda son las siguientes:

$$f = \frac{1}{T}; \quad T = \frac{\lambda}{v}; \quad f = \frac{v}{\lambda} \quad (16)$$

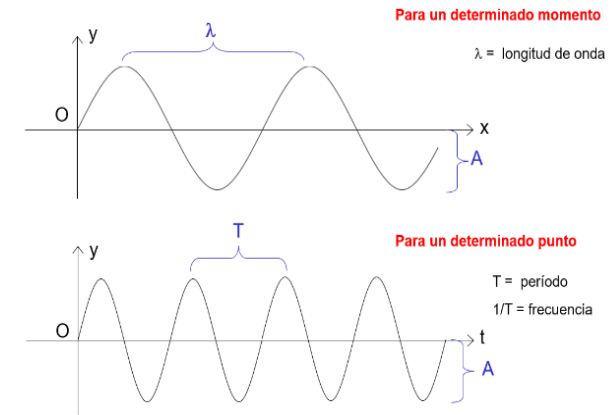
En ocasiones se prefiere hablar de *número de onda* $\tilde{\nu}$ para caracterizar una vibración. Esta variable es el número de veces que vibra la onda en una unidad de distancia y equivale a la inversa de la longitud de onda:

$$\tilde{\nu} = \frac{1}{\lambda} \quad (17)$$

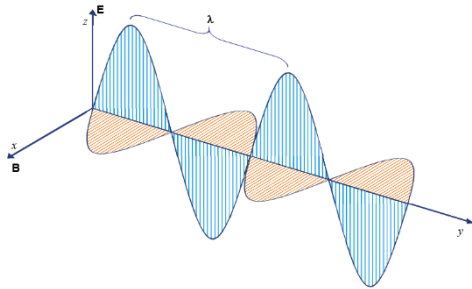
Cuando la propagación de la onda se efectúa en las tres dimensiones, la perturbación tendrá por ecuación $\xi = \xi(x, y, z, t)$. Se denomina frente de onda a la superficie alcanzada por la perturbación en el instante t_0 , cuya ecuación es $\xi(x, y, z, t_0) = \xi_0$, siendo ξ_0 una constante. Los frentes de ondas derivados por ondas cilíndricas se denominan frentes cilíndricos y, finalmente, los frentes ocasionados por ondas esféricas son a su vez esferas concéntricas.



Desplazamiento de la perturbación.



Longitud de onda y período.



Propagación de la onda electromagnética.

La onda electromagnética

Sea un sistema de referencia tridimensional ortogonal definido por los ejes X, Y, Z , cuyos vectores directores son, respectivamente $\vec{i}, \vec{j}, \vec{k}$. Para que un campo electromagnético satisfaga las condiciones de Maxwell en el vacío debe existir un campo eléctrico paralelo al eje Z en todos los puntos del espacio y un campo magnético paralelo al eje X , perpendicular al campo eléctrico, cuyas intensidades dependan solamente de su coordenada y y del tiempo t , de la forma:

$$\begin{aligned}\vec{E} &= E_0 \operatorname{sen}(y - vt) \vec{k} \\ \vec{B} &= B_0 \operatorname{sen}(y - vt) \vec{i}\end{aligned}\quad (18)$$

donde E_0, B_0 y v son constantes.

Definido de esta manera, el campo electromagnético satisface las ecuaciones de Maxwell bajo ciertas condiciones. En primer lugar, es inmediato comprobar que $\operatorname{div} \vec{E}$ y $\operatorname{div} \vec{B}$ son nulas, ya que la única componente de \vec{E} es E_z , que no depende de z , sino de y , y la única componente de \vec{B} es B_x , que no depende de x , sino también de y .

Las condiciones para dar cumplimiento a las otras dos ecuaciones son:

$$\begin{aligned}E_0 &= vB_0 \\ B_0 &= \mu_0 \varepsilon_0 v E_0\end{aligned}\quad (19)$$

Por ello debe ser:

$$v^2 = \frac{1}{\mu_0 \varepsilon_0} \quad (20)$$

Y como:

$$\mu_0 = 4\pi \cdot 10^{-7} \text{ N} \cdot \text{A}^{-2}$$

$$\varepsilon_0 = 8,854 \cdot 10^{-12} \text{ F} \cdot \text{m}^{-1}$$

en definitiva, se tiene:

$$\pm v = 2,99792 \cdot 10^8 \text{ m} \cdot \text{s}^{-1} = c$$

$$B_0 = \frac{E_0}{c} \quad (21)$$

De lo anterior se deducen las tres propiedades siguientes, válidas para cualquier onda electromagnética plana en el vacío:

1. El modelo de campo propuesto se propaga con velocidad c , siendo ésta la velocidad de la luz en el vacío ($c \approx 3 \cdot 10^8 \text{ m/s}$).
2. En cada punto de la onda y en cada instante, la intensidad del campo eléctrico es igual a c por la del magnético. En consecuencia, ambos campos están en fase, es decir toman valores extremos y nulos simultáneamente.
3. El campo eléctrico y el magnético son perpendiculares entre sí y a la dirección de propagación, siendo su sentido el resultante del producto vectorial $\vec{E} \wedge \vec{B}$.

La onda se dice que está polarizada cuando la dirección de los vectores \vec{E} y \vec{B} permanece constante a lo largo de la dirección de propagación. Si, por el contrario, el extremo del vector \vec{E} describe a lo largo de

la dirección de propagación una espiral inscrita en un cilindro de sección recta circular o elíptica se dice que tiene una polarización circular o elíptica respectivamente.

La teoría cuántica

La teoría de Maxwell presenta deficiencias cuando trata de explicar algunos fenómenos de interacción entre la radiación electromagnética y la materia. La teoría cuántica formulada por Max Planck vino a paliar tales deficiencias partiendo de la consideración de que todo cuerpo radiante emite energía electromagnética de forma discreta y no continua, cuya magnitud es múltiplo entero de un determinado número, originalmente denominado *cuanto* y, además, es directamente proporcional a la frecuencia de la radiación. Así, puede decirse que $\Delta E = h\nu$, donde h es la constante de Planck y ν es la frecuencia de la onda expresada en hercios, siendo $\nu = c/\lambda$, c la velocidad de la luz expresada en m/s y λ la longitud de onda expresada en m. La mínima unidad de energía —el *cuanto*— se llama *fotón* en la actualidad y, como puede apreciarse, transporta una cantidad de energía que depende de la longitud de onda de la radiación.

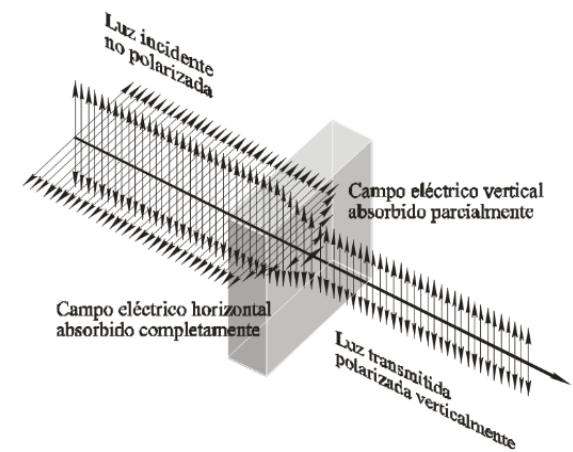
La constante de Planck tiene un valor de

$6,62607015 \cdot 10^{-34}$ J·s. Frecuentemente se presenta de la forma:

$$\hbar = \frac{h}{2\pi} \quad (22)$$

con un valor de $(1,05457182 \cdot 10^{-34}$ J·s.

La controversia surgida desde la postulación de las teorías de Maxwell sobre la naturaleza de la radiación electromagnética ha sido armonizada con la teoría cuántica considerando una cualidad dual de la radiación electromagnética: ondulatoria y corpuscular. La primera explica los fenómenos de propagación, que se manifiestan por ejemplo en las interferencias de la luz al pasar una doble rendija, y la segunda, los de interacción con la materia a nivel submolecular y subatómico, como por ejemplo el efecto fotoeléctrico descubierto por Einstein.

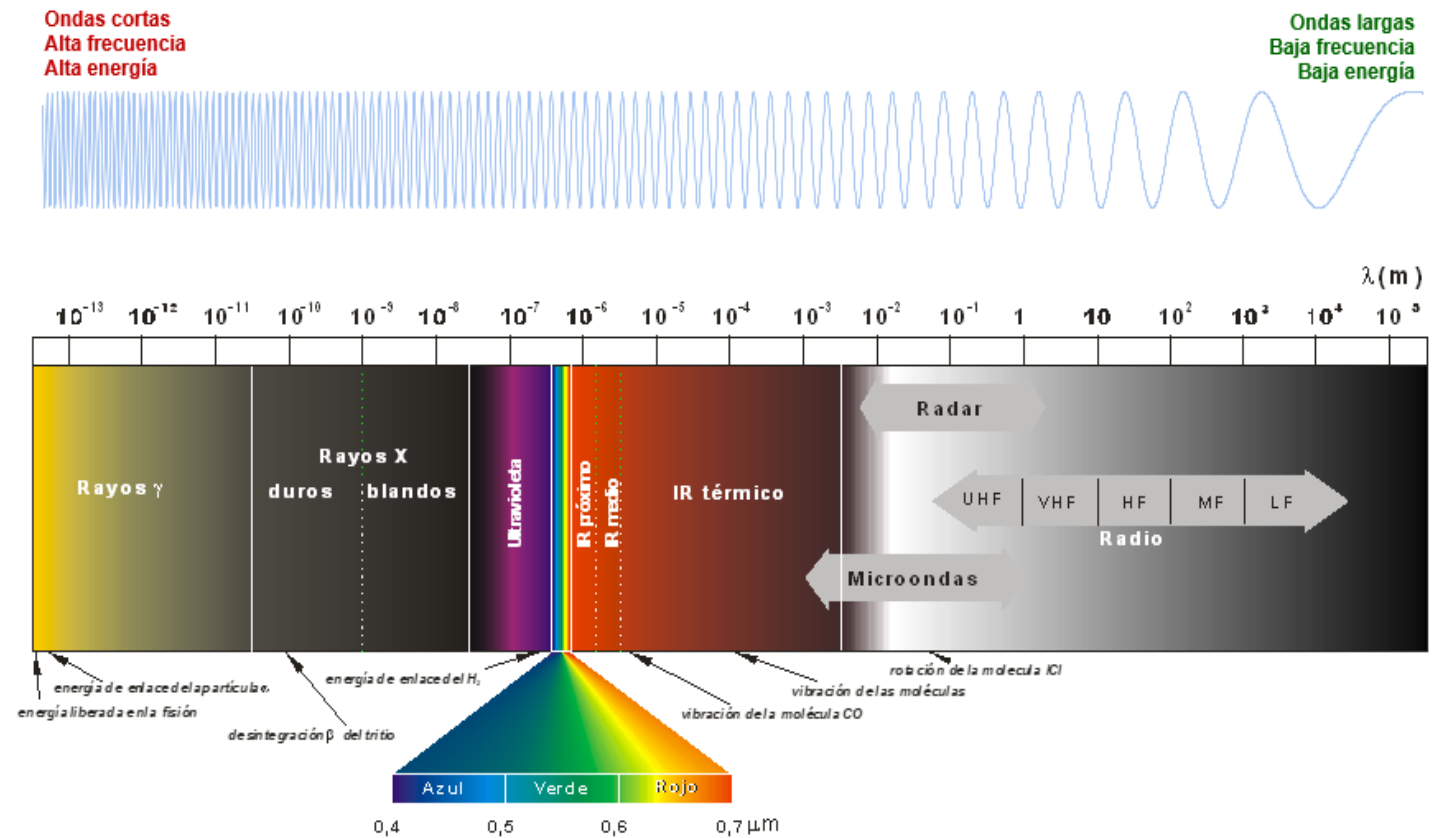


Polarización.

El espectro electromagnético

La radiación electromagnética se presenta como una distribución continua de frecuencias, desde longitudes de onda cortas y altas frecuencias (rayos gamma) hasta longitudes de onda muy largas y bajas frecuencias (ondas de radio). Dentro de esta continuidad es

posible agrupar los distintos tipos de radiación en familias de diferente comportamiento, en función de su longitud de onda. Se denomina espectro electromagnético a la sucesión de longitudes de onda de todas las radiaciones conocidas. Las familias o regiones del



El espectro electromagnético.

espectro son las siguientes: rayos cósmicos, rayos gamma, rayos X, rayos ultravioleta (UV), espectro visible, infrarrojo próximo (IRp), infrarrojo térmico (IRt), microondas y ondas de televisión y radio (UHF y VHF).

Recordando que $1 \text{ \AA} = 10^{-10} \text{ m}$, y que $1 \mu\text{m} = 10^{-6} \text{ m}$, los rayos cósmicos, compuestos por muones, fotones y neutrones en proporciones variables, cubren la región espectral de longitudes de onda inferiores a

10^{-3} \AA , los rayos γ la comprendida entre la anterior y $0,3 \text{ \AA}$, y los rayos X la que va desde $0,3$ hasta 300 \AA . Su aplicación a la teledetección está muy limitada debido a la opacidad de la atmósfera a este tipo de radiación. Solamente han sido utilizadas en plataformas de vuelo a baja altura, con poco espesor atmosférico entre el sensor y la superficie terrestre, y en el estudio de superficies sin atmósfera, como en misiones lunares.

Tabla 1. El espectro electromagnético.

Banda	Longitud de onda (m)	Frecuencia (Hz)	Energía del fotón (J)
Rayos gamma	$< 10 \cdot 10^{-12}$ m (= 10 pm)	$> 30,0$ EHz	$> 20 \cdot 10^{-15}$ J
Rayos X	$< 10 \cdot 10^{-9}$ m (= 10 nm)	$> 30,0$ PHz	$> 20 \cdot 10^{-18}$ J
Ultravioleta extremo	$< 200 \cdot 10^{-9}$ m (= 200 nm)	$> 1,5$ PHz	$> 993 \cdot 10^{-21}$ J
Ultravioleta cercano	$< 380 \cdot 10^{-9}$ m (= 380 nm)	> 789 THz	$> 523 \cdot 10^{-21}$ J
Visible	$< 780 \cdot 10^{-9}$ m (= 780 nm)	> 384 THz	$> 255 \cdot 10^{-21}$ J
Infrarrojo próximo o cercano	$< 3,0 \cdot 10^{-6}$ m (= 3,0 μ m)	> 100 THz	$> 66,2 \cdot 10^{-21}$ J
Infrarrojo medio	$< 50 \cdot 10^{-6}$ m (= 50 μ m)	$> 6,00$ THz	$> 4 \cdot 10^{-21}$ J
Infrarrojo lejano/submilimétrico	$< 10^{-3}$ m (= 1 mm)	> 300 GHz	$> 19,9 \cdot 10^{-23}$ J
Microondas	$< 30 \cdot 10^{-2}$ m (= 30 cm)	> 1 xGHz	$> 2 \cdot 10^{-24}$ J
Radio de Ultra Alta Frecuencia (UHF)	< 1 m	> 300 MHz	$> 19,9 \cdot 10^{-26}$ J
Radio de Muy Alta Frecuencia (VHF)	< 10 m	> 30 MHz	$> 19,9 \cdot 10^{-28}$ J
Radio de Onda Corta (HF)	< 180 m	$> 1,7$ MHz	$> 11,22 \cdot 10^{-28}$ J
Radio de Onda Media (MF)	< 650 m	> 650 kHz	$> 42,9 \cdot 10^{-29}$ J
Radio de Onda Larga (LF)	$< 10^4$ m (= 10 km)	> 30 kHz	$> 19,9 \cdot 10^{-30}$ J
Radio de Muy Baja Frecuencia (VLF)	$> 10^4$ m (= 10 km)	< 30 kHz	$< 19,9 \cdot 10^{-30}$ J

La región del ultravioleta ocupa el intervalo del espectro comprendido entre 300 Å y 0,4 μ m. La opacidad de los gases que componen la atmósfera terrestre a este tipo de radiación también limita su utilización con fines de teledetección. Sin embargo, ha sido utilizada con éxito en el estudio de atmósferas de otros

planetas y de las superficies de planetas sin atmósfera.

Aunque los límites entre unas familias de radiación y otras son difusos, la clasificación completa del espectro es la recogida en la Tabla 1.

Tabla 2. Clasificación de la radiación infrarroja.

Banda	Longitud de onda (m)	Frecuencia (Hz)	Energía del fotón (J)
Infrarrojo próximo o cercano (NIR)	$0,75-1,4 \cdot 10^{-6}$ m (0,75-1,4 μ m)	400-214 THz	$2,65-1,42 \cdot 10^{-19}$ J
Infrarrojo de onda corta (SWIR)	$< 3 \cdot 10^{-6}$ m (= 3,0 μ m)	> 100 THz	$> 66,21 \cdot 10^{-21}$ J
Infrarrojo de onda media (MWIR)	$< 8 \cdot 10^{-6}$ m (= 8,0 μ m)	> 37 THz	$> 24,83 \cdot 10^{-21}$ J
Infrarrojo de onda larga (LWIR)	$< 15 \cdot 10^{-6}$ m (= 15,0 μ m)	> 20 THz	$> 13,24 \cdot 10^{-21}$ J
Infrarrojo lejano (FIR)	$< 1000 \cdot 10^{-6}$ m (= 1 mm)	300 GHz	$> 19,86 \cdot 10^{-23}$ J

Con respecto al infrarrojo, son varias las clasificaciones de la radiación en este rango espectral. La subdivisión de la citada tabla corresponde al esquema ISO 20473. El esquema de subdivisión astronómica ordena el infrarrojo en tres grupos, próximo, medio y lejano con rangos respectivos desde el final del visible hasta 5 μm , de ahí hasta 25 o 40 μm y, finalmente desde ese límite hasta 200-350 μm . Según la subdivisión más generalmente aceptada, la radiación infrarroja se clasificaría como se recoge en la Tabla 2.

Los intervalos espectrales más utilizadas en teledetección corresponden a las regiones del visible, las distintas familias del infrarrojo y las microondas, caracterizados por las siguientes longitudes de onda, y se resumen en la Tabla 3.

En la región óptica del espectro tienen su manifestación energética las rotaciones y vibraciones moleculares de la materia, así como la excitación de los electrones en sus orbitales atómicos o moleculares como consecuencia de la incidencia de la radiación electromagnética.

La región espectral de las microondas cubre el intervalo de longitudes de onda comprendido entre 1 mm

Tabla 3. Ventanas atmosféricas.

Visible:	0,4	-	0,7	μm
Infrarrojo próximo:	0,7	-	2,5	μm
Infrarrojo medio:	2,5	-	5	μm
Infrarrojo térmico (IR medio y lejano):	5	-	1000	μm
Microondas	10^{-3}	-	1	m
Ondas de radio	1	-	10^2	m

y 1 m. En esta región tienen su manifestación los fenómenos de rotación molecular. La exploración de esta región se realiza por medio de radiómetros de microondas y de los sistemas de radar.

Las bandas de radio ocupan las regiones de longitudes de onda superiores a 1 m, siendo éstas las radiaciones fundamentalmente utilizadas por los sensores activos como los radares, altímetros, sondas y, en menor medida, por los radiómetros pasivos.

Generación de la radiación

Se denomina energía radiante a la energía transportada por las ondas electromagnéticas. Al proceso físico generador de energía radiante se le llama radiación. Como ocurre con cualquier otro tipo de energía, la radiación electromagnética se produce a partir de la transformación de otras clases de energía, como eléctrica, magnética, térmica, química, cinética o nuclear.

Los mecanismos de transformación dependen del tipo de radiación deseado, dando origen a gran diversidad de artilugios generadores. El mecanismo es tanto más organizado, es decir su funcionamiento reviste tanta menos aleatoriedad, cuanto más coherente sea la radiación generada, o lo que es lo mismo, cuanto menor sea la anchura de banda producida.

Las ondas de radio son generadas mediante osciladores, en los cuales se hacen circular corrientes eléctricas por oscilación periódica de cargas. El condensador, a modo de antena emisora, produce un campo electromagnético variable cuya frecuencia es la misma que la de la oscilación de aquél. Las microondas se generan en el interior de tubos electrónicos mediante la interacción de la energía transportada en chorros de electrones con diversos materiales. El mismo método puede utilizarse para la generación de luz ultravioleta, visible o incluso infrarroja. Dependiendo del estado energético inicial de la materia, de la excitación a la que se la someta y del nivel energético al que descendan las moléculas excitadas, la emisión de energía tendrá una diferente longitud de

onda, quedando ésta regida por la teoría cuántica y el *principio de exclusión de Pauli*¹. La energía térmica es debida a la energía cinética de las moléculas. Este movimiento adopta una forma errática en los gases y en los líquidos y es, en cambio, un movimiento vibratorio en los sólidos, pues las moléculas solamente se desplazan alrededor de su posición de equilibrio, sin abandonarla. Los choques producidos por las partículas liberan energía radiante en la región del infrarrojo, la cual se caracteriza por su escasa coherencia, como consecuencia de su naturaleza aleatoria.

Leyes de la radiación electromagnética

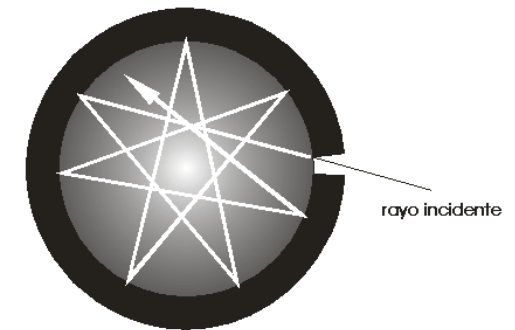
El cuerpo negro

La superficie de un cuerpo mantenido a una cierta temperatura emite radiación en todas las longitudes de onda. Se denomina cuerpo negro a un emisor y receptor de energía perfecto.

El cuerpo negro es una entidad ideal. En la naturaleza solamente existen superficies o cuerpos con mayor o menor similitud al cuerpo negro en ciertas longitudes de onda. Aun así, puede simularse con bastante eficacia un cuerpo negro disponiendo una cavidad esférica de superficie interior negra, rugosa y absorbente, con un pequeño orificio por el que existe

una probabilidad muy baja de que escape un fotón introducido a través de él. Esta esfera se comporta en la práctica como un cuerpo negro perfecto.

Si se mide la cantidad de energía radiante que emite un cuerpo negro por unidad de superficie, de tiempo y de intervalo de longitud de onda, en función de la propia longitud de onda, puede observarse cómo es muy pequeña y tendente a cero para longitudes de onda muy cortas o muy largas, y en cambio es más alta en longitudes de onda intermedias. Esa energía es máxima sólo a una cierta longitud de onda que depende de la temperatura del cuerpo.



El cuerpo negro.

¹ El principio de exclusión de Pauli es una regla de la mecánica cuántica que establece que no puede haber dos fermiones con todos sus números cuánticos idénticos (esto es, en el mismo estado cuántico) dentro del mismo sistema

cuántico. Esto quiere decir que dos electrones no pueden ocupar un mismo estado energético.

Leyes de la radiación

Del estudio de la radiación emitida por un cuerpo negro simulado pueden enunciarse las denominadas leyes de la radiación.

Ley de radiación de Planck

Establece el patrón de emisión de energía del cuerpo negro en función de su temperatura y de la longitud de onda.

La densidad de energía radiante desde el cuerpo negro por intervalo de longitud de onda, a la temperatura T depende de la longitud de onda λ y de la temperatura:

$$E_{\lambda,T} = \frac{8\pi hc}{\lambda^5} \frac{1}{e^{(hc/\lambda kT)} - 1} \quad (23)$$

En cuya expresión, k es la constante de Boltzmann ($1,3806488 \cdot 10^{-23} \text{ J}\cdot\text{K}^{-1}$) y h la constante de Planck.

Ley de desplazamiento de Wien

Para conocer la longitud de onda para la cual la emisión de energía desde el cuerpo negro es máxima a una cierta temperatura debe derivarse la expresión de la ley de radiación de Planck con respecto a λ e igualarla a cero, considerando T constante:

$$\frac{dE_{\lambda,T}}{d\lambda} = \frac{d}{d\lambda} \left(\frac{8\pi hc}{\lambda^5} \frac{1}{e^{(hc/\lambda kT)} - 1} \right)_{T=cte} = 0 \quad (24)$$

El resultado de esta operación permite encontrar la longitud de onda para la cual la radiación espectral del cuerpo negro es máxima y es conocida como ley de desplazamiento de Wien (fonéticamente, ['vi:n]):

$$\lambda|_{E=\max} (\mu\text{m}) = \frac{C_0}{T} \quad (25)$$

siendo:

$$C_0 = 0,2014 \frac{hc}{k} \quad (26)$$

es decir:

$$\lambda|_{E=\max} (\mu\text{m}) = 0,2014 \frac{hc}{kT} = \frac{2898 (\mu\text{m}\cdot\text{K})}{T (\text{K})} \quad (27)$$

Así, por ejemplo, el máximo de la curva de la radiación emitida por el Sol, cuya superficie se encuentra a una temperatura de unos 6 000 K, se presenta en torno a los 0,48 μm . Esta longitud de onda pertenece al espectro visible, en cuya región se concentra aproximadamente el 49 % de la energía emitida por el sol. El 99 % de la radiación solar se emite en una longitud de onda inferior a 5 μm , correspondiendo alrededor de un 42 % a la emisión en el infrarrojo y un 9 % a la que pertenece a las regiones del ultravioleta, rayos X y rayos γ .

La Tierra, cuya temperatura media superficial es de alrededor de 300 K, presenta un pico de radiación en torno a los 9,66 μm , esto es, en el infrarrojo térmico. Conocer la temperatura de la superficie emisora tiene importancia en teledetección, pues permite seleccionar las bandas más adecuadas, en las cuales el fenómeno a estudiar emite mayor cantidad de energía.

Ley de Stefan-Boltzmann

Para conocer la densidad total de energía radiante del cuerpo negro, esto es, la energía total que irradia un cuerpo negro por unidad de área, a una cierta temperatura en todas las longitudes de onda posibles y por unidad de tiempo, es necesario integrar la expresión de la Ley de Planck extendida a todo el intervalo de longitudes de onda:

$$P = \int_0^{\infty} E_{\lambda,T} d\lambda = \int_0^{\infty} \frac{8\pi hc}{\lambda^5} \frac{1}{e^{(hc/\lambda kT)} - 1} d\lambda \quad (28)$$

Haciendo el cambio de variable:

$$x = \frac{hc}{\lambda kT} \quad (29)$$

Se tiene:

$$d\lambda = \frac{hc}{kTx^2} dx \quad (30)$$

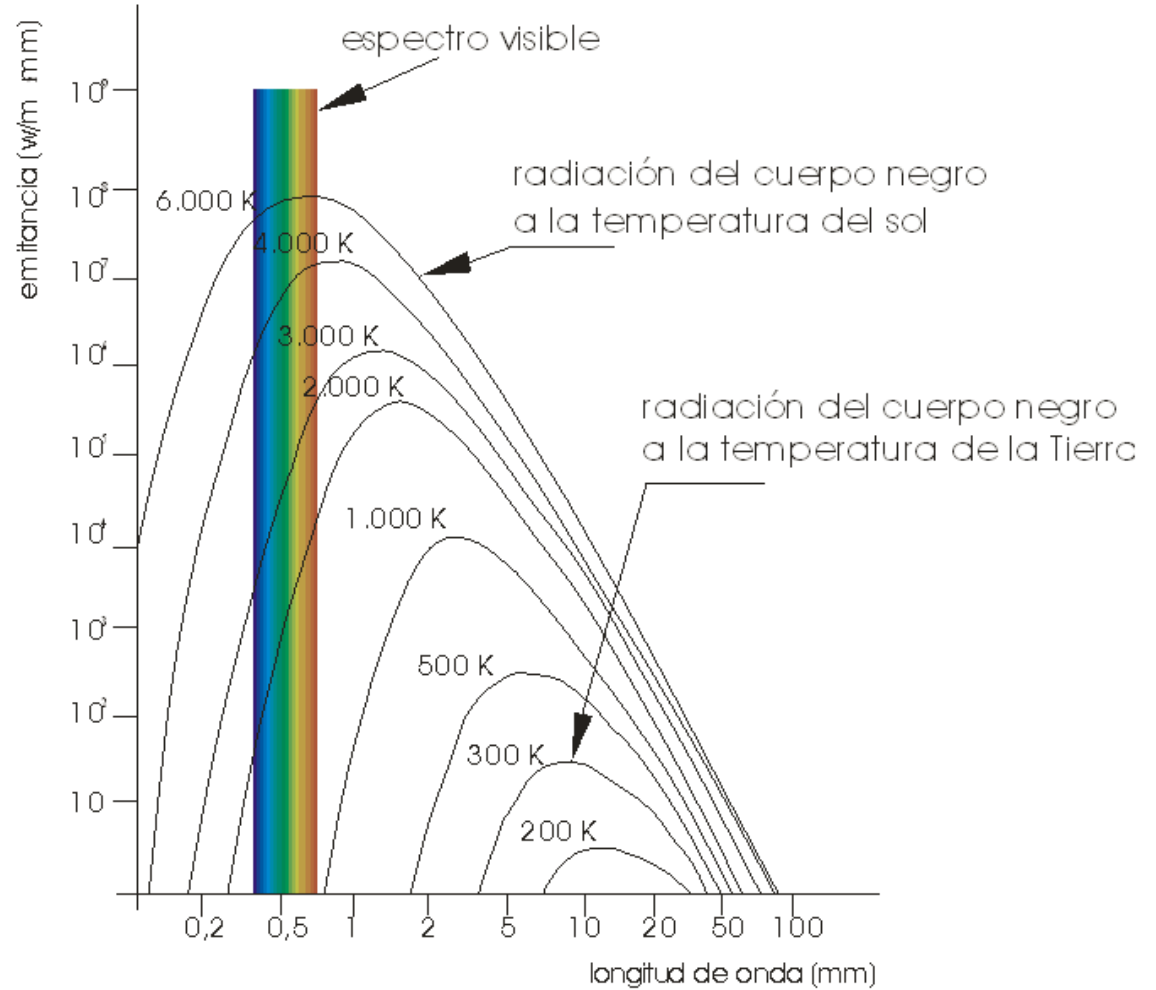
Con lo que la integral anterior quedará:

$$P = \frac{2\pi hc^2 T^4}{(hc/k)^4} \int_0^{\infty} \frac{x^3}{e^x - 1} d\lambda \quad (31)$$

Sólo queda ahora integrar la expresión desarrollando en serie $(e^x-1)^{-1}$ y tomando solamente sus cuatro primeros términos:

$$P = 5,67 \cdot 10^{-8} T^4 \quad (32)$$

En definitiva, la energía irradiada por un cuerpo negro por unidad de área y por unidad de tiempo es



Representación de la Ley de radiación del Planck

proporcional a la cuarta potencia de su temperatura absoluta:

$$P = \sigma T^4 \quad (33)$$

donde σ es la constante de Stefan-Boltzmann ($5,67 \cdot 10^{-8} \text{ W K}^{-4}$). Todos los cuerpos cuya temperatura

no sea el cero absoluto emiten radiación electromagnética. A temperaturas inferiores a los 700 K, estos cuerpos emiten radiación infrarroja. Sin embargo, cuando su temperatura excede de 800 K, la emisión de radiación se realiza en el espectro visible y en longitudes de onda progresivamente más cortas.

Terminología radiométrica

Energía radiante

La energía radiante Q_e es la transportada por la onda electromagnética procedente del emisor. Mide la capacidad de la onda para realizar un trabajo, ya sea calorífico, químico, fotoeléctrico o de cualquier otra naturaleza. Se mide en julios (J).

Flujo radiante

Se llama flujo radiante Φ_e a la energía radiante emitida, reflejada, transmitida o recibida, por unidad de tiempo. Como tiene dimensiones de potencia, se le conoce también por potencia radiante. Se mide en vatios (W).

$$\Phi_e = \frac{Q_e}{t} \quad (34)$$

Densidad de flujo radiante

La densidad de flujo radiante es el flujo radiante interceptado por la unidad de superficie plana. Se mide en vatios por metro cuadrado ($\text{W} \cdot \text{m}^{-2}$). Este es un concepto genérico que engloba a dos parámetros físicos de gran importancia en teledetección:

- Irradiancia E_e es el flujo radiante recibido por una superficie, por unidad de área.
- Emitancia o excitancia radiante M_e es de flujo radiante emitido por una superficie, por unidad de área.

$$E_e, M_e = \frac{\Phi_e}{A} \quad (35)$$

Intensidad radiante

Interesa conocer la concentración de energía en cada región del espacio que rodea a una fuente radiante. Un ángulo sólido es el ángulo espacial abarcado por un objeto desde un determinado punto y se asimila a la zona espacial limitada por una superficie cónica. El estereorradián es la unidad de medida del ángulo sólido y es el subtendido por un sector esférico de 1 m^2 de superficie, perteneciente a una esfera de 1 m de radio. En general se calcula como el cociente entre la superficie esférica subtendida y el cuadrado del radio de la esfera. La esfera completa, pues, ocuparía un ángulo sólido de $4\pi \text{ sr}$. Para calcular el ángulo sólido de una superficie vista desde un punto se proyecta sobre una esfera de radio unidad centrada en ese punto.

La intensidad radiante $I_{e,\Omega}$ es el flujo radiante emitido reflejado, transmitido o recibido por unidad de ángulo sólido. Se trata de una magnitud direccional y se mide en vatios por estereorradián ($\text{W}\cdot\text{sr}^{-1}$).

$$I_{e,\Omega} = \frac{\Phi_e}{\Omega} \quad (36)$$

La unidad de intensidad radiante es la candela (cd), y se define como la intensidad luminosa que tiene una fuente de luz monocromática de frecuencia $540\cdot 10^{12}$ Hz (555 nm) en una determinada dirección cuando su intensidad energética en esa dirección es de 1/683 vatios por estereorradián

Radiancia

La radiancia $L_{e,\Omega}$ es el flujo radiante emitido, reflejado, transmitido o recibido por una superficie, por unidad de ángulo sólido y por unidad de área perpendicularmente proyectada en una determinada dirección. Se mide en vatios por metro cuadrado y estereorradián ($\text{W}\cdot\text{m}^{-2}\cdot\text{sr}^{-1}$).

$$L_{e,\Omega} = \frac{\Phi_e}{A \cdot \Omega \cdot \cos \theta} = \frac{I_{e,\Omega}}{A \cos \theta} \quad (37)$$

siendo θ el ángulo formado por la normal a la superficie radiante y la dirección considerada. Cuando la radiancia emitida o reflejada por una superficie es isotrópica, esto es, independiente de la dirección, la superficie se denomina lambertiana.

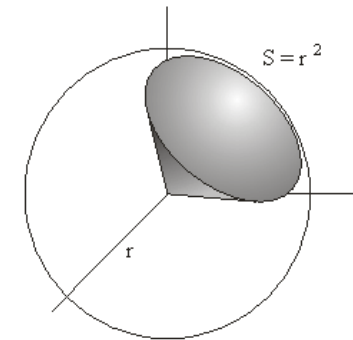
El concepto de radiancia, así como los de emitancia e irradiancia, pueden ser matizados con el calificativo de espectral, dando a entender entonces que se trata de la magnitud considerada, pero solamente para la

radiación perteneciente a un intervalo o a una determinada longitud de onda.

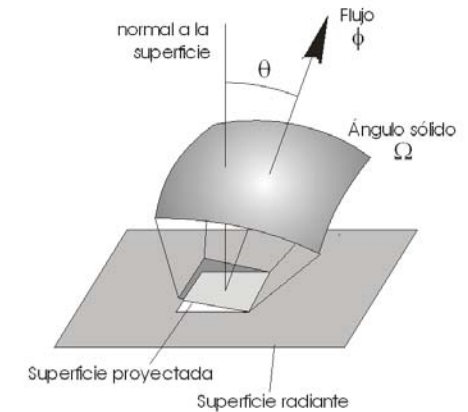
Por ejemplo, el flujo espectral de una radiación sobre una banda del espectro entre λ_1 y λ_2 se obtiene integrando el flujo radiante en el mencionado intervalo de longitudes de onda:

$$\Phi_{e,\lambda_1,\lambda_2} = \int_{\lambda_1}^{\lambda_2} \Phi_e d\lambda \quad (38)$$

El flujo espectral es, pues, el flujo radiante por unidad de frecuencia o de longitud de onda. En este último caso se mediría en vatios por estereorradián, metro cuadrado y nanómetro ($\text{W}\cdot\text{sr}^{-1}\cdot\text{m}^{-2}\cdot\text{nm}^{-1}$).



El ángulo sólido.



Radiancia.

3

Interacciones de la radiación con la materia

Introducción

La teledetección está basada en la adquisición de información procedente de un objeto lejano sin mediación de contacto físico, a partir de la medida de la energía electromagnética que, procedente del blanco, alcanza el sensor. Esta energía puede estar generada por el propio objeto mediante la transformación de otras formas de energía, o ser simplemente la que refleje su superficie tras ser iluminada por una fuente exterior como el Sol o bien un sistema generador instalado en la misma plataforma.

La capacidad de la teledetección para discriminar e identificar superficies o coberturas diferentes reside en que cada una de ellas responde de un modo característico y diferente a la radiación en las distintas zonas del espectro electromagnético, de acuerdo con su composición química y en parte también con su estructura física, las cuales se manifiestan en *signaturas espectrales* diferentes.

Estudiar de un modo general el comportamiento espectral de las superficies que pueden ser detectadas

por un sensor en órbita o aerotransportado viene a concretarse en el estudio de la respuesta de los suelos desnudos –comprendiendo en un sentido amplio todas las posibilidades de presentación del material litológico, desde las rocas vivas hasta los suelos más evolucionados a partir de ellas–, de la vegetación, de la mezcla de ambas en toda su gama de proporciones, del agua y de las superficies de hielo o nevadas, ya que en ellas consiste la totalidad de la superficie firme del planeta, excluidas las áreas urbanizadas o ciertos tipos de superficies artificiales o minoritarias. Por ello, el objeto de este apartado es profundizar en las bases de la respuesta espectral de los suelos, de la vegetación y de sus mezclas, del agua y de la criosfera en amplias regiones del espectro.

La interacción entre la radiación electromagnética y la materia puede estudiarse a nivel macroscópico y a nivel atómico y molecular. Conocer los principios básicos de lo que sucede en estas dos escalas es el objeto de los siguientes apartados.

Interacción macroscópica

Las interacciones macroscópicas entre la radiación y el suelo son debidas a fenómenos tales como la reflexión, la absorción, la transmisión, y la refracción. Cuando un flujo de energía radiante Φ_e^i incide sobre una superficie, parte de él es reflejado hemisféricamente, originando un flujo radiante reflejado Φ_e^r .

Otra parte es absorbida por el material, constituyendo un cierto flujo radiante Φ_e^a . Por último, una fracción Φ_e^t del flujo radiante incidente puede ser transmitida, hacia capas inferiores, a veces en otras formas de energía. Por el principio de conservación, debe verificarse que:

$$\Phi_e^r + \Phi_e^a + \Phi_e^t = \Phi_e^i \quad (1)$$

o lo que es lo mismo:

$$\frac{\Phi_e^r}{\Phi_e^i} + \frac{\Phi_e^a}{\Phi_e^i} + \frac{\Phi_e^t}{\Phi_e^i} = 1 \quad (2)$$

es decir:

$$\rho + \alpha + \tau = 1 \quad (3)$$

siendo:

$$\rho = \frac{\Phi_e^r}{\Phi_e^i}; \quad \alpha = \frac{\Phi_e^a}{\Phi_e^i}; \quad \tau = \frac{\Phi_e^t}{\Phi_e^i} \quad (4)$$

Los términos ρ , α , y τ reciben respectivamente los nombres de *reflectancia*, *absorbancia* y *transmitancia*. Algunos autores enfatizan el carácter relativo de estos parámetros denominándolos respectivamente *reflectividad*, *absortividad* y *transmisividad*.

La emisión de un flujo con posterioridad a la iluminación del terreno se hace a expensas de la energía absorbida previamente y, por tanto, no contabiliza en el trinomio de conservación de la energía.

Si se considera un cuerpo negro, dado que se trata de un receptor de energía perfecto, su reflectancia y su transmitancia deben ser nulas. Por lo general, si el objeto considerado es suficientemente grueso, la transmitancia deberá ser muy pequeña. En ciertos casos, en aras de simplificar la formulación, puede despreciarse esta componente del trinomio si se amplía suficientemente el espesor de la capa estudiada. Por otro lado, si la superficie receptora es oscura y rugosa, la reflectancia será casi nula. Expresada como

porcentaje, la reflectancia también se conoce con el nombre de albedo.

Cuando la superficie de incidencia es suficientemente lisa en relación con la longitud de onda incidente, la reflexión producida se llama especular y su proporción depende del índice complejo de refracción del material y del ángulo de incidencia del haz de radiación.

Numerosas propiedades ópticas del suelo están relacionadas con el índice complejo de refracción:

$$m = n - iK \quad (5)$$

donde la parte real (n) es el índice de refracción real, que depende de las velocidades de la luz en el vacío (c) y en el material en cuestión (v):

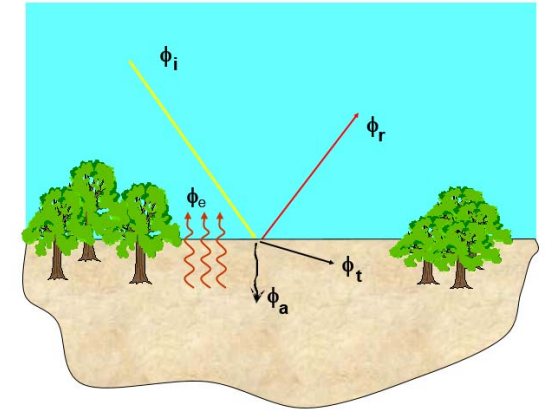
$$n = \frac{c}{v} \quad (6)$$

La parte imaginaria (K) se denomina coeficiente de extinción. Cuando un haz de luz atraviesa una capa isotropa, la disminución relativa de la intensidad de la onda es, según la ley de Lambert-Beer-Bouguer:

$$\frac{-dE}{E} = a dx \quad (7)$$

donde a es una constante que recibe el nombre de *coeficiente de absorción del material* y x es el espesor de la capa atravesada. Integrando se tiene:

$$E_x = E_0 e^{-ax} \quad (8)$$



Distribución del flujo incidente sobre una superficie.

siendo E_x y E_0 los flujos emergente e incidente, respectivamente. Los coeficientes de extinción y de absorción están relacionados mediante la ecuación:

$$K = a\lambda/4\pi \tag{9}$$

en la que λ es la longitud de onda de la radiación. Sustituyendo K :

$$a = \frac{4\pi}{\lambda} K = \frac{4\pi}{\lambda} \frac{m-n}{i} = -\frac{4\pi}{\lambda} (m-n)i \tag{10}$$

con lo cual, la energía emergente es:

$$E_x = E_0 e^{-\frac{4\pi}{\lambda} x(m-n)i} \tag{11}$$

expresión que muestra que la energía emergente en relación con la incidente depende del índice complejo de refracción.

En la naturaleza generalmente la reflexión no es especular, sino que presenta un grado de difusión más o menos acentuado, dependiendo de la rugosidad de la superficie. En este contexto, la rugosidad debe entenderse como un concepto relativo que tiene que ver con la longitud de onda incidente. Superficies que pueden considerarse lisas ante radiaciones de longitudes de onda larga son realmente rugosas estudiadas bajo un haz de radiación de onda corta. El criterio de Rayleigh sirve para establecer cuantitativamente la rugosidad:

$$g = 4\pi \frac{s_h}{\lambda} \text{sen}\theta \tag{12}$$

donde s_h es la desviación típica de la altura en torno al valor medio local, λ la longitud de onda y θ el ángulo de depresión del haz radiante. En general una superficie se considera lisa si $g < 0,3$. En la práctica, el límite a partir del cual la superficie puede considerarse lisa puede calcularse así:

$$h < \frac{\lambda}{8} \cos \theta \tag{13}$$

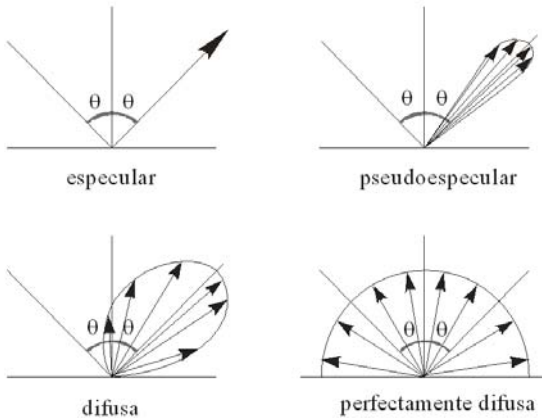
donde h es la altura media de las imperfecciones de la superficie.

Según sea el patrón de la dispersión del haz reflejado, la reflexión puede ser perfectamente difusa, difusa, pseudoespecular o especular. Un reflector es perfectamente difuso cuando refleja isotrópicamente la radiación, es decir, cuando dispersa igual cantidad de energía en todas las direcciones, con independencia de la dirección de incidencia. La reflectancia puede medirse desde todo el hemisferio superior a la superficie, hablándose de reflectancia hemisférica, o solamente desde una dirección determinada en la que se coloca el sensor, hablándose entonces de reflectancia direccional. Este último es el caso de las mediciones efectuadas desde satélite.

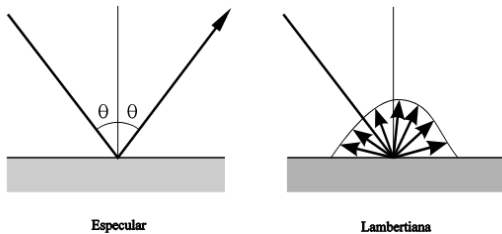
La razón que hay entre el flujo radiante reflejado por una determinada superficie y la que reflejaría una superficie lambertiana ideal con el 100 % de reflectividad se denomina factor de reflectancia.

Función de distribución de reflectancia bidireccional

El brillo de cualquier superficie es diferente cuando se observa desde posiciones distintas y cuando se ilumina desde diferentes direcciones. Este hecho tiene especial repercusión en teledetección, mediante la



Tipos de reflexión sobre una superficie.



Tipos de superficie por su hábito reflexivo.

cual se pretende deducir las características inherentes a la superficie observada, con independencia de su orientación, de su iluminación o de la posición relativa del sensor. Se conoce como función de distribución de reflectancia bidireccional (*Bidirectional Reflectance Distribution Function*, BRDF) la ecuación que define la proporción de reflexión en cada dirección proyectada del rayo reflejado sobre el plano de la superficie, en función de la dirección de incidencia. Tal ecuación viene definida como sigue:

$$f_r(\omega_i, \omega_r) = \frac{L_e(\omega_r)}{E_e(\omega_i)} \quad (14)$$

donde:

ω_i es la dirección del rayo incidente.

ω_r es la dirección del rayo reflejado.

E_e es la irradiancia.

L_r es la radiancia.

La BRDF tiene unidades de sr^{-1} . La irradiancia depende del ángulo cenital θ y del acimut de la proyec-

ción horizontal del flujo incidente ϕ . La propia función f_r depende de esos mismos ángulos y también de los ángulos homólogos que se refieren al flujo emergente, θ' y ϕ' . La radiancia depende de los ángulos anteriores y de la magnitud de la irradiancia. Por ello se puede escribir:

$$f_r(\theta, \phi, \theta', \phi') = \frac{L_r(\theta, \phi, \theta', \phi', E_e)}{E_e(\theta, \phi)} \quad (15)$$

La BRDF depende de la longitud de onda y está determinada por las propiedades estructurales y ópticas de la superficie, tales como la distribución de las sombras proyectadas o del sombreado mutuo, la dispersión múltiple, la transmisión, reflexión, absorción y emisión por los elementos de la superficie, la densidad y la distribución de la orientación de las facetas reflectoras de la superficie. En teledetección, la BRDF es útil para corregir los efectos derivados del punto de vista y de los ángulos de iluminación, a partir de lo cual se puede deducir el albedo, realizar clasificaciones más precisas de la cubierta vegetal, etc.

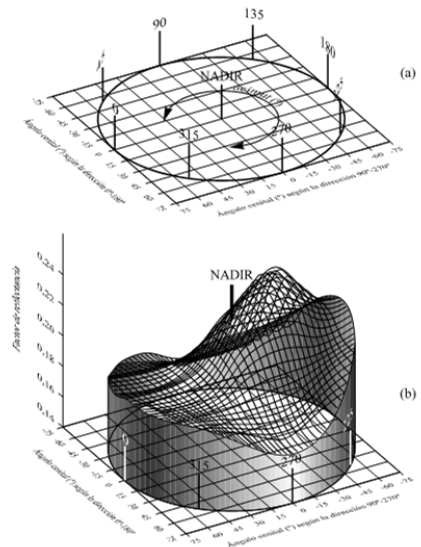
Emisividad

Los cuerpos reales no son receptores ni emisores perfectos y en general su comportamiento radiativo diverge del que tiene el cuerpo negro, en unas longitudes de onda más que en otras. Esto se traduce en diferencias entre la densidad de flujo radiante emitido por una superficie y el emitido por el cuerpo negro para cada temperatura y cada longitud de onda. La relación entre la capacidad radiante de una superficie real en comparación con la del cuerpo negro se denomina emisividad hemisférica ε :

$$\varepsilon = \frac{M_e}{M_e^o} \quad (16)$$

donde M_e y M_e^o son, respectivamente, la emitancia del cuerpo en cuestión y la del cuerpo negro a la misma temperatura.

Como la emisividad de una superficie puede variar



Representación de la BRDF: (a) Situación planimétrica. (b) Envoltura de la energía reflejada.

con la longitud de onda, es necesario definir la emisividad espectral ϵ_λ :

$$\epsilon_\lambda = \frac{M_{e,\lambda}}{M_{e,\lambda}^o} \tag{17}$$

donde el numerador representa la emitancia espectral de la superficie estudiada y el denominador la del cuerpo negro a la misma temperatura que la superficie.

Puede definirse también la emisividad direccional de una superficie ϵ_Ω , utilizando para ello las razones de radiancia de la superficie y del cuerpo negro en lugar de la emitancia:

$$\epsilon_\Omega = \frac{L_{e,\Omega}}{L_{e,\Omega}^o} \tag{18}$$

y la emisividad espectral direccional $\epsilon_{\lambda,\Omega}$:

$$\epsilon_{\lambda,\Omega} = \frac{L_{e,\lambda,\Omega}}{L_{e,\lambda,\Omega}^o} \tag{19}$$

En función del valor de su emisividad hemisférica ϵ_λ , los cuerpos se pueden clasificar en:

- Radiadores perfectos, correspondiendo esta calificación sólo al cuerpo negro, para el que $\epsilon_\lambda = 1$.
- Cuerpos grises, cuando $0 < \epsilon_\lambda < 1$, siendo ϵ_λ constante.
- Reflectores perfectos, en los que se verifica que $\epsilon_\lambda = 0$.

Cuando la emisividad espectral de una superficie es variable de acuerdo con la frecuencia de la emisión,

esto es, cuando $\epsilon_\lambda = f(\lambda)$, se trata de un radiador selectivo. Es el caso más frecuente en la naturaleza. Precisamente en esta propiedad se basa la identificación de superficies mediante teledetección, pues cada cuerpo se comporta de forma diferente frente a distintas frecuencias de radiación incidente, haciendo con ello posible su identificación. Cada superficie queda caracterizada por su signatura o firma espectral.

Ley de Kirchoff de la radiación

Si para una superficie con un determinado espesor puede considerarse que $\tau = 0$, la parte no reflejada del flujo incidente sobre ella se absorbe y se transforma en calor, como consecuencia de lo cual se incrementa su temperatura. En un cierto momento se alcanza el equilibrio térmico y, para ello pueda suceder, es necesario que la superficie emita el exceso de energía absorbida. Una vez establecido el equilibrio, se verificará que:

$$\rho = \frac{\Phi_e^r}{\Phi_e^i}; \quad \alpha = \frac{\Phi_e^a}{\Phi_e^i}; \quad \tau = \frac{\Phi_e^t}{\Phi_e^i} \tag{20}$$

$$\Phi_e^a = \Phi_e^{em} = M_e A \tag{21}$$

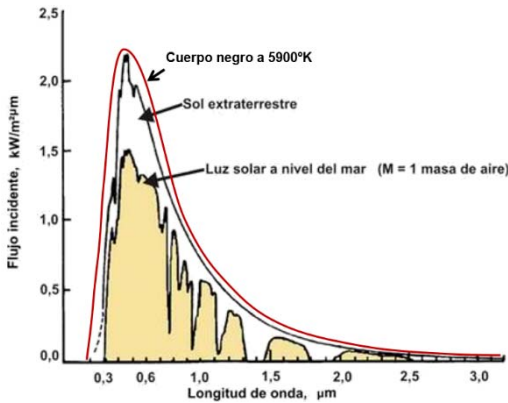
donde:

Φ_e^{em} es el flujo radiante emitido.

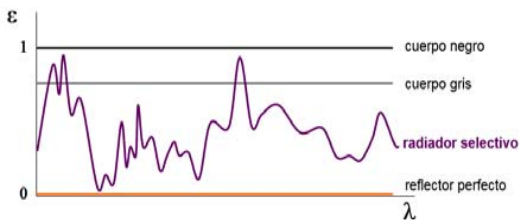
A es el área.

En el caso del cuerpo negro, todo el flujo radiante incidente es absorbido, por tratarse de un receptor de energía perfecto, por lo que se puede escribir:

$$\Phi_e^i = \Phi_e^{o,a} = \Phi_e^{o,em} = M_e^o A \tag{22}$$



Emisancia del cuerpo negro e irradiancia sobre otras superficies.



Emisividad de los distintos tipos de cuerpos.

donde

Φ_e^i es el flujo radiante incidente y todos los demás términos tienen el mismo significado que en la expresión anterior, únicamente que ahora son aplicados al cuerpo negro, que es lo que indica el superíndice 'o'.

Dividiendo miembro a miembro las dos últimas expresiones:

$$\frac{\Phi_e^a}{\Phi_e^i} = \frac{M_e}{M_e^o} \quad (23)$$

es decir:

$$\alpha = \varepsilon$$

que es la expresión de la ley de Kirchhoff de la radiación. En condiciones de equilibrio térmico, la absorbancia de una superficie es igual a su emisividad. En realidad, el enunciado de la ley debe aplicarse a cada longitud de onda, de esta forma:

$$\alpha_\lambda = \varepsilon_\lambda \quad (24)$$

Esta relación, válida para los radiadores selectivos en las bandas del espectro en las que son opacos, puede transformarse en esta otra:

$$M_\lambda = \alpha_\lambda M_\lambda^o \quad (25)$$

es decir, que para una determinada longitud de onda, el poder emisor de un cuerpo cualquiera es igual al del cuerpo negro afectado de un coeficiente igual a su absorbancia. Además, si la transmitancia no se considera, sustituyendo la absorbancia por la emisividad, puede reescribirse la ecuación de conservación de la energía de esta manera:

$$\rho_\lambda + \varepsilon_\lambda = 1 \quad (26)$$

lo que significa que reflectancia y emisividad son magnitudes complementarias, esto es, que superficies altamente reflectantes como la nieve en el espectro visible, cuyo albedo es superior al 60 %, son poco emisivas, en tanto que las poco reflectoras son altamente emisivas, como es el caso del agua líquida.

Debe hacerse la salvedad, por otro lado, de que las expresiones anteriores proporcionan la emisividad hemisférica. Sin embargo, la emisividad de una superficie real puede depender de la dirección de emisión. Por esta razón, el parámetro a estudiar debe ser la emisividad direccional, que es la relación de emitancia entre la superficie de estudio y el cuerpo negro en la dirección deseada.

Interacción atómica y molecular

La energía electromagnética que incide sobre un cuerpo es parcialmente absorbida, como se ha visto. De acuerdo con la física cuántica, esta absorción de energía produce alteraciones en el estado energético

de la materia que se corresponden con niveles energéticos discretos. La energía radiante absorbida por los átomos excita su corteza y promueve los electro-

nes a niveles energéticos superiores dentro de los orbitales atómicos disponibles. El salto energético ΔE viene cuantificado por:

$$\Delta E = h\nu \quad (27)$$

siendo h la constante de Planck y ν la frecuencia de la radiación.

Los orbitales atómicos son las soluciones de la ecuación de Schrödinger, la cual define la probabilidad de la distribución electrónica. Para el átomo más sencillo con un solo electrón, el de hidrógeno, la ecuación de Schrödinger es:

$$\nabla^2\Psi + \frac{2m}{\hbar^2}(E - V)\Psi = 0 \quad (28)$$

en la que Ψ^2 representa la probabilidad de encontrar un electrón en una posición determinada (x,y,z) , m es la masa del electrón, V su energía potencial, E la energía total en el instante considerado, ∇ el operador laplaciano y:

$$\nabla^2 = \frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} + \frac{\partial^2\Psi}{\partial z^2} \quad (29)$$

Cada una de las soluciones de la ecuación de Schrödinger constituye un orbital atómico. Debido al *principio de indeterminación de Heisenberg*, no es posible conocer simultáneamente y con precisión absoluta la posición y la energía (momento) de un electrón. Por ello, los orbitales sólo son funciones probabilísticas y no deterministas. Los electrones siempre tienden a ocupar el nivel inferior de energía de entre aquéllos que estén disponibles. Por el principio de

exclusión de Pauli, ningún estado energético puede estar ocupado por más de un electrón. Los saltos de los electrones excitados mediante una radiación determinada solamente se realizan entre dichos orbitales. Posteriormente los electrones vuelven a su estado original cediendo energía en forma de radiación electromagnética, en la misma o en diferente longitud de onda dependiendo de los saltos intermedios que pueda efectuar.

La formación de enlaces covalentes entre átomos para constituir moléculas origina la aparición de orbitales moleculares. El significado de éstos es análogo al de los orbitales atómicos, con la salvedad de que en las nubes de probabilidad se encuentran los electrones compartidos por los átomos constituyentes de la molécula. La estabilidad de la molécula queda explicada por la menor energía de los electrones en el orbital molecular que en la suma de los orbitales atómicos originales. Las transiciones electrónicas entre distintos orbitales moleculares se producen con la absorción o emisión de energía electromagnética. La energía necesaria para promover el salto, o la que el electrón devuelve al medio en forma de radiación electromagnética tras caer a su estado original o a cualquier otro estado energético inferior viene dada por:

$$\begin{aligned} E\Psi &= \mathbf{H}\Psi = \\ &= \frac{-\hbar^2}{2m} \left[\frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} + \frac{\partial^2\Psi}{\partial z^2} \right] + V\Psi = \\ &= \frac{-\hbar^2}{2m} \nabla^2\Psi + V\Psi \end{aligned} \quad (30)$$

donde \mathbf{H} es el operador Hamiltoniano.

En virtud de la excitación electromagnética, los átomos pueden adquirir una cierta energía que se traduce en movimientos del núcleo o en un cambio de estado energético de los electrones. Los movimientos nucleares pueden ser traslaciones, rotaciones o vibraciones. En el material inerte de los suelos, la interacción de la luz solar se traduce predominantemente en cambios de los estados vibracionales y de los niveles energéticos de los electrones, quedando muy restringidos los fenómenos de traslación y rotación nuclear. Las vibraciones moleculares se manifiestan en la zona infrarroja del espectro.

En general, las bandas de absorción pueden ser fundamentales, armónicas o combinaciones. Las primeras se refieren a la absorción de energía por parte de las moléculas, promoviéndose al nivel energético vibracional inmediatamente superior al de reposo. Si en cambio, la transición energética incluye dos o más niveles sobre el fundamental, se habla de bandas armónicas de absorción. Por último, las bandas de combinación se deben al reparto de la energía de un fotón para elevar el nivel de energía molecular de más de un estado vibracional. Los picos principales en los espectros de absorción de los minerales se deben a vibraciones fundamentales, por ser las más frecuentes. Por el contrario, los picos debidos a los armónicos y a las bandas de combinación producen picos más débiles por la razón contraria. Las transiciones fundamentales de los minerales del suelo se manifiestan en la región del infrarrojo, por cuya razón es en este intervalo espectral en el que suelen ser estudiados en laboratorio. Existen numerosas librerías de espectros de absorción/transmisión de minerales y vegetales realizados en laboratorio en condiciones controladas y sobre muestras homogéneas y pequeñas.

Sin embargo, las bandas de absorción que caracterizan la teledetección de suelos en la región solar del espectro son sobre todo armónicas y de combinación.

Por ello, la identificación de la composición del suelo a partir de datos remotos no es tan evidente como en los ensayos de laboratorio. De otra parte, las transiciones electrónicas también producen bandas de absorción, si bien éstas son mucho más anchas que las correspondientes a los fenómenos vibracionales y de un mayor nivel energético, por lo que se manifiestan en el visible y el ultravioleta.

A título de ejemplo, la captura de la energía solar por parte del material vegetal vivo se realiza precisamente mediante la excitación de los electrones de un cierto tipo de orbitales moleculares, denominados π , de las moléculas de dos tipos de clorofila, *a* y *b*, que se promueven a estados energéticos superiores. Posteriormente la energía es cedida (fenómeno de transmisión) a otras moléculas mediante procesos de reducción química. El máximo de absorción de energía se verifica para longitudes de onda de 673 y 683 nm. La clorofila aislada de otras moléculas en laboratorio emite la energía captada mediante fluorescencia, es decir, devuelve la energía previamente absorbida en otras longitudes de onda.

La firma espectral es, en definitiva, consecuencia de la absorción selectiva de la radiación por parte de los componentes de una superficie, entre los que, casi siempre, el agua juega un papel destacado.

4

Interacción atmosférica

Introducción

La atmósfera es un cuerpo interpuesto entre la superficie objeto de estudio y el sensor. Para algunos intervalos de longitudes de onda resulta prácticamente diáfana y su presencia no altera la detección de la radiación electromagnética. Sin embargo, en determinadas bandas del espectro, en las que puede resultar completamente opaca, la presencia de la atmósfera afecta significativamente a la señal. Es necesario conocer cómo interacciona la radiación en el proceso de transferencia radiativa, tanto en lo que se refiere a su camino descendente, hasta iluminar la superficie terrestre, como en el ascendente hasta ser recibida a bordo del sensor.

La teledetección basada en plataformas a baja altura, como es el caso de los vehículos no tripulados, o incluso desde aeronaves, no requiere correcciones atmosféricas importantes, o bien ni si quiera se llega a tomar en consideración la existencia de la atmósfera. Sin embargo, cuando la señal se captura orbitalmente, hay que tener en cuenta que el satélite tiene interpuesta entre su posición y la superficie terrestre la atmósfera en su totalidad, y los efectos de ésta no pueden ser despreciados.

Las causas por la que la atmósfera interfiere en la transferencia radiativa hay que buscarlas tanto en su

composición química como en los aerosoles que contiene. Los aerosoles son soluciones coloidales, con partículas en suspensión con tamaños comprendidos entre 1 y 200 μm , en las que el medio de dispersión es un gas, como en el caso de la propia atmósfera. La fase dispersa, por el contrario, puede estar constituida por sólidos, como por ejemplo los humos, las polvaredas o la calima, o por líquidos, como es el caso de las nubes.

Los principales fenómenos físicos que se producen por la interposición de la atmósfera entre la superficie reflectante y el sensor son la *absorción* y la *dispersión*. Por un lado, la absorción reduce la energía disponible en una determinada longitud de onda. Por otro, la dispersión altera la dirección del flujo electromagnético, produciendo una disminución del contraste de la escena producida por el desvío de los rayos procedentes de la superficie y a la suma de rayos procedentes de otros lugares de la superficie terrestre e incluso de la misma atmósfera. La interacción atmosférica se manifiesta de modo especial en las bandas del infrarrojo, visible y ultravioleta, y es en cambio poco significativa en la región de las microondas.

Componentes de la atmósfera

La atmósfera es una mezcla de gases y aerosoles en proporciones variables según la altura. Se habla de *homosfera* al considerar los primeros 80 o 100 km

de altura desde la superficie. Esta capa se caracteriza por la homogeneidad en su composición como consecuencia de la mezcla originada por fenómenos de

convección y turbulencia. La composición volumétrica (v/v) para una atmósfera seca es:

Gas	Volumen
Nitrógeno (N ₂)	780 840 ppmv (78,084 %)
Oxígeno (O ₂)	209 460 ppmv (20,946 %)
Argón (Ar)	9 340 ppmv (0,934 %)
Dióxido de carbono (CO ₂)	400 ppmv (0,04 %)
Neón (Ne)	18,18 ppmv (0,001818 %)
Helio (He)	5,24 ppmv (0,000524 %)
Metano (CH ₄)	1,79 ppmv (0,000179 %)
Kriptón (Kr)	1,14 ppmv (0,000114 %)
Hidrógeno (H ₂)	0,55 ppmv (0,000055 %)
Óxido nitroso (N ₂ O)	0,3 ppmv (0,00003 %)
Xenón (Xe)	0,09 ppmv (9x10 ⁻⁶ %)
Ozono (O ₃)	0,0 - 0,07 ppmv (0 % a 7x10 ⁻⁶ %)
Dióxido de nitrógeno (NO ₂)	0,02 ppmv (2x10 ⁻⁶ %)
Yodo (I)	0,01 ppmv (1x10 ⁻⁶ %)
Monóxido de carbono (CO)	0,1 ppmv
Amoniaco (NH ₃)	trazas

El vapor de agua, que se encuentra en unas proporciones en torno a 0,40 %, llega a alcanzar entre el 1 y el 4 % del volumen a nivel de superficie. También hay presencia de diversos aerosoles. La densidad atmosférica varía con la altura como consecuencia de la compresión diferencial a que es sometida por acción de la gravedad.

Por encima de la homósfera se halla la *heterósfera* que se extiende hasta los 10 000 km de altura aproximadamente. La composición gaseosa ya no es constante debido a fenómenos de difusión molecular, en-

contrándose mayor concentración de componentes ligeros y de forma estratificada, de acuerdo a su peso atómico o molecular: N₂ entre 80 y 400 km, O a partir de ahí hasta 1100 km, He hasta 3500 km y finalmente hidrógeno, en sus formas atómica y molecular, hasta 10 000 km. La heterósfera recibe también los nombres de *ionósfera*, dado que los gases se encuentran ionizados por efecto de la radiación, y de *termósfera*, debido a que el oxígeno atómico, al absorber la radiación ultravioleta, calienta la atmósfera hasta temperaturas de 200 y 300 °C.

La absorción atmosférica es atribuible a la absorción individual de sus componentes. La dispersión causada por los gases, por su parte, tiene que ver con las variaciones de la densidad molecular, por lo que el efecto resultante depende en último término de la presión atmosférica local.

En una atmósfera isoterma, la presión depende de la altura de acuerdo con la ley:

$$P = P_0 e^{-\frac{M \cdot g \cdot (h-h_0)}{R \cdot T}} \quad (1)$$

donde M es la masa molecular del gas, g la aceleración de la gravedad, R la constante de los gases perfectos (8,313 J·K⁻¹·kmol⁻¹), T la temperatura absoluta media entre los dos niveles, h y h_0 son las alturas a las que les corresponden respectivamente las presiones P y P_0 .

La densidad ρ de la atmósfera en ausencia de humedad es una función exponencial de la altura z desde el nivel del mar:

$$\rho = \rho_0 e^{-\frac{h-h_0}{H}} \quad (2)$$

donde H representa la *escala de altura*:

$$H = \frac{RT}{Mg} \quad (3)$$

H es la altura a la que hay que elevarse para que la presión atmosférica disminuya en un factor $e=2,718281$, o sea al 63,2 %. En la Tierra, la escala de altura es de 8,42 km. En función de la escala de altura, también se puede escribir la presión:

$$P = P_0 e^{-\frac{h-h_0}{H}} \quad (4)$$

donde P_0 es la presión atmosférica a nivel del mar.

Se conoce como *incremento de altura* a la altura a la que es necesario elevarse en la atmósfera para que la presión disminuya a la mitad. En la Tierra, el incremento de altura es de 5,8 km.

La presión atmosférica varía ligeramente con el tiempo y de un lugar a otro; sin embargo, es una variable relativamente estable, pues la desviación típica sobre sus valores promedios no supera el 1%. La concentración de vapor de agua, por el contrario, es más variable, y se encuentra principalmente en altitudes inferiores a 1 o 2 km. La variabilidad de la concentración de vapor de agua es el resultado de la heterogeneidad en la distribución espacial de las fuentes productoras –superficies evapotranspirantes– y de los sumideros de humedad –principalmente nubes–. El anhídrido carbónico se distribuye más uniformemente en la atmósfera seca, a excepción de las zonas próximas a focos de emisión, como son las grandes aglomeraciones urbanas o industriales. El ozono, por último, se concentra sobre todo en la estratosfera, a una distancia de entre 20 y 50 km de la superficie.

Una solución coloidal es aquella cuyas partículas en suspensión tienen tamaños comprendidos entre 1 y 200 μm . Los aerosoles son soluciones coloidales en la que el medio de dispersión es un gas o una mezcla de gases, como por ejemplo la atmósfera. La fase dispersa puede ser un sólido, como los humos provocados por un incendio o las polvaredas, o un líquido, como las nubes, por ejemplo.

Los aerosoles atmosféricos tienen diversos orígenes. Las principales fuentes de formación de aerosoles atmosféricos son las siguientes:

- **Partículas procedentes de superficies sólidas.** Las fuerzas mecánicas debidas al viento producen el desprendimiento de partículas microscópicas de la superficie terrestre. El período de vida de la partícula de polvo en suspensión depende de su tamaño.
- **Partículas procedentes del mar.** El principal proceso de formación de aerosoles de procedencia directamente marina es el estallido de burbujas de agua en la superficie.
- **Partículas procedentes de gases.** Existen gases atmosféricos químicamente activos (sulfatos y nitratos, por ejemplo), emitidos bien de forma natural por las plantas y las erupciones volcánicas, bien artificialmente como consecuencia de procesos industriales o de incineración de residuos, que, mediante reacciones químicas en el seno de la atmósfera, pasan a fase líquida y quedan en suspensión.

El efecto óptico causado por los aerosoles atmosféricos depende de las características físicas de las partículas componentes:

- Distribución de tamaños.
- Índice de refracción, relacionado con la composición química de las partículas, que no han de ser

necesariamente homogéneas en su interior y que puede variar en función de su propio tamaño.

- Forma de la partícula, especialmente su carácter esférico o no.
- Distribución espacial de la concentración de partículas. La distribución de tamaños de las partículas de un aerosol atmosférico, asumiendo la simplificación de considerarlas esféricas, sigue una ley lognormal:

$$\frac{dn}{dr} = \frac{N_0}{\ln(10) \cdot r \cdot \sigma \cdot \sqrt{2\pi}} e^{-\frac{(\log r - \log r_0)^2}{2\sigma^2}} \quad (5)$$

donde N_0 es la densidad de las partículas del aerosol, r es su radio, σ es la desviación típica de $\ln r$, y r_0 es la media geométrica de los radios.

Dispersión atmosférica

El fenómeno de dispersión de la radiación electromagnética o *scattering* es fruto de su interacción con los gases y aerosoles. Los choques producidos entre las ondas y los componentes atmosféricos, moléculas o partículas en suspensión, ocasionan el desvío de las mismas de sus trayectorias primitivas, originando un incremento de la radiancia procedente de la misma atmósfera, que da lugar a lo que se conoce como luz atmosférica. La consecuencia directa en teledetección es la disminución del contraste en las imágenes adquiridas desde una plataforma a gran altura, tanto más severa cuanto mayor sea. La dispersión puede ser selectiva o no selectiva, según afecte o no a determinadas longitudes de onda con preferencia sobre otras. La dispersión es no selectiva cuando las partículas dispersantes son suficientemente grandes en comparación con la longitud de onda de la radiación, y es selectiva en caso contrario.

Existen dos tipos de dispersión selectiva. La dispersión de *Rayleigh* (fonéticamente, ['reili]) es la que se produce cuando el tamaño de las partículas es mucho menor que longitud de onda de la radiación, y afecta de modo especial a las longitudes de onda más cortas.

Dispersión de Rayleigh

El campo eléctrico de la radiación polariza las partículas que intercepta, al actuar sobre las cargas superficiales y provocar la oscilación en su misma frecuencia. Esto las convierte en dipolos que emiten radiación en distintas direcciones, lo que da origen a la luz dispersada.

La intensidad I de la luz dispersada por una partícula bajo un flujo de energía radiante de intensidad I_0 y de longitud de onda λ viene cuantificada por:

$$I = I_0 \frac{1 + \cos^2 \theta}{2R^2} \left(\frac{2\pi}{\lambda} \right)^4 \left(\frac{D}{2} \right)^6 \quad (6)$$

donde:

I_0 = intensidad de la luz incidente

R = distancia a la partícula

θ = ángulo de dispersión

λ = longitud de onda.

D = diámetro equivalente de la partícula

La distribución angular de la dispersión de Rayleigh es simétrica dentro del plano perpendicular al rayo incidente. La contribución de la partícula a la dispersión se evalúa con la sección transversal de dispersión de Rayleigh σ_s , que es la sección de una esfera que proporcionase idéntica dispersión que la partícula:

$$\sigma_s = \frac{2\pi^5 D^6}{3\lambda^4} \left(\frac{n^2 - 1}{n^2 + 2} \right)^2 \quad (7)$$

donde n es el índice complejo de refracción.

Esta última expresión es coherente con las medidas experimentales que muestran que el espesor óptico de Rayleigh es función de la longitud de onda λ de la radiación. Empíricamente se ha encontrado que para una presión atmosférica $P_0=1013,25$ hPa:

$$\tau_r = 0,008569 \lambda^{-4} \left(1 + 0,0113 \lambda^{-2} + 0,00013 \lambda^{-4} \right) \quad (8)$$

donde λ viene expresada en μm y se ve de nuevo la proporcionalidad inversa del efecto atmosférico sobre todo con la cuarta potencia de la longitud de onda.

Considerando que una unidad de volumen contuviera N partículas de igual tamaño, el coeficiente de dispersión en ese medio vendría dado por:

$$\alpha_s = \sigma_s N \quad (9)$$

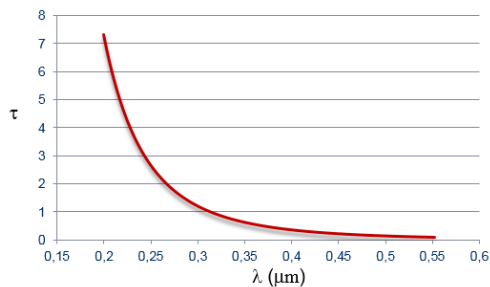
Este tipo de dispersión es omnidireccional, con un predominio ligero de la dirección de propagación, en ambos sentidos, sobre la perpendicular y es fuerte-

mente dependiente de la longitud de onda (inversamente proporcional a la cuarta potencia de la longitud de onda). Por ello resultan más dispersadas las radiaciones de onda corta, y de este modo se convierte en la responsable del color azul del cielo, y del efecto de neblina azulada cuando se divisan objetos a grandes distancias bajo atmósferas sucias. De no existir partículas dispersantes, el color del cielo fuera del círculo solar sería negro, tal como se observa desde el exterior de la atmósfera.

Dispersión de Mie

La dispersión de *Mie* (fonéticamente, [mi:]) se produce cuando el tamaño de las partículas es del mismo orden de magnitud que la longitud de onda de la radiación, desde tamaños superiores a 1/10 de la longitud de onda. Aunque puede estar causada también por aerosoles, es más propia del efecto que causan las moléculas de los gases atmosféricos. La dispersión de Mie dispersa más radiación hacia la dirección y el sentido del haz luminoso y crece con el tamaño de las partículas., aunque no depende tan fuertemente de la longitud de onda como la dispersión de Rayleigh. La propuesta de Mie a principios del S XX es la formulación completa de las ecuaciones de Maxwell para la dispersión de ondas electromagnéticas en partículas esféricas, pero carece de una solución analítica, de modo que ha tenido que esperar al desarrollo del cálculo numérico para cuantificar su efecto. La dispersión de Mie es la responsable del resplandor casi blanco alrededor del disco solar, del color blanco de la bruma y la calima y del color de algunos amaneceres, más rojizo de lo que sería explicable solamente por la dispersión de Rayleigh.

Por su parte la dispersión no selectiva es la suma de los fenómenos de reflexión con o sin penetración en el interior de las partículas, de refracción y de difracción y está causada por partículas significativamente



Variación del espesor óptico con la longitud de onda.

más grandes que la longitud de onda de la radiación. Afecta de modo generalizado a todas las longitudes de onda y produce una disminución del flujo radiante con la consiguiente atenuación de la señal detectada

por el sensor. La dispersión no selectiva está ocasionada por las nubes y el polvo en suspensión. El color blanco de las nubes se debe precisamente a la dispersión indiscriminada de todas las longitudes de onda de la luz visible que las atraviesa.

Absorción atmosférica

El fenómeno de absorción atmosférica se produce cuando los átomos o las moléculas de los componentes de la atmósfera son excitados por la radiación electromagnética. Los fotones incidentes ceden su energía a las partículas interceptoras, cuyos electrones corticales son promovidos a niveles energéticos más elevados. La absorción de energía se lleva a cabo en diversas bandas del espectro de acuerdo con la composición química de la partícula. El flujo emergente, esto es, el que consigue atravesar el material, presenta en general una composición de longitudes de onda diferente a la del haz incidente, como consecuencia de la absorción selectiva. Su espectro contendrá unos ciertos mínimos centrados en las longitudes de onda absorbidas por el material. Estos mínimos se denominan *bandas de absorción*. La atmósfera se comporta como un filtro selectivo a distintas longitudes de onda y, en algunos intervalos, la absorción llega a ser total, limitando o imposibilitando la observación remota de la superficie. Los principales absorbentes atmosféricos son el vapor de agua, anhídrido carbónico, metano, óxido de nitrógeno, oxígeno atómico y molecular, y ozono.

De todos los componentes de la atmósfera, el vapor de agua es el mayor absorbente y presenta diversas bandas de absorción entre 0,7 y 8 μm , particularmente importantes las situadas en 1,4, 1,9, 3 μm y

por encima de 27 μm . La situada en 6 μm puede llegar a ser hasta del 100 % si la humedad atmosférica es alta. En la región de las microondas, la atmósfera absorbe el 100 % de la radiación en 1,7 μm y un 12 % en 15 mm (176 y 20 GHz, respectivamente). El vapor de agua es un gas inodoro e incoloro principal responsable del mantenimiento de la temperatura en el planeta mediante el denominado *efecto invernadero*.

Este fenómeno —que toma su nombre del efecto causado por las láminas plásticas o de vidrio que cubren estas construcciones destinadas a aumentar y conservar la temperatura interior para facilitar los cultivos agrícolas— se caracteriza porque la radiación de onda corta, tras atravesar una capa transparente como la atmósfera, incide y es absorbida por la superficie. Alcanzado el equilibrio térmico, la propia superficie comienza a emitir radiación de onda larga que no consigue atravesar la atmósfera, por ser ésta opaca en esas longitudes de onda. Los gases de efecto invernadero absorben la energía infrarroja, pero los fotones infrarrojos no son lo suficientemente energéticos como para romper los enlaces covalentes de sus moléculas, sino que solo incrementan su energía rotacional y vibracional. La energía absorbida se transfiere mediante colisiones a otras moléculas, de tal modo que se incrementa la energía cinética del conjunto, esto es, sube la temperatura de la atmósfera.

Este fenómeno es reversible cuando la atmósfera se enfría transfiriendo energía en forma de calor a la superficie. Todos los gases implicados en el efecto invernadero están constituidos por moléculas dipolares, con distribución no simétrica de sus cargas, ya que las transiciones energéticas citadas requieren cambios en el momento dipolar de las moléculas. Por esta razón los principales constituyentes atmosféricos, como el nitrógeno y el oxígeno no participan en el mantenimiento de la temperatura, ya que sus moléculas carecen de momento dipolar, al estar formadas por dos átomos iguales.

El anhídrido carbónico es otro gas de efecto invernadero que presenta la mayor banda de absorción en torno a 15 μm y otras menores entre 2,5 y 4,5 μm . Otros gases absorbentes en la región de infrarrojo térmico son el metano y el óxido de nitrógeno. El gradiente térmico que produce la absorción de radiación por parte de los gases de efecto invernadero es de -6,5 K/km de altura, hasta llegar a la tropopausa, a partir de la cual el gradiente comienza a ser positivo.

Las longitudes de onda cortas son absorbidas por las capas altas de la atmósfera y por ello en la superficie terrestre no se registra radiación de longitud de onda inferior a 0,3 μm . Las longitudes de onda inferiores a 0,1 μm son absorbidas por el oxígeno atómico (O) y el molecular (O₂). El oxígeno absorbe también en la región de las microondas, presentando dos profundas bandas de absorción del 100 % en 2,5 y 5,23 mm (120 y 58 GHz, respectivamente). Las longitudes de onda comprendidas entre 0,1 y 0,3 μm son absorbidas por la capa de ozono (O₃) situada a 35 km de altura, la cual preserva a las células vivas de la radiación ultravioleta. El ozono también presenta bandas de absorción en la región del térmico, en torno a 9,5 μm y varias más en la de las microondas. En esta región, el oxígeno

presenta bandas de absorción profundas en torno a 2,6 y 5,0 mm, aunque para longitudes de onda superiores a 1 m la atmósfera es prácticamente transparente a la radiación.

Como consecuencia de su absorción selectiva para determinadas longitudes de onda, la atmósfera presenta una serie de zonas espectrales en las que es transparente a la radiación electromagnética. En cambio, actúa como un eficaz filtro en otras longitudes de onda en las que resulta completamente opaca a la radiación. Las regiones espectrales en las que la absorción es escasa reciben el nombre de *ventanas atmosféricas*, y es a través de ellas por las que es posible la teledetección de la superficie terrestre. La zona del visible resulta prácticamente transparente, sin embargo, en la región de infrarrojo, la atmósfera presenta diversas ventanas de menor anchura espectral. Las ventanas atmosféricas principales son:

Intervalo espectral (μm)		
0,3	-	1,3
1,5	-	1,8
2,0	-	2,6
3,0	-	3,6
4,2	-	5,0
7,0	-	15,0

En el espectro visible, la transmitancia de la atmósfera está afectada principalmente por la absorción del ozono y por la dispersión de *Rayleigh*. Su contribución a la transmitancia vertical T_r se expresa mediante espesor óptico τ_r :

$$\tau_r = -\ln T_r \quad (10)$$

O, dicho de otra forma:

$$T_r = e^{\tau_r} \quad (11)$$

Puesto que, según la Ley de Boyle, la densidad del aire es proporcional a su presión P , el espesor óptico vertical entre el exterior de la atmósfera y un punto de presión P puede ser expresado por:

$$\tau_r(P) = \tau_r(P_0) \frac{P}{P_0} \quad (12)$$

donde P_0 es la presión atmosférica a nivel del mar.

Óptica atmosférica

La variabilidad de las propiedades ópticas de la atmósfera son principalmente consecuencia de la variabilidad de la absorción gaseosa y de la variabilidad en el contenido y características de los aerosoles atmosféricos. Una porción de aire con aerosoles en suspensión puede absorber parte de la radiación incidente y dispersar el resto. De este modo, las propiedades ópticas de un pequeño volumen de aerosol pueden ser descritas por los parámetros siguientes:

Coefficiente de extinción K_e .

Describe la fracción de radiación que queda confinada dentro del aerosol con respecto a la radiación incidente. Se mide en km^{-1} .

Albedo debido a la dispersión simple ω_0 .

Es la proporción de dispersión en relación con la extinción total. Se calcula como el cociente entre el coeficiente de dispersión y el de extinción.

Función de fase de la dispersión $P(\theta)$.

Representa la distribución angular de la radiación dispersada, siendo θ el ángulo considerado.

La relación entre las características físicas del aerosol, como la distribución de tamaños y los índices de refracción, y sus propiedades ópticas fueron definidas

para partículas esféricas homogéneas por la teoría de Mie, quien demostró que si el radio de las partículas es mucho menor que la longitud de onda de la radiación ($r \ll \lambda$), entonces la sección transversal de extinción β (cm^2), definida como la sección transversal efectiva que interacciona con la radiación, es mucho más pequeña que la sección geométrica de la partícula. Por consiguiente el factor de eficiencia en la extinción Q , calculado como la razón entre la sección transversal de extinción y la total, es: $Q(r/\lambda) = \beta(r)/\pi r^2 < 1$. Sin embargo, para partículas de tamaño considerablemente mayor que la longitud de onda ($r \gg \lambda$), la sección transversal de extinción es alrededor del doble que el valor de la sección geométrica. Entre ambos extremos, Q se comporta como una función del cociente entre el tamaño de la partícula y la longitud de onda, alcanzando el valor $Q = 4$ para $r/\lambda \approx 1$.

El coeficiente de extinción para un pequeño volumen de aire se calcula como la contribución de todas las partículas confinadas en ese volumen:

$$K_e = \int \beta(r) n(r) dr \quad (13)$$

siendo $\beta(r)$ la sección transversal de extinción para cada radio r de la partícula y $n(r)$ la densidad para cada intervalo de radio dr , donde:

$$\int n(r) dr = N_0 \tag{14}$$

En cuya expresión, N_0 es la densidad de partículas (cm^{-3}).

Cada partícula posee una diferente función de fase de dispersión. La función de fase promedio para el volumen de aerosol es:

$$P_a(\theta) = \frac{1}{N_0} \int P(r, \theta) n(r) dr \tag{15}$$

Espesor óptico

Las propiedades estudiadas para pequeños volúmenes de aire confinados pueden ser utilizadas para inferir las propiedades de capas más amplias de aerosol o para la atmósfera completa. La relación entre el coeficiente de extinción K_e y la transmisión $T(z)$ desde el borde exterior de la atmósfera ($z=\infty$) hasta una altura z es:

$$T(z) = e^{-\int_z^\infty K_e(z) dz} \tag{16}$$

y la densidad óptica puede ser explicado por el coeficiente de extinción:

$$\tau(z) = -\int_z^\infty K_e(z) dz \tag{17}$$

o lo que es lo mismo:

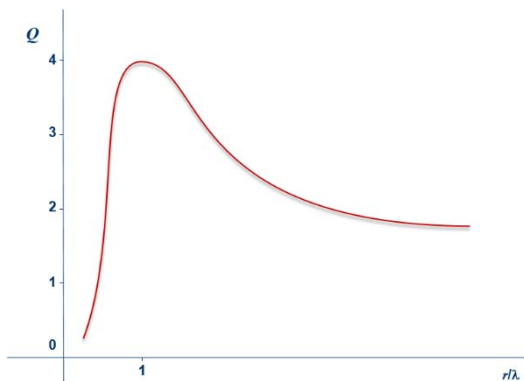
$$K_e(z) = \frac{d\tau}{dz} \tag{18}$$

siendo $z = 0$ en la superficie terrestre y $z = \infty$ en el borde exterior de la atmósfera.

Albedo de dispersión simple

El albedo de dispersión simple (*single-scattering albedo*) es la razón entre el coeficiente de dispersión del aerosol y el coeficiente de extinción total debido a la dispersión y a la absorción. Representa, pues, la fracción de energía incidente que vuelve a aparecer como dispersada y constituye una medida de la absorción por parte del aerosol. Un aerosol completamente absorbente tendrá un valor nulo de albedo de dispersión simple y, en el extremo opuesto, un aerosol perfectamente dispersante poseerá un albedo de dispersión simple igual a la unidad. El albedo de dispersión simple viene definido por la parte imaginaria del índice complejo de refracción de una partícula y por su radio.

Para aerosoles de composición homogénea, las partículas en suspensión, cuyo tamaño sea mucho menor o mucho mayor que la longitud de onda de la radiación, son más absorbentes que aquellas otras cuyo tamaño sea intermedio:



Factor de eficiencia de la extinción en función relativo del tamaño de las partículas.

Tamaño relativo de la partícula r/λ	Factor de eficiencia de extinción $Q(r/\lambda)$	Albedo de dispersión simple ω_0
10,000	$0,218 \cdot 10^1$	0,743
3,167	$0,255 \cdot 10^1$	0,895
1,000	$0,338 \cdot 10^1$	0,969
0,317	$0,129 \cdot 10^1$	0,977
0,100	$0,328 \cdot 10^{-1}$	0,834
0,032	$0,177 \cdot 10^{-2}$	0,158

Índice complejo de refracción $n=1,43 - 0,0035 i$.

Para los aerosoles cuya composición es heterogénea, el valor del albedo de dispersión simple depende de las características de la mezcla de componentes. Medidas efectuadas para alturas comprendidas entre 1 y 2 km sobre la superficie, en cuyo intervalo se encuentra la mayor concentración de aerosoles atmosféricos, han arrojado los resultados experimentales de ω_0 regidos siguientes.

Zona	ω_0
Industrial	0,50 - 0,65
Residencial	0,73 - 0,87
Virgen	0,89 - 0,95

Función de fase de la dispersión

La función de fase de la dispersión representa la distribución angular de la radiación dispersada en cada dirección del espacio. Esta función depende de las dimensiones de las partículas esféricas del aerosol. Cuando son de muy pequeño tamaño, la función de fase apenas varía con la dirección. Si la concentración de partículas en el aerosol aumenta, el valor de la función de fase de la dispersión en la dirección de la propagación (ángulo de dispersión $\theta < 30^\circ$) es tanto mayor cuanto mayor sea el espesor óptico, o lo que es lo mismo, menor sea la visibilidad; para direcciones progresivamente más alejadas de la de propagación ($30^\circ < \theta < 60^\circ$), el comportamiento de la función de fase es intermedio; por último, cuando $\theta > 60^\circ$, la función de fase de la dispersión es menor conforme aumenta el espesor óptico. La variación de la función de fase de la dispersión es mínima para ángulos de dispersión próximos a 40° .

Transferencia radiativa

Hasta este punto, todo lo dicho sobre la interacción entre radiación y la atmósfera y sus componentes se basa en estudios y modelos microscópicos, en los que se utilizan pequeños volúmenes de aire para describir los parámetros atmosféricos ya expuestos. La descripción macroscópica del fenómeno de radiación considerando el conjunto completo Tierra-atmósfera re-

quiere integrar las características radiativas de pequeños volúmenes extendiéndolas a la atmósfera completa. Esta integración es conocida como *ecuación de transferencia radiativa*, denominada usualmente *RT equation*.

Las propiedades ópticas de la atmósfera, suponiéndola horizontal y uniforme, y asumiendo también la

misma uniformidad para la superficie reflectante, vienen dadas por el coeficiente de extinción $K_e(z)$, por el albedo de dispersión simple $\omega_0(z)$ y por la función de fase de la dispersión $P(\theta, z)$.

Sustituyendo la coordenada z por una magnitud adimensional como el espesor óptico vertical τ , obtenido mediante la expresión (17), puede enunciarse la ecuación de transferencia radiativa, que describe la variación de la radiancia total $L'(\theta')$ durante su transmisión a través de una capa delgada de aire con un espesor óptico vertical τ obtenido por integración de la ecuación $d\tau = K_e(z) dz$:

$$dL' = -L' \frac{d\tau}{u} + (J + J_0) \frac{d\tau}{u} \quad (19)$$

donde $u = \cos\theta'$, siendo θ' el ángulo cenital de $L'(\theta')$. Al estar θ' comprendido entre 0° y 90° para las trayectorias descendentes de la radiación, resulta que $u > 0$. Por el contrario, $u < 0$ para las trayectorias ascendentes, por ser ahora $90^\circ \leq \theta' \leq 180^\circ$. Teniendo en cuenta el signo de z , la fracción $d\tau/u$ es siempre positiva. El primer término del segundo miembro de la expresión (19) representa la atenuación de la radiación a su paso a través de la capa considerada. La transmitancia a través de la capa en la dirección θ' es:

$$T = e^{-d\tau/u} \quad (20)$$

Utilizando los dos primeros términos de su desarrollo en serie, el valor de la transmitancia puede aproximarse a:

$$T \approx 1 - d\tau/u \quad (21)$$

En consecuencia, la variación de la radiancia $L'(\theta')$

como consecuencia de la atenuación es $L' d\tau/u$. El segundo término del segundo miembro de la expresión (19) describe el incremento de radiancia como consecuencia de la luz solar directa y de la difusa en la dirección de la radiación. De entre los dos sumandos de este término, J define la luz difusa atmosférica y la radiación reflejada desde la superficie, dispersadas en la dirección (θ', ϕ) :

$$J(u, \phi, \tau) = \iint P(u, u', \phi, \phi') \cdot L'(\tau, u', \phi') du' d\phi' \quad (22)$$

Y J_0 explica la dispersión de la luz solar directa:

$$J_0(u, \phi, \tau) = \frac{\omega_0}{4\pi} P(u, \mu_0, \phi, \phi_0) E_0 e^{-\frac{\tau}{\mu_0}} \quad (23)$$

donde $e^{-\tau/\mu_0}$ representa la atenuación de la radiación solar desde el borde exterior de la atmósfera hasta la capa considerada, μ_0 es el coseno del ángulo cenital solar θ_0 , P es la función de fase de dispersión desde la dirección del sol hasta la definida por (θ', ϕ) , y E_0 es la irradiancia solar extraterrestre. Por último el divisor 4π , que es el valor en estereorradianes del ángulo sólido que corresponde a una esfera completa, recupera el valor unitario de J_0 , ya que la integral de P extendida a una esfera adquiere ese valor.

Dividiendo los dos miembros de la ecuación (19) por $d\tau/u$ queda:

$$\frac{u}{d\tau} dL' = -L' + J + J_0 \quad (24)$$

apareciendo la radiancia L' en forma diferencial en el primer miembro e integrada en el segundo. Esta ecuación, además, no puede resolverse analíticamente debido a la presencia del término de radiación

difusa J . Una solución aproximada de la misma consiste en suponer (i) que la superficie subyacente es completamente absorbente, es decir, que no refleja radiación en absoluto y (ii) que no exista dispersión múltiple en el seno de la atmósfera. De esta manera, si la radiación solamente puede ser dispersada una vez, la luz atmosférica dispersada no podrá contribuir de nuevo sobre la propia atmósfera, por lo cual $J=0$. En estas condiciones, la ecuación (24) puede ser resuelta para la radiancia ascendente:

$$L'_{\uparrow} = \frac{E_0 \mu_0 \omega_0}{4\pi(\mu + \mu_0)} P(-\mu, \mu_0, \phi) \left[1 - e^{-\tau_0 \left(\frac{1}{\mu} + \frac{1}{\mu_0} \right)} \right] \quad (25)$$

y para la radiancia descendente:

$$L'_{\downarrow} = \frac{E_0 \mu_0 \omega_0}{4\pi(\mu - \mu_0)} P(\mu, \mu_0, \phi) \left[e^{-\frac{\tau_0}{\mu}} - e^{-\frac{\tau_0}{\mu_0}} \right] \quad (26)$$

en cuyas ecuaciones τ_0 es el espesor óptico atmosférico total y $\mu = |u|$.

La aproximación de una sola dispersión solamente es válida si se considera un espesor óptico pequeño, esto es, si $\tau_0 \left(\frac{1}{\mu} + \frac{1}{\mu_0} \right) \ll 1$. Asumiéndolo así, los exponentes de las expresiones (25) y (26) pueden sustituirse por los dos primeros términos de su desarrollo en serie, quedando finalmente:

$$\begin{aligned} L'_{\uparrow} &= \frac{E_0 \mu_0 \omega_0 \tau_0 P(-\mu, \mu_0, \phi)}{4\pi\mu\mu_0} \\ L'_{\downarrow} &= \frac{E_0 \mu_0 \omega_0 \tau_0 P(\mu, \mu_0, \phi)}{4\pi\mu\mu_0} \end{aligned} \quad (27)$$

expresiones que son válidas para atmósferas limpias. En ellas puede notarse que la distribución angular de la radiancia depende directamente de la función de fase de la dispersión. Por esta razón el Sol presenta una aureola clara observado desde la superficie terrestre y el cielo es más oscuro conforme más alejado esté del disco solar.

Efecto atmosférico

El efecto atmosférico que afecta a la radiancia ascendente en una atmósfera despejada de nubes puede ser abordado como una solución de la ecuación de transferencia radiativa. El planteamiento más general de la ecuación de transferencia radiativa exige que la superficie observada sea heterogénea y no lambertiana. Bajo estas condiciones, la solución es compleja y ha de ser acometida por métodos aproximados como el de una sola dispersión, o mediante el análisis

de Fourier de la radiancia para varias frecuencias espaciales. Sin embargo, en la mayoría de los casos, para la estimación y corrección del efecto atmosférico es suficiente considerar la superficie como uniforme, o bien como heterogénea, pero lambertiana. Existen modelos alternativos que contemplan el binomio atmósfera-cubierta, en los que cada uno de los elementos intervinientes se considera como un conjunto de capas con características radiativas diferentes.

La radiancia ascendente captada por un sensor con un campo de visión instantáneo suficientemente estrecho se puede considerar como la contribución de varios sumandos:

$$L' = L'_0 + L'_s + L'_{d1} + L'_{d2} \quad (28)$$

donde:

L'_0 es la radiancia directa producida por la incidencia de la luz solar en el campo de visión instantáneo del instrumento de adquisición, previa dispersión en el seno de la atmósfera y sin haber sido reflejada por la superficie. Este componente es independiente de la naturaleza de la superficie y produce una disminución del contraste de la escena debido a la suma homogénea de luminancia para toda la imagen.

L'_s es la señal atenuada producida por la radiancia de la luz que atraviesa directamente la atmósfera en sentido descendente, que se refleja sobre la superficie y que es transmitida directamente hacia el sensor de nuevo a través de la atmósfera y en sentido ascendente. Depende solamente de la reflectancia de la superficie que se encuentra bajo el campo de visión instantáneo del sensor. Es la única información útil para la caracterización de dicha superficie. El resto de los componentes pueden considerarse como *contaminantes* de éste, si lo que se persigue son exclusivamente medidas de la superficie terrestre. L'_{d1} es la radiancia de la luz reflejada por la superficie en la dirección del sensor como consecuencia de su iluminación a partir de la luz difusa atmosférica. Sobre este componente tiene influencia el eventual carácter no lambertiano de la superficie. L'_{d2} es la

radiancia de la luz solar reflejada por la superficie y dispersada al menos una vez en el seno de la atmósfera antes de ser captada por el sensor. Para sensores de alta resolución, este componente depende de la reflectancia de la superficie fuera del campo de visión instantáneo, por lo que estará afectado por la heterogeneidad de la superficie. Como también depende de la dirección de la reflexión de la luz incidente, está afectado pues de las características no lambertianas de la superficie. Este término L'_{d2} incluye la luz reflejada más de una vez por la superficie.

Superficies lambertianas uniformes

Cuando la observación se realiza sobre una superficie lambertiana, la ecuación de transferencia radiativa puede ser considerada como una función de la reflectancia de la superficie:

$$L'(\rho, \mu, \phi) = L'_0(\rho, \phi) + \frac{\rho E'_d(\theta_0) T(\mu)}{\pi(1 - s\rho)} \quad (29)$$

donde ρ es la reflectancia hemisférica de la superficie en cuestión; L' es la radiancia ascendente en el borde de la atmósfera procedente de una superficie de reflectancia ρ ; L'_0 es la radiancia para $\rho=0$; E'_d es la irradiancia total de la luz incidente sobre la superficie, considerando $\rho=0$; T es la transmitancia de la atmósfera desde la superficie hasta su límite exterior; y s es la reflectancia de la atmósfera para la luz isotrópica procedente de la superficie. El primer término del segundo miembro explica la radiancia de la atmósfera si la superficie fuese no reflectora. El segundo describe transmisión directa y difusa de la luz reflejada por la superficie hacia el sensor.

Los valores absolutos de radiancia L' y L'_0 pueden ser normalizados a valores de reflectancia refiriéndolos a la irradiancia solar incidente sobre la atmósfera, obteniéndose como resultado L y L_0 :

$$L = \frac{\pi L'}{E_0 \cos \theta_0}; \quad L_0 = \frac{\pi L'_0}{E_0 \cos \theta_0} \quad (30)$$

donde el divisor $E_0 \cos \theta_0$ es la irradiancia espectral solar sobre la unidad de superficie horizontal colocada en el borde de la atmósfera.

La irradiancia normalizada E'_d se calcula por el mismo procedimiento a partir de la irradiancia total:

$$E'_d = \frac{E'_d}{E_0 \cos \theta_0} \quad (31)$$

Con ello, la ecuación (29) se puede poner en función de los valores normalizados de sus variables:

$$L(\rho, \mu, \phi) = L_0(\rho, \phi) + \frac{\rho E_d(\theta_0) T(\mu)}{1 - s\rho} \quad (32)$$

Suponiendo que la atmósfera tuviese un espesor óptico $\tau = 0$, entonces $L_0 = 0$, $E_d = 1$, $T = 1$, y en consecuencia, la radiancia normalizada sería igual a la reflectancia de la superficie:

$$L(\rho, \mu, \phi) = \rho$$

Utilizando unidades de reflectancia, la diferencia entre la radiancia y la reflectancia muestra el efecto atmosférico neto:

$$L - \rho = L_0 - \rho \left[1 - \frac{E_d T}{1 - s\rho} \right] \quad (33)$$

siendo en este caso L la reflectancia aparente de la superficie observada desde el espacio.

Superficies no lambertianas

Las superficies perfectamente lambertianas son ideales. Las naturales distan de la lambertianidad en mayor o menor medida. Por simplificación de los modelos de transferencia, en determinados casos se pueden considerar lambertianas superficies que no lleguen a serlo completamente. Tanto el agua como la vegetación son superficies cuya capacidad reflectante tiene características no lambertianas derivadas de la reflexión especular, en el caso de la primera y de la retrodispersión, de las sombras internas y del efecto de fondo del suelo subyacente, en el caso de la vegetación. Diversas medidas experimentales han demostrado que para superficies de pequeña reflectancia, como por ejemplo las praderas en la zona roja del espectro visible, observadas a través de atmósferas de gran espesor óptico, las propiedades de la superficie quedan ocultas por el efecto atmosférico. La radiancia ascendente en este caso es independiente de la reflectancia de la superficie y muestra una fuerte dependencia del ángulo cenital de observación. Por el contrario, para el infrarrojo próximo, en el cual la reflectancia de este tipo de superficies es alta, la distribución angular de la radiancia ascendente muestra estrechamente las características de la superficie salvo para ángulos cenitales de visión próximos a 180° , es decir, con el sol de espaldas al observador.

Superficies lambertianas no uniformes

El término de radiancia difusa L'_{d1} de la ecuación (28) proporciona las diferencias existentes entre la observación de una superficie uniforme y otra no uniforme. La radiancia no sólo está compuesta por la radiación reflejada por el área correspondiente al

campo de visión instantáneo, sino también por la radiación dispersada en la atmósfera procedente de zonas de la superficie próximas a una que se considere uniforme. Este fenómeno, denominado *efecto de adyacencia* introduce información de áreas no observadas, contaminando en consecuencia la señal detectada. La radiancia puede aumentar o disminuir en función de la reflectancia de las zonas aledañas. Por otro lado, el efecto de adyacencia depende del tamaño de la zona observada que se considere uniforme. Es significativo que el efecto fondo circundante sobre un píxel determinado en una imagen Landsat de 30 m de resolución espacial puede extenderse hasta 1 km. En general puede considerarse que el efecto de adyacencia es apreciable para zonas de hasta 2 km de lado y ligero para zonas de hasta 10 km.

Si se considera un infinitésimo de superficie de reflectancia ρ_f en el seno de un fondo superficial de reflectancia ρ_b , la radiancia correspondiente al infinitésimo puede extraerse de la expresión (32):

$$L_{in} = L_0 + \frac{E_d (\rho_f T_s + \rho_b T_d)}{1 - s\rho_b} \quad (34)$$

donde los términos T_s y T_d son respectivamente la transmitancia directa de la luz reflejada desde la superficie infinitesimal y la difusa procedente de las zonas circundantes.

Para desarrollar una expresión que relacione la radiancia con la reflectancia de una superficie no uniforme $\rho(x,y)$, se parte de una superficie negra con un sólo punto reflectante situado en la posición (x_0, y_0) , que puede ser considerado como un impulso unitario o función delta de Dirac:

$$\rho(x, y) = \delta(x - x_0, y - y_0) \quad (35)$$

En ausencia de efecto atmosférico, la radiancia ascendente en cualquier dirección puede ser considerada como una función delta. La dispersión atmosférica redistribuye la radiancia según una función de dispersión puntual atmosférica:

$$L(x, y) = m(x - x_0, y - y_0) + L_0 \quad (36)$$

Para superficies de reflectancia variable, la radiancia ascendente puede ser considerada como la convolución entre la reflectancia superficial $\rho(x,y)$ y la función de dispersión puntual atmosférica:

$$L(x, y) = \mathcal{F}^{-1} \{ \mathcal{F}[\rho(x, y)] M(k_x, k_y) \} + L_0 + L_\beta \quad (37)$$

donde L_0 es la radiancia ascendente correspondiente a una superficie de reflectancia cero; M es la *Función de Modulación de la Transferencia Atmosférica* (MTF), esto es, la transformada de Fourier de la función de dispersión puntual atmosférica; k_x y k_y son las frecuencias espaciales en las direcciones respectivas x e y ; y L_β es un término aditivo que representa la radiancia de la luz reflejada por una superficie y dispersada en la atmósfera más de una vez. La función MTF bidimensional, $M(k_x, k_y)$ puede ser enunciada de modo unidimensional, $M(k)$, mediante el paso de coordenadas cartesianas a polares, considerando las distancias nadirales, mediante el cambio:

$$k = \sqrt{k_x^2 + k_y^2} \quad (38)$$

de tal modo que:

$$M(k_x, k_y) = M(k) \quad (39)$$

La MTF atmosférica es la modulación dispersiva que introduce la atmósfera sobre las frecuencias espaciales propias de la superficie observada. Una frecuencia cero indica una superficie uniforme infinitamente grande. La MTF de una zona de estas características representa la forma de transmisión del contraste entre dos grandes áreas limítrofes a través de la atmósfera:

$$M(0) = \frac{L(\rho) - L(\rho=0)}{\rho} \quad (40)$$

La MTF normalizada se define como:

$$M_n(k_x, k_y) = \frac{M(k_x, k_y)}{M(0)} \quad (41)$$

y representa el efecto neto de la atmósfera sobre una superficie de reflectancia no uniforme. Una expresión empírica del valor de la MTF es la siguiente:

$$M_n(k) = 1 - 0,5\tau_r \left[1 - e^{-2,5kH_r} \right] - 0,7\lambda^{-0,2}\tau_a \left[1 - e^{-1,3kH_a} \right] \quad (42)$$

con λ expresada en μm , y donde τ_r y τ_a son los espesores ópticos molecular y aerosólico, respectivamente, H_r y H_a son las escalas de alturas respectivas adoptadas para las moléculas (~ 8 km) y para los aerosoles (~ 1 km).

Nubes

La presencia de nubes en la escena produce alteraciones en las medidas de la reflectancia de una superficie y en las magnitudes derivadas de ellas, como es el caso de los índices de vegetación y de la clasificación digital de las cubiertas.

Las nubes de gran tamaño en relación a la celda de la imagen pueden ser fácilmente aisladas por su alta reflectancia en el visible y en el infrarrojo próximo. Además, dado que suelen estar a temperaturas considerablemente más bajas que las de los elementos contiguos de la imagen, son detectables sin dificultad en el infrarrojo térmico. Sin embargo, cuando se trata de nubes de tamaño inferior al de la celda, su detección es difícil. Las propiedades ópticas de una capa nubosa pueden ser descritas por la distribución de

tamaños de las nubes. La reflexión y transmisión características resultan ser función del tamaño de la nube y de la fracción nubosa, definida ésta como la proporción de superficie cubierta por la proyección ortogonal de las nubes sobre el terreno.

Se ha encontrado una relación empírica entre la transmitancia de un cielo nublado T_{cn} y la fracción nubosa χ_n medida desde la superficie:

$$T_{cn} = 1 - (1 - C_i)\chi_n^2 \quad (43)$$

donde C_i es la transmitancia para un cielo cubierto por un determinado tipo de nube i , para un ángulo solar cenital θ_0 , y cuya expresión es:

$$C_i = \frac{A_i}{94,4} e^{-(B_i - 0,059)/\cos\theta_0} \tag{44}$$

obteniéndose los valores A_i y B_i de la siguiente tabla:

Tipos de nubes	A_i	B_i	C_i ($\theta_0 = 45^\circ$)
Cirros	82,2	0,079	0,85
Cirro-estratos	87,1	0,148	0,85
Alto-cúmulos	52,5	0,112	0,51
Alto-estratos	39,0	0,063	0,41
Estrato-cúmulos	34,7	0,104	0,35
Estratos	23,8	0,159	0,22
Nimboestratos	11,2	-0,167	0,16
Brumas	15,4	0,028	0,17

Fuente: Haurwitz (1948).

El promedio de la transmitancia de las propias nubes puede deducirse a partir de la expresión (43), ya que T_{cn} es la media ponderada de la transmitancia de las zonas cubiertas (T_n) y despejadas ($T=1$):

$$T_{cn} = T_n \chi_n + (1 - \chi_n) \tag{45}$$

por lo que:

$$T_n = \frac{T_{cn} - 1 + \chi_n}{\chi_n} = \frac{1 - (1 - C_i) \chi_n^2 - 1 + \chi_n}{\chi_n} \tag{46}$$

es decir:

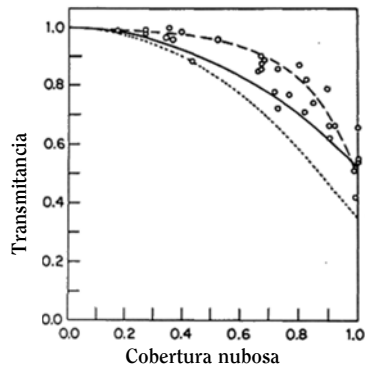
$$T_n = 1 - (1 - C_i) \chi_n$$

Nubes de tamaño inferior al de la celda

La presencia de nubes en la escena condiciona la calidad de las imágenes adquiridas por los sensores en órbita. Para caracterizar mejor el efecto producido

deben distinguirse cuatro clases de nubes:

- Nubes de tamaño superior al campo de visión del sensor y de gran espesor óptico. Su presencia se detecta fácilmente en el visible, ya que la dispersión que producen es del tipo no selectiva y, por tanto, independiente de la longitud de onda de la observación.
- Nubes altas y poco espesas, del tipo cirros, que no son visibles en los canales ópticos, pero que sí son detectables en el infrarrojo debido a sus bajas temperaturas.
- Nubes bajas y estratificadas cuya reflectancia no es lo suficientemente alta como para ser detectada en el visible y además con una temperatura muy parecida a la de la superficie. Este tipo de nubes pueden ser consideradas como una capa de aerosol, únicamente que con partículas de mayor tamaño que las típicas de los aerosoles atmosféricos, por lo que la corrección del efecto que ocasionan puede acometerse como si de cualquier otro aerosol se tratase.
- Nubosidad muy dispersa con gran cantidad de nubes mucho más pequeñas que el campo de visión del instrumento de adquisición, cuyo efecto sobre la radiancia final detectada por el sensor difiere del atribuido a los tipos anteriores, y por lo cual merece un estudio más detallado.



Modelos empíricos de la transmitancia solar a través de una capa rota de estratocúmulos en función de la fracción nubosa.

5

Respuesta espectral de la superficie

Introducción

En los capítulos anteriores se ha estudiado la interacción de la radiación con la materia en general desde el punto de vista macroscópico y la repercusión que la atmósfera terrestre tiene sobre el proceso de transferencia radiativa. Salvo que el objeto de estudio sea la propia atmósfera, ésta se considera un obstáculo para la observación de la superficie, y de ahí que haya que arbitrar procedimientos de corrección para minimizar su efecto. Queda por abordar lo que sucede cuando la radiación intercepta la superficie observada.

Desde este punto de vista, en la Tierra se encuentran tres tipos de superficies de marcadas características comunes, aun dentro de la diversidad que cada una de ellas presenta. De un lado abunda el agua en sus estados líquido y sólido. El agua en forma de vapor está incluida dentro de la atmósfera y, si bien su presencia es decisiva en el mantenimiento de las condiciones físicas del planeta, su cantidad es despreciable en comparación con las otras dos formas, presentes en los océanos, mares y aguas interiores y en los casquetes polares, glaciares y cubiertas de nieves perpetuas o estacionales de las cotas altas.

Otro conjunto de superficies, bastante heterogéneas, lo constituyen el material inerte de los suelos desnudos, constitutivos del sustrato de los continentes. Sin embargo, este estado es en la práctica inexistente por la colonización inevitable que la vida en todas sus formas ejerce sobre esas superficies. Con todo, con-

viene estudiar la respuesta de los materiales litológicos y edáficos ante la radiación para comprender su interacción con la superficie terrestre.

El tercer gran grupo de superficies es el constituido por la vegetación, tanto natural como cultivada. En determinadas áreas, la vegetación predomina de tal forma que es la que determina la respuesta espectral. En otros lugares, la vegetación no es tan tupida y deja entrever el suelo subyacente en todas las proporciones posibles, obteniéndose como resultado signatures mixtas con mayor o menor incidencia de un tipo de superficie u otro.

En definitiva, sobre la superficie terrestre cabe encontrar agua, suelo, vegetación y las mezclas de los tres componentes, y de su naturaleza y proporciones dependerá la señal recibida a bordo de los sensores de los satélites de observación de la Tierra, descontado el efecto atmosférico.

Para el estudio del agua líquida y las superficies heladas o nevadas se dispone de metodologías propias que serán examinadas en los capítulos de aplicaciones oceanográficas y en la criosfera. En los apartados siguientes se abordan los rasgos más importantes de las superficies de suelo desnudo, de la vegetación y de los sistemas suelo-vegetación que habitualmente se presentan en diferentes proporciones de cobertura.

El suelo

Los suelos son sistemas heterogéneos polifásicos y dispersos derivados de la meteorización de rocas parentales. Están constituidos por una fase sólida, con partículas de muy diferentes tamaños, formas y naturaleza, una fase líquida tiende a llenar los huecos dejados por las anteriores y una fase gaseosa compuesta por el aire que ocupa los poros libres de agua.

Composición

La composición química de los suelos es muy variable y solamente pueden hacerse algunas generalizaciones. Los minerales constituyen alrededor del 95 % de la masa total y entre ellos destacan los óxidos de silicio (SiO_2), los óxidos de aluminio (Al_2O_3) y los óxidos de hierro (Fe_2O_3). Pueden presentarse libres o como como minerales de la arcilla en forma de silicatos aluminico-potásicos de magnesio e hierro. La materia orgánica está presente en los horizontes superiores del suelo, llegando a constituir hasta el 20 % del peso en ocasiones.

Los minerales del suelo son materiales inorgánicos cristalinos y no cristalinos procedentes de disgregación de la roca madre. El tamaño de las partículas sólidas del suelo depende de la composición química de sus componentes, y a su vez la condiciona. Tradicionalmente se ha dividido el suelo en varias fracciones de tamaños diferentes, cuya distribución según el Departamento de Agricultura de los Estados Unidos (USDA) es la recogida en la Tabla 1.

Los minerales primarios son los que proceden directamente de la disgregación de las rocas de las que procede el suelo y están presentes en las fracciones

Tabla 1. Tamaño de las partículas de las distintas fracciones del suelo (USDA).

Fracción	Tamaño de las partículas (mm)	
Arcilla	< 0,002	
Limo	0,002	- 0,05
Arena muy fina	0,05	- 0,10
Arena fina	0,10	- 0,25
Arena media	0,25	- 0,50
Arena gruesa	0,50	- 1,00
Arena muy gruesa	1,00	- 2,00
Grava	2,00	- 500
Piedras	> 500	

arena y limo, esto es, en las más gruesas. De entre ellos, el cuarzo (SiO_2) es el más abundante, sobre todo en las regiones húmedas, aunque en las rocas ígneas su presencia es inferior al 12 %. La elevada abundancia se debe a su gran resistencia a la meteorización química. Los minerales primarios más frecuentes en los suelos se recogen en la Tabla 2.

Muchos de los minerales primarios constituyen en realidad grupos de minerales de composición muy parecida, que se diferencian entre sí solamente por sustituciones isomorfas que son posibles cuando los radios catiónicos tienen magnitudes similares, aunque no lo sean sus cargas. Así, por ejemplo, el grupo de los piroxenos está integrado por minerales con distintas proporciones de calcio, magnesio e hierro, dando así lugar a la enstatita, broncita o hiperstena (piroxenos rómbicos) y al diópsido, dialaga, augita o egrina (piroxenos monoclinicos). Además, durante el proceso de cristalización, es frecuente la contaminación del cristal por otros elementos que influyen en el color resultante. Las características reflectoras del

Tabla 2. Composición de los minerales primarios más abundantes en suelos.

Nombre	Composición	Sistema de cristalización
Cuarzo	SiO ₂	Trigonal
Feldespatos		
Ortoclasas	KAl Si ₃ O ₈	Monoclínico
Plagioclasas		
Albita	NaAlSi ₃ O ₈	Triclínico
Anortita	CaAl ₂ Si ₂ O ₈	Triclínico
Silicatos laminados o filosilicatos		
Micas		
Moscovita	KAl ₃ Si ₃ O ₁₀ (OH) ₂	Monoclínico
Biotita	KAl(Mg, Fe) ₃ Si ₃ O ₁₀ (OH) ₂	Monoclínico
Clorita	(Mg, Fe, Al) ₆ (Si, Al) ₄ O ₁₀ (OH) ₈	Monoclínico
Silicatos férrico-magnésicos		
Anfíboles	(Na, Ca, Mg, Fe) ₂₋₃ (Mg, Fe) ₅₋₇ (Al _{2-x} Si _{6+x})O ₂₂ (OH) ₂	Monoclínico
Piroxenos	(Ca, Mg, Fe) ₂ Si ₂ O ₆	Monoclínico o rómbico
Olivino	(Mg, Fe) ₂ SiO ₄	Rómbico

suelo dependen en gran medida de la contaminación que hayan sufrido los cristales, aparte de la gran diversidad que de por sí muestran en su propia composición debido a las mencionadas sustituciones isomórficas.

Casi todos los minerales primarios que ordinariamente forman parte de las rocas parentales se presentan en las fracciones arena y limo del suelo derivado de ellas. De entre ellos, el cuarzo (SiO₂) es el más abundante, si bien en las rocas ígneas –de las que procede preferentemente, aunque también de las metamórficas– no representa más del 12 %. La causa hay que buscarla en su elevada resistencia a la meteorización química. Efectivamente, en las regiones húmedas, el cuarzo se acumula a expensas de los piroxenos menos resistentes, anfíboles, micas y feldespatos, que son meteorizados con mayor facilidad y se transforman en minerales secundarios. En cambio, en

la fracción arena de los suelos de las regiones semi-húmedas y áridas es mayor el contenido de minerales distintos del cuarzo, con promedios del 20 y 37 % respectivamente. Todo ello viene a indicar que el agua juega un papel muy importante en la transformación química de los minerales primarios.

La acción química de la atmósfera, en unión de otros procesos químicos o de recristalización, transforma los minerales primarios en los denominados *secundarios*. Este tipo de minerales es el más frecuente en la fracción arcilla del suelo. Los más abundantes son los llamados minerales de la arcilla y ciertos óxidos e hidróxidos. El proceso de degradación también produce carbonatos, sulfatos y fosfatos. En los suelos de ciertas zonas húmedas y frías o semihúmedas y templadas, la ilita es el mineral predominante. En regiones húmedas y templadas se presentan mezclas de caolinita, vermiculita y otros minerales interestratificados.

En regiones cálidas y húmedas predomina la caolinita con mezclas en menor proporción de gibsita y hematites. El clima tropical beneficia especialmente la presencia de sesquióxidos, llegándose en algunos casos extremos a encontrarse sólo este tipo de minerales. Las condiciones básicas y áridas favorecen la formación de montmorillonita, también presente en los suelos de desarrollo retardado como los que presentan un drenaje imperfecto. Una de las principales características de los minerales de la arcilla es su estructura laminar en dos o tres capas formadas por tetraedros silícicos y octaedros aluminico-magnésicos, de la cual derivan sus principales propiedades edáficas, es decir, la moldeabilidad y las capacidades de intercambio catiónico y de retención de agua. Los minerales secundarios más abundantes en los suelos son los recogidos en la Tabla 3.

La materia orgánica procede de la descomposición química de las excreciones de animales y microorganismos, de residuos de plantas o de su degradación tras su muerte. Puede distinguirse entre compuestos no húmicos y humus. En los primeros persiste todavía la composición química e incluso la estructura física de los tejidos animales o vegetales originales, puesto que los organismos del suelo descomponen las sustancias orgánicas dejando solamente residuos difícilmente atacables, como algunos aceites, grasas, ceras y ligninas procedentes de las plantas superiores de origen. El humus es el producto de la transformación de la materia orgánica metabolizable. Consiste en una mezcla compleja de sustancias coloidales y amorfas de color negro o marrón oscuro. En los llamados *suelos minerales*, es decir los que tienen un contenido de materia orgánica inferior al 20 %, el humus constituye entre el 65 y el 75 % de la materia orgánica total. Este tipo de suelos ocupa el 95 % de la superficie terrestre mundial. Los suelos con un mayor contenido en materia orgánica se denominan *sue-*

los orgánicos. Como referencia útil puede considerarse que el contenido medio aproximado de materia orgánica en los suelos de cultivo ordinarios oscila entre 1 y 6 %. Los suelos desarrollados en condiciones de semiaridez suelen presentar altos contenido en materia orgánica, y ofrecen una pigmentación muy oscura. En regiones templadas y húmedas la pigmentación es menos acusada y es escasa en los suelos de las regiones tropicales y subtropicales. La consecuencia radiométrica de mayor interés derivada del contenido en materia orgánica es la pérdida de reflectancia del suelo en el espectro visible, produciendo un oscurecimiento típico.

Los componentes sólidos del suelo constituyen una matriz sólida, frecuentemente no rígida, cuyos intersticios irregulares, de tamaños diversos, están más o

Tabla3. Composición de los minerales secundarios más abundantes en los suelos.

Nombre	Composición	Sistema de cristalización
Minerales de la arcilla		
Caolín		
Caolinita	$\text{Al}_2 \text{Si}_2\text{O}_5(\text{OH})_4$	Monoclínico o triclínico
Esmectita		
Montmorillonita	$\text{Na}_{0,33}(\text{Al}_{1,67}\text{Mg}_{0,33})\text{Si}_4\text{O}_{10}(\text{OH})_2$	Monoclínico
Illita	$\text{Al}_2(\text{Si}_{3,85}\text{Al}_{0,15})\text{O}_{10}(\text{OH})_2$	Monoclínico
Vermiculita	$\text{Mg}_3(\text{Al},\text{Si})_4 \text{O}_{10}(\text{OH})_2 \cdot 4\text{H}_2\text{O}$	Monoclínico
Clorita	$(\text{Mg}, \text{Fe}, \text{Al})_6(\text{Si}, \text{Al})_4\text{O}_{10}(\text{OH})_8$	Monoclínico
Óxidos e hidróxidos		
Gibsita	$\text{Al}(\text{OH})_3$	Monoclínico
Goethita	$\text{FeO}(\text{OH})$	Rómbico
Hematites	Fe_2O_3	Trigonal
Carbonatos		
Calcita	CaCO_3	Trigonal
Dolomita	$\text{CaMg}(\text{CO}_3)_2$	Trigonal
Sulfatos		
Yeso	$\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$	Monoclínico
Fosfatos		
Apatito	$\text{Ca}_5(\text{PO}_4)_3(\text{Cl}, \text{F}, \text{OH})$	Hexagonal

menos conectados entre sí y en general son ocupados por la fase líquida o *solución del suelo*, integrada por agua con sólidos y gases disueltos. Como el vapor de agua en la fase gaseosa es despreciable comparativamente, suele hablarse de *agua del suelo* como sinónimo de *solución del suelo*. La fase gaseosa o *aire del suelo* es la mezcla de gases, incluido el vapor de agua que ocupa los huecos no saturados por agua. La humedad relativa de la fase gaseosa del suelo es prácticamente del 100 % de saturación. El aire del suelo difiere en su composición de la atmósfera, debido al alto contenido en CO₂ y la baja proporción de oxígeno debida a la actividad microbiana. La fluctuación natural del contenido relativo de agua y aire en el suelo causa variaciones temporales de la reflectancia del suelo.

El agua es retenida en los poros del suelo en virtud de fenómenos de capilaridad. Debido a su polaridad, las moléculas de agua en suelos subsaturados, son atraídas por las moléculas superficiales de la fase sólida mediante fuerzas de adhesión. Por su parte, las moléculas de agua sufren atracciones mutuas debidas a las fuerzas de cohesión. El resultado es un equilibrio de fuerzas de adsorción producidas en las interfases sólido-agua y de capilaridad en los meniscos curvos de las interfases agua-aire que posibilitan la retención del agua, en antagonismo con la fuerza de la gravedad que origina la percolación. La fluctuación natural del contenido relativo de agua y aire en el suelo causa variaciones temporales de la reflectancia del suelo.

Textura

Además de la composición química, hay otros factores que afecta a la reflectancia del suelo. Uno de los principales es el tamaño de las partículas y su modo de organización. La textura del suelo es la distribución de tamaños de las partículas constituyentes y

hace referencia a las fracciones enunciadas arriba. Las clases texturales se determinan mediante diagramas triangulares de texturas, en el que los lados representan las transiciones arena-arcilla, arcilla-limo y limo-arena, y la superficie central cualquiera de las posibles proporciones posibles entre las tres fracciones. La textura es esencial para comprender las características de un suelo, ya que influye en la composición mineral del suelo, la fertilidad, su densidad aparente, la cultivabilidad, la conductividad hidráulica y térmica, y en las propiedades reflectoras del suelo. De hecho, la reflectancia del suelo no está sólo influenciada por su composición química, sino que depende del tamaño de las partículas y de su modo de organización. Para la determinación de la textura se dejan al margen los elementos gruesos o superiores a 2 mm y se consideran tres grupos o fracciones: arena, limo y arcilla, cuyos tamaños en los sistemas USDA y la *International Society of Soil Science* (ISSS) son los recogidos en la Tabla 4:

Tabla 4. Grupos texturales

Fracción	Tamaño de las partículas (mm)			
	USDA		ISSS	
Arena	0,05	- 2,00	0,02	- 2,00
Limo	0,002	- 0,05	0,002	- 0,02
Arcilla	< 0,002		< 0,002	

Estructura

La estructura del suelo es una propiedad que se refiere a la forma que tienen de agruparse entre sí las partículas primarias de las fracciones arena, limo y arcilla en partículas compuestas, llamadas agregados o *pedios*, que constituyen unidades estructurales y que están separadas de los agregados adyacentes por superficies de discontinuidad. Las propiedades del

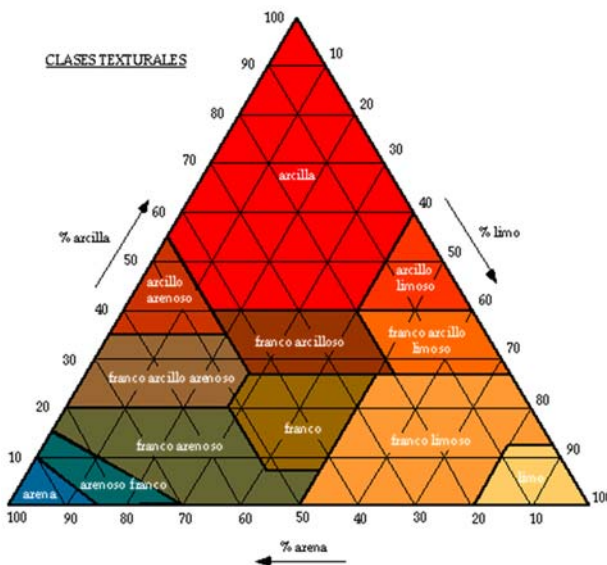


Diagrama triangular de texturas.

conjunto difieren de las que presenta una masa de similar composición y textura, pero de partículas primarias no agregadas. Los extremos en la gama de estructuras de un suelo son la estructura masiva, que es aquella en la que la agregación es tan acusada que no existen discontinuidades, y la estructura de grano simple, en la que las partículas aparecen completamente desunidas. Lo más frecuente es encontrar una agregación parcial o una cierta estructura que determina la geometría de la matriz sólida del suelo y del espacio de poros que la acompaña. Así como la textura es una propiedad prácticamente permanente, la estructura del suelo, por el contrario, puede variar considerablemente a corto plazo, en respuesta a los cambios de las condiciones naturales o de cultivo.

La estructura se define de acuerdo a tres características. El grado expresa la intensidad y estabilidad de

la agregación, estableciendo la diferencia entre la cohesión dentro de los agregados y la adhesividad entre agregados. En base al grado, el suelo se define como *sin estructura*, o con estructura *débil*, *moderada* o *fuerte*. La clase describe el tamaño de los pedios. Finalmente, el tipo señala la forma y ordenación de los agregados. En función de las tres características, la estructura puede ser clasificada según el criterio del Servicio de Conservación de Suelos de los EEUU (USDA) y adoptada por la FAO para la descripción de la estructura del suelo, recogida en la Tabla 5.

La estructura deriva en gran parte de la composición mineral del suelo, ya que en la agregación de las partículas individuales juega un papel esencial la presencia de una matriz arcillo-húmica y de otros minerales secundarios que establecen enlaces electrostáticos

Tabla 5. Clasificación de la estructura de los suelos.

CLASE (tamaño de los agregados)	TIPO (forma y ordenación de los agregados)										
	En placas Ejes horizontales más largos que el vertical	En bloques-Poliédricos-Esféroidales								Poliédricos-Esféroidales Superficies planas o curvas no acomodadas a las caras de los pedios circundantes	
		Prismáticos Ejes horizontales más cortos que el vertical. Dispuestos alrededor de línea vertical. Vértices angulares				En bloques-Poliédricos Superficies planas o curvas acomodadas a las caras de los pedios circundantes					
		En placas	Prismáticos		Columnares		En bloques		En bloques subangulares		Granulares
1. Muy fino	<1 mm	<10 mm	<10 mm	<10 mm	<5 mm	<5 mm	<5 mm	<5 mm	<5 mm	<1 mm	<1 mm
2. Fino	1-2 mm	10-20 mm	10-20 mm	10-20 mm	5-10 mm	5-10 mm	5-10 mm	5-10 mm	5-10 mm	1-2 mm	1-2 mm
3. Medio	2-5 mm	20-50 mm	20-50 mm	20-50 mm	15-20 mm	15-20 mm	15-20 mm	15-20 mm	15-20 mm	2-5 mm	2-5 mm
4. Grueso	5-10 mm	50-100 mm	50-100 mm	50-100 mm	20-50 mm	20-50 mm	20-50 mm	20-50 mm	20-50 mm	5-10 mm	-
5. Muy grueso	>10 mm	>100 mm	>100 mm	>100 mm	>50 mm	>50 mm	>50 mm	>50 mm	>50 mm	>10 mm	-

GRADO (estabilidad de los agregados)	
0. Sin estructura	Sin agregación ni arreglo ordenado.
1. Débil	Agregados mal formados, indistintos, no durables, que se fragmentan en una mezcla de pocos pedios, muchos pedios rotos y mucho material disgregado.
2. Moderada	Agregados bien formados, medianamente durables, indistintos en el suelo no perturbado, que se rompen en muchos pedios enteros, algunos rotos y poco material disgregado.
3. Fuerte	Agregados claros, bien formados, durables, débilmente unidos unos a otros, que se rompen casi completamente en pedios enteros.

débiles entre ellas. La estructura interviene, junto a la textura, en la determinación de las propiedades del suelo que controlan el movimiento del agua, la transferencia de calor, la porosidad, la aireación, la densidad aparente y la reflectancia.

Color

Probablemente, el color es la característica más evidente cuando se observa la superficie o el perfil de un suelo, y constituye su respuesta a la radiación electromagnética incidente en la región visible del espectro. Es, además, una característica muy utilizada en edafología para obtener información sobre la génesis del suelo e, indirectamente, sobre sus propiedades físicas y químicas.

La descripción del color del suelo se realiza sobre el sistema de especificación Munsell, que utiliza las componentes cromáticas intensidad, tono o matiz y saturación, que en su expresión anglosajona (*intensity, hue, saturation*) proporcionan las iniciales para el acrónimo IHS con la que se conoce también este sistema. El tono o matiz corresponde al *color* primario predominante, representado por la escala nominal rojo-amarillo-verde-azul-púrpura (R-Y-G-B-P) y las transiciones de uno a otro. La designación de los tonos intermedios se hace con las dos letras respectivas, precedidas de un número que indica las proporciones de la mezcla. La intensidad se refiere al valor o claridad del suelo en una escala acromática comprendida entre 0 y 10, de la cual se le asigna el valor que identifica al gris que exhibiría visualmente la misma claridad que el suelo en cuestión. La saturación, finalmente, es el grado de pureza del color. Puede oscilar entre el 0 (gris equivalente de la misma intensidad) y diferentes niveles máximos (color puro) según el color de que se trate, correspondiendo los niveles intermedios a manifestaciones *sucias* del color. La intensidad y saturación se indican mediante

dos números separados por una barra tras las letras del tono. Así, por ejemplo, la notación 2,5 YR 4/3 representa un color intermedio al amarillo y el rojo con una intensidad 4 y una saturación 3.

El color del suelo viene dado por la existencia y la proporción de compuestos orgánicos y minerales. La materia orgánica produce colores oscuros, generalmente negruzcos o pardos, como consecuencia de la presencia de ácidos húmicos. La acumulación de materia orgánica tras la muerte y descomposición de los organismos vivos (hojas, raíces, microorganismos y pequeños animales) les confiere a los horizontes superiores de los suelos unos colores más oscuros que los de los materiales más profundos. Una excepción a esta regla la constituyen los suelos llamados podsoles, en los que un proceso continuado de lixiviación transporta la materia orgánica hacia los horizontes inferiores, provocando en ellos un color oscuro característico.

Muchos de los silicatos que se encuentran abundantemente en los suelos presentan colores claros, como es el caso del cuarzo y los feldespatos. Otros, por el contrario, son oscuros, como la hornblenda, la biotita o los piroxenos. Los minerales de la arcilla puros presentan colores claros (blanco o grisáceo) si bien su contaminación con minerales de hierro o con materia orgánica les hace perder esa característica. Sin embargo, entre los minerales que tienen más influencia en el color resultante del suelo se encuentran los óxidos de hierro —término con el que se suelen englobar compuestos de hierro de diversa naturaleza como los propios óxidos, hidróxidos e incluso hasta ciertos carbonatos— y los óxidos de manganeso. Su presencia en el suelo puede estar relacionada con la herencia de la roca madre o bien puede ser consecuencia de ciertos procesos edáficos.

La meteorización de los minerales primarios que contienen hierro (biotita, anfíboles, piroxenos, etc.) origina la aparición de minerales secundarios de hierro, la mayor parte de los cuales son óxidos e hidróxidos. Los más importantes, sus características cromáticas y su hábito de aparición se recogen en la tabla 6.

Pero además de minerales, en los suelos aireados existen también ciertos compuestos más o menos amorfos. El color depende del tipo de óxido presente, de su cristalinidad y del grado de hidratación. En general, en ambientes más húmedos y fríos, los colores debidos al hierro son amarillos o pardos, mientras que en otros ambientes son más rojos. En condiciones reductoras el hierro se presenta en forma ferrosa o bien en forma de compuestos ferroso-férricos. En este caso los colores predominantes son verdosos e incluso llegan a ser azulados cuando se forman sulfuros. En ambiente reductor la solubilidad del hierro aumenta, pudiendo individualizarse de los minerales de la arcilla y originando manchas grises o de colores pálidos. En suelos sometidos a un hidromorfismo continuado los colores resultantes son azulados más o menos oscuros. Si, por el contrario, el hidromorfismo es temporal, y por tanto también la reducción,

los colores van alternando entre amarillos o pardos y verdes, grises o azulados.

La meteorización de los silicatos que contienen manganeso en estado de valencia 2 produce diversos óxidos de manganeso tri y tetravalente. Solamente en ambientes persistentemente reductores permanece el manganeso en su forma divalente. Los colores de los compuestos oxidados del manganeso son muy oscuros, en ocasiones prácticamente negros, como es el caso de la pirolusita ($\beta\text{-MnO}_2$). Sin embargo, para que dicho color llegue a condicionar el del suelo se necesitan unas condiciones de alternancia de oxidación y reducción que permitan al manganeso moverse y acumularse suficientemente, lo cual solamente es posible en su estado divalente, que es el único soluble.

Por último, entre los compuestos secundarios que condicionan el color del suelo, cabe citar por su notable influencia el carbonato cálcico (CaCO_3) o calcita, el sulfato cálcico hidratado ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$) o yeso y la sílice (SiO_2), los cuales confieren un característico color blanco. La elevada salinidad de determinados suelos también produce colores blancos como consecuencia de la aparición de

Tabla 6. Principales minerales secundarios de hierro.

	Composición	Color	Condiciones de los suelos donde aparece
Goethita	$\alpha\text{-FeO(OH)}$ Cristalino	Amarillento 7,5YR a 10YR	Climas templados y frescos, húmedos a subhúmedos. Climas tropicales. Suelos moderadamente hidromorfos en climas subtropicales.
Hematites	$\alpha\text{-Fe}_2\text{O}_3$ Cristalino	Rojo 2,5YR a 7,5 YR	Suelos tropicales y subtropicales. Climas mediterráneos.
Lepidocrocita	$\gamma\text{-FeO(OH)}$ Cristalino	Anaranjado 5YR a 7,5YR	Suelos no calizos, hidromorfos, de climas templados.
Maghemita	$\gamma\text{-Fe}_2\text{O}_3$ Cristalino	Pardo rojizo Oscuro	Suelos tropicales y subtropicales.
Ferrihidrita	Fe(OH)_3 Microcristalino o amorfo.	Pardo rojizo 5YR a 7,5YR	Suelos ácidos de las zonas frías y templadas, ricos en materia orgánica.

eflorescencias en su superficie.

Es evidente el interés que, en teledetección, tiene el conocimiento de la relación entre el color del suelo y su composición, ya que, a partir del estudio de los niveles de reflectancia en las bandas del espectro visible podrán extraerse conclusiones acerca de la naturaleza del suelo, o se podrán identificar superficies salinas o con altos contenidos en yesos o en carbonato cálcico por su alto brillo en esa región espectral.

Variabilidad

La acción de los agentes naturales produce variaciones espaciales y temporales en la naturaleza química y en las propiedades físicas de los suelos. Se denomina perfil de un suelo a una sección vertical del mismo. Como el suelo procede de la degradación de un determinado material parental –normalmente subyacente en las capas profundas–, es frecuente encontrar en los perfiles toda una gama de capas de diferentes características que se denominan horizontes: desde los más parecidos a la roca madre situados en los niveles inferiores hasta aquellos otros que apenas recuerdan su procedencia o que prácticamente están compuestos por materia orgánica de origen vegetal.

Las variaciones espaciales de los suelos se recogen

en mapas edafológicos. Este tipo de cartografía temática es de carácter cualitativo, pues las variables geográficas representadas –tipos de suelo– han sido simplificadas a la escala nominal mediante la sustitución los gradientes de cambio por fronteras absolutas y la adopción de la clase modal. Los resultados son mapas de clases discretos que difieren del resultado de la observación remota. Precisamente una de las aplicaciones más familiares de la teledetección aplicada a la descripción de la variación espacial de los suelos es el uso rutinario de la fotografía aérea –considerando ésta una forma más de la teledetección– para la cartografía precisa de los suelos.

La evolución de un suelo es un proceso continuo, pero demasiado lento para la escala temporal utilizada en teledetección. Por ello, esta disciplina está más bien indicada para la medida y seguimiento de cambios más rápidos y frecuentes que los que en rigor pertenecen a la pedología. Las variaciones de algunas de las propiedades de los suelos son detectables desde plataformas orbitales, como la rugosidad, la estructura y la humedad, si bien es cierto que frecuentemente se ven alteradas por las condiciones meteorológicas, los métodos de cultivo, la erosión y la cubierta vegetal. Estas variaciones temporales afectan decisivamente a las propiedades reflectivas de la superficie, por lo que dificultan la caracterización de los suelos mediante teledetección.

Características reflexivas del suelo desnudo

Los suelos no son cuerpos simples sobre los cuales puedan establecerse modelos sencillos para explicar las interacciones entre la radiación y la materia. Su superficie es rugosa y su composición muy diversa. Las partículas presentes adoptan gran diversidad de

formas y tamaños y su propia naturaleza les confiere distintos índices de refracción. Por todo ello, el resultado de la incidencia de un fotón sobre un suelo no puede ser caracterizado *a priori* de un modo absoluto. Ha de recurrirse a modelos que simplifiquen la

realidad y puedan proporcionar soluciones aproximadas al problema planteado.

Además, para que la teledetección sea útil en este contexto, es necesario relacionar las características reflectoras de los suelos con su composición física y química y, si es posible, también con el resto de las propiedades descritas en los apartados anteriores.

Modelos de reflectancia

Los modelos de reflectancia del suelo están contruidos sobre el concepto de dispersión múltiple. Cuando un fotón incide sobre una partícula de materia del suelo, sus cargas eléctricas oscilan por acción del campo inducido por la onda electromagnética. Las cargas excitadas ocupan estados energéticos superiores y su vuelta al estado primitivo puede realizarse emitiendo energía en otra longitud de onda o transformando la radiación incidente en otras formas de energía, hablándose en ese caso de energía absorbida. Pero la transición energética puede también llevarse a cabo irradiando de nuevo la energía en la misma longitud de onda que la radiación incidente, aunque en otras direcciones. Ésta es la radiación dispersada.

Si en lugar de una sola partícula se considera un gran número de ellas confinadas en un espacio reducido, de forma que estén muy próximas unas a otras, la radiación dispersada por una partícula puede sufrir interacciones con otra partícula próxima, de tal forma que la promoción de los niveles energéticos de las partículas de suelo no sólo sea fruto de la presencia de un campo electromagnético inducido por una fuente externa, sino también del producido por la dispersión de partículas adyacentes. Este fenómeno se conoce como dispersión múltiple. Dada la naturaleza compleja de la composición y la estructura de los sue-

los, la reflexión de la luz solar sobre ellos puede considerarse un caso de dispersión múltiple. Toda vez que este tipo de dispersión puede ser abordada considerando que el medio dispersante resulta ser la suma de muchos *difusores* individuales, es habitual abordar el estudio de la dispersión múltiple como la composición de los efectos de una dispersión simple sobre cada difusor individual.

El fenómeno de dispersión simple (*single scattering*) se conoce bien, desde que Mie enunció su teoría de la dispersión para partículas esféricas de un tamaño similar a la longitud de onda incidente. Las partículas del suelo no tienen esa forma regular, pero a pesar de ello, la teoría de Mie puede ser aplicada sustituyendo las partículas irregulares por esferas de radio, superficie o volumen equivalente, teniendo en cuenta que los resultados son sólo aproximados. Mediante la aplicación de la teoría de la dispersión simple se obtienen los parámetros básicos que describen la radiación dispersada, como son (i) la sección transversal de dispersión σ , (ii) el albedo de dispersión simple ω y (iii) la función de fase de la dispersión P . La sección transversal de dispersión es la proporción de la energía dispersada en todas las direcciones con respecto a la incidente, y tiene dimensiones de área, ya que la energía dispersada está en proporción directa a la superficie normal interceptora. Dicho de otro modo, σ corresponde al área necesaria para dispersar una cantidad de energía igual a la radiación que dispersa una unidad de área de otra superficie que fuera completamente reflectora.

Dispersión múltiple

La dispersión múltiple se estudia bajo el enfoque de la transferencia radiativa. De forma análoga a la exposición hecha para la atmósfera, la ecuación de transferencia radiativa para cualquier otro medio

confinado entre dos planos paralelos puede escribirse en función de la intensidad radiante de este modo:

$$\mu \frac{\partial I}{\partial \tau}(\tau, \mu, \phi) = -I(\tau, \mu, \phi) + J(\tau, \mu, \phi)$$

donde:

μ es el coseno del ángulo formado por la dirección de propagación y la normal al plano de la superficie.

ϕ es el acimut de la dirección de propagación.

τ es el espesor óptico del medio.

I es la intensidad radiante.

J es un término fuente que engloba la dispersión del flujo difuso y la atenuación de las fuentes externas. Este término puede expresarse como:

$$J(\tau, \mu, \phi) = \frac{\omega}{4\pi} \int_0^1 \int_{-1}^1 P(\mu', \phi', \mu, \phi) I(\tau, \mu', \phi') d\mu' d\phi' + \frac{\omega}{4\pi} P(\mu_0, \phi_0, \mu, \phi) \pi F_0 e^{-(\tau-\tau_0)/\mu}$$

siendo:

ω el albedo de dispersión.

τ_0 el espesor óptico del medio.

P la función de fase.

F_0 la radiancia incidente desde una fuente externa.

La reflexión especular está regida por las ecuaciones de Fresnel. En el caso en el que los dos medios que forman la interfase de reflexión sean dieléctricos se tiene:

$$r_{\perp} = \frac{n_i \cos \theta_i - n_t \cos \theta_t}{n_i \cos \theta_i + n_t \cos \theta_t}$$

$$t_{\perp} = \frac{2n_i \cos \theta_i}{n_i \cos \theta_i + n_t \cos \theta_t}$$

$$r_{\parallel} = \frac{n_t \cos \theta_i - n_i \cos \theta_t}{n_i \cos \theta_i + n_t \cos \theta_t}$$

$$t_{\parallel} = \frac{2n_i \cos \theta_t}{n_i \cos \theta_i + n_t \cos \theta_t}$$

donde el subíndice \perp indica que el campo eléctrico E es perpendicular al plano de incidencia y el subíndice \parallel que el campo eléctrico está contenido en dicho plano. En las ecuaciones de Fresnel, para cada uno de los dos planos descritos, r es el coeficiente de reflexión, es decir, la proporción de radiación reflejada E_r en relación a la incidente E_i , t es el coeficiente de transmisión o relación entre la radiación transmitida E_t y la incidente, esto es:

$$r_{\perp} \equiv \left(\frac{E_r}{E_i} \right)_{\perp}; \quad t_{\perp} \equiv \left(\frac{E_t}{E_i} \right)_{\perp}$$

$$r_{\parallel} \equiv \left(\frac{E_r}{E_i} \right)_{\parallel}; \quad t_{\parallel} \equiv \left(\frac{E_t}{E_i} \right)_{\parallel}$$

n_i es el coeficiente de refracción del medio desde el que incide la luz y n_t el correspondiente al medio reflector y transmisor. Según la ley de la refracción de Snell, los ángulos de incidencia y refracción están relacionados de la siguiente manera:

$$n_i \sin \theta_i = n_t \sin \theta_t$$

por lo que las ecuaciones de Fresnel pueden ser simplificadas, quedando estas otras:

$$r_{\perp} = -\frac{\text{sen}(\theta_i - \theta_t)}{\text{sen}(\theta_i + \theta_t)}$$

$$r_{\parallel} = \frac{\text{tg}(\theta_i - \theta_t)}{\text{tg}(\theta_i + \theta_t)}$$

$$t_{\perp} = \frac{2 \text{sen} \theta_t \cos \theta_i}{\text{sen}(\theta_i + \theta_t)}$$

$$t_{\parallel} = \frac{2 \text{sen} \theta_t \cos \theta_i}{\text{sen}(\theta_i + \theta_t) \cos(\theta_i - \theta_t)}$$

En el caso de una incidencia normal, el coeficiente de reflexión será:

$$r_h = r_v = r = \frac{n - 1}{n + 1}$$

siendo n el índice complejo de refracción, de componentes real e imaginario N_r y N_i respectivamente. Por tanto:

$$r = \frac{N_r + jN_i - 1}{N_r + jN_i + 1}$$

y

$$|r|^2 = \frac{(N_r - 1)^2 + N_i^2}{(N_r + 1)^2 + N_i^2}$$

En las regiones espectrales alejadas de bandas de absorción fuertes se verifica que $N_r \gg N_i$, por lo que:

$$|r| = \frac{N_r - 1}{N_r + 1}$$

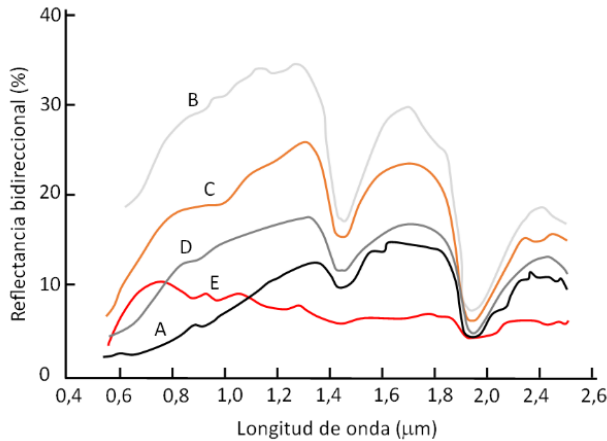
En cambio, en el dominio de las bandas de absorción $N_r \ll N_i$, y entonces:

$$|r| \approx 1$$

Sin embargo, la reflexión de la luz en el suelo no puede considerarse una simple reflexión especular. Consiste, sí, en una serie de reflexiones especulares, pero también de refracciones y dispersiones simples en cadena, como consecuencia de las cuales la radiación emerge de la superficie con diferente longitud de onda y dirigida hacia muchas direcciones, de entre las que suele salir beneficiada la correspondiente a la reflexión especular. Un modelo que permite caracterizar la dispersión múltiple en el suelo es el de Kubelka-Munk, basado en la resolución del sistema:

$$\left. \begin{aligned} dI &= -(\alpha + \sigma)a_0 I dx + \sigma b_0 J dx \\ dJ &= (\alpha + \sigma)b_0 J dx + \sigma a_0 I dx \end{aligned} \right\}$$

donde I es el flujo radiante en la dirección considerada positiva de una capa de difusores elementales, J lo mismo, pero en la dirección contraria, α es la proporción de radiación absorbida por unidad de espesor de la capa, σ es la proporción de radiación dispersada y a_0 y b_0 dos constantes. Para una capa infinitamente gruesa, la solución de las ecuaciones diferenciales planteadas proporciona un valor nulo para la transmisión y una reflectancia dada por:



Signaturas espectrales de distintos tipos de suelos (ver texto).

$$\rho = \frac{1 - \sqrt{\alpha/(\alpha + 2\sigma)}}{1 + \sqrt{\alpha/(\alpha + 2\sigma)}}$$

Reflectancia espectral

La variación de la reflectancia del suelo con la longitud de onda es lo que, en definitiva, permite caracterizar su composición mediante teledetección. Se dispone de abundante información sobre los espectros de transmisión y reflexión de muy diversos minerales, a partir de análisis de laboratorio. Sin embargo, la identificación de las firmas espectrales de cada componente en el diagrama de firmas de un suelo reviste graves dificultades. Por un lado, se encuentran las inherentes a la anchura de banda de los canales empleados en la adquisición de la imagen, si no se trata de un sensor hiperespectral. En estos casos, el sensor no tiene suficiente resolución espectral, y los datos obtenidos se reducen a los valores medios de reflectancia en los intervalos de longitud de onda de cada canal. Por otro, la mezcla de numerosos componentes puros (*endmembers*) con distintas firmas espectrales *ensucia* la gráfica resultante modificando su forma hasta hacer desaparecer los rasgos identificativos. Es la consecuencia del modelo lineal de mezcla, en el que la reflectancia resultante es la suma ponderada de la correspondiente a cada componente, de acuerdo con su abundancia o proporción dentro de la celda. A pesar de ello, existen ciertos rasgos que permanecen a pesar de la adición de las firmas. Con ellos se pueden definir cinco grupos de suelos, clasificados a partir de su textura y de sus contenidos en materia orgánica y óxidos de hierro, cuyas curvas de reflectancia en el ámbito de la región solar del espectro tienen formas claramente distinguibles:

A. Suelos con alto contenido en materia orgánica (>2%) y textura fina.

- B. Suelos con bajo contenido en materia orgánica (<2%) y en óxido de hierro (<1%).
- C. Suelos con bajo contenido en materia orgánica (<2%) y contenido medio en óxidos de hierro (1 a 4%).
- D. Suelos con alto contenido en materia orgánica (>2%), bajo contenido en óxidos de hierro (<1%) y textura moderadamente gruesa.
- E. Suelos con alto contenido en óxidos de hierro (>4%) y textura fina.

El espectro de reflectancia de la mayoría de los suelos se ve afectado por la presencia de dos profundas bandas de absorción centradas sobre 1,4 y 1,9 μm . Estas bandas son, junto a otras de menor intensidad en 0,97, 1,20 y 1,77 μm son armónicas y de combinación, y corresponden a la molécula de agua. La anchura de las bandas de absorción de 1,4 y 1,9 μm indica una disposición desordenada de las moléculas de agua en la matriz líquida del suelo. Estas bandas están presentes en casi todos los espectros de los suelos –y cualquier otra cobertura que contenga agua, como la vegetación– y no comienzan a desaparecer hasta la humedad no baja de un 0,5 %. Otro efecto de la humedad del suelo es la disminución indiscriminada de la reflectancia espectral. Contenidos crecientes de humedad en el suelo producen una disminución indiscriminada de la reflectancia espectral en todas las longitudes de onda, lo cual suele explicarse por el incremento de las reflexiones internas en la película de agua adherida a cada partícula sólida que disminuyen la radiación emergente. Es la razón de por qué un suelo humedecido suele presentar un color más oscuro que el mismo suelo seco.

La presencia de hierro es un factor que condiciona significativamente la firma espectral de los suelos. Este metal puede estar presente en solución, sustituyendo isomórficamente algunos iones Al^{3+} y Si^{4+}

de los minerales de la arcilla o formando parte de otros minerales. La consecuencia cromática de la presencia de hierro suele ser muy espectacular, ya que solamente una pequeña proporción, inferior incluso al 1%, le confiere al suelo un color muy característico. Las asociaciones de hierro con otros elementos muestran varias bandas de absorción. Así, el enlace Fe-O absorbe energía en la región del ultravioleta, el Fe³⁺ absorbe entre 0,7 y 0,87 μm y el Fe²⁺ en 1,0 μm. Existen otras bandas de menor importancia situadas entre 0,4 y 0,55 μm. El resultado de este comportamiento es el color rojizo propio de los suelos férricos (Fe³⁺) y verde-azulada de los ferrosos (Fe²⁺).

Otro de los factores determinantes de la respuesta espectral del suelo es su contenido en materia orgánica. En términos generales, cuando la proporción de ésta aumenta, la reflectancia disminuye en todas las longitudes de onda. La influencia es tan decisiva que contenidos superiores al 2 % producen tal disminución de la reflectancia que impiden manifestarse las bandas de absorción de otros elementos presentes en el mismo suelo, quedando de este modo su presencia enmascarada. Contenidos en materia orgánica superiores al 5 % confieren a la signatura espectral del suelo una forma cóncava característica entre los 0,5 y 1,3 μm que los distingue del resto, que presentan formas convexas en este mismo intervalo.

Seguramente la fuente de peculiaridades de respuesta espectral más variadas es la composición mineralógica del suelo. A modo de resumen de las extensas publicaciones sobre los espectros de numerosos minerales, puede decirse que el cuarzo tiene alta reflectancia en toda la región de las ondas cortas y no presenta bandas de absorción a menos que esté contaminado con impurezas; los anfíboles exhiben una banda de absorción característica en 1,0 μm; los silicatos laminados presentan bandas debidas a los

grupos OH⁻ en 1,4 y 2,2 μm; la calcita absorbe entre 1,8 y 2,5 μm debido al CO₃⁼ de su composición; el yeso muestra bandas de absorción entre 1,8 y 2,3 μm debido a las moléculas de agua de su composición.

También la textura del suelo afecta a la respuesta espectral debido a que el tamaño de las partículas influye en la reflectancia bidireccional. De forma general, para los silicatos y los minerales transparentes la reflectancia bidireccional aumenta cuando disminuye el tamaño de sus partículas, pero con los materiales opacos sucede a la inversa: la reflectancia bidireccional disminuye al disminuir el tamaño de las partículas. Según este comportamiento general, los suelos arcillosos deberían aparecer más claros que los arenosos, y sin embargo la realidad es la contraria: los suelos arenosos suelen ser muy claros (alta reflectancia) y los arcillosos más oscuros (baja reflectancia) como resultado del tamaño de los agregados de los suelos arcillosos, considerablemente mayor que el de las partículas de arena. De este modo pasa a ser la estructura y no la textura el factor determinante en la reflectancia final.

Direccionalidad

El suelo no es una superficie lambertiana perfecta. Por ello, la reflectancia no es solamente función de la longitud de onda, como correspondería a su signatura espectral, sino que depende también de condiciones como la dirección de iluminación y la de visión o medida. Se denomina *factor de reflectancia* a la razón que hay entre el flujo radiante reflejado por una determinada superficie y la que reflejaría una superficie lambertiana ideal con el 100 % de reflectividad.

Albedo

Se define como albedo a la proporción del flujo solar incidente sobre una superficie que es reflejado en cualquier dirección. Esta definición abarca a toda la gama de longitudes de onda del espectro. Sin embargo, es frecuente hablar de albedo en estudios geológicos, refiriéndose solamente al intervalo del espectro comprendido entre el ultravioleta y el infrarrojo medio que, de cualquier modo, concentra la mayor parte de la radiación solar sobre la superficie terrestre. Es posible, con todo, definir un albedo monocromático, que coincide con el concepto de reflectancia espectral.

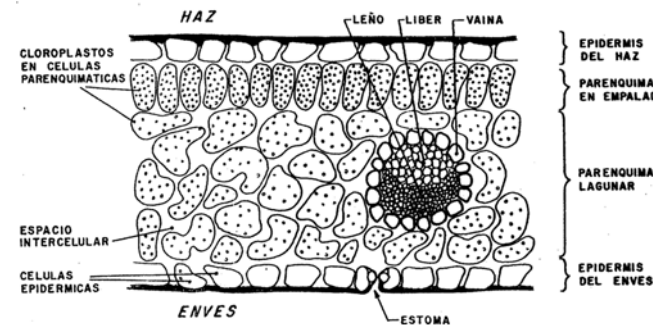
Ordinariamente, el albedo de los suelos oscila entre el 10 y el 40 % y es una variable que, como se ha dicho, depende de la humedad y también de las condiciones de iluminación. En el transcurso del día, el albedo comienza siendo máximo para posiciones solares de mayor ángulo cenital, esto es, al amanecer. Desde ese momento comienza a disminuir hasta alcanzar su mínimo cuando en ángulo cenital es el mínimo posible, a mediodía, y a partir de ese momento vuelve a aumentar de nuevo hasta alcanzar el valor máximo en el crepúsculo. Este fenómeno obedece a la aplicación de las leyes de Fresnel para la reflexión, las cuales establecen la relación entre la reflectancia y el ángulo de incidencia de la radiación.

La vegetación

Visto el comportamiento que tiene el componente inerte de la superficie observada ante la radiación electromagnética, corresponde ahora centrar la atención sobre el otro factor presente casi en la totalidad de la superficie sólida del planeta, la vegetación. En una primera aproximación, se considerará solamente la vegetación aislada, desde el punto de vista de los estudios de laboratorio, o bien su presencia real en el suelo, pero considerando una ocupación exclusiva de la superficie.

Con carácter general, en fitotecnia se habla de plantas de hoja estrecha y de hoja ancha, correspondiendo estas denominaciones a las categorías botánicas de mono y dicotiledóneas. En la sección transversal de una hoja pueden distinguirse varias capas de tejidos diferentes. La más externa de ambas caras se denomina epidermis y está formada principalmente por células epidérmicas, si bien también se presentan

algunas otras entre las que destacan las células oclusivas de los estomas u orificios de intercambio gaseoso, que abundan más en el envés que en el haz. Las paredes de las células epidérmicas suelen estar cubiertas de una cutícula, es decir, de una fina película protectora que juega un papel importante en la reflectancia resultante de la hoja. El mesófilo es la zona media de la hoja, situada entre la epidermis de las caras superior e inferior, y contiene las células especializadas en realizar la fotosíntesis. Se distinguen dos tipos de tejidos: el parénquima en empalizada y el parénquima lagunar. El primero está situado inmediatamente debajo de la epidermis del haz, aunque en algunas especies con hojas verticales también aparecen en el envés, y sus células tienen forma de prismas muy bien ordenados. Debajo se sitúa el parénquima lagunar, caracterizado por contener células irregulares que dejan amplios espacios intersticiales que se comunican con los estomas. Los cloroplastos, ricos en clorofilas, son los corpúsculos



Sección de una hoja vegetal.

celulares especializados en la fotosíntesis y se concentran especialmente en el parénquima en empalizada que, por su situación dentro de la hoja, recibe mayor cantidad de energía radiante.

Características espectrales de la vegetación

Las propiedades reflectivas de la vegetación dependen de tres tipos de variables:

1. Estructura de la cubierta vegetal, principalmente caracterizada por su índice de área foliar, por la orientación de las hojas, por su distribución y tamaño.
2. Propiedades ópticas de los elementos reflectantes, como son los tallos, las hojas, las flores y los frutos.
3. Geometría de la observación, determinada por la orientación relativa entre el sol y la superficie y por la situación del sensor con respecto a la última.

El *índice de área foliar* (*Leaf Area Index*, LAI) es la proporción de suelo cubierto eficazmente por hojas u otro material vegetal. Aunque en mayor medida son las hojas las responsables de la respuesta espectral de las cubiertas vegetales, también en determinados casos cobran gran relevancia los tallos, las flores y los frutos, cuyas proporciones determinarán en definitiva la signatura espectral de la cubierta. El estudio de la firma espectral de una masa de hojas permite extraer algunas conclusiones generales:

1. En el intervalo espectral que se extiende desde el visible hasta el infrarrojo medio, las hojas presentan tres zonas diferenciadas: una en el visible (0,4-0,7 μm), otra en el infrarrojo próximo (0,7-1,3 μm) y otra en el infrarrojo medio (1,3-2,5 μm).

2. En el visible, la reflectancia de las hojas es relativamente baja, no superando apenas el 15 % y la transmitancia es también pequeña. Esto es debido a que los pigmentos fotosintéticos (clorofilas, xantofila, antocianinas y carotenos) tienen fuertes bandas de absorción en este intervalo espectral. Entre ellos, las clorofilas a y b son las más absorbentes –un 65 % del total– y presentan dos bandas de absorción centradas en el azul y en el rojo, por lo que producen un máximo relativo en la signatura espectral en torno a 0,55 μm , que es el responsable del color verde de las hojas.
3. En el infrarrojo próximo las hojas presentan una meseta de alta reflectancia, que puede superar el 50 %. De la energía restante, la mayor parte es transmitida, ya que la absorbancia es sólo del 10 %, por lo que puede considerársele prácticamente transparente en esta región del espectro. Además, hay que considerar que la energía transmitida puede ser interceptada por otra hoja y será reflejada y transmitida en la misma proporción, y así sucesivamente.
4. En el infrarrojo medio, el agua marca su presencia en las hojas produciendo tres fuertes bandas de absorción en la signatura resultante, al igual que ocurre con los suelos: 1,4; 1,9 y 2,5 μm .

Los estudios de laboratorio muestran las características principales de los espectros de absorción de la clorofila disuelta en éter, en los que pueden apreciarse las bandas de absorción. La molécula de clorofila es muy eficaz absorbiendo energía electromagnética en la región del visible, y en especial en el rojo espectral. Las longitudes de onda en la que las cloro-

filas tienen mayor absorbancia se muestran en la Tabla 7.

Tabla 7. Absorción de la molécula de clorofila.

Tipo	Longitud de onda del máximo de absorción (μm)
Clorofila a	0,662
Clorofila b	0,654
Clorofila c	0,627
Clorofila d	0,688

Factores internos que afectan a la reflectancia vegetal

Como ocurre en toda superficie parcialmente lambertiana, la radiación incidente es en parte reflejada especularmente y en parte dispersada. La responsable de la reflexión especular es la cutícula que cubre las hojas, que le confiere un brillo característico a algunas especies vegetales. La luz no reflejada penetra en el interior de la hoja y es sucesivamente refractada, reflejada y difractada. Como consecuencia de ello, la radiación emergente tiene un carácter difuso. Esta dispersión depende de la propia estructura de la hoja, de la composición de sus células, –la presencia de cristales de almidón, por ejemplo– y de su grado de humedad.

En el infrarrojo próximo, las hojas con mayor desarrollo del parénquima lagunar producen mayor dispersión. Por esta razón las plantas de hoja ancha suelen ser más reflectantes en esa región espectral que las de hoja estrecha. De otra parte, la presencia de cloroplastos incrementa la absorción, por lo que necesariamente debe disminuir la dispersión. Esta es la

razón por la cual las hojas suelen tener mayor reflectancia en el envés que en el haz, ya que el parénquima en empalizada se concentra preferentemente en las proximidades del haz. La vellosidad de algunos tipos de hojas incrementa también la reflectancia en el visible y el infrarrojo próximo, no alterando sin embargo la del infrarrojo medio.

Un fenómeno interesante de resaltar es la senescencia de las hojas, es decir el proceso de envejecimiento que conduce a la clorosis (amarilleamiento) y a la marchitez permanente. Durante este proceso, el contenido en clorofilas decrece en favor de los carotenos, que producen colores rojizos y marrones, y de las xantofilas, que producen colores amarillentos. En el infrarrojo próximo la reflectancia solamente varía cuando, como consecuencia de la muerte de la hoja, cambia su estructura interna. Sin embargo, en el infrarrojo medio sí que se registra un incremento de la reflectancia en todo el intervalo espectral, debido a la progresiva pérdida de agua, aunque las características bandas de absorción del agua solamente desaparecen completamente cuando la humedad se hace prácticamente inexistente.

Pero el contenido de agua del material vegetal influye en la reflectancia no sólo de forma directa, dejando su impronta en la signatura espectral mediante las bandas de absorción que le son propias, sino también de forma indirecta, ya que de la humedad depende la turgencia celular y este factor determina la estructura interna de la hoja. En general, en laboratorio se obtienen reflectancias progresivamente mayores en todo el espectro al ir disminuyendo el contenido en agua. En la práctica, sin embargo, este efecto solamente se hace evidente cuando el estrés hídrico es muy acusado.

El estado nutritivo de las plantas es otro factor que

también influye en la reflectancia espectral. La característica principal que puede apreciarse en relación con el vigor vegetal es el amarilleamiento o clorosis, y bien sabido es que numerosos estados carenciales dan como resultado la clorosis. Salvo las plantas leguminosas, el resto del material vegetal es extremadamente sensible a la carencia de nitrógeno soluble en el suelo. Una escasa dotación de este elemento en forma amoniacal produce un típico color amarillo en la planta debido a la disminución del contenido en clorofila. Ello produce un incremento de la reflectancia en el visible, mientras que, como consecuencia de la pérdida de capas celulares, se produce un descenso de la reflectancia en el infrarrojo próximo y medio. Pero la clorosis también puede estar producida por la carencia de algún elemento no tan esencial como el nitrógeno. Tal es el caso del hierro, que interviene en la síntesis de la molécula de clorofila y cuya deficiencia da lugar a la denominada clorosis férrica, caracterizada por un amarilleamiento de las hojas que, al igual que antes, causa un incremento de la reflectancia en el visible. En definitiva, la pérdida de clorofila conduce a un aplanamiento de la signatura espectral característica del material vegetal. Efectos similares producen las plagas y/o enfermedades sobre las plantas, precisamente por lo cual es posible el seguimiento de sus efectos mediante teledetección.

Determinados cultivos de floración muy perceptible durante un período dilatado de tiempo modifican su signatura espectral en estado de floración con respecto a otros períodos vegetativos. Por lo tanto deberá tenerse en cuenta este factor en los períodos afectados, contrastando los registros digitales obtenidos desde satélite con los resultados de la radiometría de campo que, en estas circunstancias, se hace especialmente útil. En especies leñosas también cobra importancia el color y la distribución del fruto por los motivos expuestos. El caso es patente en las coníferas, en las que la proporción de infrutescencias

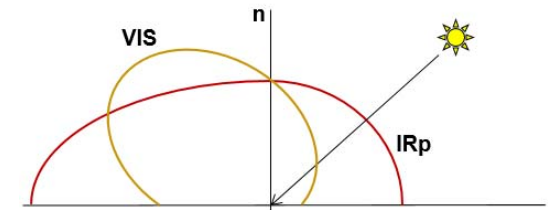
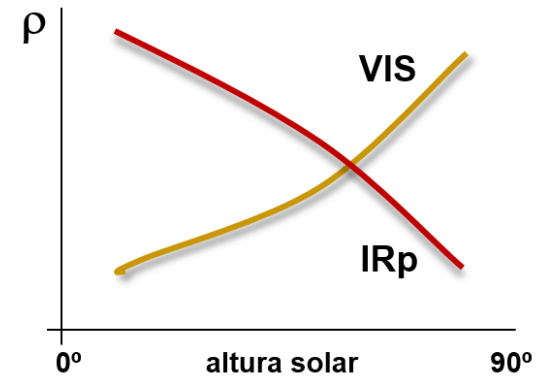
(piñas) tiene, junto con la corteza de troncos y ramas, gran repercusión espectral.

Finalmente, aunque no se trata de reflectancia sino de emisión propia, es oportuno señalar aquí que un déficit acusado de humedad puede conducir a la planta a un estado de estrés hídrico que induce el cierre estomático y consecuentemente el incremento de temperatura, al disminuir o cesar completamente la transpiración del dosel vegetal. Este fenómeno puede detectarse midiendo la temperatura mediante lecturas en el infrarrojo térmico desde el sensor. Una consecuencia del aumento de temperatura es el bloqueo metabólico que sufre la hoja, que interrumpe la fotosíntesis, y permite manifestarse el fenómeno de fluorescencia clorofílica. La medición de la radiancia vegetal en torno a los 760 nm, longitud de onda a la que acontece este fenómeno, permite realizar un seguimiento del estado de humedad del material vegetal.

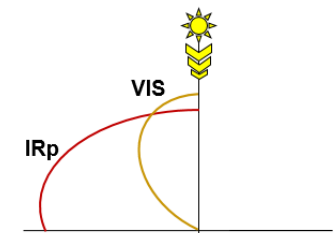
Factores externos que afectan a la reflectancia vegetal aparente

Existen numerosos factores que, si bien no modifican sustancialmente la reflectancia espectral aparente del material vegetal, sí condicionan en cierta medida sus peculiaridades. Uno de ellos es la orientación relativa de la superficie vegetal con respecto a la luz incidente, que depende de cuatro variables: elevación y acimut solar, pendiente y orientación geográfica de la ladera.

La variación horaria de la reflectancia aparente de una cubierta vegetal durante el día está, en principio, sujeta a lo dicho sobre el albedo. Sin embargo, el efecto de incremento de la reflectancia aparente para alturas solares pequeñas se ve en cierto modo contrarrestado por las sombras producidas por las propias plantas sobre las vecinas, especialmente cuando



Sensor moviéndose en el plano de iluminación



Sensor moviéndose en el plano perpendicular al de iluminación

Variación de la reflectancia aparente de las cubiertas vegetales bajo diferentes condiciones de iluminación.

se trata de cubiertas forestales. Los estudios realizados por diferentes investigadores demuestran que la reflectancia aparente en el visible aumenta a medida que el sol asciende, disminuyendo, por el contrario, la correspondiente al infrarrojo próximo. Sin embargo, esta tendencia general puede alterarse en función de la densidad de los árboles y de la forma de sus copas.

La reflectancia aparente no es isotrópica, como lo demuestra la variación registrada cuando se varía la posición relativa del sensor con respecto a la dirección de iluminación. Considerando el sensor en el plano de incidencia y en el lado de los rayos reflejados, la reflectancia aparente en el visible aumenta conforme aumenta el ángulo cenital de la posición del sensor hasta alcanzar el máximo en una posición próxima a los 45°, a partir de cuyo momento disminuye. La del infrarrojo próximo, en cambio, sigue aumentando con el ángulo cenital. El efecto es similar a mantener el sensor fijo y dejar que varíe la elevación solar con las horas. Si el sensor está colocado del lado de incidencia, la reflectancia aparente en el infrarrojo próximo se mantiene constante y la del visible disminuye al aumentar el ángulo cenital. Si se considera el sensor situado en un plano perpendicular al de incidencia de los rayos solares, la reflectancia aparente en el visible disminuye y la del infrarrojo próximo aumenta al aumentar el ángulo cenital del sensor. La explicación a tales fenómenos reside por un lado en el hecho

de que las sombras son menos reflectantes en el visible que en infrarrojo próximo, y por otro, que el suelo en el visible es más claro que la vegetación. Además, cuando aumenta el ángulo cenital del sensor, la contribución del suelo disminuye al ir perdiéndose visibilidad, en favor de la de las sombras.

La reflectancia aparente, como se ha visto, depende de la geometría del triedro formado por la normal a la superficie, la dirección de incidencia de los rayos solares y la de posición del sensor. Sin embargo, no afecta sólo la orientación general de la superficie con respecto al sol, sino también la orientación relativa hoja a hoja. Por ello es importante considerar la inclinación de las hojas, caracterizando la distribución espacial de dicha inclinación —materializada por la función LIDF (*Leaf Inclination Directional Factor*)—, pues no todas las hojas presentan la misma orientación con respecto al sol. Ciertas plantas cultivadas presentan fototropismo, en virtud del cual las hojas siguen diariamente el movimiento del sol, modificando su LIDF, y con ella, la reflectancia resultante.

En cualquiera de los casos, en superficies vegetales naturales es necesario considerar la influencia que la proporción de suelo no cubierto eficazmente por el dosel vegetal puede tener en la signatura resultante, cuando la resolución del sensor es inferior a la tasa de variación de la cubierta y el resultado es una imagen de píxeles mixtos.

Respuesta espectral del sistema suelo-vegetación

La gama de cobertura vegetal de una superficie en términos porcentuales puede oscilar entre 0 (suelos desnudos) y 100 (vegetación densa y tupida) y se

mide por la *fracción de cabida cubierta* (Fcc). Este parámetro contabiliza el grado de recubrimiento del suelo por la proyección vertical de la masa vegetal,

Tabla 8. Fracciones de cabida cubierta en zonas forestales.

Cobertura	Fracción de cabida cubierta (%)
Espesura completa	85-100
Espesura defectiva	70-85
Espesura abierta	40-70
Masas claras	10-40
Rasos	0-10

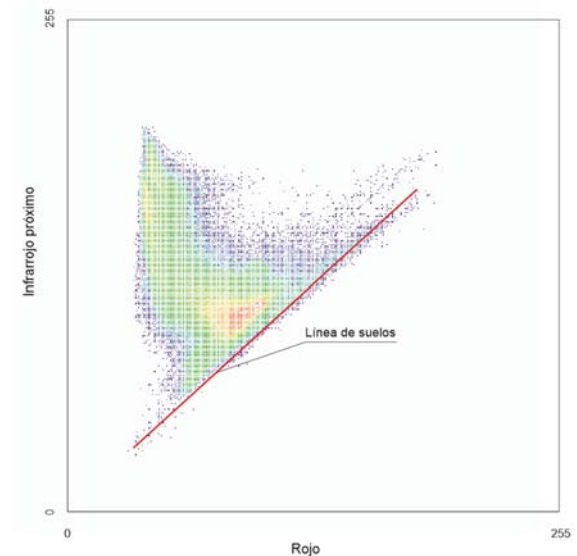
copas de los árboles en las cubiertas forestales y en cultivos leñosos y, por extensión, material vegetal en cultivos herbáceos. Así pues, tal como se recoge en la Tabla 8, una Fcc de entre 85 y 100 % indica espesura completa, de 70 a 85 % espesura defectiva, entre 40 y 70 % espesura abierta, las masas claras presentan Fcc de entre 10 y 40 % y finalmente los rasos, de 0 a 10 %. Hasta ahora solamente se han considerado superficies ideales situadas en uno u otro extremo. Sin embargo, la realidad proporciona coberturas vegetales con fracciones de cabida cubierta de valores intermedios, por lo que todas las propiedades reflectoras anteriores deben ser matizadas.

Los estudios encaminados a caracterizar el comportamiento del sistema suelo-vegetación se han realizado tradicionalmente analizando el comportamiento espectral de superficies de cultivos anuales durante todo su ciclo vegetativo, desde el estado de siembra hasta el de marchitez. En los primeros estadios del cultivo, la signature espectral de la superficie es netamente la correspondiente al suelo sobre el que se asienta. En el estado de nascencia comienzan a manifestarse algunos rasgos de la signature del cultivo, si bien las bandas de absorción típicas del material vegetal solamente se perciben como mínimos incipientes. Estos rasgos característicos se van afianzando a medida que el cultivo crece y cubre el suelo.

Evolución de los cultivos

Conforme se desarrolla una cubierta vegetal, el índice de área foliar se incrementa hasta llegar a un máximo, normalmente próximo a la fecha de la floración, a partir de cuyo momento decrece hasta alcanzar el punto correspondiente a su marchitez permanente. Durante el crecimiento del cultivo la contribución del suelo disminuye en favor del material fotosintéticamente activo. En ese período, la reflectancia en el visible y en el infrarrojo medio disminuyen progresivamente, mientras que la del infrarrojo próximo aumenta notablemente, como corresponde a cualquier vegetal hasta su plena madurez. En la senescencia del cultivo la reflectancia en el visible aumenta, como se ha visto, y la del infrarrojo próximo decrece ligeramente.

Si se analizan las posiciones que van dejando en un espacio multidimensional los extremos de los vectores de características de las celdas de la imagen multiespectral de una superficie cultivada desde el estado de prenascencia hasta el de marchitez del cultivo, la nube de puntos obtenida en el transcurso del tiempo adquiere la forma de una especie de gorro borlado (*tasseled cap*). Si sólo se considera un espacio bidimensional con las bandas roja e infrarroja, la base de tal gorro está definida por una línea denominada de suelos, a lo largo de la cual se distribuyen los distintos tipos de suelos de acuerdo a sus características propias y a su humedad. A medida que el cultivo se desarrolla y llega a su pleno vigor, los puntos que lo representan tienden a aproximarse al eje del infrarrojo, retirándose de la línea de suelos. Cuando se inicia el período de senescencia, la nube de puntos va convergiendo hacia un punto único, al que alcanza cuando el cultivo llega al estado de marchitez permanente, y cuya posición tiende a regresar de nuevo a la línea de suelos. A resultados similares se llega si se



construyen diagramas de dispersión a partir de subescenas en las que existan parcelas con toda la gama de proporciones de ocupación del suelo por parte del material vegetal.

Es posible reducir la dimensionalidad de la imagen multispectral a sólo tres variables representativas del estado fenológico del material vegetal mediante combinaciones lineales de sus bandas originales. De esta manera se han desarrollado transformaciones que, a partir de los datos originales deducen componentes como el brillo, el verdor y la humedad, que se estudiarán en el capítulo dedicado al álgebra de bandas.

Grado de cobertura

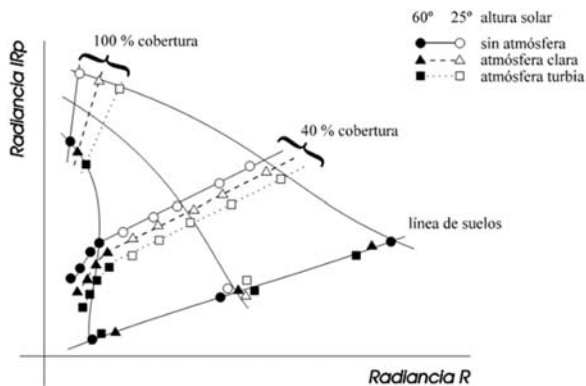
La representación bidimensional de la reflectancia de una escena según las bandas del rojo (R) y el infrarrojo próximo (IRp) proporciona una nube de puntos convergentes hacia el eje IRp a medida que la fracción de cabida cubierta va creciendo. Sobre esa nube pueden trazarse isolíneas de cobertura, lo que equivale a decir de igual índice de área foliar, cuya posición puede desplazarse ligera y paralelamente según sea la turbidez atmosférica. Las distintas situaciones posibles dentro de cada isolínea corresponden en general a diferentes estados de humedad de la celda. De entre todas las posibles líneas *isoLAI* que pueden aparecer en la representación R/IRp, la más próxima al eje R contiene los vectores de características de las celdas ocupadas por suelo desnudo (LAI=0) y se denomina línea de suelos.

En plantas cultivadas en hileras, la orientación de los surcos con respecto a la dirección de visualización puede ser determinante en la reflectancia que mide el sensor. Este efecto suele ser más severo en el visible, en el cual el suelo es más reflexivo que el mate-

rial vegetal, que en el infrarrojo próximo. La influencia ejercida por la orientación de las líneas depende, a su vez, del grado de cobertura, llegando a ser máximo cuando está entre el 40 y el 60 %.

Tanto en el caso del estudio evolutivo de una misma parcela que progresivamente se va cubriendo de material vegetal como de la imagen instantánea de un área más extensa en la que simultáneamente coexisten parcelas con diferentes grados de cobertura vegetal, el diagrama de dispersión IRp/R, es decir, la representación gráfica bidimensional de la reflectancia en esas bandas presenta la característica forma de gorro borlado citada anteriormente. En esta nube de puntos, las celdas de suelo desnudo se ubican en su base, el agua, de haberla, ocupa el lugar más próximo al origen de coordenadas, la vegetación vigorosa se sitúa cerca del eje IRp (valores muy bajos de R) y en la posición más alejada del eje R (valores de IRp que pueden superar el 50 o el 60 % en algunos casos), las infraestructuras urbanas suelen ocupar posiciones próximas a la línea de suelos, y tanto las nubes como las superficies de nieve limpia aparecen en las posiciones de coordenadas más altas en dentro del diagrama.

Los índices de vegetación son medidas basadas en la situación espectral de una determinada celda de terreno, especialmente en la región del visible y del infrarrojo próximo, en las cuales la vegetación tiene rasgos peculiares. Su estudio se abordará dentro del álgebra de banda



Variación de la radiancia de una superficie natural con diferentes proporciones de vegetación en el infrarrojo próximo y en el rojo espectral.

6

La imagen digital

Introducción

El análisis de las imágenes de satélite puede abordarse desde un punto de vista puramente visual, con los mismos criterios que se aplican a la fotointerpretación, aunque con los matices que introduce el haber sido adquiridas desde plataformas orbitales en lugar desde aeronaves volando a baja altura. Sin embargo, el potencial de información que proporcionan las imágenes de satélite requiere un estudio más profundo que solo puede realizarse mediante análisis digital. Para ello, las imágenes deben ser consideradas objetos matemáticos susceptibles de ser manipuladas

en sistemas informáticos para poder realizar sobre ellas operaciones matemáticas, a partir de las cuales se puedan deducir propiedades inherentes a la imagen que en el simple análisis visual pasan inadvertidas.

Para abordar el análisis digital de las imágenes es muy útil considerar el conjunto [imagen bruta - sistema de tratamiento - imagen derivada] como un sistema. En el apartado siguiente se realizará una aproximación a la ingeniería de sistemas.

Sistemas

Un sistema, es una estructura, dispositivo, esquema o procedimiento —real o abstracto— que relaciona, en un tiempo de referencia dado, una entrada, causa o estímulo de materia, energía o información y una salida, efecto o respuesta de información, energía o materia.

Esta definición genérica, destinada en su origen al campo de la hidrología, engloba la práctica totalidad de los espacios de aplicación de la ingeniería de sistemas y, en particular, al análisis digital de imágenes. En el ámbito de la teledetección, el factor tiempo incluido en la definición suele carecer de importancia al aplicarse generalmente a imágenes estáticas. Sin embargo, en otros campos, sí que es algo para tener en cuenta, como sucede en el análisis y el manejo de imágenes en movimiento, como por ejemplo en visión artificial o en televisión.

Es importante considerar que el sistema puede estar constituido por uno o más elementos reales sustentado en soportes físicos, o bien ser una estructura abstracta que transforma *entradas* en *salidas*, y que éstas pueden ser de la misma naturaleza o no. En el ámbito de las finanzas, una entrada de naturaleza informativa, que puede considerarse abstracta, puede producir, tras las operaciones bursátiles necesarias, una salida basada en el pánico que resulte en una caída masiva de las cotizaciones, aspecto que sí tiene naturaleza real y contable. En otras ocasiones, la entrada y la salida son del mismo tipo, como sucede en el campo de los estudios hidrológicos. La entrada a una cuenca hidrográfica puede ser la precipitación registrada durante un aguacero o en un episodio de lluvia extrema, y la salida es el caudal aforado o la cota de la lámina de agua en la llanura de inundación medida en un punto sensible de la cuenca aguas



Transformación de una función de entrada f_e en otra de salida f_s por la acción de un sistema \mathcal{K} .

abajo. Bajo la teoría de sistemas se analizan aspectos muy diversos de la vida cotidiana, desde la sociología pasando por la geografía, la economía y la informática hasta numerosas especialidades de la ingeniería.

En términos generales, un sistema está constituido por tres componentes:

- La entrada $X(t)$.
- El operador o función de transferencia \mathcal{H} .
- La salida $Y(t)$.

La relación existente entre la entrada y la salida se expresa mediante:

$$Y(t) = \mathcal{H}[X(t)] \quad (1)$$

que indica que $Y(t)$ es la respuesta de \mathcal{H} ante la entrada $X(t)$.

Los sistemas pueden ser lineales o no lineales. Aunque en el ámbito del tratamiento digital de imágenes existen aplicaciones basadas en estos últimos, en la mayoría de los casos, los que se utilizan son los lineales. Para que un sistema sea lineal debe cumplir dos propiedades:

- Principio de proporcionalidad: si a la entrada X_1 al sistema le corresponde la salida Y_1 , la entrada $C \cdot X_1$ debe producir una salida $C \cdot Y_1$, siendo C una constante.
- Principio de aditividad: si dos entradas individuales X_1 y X_2 producen, respectivamente, dos salidas Y_1 e Y_2 , la entrada $X_1 + X_2$ debe producir la salida $Y_1 + Y_2$.

Estas dos propiedades se pueden combinar escribiendo:

$$\begin{aligned} \mathcal{H}(C_1 X_1 + C_2 X_2) &= \\ &= \mathcal{H}(C_1 X_1) + \mathcal{H}(C_2 X_2) = C_1 \mathcal{H}(X_1) + C_2 \mathcal{H}(X_2) = \\ &= C_1 Y_1 + C_2 Y_2 \end{aligned} \quad (2)$$

que significa que el resultado de aplicar el sistema a la suma de dos entradas amplificadas es el mismo que si se suman los resultados amplificados de las salidas de cada una de las entradas sometidas al sistema. En palabras más sencillas, considerando el filtrado de imágenes como un sistema lineal y teniendo en cuenta solamente la propiedad de aditividad, se puede decir que el filtrado de la suma de dos imágenes produce el mismo resultado que la suma de los filtrados de cada una de ellas. Más adelante se verá que también hay filtros, así como otras operaciones interesantes en teledetección, que no son lineales.

Los instrumentos de adquisición de imágenes son, por lo general, sistemas lineales invariables en el tiempo, en los cuales la magnitud de la salida no depende del tiempo sino solo de la entrada. En concreto, en un sistema formador de imágenes, la entrada es la radiancia de la escena, el sistema está constituido por el sensor y la salida es la irradiancia sobre el plano focal del sensor o, si además de la óptica se considera la electrónica del instrumento, su traducción a voltaje o niveles digitales de la imagen formada. En teledetección se considera a menudo la atmósfera como una parte más del sistema, ya que su presencia modula la señal procedente de la superficie.

La imagen digital

En su definición académica, la imagen es la reproducción de la figura de un objeto por la combinación de los rayos de luz que proceden de él. El diccionario realmente no aclara gran cosa, puesto que proporciona una descripción circular del concepto cuando habla de *reproducir la figura*. Más allá de las definiciones formales, aquí lo que interesa es su descripción física, y desde el punto de vista óptico, una imagen es una distribución bidimensional de energía electromagnética, particularmente circunscrita al espectro visible. Aun así, la teledetección amplía la noción a cualquier otro intervalo de longitudes de onda del espectro sin ningún problema, y de este modo, aunque no sea visible por el ojo humano, se puede configurar perfectamente una imagen a partir de la radiancia espectral extendida a cualquier tipo de radiación electromagnética. De manera más sucinta, la imagen digital es un modelo digital de radiancia.

Esto quiere decir que la energía que puede registrarse como consecuencia de la reflexión de la radiación sobre una superficie bajo observación depende de la composición de la luz incidente, de la reflectancia espectral de la superficie y del intervalo de longitudes de onda estudiado. Pero es importante considerar, además, que esta energía —y en definitiva, la imagen formada a partir de su registro— dependen de cada posición en el terreno, puesto que cada punto puede tener una reflectancia diferente. De este modo se puede proponer un modelo matemático para una imagen continua, monocroma y fija, esto es, considerando una sola longitud de onda y además prescindiendo de cambios o de movimiento:

$$z = f(x, y) \quad (5)$$

siendo $x, y, z \in \mathfrak{R}$, esto es, pudiendo tomar cualquiera de las variables valores reales arbitrariamente próximos.

Es posible normalizar la imagen adoptando un rango de existencia para la variable z de modo que $0 \leq z \leq 1$ y atribuyendo el color negro al 0 y el blanco al 1. Entre ambos extremos se distribuirá toda la gama de grises proporcional al valor de la imagen en cada punto.

Un sistema de tratamiento digital de imágenes no puede procesar una imagen de este tipo, puesto que las posiciones de sus puntos difieren infinitesimalmente en distancia y su tamaño en consecuencia tiene al infinito. Se hace necesario *discretizar* la imagen. Se entiende por imagen discreta aquella cuyas



Una imagen real.

Considérese una fuente radiante cuya distribución espectral de radiación esté dada por una función c_λ . Si la energía que procede de ella incide sobre una superficie de reflectancia espectral ρ_λ , la energía devuelta para cada longitud de onda será:

$$e_\lambda = c_\lambda \rho_\lambda \quad (3)$$

y la devuelta para el intervalo de longitudes de onda comprendidos entre λ_1 y λ_2 :

$$e_{[\lambda_1, \lambda_2]} = \int_{\lambda_1}^{\lambda_2} c_\lambda \rho_\lambda d\lambda \quad (4)$$

posiciones planimétricas son discretas, es decir, que solo pueden ocupar ciertas posiciones y no las intermedias. Para construir una imagen así a partir de una real es preciso muestrear el espacio objeto generando una malla regular y atribuir un solo valor de brillo a cada nodo, promediando la iluminación procedente de la región dominada por cada uno de esos puntos. El modelo matemático para esta imagen será:

$$z = f(c, l) \tag{6}$$

siendo ahora $c, l \in \mathbb{Z}$ y $z \in \mathfrak{R}$, o sea, las posiciones sobre el plano son números enteros, que pueden organizarse en columnas (c) y líneas (l), aunque el brillo de cada punto todavía puede adoptar cualquier valor real entre 0 y 1. Así pues, en la imagen discreta, las coordenadas de los puntos son discretas y la luminancia puede variar de forma continua. Si las coordenadas (c, l) cumplen que $0 \leq c \leq C-1$ y $0 \leq l \leq L-1$, entonces la imagen, además de discreta, es finita y puede representarse por una matriz de dimensiones $C \times L$, siendo C el número total de columnas y L el de líneas de la imagen. Cada celda de la matriz corresponderá a una celda de imagen. Pero sigue siendo esta imagen imposible de someter a tratamiento digital debido a la continuidad de los valores de luminancia. Por ello es preciso hacer discretos también los valores de luminancia. De esta manera se enuncia el modelo de imagen digital:

$$ND = f(c, l) \tag{7}$$

donde ND significa *Nivel Digital*, que es el brillo o valor de gris discreto adoptado en sustitución de la variable continua anterior z . Ahora, en consecuencia, $c, l, ND \in \mathbb{Z}$. Finalmente, para que un sistema informático pueda tratar estas imágenes es necesario imponer una condición más: el nivel digital también

debe adoptar valores enteros en la forma:

$$0 \leq ND \leq 2^k - 1 \tag{8}$$

y por lo tanto tener un valor mínimo igual a 0 y unos valores máximos que dependerán de la resolución radiométrica del instrumento de adquisición, tal como muestra la Tabla 1.

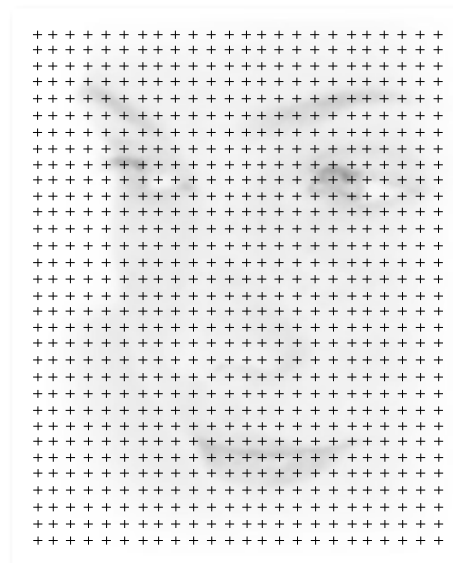
Tabla 1. Número de ND en función del número de bits.

Núm. bits	Núm. ND	Núm. bits	Núm. ND
1	2	9	512
2	4	10	1 024
3	8	11	2 048
4	16	12	4 096
5	32	13	8 192
6	64	14	16 384
7	128	15	32 768
8	256	16	65 536

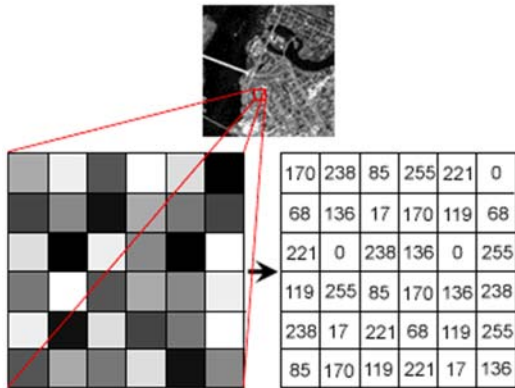
Se ha dicho anteriormente que la imagen de satélite puede explorar regiones de longitudes de onda del espectro más allá del visible. Esto hace que frecuentemente las imágenes almacenen la información obtenida del terreno en distintas bandas, cada una de ellas reservada a un intervalo espectral concreto. En consecuencia, la imagen en realidad no contiene una sola *capa* de información, sino tantas como bandas. Se habla entonces de una imagen multiespectral. Cada celda conserva, pues, tantos niveles digitales como bandas tenga la imagen. El modelo matemático para la imagen multiespectral será:

$$\overline{ND} = \overline{f}(c, l) \tag{9}$$

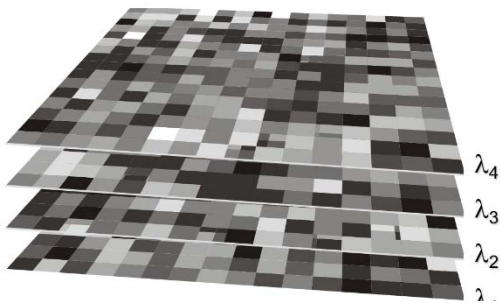
es decir:



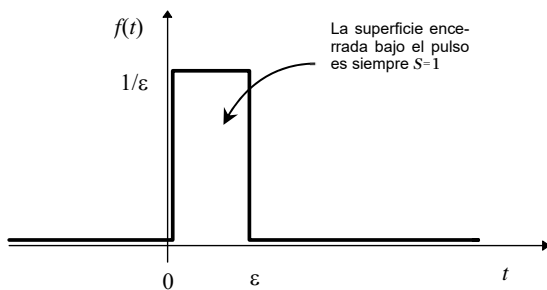
Proceso de discretización de la imagen e imagen discreta resultante.



La imagen digital.



La imagen multispectral.



Definición de la función δ de Dirac como límite del pulso unitario.

$$\begin{pmatrix} ND_1 \\ ND_2 \\ \vdots \\ ND_n \end{pmatrix} = \begin{pmatrix} f_1(c,l) \\ f_2(c,l) \\ \vdots \\ f_n(c,l) \end{pmatrix} \quad (10)$$

La función de dispersión puntual

En un sistema formador de imágenes, el nivel digital atribuido a una celda concreta corresponde al promedio de los valores de radiancia de la región del terreno homóloga. Se puede hablar de espacio objeto y espacio imagen para distinguir entre el terreno y la imagen formada en el plano focal del instrumento de adquisición. Desde la óptica de la ingeniería de sistemas puede considerarse el sistema de adquisición de imágenes como el conjunto formado por el espacio objeto, la atmósfera interpuesta entre el terreno y el propio instrumento formador de imágenes.

Abandonando por un momento la formación de imágenes bidimensionales y recurriendo a las formulaciones clásicas de teoría de señales, se define un impulso unitario o función delta de Dirac¹ como una función pico de intensidad o amplitud infinita y una duración infinitesimal. Puede definirse el impulso unitario como el límite de un pulso rectangular de duración ϵ y amplitud $1/\epsilon$ cuando $\epsilon \rightarrow 0$:

¹ El Nobel de física inglés Paul Dirac descubrió el electrón y predijo la existencia del positrón.

El conjunto de niveles digitales de cada celda se denomina *vector de características* y representa una posición en un espacio espectral de tantas dimensiones como bandas tenga la imagen multispectral.

$$\delta(t) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \quad (11)$$

La definición matemática del impulso unitario viene dada por:

$$\begin{cases} \delta(t) = \begin{cases} 0 & \forall t \neq t_0 \\ \infty & t = 0 \end{cases} \\ \int_{-\infty}^{\infty} \delta(t) dt = \int_{-\epsilon}^{\epsilon} \delta(t) dt = 1 \quad \epsilon > 0 \end{cases} \quad (12)$$

Y para el instante t_0 :

$$\begin{cases} \delta(t - t_0) = 0 \quad \forall t \neq t_0 \\ \int_{-\infty}^{\infty} \delta(t - t_0) dt = 1 \end{cases} \quad (13)$$

El impulso unitario tiene propiedades muy interesan-

tes en el tratamiento de señal en general, y de imágenes digitales en particular. El producto del impulso por una función de prueba cualquiera equivale a aplicarlo solo en el instante en el que el impulso no es nulo:

$$\delta(t - t_0)f(t) = \delta(t - t_0)f(t_0) \tag{14}$$

Y como consecuencia de ello,

$$\begin{aligned} \int_{-\infty}^{\infty} \delta(t - t_0)f(t) dt &= \int_{-\infty}^{\infty} \delta(t - t_0)f(t_0) dt = \\ &= f(t_0) \int_{-\infty}^{\infty} \delta(t - t_0) dt \end{aligned} \tag{15}$$

o lo que es lo mismo, si se introduce el intervalo $\tau = t - t_0$:

$$\begin{aligned} \int_{-\infty}^{\infty} \delta(t - t_0)f(t) dt &= \int_{-\infty}^{\infty} \delta(\tau)f(\tau + t_0) d\tau = \\ &= \int_{-\infty}^{\infty} \delta(t)f(t + t_0) dt = f(t_0) \end{aligned} \tag{16}$$

Es decir, la integración del producto de un impulso en el entorno del instante t_0 por una función o señal da como resultado el valor de la función en ese instante. Esta peculiaridad se conoce como *propiedad de exploración o muestreo* del impulso unitario.

Mediante ella, el impulso unitario permite extraer, retener o cribar el valor de una señal en el instante considerado. Por ello sirve para muestrear (*sample*) el comportamiento de una función extendida durante un intervalo. Es la forma de convertir en digitales y discretas señales analógicas continuas.

Es interesante considerar que el impulso unitario también se puede definir como la derivada de la función *escalón unitario de Heaviside*:

$$\delta(t) = \frac{du(t)}{dt} \tag{17}$$

definiéndose tal función como sigue:

$$\int_{-\infty}^{\infty} u(t)f(t)dt = \int_0^{\infty} f(t)dt \tag{18}$$

Es decir, cuando se aplica la función de Heaviside a una señal de prueba sólo extraen los valores positivos de esta última. Para ello, debe ser:

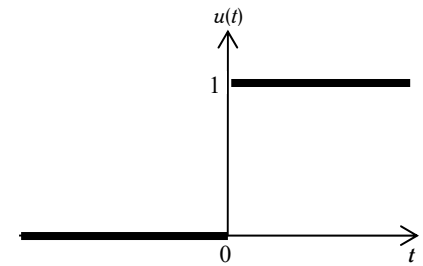
$$u(t) = \begin{cases} 1 & \forall t > 0 \\ 0 & \forall t < 0 \end{cases} \tag{19}$$

razón por la que se denomina escalón unitario. Para un instante genérico t_0 :

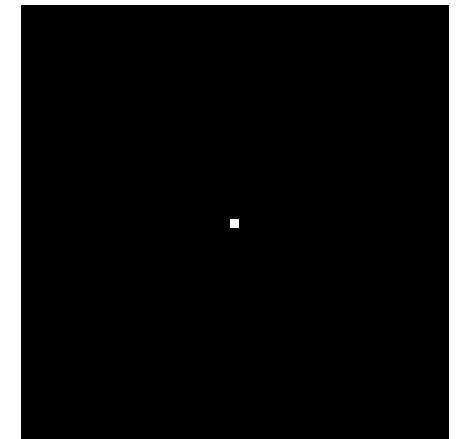
$$u(t - t_0) = \begin{cases} 1 & \forall t > t_0 \\ 0 & \forall t < t_0 \end{cases} \tag{20}$$

La función de Heaviside sólo extrae los valores de cualquier función de prueba $f(t)$, a partir del instante considerado y no está definida para $t = t_0$.

En el contexto del tratamiento digital de imágenes, el concepto de impulso unitario ha de extenderse a dos dimensiones, y además será necesario enunciarlo de forma discreta y no continua. El impulso unitario bidimensional puede asimilarse a un punto o celda del máximo brillo sobre un fondo completamente oscuro.

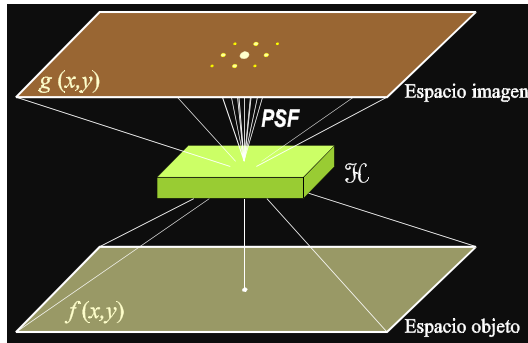


Escalón unitario de Heaviside.



Impulso unitario en una imagen.

Cualquier perfil radiométrico que contenga a esa celda brillante tendrá un valor 0 hasta llegar a ella, en cuyo momento alcanzará el máximo posible para volver a 0 de nuevo. La duración infinitesimal en las nuevas circunstancias es un segmento espacial del tamaño de una celda y la magnitud infinita se convierte ahora en nivel de brillo máximo del sistema.



Función de dispersión puntual en la formación de la imagen.

Si la entrada al sistema formador de imágenes fuese un impulso unitario, en ausencia de elementos perturbadores entre el espacio objeto y el espacio imagen, la imagen formada debería ser también un impulso unitario, es decir, la imagen de una celda de terreno brillante habría de ser un píxel brillante. En la práctica esto no sucede así por varias razones entre las que las más importantes son la presencia de atmósfera, con efecto dispersivo debido a los aerosoles y moléculas que la constituyen, y la óptica del instrumento de adquisición, que puede introducir aberraciones y otros efectos dispersivos. El resultado es que la imagen formada recoge principalmente el píxel brillante, pero también forma un halo a su alrededor de mayor o menor entidad dependiendo del alcance de la dispersión operante. La forma en la que se degrada la señal unitaria original se conoce como Función de Dispersión Puntual (*Point Spread Function*, PSF).

Es necesario en este punto introducir otra operación matemática para comprender cómo se forman realmente las imágenes. Se trata de la integral de convolución. Se define la operación de convolución de dos funciones $f_1(t)$ y $f_2(t)$ como la nueva función $f(t)$:

$$f(t) = f_1(t) \otimes f_2(t) = \int_{-\infty}^{\infty} f_1(x) f_2(t-x) dx \quad (21)$$

que en el caso particular de ser

$$\left. \begin{aligned} f_1(t) &= 0 \\ f_2(t) &= 0 \end{aligned} \right\} \forall t < 0 \quad (22)$$

sería:

$$f(t) = \int_0^t f_1(x) f_2(t-x) dx \quad (23)$$

La operación de convolución tiene las propiedades conmutativa y asociativa. Así pues:

$$f_1(t) \otimes f_2(t) = f_2(t) \otimes f_1(t) \quad (24)$$

y

$$[f_1(t) \otimes f_2(t)] \otimes f_3(t) = f_1(t) \otimes [f_2(t) \otimes f_3(t)] \quad (25)$$

Cuando una de las dos funciones es un impulso unitario, se tiene:

$$f(t) \otimes \delta(t) = \int_{-\infty}^{\infty} f(x) \delta(t-x) dx \quad (26)$$

y por su propiedad conmutativa:

$$f(t) \otimes \delta(t) = \delta(t) \otimes f(t) \quad (27)$$

Echando mano de la propiedad de muestreo del impulso unitario vista en (16):

$$\int_{-\infty}^{\infty} \delta(t) f(t+t_0) dt = f(t_0) \quad (28)$$

se tiene que

$$\delta(t) \otimes f(t) = \int_{-\infty}^{\infty} \delta(x) f(t-x) dx = f(t) \quad (29)$$

y en consecuencia:

$$f(t) \otimes \delta(t) = f(t) \quad (30)$$

Esto quiere decir que la convolución de una señal o función con un impulso unitario devuelve como resultado la misma señal. Se trataría así el impulso como el elemento neutro de la operación de convolución.

Pasemos ahora a las coordenadas espaciales x e y , en lugar de la variable lineal t . Como el espacio objeto es, en general, diferente al espacio imagen, establezcamos para él un sistema de coordenadas ξ, η . Pues bien, si consideramos un sistema de formación de imágenes \mathcal{H} que cumpla las propiedades de linealidad, su respuesta $h(x, y)$ ante la excitación producida por un impulso unitario será:

$$\mathcal{H}[\delta(\xi, \eta)] = h(x, y) \quad (31)$$

Y si el sistema es invariante o de parámetros constantes, cuando la señal de entrada sufra un desplazamiento espacial, la imagen de salida registrará el mismo desplazamiento sin modificación de la función:

$$\mathcal{H}[\delta(\xi - \xi', \eta - \eta')] = h(x - x', y - y') \quad (32)$$

Por la propiedad del impulso se puede expresar cualquier función de entrada $f_c(x, y)$ como la convolución de ella misma con un impulso unitario, en este caso extendida la integral de convolución a las dos dimensiones planimétricas:

$$\begin{aligned} f(\xi, \eta) &= f(\xi, \eta) \otimes \delta(\xi, \eta) = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) \delta(\xi - u, \eta - v) dudv \end{aligned} \quad (33)$$

La función salida o respuesta del sistema formador de imágenes a la señal de entrada es:

$$\begin{aligned} g(x, y) &= \mathcal{H}[f(\xi, \eta)] = \\ &= \mathcal{H}\left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) \delta(\xi - u, \eta - v) dudv\right] \end{aligned} \quad (34)$$

Y si el sistema es lineal e invariante:

$$\begin{aligned} \mathcal{H}\left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) \delta(\xi - u, \eta - v) dudv\right] &= \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) \mathcal{H}[\delta(\xi - u, \eta - v)] dudv &= \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) \cdot h(x - u, y - v) dudv \end{aligned} \quad (35)$$

Esta última expresión es la de la convolución entre la señal de entrada y la función $h(x, y)$, que es la respuesta del sistema ante un impulso unitario, de modo que finalmente se puede escribir:

$$g(x, y) = h(x, y) \otimes f(\xi, \eta) \quad (36)$$

Es decir, que la respuesta de un sistema lineal consiste en la convolución de la entrada con la respuesta del sistema al impulso unitario y por lo tanto, está determinada por esta última. Puede conocerse con exactitud la respuesta del sistema ante una entrada

de gran complejidad sólo conociendo cómo transforma el sistema una entrada unitaria. En la formación de imágenes, la función $h(t)$ es la PSF vista anteriormente y la imagen formada es en realidad la convolución entre la imagen real, la correspondiente al espacio objeto, y la PSF. Cuanto mayor sea la dispersión atmosférica, unida a las distorsiones causadas por la óptica del instrumento, mayor será el efecto emborronador (*blurring*) de la PSF y peor será la nitidez de la imagen.

Se denomina función del sistema a la transformada de Fourier de la respuesta del sistema al impulso unitario:

$$H(\rho, \sigma) = \mathcal{F}[h(x, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) e^{-i(\rho x + \sigma y)} dx dy \quad (37)$$

La función del sistema lo caracteriza completamente y su conocimiento permite predecir la degradación a que se verá sometida la imagen. La transformada de Fourier de una imagen permite definirla en el dominio de la frecuencia en lugar de hacerlo en el dominio

espacial convencional. Una de las ventajas de operar en el dominio de la frecuencia es que, en virtud del teorema de la convolución, la operación de convolución de dos funciones en el dominio espacial equivale al producto de sus respectivas transformadas de Fourier, y multiplicar imágenes, aunque sean éstas en el dominio de la frecuencia es una operación más simple y permite mayor flexibilidad. Todo ello se verá en mayor detalle en el apartado dedicado a la imagen en el dominio de la frecuencia, dentro de la lección de tratamiento digital de imágenes.

Adicionalmente, si se conoce la PSF o se tienen medios para estimarla, operando en el dominio de la frecuencia, aunque también en el espacial, es posible aplicar funciones capaces de introducir el efecto opuesto al inducido por aquella en la imagen. De esta manera se pueden restaurar imágenes degradadas por la PSF para obtener otras más parecidas a la realidad aplicando funciones restauradoras. Estos procedimientos también se conocen genéricamente como deconvolución, haciendo referencia al efecto contrario al que causa la PSF sobre la imagen, y se verán en el capítulo 10.

Grabación

Los satélites adquieren líneas completas de imagen sucesivamente y sin solución de continuidad mientras se desplazan en su órbita alrededor de la Tierra. La escena formada por un cierto paquete de líneas se forma en el segmento terrestre de la misión para comercializar la información. Las celdas que constituyen cada línea de imagen pueden ser adquiridas simultáneamente o mediante barrido secuencial, según sea el diseño del sensor. Incluso las líneas pueden ser

adquiridas una a una o bien en bloques de varias de ellas.

El instrumento está dotado de una óptica y unos sistemas electrónicos que permiten dirigir y registrar la señal recibida en su campo de visión en distintos planos focales dedicados a cada longitud de onda. En cada uno de ellos, el dispositivo se encarga de traducir la irradiancia espectral en niveles digitales de cada celda para configurar la imagen digital. Finalmente,

esa imagen se graba a bordo del satélite y se transmite a la estación de recepción terrestre, bien directamente, bien a través de constelaciones de satélites repetidores.

Ruido

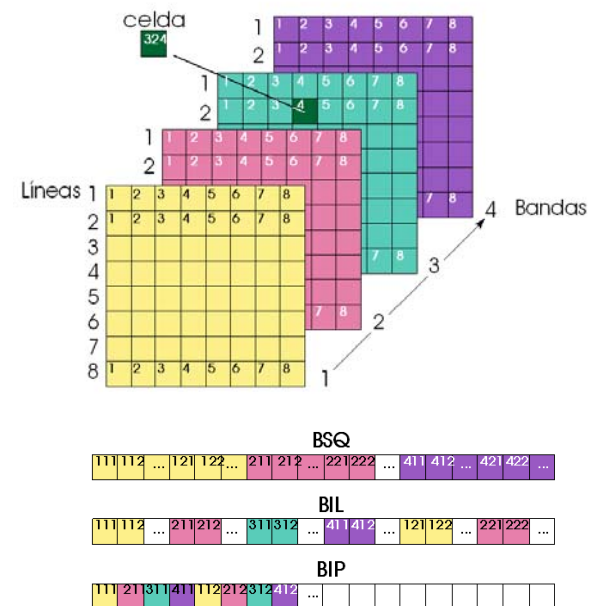
Se denomina ruido electrónico a la energía electromagnética que aparece en cualquier circuito electrónico, además de la señal útil a procesar. La calidad de la imagen adquirida con cualquier sistema formador de imágenes está en general afectada por un determinado nivel de ruido, cuya existencia viene a perturbar el procesamiento e interpretación de la señal original. Por extensión, se considera ruido a toda información espuria añadida por cualquier vía a la información verdaderamente útil que procede de la superficie observada y que, en su caso, generaría una reproducción fidedigna en el espacio imagen de la realidad del espacio objeto. Lo cierto es que durante el proceso de registro de la imagen se introduce inevitablemente ruido como un factor aditivo, y por esta razón el sistema en su conjunto no puede considerarse exactamente lineal.

$$g(x, y) = Grab[f(\xi, \eta)] = h(x, y) \otimes f(\xi, \eta) + n(\xi, \eta) \quad (38)$$

Mediante operadores de deconvolución se puede restaurar la imagen, si bien no completamente, ya que el término de ruido es aditivo y no está dentro de la función de convolución degradante de la imagen. Al aplicar operadores de restauración es inevitable perder parte de la señal útil mientras se busca eliminar solo el ruido. Por lo tanto, en la restauración de una imagen se buscará mantener un compromiso entre señal y ruido acorde con las exigencias del trabajo.

Formatos

La imagen multiespectral, como se ha visto, consiste en un pila de imágenes monocanal. En los sistemas de almacenamiento, discos, cintas, memorias USB, etc. la imagen es un archivo de carácter lineal donde, si una celda se almacena en un byte, el registro de la imagen completa consiste en un número determinado de bytes, idealmente contiguos, igual al número de celdas contenido en el conjunto de todas las bandas. Existen tres formatos de grabación para este tipo de imágenes: bandas secuenciales (BSQ), bandas intercaladas por línea (BIL) y bandas intercaladas por píxel (BIP). En el primero de ellos se escribe la imagen registrando todas las celdas de una en una por líneas consecutivas y una banda tras otra. Cuando el almacenamiento se realizaba en los inicios del tratamiento digital sobre cintas magnéticas, estudiar un determinado píxel multiespectral requería la búsqueda del byte en cuestión a lo largo de toda la cinta, ya que una banda no era grabada hasta no haber terminado completamente la anterior. Lógicamente el tiempo de proceso se dilataba considerablemente. Para agilizar estos procesos se recurrió al tercero de los sistemas (BIP) y comenzaron a almacenarse los niveles digitales de cada celda en todas las bandas de la imagen en las posiciones sucesivas del archivo informático. Concluidos los registros de los n valores de la celda en las n bandas, se pasaba a la siguiente celda, y así sucesivamente. Si con este procedimiento, el acceso a un píxel concreto era más rápido, en cambio la lectura de una banda retrasaba notablemente el proceso. Una solución de compromiso consistió en grabar una línea completa de una banda, pasar a continuación a grabar la misma línea de la segunda banda, y así sucesivamente hasta completar todas las bandas. Terminado el registro de la primera línea multiespectral se pasaba a la segunda línea y se continuaba el proceso hasta terminar todas las líneas de la imagen (BIL). En la actualidad coexisten los tres



Formatos de grabación de la imagen digital.

sistemas y los paquetes de tratamiento digital de imágenes permiten cómodamente la codificación de cualquiera de ellos y la traducción de uno a otro.

Calibración

En determinados estudios basados en la teledetección no basta con medir las diferencias en la respuesta espectral de las diferentes zonas de la imagen, sino que es necesario conocer de modo absoluto el comportamiento físico de la cubierta.

Para ello es necesario encontrar la función genérica de transformación que existe entre la variable física objeto de estudio y el ND asignado finalmente por el sistema de adquisición. Una vez conocida esa función, la aplicación de su inversa a los ND de la imagen permite recuperar la variable física y poder así caracterizar la cubierta estudiada

Calibrar una imagen consiste en transformar sus niveles digitales en variables que tengan un significado físico como son la radiancia, la reflectancia, la temperatura de la superficie, etc.

La relación entre ND y la magnitud física habitualmente implica dos etapas:

1. Convertir los ND en valores de radiancia espectral y, tras ello, en reflectancia espectral de la cubierta: La relación entre unos y otros está gobernada por funciones matemáticas simples, normalmente lineales, las llamadas funciones de calibración.
2. Establecer una relación entre la reflectancia hallada y la magnitud física buscada. Este paso se

basa en dependencias estadísticas, según la correlación que exista entre aquélla y ésta

Función de conversión

Los sensores integrados en el instrumento de adquisición convierten la irradiancia recibida en una señal eléctrica variable. El voltaje producido es proporcional a la energía incidente.

Dado que se conocen los parámetros orbitales del satélite, y con ello su situación con respecto a la superficie observada, particularmente su distancia y el ángulo de observación, es inmediato deducir la radiancia que procede del terreno a partir de la irradiancia sobre el plano focal del instrumento. En consecuencia, se puede decir que el sensor convierte la radiancia en una señal analógica, materializada por una corriente eléctrica.

Otro elemento constitutivo del instrumento es un convertidor analógico-digital que transforma esa corriente en niveles digitales discretos, en función del voltaje de la primera.

Existen diversos tipos de sensores, cuya descripción queda fuera del alcance de este manual. Los más importantes son de tipo CCD (*Charge-Coupled Device*, dispositivo de carga acoplada) y los de tipo CMOS (*Complementary Metal-Oxide Semiconductor*, semiconductor complementario de óxido metálico).

Basándose ambos en el efecto fotoeléctrico, el detector contiene tantos fotodiodos o *fotosites* como píxeles, los cuales producen una corriente eléctrica proporcional al número de fotones incidentes y, en definitiva, a la intensidad de luz recibida.

En los sensores CCD, esa corriente se transfiere a un circuito integrado que contiene una serie de condensadores acoplados, cada uno de los cuales se carga en función de la corriente recibida. En electrónica independiente, la señal se amplifica, y en el convertidor analógico-digital la carga inicial es convertida en niveles digitales que son finalmente almacenados en forma de imagen digital.

Los sensores CMOS son en realidad sensores de píxeles activos basados en la tecnología CMOS. A diferencia del CCD incorpora un amplificador de la señal eléctrica en cada fotodiodo y suele incluir el convertidor analógico-digital en el propio chip, prescindiendo así de parte de la electrónica externa.

Los sensores, con independencia de la familia a la que pertenezcan, transforman en una señal de salida $Q(V)$, materializada por un voltaje V , la radiancia espectral de la superficie L_λ , medida en $W \cdot m^{-2} \cdot sr^{-1}$. Esta correspondencia es prácticamente lineal en el intervalo espectral de cada una de las bandas exploradas por el sensor.

$$Q = k \int_{d\lambda} L_\lambda r_\lambda d\lambda + O \quad (1)$$

Donde k es un factor de proporcionalidad, r_λ es la respuesta espectral del instrumento y O es una constante aditiva. De forma simplificada puede escribirse:

$$Q = G_1 L_\lambda + O_1 \quad (2)$$

de donde se deduce que: $L_\lambda = \frac{Q - O_1}{G_1}$

Es decir:

$$Q = f_1(L_\lambda); \quad L_\lambda = f_1^{-1}(Q) \quad (3)$$

El voltaje es función de la radiancia y viceversa, y, como puede verse, la relación entre la señal de salida y la radiancia tiene la forma de una recta.

De otra parte, la relación entre la radiancia y el nivel digital es también lineal:

$$ND = G_2 L_\lambda + O_2 \quad (4)$$

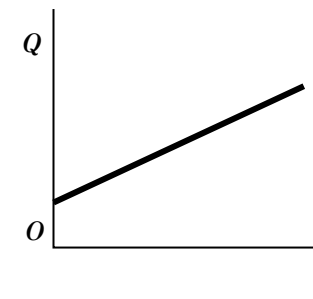
y

$$L_\lambda = \frac{ND - O_2}{G_2}$$

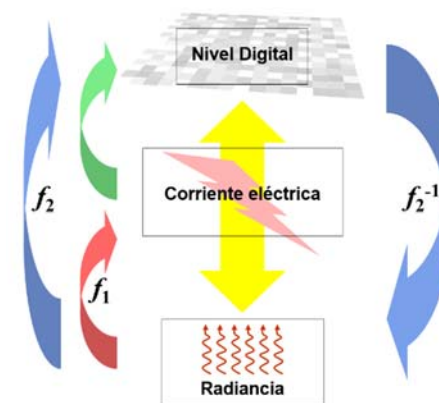
por lo que:

$$ND = f_2(L_\lambda); \quad L_\lambda = f_2^{-1}(ND) \quad (5)$$

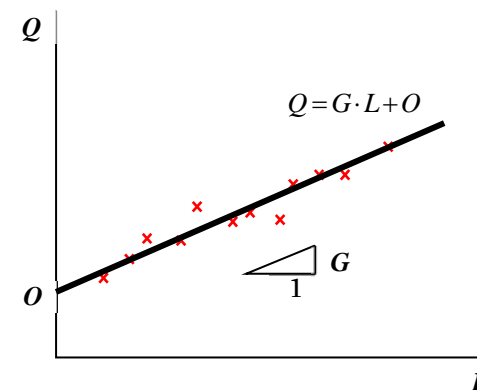
Es decir, también puede ponerse en relación directamente mediante funciones de calibración la radiancia del terreno y el nivel digital de cada una de las celdas de una imagen. En la práctica, los sensores se calibran exponiéndolos, al menos, a dos niveles conocidos de radiancia y calculando los coeficientes de ganancia G (*gain*) y sesgo O (*offset*). Lo habitual es deducir la función de calibración por ajuste a la recta de regresión del conjunto de numerosas lecturas. Estas operaciones se realizan en tierra durante la fase de prelanzamiento del satélite, antes de su



Linealidad del voltaje con respecto a la radiancia.



Funciones de conversión de radiancia en nivel digital.



Ajuste por regresión de los parámetros de calibración.



Distintos modelos de espectrorradiómetros.

puesta en órbita. Pero una vez orbitando, vuelve a recalibrarse periódicamente para ajustar los parámetros de ganancia y sesgo, ya que por el envejecimiento de los materiales suelen variar, aunque mínimamente. Esta recalibración se lleva a cabo realizando medidas simultáneas de radiancia desde el satélite y desde tierra, mediante espectrorradiómetros, sobre blancos conocidos.

Radiometría de campo

La radiometría consiste en medir la radiación emitida o reflejada desde una determinada superficie objeto de estudio, mediante espectrorradiómetros, sometién-dola a distintas condiciones de observación. Las campañas de radiometría son útiles y necesarias para caracterizar desde el punto de vista radiométrico la verdad-terreno (*ground-truth*) con distintos objetivos como son la calibración de los instrumentos de adquisición, la validación de los datos adquiridos por los satélites y la interpretación de los resultados.

El radiómetro es un instrumento destinado a realizar medidas en campo o en laboratorio, de lo cual dependerá su tamaño y complejidad, capaz de medir la radiancia espectral en un cierto intervalo de longitudes de onda. Suele ser un instrumento hiperespectral, esto es, realiza medidas de radiancia en intervalos muy estrechos y contiguos de longitudes de onda, frecuentemente de entre 1 y 5 nm.

Mediante la radiometría de campo es habitual realizar medidas bidireccionales, con el fin de caracterizar el hábito reflexivo de la superficie, su eventual carácter lambertiano, deducir la función de distribución de reflectancia bidireccional, etc. Para ello se mueve el radiómetro haciéndolo girar dentro de un plano vertical, que a su vez también puede girar alrededor del eje normal a la superficie. De esta manera, manteniéndose la fuente de iluminación fija, el sol, se mide

la radiancia en todas las direcciones del hemisferio superior a la superficie.

También es frecuente realizar medidas manteniendo fijo el espectrorradiómetro y dejando que el sol realice su órbita diurna, a fin de comprobar la respuesta de la superficie bajo distintas condiciones de azimut y elevación solar.

Calibración de sensores orbitales

Genéricamente, la función de calibración de un sensor formador de imágenes situado en órbita tiene la forma:

$$L_{sen}(k) = a_0(k) + a_1(k)ND(k) \quad (6)$$

en cuya expresión L_{sen} es la radiancia medida por el sensor, $a_0(k)$ y $a_1(k)$ son los coeficientes de calibración del sensor y $ND(k)$ es el nivel digital, todo ello referido a una banda genérica k de la imagen.

En el espectro óptico, la emisión propia de la superficie terrestre es despreciable y, por ello, se considera que la radiación procedente de la tierra en ese intervalo espectral se debe a la reflexión. Los valores de radiación registrados dependen entonces de cuatro factores:

- La irradiancia solar, que depende a su vez de tanto de la altura solar como de la distancia Tierra-Sol en el momento de la adquisición de la imagen.
- La reflectancia de la superficie, que es un parámetro invariable, siempre que la naturaleza o la cobertura de la superficie se mantenga en el tiempo.
- La transmitancia atmosférica en las longitudes de onda consideradas, que depende de la

composición química de los gases atmosféricos y del contenido en aerosoles.

- La radiancia difusa atmosférica, que depende de los aerosoles en suspensión.

Puede decirse que, en su tránsito a través de la atmósfera, parte de la energía solar incidente sobre el borde exterior de la atmósfera se reduce, debido a fenómenos de absorción, hasta llegar a la superficie, donde es reflejada en virtud de la reflectancia inherente al terreno y redirigida en numerosas direcciones, entre ellas también la del sensor. De nuevo, al atravesar la atmósfera, parte de la energía es absorbida, pero el fenómeno se ve compensado parcialmente por la adición de la radiancia difusa atmosférica que ensucia la señal y termina penetrando en el campo de visión del instrumento.

La radiancia medida por el sensor es:

$$L(k) = L_{sen}(k) - L_{atm}(k) = \rho(k)L_{sue}(k)\tau(k) \quad (7)$$

donde $L_{sen}(k)$ es la radiancia espectral medida en el sensor, $L_{sue}(k)$ es la radiancia espectral solar recibida sobre el suelo, $\rho(k)$ es la reflectancia espectral de la cubierta, $\tau(k)$ es la transmitancia espectral de la atmósfera y $L_{atm}(k)$ es la radiancia espectral aportada por la atmósfera.

Despejando:

$$\rho(k) = \frac{L_{sen}(k) - L_{atm}(k)}{L_{sue}(k)\tau(k)} \quad (8)$$

Si se considera la superficie como si fuera lambertiana, para un determinado día del año, la radiancia incidente sobre el suelo es:

$$L_{sue}(k) = \frac{E_0(k)\cos\theta}{K\pi} \quad (9)$$

En donde $E_0(k)$ es la irradiancia espectral solar extraterrestre en la banda k , θ es el ángulo cenital solar y K es un factor corrector de la distancia Tierra-Sol que puede obtenerse de un anuario astronómico. Combinando las dos expresiones anteriores:

$$\rho(k) = K\pi \frac{L_{sen}(k) - L_{atm}(k)}{E_0(k)\tau(k)\cos\theta} \quad (10)$$

Si se corrige el efecto atmosférico, $L_{atm}(k)=0$ y $t(k)=1$, por lo que:

$$\rho(k) = K\pi \frac{L_{sen}(k)}{E_0(k)\cos\theta} \quad (11)$$

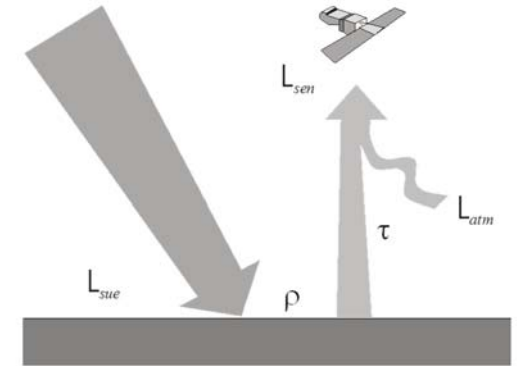
Y considerando (6):

$$\rho(k) = K\pi \frac{a_0(k) + a_1(k)ND(k)}{E_0(k)\cos\theta} \quad (12)$$

expresión que puede ponerse de la forma:

$$\rho(k) = b_0(k) + b_1(k)ND(k) \quad (13)$$

y que indica que, de forma análoga a la radiancia, también la reflectancia muestra una relación lineal con el nivel digital de cada celda en la imagen, a través, en este caso, de los coeficientes de calibración $b_0(k)$ y $b_1(k)$ de cada banda. Al igual que los coeficientes de calibración de radiancia, los de reflectancia son proporcionados por el constructor del sensor y se actualizan periódicamente.



Transferencia radiativa a través de la atmósfera.

7

Sensores y plataformas

Introducción

La adquisición de la información en forma de imagen digital requiere de sistemas electrónicos u óptico-electrónicos capaces de convertir la irradiancia incidente en el instrumento en una señal susceptible de ser guardada o transmitida. A estos sistemas toman el nombre de sensores o instrumentos de adquisición y constituyen la carga útil del satélite. Por su parte, en este último cabe distinguir la plataforma portadora propiamente dicha, que integra los sistemas de suministro de energía, comunicaciones, telemetría, estabilización, etc. y la carga útil, que es el

instrumento o grupo de instrumentos de adquisición de información.

El objeto de este capítulo es proporcionar una noción de los requisitos del sensor, de sus distintos tipos de resolución, de los tipos de sensores, de la forma en que éstos adquieren las imágenes y, finalmente, proporcionar muy sucintamente una panorámica de algunos de los programas de teledetección de propósito general más interesantes.

Resolución

Para abordar el estudio de los tipos de sensores es necesario introducir previamente el concepto de resolución, que en el caso de las plataformas orbitales adquiere dimensiones propias, más allá de la puramente espacial, que es a la que se está más acostumbrado.

La agencia estadounidense para la administración oceánica y atmosférica, NOAA define la resolución como el valor mínimo determinado para alguna de las variables que definen a una imagen digital. Puede decirse que es la capacidad que ha tenido el sistema formador de imágenes para discriminar información de detalle en un objeto. En los instrumentos ópticos tradicionales la resolución se identifica con el poder de separación espacial del sistema de lentes y el tamaño de grano de la emulsión fotográfica. Una imagen fotográfica tiene tanta más resolución cuanto menores son los objetos que es capaz de discriminar.

Sin embargo, al aparecer los sensores orbitales, a las variables planimétricas se añaden otras nuevas, como son la altimetría, la espectralidad y la temporalidad.

Cuando se diseña un sensor que va a ser puesto en órbita, el constructor atiende las demandas, planteadas a menudo por la comunidad de usuarios, en función del objeto de estudio. Puede ser interesante explorar con un alto grado de detalle pequeñas superficies, como es el caso de un satélite espía, o bien por el contrario puede buscarse abarcar grandes zonas geográficas a nivel regional o continental, aunque no sea tan importante el detalle. Por otro lado, se puede estar interesado en identificar minuciosamente la signatura espectral de las superficies observadas o bien la vocación del sistema se vuelca sobre el estudio e identificación de las formas geométricas, sin ser tan trascendente conocer la respuesta de la superficie en las distintas zonas del espectro. Si el sensor ha de

poder identificar y registrar ligeras diferencias de radiancia entre diferentes zonas, ello dará lugar a un diseño diferente al que habría adoptado el constructor si solamente se hubiera conformado con apreciar unos cuantos niveles de gris. Finalmente, dependiendo de cuál vaya a ser el uso del sensor, puede ser algo esencial que el satélite sea capaz de visitar un determinado punto de la superficie terrestre cada pocos minutos, horas, días o semanas. Con estas exigencias se da lugar a los cuatro tipos de resolución que se atribuyen a los sensores puestos en órbita: espacial, espectral, radiométrica y temporal.

Resolución espacial

La resolución espacial coincide con la formulación tradicional de resolución, tal como se aplica a otros sistemas analógicos. Es una resolución de tipo puramente geométrica y viene determinada por el tamaño de la celda sobre el terreno. Depende de dos factores: el campo de visión instantáneo y la altura de la plataforma.

El campo de visión instantáneo (*Instantaneous Field of View*, IFOV) es el ángulo sólido, la radiancia contenida en el cual se graba en una única celda de imagen. El campo de visión instantáneo en tierra (*Ground Instantaneous Field of View*, GIFOV) es el área de terreno que subtiende un ángulo sólido igual al IFOV del instrumento y equivale a la distancia de muestreo en tierra (*Ground Sample Distance*, GSD) y determina la resolución espacial. El IFOV de un instrumento es algo fijo que responde a un requerimiento de diseño determinado del sistema óptico electrónico. En función de cuál sea la altura en la que orbite el satélite, el GIFOV será mayor o menor, dando lugar a celdas de mayores proporciones o celdas más pequeñas, es decir, el sensor tendrá menor

o mayor resolución espacial, respectivamente. Se conoce como campo de visión (*Field of View*, FOV) al ángulo abarcado por una línea completa de imagen.

La resolución espacial tiene trascendencia no solamente en cuanto a la definición de las formas en la imagen, sino también en la identificación del terreno observado según su nivel de radiancia. Asumiendo un modelo aditivo de radiancia, en el que cada punto de la superficie contribuye a la radiancia total de acuerdo a su proporción dentro de un área determinada, cuando el terreno no sea uniforme y la celda sea muy grande, la radiancia que alcanza el sensor será la mezcla de la que emita cada pequeña zona homogénea de la celda. Solamente se obtendrá un nivel digital promedio como respuesta de esa celda. Sin embargo, a medida que la resolución aumente, la gran celda contendrá a su vez cuatro, nueve, dieciséis o más celdas pequeñas, cada una de las cuales reflejará con mayor fidelidad la respuesta de esa superficie abigarrada.

Resolución espectral

La resolución espectral es la capacidad del sensor para discriminar la radiancia en distintas longitudes de onda del espectro electromagnético. Cada una de las bandas o canales de un sensor multiespectral exploran la respuesta de la superficie en un determinado intervalo de longitudes de onda. En realidad, la resolución espectral viene determinada por el número de canales de la imagen y por la anchura de banda de cada canal.

La identificación de las superficies bajo observación, haciendo ahora caso omiso de las formas, se basa en la reconstrucción de su signatura espectral y en una correcta caracterización de la curva correspondiente.

Cuanto más canales ofrezca el sensor, más componentes tendrá el vector de características de cada celda y, con ello, mejor se podrán delinear los rasgos de la signatura espectral de la superficie. Téngase en cuenta que, si el barrido espectral no es tan minucioso, el ND de la celda en un determinado canal de gran anchura de banda es el resultante de integrar los niveles de radiancia espectral extendida a todo el intervalo espectral explorado en ese canal. Evidentemente, la identificación de la superficie se hace más difícil en estas condiciones.

Una situación ideal es disponer de gran cantidad de bandas en la imagen. La teledetección hiperespectral maneja imágenes de decenas, e incluso de centenares de bandas. Bajo esta modalidad, en general, no sólo se requieren un gran número de canales, sino algunas condiciones adicionales, como son una alta resolución espectral, es decir, bandas muy estrechas, de uno o pocos nanómetros, y contigüidad de las bandas en sus longitudes de onda. En algunos textos, especialmente los relacionados con la astronomía y la astrofísica, la técnica se denomina espectrometría o espectroscopia de imágenes. Las técnicas de barrido espectral, aunque sin formación de imágenes, son también muy familiares en química.

Una de las principales ventajas de la teledetección hiperespectral es que permite construir el perfil espectral detallado de las superficies observadas. Con la signatura espectral se pueden identificar ciertos compuestos directamente, como es el caso de los minerales en los suelos. Sin embargo, la identificación es bastante difícil con el uso de imágenes multiespectrales convencionales. Los perfiles espectrales obtenidos a partir de imágenes de satélite también permiten reemplazar las pruebas de laboratorio o las campañas de espectralradiometría de campo.

Sin embargo, a pesar de las ventajas que supone el

empleo de imágenes hiperespectrales, es necesario tener en cuenta que los sensores sobremuestran el espectro para evitar perder rasgos de absorción de banda estrecha. Esto ocasiona que la información manejada sea de un gran peso, especialmente si el estudio abarca superficies medianas o grandes. Por ello en ciertas ocasiones, para hacer posible el tratamiento digital de las imágenes es necesario reducir el número de dimensiones del vector de características de las celdas para mejorar la eficiencia en tiempo y precisión de los algoritmos. Entre estos procedimientos se encuentran la transformación a componentes principales (*Principal Components Analysis*, PCA) y la transformación de la fracción de mínimo ruido (*Minimum Noise Fraction Transform*, MNF).

Resolución radiométrica

La resolución radiométrica es la capacidad del sensor para discriminar niveles de intensidad de radiancia espectral dentro de un intervalo dado, esto es, en un determinado canal. En los sistemas analógicos como los fotográficos, la resolución radiométrica viene determinada por el número de niveles de gris, que también se conoce como profundidad de color. En los sistemas óptico-electrónicos, a cada celda se le asigna un nivel digital (ND) que guarda una relación directa con la cantidad de energía recibida.

La imagen digital almacena unos niveles digitales que pueden exhibir un valor máximo igual a $2^k - 1$, siendo k el número de bits con el que codifica dichos ND. En consecuencia, dependiendo de la resolución radiométrica de la imagen, el rango dinámico de niveles de gris puede ser de 2, 4, 8, 16, 128, 256, 512, 1024, 2048, 4096, 8192, etc., o sea, los niveles digitales pueden estar comprendidos respectivamente entre 0 y 1, 3, 7, 15, 31, 63, 127, 255, 511, 1023, 2047, 4095,

8191, etc. Las imágenes más habituales hasta hace relativamente poco tiempo eran de 8 bits y, con ello, su rango dinámico alcanzaba 256 niveles de gris, que oscilaban entre 0 y $2^8-1=255$. Con este margen era posible almacenar cada celda de imagen en un byte, que es la unidad de tratamiento informático de ficheros, con la cual se pueden codificar 256 valores diferentes con 8 posiciones binarias. Márgenes dinámicos superiores exigen reservar al menos 2 bytes para codificar los ND de cada celda.

La irrupción de sensores de muy alta resolución espacial en el panorama de la teledetección ha permitido el desarrollo de la cartografía urbana a partir de imágenes de satélite. Pero hay que tener en cuenta que, en estos ámbitos, las sombras ocupan una gran proporción del campo de la imagen. Distinguir con eficacia objetos mal iluminados requiere dotar de amplio margen dinámico las zonas sombreadas. Los sensores de 8 bits, en los cuales las sombras ocupan tan solo unos cuantos niveles digitales cercanos al 0 se mostraban ineficaces para este cometido. Por esta razón fue necesario el perfeccionamiento de los sistemas de adquisición dotándolos de mayor margen dinámico. Hoy día son habituales las imágenes de satélite de 10, 11 y 12 bits, con las cuales, la discriminación de objetos dentro de las áreas sombreadas no reviste problema.

En la presentación monocroma de una imagen impresa o en pantalla, los ND almacenados numéricamente son traducidos a niveles visuales de gris (NV) de forma lineal, correspondiendo el valor 0 al negro, el valor 2^k-1 al blanco y los intermedios, a un nivel de gris proporcional.

Resolución temporal

La resolución temporal es la capacidad del sistema

Carlos Pinilla

para discriminar los cambios sufridos en el tiempo por la superficie estudiada y hace referencia a la periodicidad con que el sensor, gracias a la plataforma portadora y a su propio diseño, puede adquirir una nueva imagen del mismo punto de la superficie terrestre o, dicho de otra forma, es el tiempo necesario para que el sensor revise un determinado punto de la superficie. Para una determinada latitud, la periodicidad de paso por la vertical de un lugar solamente depende de dos factores: de la altura de la órbita y del campo de visión del instrumento. La anchura de la franja de superficie barrida por el campo de visión del sensor se conoce como franja explorada o *swath*.

La relación entre la periodicidad de revisita y la altura orbital del satélite se fundamenta en las leyes de Johannes Kepler, quien en 1609 enunció sus tres leyes a partir de las observaciones realizadas por Tycho Brahe, con las que se rige el sistema solar, y a partir de las cuales Newton desarrolló la ley de la gravitación universal. En su obra *Astronomia Nova* estableció que:

1. Todos los planetas describen órbitas elípticas directas, uno de cuyos focos es el Sol.
2. La velocidad areolar de un planeta es constante o, lo que es lo mismo, las áreas barridas por los radios de los planetas son proporcionales al tiempo empleado por estos en recorrer el perímetro de dichas áreas.
3. El cuadrado del período de un planeta es proporcional al cubo del semieje mayor de su órbita, es decir, $T^2 = ka^3$.

La estabilidad de la órbita del satélite requiere que la suma de las dos fuerzas que actúan sobre el satélite sea nula, por un lado, la fuerza gravitatoria ejercida por la Tierra y, por otro, la fuerza centrífuga originada por el movimiento circular:

$$\left. \begin{aligned} F &= K \frac{Mm}{r^2} \\ f &= m \frac{v^2}{r} = m\omega^2 r \end{aligned} \right\} \quad (1)$$

donde:

K es la constante de la gravitación universal, M es la masa de la Tierra, m es la masa del satélite, v y ω son las velocidades lineal y angular del satélite, respectivamente (medido el ángulo en radianes) y r el radio de la órbita.

Igualando:

$$K \frac{Mm}{r^2} = m\omega^2 r \quad (2)$$

Es decir:

$$r^3 \omega^2 = KM \quad (3)$$

Y como $\omega = \frac{2\pi}{T}$, resulta:

$$r^3 = \frac{KM}{4\pi^2} T^2 \quad (4)$$

o sea, $T^2 = \frac{4\pi^2}{KM} r^3$, que es una condición que cumple la tercera ley de Kepler, al ser una forma de escribir $T^2 = ka^3$.

Existen dos tipos de órbitas en las cuales se aparcan la mayoría de los satélites de observación de la Tierra,

la denominada geostacionaria y las órbitas polares. La primera de ellas es una órbita ecuatorial directa, esto es, en la que el plano orbital contiene al ecuador terrestre y el satélite se mueve de oeste a este. La cualidad de los satélites geostacionarios, además de cumplir esas dos condiciones, es que su órbita es geosíncrona, o sea su velocidad angular coincide con la de la Tierra, o sea, su período de revolución es de 23 horas, 56 minutos y 4,09 segundos. Por ello se encuentra siempre sobre la vertical de cualquier punto del ecuador en el que se aparque. Para que se cumpla esta última condición, suponiendo la órbita completamente circular:

$$\begin{aligned} r(\text{km}) &= \sqrt[3]{\frac{KM}{4\pi^2} T^2 \cdot 10^{-3}} = \\ &= \sqrt[3]{\frac{6,6738431 \cdot 10^{-11} \cdot 5,972 \cdot 10^{24}}{4\pi^2} \cdot 86164,09^2 \cdot 10^{-3}} = \\ &= 42164 \text{ km} \end{aligned}$$

Quiere esto decir que si se desea que un satélite sea geostacionario debe colocarse necesariamente a una altura de $42164 - 6378 = 35786$ km sobre la superficie, siendo el sustraendo el valor del radio de la Tierra en el ecuador.

Dentro del otro grupo de órbitas se engloban tanto las órbitas absolutamente polares como las denominadas cuasipolares. La característica de ellas es que el plano orbital incluye el eje terrestre, o forma un ángulo pequeño con él. Con este tipo de circulación, el satélite, que se mantiene siempre dentro de su plano orbital, va barriendo diferentes partes del planeta en tanto éste sigue su rotación. El hecho de inclinar unos grados la órbita con respecto al eje de la Tierra tiene que ver con la distribución irregular de masas en el interior del planeta, en el que, debido a

su ligero achatamiento en los polos, distribuye un exceso de masa cerca del ecuador. Si se aprovecha bien esta asimetría, es posible conseguir que el plano orbital gire sobre sí mismo con un período de rotación de 365 días, y de esta manera queda orientado al sol bajo el mismo ángulo durante todo el año. Para una determinada latitud, el satélite pasará siempre a la misma hora solar local. Se consigue así una órbita heliosincrónica, para lo cual el plano orbital está en precesión —en giro— aproximadamente un grado al día hacia el este. La velocidad de la precesión depende tanto de la inclinación de la órbita como de la altitud del satélite. Fijada la segunda por necesidades del diseño y objetivo del éste, es posible determinar la primera para obtener la precesión requerida. Órbitas típicas heliosincrónicas de satélites de observación de la Tierra están entre 600 y 800 km de altitud. Para esa altura orbital, los períodos vienen a ser de entre 96 y 100 minutos. Para estas órbitas, se consigue el sincronismo con el Sol inclinándolas alrededor de 98° , esto es, haciéndolas ligeramente retrógradas.

Efemérides astronómicas del satélite

Se denomina pericentro al punto de la órbita del satélite más próximo al astro. En el caso de satélites artificiales se habla de perigeo. En contraposición, el apogeo es el punto de la órbita del satélite más alejado de la Tierra.

Traza es la proyección central de la órbita del satélite sobre la superficie terrestre. Si la órbita es ecuatorial, la traza coincide con el Ecuador, y podría reducirse a un solo punto si el satélite es geoestacionario. Si, por el contrario, es polar, el plano orbital contiene a un meridiano, aunque la traza no puede coincidir con el propio meridiano debido a la rotación terrestre. En

su lugar describe una curva característica que queda recogida con formas diferentes según la proyección elegida para la representación. Se llama traza descendente la mitad de la traza de una órbita no ecuatorial cuyas latitudes son decrecientes en el hemisferio norte y crecientes en el sur. La traza ascendente de un satélite es la que se progresa desde el hemisferio sur hacia el hemisferio norte.

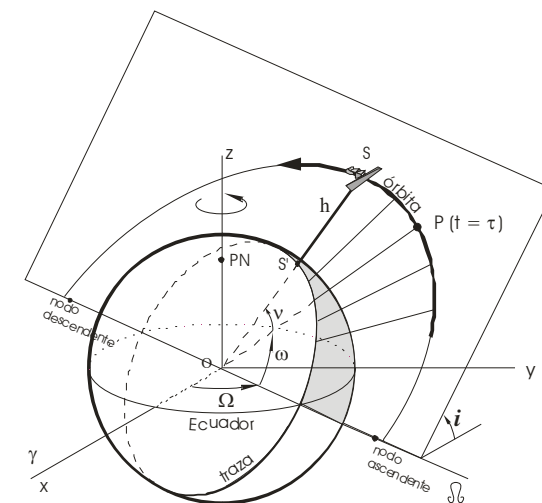
Se denominan nodos ascendente y descendente, respectivamente, a las intersecciones de las ramas ascendente y descendente de la órbita con el plano ecuatorial. La ascensión recta del nodo ascendente Ω es el ángulo medido en el plano ecuatorial entre el equinoccio vernal, —determinado por el punto Aries, intersección del plano ecuatorial terrestre con el de la eclíptica en primavera— y el punto en el que el satélite cruza el plano ecuatorial, de sur a norte.

El argumento del perigeo ω es el ángulo medido en el plano orbital entre el nodo ascendente y el perigeo de la órbita, medido en la dirección del movimiento.

Por su parte, la inclinación de la órbita i es el ángulo entre el plano orbital y ecuatorial. Es el menor de los ángulos formados por el eje de rotación terrestre y el vector momento cinético del satélite. En función de su inclinación, las órbitas se clasifican en:

Órbita	Inclinación
ecuatorial directa	$i = 0^\circ$
directa	$0 < i < 90^\circ$
polar	$i = 90^\circ$
retrógrada	$90^\circ < i < 180^\circ$
ecuatorial retrógrada	$i = 180^\circ$

La mayoría de los satélites de observación de la Tierra mantienen una órbita cuasi-polar de entre 600 y



Efemérides de un satélite.

800 km de altitud, con una inclinación de alrededor de 98° , es decir, siguen órbitas ligeramente retrógradas. El motivo de esto es conseguir una precesión del plano orbital de algo más de un grado sexagesimal por día, causada por la esfericidad imperfecta de la Tierra. El lento giro permite que dicho plano mantenga su orientación con respecto al sol durante todo el año. Este tipo de órbitas se denominan heliosincrónicas y hacen que el satélite cruce el plano ecuatorial todos los días a la misma hora, de modo que las imágenes conservan una iluminación similar en cualquier fecha.

La excentricidad e define la forma de la órbita, siendo

$$e^2 = \frac{a^2 - b^2}{a^2} \text{ donde } b \text{ es semieje menor de la elipse.}$$

Un satélite destinado a la teledetección requiere mantener una órbita lo más circular posible. Por ello el control terrestre de la misión necesita un tiempo desde su lanzamiento hasta dejarlo en estado operativo, estabilizando su órbita mediante la reducción paulatina de su excentricidad hasta valores no significativos.

Finalmente, el movimiento del satélite queda determinado con exactitud añadiendo un último parámetro, que es el instante de paso por el perigeo τ , que como su propio nombre indica es el instante medido en la hora local del momento en el que el satélite, moviéndose en su órbita, alcanza el perigeo.

Relación entre las distintas resoluciones

Existe un cierto antagonismo entre algunos tipos de resolución. En unas ocasiones, la rivalidad deriva de las condiciones físicas del diseño, en otras se trata de

una competencia definida en términos económicos. En general, puede ser difícil, costoso o virtualmente imposible armonizar los cuatro tipos de resolución.

Ciertamente, sería deseable poder diseñar un sistema dotado de la máxima resolución en las cuatro variedades que hemos citado. Sin embargo, unos tipos de resolución están enfrentados a otros. Por ejemplo, para tener una mayor resolución espacial, una forma de adquirirla podría ser acercar la órbita a la superficie, pero esto conduciría a una pérdida de resolución temporal que podría ser importante. Otra forma sería reducir el IFOV del instrumento, pero esta vía choca con aspectos técnicos difíciles de resolver. Disfrutar de una gran resolución espacial puede ser ventajoso para cierto tipo de aprovechamientos de la imagen, pero, sin embargo, para la observación y medición de variables geofísicas sobre superficies homogéneas este aspecto carece de interés. Por ejemplo, no es necesario tener datos de temperatura de superficie marina cada metro cuadrado, puesto que apenas se registrarán variaciones y, en su contra, el volumen de los datos recogidos y procesados aumenta con el cuadrado de la resolución.

De otra parte, no siempre es necesario e incluso deseable dotarse de la máxima resolución en las imágenes a procesar. La resolución temporal es una característica muy apreciada dentro de la meteorología, puesto que ello proporciona mayor riqueza en el estudio de la evolución de las condiciones atmosféricas. La mayor resolución temporal la proporcionan los satélites de órbita geoestacionaria, pero su distancia a la superficie impide que puedan tener resoluciones espaciales de detalle, por otra parte, poco interesantes para el meteorólogo. El cartógrafo o quien se dedica a la actualización catastral de rústica, por ejemplo, requieren para su trabajo imágenes de gran riqueza espacial, y en cambio apenas necesitan que

sean multiespectrales, y mucho menos, hiperespectrales. Sin embargo, dentro de la edafología o para quien que se dedica a la agricultura de precisión son imprescindibles estas últimas. Los requerimientos de resolución no son lo mismo para un satélite espía que para otro destinado al estudio de la evolución de la vegetación a escala regional.

Tipos de sensores

Antes de clasificar los sensores es necesario distinguir entre dos tipos de teledetección: la pasiva y la activa. En la teledetección pasiva el instrumento solamente recibe radiación. Habitualmente se aprovecha la reflectividad de la superficie y se mide la radiancia espectral reflejada por ella, una vez está siendo irradiada desde una fuente externa. El sol ilumina la superficie terrestre en toda la gama de longitudes de onda, entre las que predominan las correspondientes al espectro visible y sus vecinas más cortas y más largas, lo que se denomina *espectro solar*. La superficie, en virtud de su reflectancia espectral reenvía al espacio parte de esa radiación, que es la fracción que captura y mide el sensor. Pero también hay que considerar que la propia superficie terrestre, debido a su temperatura emite radiación de onda larga, que puede ser medida de día o de noche. La radiación térmica puede diferenciarse de la reflejada si se mide en condiciones de oscuridad. Tal es el caso de los satélites meteorológicos cuando adquieren imágenes nocturnas en el infrarrojo térmico en torno a 10 μm . También emite microondas, radiación que puede ser registrada tanto de día como de noche, dado que la reflexión de la radiación solar en esta región del espectro es poco significativa. En la teledetección activa se obvia la radiación externa y es el

En definitiva, dentro de la gama disponible, debe elegirse la plataforma y el sensor de acuerdo con las características propias del proyecto de teledetección.

propio instrumento el que ilumina la superficie y registra la parte reflejada de esa radiación. Así funcionan los altímetros y los sistemas de radar formadores de imágenes.

Dentro de los sensores pasivos puede establecerse la siguiente clasificación:

- **No escáneres.**
 - No formadores de imagen
 - Radiómetros de microondas.
 - Sensores magnéticos.
 - Gravímetros.
 - Espectrómetros de Fourier.
 - Sónar pasivo.
 - Otros.
 - Formadores de imagen
 - Cámaras.
 - monocromas.
 - de color natural
 - infrarrojo.
 - infrarrojo color.
 - térmicas.
 - otras.
- **Escáneres**
 - Formadores de imagen

- Escáner del plano imagen.
 - ▾ cámaras de TV.
 - ▾ escáneres sólidos.
- Escáneres del plano objeto.
 - ▾ escáneres óptico-mecánicos.
 - ▾ radiómetros de microondas.

Entre todos estos sistemas, dentro de los instrumentos destinados a la formación de imágenes destacan los escáneres. Estos instrumentos exploran la superficie mediante el barrido de la escena, gracias a un espejo basculante, cuyo eje de oscilación es paralelo a la trayectoria del satélite portador. Antes de ser registrada, la señal puede ser descompuesta ópticamente en varias bandas espectrales, cada una de las cuales se envía a un detector para formar una imagen multiespectral. Si se conocen los parámetros de calibración del sensor y las condiciones de adquisición, es posible transformar los niveles digitales de la imagen en valores físicos de radiancia. Los valores numéricos son almacenados a bordo del satélite para su emisión posterior cuando en su trayectoria orbital entre dentro en el cono de recepción de las estaciones de seguimiento, o transmitidos inmediatamente a una constelación de satélites repetidores.

Con la puesta en órbita del satélite SPOT en 1986 se introdujo un nuevo sistema de exploración que prescinde del espejo oscilante. Los detectores se disponen en líneas, con tantos elementos como celdas tenga cada línea de imagen, de modo que abarquen simultáneamente todo el campo de visión. Estas líneas de detectores se van excitando ordenadamente con el movimiento del satélite y la captura de datos no se realiza píxel a píxel sino línea a línea. En general, estos sistemas aumentan la resolución espacial del sensor, pues al eliminar partes móviles minimizan el error introducido por la falta de sincronía entre su movimiento y el de la plataforma. Presentan, sin embargo, el inconveniente de la dificultad de calibrar

toda la cadena de detectores, de modo que todos mantengan los mismos coeficientes de conversión de radiancia a nivel digital. No obstante, es la tendencia en el diseño y fabricación actual tanto de sensores orbitales como de cámaras aéreas.

En cuanto a los sensores activos, se pueden clasificar del siguiente modo:

- **No escáneres.**
 - No formadores de imagen
 - Radiómetros de microondas.
 - Altimetros de microondas.
 - Batímetros láser.
 - Distanciómetros láser.
- **Escáneres**
 - Formadores de imagen
 - Escáneres de plano objeto.
 - ▾ radar de apertura real.
 - ▾ radar de apertura sintética (SAR).
 - ▾ sónar activo.
 - Escáneres del plano imagen.
 - ▾ radar pasivo de antena en fase (*phase array*).

En el ámbito de la teledetección es ampliamente utilizado el radar, que es un sistema activo cuya antena emite pulsos de energía de longitud de onda comprendida aproximadamente entre 1 mm y 1 m, y también mide la radiación reflejada por la superficie. El sistema consiste en un mecanismo generador de pulsos electromagnéticos que controla la emisión de energía desde el transmisor y, además, acciona un mecanismo de registro. La antena emite un pulso y recibe la respuesta (eco) del terreno. Un conmutador electrónico (duplicador) evita la interferencia entre ambos pulsos. El receptor amplifica la señal captada por la antena, que es muy inferior a la emitida en

razón a que este tipo de instrumentos operan emitiendo un haz inclinado para evitar ambigüedades en el retorno de la señal. De esta forma la mayor parte de la radiación emitida hacia la superficie se pierde en la dirección de reflexión especular y tan solo una fracción de ella retorna hacia el sensor, fenómeno que se conoce como retrodispersión, y cuya magnitud depende de la longitud de onda del radar, de la inclinación del haz radar y de la rugosidad del terreno. El pulso recibido se graba para su procesamiento posterior en el segmento terrestre de la misión. En el rango espectral del radar las cubiertas nubosas, tanto

de hielo como de agua, presentan una alta transmisividad, lo que hace este sistema muy adecuado para el estudio de las zonas cubiertas casi permanentemente por nubes, como son las tropicales y ecuatoriales. Entre las modalidades de instrumentos se pueden encontrar sistemas de radar lateral aerotransportado, denominado SLAR (*Side Looking Airborne Radar*) y radares de apertura sintética (SAR, *Synthetic Aperture Radar*), tanto orbitales como aerotransportados.

Misiones y programas de teledetección

Abarcar todo el panorama de satélites y programas de teledetección queda fuera del reducido ámbito de este capítulo. A finales de 2018 había catalogados cerca de 2000 satélites en órbita, que están recogidos en diversas bases de datos accesibles en internet¹. Lógicamente, cada familia de satélites está destinada a cubrir un objetivo concreto y, en función de él, se diseñan sus parámetros orbitales y se les provee de uno o varios instrumentos con las características específicas apropiadas.

En 1972 se puso en órbita el primero de la serie de satélites ERTS (*Earth Resource Technology Satellite*) destinados al estudio de los recursos naturales. A partir del segundo lanzamiento, el programa se red denominó Landsat y, con él, se han llegado a poner en órbita ocho generaciones de satélites. Constituye esta misión un ejemplo paradigmático de lo que se conoce como satélites de observación de la Tierra.

Son misiones multipropósito orientadas a estudiar la superficie terrestre en la región solar del espectro, incluido el infrarrojo térmico. Las amplias utilidades de sus imágenes abarcan campos tan variados como los inventarios agronómicos y forestales, el pronóstico temprano de cosechas, la evaluación y control de zonas regables, la planificación hidrológica de cuencas, la cartografía de los usos del suelo, el estudio de los recursos litorales, los estudios geológicos y edafológicos, la glaciología, el control de la contaminación de aguas y suelos, etc. El actual Landsat 8 está dotado con dos instrumentos, el primero de los cuales, OLI (*Operational Land Imager*), con una resolución espacial de 30 m, explora el rango de longitudes de onda comprendido entre el visible y el infrarrojo medio, con cuatro canales en el visible más uno pancromático —éste con 15 m de resolución espacial—, uno en el infrarrojo próximo y tres en el infrarrojo medio, uno de los ellos en la banda de absorción del

¹ Una de ellas, por ejemplo, es: <https://www.ucsusa.org/resources/satellite-database#.WIMrWVXiZpg>

agua, en 1,38 μm . El otro instrumento, TIRS (*Thermal Infrared Sensor*) tiene dos canales entre 10 y 13 μm con una resolución espacial de 100 m.

A partir del final de la década de los 70 comenzaron a desarrollarse programas europeos de mayor resolución espacial que constituyeron un hito en la observación de la Tierra. El programa SPOT, desarrollado por el CNES (Centro Nacional de Estudios Espaciales francés) en colaboración con Bélgica y Suecia puso su primer satélite en órbita en 1978 y desde entonces ha renovado hasta siete veces sus plataformas. Spot 7 monta dos instrumentos orientables que permiten ampliar el campo de visión, tomar imágenes simultáneas de áreas alejadas, o generar estereoscopia adquiriendo la imagen de un lugar desde una órbita contigua. En el pancromático tiene una resolución espacial de 2,5 m, aunque con procesamiento digital el constructor asegura alcanzar 1,5 m de GSD, y en modo multiespectral de 6 m. Cuenta con cinco bandas espectrales desde el visible hasta el infrarrojo próximo.

En septiembre de 1999 hizo su aparición el primero de una serie de satélites llamados entonces de muy alta resolución espacial. IKONOS es un satélite operado por una compañía privada, *Space Imaging*, con una resolución espacial de 1 m capaz de adquirir escenas de 11x11 km^2 . Su instrumento es orientable, de modo que puede programarse para tomar imágenes con una resolución temporal de tan solo 3 días. Un par de años más tarde la compañía *DigitalGlobe* puso en órbita el primer satélite *QuikBird* con resolución submétrica, concretamente con 61 cm en el pancromático. El instrumento tiene capacidad de capturar imágenes en color gracias a sus cuatro canales, en este caso con 2,44 m de resolución. Con la resolución de estos satélites ha sido posible comenzar a aplicar la teledetección en áreas urbanas, para lo cual

ha sido necesario también dotar a los instrumentos de una mayor resolución radiométrica que hiciera posible estudiar las zonas en sombra, muy abundantes en estos ámbitos. La misma compañía es la actual propietaria de *GeoEye Inc.*, anteriormente denominada *Orbimage* corporación diseñadora y gestora de una flota de satélites de muy alta resolución que incluye GeoEye1, satélite lanzado en 2008 con una resolución espacial inicial de 0,41 m en modo pancromático. En el verano de 2013 el satélite fue elevado de la órbita de diseño de 681 km de altura a la que mantiene actualmente, que es de 770 km, disminuyendo ligeramente su resolución espacial a 0,46 m en el pancromático. La resolución en modo multiespectral es de 1,84 m. La flota gestionada por *DigitalGlobe* incluye satélites de similar resolución como *WorldView 1 y 2*, que tienen una banda pancromática de 0,43 m de resolución y 8 bandas multiespectrales distribuidas entre el ultrazul y el infrarrojo próximo. En 2014 se puso en órbita el *WorldView 3* a 617 km de altitud con una resolución espacial pancromática de 31 cm, una resolución multiespectral de 1,24 m y una resolución infrarroja de onda corta de 3,7 m. El número 4 de la serie, lanzado en 2016, ha fracasado a primeros de 2019 y, aunque tenía el mismo diseño que su predecesor, se había discutido la modificación de su altura orbital para conseguir imágenes de hasta 21 cm de resolución pancromática.

Un apartado especial merecen los satélites meteorológicos, dado que fueron de los primeros en aparecer, que impulsaron la teledetección de uso civil y que abrieron paso a la generación de satélites de observación de la tierra. La familia de satélites meteorológicos se encuentra dividida en dos grupos, según sea la órbita de la plataforma. Los satélites geoestacionarios se sitúan permanentemente sobre la vertical del mismo lugar, para lo cual se les coloca a casi 36 000 km de la superficie. La pérdida de resolución espacial se compensa con la gran resolución temporal propia

de estos satélites, ya que son capaces de adquirir escenas completas del hemisferio que observan cada pocos minutos. Por otro lado, los satélites polares se sitúan en órbitas de entre 800 y 900 km de altura, mejorando la resolución espacial de las imágenes, pero perdiendo resolución temporal. Estos satélites revisitan el mismo lugar cada pocos días y, para compensar la carencia de datos, suelen operar en constelaciones de satélites de la misma familia, reduciendo con ello el tiempo de revisita a unas pocas horas. Lo que sí es característico de todos ellos, sea cual sea su órbita es la existencia de canales de observación específicos para el estudio de la atmósfera. En las primeras versiones de *Meteosat*, satélite europeo de órbita geoestacionaria, el instrumento montaba tres bandas, una pancromática visible (0,45-1 μm) para la observación de la superficie y de las cubiertas nubosas, otra en el infrarrojo medio (5,7-7,1 μm) denominado *vapor de agua (WV)*, en la que la atmósfera es poco transparente a la radiación debido a la presencia de vapor de agua, pero precisamente por eso proporciona información sobre el contenido de humedad en la atmósfera, y finalmente otro en el infrarrojo térmico (10,5-12,5 μm) destinado a medir la temperatura de la superficie o de las nubes interpuestas. Con esta combinación de canales, el meteorólogo cuenta con los elementos indispensables para realizar su pronóstico: observa las nubes y su movimiento, mide la temperatura y determina los lugares de riesgo de precipitación por las altas concentraciones de vapor de agua. La resolución espacial en el nadir era de 2,5 km en el visible y de 5 km en los otros dos canales. En la actualidad, los instrumentos herederos del radiómetro primitivo, integrantes de los *Meteosat de Segunda Generación (MSG)*, cuenta con 12 canales, uno en el pancromático visible (0,45-1,05 μm), dos en el visible (0,6 y 0,8 μm), uno en el infrarrojo próximo (1,6 μm) y ocho en el infrarrojo medio y térmico (3,9; 6,2; 7,3; 8,7; 9,7; 10,8; 12,0 y 13,4 μm). La resolución

espacial del visible ha sido mejorada hasta 1 km. A partir de 2021 se prevé la puesta en órbita del primer *Meteosat de Tercera Generación (MTG)* con nuevas mejoras.

Con características similares están los satélites meteorológicos geoestacionarios de Estados Unidos y Rusia, GOES y *Meteor*, respectivamente y de otras potencias mundiales.

En cuanto a los satélites meteorológicos de órbita polar, un ejemplo muy característico son los pertenecientes a la serie NOAA, la cual está constituida por un conjunto de al menos dos satélites operativos, aunque a menudo hay épocas en las que coexisten varios más adquiriendo y enviando datos. Toman su nombre de la agencia que los gestiona, la Administración Nacional Oceánica y Atmosférica de los Estados Unidos. El último lanzado es el NOAA-19, pero siguen operando con la mayor parte de sus instrumentos activos los dos anteriores. Heredan la serie iniciada en 1978 con el primero de los satélites TIROS-N (*Television Infrared Observatrion Satellite-Next Generation*). Los satélites NOAA orbitan a una altura de entre 833 y 870 km y tienen un período de 102 minutos. Con la configuración de su órbita, el FOV del instrumento y el hecho de que vuelan varios de ellos, se consigue un suficiente número de ocurrencias diarias para cualquier punto de la superficie terrestre como para elaborar un pronóstico meteorológico de gran exactitud. El satélite lleva una carga útil de una veintena de sensores específicamente diseñados para distintos objetivos, entre los que destaca por su gran versatilidad el AVHRR (*Advanced Very High Resolution Radiometer*) con 6 bandas espectrales (dos en VIS, uno en IRp, dos en IRm y dos más en IRT), una resolución nominal de 1100 m y una resolución radiométrica de 10 bits.

El programa de satélites de operaciones meteorológicas (MetOp) es un programa europeo conjunto de la Agencia Espacial Europea (ESA) y la Organización Europea para la Explotación de Satélites Meteorológicos (Eumetsat) coordinado también con la NOAA. Desde 2006, MetOp-A, es el primer satélite meteorológico europeo de órbita polar. En 2012 y 2018 se pusieron en órbita MetOp-B y MetOp-C, respectivamente. Los satélites MetOp vuelan a 800 km de altura en órbitas complementarias a los NOAA para completar los datos obtenidos por ambos. Cinco de sus instrumentos son de diseño europeo, en tanto que el resto heredan las características de los montados en la serie NOAA o incorporan aportaciones de la Agencia Espacial Francesa (CNES).

Entre las más de setenta misiones que gestiona, la ESA está desarrollando el programa Sentinel, a través del cual mantiene en órbita una constelación de satélites gemelos, cada uno de cuyos tipos están destinados a objetivos complementarios. En cada programa coexisten dos satélites para cubrir los requisitos de cobertura y resolución temporal, denominados A y B. Sentinel-1 es una misión de radar de órbita polar, con un primer satélite operativo desde 2014 y el segundo desde 2016. Sentinel-2 es una plataforma multiespectral de observación de la Tierra con hasta 10 m de resolución espacial, 13 bandas espectrales y una alta resolución temporal. Sus dos satélites fueron lanzados en 2015 y 2017, respectivamente. Sentinel-3 es multipropósito y embarca varios instrumentos destinados a medir la altimetría, la temperatura o el color del océano. Están en órbita desde 2016 y 2018, respectivamente. Por su parte, los instrumentos Sentinel-4 y Sentinel-5, dedicados al estudio y seguimiento de la atmósfera, de la calidad del aire, del ozono estratosférico y de la radiación solar, no constarán de plataforma propia sino que serán embarcados en satélites operados por Eumetsat. El primero

será geoestacionario y está integrado dentro del satélite MTG, en tanto que el segundo será de órbita polar y estará embarcado en la plataforma MetOp de Segunda Generación. Finalmente, Sentinel-6 será un radar altimétrico que medirá con gran precisión la topografía y en nivel del océano, prosiguiendo los estudios iniciados con el altímetro Jason y la misión CryoSat de la ESA.

Dentro de las plataformas activas de radar destaca Radarsat, satélite canadiense activo desde 1995. Radarsat-1 trabajó hasta 2013, pero desde 2008 fue primeramente complementado y luego sustituido por Radarsat-2. El satélite porta un instrumento consistente en un radar de apertura sintética (SAR) que trabaja en la banda C con una frecuencia portadora de 5,405 GHz, correspondiente a una longitud de onda de 5,55 cm). Radarsat tiene siete modos de haz con el objetivo de ofrecer diferentes resoluciones y coberturas del terreno. El modo estándar cubre un área de $100 \times 100 \text{ km}^2$ y tiene resolución espacial de 30 m, pero puede ofrecer desde 1 a 100 m. Una capacidad muy interesante de este satélite es su capacidad para evaluar la polarización horizontal o vertical del haz recibido, en dependencia de la polarización de la emisión, de modo que puede adquirir imágenes HH, HV, VH, VV y distintos modos de combinación de las anteriores. Vuela a una altura de 798 km y, al igual que otros satélites, tiene la capacidad de dirigir el haz en diferentes ángulos. Su resolución temporal es de 24 días.

8

Tratamiento digital de imágenes

Introducción

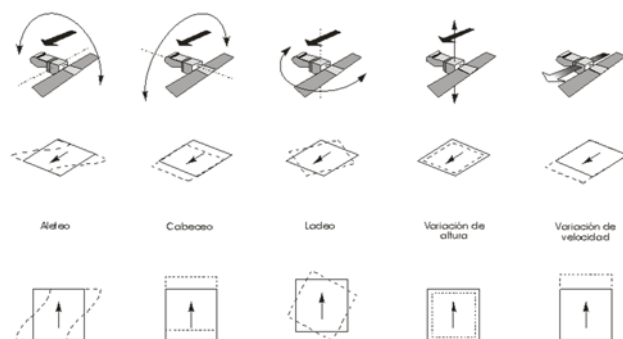
La imagen adquirida por un sensor a bordo de un satélite presenta diferencias con respecto a la imagen real tanto en la situación de sus puntos como en el nivel de brillo de cada uno. En consecuencia, la geometría de la imagen no refleja con fidelidad absoluta la de la superficie, ni tampoco pueden deducirse sus propiedades físicas a partir de la lectura directa de los niveles digitales. Es necesario aplicar unas determinadas operaciones de corrección, que están encaminadas a minimizar estas alteraciones en la medida de lo posible. Algunas de estas tareas son aplicadas

por las estaciones de recepción o los distribuidores de las imágenes, que ofrecen distintos niveles de corrección. El estudio detallado de la imagen, los estudios multitemporales o la fusión con otras fuentes cartográficas requieren correcciones precisas, que normalmente son realizadas por el analista.

Este capítulo aborda las operaciones más habituales en el pretratamiento de las imágenes de satélite antes de proceder al estudio en sí de la información que proporcionan.

Distorsiones

Las imágenes de satélite no constituyen una cartografía debido a factores como la propia naturaleza de la proyección, las distorsiones sistemáticas y las aleatorias que sufren las imágenes como consecuencia de la geometría de la adquisición, del movimiento del satélite y de la óptica y la mecánica del instrumento de adquisición. Las distorsiones que sufre una imagen pueden ser de tipo geométrico o radiométrico.



Distorsiones aleatorias en la adquisición de la imagen

Distorsiones geométricas

Las fuentes que dan origen a las distorsiones geométricas se enumeran a continuación.

Oscilación de la plataforma

Si bien el movimiento orbital del satélite es prácticamente uniforme, y por tanto, sus aceleraciones lineales son nulas en cualquier dirección salvo las derivadas de la excentricidad de su órbita, en realidad puede sufrir pequeñas variaciones en altura orbital, velocidad u orientación de cualquiera de sus tres ejes, lo que da lugar a los efectos de aleteo (*pitch*), que es el giro según el eje de la trayectoria, de cabeceo (*roll*), o giro según un eje perpendicular al primero y paralelo a la superficie terrestre, y ladeo (*yaw*), giro según el eje perpendicular a ella.

Rotación terrestre

La traza de un satélite polar no está situada a lo largo

de un mismo meridiano terrestre, debido a que mientras el satélite describe su órbita, la Tierra rota alrededor de su eje. Cuando el satélite, tras haber cruzado el polo terrestre, ha alcanzado el ecuador, el planeta ha girado y el punto ecuatorial que pertenece al meridiano contenido en el plano orbital cuando el satélite estaba sobre la latitud 90° se ha desplazado hacia el este.

Los satélites adquieren líneas completas de imagen sucesivamente y sin solución de continuidad mientras se desplazan en su órbita alrededor de la Tierra. La escena formada por un cierto paquete de líneas se forma en el segmento terrestre de la misión para comercializar la información. Una escena contiene un determinado número de líneas de imagen, que son perpendiculares a la traza del satélite. Si la imagen corresponde a una traza descendente, las líneas inferiores corresponderán a zonas geográficas situadas más al oeste que las superiores, que fueron tomadas unos instantes antes. Lo contrario ocurre si la traza es ascendente. Esto produce un efecto típico de distorsión en todas las imágenes de sensores polares o casi-polares (no afecta a los geoestacionarios): si la imagen no se corrige, tendrá forma rectangular y las zonas inferiores estarán desplazadas hacia el este (traza descendente). Tras la corrección de la distorsión debida a la rotación terrestre, la imagen adquirirá una forma romboidal con las líneas inferiores desplazadas hacia el oeste, representando zonas, ahora sí, en su posición geográfica correcta.

Tiempo de barrido

Si el sensor es del tipo escáner, cada línea o grupo de ellas es adquirida por un solo detector o un pequeño número de éstos mientras un espejo explora toda la línea. En el barrido se invierte un tiempo determinado, durante el cual, el satélite evoluciona en

su propia órbita y deja de encontrarse sobre la vertical del mismo punto en el que comenzó el barrido. El final de la línea corresponde a puntos del terreno más hacia el sur que los del inicio (en traza descendente). Es otra distorsión que debe corregirse para hacer útil la imagen desde el punto de vista cartográfico.

Distorsión panorámica

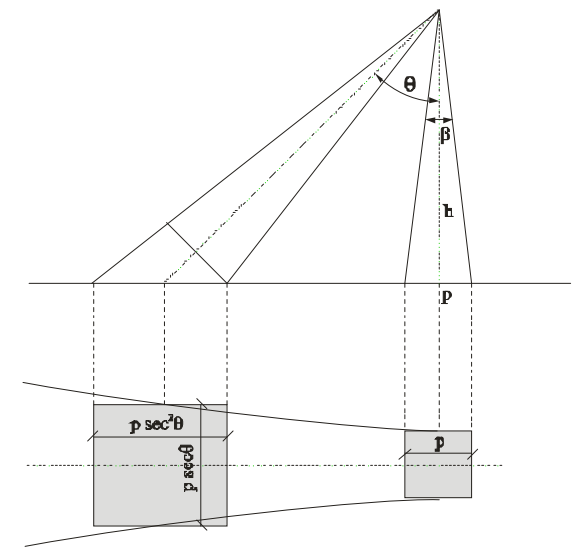
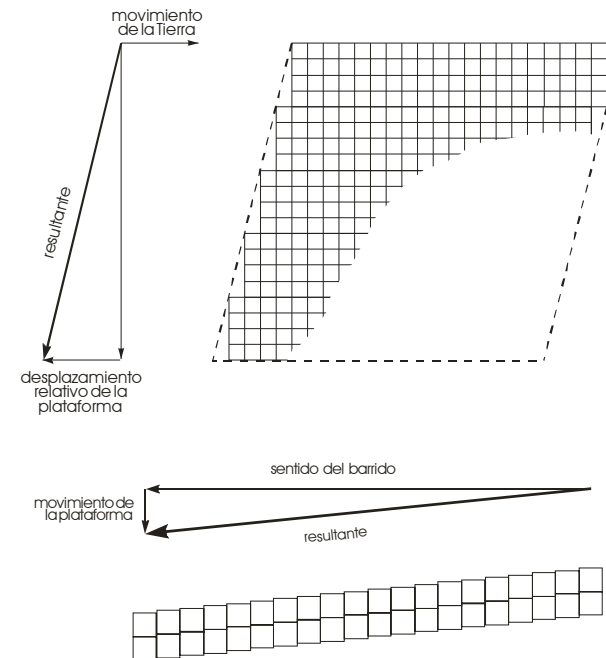
Dado que el campo de visión instantáneo es una característica propia del sensor y además es invariable, el tamaño de las celdas en puntos alejados de la traza del satélite es mayor que el correspondiente al nadir del instrumento, que coincide con su resolución espacial nominal. Realmente, ni siquiera el IFOV es cuadrado ni rectangular, sino circular o elíptico. La formar la imagen se traduce el nivel de radiancia registrado en cada área explorada instantáneamente a un nivel digital conferido a una celda cuadrada.

Curvatura de la Tierra.

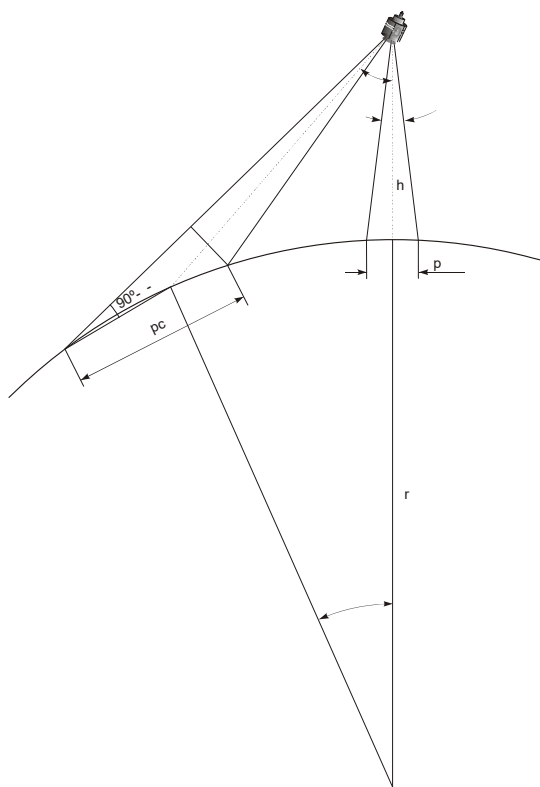
Además del efecto anterior, hay que considerar que la Tierra no es plana y que, por lo tanto, el tamaño de las celdas alejadas de la traza crece considerablemente. Este efecto es particularmente apreciable en satélites meteorológicos, de gran campo de visión, sobre todo en los geoestacionarios.

Distorsiones radiométricas.

En el otro grupo de distorsiones se encuentran las de tipo radiométrico, que no afectan a las posiciones de los puntos, sino a su nivel digital. Normalmente se trata de pérdida de celdas aisladas o de líneas completas, que aparecen en la imagen en negro o en blanco, o bien en niveles de gris anormalmente altos o bajos con respecto a su entorno. El efecto de líneas alteradas puede repetirse regularmente a lo largo de



Distorsión debida a la rotación terrestre (arriba), al tiempo de barrido (en el centro) y panorámica (abajo).



Distorsión debida da la curvatura terrestre.

la imagen en sensores de barrido cuando se descalibra algún detector, causando el efecto conocido como *bandeado*. En los sensores de empuje, la descalibración de los detectores no se presenta en las líneas sino en las columnas de la imagen y además

Operadores

Un operador es una función matemática que se aplica sobre la imagen, entendida como una matriz numérica. Un operador T es una aplicación entre el espacio inicial, por ejemplo, una imagen original $z=f(x,y)$ y el espacio final, constituido por la imagen transformada $z'=f(x',y')$:

$$z = f(x, y) \xrightarrow{T} z' = f'(x', y')$$

Esta transformación se puede desglosar en sus tres componentes, donde las variables x e y y sus transformadas x' e y' hacen referencia a las dimensiones planimétricas de la imagen y la variable z y su transformada z' se refieren al nivel digital de la celda:

$$x' = t_x(x, y, z)$$

$$y' = t_y(x, y, z)$$

$$z' = t_z(x, y, z)$$

Las dos primeras constituyen el grupo de transformaciones geométricas, que alteran la geometría de la imagen sin modificar los niveles digitales —o al menos no de forma significativa, como se verá más ade-

no suelen seguir ningún patrón repetitivo. Finalmente, cabe también hablar de una distorsión radiométrica causada por el efecto atmosférico, ya que tanto la dispersión como la absorción, aerosólica o molecular, alteran los niveles digitales, produciendo en la imagen niveles de brillo que no son representativos de la radiancia de la superficie.

lante— y la tercera las transformaciones radiométricas, que modifican el brillo sin intervenir en la geometría. Habitualmente, ambos tipos de transformaciones se abordan por separado. Las transformaciones también pueden escribirse:

$$f_1(x', y') = T_1[f(x, y)]$$

$$f_2(z') = T_2[f(z)]$$

Siendo T_1 y T_2 lineales e invariables al desplazamiento, esto es, operan del mismo modo sobre cualquier celda de la imagen, manteniéndose la función igual a sí misma:

$$\begin{aligned} T[f(x-x_0, y-y_0)] - f(x-x_0, y-y_0) = \\ = T[f(x, y)] - f(x, y) \end{aligned}$$

Según el número de celdas involucradas en la transformación, los operadores pueden ser:

- Puntuales: cuando cada elemento de la imagen final es función de un único elemento de la imagen inicial.
- Locales: cuando el nivel digital de cada elemento de la imagen final es función de los valores de las celdas de un determinado entorno de cada elemento de la imagen inicial.
- Globales: cuando la imagen final en su totalidad es función de toda la imagen inicial.

Entre los primeros se encuentran unas de las operaciones más habituales en el tratamiento digital de imágenes, las transformaciones radiométricas y geométricas. Las transformaciones radiométricas modifican la amplitud de la señal (brillo o nivel digital), manteniendo invariables las coordenadas de las celdas. Las transformaciones geométricas, por el contrario, conservan la radiometría pero modifican la geometría haciendo variar la posición de los puntos de la imagen.

Estadísticos básicos de la imagen

Para el tratamiento digital de la imagen va a ser recurrente echar mano de algunos estadísticos básicos de la imagen. Para ello se definen los siguientes conceptos:

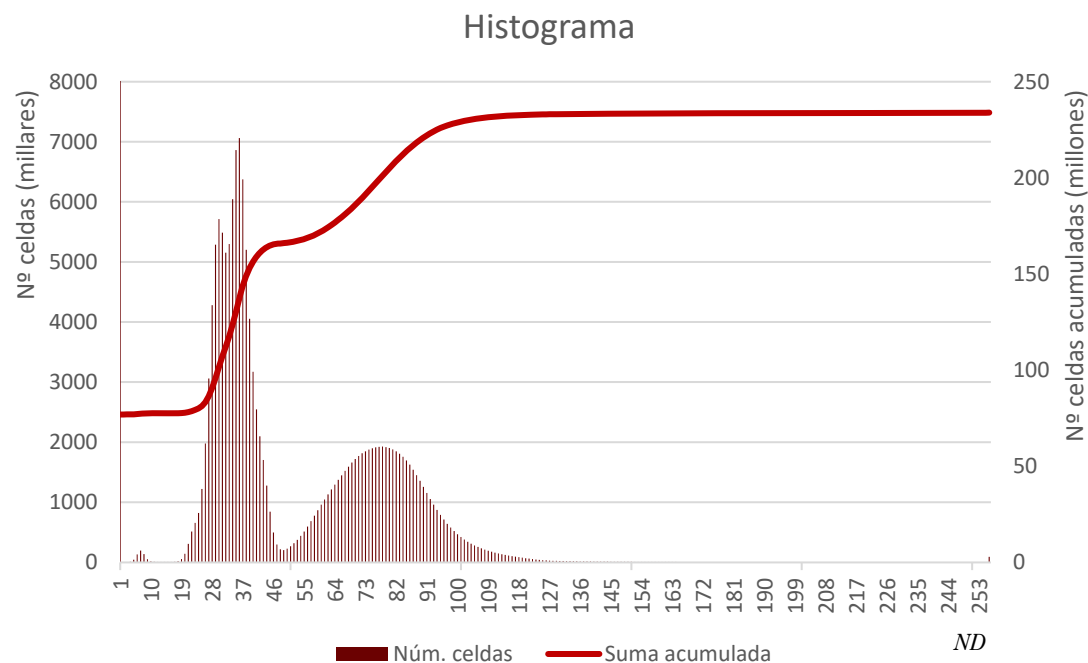
Histograma

El histograma de la imagen es la representación del número de celdas que tienen cada uno de los k niveles digitales posibles. Es el equivalente estadístico del histograma de frecuencias. Por ello también es denominado a veces *función de densidad* (fdd) o $p(z)$.

Puede asimilarse a una función probabilística si se considera el universo muestral el conjunto de todas las celdas de la imagen, cada una de ellas con un cierto ND . La frecuencia relativa vendría a ser en este caso la *probabilidad* de encontrar en toda la imagen celdas con un determinado ND .

Histograma acumulativo

Es el histograma del número de celdas que en la imagen tienen un determinado nivel digital o inferior a



Representación del histograma de una imagen y de su histograma acumulativo.

éste. Coincide con el concepto estadístico de *función de distribución* (ffD) o $P(z)$.

Media

La media de la imagen es el valor promedio de los ND del conjunto de celdas.

$$\overline{ND} = \sum_{i,j} \frac{ND}{N}$$

donde \overline{ND} es el nivel digital promedio, i,j las columnas y líneas de la imagen y N el número de celdas, esto es $i \cdot j$.

Varianza

Es una medida de dispersión de los ND en torno al nivel promedio. Encierra una forma de considerar el contraste y la información que porta una imagen.

$$\sigma^2 = \sum_{i,j} \frac{(ND - \overline{ND})^2}{N}$$

Valores extremos

Será interesante con diversos propósitos considerar los niveles digital mínimo ND_{min} y máximo ND_{max} de una imagen, es decir, el nivel digital de la celda más oscura y el de la más clara de la imagen, respectivamente, que en términos generales no ha de ser necesariamente negro (0) o blanco ($k-1$).

Transformaciones radiométricas

Las transformaciones radiométricas son utilizadas para alterar el brillo de la imagen, mediante operaciones de realce o reducción del contraste. También pueden emplearse para minimizar, que no corregir por completo, determinadas anomalías sufridas durante la adquisición o grabación de la imagen.

Se entiende por rango o margen dinámico la amplitud completa de niveles digitales que una imagen puede exhibir, según sea su resolución radiométrica.

Para un rango dinámico k :

$$ND = f(x, y); \quad 0 \leq ND \leq k-1$$

Siendo 0 el negro y $k-1$ el blanco de la celda.

En el análisis digital puede ser interesante operar con la imagen normalizada, esto es, acotada entre 0 y 1:

$$z = g(x, y)$$

de modo que:

$$ND = \text{RED}(k z' - 1)$$

Siendo RED la función de redondeo. Por ejemplo, si $z = 0,5$:

$$ND = RED[256 \cdot 0,5 - 1] = 127$$

Redondeo y truncamiento

Entre las transformaciones radiométricas más sencillas se encuentra el redondeo y truncamiento. Si la imagen original posee un rango dinámico $0 \leq ND \leq k - 1$ y la imagen final, que no tiene por qué conservarlo, tiene un rango $0 \leq ND' \leq k' - 1$, entonces:

$$ND' = \begin{cases} 0 \\ RED[ND] \\ k'-1 \end{cases} \text{ con } \begin{cases} ND' \leq 0 \\ 0 \leq ND' \leq k'-1 \\ k'-1 \leq ND' \end{cases}$$

El operador acota entre 0 y $k'-1$ la imagen, aun en el supuesto de que, como resultado de una transformación previa, los valores pudieran ser negativos o exceder del nuevo rango dinámico, y además redondea a valores enteros cualquier nivel de brillo comprendido dentro de ese rango dinámico.

En las transformaciones radiométricas, la condición necesaria para que no se produzca una inversión del contraste y para que exista la transformación inversa es que $T(z)$ sea biunívoca y monótonamente creciente, es decir que:

$$T(z_1) \geq T(z_2) \quad \text{si} \quad z_1 \geq z_2$$

para lo cual es necesario que el rango dinámico de la imagen final no sea inferior al de la imagen original:

$$k' \geq k$$

Transformaciones generales de amplitud

Existen numerosas transformaciones radiométricas, cada una de ellas orientada a conseguir un determinado efecto. Considerando el caso más general en el que la imagen original y la transformada tengan, respectivamente, rangos dinámicos k y k' , pueden citarse como unas de las más habituales la identidad y el negativo:

→ Identidad:

$$z' = z$$

lo que implica que:

$$ND' = RED \left[\frac{k'-1}{k-1} \cdot ND \right]$$

→ Negativo, en el sentido de *negativo fotográfico*, esto es, inversión del contraste:

$$z' = 1 - z$$

lo que implica que:

$$ND' = RED \left[(k'-1) \left(1 - \frac{ND}{k-1} \right) \right]$$

Otras transformaciones de amplitud podrían implicar funciones exponenciales o logarítmicas.

Correcciones radiométricas

Las correcciones radiométricas son operaciones encaminadas a eliminar o minimizar las alteraciones que la imagen original haya podido sufrir en sus valores de luminancia, bien a nivel general o bien localmente como consecuencia de algún fallo en el proceso de captura y/o grabación.

Restauración de celdas o líneas perdidas

Durante el transcurso de la adquisición o grabación de la imagen, es posible que por diversas causas algunas celdas aisladas que se queden sin un ND asignado, presentando valores claramente erróneos y apareciendo en la imagen como completamente negras ($ND = 0$) o completamente blancas ($ND = k-1$). Existen técnicas para la restauración de estas celdas que operan en el dominio de la frecuencia y que serán abordadas en el capítulo correspondiente. En el dominio espacial es muy fácil buscar un remedio a esta situación asignando a la celda perdida en la posición de columna y fila (i,j) un valor igual al promedio de las cuatro vecinas más próximas:

$$ND'_{i,j} = \text{RED} \left[\frac{1}{4} (ND_{i,j-1} + ND_{i,j+1} + ND_{i-1,j} + ND_{i+1,j}) \right]$$

o al promedio de las ocho celdas circundantes:

$$ND'_{i,j} = \text{RED} \frac{1}{8} \left[(ND_{i,j-1} + ND_{i+1,j-1} + ND_{i+1,j} + ND_{i+1,j+1} + ND_{i,j+1} + ND_{i-1,j+1} + ND_{i-1,j} + ND_{i-1,j-1}) \right]$$

En otras ocasiones, en lugar de celdas aisladas, se pierden o se alteran líneas completas de imagen. Esta

circunstancia se hace patente cuando el valor promedio de los ND de las celdas de toda la línea alterada es significativamente mayor o menor que el promedio del resto de ND de la imagen. Una primera forma de asignar un valor a la línea perdida es tomar para cada celda el ND de la celda vecina de la línea anterior:

$$ND'_{i,j} = ND_{i,j-1},$$

de la siguiente:

$$ND'_{i,j} = ND_{i,j+1},$$

o, mejor aún, del promedio de ambas:

$$ND'_{i,j} = \text{RED} \left[\frac{1}{2} (ND_{i,j-1} + ND_{i,j+1}) \right]$$

En lugar de las líneas, pueden usarse de igual modo las columnas.

Esta operación no recupera los niveles digitales deficientemente adquiridos, sino que solamente minimiza el efecto óptico o mejora en cierta medida el tratamiento digital posterior. Además, es condición imprescindible para que su aplicación esté justificada desde el punto de vista matemático que exista alta

autocorrelación¹ entre los ND de las celdas vecinas.

Hay veces que el promedio entre las líneas anterior y posterior no asegura un resultado óptimo, y esto puede asegurarse a la vista de lo que sucede en una banda de la imagen que esté altamente correlacionada con la que presenta el defecto (K). En esta banda de referencia (R) podría suceder que el promedio de las líneas próximas difiera ligeramente de los valores promedio de la línea central con la que se compara. Se puede trasladar esta diferencia a la banda defectuosa y realizar una corrección más depurada:

$$ND_{i,j}^{(K)} = \text{RED} \left[\frac{ND_{i,j-1}^{(K)} + ND_{i,j+1}^{(K)}}{2} + \frac{s_K}{s_R} \left(ND_{i,j}^{(R)} - \frac{ND_{i,j-1}^{(R)} + ND_{i,j+1}^{(R)}}{2} \right) \right]$$

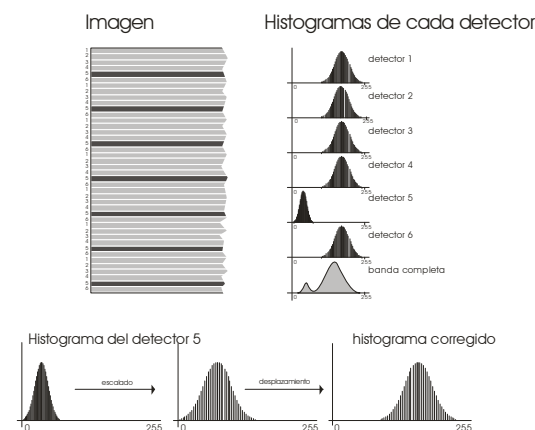
donde los superíndices (K y R hacen referencia a las bandas a corregir y la de referencia, respectivamente y s es la desviación típica en cada una de las bandas. El último término de la derecha representa el promedio de los ND de las líneas en la banda R y todo lo que está dentro del paréntesis es la diferencia entre el valor real de la línea central y el promedio de las circundantes. El primer término dentro del corchete es el promedio de ND en la banda K y el cociente de las desviaciones típicas es el factor de expansión que debe aplicarse a los resultados tomados de la banda

de referencia para homologarlos a los de la banda defectuosa.

Bandeado

Para optimizar el rendimiento de los sensores basados en sistemas de escáner, con cada oscilación del espejo no se rastrea una sola línea sino un bloque de varias. Para ello, en el plano focal se disponen tantos detectores como líneas constituyan dicho bloque. Si el detector n -simo se descalibra, todas las líneas barridas por él quedarán registradas con niveles anómalos de brillo, bien por defecto o bien por exceso. Como el fallo es permanente, la imagen exhibirá líneas anómalas en intervalos regulares de n líneas. Esta distorsión se conoce como bandeo de la imagen.

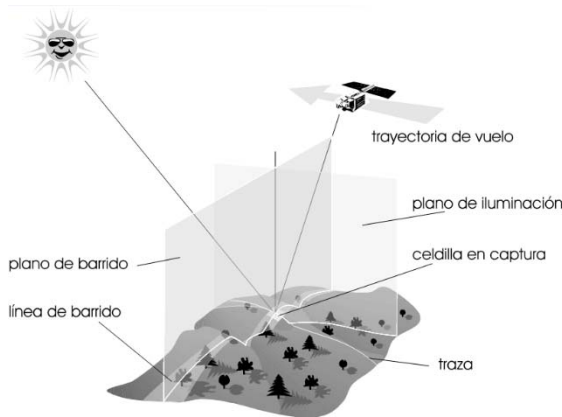
Es posible detectar si hay bandeo construyendo y comparando entre sí los histogramas parciales de la imagen que corresponden al conjunto de líneas registradas con cada detector. En condiciones normales todos los histogramas tendrán estadísticos similares, particularmente los valores de nivel digital medio desviación típica. Pero si un detector falla, es probable que todas las líneas por él registradas presenten niveles digitales inferiores, por ejemplo, y una menor varianza o desviación típica. La corrección del bandeo consiste en conseguir para el conjunto de líneas barridas por ese detector los mismos estadísticos que para el resto de la imagen. Si k es el conjunto de las líneas defectuosas:



Esquema de proceso de la corrección del bandeo de una imagen adquirida por un sensor de barrido

¹ La autocorrelación es una función estadística definida como la correlación cruzada de una señal, en este caso una imagen, consigo misma. La autocorrelación mide el grado de similitud entre una función y una versión de ésta desplazada. En el caso de la imagen, la autocorrelación de los ND viene a significar que el valor de una celda depende estrechamente de los valores de las celdas vecinas, al menos dentro de un contexto determinado. Los

modelos digitales del terreno (MDT) son funciones que tienen gran autocorrelación dentro del contexto de una ladera, limitada por las líneas de rotura. Para la mayoría de los terrenos de topografía suave, la cota en un punto depende mucho de la cota en puntos cercanos. Con la imagen de reflectancia, ocurre algo similar: es frecuente que los valores de luminancia no cambien bruscamente de una celda a la vecina.



$$ND'_{i,j} = \text{RED} \left[\frac{s}{s_k} (ND_{i,j} + \overline{ND}_k) + \overline{ND} \right]$$

donde s es la desviación típica de los ND de toda la imagen, a excepción del subconjunto defectuoso, cuya desviación típica es s_k y \overline{ND} y \overline{ND}_k son los niveles digitales promedio de la imagen, a excepción del conjunto de líneas defectuosas y el promedio de estas últimas, respectivamente.

Si se escribe:

$$ND'_{i,j} = \text{RED} [a_k ND_{i,j} + b_k]$$

puede identificarse los coeficientes a y b con la ganancia y sesgo, respectivamente:

$$a_k = \frac{s}{s_k}; \quad b_k = \overline{ND} - \frac{s}{s_k} \overline{ND}_k$$

Normalización topográfica

Con el nombre de normalización topográfica se conocen una serie de procedimientos encaminados a corregir el efecto que la topografía de la superficie induce en su reflectancia aparente. Es sabido La radiancia mantiene una relación lineal con el nivel digital de la celda y que lo que ciertamente caracteriza a la celda es su reflectancia, la cual también mantiene la linealidad con el nivel digital cuando la superficie es horizontal. Sin embargo, en condiciones de orografía muy acusada, la radiancia direccional no representa suficientemente bien a la reflectancia, puesto que aquélla depende de la irradiancia y del ángulo de visión del sensor. Hay diferencias de iluminación entre unas laderas y otras dependiendo de su orientación,

y además las celdas son vistas con diferentes grados de inclinación.

En definitiva, interesa estudiar el triedro que forman el vector característico perpendicular a la celda, la visual instantánea desde el sensor y el vector de orientación solar para corregir la radiancia simulando una superficie plana horizontal y una visual ortogonal al terreno. La radiancia estimada como real puede ser deducida a partir de la que ha registrado la imagen del siguiente modo:

$$\hat{L}_\lambda = \tau(p, q) \cdot L_\lambda = \frac{L(0,0)}{L(p, q)} \cdot L_\lambda$$

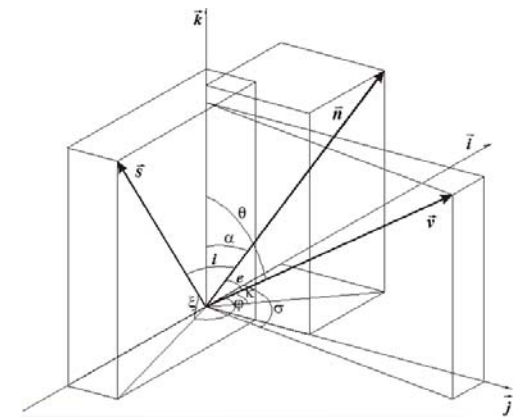
en cuya expresión, \hat{L}_λ es la radiancia estimada como real procedente de la superficie plana, L_λ es la radiancia registrada en la imagen, $\tau(p, q)$ es el factor de normalización topográfica para una celda de pendientes p y q , $L(0,0)$ es la radiancia de la superficie horizontal y $L(p, q)$ es la radiancia de la superficie inclinada.

Existen diferentes procedimientos aplicara la normalización topográfica, algunos de los cuales se muestran a continuación:

- Normalización por el coseno del ángulo de incidencia solar:

$$\hat{\tau} = \frac{\cos \theta_s}{\cos i}$$

siendo θ_s el ángulo cenital solar e i el ángulo de incidencia solar sobre la celda. Este método adolece del defecto de sobrecorregir las imágenes cuando $\cos i$ es muy pequeño.



Vectores y ángulos de interés en la adquisición de la imagen: n = vector característico; v = visual instantánea.; s = orientación solar; α = inclinación del terreno; κ = orientación del terreno; φ = acimut solar; ξ = elevación solar; θ = visión instantánea; σ = inclinación de la órbita; i = incidencia solar sobre el terreno; e = inclinación de la observación.

- Normalización de C-Huang Wei para cubiertas lambertianas:

$$\hat{L} = L - L_{\min} \cdot \left(\frac{\cos \theta_s - \cos i_{\min}}{\cos i - \cos i_{\min}} \right) + L_{\min}$$

donde L_{\min} es la radiancia mínima antes de la corrección.

- Corrección basada en la geometría sol-cubierta-sensor:

$$\hat{L} = L \cdot \left(\frac{\cos \theta_s + \cos S + C}{\cos i + C} \right)$$

donde S es la pendiente local y C es una constante

Correcciones atmosféricas

Los aerosoles atmosféricos causan un efecto de dispersión que, si bien pueden absorber radiación en determinados intervalos de longitudes de onda, comúnmente incrementan la irradiancia sobre el sensor en el espectro solar. Este efecto de la dispersión no es constante ni en el tiempo, ni en el espacio. Por esta razón, las correcciones generales sólo pueden ser consideradas como aproximaciones, en tanto que para una corrección rigurosa deberán realizarse medidas en tierra simultáneas a la adquisición de la imagen y aplicar correcciones acordes con el efecto sufrido, por lo general diferentes para cada celda de la imagen.

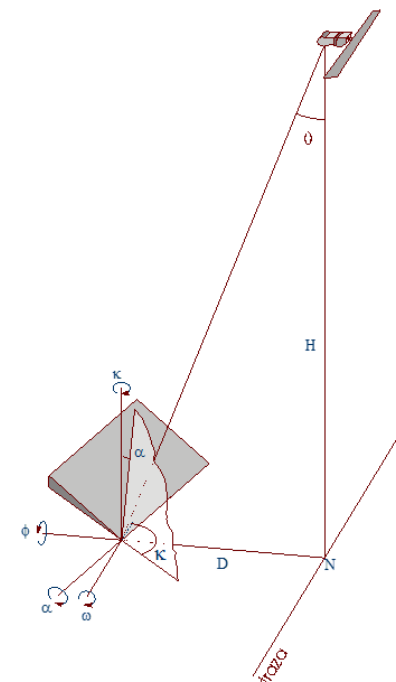
La corrección del efecto atmosférico es imprescindible cuando se desee deducir cualquier magnitud física a partir de los ND de la imagen, cuando hayan de efectuarse combinaciones lineales entre bandas, por ser la dispersión un fenómeno dependiente de la

empírica.

- Corrección por superficies de Minnaert:

$$\hat{L} = L \cdot \left(\frac{\cos e}{\cos^k i \cdot \cos^k e} \right)$$

Esta corrección introduce los ángulos de incidencia solar i y el de visión de la celda desde el sensor e , afectados por un coeficiente k de Minnaert, que tiene en cuenta el carácter lambertiano de la superficie y que oscila entre 0, para superficies absolutamente lambertianas y 1, para superficies especulares.



Relación entre la situación del sensor y el ángulo de visión de la celda.

longitud de onda y cuando se realicen estudios multitemporales, en cada una de cuyas fechas, la repercusión de la atmósfera puede haber sido diferente.

Componentes de la radiancia ascendente

Si se considera un sensor con un campo de visión instantáneo suficientemente estrecho en relación con la variabilidad de la superficie explorada, la radiancia capturada por el instrumento puede descomponerse en una suma de radiancias de diversa procedencia.

En primer lugar, como ya se dijo en el capítulo 4, se encuentra la señal L_s atenuada producida por la radiancia de la luz que atraviesa la atmósfera en sentido descendente, que se refleja en la superficie y que es transmitida hacia el sensor de nuevo a través de la atmósfera. De entre todas las demás, ésta es la señal útil que sirve para identificar la superficie y obtener

de ella sus características físicas.

Desgraciadamente, no es la única señal que incide en el plano focal del sensor. También penetra radiancia directa L_0 que procede del sol y que, previa dispersión en el seno de la atmósfera y sin haber alcanzado la superficie, termina incrementando la señal.

Y todavía hay dos fuentes más de radiancia indirecta. De una parte, la luz L_{d1} que ha sido reflejada en la dirección del sensor por una superficie que no ha sido iluminada directamente por el sol, sino a partir de la luz difusa atmosférica, siendo el caso más extremo es el de las superficies en sombra, y finalmente, puede penetrar en el instrumento radiancia L_{d2} reflejada por una superficie que ha sido iluminada directamente por el sol o por la propia atmósfera, pero cuya luz es dispersada al menos una vez dentro de la atmósfera antes de ser captada por el sensor. Dicho de otra forma, se trata de celdas iluminadas directa o indirectamente que no están bajo el IFOV del sensor.

En definitiva, se tiene que:

$$L = L_0 + L_s + L_{d1} + L_{d2}$$

y las correcciones atmosféricas que se implementen deberán ir encaminadas a estimar de la forma más aproximada el valor de L_s a partir del valor medido L .

Corrección por los valores mínimos

Una de las correcciones más rápidas y extendidas es la denominada *corrección del histograma por sus valores mínimos*. Este procedimiento asume que el ND mínimo en los canales visibles, especialmente en el rojo, debe corresponder a las zonas de agua limpia y profunda. A la vista de la signatura espectral de esta

superficie, en ausencia de efecto atmosférico la reflectancia debería ser próxima a cero. En consecuencia, los valores mínimos distintos de cero de los histogramas de las bandas afectadas han de estar causados por la radiación difusa atmosférica. El método de corrección consiste en restar de todos los ND de cada banda su ND mínimo. Para la banda k , por ejemplo:

$$ND'_{i,j(k)} = ND_{i,j(k)} - ND_{\min(k)}$$

Es importante tener en cuenta que la consideración anterior se refiere a superficies de agua con las características citadas, las cuales no siempre están disponibles en la imagen. Es relativamente frecuente encontrar cuerpos de agua más o menos profunda, como la retenida en embalses o lagos, que, sin embargo, no puede considerarse *limpia* a los efectos que interesan aquí. A menudo son aguas eutrofizadas y ricas en clorofila, cuya radiancia en el visible no es tan baja como la del agua marina. En el caso de los ríos o cabeceras de embalses hay que advertir también la presencia de fondos poco profundos que contribuyen en la reflexión de la luz solar. En estos casos hay que ser prudente a la hora de restar los valores mínimos para no incurrir en una sobrecorrección. Si no se dispone de más datos, podría considerarse la posibilidad de corregir sustrayendo tan solo $2/3$ de los niveles digitales mínimos.

Correcciones sobre superficies de tierra

En numerosas ocasiones puede no encontrarse en la imagen una sola superficie de agua apta para utilizar sus ND para la corrección atmosférica. En estos casos debe recurrirse a zonas de baja radiancia, como por ejemplo ciertas manchas de bosque denso y sombreado, delimitarlas sobre la escena y extraer los ND en las distintas bandas. A partir de estos valores se

calcula la recta de regresión entre cada banda y alguna otra de referencia en la que la dispersión atmosférica tenga menor repercusión, por ejemplo el rojo.

La ordenada en el origen de la recta de regresión indica el valor a sustraer de todos los ND de esa banda.

La vegetación densa y pobremente iluminada, como es la de los bosques situados en laderas orientadas al norte en el hemisferio boreal y en ladera sur en el austral, tiene baja reflectancia en el visible. Esta cualidad es ampliamente validable para bosques en cualquier latitud y época, salvo, por motivos evidentes, los caducifolios en invierno, y se encuentra recogida en la literatura especializada. Para ángulos cenitales solares próximos a 40° , la reflectancia de las praderas oscila entre el 2,6 % en el azul y el 3,9 % en el rojo, en tanto que para los bosques caducifolios está entre el 1 y el 2,5 % en ambos canales. Cualquier cálculo de reflectancia sobre superficies semejantes que se aleje de estas cifras deberá ser entendido como presencia de efecto atmosférico.

Para la aplicación del método solamente es necesaria una imagen que contenga alguna zona de vegetación densa y sombreada.

Corrección atmosférica en el análisis multitemporal

Los estudios multitemporales exigen comparar imágenes con similar influencia atmosférica, dado que cualquier diferencia en la radiometría de las imágenes imputable a efecto atmosférico podría ser interpretada como un cambio, aun cuando en realidad la superficie pueda haber permanecido inmutable.

La corrección puede hacerse de modo relativo determinando la reflectancia de una serie de cubiertas de reflectancia invariable a lo largo del año y calculando

los coeficientes lineales de ganancia y sesgo de esas cubiertas en cualquier fecha con respecto a la escena de referencia. Para normalizar radiométricamente la reflectancia a los valores de la fecha de referencia se aplica la transformación inversa con esos mismos coeficientes para todas las celdas de la imagen.

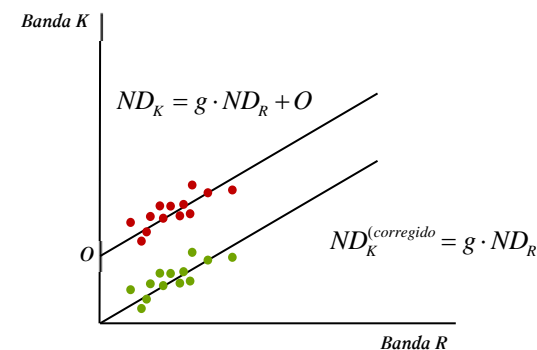
En el capítulo dedicado a la teledetección multitemporal se dan más detalles de los procedimientos a emplear.

Corrección atmosférica rigurosa

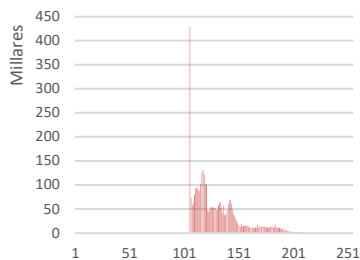
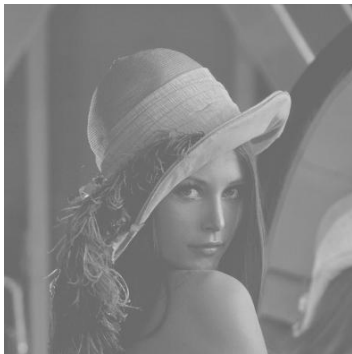
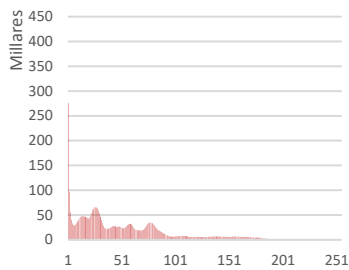
La corrección atmosférica rigurosa requiere tener un conocimiento preciso de los parámetros atmosféricos sobre la escena en el momento de la captura de la imagen. Los más importantes son el coeficiente de extinción, el espesor óptico, el albedo por dispersión simple y la función de fase de dispersión, y pueden obtenerse a partir de distintas fuentes:

- **Climatología.** Se pueden hacer estimaciones para un determinado lugar y fecha a partir de bancos de datos tabulados de las características atmosféricas para cada estación del año.
- **Medidas de campo.** Se miden los parámetros atmosféricos mediante observaciones fotométricas direccionales del sol y por recolección de partículas en suspensión en las altas capas de la atmósfera con aviones o globos.
- **Medidas sobre la imagen.** El albedo por dispersión simple y la función de fase de dispersión, pueden medirse sobre imágenes con líneas de fuerte contraste, como por ejemplo son los contactos tierra-agua.

Existen modelos de corrección atmosférica muy precisos basados en algoritmos que comparan la reflectancia de zonas de alto contraste.



Corrección por regresión sobre otra banda poco influida por el efecto atmosférico.



Imágenes ocupando todo el rango dinámico y con un contraste reducido.

Realce de imágenes

Se conocen como operaciones de realce de imágenes las diversas técnicas orientadas a mejorar o acentuar la calidad de las imágenes con el objetivo de facilitar el análisis visual, sobre papel o monitor, o para perfeccionar la producción cartográfica.

Se puede hablar de dos tipos de realce, el radiométrico, basado en tareas de expansión del contraste, y el geométrico, sustentado por el filtrado digital.

Operaciones con el contraste

El contraste es un indicativo de la amplitud del intervalo de ND presentes en una imagen, de entre la gama completa de valores que podría ofrecer.

Existen varias formas de medir el contraste en una determinada banda de la imagen:

$$c_1 = ND_{\max} / ND_{\min}$$

$$c_2 = ND_{\max} - ND_{\min}$$

$$c_3 = \sigma_{ND}^2$$

La primera de ellas, el cociente entre el ND máximo y el mínimo da una idea relativa del rango dinámico de la imagen. La diferencia proporciona la misma información, pero desde un punto de vista absoluto. Finalmente, la varianza del histograma de la imagen informa de la dispersión de los valores a partir del valor medio. Entendido de este modo, la varianza (o en su caso, la desviación típica) no sólo es una medida del

contraste sino también de la cantidad de información contenida en la imagen.

Para una determinada resolución radiométrica, la imagen puede ocupar todo el rango dinámico, o no. En el primer caso, $ND_{\min}=0$ y $ND_{\max}=k-1$, siendo k el número de niveles digitales representables. En una gama acromática, ambos valores serán representados con un color negro y blanco, respectivamente.

Sin embargo, por diversas razones, la imagen puede no abarcar todo el rango dinámico del sistema. Téngase en cuenta que, al diseñar el sensor, debe estar preparado para adquirir imágenes en lugares muy de la superficie terrestre dispares, desde las de muy alta reflectancia, como los casquetes polares en el visible, hasta otras de reflectancia prácticamente nula, como los océanos en el infrarrojo próximo o la vegetación umbrosa en el rojo espectral. Si en una escena particular no existen esas superficies extremas, la gama de ND exhibida no ocupará todo el rango dinámico, sino sólo un conjunto de valores intermedios. En consecuencia, la celda más oscura de la imagen monocroma no será negra, sino un gris más o menos oscuro, y la más brillante tampoco será blanca sino un valor de gris más o menos claro.

Las operaciones de ajuste del contraste pueden ser realizadas sobre la imagen directamente, generando una nueva imagen de salida con los ND alterados, o bien pueden ser realizadas sobre la pantalla, simplemente alterando el brillo que corresponde a cada ND , pero sin modificar el archivo original de la imagen y sin generar una nueva.

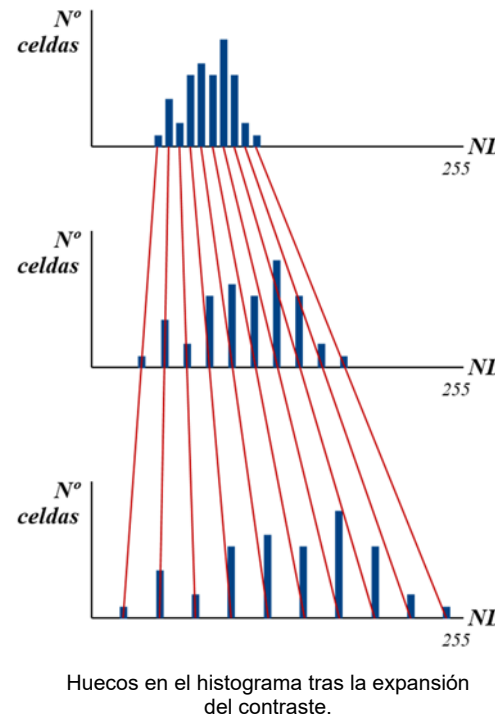
Se define el nivel visual (NV) como el brillo atribuido a una celda de imagen en el sistema de visualización o de impresión. Siempre que no haya mediado una alteración del contraste, $ND=NV$.

La relación entre ND y NV puede representarse mediante una gráfica en la que el ND se sitúa en abscisas y el NV en ordenadas. El caso anterior de ausencia de alteración viene representado por una recta de pendiente unidad. Cualquier operación que lleve aparejada una reducción del contraste implicará una relación $NV-ND$ lineal de pendiente < 1 . Por el contrario, la expansión del contraste vendrá establecida mediante una función, en este caso lineal, de pendiente > 1 .

En general no es interesante reducir el contraste de una imagen, dado que esta operación siempre lleva consigo una pérdida de información. Solamente es imprescindible realizarlo cuando el sistema de tratamiento o de visualización tiene un rango dinámico inferior a de la propia imagen. En este caso hay que reunir bajo un mismo nivel digital nuevos rangos de valores originales que implicaban más de un nivel digital. Obviamente, todas las celdas de brillo dentro de ese rango que en la imagen original eran distinguibles pasarán ahora a ser iguales.

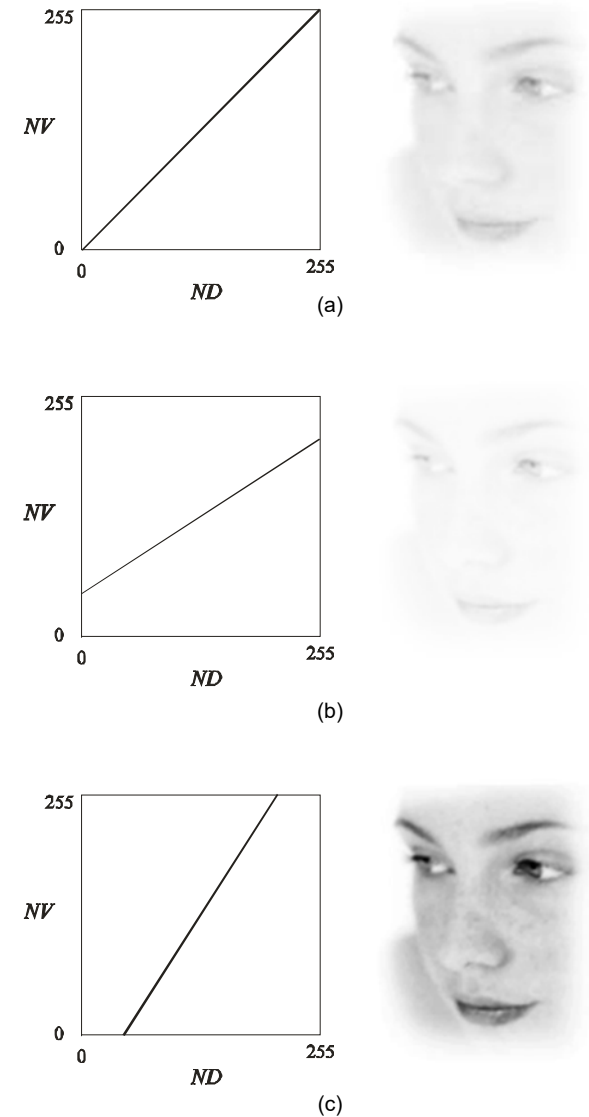
Las operaciones más frecuentes de alteración del contraste, pues, son las de expansión. Se aplican estas transformaciones para adecuar la resolución radiométrica de la imagen a la capacidad de visualización del sistema, representada por el rango dinámico de la tarjeta gráfica o del número de niveles de gris que pueda proporcionar el periférico de salida. Así, mediante la expansión del contraste de la imagen se hace corresponder un nuevo rango de ND mayor, o de NV si sólo se pretende alterar la muestra en pantalla, que le rango de ND presentes en la imagen original. Dado que el número de ND diferentes en ésta

es inferior al margen dinámico posible, cuando se haya expandido el contraste y se hayan distribuido los nuevos ND en dicho margen, podrán quedar ND sin asignar a las celdas, es decir, el histograma tenderá a presentar huecos.



Expansión lineal

El más general de los casos de expansión de contraste es el lineal. Mediante este tipo de expansión, en una gama de brillo z entre 0 y 1 se ocupará todo el margen dinámico de la imagen de salida, asignándose el valor 0 a las celdas cuyo z sea inferior o igual a un determinado valor umbral mínimo en la imagen original (m) y el valor 1 a las celdas con z iguales o superiores a un determinado valor umbral máximo



Visualización de una imagen con (a) contraste original, (b) reducido y (c) expandido.

(M). Entre ambos extremos, se distribuirán linealmente los nuevos valores de brillo expandidos (z'):

$$z' = \begin{cases} 0 & \forall z \leq m \\ \frac{z-m}{M-m} & \forall m \leq z \leq M \\ 1 & \forall z \geq M \end{cases}$$

O, si se prefiere:

$$ND' = \begin{cases} 0 & \forall ND \leq m \cdot (k-1) \\ \frac{ND-m}{M-m} \cdot (k'-1) & \forall m \cdot (k-1) \leq ND \leq M \cdot (k-1) \\ k'-1 & \forall ND \geq M \cdot (k-1) \end{cases}$$

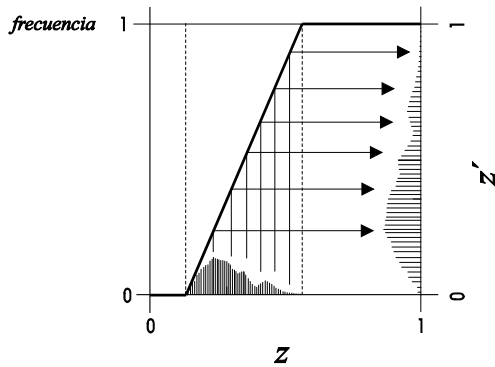
La expresión de ND' para los valores de ND comprendidos entre m y M puede descomponerse en dos sumandos, de modo que el nivel digital resultante queda expresado:

$$ND' = \frac{k'-1}{M-m} ND - \frac{(k'-1) \cdot m}{M-m} = g \cdot ND + s$$

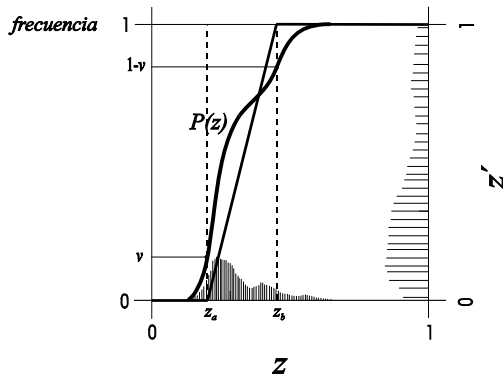
Donde la ganancia (g) y el sesgo (s) son los factores multiplicativo y aditivo de la transformación lineal:

$$g = \frac{k'-1}{M-m}; \quad s = -\frac{(k'-1) \cdot m}{M-m}$$

La forma en cómo se establezcan los valores m y M da lugar a los distintos tipos de expansión lineal.



Expansión lineal sin saturación.



Expansión lineal con saturación.

Máxima expansión lineal sin saturación

Con esta expansión se estira el histograma de la imagen original hasta hacer cero su ND mínimo y $k'-1$ su ND máximo, consiguiendo que la gama de grises primitiva, que no ocupaba todo el margen dinámico pase ahora a discurrir entre el negro y el blanco. Para ello, en la expansión genérica indicada anteriormente se fijan los valores m y M de esta forma:

$$m = ND_{\min} \\ M = ND_{\max}$$

$$ND' = \begin{cases} 0 & \forall ND \leq ND_{\min} \\ \frac{ND-ND_{\min}}{ND_{\max}-ND_{\min}} \cdot (k'-1) & \forall ND_{\min} \leq ND \leq ND_{\max} \\ k'-1 & \forall ND \geq ND_{\max} \end{cases}$$

Una forma intuitiva de presentar la expansión del contraste consiste en construir una gráfica en la que se presenta el histograma de la imagen original en abscisas. Sobre él se dibuja la función de expansión determinada por las expresiones anteriores y se hace *reflejar* la situación de cada ND (o z , en su caso) sobre ella para terminar proyectándola sobre el eje auxiliar de ordenadas, en el cual aparecerán los nuevos valores de ND' asignados. En el caso de la expansión sin saturación, dado que la recta de expansión tiene su pie en el valor mínimo m y la cabeza en el valor M máximo, del histograma resultante ocupará el margen dinámico completo, entre 0 y 1 si se habla de valores z' o entre 0 y $k'-1$ si se habla de ND .

Expansión lineal con saturación

Puede darse el caso de que una sola o muy pocas celdas de la imagen ocupen ND muy oscuros, incluso el propio valor 0, bien porque realmente existan esos

niveles de brillo real en la imagen, o bien porque han podido *perderse* celdas en el transcurso de la adquisición o grabación. Igual puede suceder con los valores altos del histograma. Unas pocas celdas pueden ser blancas completamente o tener un valor muy cercano al máximo. En cualquiera de los casos la expansión lineal anterior no proporcionará buenos resultados porque el margen dinámico ya está ocupado, aunque sea por muy pocas celdas. Si se quiere dotar de mayor margen a una gran proporción de la imagen aplicando una expansión lineal no queda otra opción que escoger otros umbrales m y M diferentes de los ND mínimo o máximo. Si se eligen valores del histograma que dejen a su izquierda un determinado porcentaje de celdas más oscuras y otro similar de celdas más claras a su derecha y se expanden linealmente los valores comprendidos entre m y M , se dice que la expansión es lineal con una saturación igual a ese porcentaje.

Para implementar el cálculo de la forma más fácil se recurre a construir el histograma acumulativo de la imagen y transformarlo en porcentaje sobre el total de celdas a partir del número de celdas acumuladas. De este modo, seleccionando la ordenada del histograma se podrá deducir directamente en abscisas el nivel digital que deja por debajo el porcentaje elegido de celdas más oscuras. De un modo equivalente se procede para las celdas brillantes.

Llamando v a la proporción de celdas que se desea saturar, en tanto por uno, y P al histograma acumulativo en valores de tanto por uno, siendo uno el total de celdas de la imagen:

$$m = z_a \quad (\neq z_{\min}) \quad \forall z < z_a \quad P(z) < v$$

$$M = z_b \quad (\neq z_{\max}) \quad \forall z > z_b \quad P(z) > 1 - v$$

en donde z_a y z_b son los valores de brillo de la imagen, o en su caso los niveles digitales, correspondientes a las proporciones v y $1-v$ del histograma acumulativo.

En la práctica este tipo de expansión permite ganar mayor margen dinámico para un cierto intervalo de niveles digitales originales, en detrimento de los valores más oscuros o más claros, que ahora quedarán saturados como negro y blanco respectivamente.

Expansión con saturación basada en la distribución normal

Este procedimiento de expansión satura los valores de gris basándose en los parámetros de la distribución normal. Asumiendo que el histograma se ajusta a este tipo de distribución, se eligen los valores m y M como un número entero de desviaciones típicas hacia la izquierda y hacia la derecha del valor medio del histograma, respectivamente:

$$m = \mu - n\sigma$$

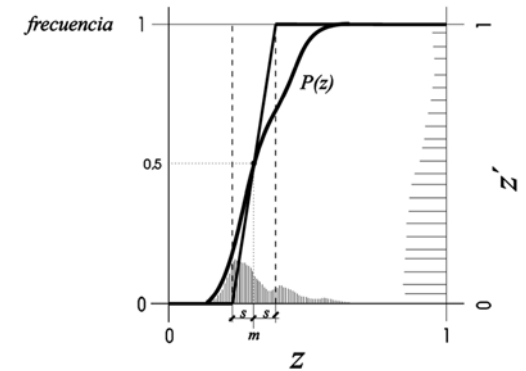
$$M = \mu + n\sigma$$

Siendo μ la media aritmética de los ND , σ la desviación típica de la distribución y n un número entero elegido por el analista.

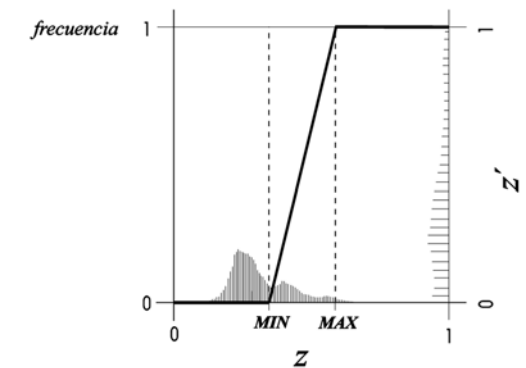
Expansión con saturación por selección directa

Indudablemente, la expansión con saturación puede realizarse seleccionando directa y arbitrariamente los valores umbral m y M , de tal modo que se consiga así destacar el rango de niveles digitales que interese.

Es frecuente necesitar estudiar zonas de la imagen



Expansión lineal con saturación basada en la distribución normal.



Expansión lineal con saturación por selección directa de los umbrales.

que resulta excesivamente oscura para el análisis visual, como por ejemplo las superficies marinas o la vegetación en el visible. En estos casos es útil recurrir a esta modalidad de expansión, aunque durante el proceso no se puedan tomar decisiones sobre el resto de la imagen.

Expansión ajustada a estadísticos

Hay ocasiones en las que se requiere que la imagen adquiera un determinado contraste y un brillo medio establecido. Tal es el caso de las tareas de *matching* de histogramas, esto es, la obtención de dos imágenes con igual media y varianza en sus *ND*, de modo que operaciones como el mosaico puedan ser llevadas a cabo sin que se perciba la desigualdad inicial de los contrastes.

Para obtener una imagen cuyo histograma presente una media y una varianza elegida con anterioridad se recurre a la siguiente transformación:

$$z' = \frac{\sigma'}{\sigma} (z - \mu) + \mu'$$

en donde μ y σ son el valor medio y la varianza de la imagen que se desea ajustar μ' y σ' los estadísticos de referencia a conseguir.

Ecualización del histograma

En teoría de señal, la ecualización es una operación que se realiza sobre una señal destinada a conseguir que su espectro de frecuencias sea uniforme, es decir, en el que la intensidad de todos los componentes sea igual, sea cual sea la frecuencia básica de éstos. Esto supone alterar los coeficientes de Fourier de la señal en el dominio de la frecuencia, o lo que es lo mismo,

potenciar las amplitudes de los componentes de baja o de alta frecuencia.

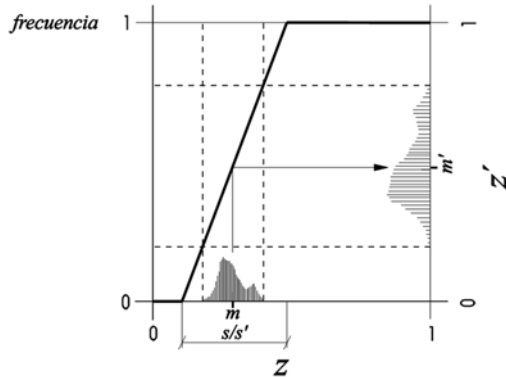
Por analogía, en el procesamiento digital de imágenes, la ecualización del histograma de una imagen tendría como objetivo obtener una nueva imagen cuyo histograma tuviera una distribución uniforme, es decir, que existiera el mismo número de celdas para cada nivel digital. Sin embargo, en la práctica esto no es posible, dado que la función de niveles de gris, el histograma, es una función discreta, a diferencia de lo que sucede en sonido.

En su lugar, y con el objeto de que el balance de niveles de gris sea más equilibrado lo que se hace en realidad es dotar de mayor margen dinámico a los *ND* más frecuentes en la imagen, que suelen ser los de gama media, en detrimento de los muy bajos o muy altos, a los que se les roba parte de su margen dinámico inicial. El resultado maximiza el contraste de la imagen sin perder información de tipo estructural, es decir, que si en la imagen original dos celdas exhiben diferentes *ND*, tras la ecualización seguirán teniendo niveles de gris diferentes, aunque con otros valores.

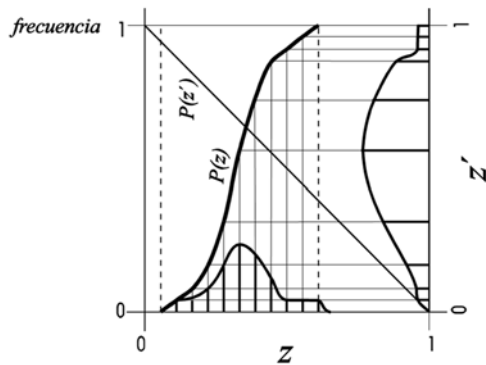
La función de expansión es el propio histograma acumulativo $P(z)$. Esta función se caracteriza por tener mayor pendiente cuanto mayor sea el número de celdas de cada nivel digital. Como se utiliza para *reflejar* en nivel digital original sobre el eje de ordenadas auxiliar, la separación entre los nuevos niveles digitales será mayor cuanto los *ND* originales sean más frecuentes.

$$z' = P(z)$$

Una propiedad de esta modalidad de ecualización es que una segunda ecualización del histograma previamente ecualizado proporciona la misma imagen, sin



Expansión lineal ajustada a estadísticos prefijados.



Expansión mediante ecualización del histograma.

variación del contraste, dado que la abscisa y la ordenada de cada punto de la nueva función $P(z')$ son iguales y convierten a esta función en una recta de pendiente unidad.

Expansión especial

Cuando se necesite expandir con diferente amplitud dinámica unas zonas y otras de la imagen se puede recurrir a emplear una función que mantenga pendientes locales seleccionadas por el analista. Para ello en lugar de una recta se puede utilizar una polilínea de expansión, con una selección arbitraria de los puntos de cambio de pendiente. De este modo se consiguen distintos márgenes dinámicos para las zonas de interés.

Binarización

La binarización es un caso extremo de la expansión con saturación en la que los valores m y M se hacen coincidir en un solo valor umbral u y se adopta una función escalonada. Desde este nivel hacia la izquierda, todas las celdas adquieren el valor 0 y todas las de la derecha, incluidas las que ostenten el valor umbral, adquieren el valor 1. En la práctica la imagen original queda convertida en una imagen binaria de celdas negras y blancas configurada a partir del valor umbral elegido.

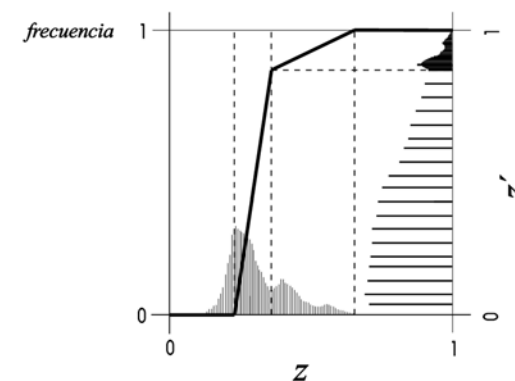
$$z' = \begin{cases} 0 & \forall z < u \\ 1 & \forall z \geq u \end{cases}$$

Es el procedimiento para construir máscaras en una imagen. Si, por ejemplo, se desea obtener una máscara del contorno continental, eliminando de la imagen la zona marina, es posible construir una máscara binaria a seleccionando el umbral en la imagen tér-

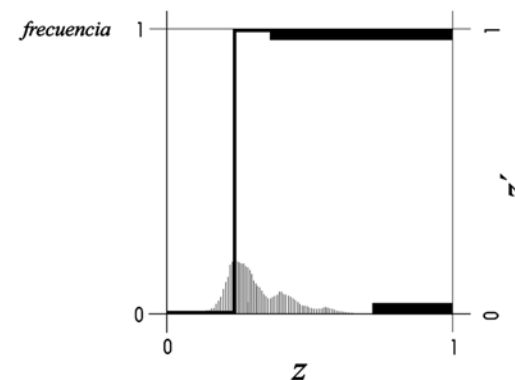
mica mediante una temperatura de corte, o en el infrarrojo próximo, mediante un nivel de brillo bajo que sólo aisle el agua. Para emplear la máscara obtenida se multiplicará por la imagen original, eliminando de este modo la zona no deseada, ya que las celdas de esta zona en la máscara tendrán valor 0, en tanto que el resto valdrá 1.

Inversión

La inversión del contraste se lleva a cabo mediante una recta de expansión de pendiente negativa. La inversión puede realizarse aprovechando todo el margen dinámico o se puede ajustar a cualquiera de los valores explicados anteriormente.



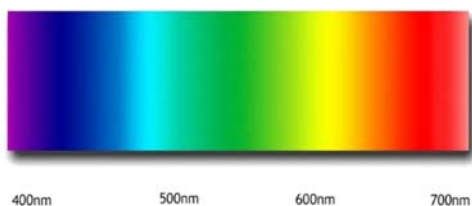
Expansión especial con saturación en los niveles bajos y con dos intensidades diferentes.



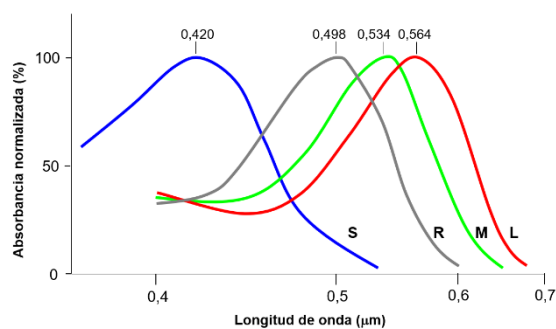
Binarización del histograma.

9

Empleo del color



Longitudes de onda del espectro visible.



Curvas de sensibilidad espectral de las células receptoras de la retina humana: conos S, M y L y bastones R.

Introducción

Las imágenes de satélite pueden ser monobanda o multispectrales. Esta particularidad hace que con ellas se pueda representar la escena de dos modos diferentes. Si la imagen consiste en una sola banda, o bien se utiliza solamente una de ellas cuando la imagen es multispectral, la representación será en gama de grises. Si, por el contrario, se aprovechan

distintos canales de la imagen, entonces la representación puede ser en color. En este capítulo se verá sucintamente la teoría del color, los modos de especificación y la aplicación del color en el campo de la representación de imágenes monocanal y multispectrales.

El color

El espectro visible abarca el intervalo espectral comprendido entre 400 y 700 nm. Las ondas más cortas de ese segmento corresponden al azul espectral y las más largas al rojo. Es el intervalo de luz capaz de estimular las células retinianas humanas.

La tonalidad con la que se percibe cada intervalo es la interpretación que el cerebro realiza cuando la retina captura luz en las distintas longitudes de onda. Dicho de otra forma, el azul, verde o rojo no tienen existencia como tales. No son *reales*. Lo único ciertamente real es una luz de 420 nm, que el ojo interpreta como azul, de 560 nm, que se interpreta como verde, o 650 nm que sería identificada como roja.

La capacidad de discriminar estas longitudes de onda del visible les proporciona a los animales la posibilidad de adquirir información relevante de su entorno muy relacionada con su seguridad. Dentro de la teledetección y, en general, del análisis de imágenes, será aprovechada para interpretar la información que encierran las imágenes y para codificar detalles, sintetizando en una sola imagen los procedentes de varias.

Esta habilidad visual se basa en la existencia de diferentes familias de células sobre la retina. Entre las principales destacan los bastones y los conos y, dentro de éstos, tres tipos diferentes: los conos S, M y L, referencia a la longitud de onda a la que son más sensibles (*short, medium y large*). Cada uno de estos tipos está especializado en capturar luz de unas determinadas longitudes de onda gracias a sus fotopigmentos cianopsina, cloropsina y eritropsina, respectivamente. Los conos S absorben radiación con un pico centrado en torno a 420 nm, los conos M en 534 nm y los conos L en 564 nm. Cada luz es transformada en una señal nerviosa que, interpretada en el cerebro, se corresponde con el azul, el verde y el rojo. La respuesta espectral de los bastones es, por el contrario, más inespecífica, con una curva de sensibilidad espectral en torno a 498 nm.

Tal como se deduce al examinar las curvas de sensibilidad espectral, la respuesta de cada tipo de conos no es muy coherente, es decir, en lugar de responder al estímulo de un intervalo reducido de longitudes de onda, abarca un rango absorción de radiación de

gran anchura de banda. Esto permite que los conos M, si bien son estimulados con mayor eficiencia por la luz de 530 nm, también envíen señales eléctricas al cerebro cuando son iluminados con luz de 450 o

de 600 nm, aunque sea más débilmente. Esta circunstancia posibilita la síntesis del color mediante la teoría aditiva de formación del color, como se verá más adelante.

Síntesis del color

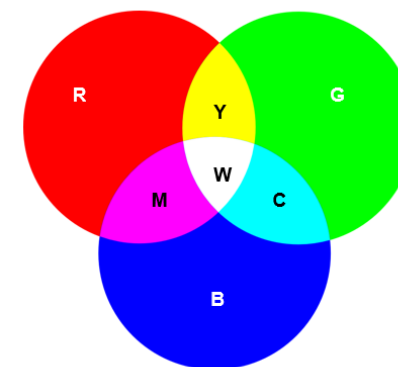
Existen dos procedimientos para la formación del color a partir de colores simples o primarios, la síntesis aditiva y la sustractiva.

En la síntesis aditiva se parte de los colores básicos R, G y B. Mediante haces de luz monocromática de las longitudes de onda que corresponden a esos colores, las distintas mezclas proporcionan los colores secundarios. La mezcla en igual proporción de R y B da como consecuencia un nuevo color que el cerebro interpreta como magenta (M). La mezcla de G y B produce cian (C) y, finalmente, la suma de R y G origina el amarillo (Y). A los efectos de la interpretación que el sistema ojo-cerebro da de los colores, y debido a las curvas de sensibilidad de los conos retinianos es lo mismo iluminar el ojo con una mezcla de B y G que hacerlo con una luz monocromática de una longitud de onda intermedia, 460 nm, por ejemplo, es decir, cian. La suma de los tres colores básicos en la síntesis aditiva proporciona el color blanco. La luz solar es blanca porque contiene todas las longitudes de onda del espectro visible y el color blanco en una pantalla iluminada mediante un cañón de proyección se consigue mediante la mezcla de radiación roja, verde y azul exclusivamente. Este sistema de formación de color es el útil en informática y en todas las operaciones que se realicen sobre un monitor.

La síntesis sustractiva requiere de una fuente de iluminación blanca que contenga todas las longitudes de onda y de un soporte que las refleje todas por igual, esto es, que se perciba como blanco. Si sobre el soporte se vierte una tinta capaz de absorber todas las longitudes de onda excepto las correspondientes al C, la superficie se verá de color cian. La tinta Y será la que tenga la propiedad de absorber todas las longitudes de onda excepto las correspondientes al amarillo espectral y lo mismo cabe decir para la tinta M. Si se mezclan tintas C e Y, la radiación emergente de la superficie, siempre iluminada con luz blanca, contendrá las longitudes de onda del cian (entre el azul y el verde) y del amarillo (entre el verde y el rojo), y la retina la interpretará como si de G se tratara. De forma análoga, la mezcla Y+M produce R y la de M+C proporciona B. Finalmente, la mezcla aditiva de las tres tintas básicas C+M+Y teóricamente produce negro, puesto que todas las longitudes de onda serían absorbidas. En la realidad, como las tintas no son cuerpos negros, alguna reflectancia tienen y el color resultante de la mezcla a tres no es absolutamente negro. Por esta razón en artes gráficas, donde es más útil este sistema de síntesis de color, el negro se refuerza con la adición de tinta negra, dando lugar al sistema de especificación de color CMYK, donde el K designa al negro añadido.

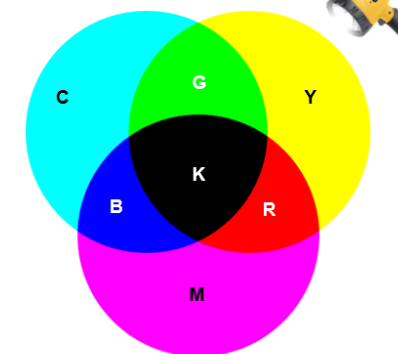
Síntesis aditiva

Adición de luces monocromáticas



Síntesis sustractiva

Adición de tintas monocromáticas sobre una superficie blanca iluminada con luz blanca



Síntesis aditiva y sustractiva del color en los sistemas RGB y CMYK, respectivamente.

Paletas de color

Se llama Nivel visual (NV) al nivel de gris que le atribuye a cada ND de la imagen en una representación gráfica, bien sea sobre un monitor o sobre algún soporte de impresión. La paleta de color es una tabla de referencia mediante la cual se define el color que se aplica a cada ND. Esta expresión equivale a otras denominaciones habituales como *Color Look-up Table* (CLUT), tabla de color, etc.

Las paletas pueden ser:

- Monocromáticas: constituidas por una gama de grises.
- Cromáticas: elaboradas con mezclas de los tres colores primarios rojo (R), verde (G), azul (B), cuyas abreviaturas proceden de sus nombres en inglés.

Dependiendo de la resolución radiométrica de la imagen, a cada ND procesable (64, 256, 1024, etc.) se le hace corresponder un cierto nivel de R, de G y de B.

En imágenes cuantitativas monobanda es habitual emplear paletas monocromáticas, en una gama de grises que puede abarcar desde el negro hasta el blanco, o bien una rampa cromática que incluya proporciones de color en mezclas bien elegidas, de modo que se mantenga intuitivamente la idea de cantidad

Por el contrario, en imágenes cualitativas monobanda, como pueden ser las imágenes clasificadas, los mapas temáticos, etc., es preferible emplear paletas cromáticas que el cerebro no asocie a la idea de *mayor-menor*, sino a *diferente*.

ND	nivel de tono primario (NV)			color resultante
	Rojo	Verde	Azul	
0	0	0	0	negro
1	255	0	0	rojo
2	0	255	0	verde
3	0	0	255	azul
4	255	255	0	amarillo
5	255	0	255	magenta
6	0	255	255	cian
7	128	0	0	marrón
8	128	128	0	caqui
9	128	0	128	púrpura
10	0	128	0	verde oliva
11	0	128	128	azul verdoso
12	0	0	128	azul marino
13	255	128	0	naranja
14	128	255	0	verde limón
15	255	255	255	blanco

Ejemplo de paleta cualitativa asignada a una imagen de 4 bits de resolución radiométrica.

Imágenes en color compuesto

Las composiciones en color de imágenes constituyen caso particular de combinación lineal analógica en el periférico de visualización, pantalla o impresora. El color se forma sumando las contribuciones de brillo de tres bandas de una imagen multispectral, a cada una de las cuales se le asigna uno de los tres tonos primarios, R, G o B.

Considerando el espacio tridimensional RGB que generan tres ejes ortogonales cuyas variables son los tres colores básicos, la luminancia cromática de una celda es:

$$\vec{z} = R z_1 \vec{r} + G z_2 \vec{g} + B z_3 \vec{b}$$

en donde \vec{z} es el color resultante, z_1 , z_2 y z_3 , son los valores de la celda en las tres bandas elegidas y \vec{r} , \vec{g} y \vec{b} son los vectores unitarios de los ejes cromáticos.

Cualquier color posible queda especificado en el sistema RGB componiendo en el periférico de salida la suma de las señales que alimentan los componentes rojo, verde o azul. En el caso de las imágenes en color compuesto se utilizan los niveles digitales de tres

bandas, elegidas de entre las posibles de acuerdo a cualquier criterio, para alimentar los cañones R, G y B. Niveles digitales bajos de una cierta banda proporcionarán luminancias bajas en el color que se le asigne, niveles altos producirán un efecto de prevalencia de ese color sobre los demás. Si los niveles digitales son altos en las tres bandas el color resultante de la celda tenderá al blanco.

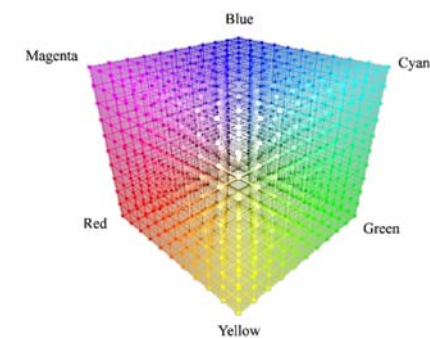
Se pueden generar imágenes en color compuesto utilizando cualesquiera de las bandas disponibles de una imagen multispectral de más de tres bandas, pero de entre todas, las combinaciones en *color verdadero* son las que utilizan las bandas que por su longitud de onda pertenecen al rojo, verde y azul espectral para alimentar los cañones R, G y B, en ese orden. Cualquier otra combinación se denomina de *falso color*. Entre ellas, la más difundida por haber sido la primera en usarse es la correspondiente a infrarrojo próximo, rojo y verde espectral en el orden R-G-B. La razón era que por entonces los sensores carecían de canal azul. En esta combinación es característico el color rojo de la vegetación, tanto más intenso cuanto más vigorosa sea. Si, en cambio, se desea proporcionar una impresión de verdor intenso para la vegetación vigorosa, es útil recurrir a combinaciones del tipo infrarrojo medio-infrarrojo próximo visible para el orden R-G-B.

Cuando se sintetiza en una sola imagen en color compuesto la información procedente de tres bandas, se está prescindiendo de la que queda en el resto de las bandas no utilizadas de la imagen. Si no se tiene un requerimiento expreso para la elección de bandas, surge entonces la pregunta de cuáles escoger de modo que se pierda la menor cantidad de informa-

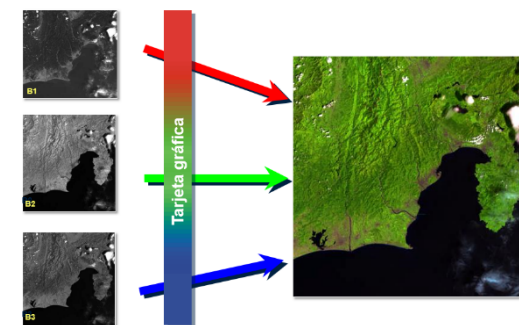
ción posible. Para responder eficazmente a esta pregunta es necesario reparar en que el conjunto de bandas de una imagen suele presentar redundancia en su información. Cuanto más cercanos sean los intervalos de longitud de onda de dos bandas, mayor parecido habrá en la respuesta de una superficie. Esto se manifiesta en una gran correlación entre sus niveles digitales. Si se pretende generar una imagen en color compuesto, es conveniente que la información contenida en cada una de las tres bandas sea lo más independiente posible. De otra parte, si ha de elegirse una banda de entre varias redundantes, es ventajoso recurrir a la que tenga un contraste más acusado, puesto que eso es signo de una gran varianza, lo cual quiere decir que sus niveles digitales presentan valores muy dispersos en torno a la media y, en definitiva, esa banda aporta más información que cualquier otra que exhiba un rango dinámico menor. Así, pues, la combinación preferible será la que contenga tres bandas de gran varianza y, simultáneamente, muestre una menor correlación entre ellas. Este doble requerimiento queda sintetizado en el *Factor de Índice Óptimo* (*Optimum Index Factor*, OIF):

$$OIF = \frac{\sum_{i=1}^3 s_i}{\sum_{i,j=1}^3 ABS(r_{i,j})}$$

donde s_i son las desviaciones típicas de cada una de las bandas involucradas y $r_{i,j}$ son los coeficientes de correlación de todos los pares posibles de bandas.



El espacio de color RGB.



Generación de una imagen en color compuesto.

Imágenes en pseudocolor

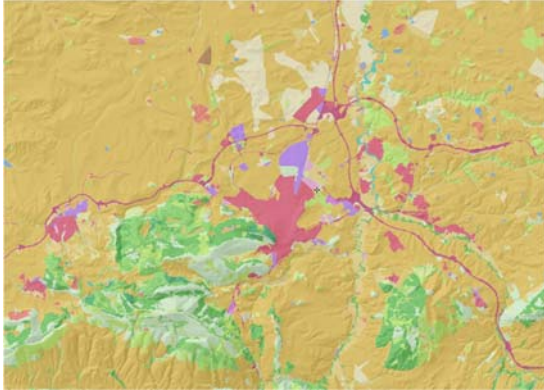


Imagen en pseudocolor mostrando una clasificación de usos del suelo.

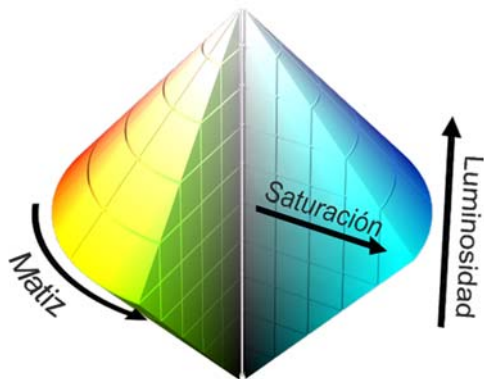
Aunque el prefijo *pseudo* significa *falso*, en el ámbito de aplicación del color a las imágenes se utiliza para referirse a la aplicación de una paleta cromática a una imagen monobanda, el lugar aludir a las imágenes en falso color que, como se ha visto, son imágenes en color compuesto. Las imágenes en pseudocolor están constituidas por una sola banda a la que, siguiendo algún criterio de equivalencia entre NV y ND, se le aplica una paleta cromática.

En general, es una opción muy adecuada para representar imágenes cualitativas como por ejemplo una clasificación de usos del suelo. En ciertas ocasiones se puede aplicar también a imágenes cuantitativas, siempre que la elección de la paleta sugiera una cierta graduación de los valores representados. Así

pues, los mapas térmicos recurren a paletas con rampas de color desde tonos fríos como los azules hasta colores muy cálidos como los rojos, pasando por las gamas intermedias. Por el contrario, en los mapas cualitativos la mejor paleta es la que solamente trasladada al usuario la idea de diferencia, no de graduación.

Si se utiliza una imagen monobanda cuantitativa cruda o en la que no es muy inequívoca la asignación de una determinada paleta cromática, es siempre preferible presentarla en gama de grises, puesto que cualquier usuario sin mayor preparación comprenderá la equivalencia entre los tonos oscuros con valores bajos de la celda y entre tonos muy brillantes con los de nivel digital elevado.

Otros sistemas de especificación del color



El sistema IHS.

Existen otros sistemas de especificación de color además del RGB. Entre los más difundidos y utilizados en teledetección se encuentra el llamado IHS (*Intensity-Hue-Saturation*; Intensidad-Tono-Saturación), también denominado HLS (*Hue-Saturation-Lightness*). Como cualquier sistema de especificación, incluido el RGB, también necesita tres componentes para especificar sin ambigüedad cualquier color. Se representa gráficamente como un doble cono o un doble hexágono en el que el vértice inferior corresponde al negro y el superior al blanco. Sobre el eje del doble cono se mide la variable Intensidad o Luminosidad. Las variables, H y S se establecen en coordenadas polares sobre planos perpendiculares al eje

Intensidad. El tono se mide angularmente y la distancia desde el eje a un color determinado queda fijada por la Saturación.

De este modo, la gama de grises vendrá establecida por el eje del doble cono y el tono predominante estará determinado por la generatriz que se elija. El color final podrá oscilar entre el más puro, completamente saturado y situado sobre la superficie del cono, y otro color que irá perdiendo saturación a medida que se desplace horizontalmente hacia el eje, hasta convertirse en un gris de la misma luminosidad que el color saturado del que partió.

Se puede pasar del espacio RGB al IHS realizando una rotación de los ejes cromáticos. Considerando $0 \leq R \leq 1$; $0 \leq G \leq 1$; $0 \leq B \leq 1$, las funciones de transformación son las siguientes:

$$I = R + G + B$$

$$H = \begin{cases} (G - B)/(I - 3B) & \text{si } B = \text{mín}(R, G, B) \\ (B - R)/(I - 3R) + 1 & \text{si } R = \text{mín}(R, G, B) \\ (R - G)/(I - 3G) + 2 & \text{si } G = \text{mín}(R, G, B) \end{cases}$$

$$S = \begin{cases} (I - 3B)/I & \text{si } 0 \leq H \leq 1 \\ (I - 3R)/I & \text{si } 1 \leq H \leq 2 \\ (I - 3G)/I & \text{si } 2 \leq H \leq 3 \end{cases}$$

De igual forma, puede arbitrarse un algoritmo para volver del espacio IHS al RGB, siendo $0 \leq I \leq 1$; $0 \leq H \leq 3$; $0 \leq S \leq 1$:

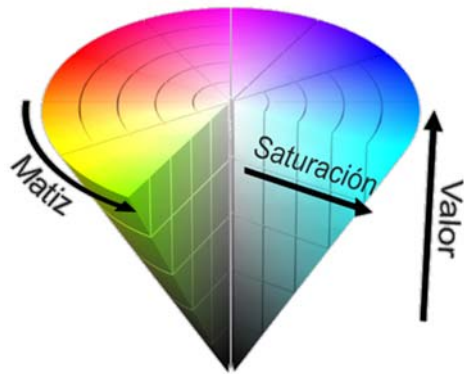
$$R = \begin{cases} I(1 + 2S - 3HS)/3 & \text{si } 0 \leq H \leq 1 \\ I(1 - S)/3 & \text{si } 1 \leq H \leq 2 \\ I(1 - S + 3S(H - 2))/3 & \text{si } 2 \leq H \leq 3 \end{cases}$$

$$G = \begin{cases} I(1 - S + 3HS)/3 & \text{si } 0 \leq H \leq 1 \\ I(1 + 2S - 3S(H - 1))/3 & \text{si } 1 \leq H \leq 2 \\ I(1 - S)/3 & \text{si } 2 \leq H \leq 3 \end{cases}$$

$$B = \begin{cases} I(1 - S)/3 & \text{si } 0 \leq H \leq 1 \\ I(1 - S + 3S(H - 1))/3 & \text{si } 1 \leq H \leq 2 \\ I(1 + 2S - 3S(H - 2))/3 & \text{si } 2 \leq H \leq 3 \end{cases}$$

Además de las aplicaciones puramente cromáticas que puede tener la transformación IHS, en teledetección es muy útil para fusionar imágenes de diferente resolución espacial, bien adquiridas por el mismo instrumento o bien procedentes de sensores diferentes. Habitualmente se utiliza esta transformación para reunir en una sola imagen las características cromáticas propias de una imagen en color compuesto, procedente de un sensor multiespectral de menor resolución, con la riqueza de detalles espaciales propia de una imagen pancromática de mayor resolución espacial.

Para llevar a cabo la fusión, en primer lugar, deben corregirse geoméricamente las dos imágenes de la forma más rigurosa posible para permitir su superposición exacta. En un segundo paso, se ha de remuestrear la imagen de menor resolución al tamaño de la celda de la otra.



El sistema HSV.

Se transforma la imagen RGB en sus componentes IHS y se sustituye la banda Intensidad por la imagen pancromática de alta resolución. Por último, se revierte la transformación del espacio de color aplicando su inversa y consiguiendo de nuevo una imagen multispectral con tres bandas correspondientes a las coordenadas RGB. En la nueva imagen permanecerán los tonos originales de la imagen multispectral, pero se habrán añadido los detalles geométricos de la pancromática.

El sistema de especificación HSV o HSB (Tono o Matiz-Saturación-Valor o Brillo, *Hue-Saturation-Value* o *Brightness*) constituye un espacio de color similar al IHS, pero formado por un sólo cono de color invertido. El color negro se sitúa en el vértice inferior y el blanco en el centro de la base, que se presenta en la parte superior.

10

Realce geométrico

Introducción

Se entiende como realce geométrico un conjunto de operaciones destinadas a mejorar la calidad visual de la imagen alterando algunas de sus características geométricas, sin alterar sustancialmente a su radiometría. Este objetivo puede abordarse por diversos caminos:

- Modificando el tamaño de los objetos en la salida gráfica de la imagen.
- Minimizando defectos localizados o generalizados de la imagen.
- Realzando rasgos que mejoren la percepción de los elementos de la imagen.

Estas operaciones de realce geométrico son útiles para el análisis visual de la imagen sobre pantalla o sobre papel. Son deseables e incluso necesarias en numerosas ocasiones de cara a la producción cartográfica o a la presentación de los resultados de operaciones realizadas con imágenes sobre cualquier tipo de soporte, especialmente las dos últimas. Sin embargo, no deben llevarse a cabo si tras ellas se pretende obtener resultados con valor físico, puesto que comportan la alteración, siquiera mínimamente, de los ND originales.

Entre las operaciones de realce geométrico, una de las más empleadas, aunque no la única, es sin duda el filtrado digital.

Cambios de escala

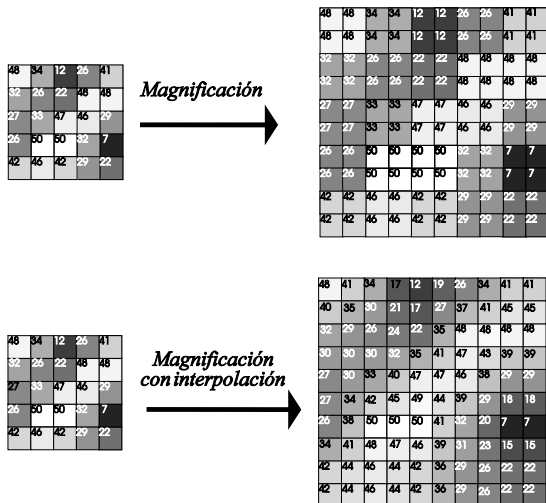
Es impropio aplicar el concepto de escala a la imagen digital, porque sólo es un registro electrónico almacenado en forma de archivo, incluso sin estructura matricial sino lineal. La imagen digital, pues no tiene escala. Sólo tiene sentido hablar de escala cuando esa imagen es representada en algún soporte plano e indeformable.

En general se prefiere utilizar el concepto de resolución, sobreentendida ésta como espacial, es decir, el tamaño nominal de la celda en el terreno y en el nadir de la imagen. Corresponde a las denominaciones de GSD (*Ground Sample Distance*, distancia de mues-

treo en tierra), GIFOV (*Ground-projected Instantaneous Field of View*, campo de visión instantáneo proyectado en tierra). Este texto utilizará preferentemente la denominación de tamaño de la celda, entendiéndose ésta en el terreno.

Existen dos tipos de cambio de escala:

- La magnificación o *zoom*. Consiste en replicar cada celda una o más veces de modo que, del pixel original de pantalla que representa cada celda en una presentación *natural* de la imagen, cada celda pase a estar representada por n^2 píxeles, siendo n el grado de magnificación. Es importante insistir



Operaciones de magnificación de la imagen.

en que, aunque la celda se perciba con mayor tamaño en la pantalla, la resolución espacial sigue siendo la misma. No se incrementa el nivel de detalle, solamente se han repetido valores de celda y se han utilizado píxeles contiguos para simular una celda más grande. Ciertos sistemas de tratamiento digital de imágenes disponen de una modalidad de magnificación con interpolación que permite suavizar las transiciones entre los valores replicados de modo que se reducen los saltos bruscos de valores. Salvo que la variable representada

en la imagen tenga alta autocorrelación, el resultado obtenido no tiene por qué ser más fiel a la imagen original.

- La reducción. Consiste en prescindir de una o más celdas siguiendo un patrón de elección, de modo que resulte una imagen reducida compuesta por una muestra de la original. Esta operación es necesaria para poder tener una visión de conjunto de una imagen cuya extensión real en número de celdas (columnas x líneas) supere el tamaño del monitor o de la salida gráfica que se desea obtener.

El filtrado digital

El filtrado digital es una transformación de la imagen mediante operadores locales que involucran a la celda que se pretende modificar y a un entorno cercano de ella. El nuevo ND de la celda filtrada será función no sólo del valor de la celda original, sino también de los ND de las celdas contiguas.

Los filtros son operadores que se aplican con objetivos opuestos, dando origen a dos grandes familias, los filtros de paso bajo, que suavizan la imagen final, aproximando los ND originales a los de sus vecinos, bien para destacar elementos de interés, y los filtros de paso alto, que exageran las diferencias y aumentan el contraste local.

Los filtros digitales pueden ser lineales y no lineales.

El filtrado digital es una operación de convolución. Se define la convolución de dos funciones unidimensionales $f_1(t)$ y $f_2(t)$ como el resultado de integrar el producto de las dos funciones después de desplazar una de ellas una distancia x , que varía entre $-\infty$ e ∞ ,

es decir:

$$f(t) = f_1(t) \otimes f_2(t) = \int_{-\infty}^{\infty} f_1(x) f_2(t-x) dx \quad (1)$$

en donde \otimes es el símbolo de la convolución.

El filtro digital es una operación de convolución entre la imagen original $z(x,y)$ y la función filtro $h(x,y)$. Pero para que esta operación sea aplicable dentro del ámbito de las imágenes digitales se requieren tres requisitos:

- Debe aplicarse bidimensionalmente.
- Debe ser una función finita.
- Debe ser una función discreta.

La convolución (1) se puede enunciar de modo bidimensional, que es más apropiada para el ámbito del tratamiento de imágenes. La convolución entre dos

funciones bidimensionales $h(x,y)$ y $z(x,y)$ es:

$$z'(x,y) = h(x,y) \otimes z(x,y) = \int \int_{-\infty}^{+\infty} h(x,y) f(x-u, y-v) dudv \tag{2}$$

donde la variable auxiliar móvil x de la ecuación (1) se ha cambiado por las variables u y v .

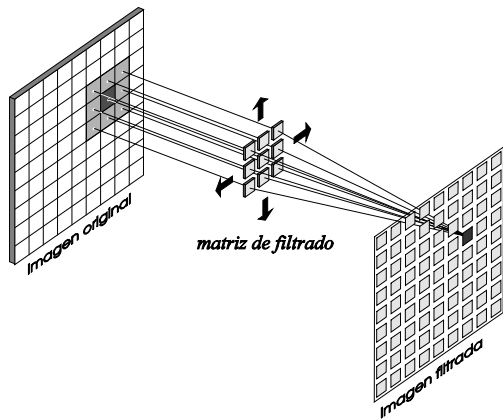
Para cumplir la segunda y tercera condición es necesario cambiar los límites de integración y acotarlos según la dimensión de la imagen y, además, ya que la imagen digital es discreta y no continua, es preciso sustituir la doble integral por una doble sumatoria. Normalmente, el filtro consiste en una matriz deslizante de dimensión impar $(2g+1) \times (2h+1)$ y su convolución con la imagen se expresa:

$$z'(m,n) = w(k,l) \otimes f(m,n) = \sum_{k=-g}^g \sum_{l=-h}^h w(k,l) f(m+k, n+l) \tag{3}$$

Donde m y n son, respectivamente, las columnas y líneas de la imagen, y k y l son las columnas y líneas de la matriz de filtrado.

Los elementos w de la matriz de filtrado se denominan coeficientes de peso. El entorno $[-g, g] \cdot [-h, h]$ es la ventana del filtro, en la cual usualmente $g = h$, es decir, la ventana suele ser cuadrada. Según sea el valor de g y h , la influencia del entorno de las celdas en el ND final podrá ser de mayor o menor alcance. Los filtros más utilizados son de 3x3 elementos, y en algunos casos de 5x5, aunque pueden aplicarse filtros mayores en determinados casos.

La aplicación de un filtro:



El filtrado digital.

$$\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \tag{4}$$

sobre una imagen $z(m,n)$ da lugar a la imagen:

$$z'(m,n) = w_{11}z(m-1,n-1) + w_{12}z(m,n-1) + w_{13}z(m+1,n-1) + w_{21}z(m-1,n) + w_{22}z(m,n) + w_{23}z(m+1,n) + w_{31}z(m-1,n+1) + w_{32}z(m,n+1) + w_{33}z(m+1,n+1)$$

Una vez obtenido el valor de $z'(m,n)$, se repite la operación para las posiciones $(m+1,n)$, $(m+2,n)$, etc. hasta terminar con la línea n y seguir con la $n+1$, y así sucesivamente hasta recorrer la imagen completa.

Conviene que la matriz de filtrado sea normalizada para no variar la luminancia general de la imagen de salida. La normalización consiste en hacer que la suma de todos sus elementos sea igual a la unidad. Si la suma de los elementos de peso de la matriz de filtrado no cumple esa condición, se normaliza dividiendo el resultado por:

$$\sum_{k=1}^3 \sum_{l=1}^3 w_{k,l} \tag{5}$$

El filtro puede ser definido por el usuario, de acuerdo a sus necesidades. Según la naturaleza y el efecto que produzcan sobre la imagen, los filtros se clasifican de diversas maneras. Entre los filtros lineales se encuentran:

- Filtros de paso bajo.
- Filtros de paso alto.
- Filtros de muesca.
- Filtros direccionales.

→ etc.

Entre los no lineales, los más habituales son:

- Filtros de moda o de mayoría.
- Filtros de mediana.
- Filtros de refuerzo o de extracción de bordes.
- etc.

Filtros de paso bajo

Los filtros de paso bajo son filtros que suavizan la imagen uniformizando cada ND original en relación con los de su entorno. La denominación *de paso bajo* es una reminiscencia del tratamiento de señales electrónicas unidimensionales, del cual deriva el tratamiento digital de imágenes habitual en teledetección. Las señales electrónicas son manipuladas en el dominio de la frecuencia y, dentro de él, los filtros se destinan a eliminar los componentes de altas o de bajas frecuencias, que en correspondencia con los sonidos transformados en señal equivalen a sonidos agudos y graves respectivamente. Los filtros de paso bajo eliminan las altas frecuencias y dejan pasar solamente las bajas y medias.

Ejemplos de filtros de paso bajo son los siguientes:

$$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \frac{1}{10} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (6)$$

El primero de ellos se denomina filtro de media porque devuelve el valor medio calculado entre el central y los ocho más próximos. Los otros dos filtros atribuyen más importancia al ND central que a su entorno inmediato y, por tanto, en la imagen de salida

se percibirá un efecto de suavizado menos intenso que con el filtro de media.

Se aplican para eliminar ruido de la imagen y para suprimir el bandedo. Se entiende por *ruido* el conjunto de errores aleatorios introducidos en la imagen durante el proceso de adquisición, transmisión o grabación y puede materializarse en píxeles anormalmente claros u oscuros salpicados de una forma más o menos aleatoria, efecto que se conoce con el nombre de *sal y pimienta*.

Este tipo de filtros son también utilizados para suavizar contrastes entre diferentes zonas de la imagen y para reducir la variabilidad de la imagen antes de proceder a una clasificación digital. Presentan en general el inconveniente de difuminar excesivamente los bordes.

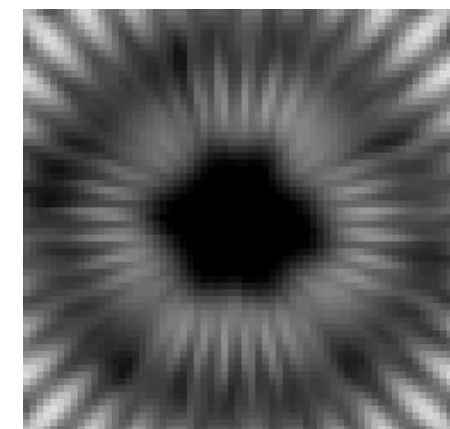
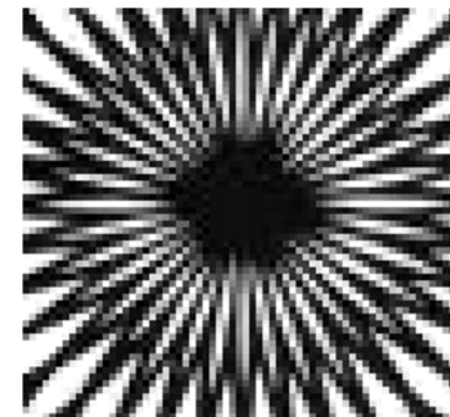
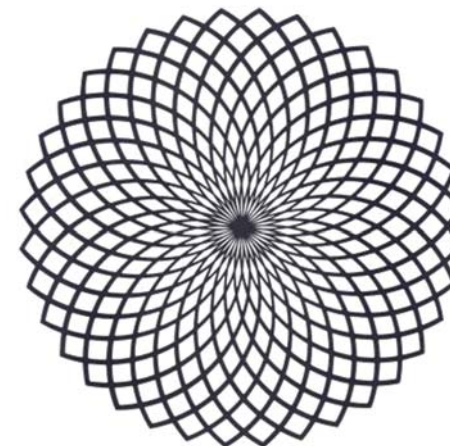
Filtros de paso alto.

Son operadores que realzan los rasgos lineales de la imagen aumentando los contrastes locales. Se llaman así porque refuerzan los componentes de alta frecuencia en la imagen, aunque pueden no eliminar los de baja frecuencia.

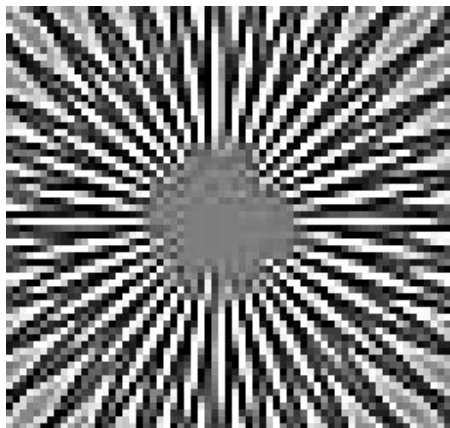
El procedimiento es el contrario al seguido anteriormente: se trata ahora de aumentar digitalmente las diferencias entre ND vecinos. Existen dos formas de obtener un refuerzo de las zonas de alta variabilidad: un procedimiento sencillo, aunque poco utilizado consiste en sumar a la imagen primitiva la diferencia entre ella misma y la producida por un filtro de paso bajo:

$$z'' = z_{i,j} + (z_{i,j} - z'_{i,j}) = 2z_{i,j} - z'_{i,j} \quad (7)$$

donde z'' es la imagen realzada, z la original y z' es



La imagen central es la parte del centro del roseón superior. La imagen inferior es el resultado de aplicarle un filtro de media.



Resultado de aplicar un filtro de paso alto.

la imagen filtrada con un filtro de paso bajo.

La otra forma es aplicar filtros de convolución de refuerzo de bordes. Para realzar las fronteras entre áreas contiguas de la imagen independientemente de su dirección, se necesita que el filtro tenga un gradiente isótropo de pesos en las ocho direcciones posibles a partir del elemento central de la matriz, esto es, que los ocho elementos adyacentes al central tengan pesos idénticos. Estos filtros se caracterizan por contener valores negativos y positivos. Si solamente se desea detectar los bordes sin mantener la imagen, la suma de todos los coeficientes periféricos, debe ser igual, pero de signo contrario al peso asignado a la celda central.

Algunos filtros comúnmente usados son:

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad (8)$$

de los cuales, el primero es conocido como filtro laplaciano. Obsérvese que la suma de sus elementos de peso es cero. Esto quiere decir que el valor del brillo final de la imagen filtrada será muy reducido. La imagen habrá desaparecido y en su lugar solamente destacarán los bordes que en ella hubiera. Si se desea conservar la imagen, pero enriquecerla con los rasgos lineales realzados se puede sumar la imagen original al resultado del filtro laplaciano. Pero como el filtrado es una operación lineal, lo mismo da sumar las imágenes al final que sumar al filtro la función identidad, que sería:

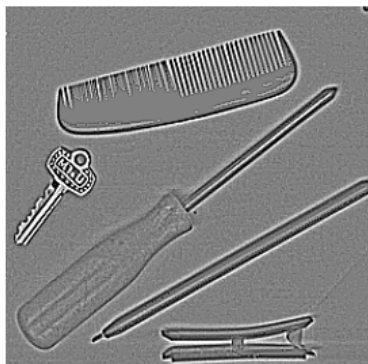


Imagen real y filtro laplaciano.

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (9)$$

Su suma con el laplaciano origina el segundo de los filtros (8).

Otros posibles filtros pueden ser:

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad \begin{bmatrix} 1 & -2 & 1 \\ -2 & 5 & -2 \\ 1 & -2 & 1 \end{bmatrix} \quad \begin{bmatrix} -1 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & -1 \end{bmatrix} \quad (10)$$

cuya selección dependerá del valor que pretenda dársele al ND original de cada posición en relación con los circundantes.

Filtros direccionales

Los filtros direccionales son operadores de gradiente anisótropo que tienden a realzar los rasgos lineales de la imagen primando determinadas direcciones. El efecto de refuerzo lineal puede ser unidimensional:

$$z'(m,n) = z(m,n) - z^{(+1)}(m,n) \quad (11)$$

cuya expresión significa sustraer de sí misma la imagen desplazada una posición en cualquier dirección elegida. Si no hay cambios en esa dirección, el resultado para la celda será cero, y será tanto más alto en valor absoluto cuando mayor diferencia de ND haya en dicha dirección.

Este reforzador lineal también puede operar sobre los dos sentidos de la dirección elegida:

$$z'(m,n) = [z(m,n) - z^{(+1)}(m,n)] + [z(m,n) - z^{(-1)}(m,n)] \tag{12}$$

Igual que antes, valores de z' próximos a cero indican zonas de escasa variación y viceversa, valores altos indican que ha habido un cambio local de luminosidad, es decir, que existe un borde.

Sin embargo, el filtrado direccional más utilizado está basado en la convolución mediante los filtros de Prewitt:

$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & -1 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & -1 & 1 \end{bmatrix}$
NO	N	NE
$\begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & -1 \end{bmatrix}$	$\begin{matrix} \leftarrow & \uparrow & \rightarrow \\ \leftarrow & & \rightarrow \\ \leftarrow & \downarrow & \rightarrow \end{matrix}$	$\begin{bmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{bmatrix}$
O		E
$\begin{bmatrix} 1 & -1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 & -1 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 & 1 \\ -1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$
SO	S	SE

(13)

Que realzan los bordes detectados en las direcciones norte, sur este, oeste e intermedias, dependiendo del gradiente de sus elementos. Obsérvese que siempre guardan una simetría respecto a un eje perpendicular a la dirección en la que detectan el borde.

Filtros no lineales

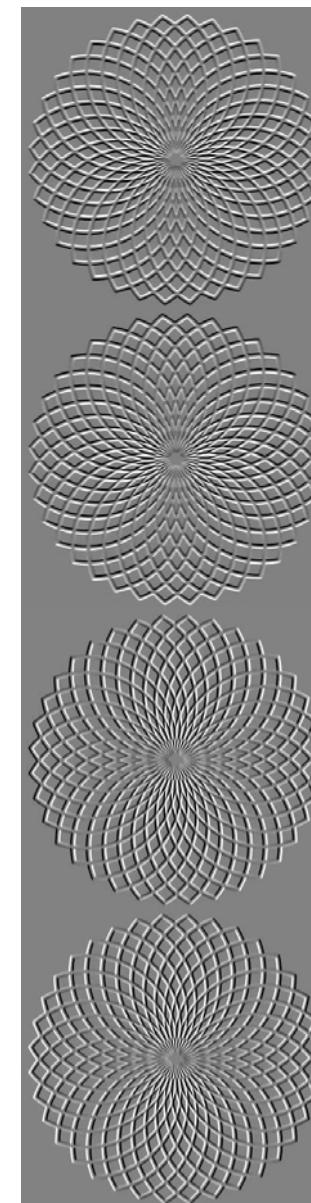
Existen filtros que no mantienen la propiedad de linealidad matemática, fundamentalmente porque sus elementos de peso son variables con la posición sobre la que se aplican. En estas circunstancias, aventurar el resultado del filtrado puede ser difícil sin probar antes.

Uno muy difundido es el *filtro de mediana*. Se conoce también como ventana de Tukey, y fue ideado para suprimir ruido de la escena sin difuminarla demasiado y respetando así los bordes. Consiste en asignar a la posición central de la ventana de filtrado, que suele ser de 3x3, un valor igual a la mediana estadística de los ND incluidos en ella. No existe, pues, una ventana de filtrado única para toda la imagen, sino que ésta cambia de coeficientes según se desplaza a través de la escena. Para una ventana de 3x3, el filtro sería:

$$z'(m,n) = \text{Med} \begin{bmatrix} z(m-1,n-1) & z(m-1,n) & z(m-1,n+1) \\ z(m-1,n) & z(m,n) & z(m+1,n) \\ z(m-1,n+1) & z(m,n+1) & z(m+1,n+1) \end{bmatrix} \tag{14}$$

Tiene un efecto de paso bajo y normalmente respeta mejor las formas que el filtro de media, pero al tratarse de un operador no lineal, sus resultados no son previsibles de forma general.

Otro muy utilizado es el *filtro de moda o de mayoría*. A la posición central de la ventana de filtrado se le asigna un valor igual a la moda estadística de los ND incluidos en ella. No existe tampoco en este caso una ventana de filtrado única para toda la imagen, sino que ésta cambia de coeficientes según se desplaza a través de la escena. Para una ventana de 3x3, este



Resultados de aplicar filtros de Prewitt N, S, E y O.

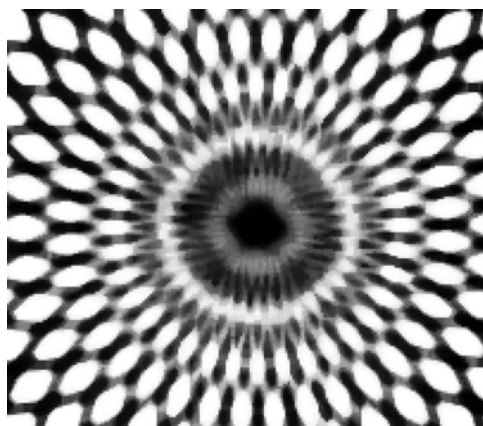
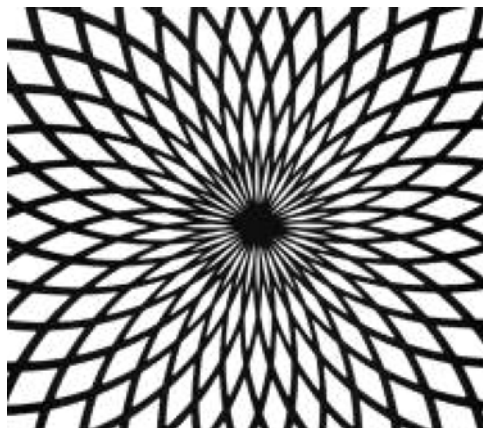
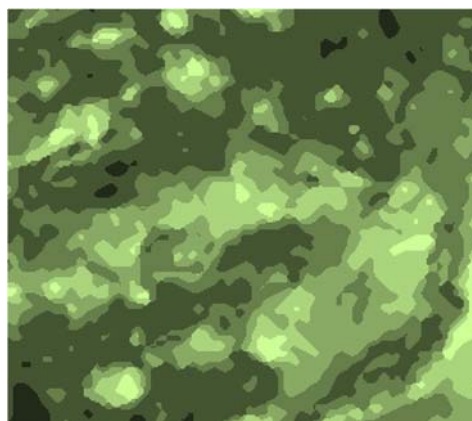
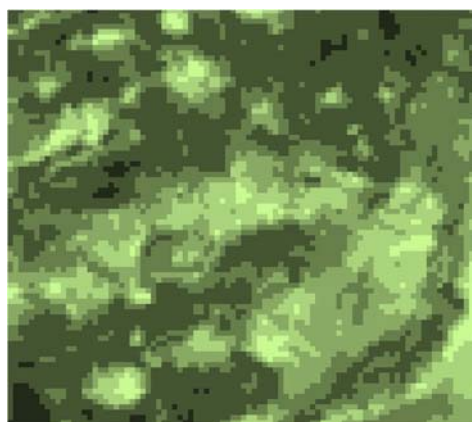


Imagen original y filtro de mediana.

filtro puede enunciarse del siguiente modo:

$$z'(m,n) = \text{Mod} \begin{bmatrix} z(m-1,n-1) & z(m-1,n) & z(m-1,n+1) \\ z(m-1,n) & z(m,n) & z(m+1,n) \\ z(m-1,n+1) & z(m,n+1) & z(m+1,n+1) \end{bmatrix} \quad (15)$$

Su efecto es el de eliminar píxeles sueltos de un valor



Resultados de aplicar un filtro de moda sobre la imagen superior

diferente al de su entorno. Se aplica fundamentalmente sobre imágenes cualitativas. Es frecuente utilizarlo sobre el resultado de una clasificación digital para obtener un aspecto más generalizado de la distribución de clases, ofreciendo grandes zonas homogéneas que el usuario puede interpretar más fácilmente, aun a riesgo de perder una cuota de veracidad en los resultados.

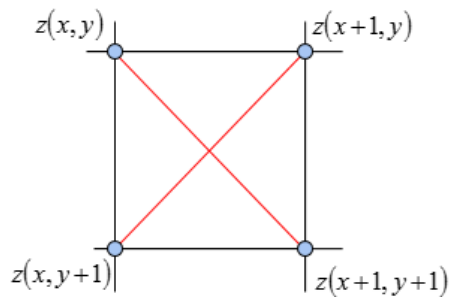
Existe un gran grupo de filtros no lineales destinados a detectar contornos. Para reconocer automáticamente los contornos, estos operadores calculan las diferencias entre los ND de unja celda y las contiguas en todas las direcciones. El trazado de los contornos se basa en representar los vectores direccionales cuyos módulos son las diferencias en todas las direcciones que superen un determinado umbral seleccionado. Precisamente en la elección de los umbrales depende en gran medida la eficacia de estos filtros. Pueden aparecer bordes no deseados o perderse otros realmente buscados.

El filtro de *Roberts* se trata de un filtro no lineal aplicado mediante una ventana de 2x2:

$$z'(i,j) = \sqrt{[z(i,j) - z(i+1,j+1)]^2 + [z(i,j+1) - z(i+1,j)]^2} \quad (16)$$

o bien:

$$z'(i,j) = |z(i,j) - z(i+1,j+1) + z(i,j+1) - z(i+1,j)| \quad (17)$$



Posiciones para el filtro de Roberts.

Para la extracción de bordes también se aplica el filtro de *Sobel*:

$$z'(i, j) = \sqrt{C^2 + F^2} \tag{18}$$

siendo la matriz representada una ventana de la imagen de 3x3 centrada sobre la celda (i, j) y donde los coeficientes C y F se calculan del siguiente modo:



Original y filtro de Roberts.

$$\begin{bmatrix} z_0 & z_1 & z_2 \\ z_7 & z[i, j] & z_3 \\ z_6 & z_5 & z_4 \end{bmatrix} \quad \begin{aligned} C &= (z_2 + 2z_3 + z_4) - (z_0 + 2z_7 + z_6) \\ F &= (z_0 + 2z_1 + z_2) - (z_6 + 2z_5 + z_4) \end{aligned} \tag{19}$$

Para usar el filtro de *Kirsh* se definen las variables S_n y T_n :

$$S_n = z_n + z_{n+1} + z_{n+2} \tag{20}$$

$$T_n = z_{n+3} + z_{n+4} + z_{n+5} + z_{n+6} + z_{n+7}$$

siendo n cualquiera de las ocho posiciones señaladas alrededor de la celda (i, j) :

$$\begin{matrix} z_1 & z_2 & z_3 \\ z_8 & z[i, j] & z_4 \\ z_7 & z_6 & z_5 \end{matrix} \tag{21}$$

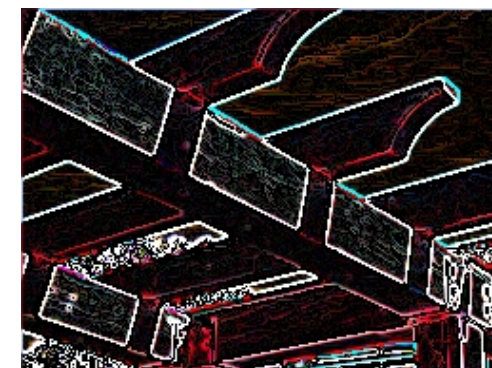
Se calculan los ocho valores de S_n y T_n para n variando entre 0 y 7 y se adopta el valor del filtro como:

$$K(i, j) = \text{MAX} [|5S_n - 3T_n|] \tag{22}$$

Finalmente, otro de los filtros de extracción de bordes utilizado con relativa frecuencia es el de *Wallis*. Este operador considera una celda de imagen como borde si el logaritmo de su ND excede de la media de los logaritmos de los cuatro vecinos. Asumiendo la nomenclatura de (21):



Resultado del filtro de Sobel.



Original y filtro de Kirsh.

$$z'(i, j) = \log[z(i, j)] - \frac{1}{4}[\log(z_2) + \log(z_4) + \log(z_6) + \log(z_8)] = \log[z(i, j)] - \frac{1}{4}\log(z_2 z_4 z_6 z_8)$$

$$z'(i, j) = \frac{1}{4} \log \left(\frac{[z(i, j)]^4}{z_2 z_4 z_6 z_8} \right) \quad (23)$$

o, lo que es lo mismo:

Filtrado en el dominio de la frecuencia

Cualquier señal periódica puede enunciarse como una sumatoria infinita de funciones periódicas de distintas frecuencias. Este nuevo espacio de referencia para la descripción de la función se conoce como dominio de la frecuencia. La serie de Fourier permite realizar el cambio de dominio temporal al de la frecuencia. Una función periódica en forma de serie de Fourier queda unívocamente definida mediante la especificación de los coeficientes de dicha serie. La representación gráfica de los coeficientes complejos de la serie en función de la frecuencia angular se denomina espectro de amplitud y la representación gráfica del ángulo de fase en función de la frecuencia es el espectro de fase:

La generalización del análisis de Fourier y su utilización en el campo de la teledetección requiere desarrollar un método aplicable a funciones no periódicas, a las cuales puedan ser asimiladas las imágenes digitales. La integral o transformada de Fourier permite describir funciones no periódicas en el dominio de la frecuencia. De forma análoga, la transformada inversa de Fourier permite pasar de nuevo al dominio temporal original.

En el análisis digital de imágenes todavía hay que

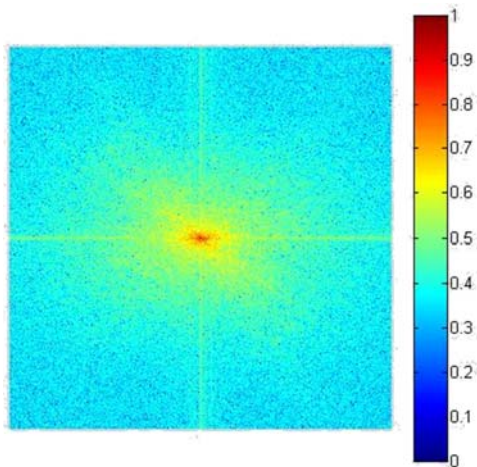
realizar algunas adaptaciones más para poder aplicar la transformada de Fourier. En primer lugar, las funciones deben ser finitas y discretas, como lo son las imágenes digitales. Además, la variable temporal debe ser reemplazada por la espacial, como corresponde a las imágenes de satélite, en contraste con las señales electrónicas o de audio. Por último, tampoco son útiles las funciones unidimensionales, puesto que la imagen se desarrolla en el plano bidimensional. Por lo tanto, debe reconsiderarse la teoría y adaptar el análisis de Fourier al caso de las imágenes digitales. Para ello se enuncia la transformada discreta y bidimensional de Fourier.

Sea $\phi(i, j)$; $i, j = 0, \dots, K-1$, el brillo de la celda (i, j) de una imagen cuadrada de K^2 filas por columnas.

Su transformada discreta de Fourier es:

$$\Phi(r, s) = \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} \phi(i, j) e^{-i2\pi(ir+js)/K} \quad (24)$$

En esta expresión $i = \sqrt{-1}$ y r y s son las frecuencias espaciales en las dos dimensiones de la imagen.



Transformada de Fourier de una imagen mostrando las frecuencias espaciales en una escala de 0 a 1.

La imagen puede ser reconstruida en el dominio espacial a partir de su transformada de Fourier aplicando la transformada inversa:

$$\phi(i,j) = \frac{1}{K^2} \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} \Phi(r,s) e^{i2\pi(ir+js)/K} \quad (25)$$

La expresión (22) también suele escribirse del siguiente modo:

$$\Phi(r,s) = \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} \phi(i,j) W^{ir+js} \quad (26)$$

donde $W = e^{-i2\pi/K}$. (27)

La aplicación de la transformada de Fourier permite pasar la imagen del dominio espacial al de la frecuencia. De forma similar a lo que sucede con las series de Fourier, también con la transformada cabe hablar de espectro de magnitud y espectro de fase de la función. El sistema de coordenadas en una imagen es el dominio espacial de la imagen e indica la posición de cada celda. En el dominio de la frecuencia, las coordenadas r o s representan la frecuencia con que un determinado ND aparece en la imagen.

El concepto de frecuencia espacial de una imagen estática es análogo al de frecuencia temporal en el ámbito del procesamiento de señales. Una señal de audio está constituida por una composición más o menos extensa de funciones periódicas. La frecuencia temporal indica la velocidad de cambio en los valores instantáneos de la señal. En la imagen, la función a estudiar es la sucesión de niveles digitales correspondientes a las celdas de una misma línea o de una misma columna. La frecuencia espacial indica la frecuencia de alteración o cambio de los niveles digitales. Imágenes muy abigarradas manifiestan una alta

frecuencia espacial. En cambio, imágenes muy uniformes y monótonas, están caracterizadas por su baja frecuencia espacial. La frecuencia espacial en cada dirección habitualmente presenta valores distintos.

La transformada de la imagen en dominio de la frecuencia produce una nueva imagen $\Phi(r,s)$ en la que las coordenadas de sus celdas representan la frecuencia espacial en ambas direcciones. La celda $\Phi(0,0)$ muestra un nivel digital igual al promedio de luminosidad de la imagen, único componente del espectro con frecuencia espacial cero en ambas direcciones. Las demás celdas representan los componentes de la imagen con frecuencias que van aumentando en un factor $1/K$.

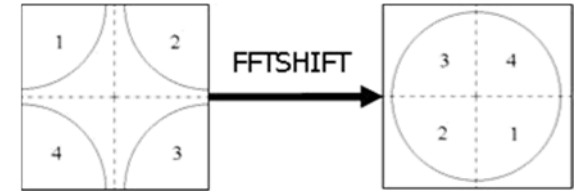
La transformada de Fourier tiene la propiedad de periodicidad por la cual el espectro de amplitud de la imagen es simétrico en el entorno de $K/2$ y el espectro de fase, antisimétrico. De ello se deduce que la imagen transformada no presentará más información por encima de $K/2$, ni horizontal ni verticalmente. Por esto, se acostumbra a colocar la celda $\Phi(0,0)$ en el centro de imagen, de tal modo que la imagen frecuencial resulta ser simétrica respecto de su centro

Teorema de la convolución

El teorema de la convolución en el espacio establece que la transformada de Fourier de la convolución entre dos funciones en el dominio del espacio es igual al producto de las transformadas de cada una de ellas. Enunciándolo para una función unidimensional en el dominio temporal:

$$\mathcal{F}[f_1(t) \otimes f_2(t)] = F_1(\omega) \cdot F_2(\omega) \quad (28)$$

donde $F_1(\omega) = \mathcal{F}[f_1(t)]$ y $F_2(\omega) = \mathcal{F}[f_2(t)]$ y \mathcal{F} es



Simetría central de la transformada de Fourier y aplicación a una imagen real. El origen de coordenadas frecuenciales se sitúa en el centro de la transformada.

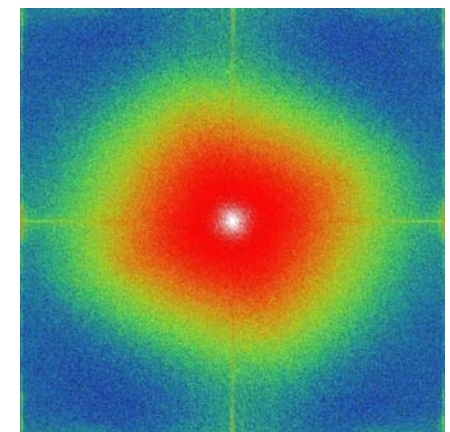
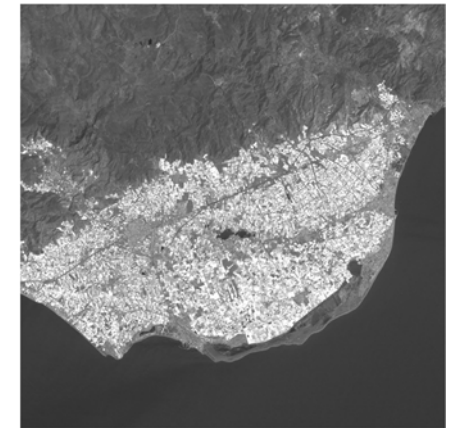
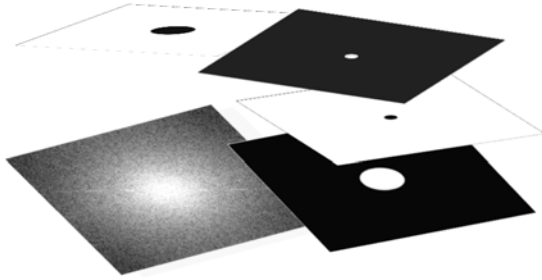
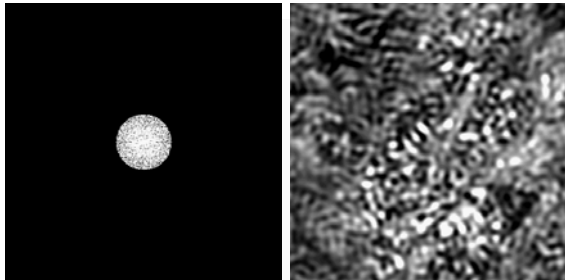


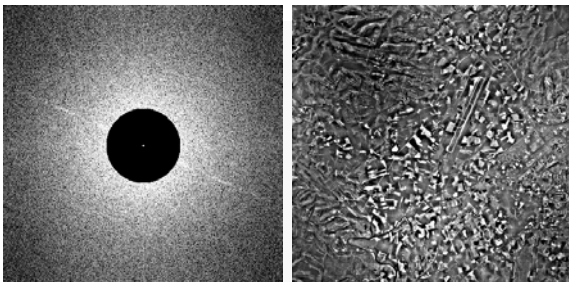
Imagen real y su transformada de Fourier, con el origen de coordenadas frecuenciales en el centro del espectro.



Filtros de paso bajo y de paso alto para aplicar a una imagen en el dominio de la frecuencia mediante su producto.



Plantilla de filtro de paso bajo sobre la imagen frecuencial y el resultado del filtrado devuelto al dominio espacial mediante la transformada inversa de Fourier.



Plantilla de filtro de corte de banda sobre la imagen frecuencial y el resultado del filtrado devuelto al dominio espacial mediante la transformada inversa de Fourier.

la transformada de Fourier.

De forma análoga se puede expresar el teorema de la convolución en el de la frecuencia:

$$\mathcal{F}^{-1} [F_1(\omega) \otimes F_2(\omega)] = 2\pi f_1(t) f_2(t) \quad (29)$$

siendo ahora:

$$\mathcal{F}^{-1} [F_1(\omega)] = f_1(t) \text{ y } \mathcal{F}^{-1} [F_2(\omega)] = f_2(t)$$

La transformada inversa de la convolución de dos funciones es igual al producto de sus transformadas inversas, afectadas de una constante.

De una forma sencilla y aproximada se puede decir que lo que en el dominio del tiempo es una convolución, en el de la frecuencia equivale a un producto, y que un producto en el dominio espacial equivale a una convolución en el de la frecuencia.

Del teorema de la convolución deriva una propiedad muy interesante de la transformada de Fourier aplicable al filtrado. El filtrado digital, tal como se ha explicado arriba, consiste en una convolución entre la imagen y la función filtro. Si se pasan al dominio de la frecuencia tanto la imagen como el filtro, pueden ambas transformadas multiplicarse simplemente para realizar el filtrado. Solamente quedará para finalizar el proceso aplicar la transformada inversa de Fourier al producto para volver al dominio espacial y obtener la imagen filtrada. Este proceso se denomina filtrado en el dominio de la frecuencia.

La convolución $z'(i, j) = w(m, n) \otimes z(i, j)$ propia del filtrado en el dominio del espacio puede sustituirse por el siguiente producto en el dominio de la frecuencia:

$$Z'(r, s) = W(\rho, \sigma) \cdot Z(r, s) \quad (30)$$

donde $Z' = \mathcal{F}(z')$; $W = \mathcal{F}(w)$; $Z = \mathcal{F}(z)$.

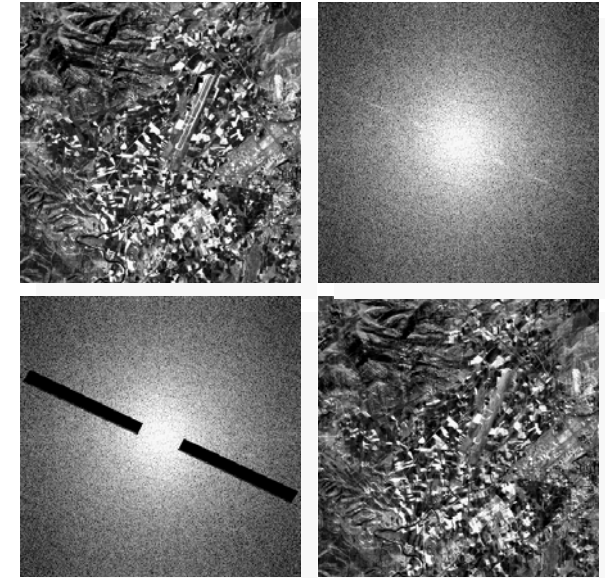
El filtro en el dominio de la frecuencia W tiene un aspecto peculiar y en la mayoría de los casos viene a ser un círculo central de $ND=1$ en un entorno de $ND=0$, si se trata de un filtro de paso bajo, o bien el caso contrario: un círculo de $ND=0$ rodeado de $ND=1$ hasta completar la imagen. Como quiera que su aplicación en el dominio de la frecuencia es mediante el producto, cuando se multiplique esta función por la imagen en el dominio de la frecuencia solamente se extraerán los valores centrales en el primer caso, o solamente los periféricos en el segundo. Dicho en otros términos, se mantienen las bajas frecuencias, que como se ha dicho se sitúan próximas al centro de la imagen frecuencial, o bien se mantienen solamente las altas frecuencias, en el segundo caso. De ahí la denominación de filtros de *paso bajo* o de *paso alto*.

El filtrado en el dominio de la frecuencia es más poderoso y versátil que el filtrado espacial, ya que permite, entre otras cosas, dejar ajenos al filtrado determinados intervalos de frecuencia espacial que convengan ser conservados como en la imagen original. Para evitar que en el proceso de filtrado de paso alto se pierdan las características geométricas y radiométricas básicas de la imagen en favor de los elementos de sólo baja o alta frecuencia, es posible idear filtros de *paso de banda* o de *corte de banda*, respectivamente. Estos últimos tienen la misión de extraer los componentes de alta frecuencia sin perder los valores próximos al centro de la imagen frecuencial. De este modo se mantiene la imagen y se refuerzan los rasgos lineales de un modo similar a como actúa el filtro laplaciano.

Se puede así mismo adoptar para la función de filtrado frecuencial valores intermedios entre 0 y 1, de modo que el efecto del filtro no sea tan drástico. Además, es posible diseñar filtros de muesca (*notch filters*) que permiten omitir determinados intervalos de frecuencia en ciertas direcciones espaciales y no en otras. Esto se consigue aplicando máscaras de filtrado frecuencial de más de una ventana en posición no centrada con respecto al espectro de amplitud a filtrar, pero colocadas simétricamente con respecto al centro del espectro. El resultado al deshacer la transformación al dominio espacial es un filtrado direccional pero sólo de los niveles digitales que muestran la frecuencia de cambio correspondiente a la situación de las ventanas muesca del filtro.

Una vez realizada la operación de filtrado en el dominio de la frecuencia, el espectro resultante debe ser transformado en imagen espacial mediante la aplicación de la transformada inversa. En total, el conjunto de las etapas de paso del dominio espacial al de la frecuencia, la propia operación de filtrado y la reversión final al dominio espacial consumen

$O_{frec} = 2K^2 \log_2 K + K^2$ multiplicaciones. En contraposición, si el filtrado es realizado en el dominio espacial utilizando una matriz deslizante de $C \times L$ (columnas \times líneas), el número de operaciones implicadas son $O_{esp} = CLK^2$. En consecuencia, la razón entre las operaciones necesarias para el filtrado en uno y otro dominio es: $R = O_{esp} / O_{frec} = CL / (2 \log_2 K + 1)$. Para una imagen de 512×512 celdas y una matriz deslizante de filtrado de 3×3 , el número de operaciones en el dominio espacial es de 2 359 296. La misma operación en el dominio de la frecuencia exige 4 980 736. La tendencia se afianza a medida que la imagen crece de tamaño. Sin embargo, se invierte cuando se utilizan plantillas de filtrado de mayor número de elementos. Así pues, para la misma escena anterior, si el filtrado espacial se realiza con una matriz de 4×4 elementos, el número de operaciones resulta ser de 4 194 304, mientras que en el dominio de la frecuencia el número de operaciones sigue siendo el mismo. Si el filtro se efectúa con una plantilla de 5×5 , el filtrado espacial consume ya 6 553 600 operaciones.



De izquierda a derecha y de arriba hacia abajo: imagen original, su espectro en el dominio de la frecuencia, filtro de muesca y resultado del filtrado en el dominio espacial. Obsérvese cómo se han borrado las pistas de rodadura y aterrizaje del aeropuerto, que son los únicos elementos geométricos perpendiculares a la orientación de las muescas del filtro.

Restauración de imágenes

La adquisición de la imagen comporta una serie de procesos que contribuyen a la degradación de la imagen real. El factor más importante suele ser la propia atmósfera, cuyos aerosoles en suspensión producen un efecto dispersivo que emborrona la nitidez de la imagen. A ello se suma la propia óptica del sensor con sus sistemas de lentes y espejos que también distorsionan la señal. Considerando el conjunto de elementos que degradan la señal como un sistema, la transferencia desde el espacio objeto al espacio imagen vendrá determinada por la *función del sistema*.

En el capítulo 6 se introdujeron los conceptos de función de dispersión puntual (*PSF*) y del impulso unitario. Pues bien, el patrón de degradación de la señal, como ya se vio, viene establecido por la *PSF*, función que describe la forma en la que un impulso unitario es transformado en su imagen final.

Una propiedad del impulso unitario no vista hasta ahora es que su transformada de Fourier es igual a una constante de valor unidad:

$$\mathcal{F}[\delta(t)] = \int_{-\infty}^{\infty} \delta(t) e^{-i\omega t} dt = 1 \quad (31)$$

y de un modo análogo, la transformada de Fourier de una constante es un impulso:

$$\mathcal{F}(c) = \int_{-\infty}^{\infty} c e^{-i\omega t} dt = 2\pi c \delta(\omega) \quad (32)$$

Ambas serán de utilidad en la restauración de imágenes.

De no existir atmósfera dispersiva ni sistemas ópticos que introdujeran distorsión, la respuesta del sistema ante el impulso unitario debería ser otro impulso unitario. Desafortunadamente, el sistema degrada la señal y convierte el impulso unitario en una PSF, que es la función dispersiva $h(x,y)$. Si se conoce con precisión la PSF de una imagen, su convolución por la imagen real debe proporcionar exactamente la imagen finalmente adquirida. Siendo esto así, cabe plantearse si fuera posible aplicar una suerte de función inversa que, aplicada a la imagen adquirida pudiese proporcionar la imagen real. Este es el campo de la restauración de imágenes.

Existen diversos métodos de restauración de imágenes, de los cuales, unos operan en el dominio del espacio, otros netamente en el de la frecuencia y, finalmente, algunos que aprovechan el dominio de la frecuencia para deducir funciones restauradoras que se aplican sobre la imagen en el dominio del espacio, de ahí que también se les llame funciones de deconvolución. Un procedimiento relativamente sencillo es el de deconvolución ciega (*blind deconvolution*) que se explica a continuación.

Se parte de la siguiente premisa: si se conociera la

función de deconvolución $g(x,y)$ y la PSF del sistema formador $h(x,y)$ debería verificarse:

$$g(x,y) \otimes h(x,y) = \delta(x,y) \quad (33)$$

es decir, la convolución de la función restauradora, por la PSF del sistema, por su propia naturaleza, debe convertir esa PSF en un impulso unitario, puesto que su acción debe ser contrarrestar el efecto dispersivo del sistema. Esa misma expresión, en el dominio de la frecuencia será:

$$\text{FFT}[g(x,y)] \cdot \text{FFT}[h(x,y)] = \text{FFT}[\delta(x,y)] \quad (34)$$

siendo FFT la transformada rápida de Fourier, un algoritmo para realizar la transformada discreta de Fourier sobre la imagen digital. Escrito de otra manera:

$$G(r,s) \cdot H(r,s) = [1] \quad (35)$$

en donde:

$$\text{FFT}[\delta(x,y)] = [1] = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix} \quad (36)$$

Aplicando la propiedad (29) a un impulso bidimensional. La constante unidad es ahora una imagen continua con todos sus ND=1.

La función de deconvolución puede obtenerse mediante la reversión al dominio del espacio de la expresión (43):

$$g(x, y) = \text{FFT}^{-1} \left\{ \frac{[1]}{H(r, s)} \right\} \quad (37)$$

Para comprobar que la aplicación de $g(x, y)$ da los resultados perseguidos, considérese que en el proceso de adquisición de la imagen se verifica:

$$h(x, y) \otimes I(x, y) = f(x, y) \quad (38)$$

en donde $I(x, y)$ es la imagen real, $f(x, y)$ es la imagen degradada y $h(x, y)$ es la PSF del sistema.

Operando con la función de deconvolución en los dos miembros:

$$g(x, y) \otimes h(x, y) \otimes I(x, y) = g(x, y) \otimes f(x, y) \quad (39)$$

o sea:

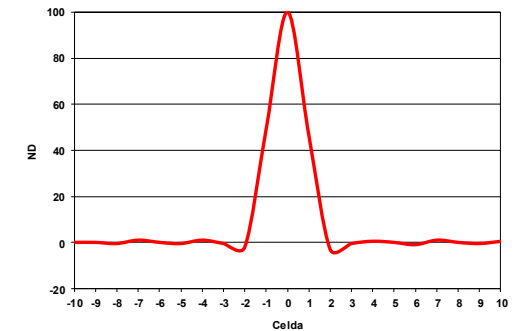
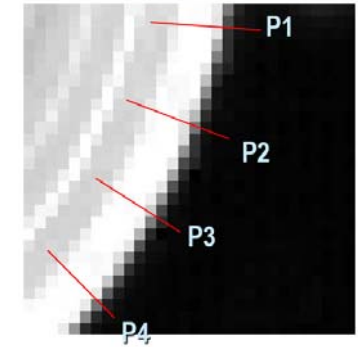
$$\delta(x, y) \otimes I(x, y) = g(x, y) \otimes f(x, y) \quad (40)$$

Y como la convolución de un impulso con una imagen da como resultado la misma imagen:

$$I(x, y) = g(x, y) \otimes f(x, y) \quad (41)$$

Expresión que viene a demostrar que la imagen puede ser restaurada mediante la convolución de la función de deconvolución con la propia imagen.

En cuanto a la forma de obtener la PSF de la imagen, también existen diversos procedimientos. Uno relativamente sencillo consiste en buscar dentro de la escena áreas de saltos bruscos de reflectancia, esto es, zonas contiguas de alto contraste, como pueden ser las fronteras agua-tierra en el visible. Un perfil radiométrico o *transecto* perpendicular a la frontera de ambas zonas en ausencia de efecto dispersivo debería parecerse a una función escalonada. Dado que la PSF degrada la señal por su efecto dispersivo, el escalón aparecerá suavizado. Las últimas celdas de la zona oscura empezarán a clarearse, en tanto que las primeras de la zona clara no serán tan brillantes como debieran. La derivada de esa curva de transición es precisamente la PSF buscada.



Arriba: transectos sobre una presa, zona de alto contraste en su cara aguas arriba.
Abajo, la PSF obtenida al derivar el promedio de los perfiles radiométricos obtenidos.

11

Georreferenciación

Introducción

La imagen adquirida por un sensor a bordo de un satélite presenta diferencias con respecto a la imagen real tanto en la situación de sus puntos como en el nivel de brillo de cada uno. En consecuencia, la geometría de la imagen no refleja con fidelidad absoluta la de la superficie, ni tampoco pueden deducirse sus propiedades físicas a partir de la lectura directa de los niveles digitales. Es necesario aplicar unas determinadas operaciones de corrección, que están encaminadas a minimizar estas alteraciones en la medida de lo posible. Algunas de estas tareas son aplicadas por las mismas estaciones de recepción, las cuales ofrecen distintos niveles de corrección. El estudio detallado de la imagen, los estudios multitemporales o la fusión con otras fuentes cartográficas requieren

correcciones minuciosas, que normalmente son realizadas por el analista.

La georreferenciación es el proceso mediante el cual se dota a la imagen de coordenadas cartográficas, en el sistema de proyección que se elija, tarea que suele abordarse simultáneamente con la corrección geométrica.

Este capítulo aborda las operaciones más habituales de corrección geométrica en el pretratamiento de las imágenes de satélite, operación necesaria antes de proceder al estudio en sí de la información que proporcionan.

Métodos polinómicos de corrección

La corrección geométrica de las imágenes requiere deducir una función de transformación para cada variable planimétrica que convierta la posición inicial (x,y) de cada celda en coordenadas corregidas (u,v) :

$$\left. \begin{aligned} u &= f_1(x, y) \\ v &= g_1(x, y) \end{aligned} \right\} \quad (1)$$

Que también se pueden poner en función de las coordenadas imagen en columna y línea de cada celda:

$$\left. \begin{aligned} u &= f_2(c, l) \\ v &= g_2(c, l) \end{aligned} \right\} \quad (2)$$

Dado que las distorsiones geométricas sufridas por la imagen obedecen a numerosas causas, algunas derivadas del efecto proyectivo, de la esfericidad terrestre, de la rotación del planeta, etc. cuyo análisis puede abordarse analíticamente, otras tantas son aleatorias y vienen a engrosar y complicar el efecto resultante. En definitiva, la distorsión geométrica estaría definida por una función complicada y desconocida. Cuando no se conoce explícitamente una función di-

ficultosa de estas características, se le suele caracterizar matemáticamente mediante su desarrollo en serie, de cuyo polinomio se toman sólo los términos necesarios, sabiendo que un cierto número de ellos proporcionan una explicación razonablemente suficiente del comportamiento de la función. En consecuencia, a efectos prácticos, se puede sustituir la función original por dicho polinomio.

Los métodos más sencillos de corrección geométrica están basados en la adopción de funciones polinómicas que transforman las coordenadas de un conjunto de puntos de control en coordenadas corregidas. Una vez encontrados los polinomios, se aplican a todas las celdas de la imagen para forzar su *deformación* hasta conseguir una imagen corregida geoméricamente.

El procedimiento sigue las etapas siguientes:

- Selección de los puntos de control.
- Cálculo de los coeficientes de los polinomios de ajuste.
- Transferencia de las celdas a sus nuevas posiciones.
- Remuestreo.

Selección de los puntos de control

Los puntos de control son puntos de los que se conocen sus coordenadas reales (coordenadas-terreno) y las coordenadas matriciales de sus homólogos en la imagen (coordenadas-imagen).

La correcta selección de los puntos de control y la determinación de sus coordenadas reales e imagen es determinante en el resultado final del proceso. Los puntos de control deben ser elegidos en accidentes del terreno permanentes, como cruces de caminos, por ejemplo, u otros hitos visibles. No son válidos las

formas definidas por sombras en la imagen, como tampoco lo son las derivadas de accidentes hidrológicos, puesto que su posición no es seguro que sea estable.

La calidad del ajuste depende de tres factores:

1. Número de puntos. El mínimo número necesario para resolver las ecuaciones es de 3 puntos, si los polinomios son de primer grado, 6, si son de segundo grado y 10 si son de tercer grado. Aunque se aconseja tomar no menos del doble de número matemáticamente necesario para el ajuste, siempre es preferible tomar un número de puntos considerablemente mayor.
2. Localización de los puntos. Es imprescindible un gran rigor en la localización geométrica exacta, tanto en la imagen como en el terreno o en la fuente cartográfica que lo sustituya.
3. Distribución de los puntos en la imagen. Es deseable una distribución uniforme. Por otra parte, siempre es preferible la interpolación a la extrapolación. Con ello se quiere decir que si se toman puntos periféricos en la imagen puede asegurarse con algún margen de seguridad que el ajuste interior pueda ser tolerable. Sin embargo, la inversa no es cierta: por muchos puntos que se tomen en una zona central de la imagen, nunca se tiene la seguridad de que las zonas periféricas de la imagen ajusten correctamente, especialmente si se utilizan polinomios de segundo o tercer grado.

Las funciones que transforman las coordenadas actuales (x,y) de los puntos de control en coordenadas

corregidas (u,v) del mapa de referencia son polinomios, cuya forma genérica es:

$$\left. \begin{aligned} u &= \sum_{p=0}^n \sum_{q=0}^{n-p} a_{pq} x^p y^q \\ v &= \sum_{p=0}^n \sum_{q=0}^{n-p} b_{pq} x^p y^q \end{aligned} \right\} \quad (2)$$

donde n representa el grado del polinomio. Los polinomios de primer grado permiten rotar, trasladar la imagen y cambiar su escala. Los de segundo incluyen ciertas deformaciones residuales aleatorias. Los de tercer grado comportan una corrección más depurada. No es habitual utilizar polinomios de grado superior, aunque en determinadas aplicaciones de corrección asistida por ordenador se utilizan polinomios de cuasi-cuarto grado y cuasi-sexto grado:

$$\left. \begin{aligned} u &= \sum_{p=0}^n \sum_{q=0}^n a_{pq} x^p y^q \\ v &= \sum_{p=0}^n \sum_{q=0}^n b_{pq} x^p y^q \end{aligned} \right\} \quad (3)$$

En cualquiera de los casos, Los coeficientes a_{pq} y b_{pq} se determinan por ajuste mínimo-cuadrático a partir de las coordenadas primitivas y corregidas de los puntos de control. El número mínimo de puntos de control que deben seleccionarse para encontrar una solución depende del grado del polinomio y sigue la regla:

$$m = \frac{(n+1)(n+2)}{2} \quad (4)$$

Donde m es el mínimo número de puntos de control

y n el grado del polinomio adoptado para el ajuste.

Siguiendo el formato (2), El caso más sencillo es el de ajuste lineal, en el que las coordenadas corregidas (u,v) se encuentran en función de las coordenadas (x,y) de los puntos de control de la imagen sin corregir:

$$\left. \begin{aligned} u &= a_0 + a_1 x + a_2 y \\ v &= b_0 + b_1 x + b_2 y \end{aligned} \right\} \quad (5)$$

Al ser 3 las incógnitas para cada una de las dos variables a_0, a_1, a_2 y b_0, b_1, b_2 , su resolución requiere al menos de un sistema de 3 ecuaciones. Se necesitarían por lo tanto un mínimo de 3 puntos de control para este ajuste y ello podría proporcionar una solución matemática, aunque no necesariamente correcta desde el punto de vista cartográfico. Siempre que el sistema de ecuaciones sea compatible se puede dar con una solución, incluso habiendo tomado alguna coordenada erróneamente. Sin embargo, el ajuste mínimo cuadrático se basa en la sobreabundancia de datos para encontrar una solución que satisfaga razonablemente al conjunto de todos ellos. Por esta razón se procederá siempre a seleccionar un número no menor del doble del mínimo matemáticamente necesario, aunque es deseable un número considerablemente superior.

Para el ajuste mediante polinomios de segundo grado, las ecuaciones a plantear serán:

$$\left. \begin{aligned} u &= a_0 + a_1 x + a_2 y + a_3 xy + a_4 x^2 + a_5 y^2 \\ v &= b_0 + b_1 x + b_2 y + b_3 xy + b_4 x^2 + b_5 y^2 \end{aligned} \right\} \quad (6)$$

Para la resolución de los coeficientes se necesitará un sistema de 6 ecuaciones con 6 incógnitas para

cada una de las variables. Por ello, el mínimo número de puntos de control a utilizar en un ajuste de segundo grado deberá ser de 6, pero en la práctica no se tomarán menos de 12 puntos.

Un ajuste de tercer grado exigiría plantear el siguiente sistema:

$$\left. \begin{aligned} u &= a_0 + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2 + a_6xy^2 + \\ &\quad + a_7x^2y + a_8x^3 + a_9y^3 \\ v &= b_0 + b_1x + b_2y + b_3xy + b_4x^2 + b_5y^2 + b_6xy^2 + \\ &\quad + b_7x^2y + b_8x^3 + b_9y^3 \end{aligned} \right\} \quad (7)$$

para cuya resolución se requiere un mínimo de 10 puntos, aunque en la práctica se tomarán al menos 20 puntos. En correcciones de imágenes completas se utilizarán por encima de 50 o 60 puntos, bien distribuidos por la escena.

En el caso de utilizar un polinomio de cuasi-cuarto grado, el sistema de ecuaciones sería:

$$\left. \begin{aligned} u &= a_0 + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2 + a_6xy^2 + \\ &\quad + a_7x^2y + a_8x^2y^2 \\ v &= b_0 + b_1x + b_2y + b_3xy + b_4x^2 + b_5y^2 + b_6xy^2 + \\ &\quad + b_7x^2y + b_8x^2y^2 \end{aligned} \right\} \quad (8)$$

Como se ve, con este formato se requieren al menos 9 puntos de control, aunque se seguirán tomando como poco el doble.

Ajuste mínimo-cuadrático

Como ha quedado dicho, para resolver los coeficientes de un polinomio de grado n no sirve plantear un

sistema de tantas ecuaciones como coeficientes. Si el sistema es compatible, siempre arrojaría una solución, con independencia de la correcta o incorrecta selección de los puntos o de la introducción de sus coordenadas reales. Para asegurar el cálculo razonablemente correcto de los coeficientes, se recurre a ajuste mínimo-cuadrático.

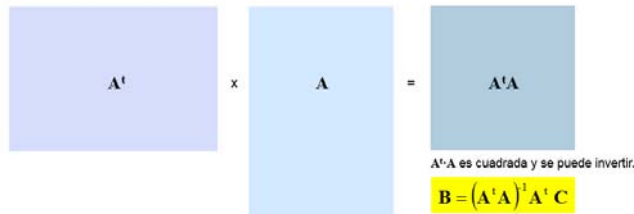
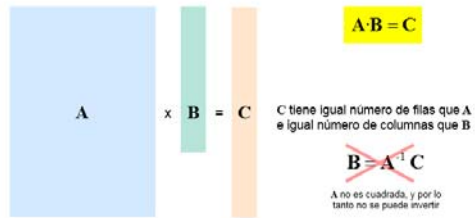
En el ajuste por mínimos cuadrados se seleccionan un número m de puntos considerablemente mayor del necesario y se plantean con ellos un sistema de m ecuaciones con n incógnitas.

Para m puntos, el sistema de ecuaciones mediante polinomios de segundo grado, sería, siguiendo (6):

$$\left. \begin{aligned} \begin{pmatrix} 1 & x_1 & y_1 & x_1y_1 & x_1^2 & y_1^2 \\ 1 & x_2 & y_2 & x_2y_2 & x_2^2 & y_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & y_m & x_my_m & x_m^2 & y_m^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} &= \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_m \end{pmatrix} \\ \begin{pmatrix} 1 & x_1 & y_1 & x_1y_1 & x_1^2 & y_1^2 \\ 1 & x_2 & y_2 & x_2y_2 & x_2^2 & y_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & y_m & x_my_m & x_m^2 & y_m^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix} &= \begin{pmatrix} v_0 \\ v_1 \\ \vdots \\ v_m \end{pmatrix} \end{aligned} \right\} \quad (9)$$

o sea:

$$\left. \begin{aligned} \mathbf{A} \cdot \mathbf{B}_u &= \mathbf{C}_u \\ \mathbf{A} \cdot \mathbf{B}_v &= \mathbf{C}_v \end{aligned} \right\} \quad (10)$$



Procedimiento para obtener el vector B.

donde **A** es la matriz de coordenadas imagen de los puntos, **B** el vector de coeficientes y **C** el vector de coordenadas terreno.

Como **A** no es cuadrada, no es invertible, por lo que hay que recurrir a un pequeño artificio matemático para poder despejar **B**, que es lo que interesa. Multiplicando ambos miembros por la matriz traspuesta de **A** se consigue en el primer miembro de cada una de las expresiones (10) una matriz que sí puede ser invertida y pasada al segundo miembro, quedando:

$$\left. \begin{aligned} \mathbf{B}_u &= (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \mathbf{C}_u \\ \mathbf{B}_v &= (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \mathbf{C}_v \end{aligned} \right\} \quad (11)$$

que es la solución buscada.

El enfoque del problema desde la perspectiva de los mínimos cuadrados conduce al mismo resultado. Se conocen las variables x e y , que son las coordenadas imagen de los puntos de control. Las incógnitas, por el contrario, son los coeficientes de los polinomios a_i y b_i . Tampoco se conocen las coordenadas u y v de cada punto corregido, puesto que todavía se desconocen los coeficientes que conforman cada polinomio.

Para formalizar el método, de cada punto se determinan sus coordenadas reales X e Y , que son las coordenadas terreno. La diferencia entre la posición (u,v) y (X,Y) es el error del ajuste. Se pretende que la suma de los cuadrados de todos ellos sea mínima.

Supóngase un ajuste de primer grado, que sólo exigiría 3 puntos y 4 puntos de control. Asumiendo que el ajuste fuera perfecto:

$$\left. \begin{aligned} X_1 &= a_0 + a_1 x_1 + a_2 y_1 \\ X_2 &= a_0 + a_1 x_2 + a_2 y_2 \\ X_3 &= a_0 + a_1 x_3 + a_2 y_3 \\ X_4 &= a_0 + a_1 x_4 + a_2 y_4 \end{aligned} \right\} \quad (12)$$

que en forma matricial sería:

$$\underbrace{\begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \\ 1 & x_4 & y_4 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}}_{\mathbf{B}} = \underbrace{\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix}}_{\mathbf{C}} \quad (13)$$

Pero en realidad sucede que:

$$\left. \begin{aligned} a_0 + a_1 x_1 + a_2 y_1 &= u_1 \\ a_0 + a_1 x_2 + a_2 y_2 &= u_2 \\ a_0 + a_1 x_3 + a_2 y_3 &= u_3 \\ a_0 + a_1 x_4 + a_2 y_4 &= u_4 \end{aligned} \right\} \quad (14)$$

que difiere de la posición teórica X_i .

Calculando el error cometido para cada punto:

$$\left. \begin{aligned} \varepsilon_1 &= u_1 - X_1 = a_0 + a_1 x_1 + a_2 y_1 - X_1 \\ \varepsilon_2 &= u_2 - X_2 = a_0 + a_1 x_2 + a_2 y_2 - X_2 \\ \varepsilon_3 &= u_3 - X_3 = a_0 + a_1 x_3 + a_2 y_3 - X_3 \\ \varepsilon_4 &= u_4 - X_4 = a_0 + a_1 x_4 + a_2 y_4 - X_4 \end{aligned} \right\} \quad (15)$$

y de igual forma para la Y .

Se suman los cuadrados de todos los errores:

$$\Phi = \sum \varepsilon_i^2 = \sum_{i=1}^4 (a_0 + a_1 x_i + a_2 y_i - X_i)^2 \quad (16)$$

y se impone la condición de mínimo igualando a cero la primera derivada:

$$\left. \begin{aligned} \frac{\partial \Phi}{\partial a_0} &= 2 \sum_{i=1}^4 (a_0 + a_1 x_i + a_2 y_i - X_i) = 0 \\ \frac{\partial \Phi}{\partial a_1} &= 2 \sum_{i=1}^4 (a_0 + a_1 x_i + a_2 y_i - X_i) x_i = 0 \\ \frac{\partial \Phi}{\partial a_2} &= 2 \sum_{i=1}^4 (a_0 + a_1 x_i + a_2 y_i - X_i) y_i = 0 \end{aligned} \right\} \quad (17)$$

Operando:

$$\left. \begin{aligned} 4 a_0 + a_1 \sum x_i + a_2 \sum y_i &= \sum X_i \\ a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i y_i &= \sum X_i x_i \\ a_0 \sum y_i + a_1 \sum x_i y_i + a_2 \sum y_i^2 &= \sum X_i y_i \end{aligned} \right\} \quad (18)$$

Y poniendo la expresión en forma matricial:

$$\underbrace{\begin{pmatrix} 4 & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i \\ \sum y_i & \sum x_i y_i & \sum y_i^2 \end{pmatrix}}_{\mathbf{A}^t \mathbf{A}} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \sum X_i \\ \sum X_i x_i \\ \sum X_i y_i \end{pmatrix} \quad (19)$$

que en realidad es:

$$\mathbf{A}^t \mathbf{A} \cdot \mathbf{B} = \mathbf{A}^t \mathbf{C} \quad (20)$$

es decir:

$$\mathbf{B} = (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \mathbf{C} \quad (21)$$

expresión análoga a (11).

Otros métodos de corrección

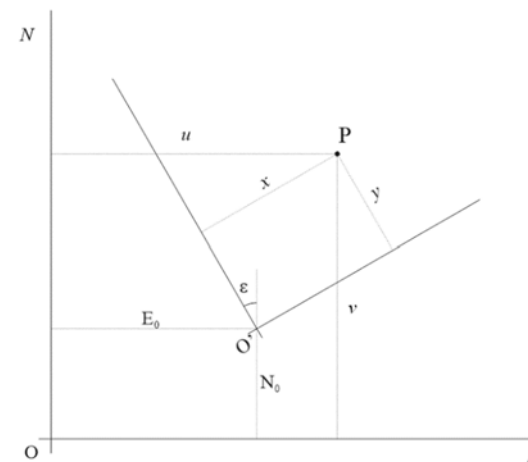
Aunque los más utilizados son los métodos de corrección polinómica, existen otras formas de asemejar la imagen a un mapa, y el analista deberá valorar si convienen emplearlos. A continuación se exponen brevemente algunos de ellos.

Transformación afín

Consiste en la combinación de una rotación, una traslación y un cambio de escala. Requiere conocer las

coordenadas de dos puntos en el sistema de referencia de la imagen y en el del mapa, es decir, cuatro coordenadas en total. Se conoce también como transformación de «4 parámetros», que son: E_0 , N_0 , f , ε , es decir, las dos traslaciones este y norte del origen de coordenadas, el factor de escala y el ángulo de rotación, respectivamente.

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} E_0 \\ N_0 \end{pmatrix} + f \begin{pmatrix} \cos \varepsilon & -\text{sen } \varepsilon \\ \text{sen } \varepsilon & \cos \varepsilon \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (22)$$



Transformación afín.

donde u y v son las coordenadas corregidas de un determinado punto y x e y son sus coordenadas originales.

Los cuatro parámetros se calculan mediante un ajuste por mínimos cuadrados de un conjunto de ecuaciones bastante superior a cuatro.

Esta transformación es demasiado simple para corregir las deformaciones habituales de una imagen de satélite. La razón se debe al hecho de que el algoritmo simplemente escala el mapa por un factor, pero no modifica la matriz de celdas. Además, el factor de escala es único para toda la escena, lo que exigiría que la imagen hubiera sido deformada por igual en toda su extensión y en todas las direcciones.

Control de calidad del ajuste

Las funciones polinómicas de corrección definen lo mejor posible los cambios de posición que han de sufrir los puntos de control. Estos cambios no obedecen en términos generales a transformaciones matemáticas simples, debido a la aleatoriedad de las distorsiones sufridas por la imagen.

Las funciones encontradas son aquéllas que mejor satisfacen las posiciones de todos los puntos de control, estableciendo un compromiso entre ellos, de tal modo que la suma de los cuadrados de las desviaciones sea mínima. El indicador más utilizado para medir la calidad del ajuste es el error medio cuadrático *RMS* (*Root Mean Square*):

Rubber-sheeting

La transformación denominada *rubber-sheeting* toma su nombre de su similitud con la adaptación de una lámina elástica a un mapa base indeformable. La imagen de satélite estaría impresa sobre la superficie de un globo hinchable de caucho, con multitud de deformaciones, y se trataría de hacer coincidir un conjunto de puntos esenciales de la imagen —de control—, sujetándolos con alfileres o chinchetas al propio mapa. Realmente se trata de un modelo polinómico de corrección basado en, al menos, cuatro puntos:

$$\left. \begin{aligned} u &= a_0 + a_1x + a_2y + a_3xy \\ v &= b_0 + b_1x + b_2y + b_3xy \end{aligned} \right\} \quad (23)$$

$$RMS = \sqrt{\frac{\sum_{i=1}^n [(u_i - X_i)^2 + (v_i - Y_i)^2]}{n}} \quad (24)$$

en cuya expresión, (X_i, Y_i) son las coordenadas reales de los puntos de control, (u_i, v_i) son las coordenadas estimadas para los puntos de control a partir de la aplicación de las funciones de ajuste y n es el número de puntos de control utilizados.

En la práctica se admite un RMS máximo de 2/3 del tamaño de la celda para cada coordenada X e Y , pero se pueden admitir desviaciones de cada punto equivalentes al tamaño de la celda tanto en X como en Y , ya que es el límite de la resolución espacial de la imagen.

Ortorectificación

El relieve del terreno tiene un carácter caprichoso y su mayor o menor aleatoriedad no responde a ninguna función sistemática. Como consecuencia, las distorsiones geométricas que introduce en la imagen persisten tras la corrección geométrica polinómica. Para eliminar el desplazamiento debido al relieve es necesario someter la imagen a un proceso de rectificación diferencial punto a punto, si bien, cuando el desplazamiento de cualquier celda sea inferior a la resolución espacial de la imagen, podría prescindirse de esta corrección.

En una proyección cónica el desplazamiento debido al relieve que sufre un elemento de una altura Δh sobre el plano de comparación situado a una cierta distancia del nadir de la imagen sigue la relación:

$$\frac{\Delta r}{r} = \frac{\Delta h}{h} \quad (25)$$

donde h es la altura del instrumento de adquisición, Δh , como se ha dicho, es la altura del objeto, r es la distancia del extremo proyectado medida desde el punto principal de la imagen y Δr es el desplazamiento debido al relieve.

La corrección implica conocer el desplazamiento debido al relieve despejando de (25):

$$\Delta r = r \frac{\Delta h}{h} \quad (26)$$

La ortorectificación es el proceso mediante el cual

se corrige el desplazamiento de cada una de las celdas de la imagen, en función de su altura relativa y su distancia a la traza del satélite.

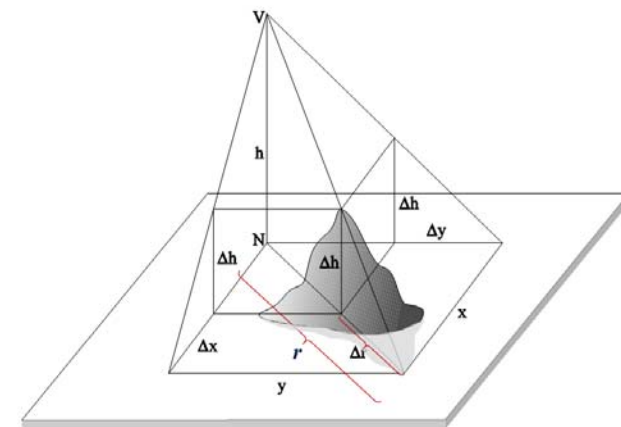
La expresión (26) puede particularizarse para cada una de las coordenadas de la siguiente forma:

$$\left. \begin{aligned} X &= x - \Delta x = x \left(1 - \frac{\Delta h}{h} \right) \\ Y &= y - \Delta y = y \left(1 - \frac{\Delta h}{h} \right) \end{aligned} \right\} \quad (27)$$

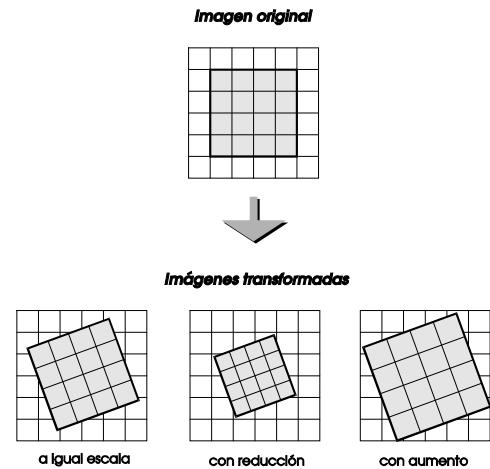
Pero las imágenes de satélite, que no son adquiridas por fotogramas completos sino por líneas o, a lo sumo, por pequeños grupos de ellas, no se adaptan a una proyección cónica sino más bien a una de tipo cilíndrico donde las distorsiones causadas por la topografía adquieren direcciones perpendiculares a la traza del satélite y son proporcionales a la distancia del punto hasta dicha traza. En consecuencia, puede prescindirse de una de las dos componentes. Si se considera que la traza mantiene la dirección Y , el desplazamiento a corregir solamente será:

$$X = x - \Delta x = x \left(1 - \frac{\Delta h}{h} \right) \quad (28)$$

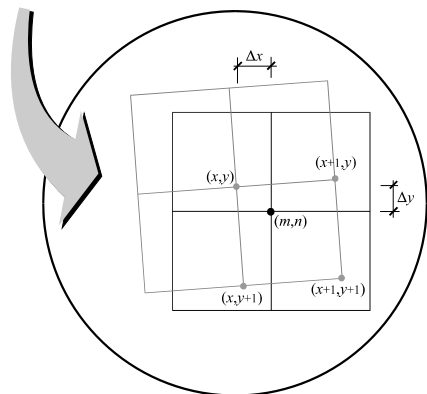
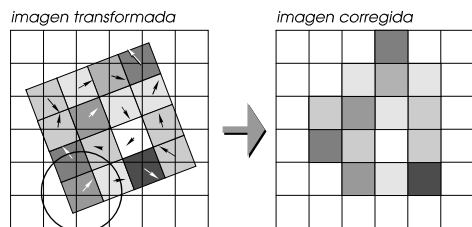
En la práctica, el valor Δh se toma de un modelo digital del terreno (MDT).



Geometría de la proyección cónica.



Necesidad del remuestreo.



Remuestreo por el vecino más próximo.

Remuestreo

Tras ser sometidas al cambio de coordenadas mediante la transformación polinómica elegida y, en su caso, corregido también el efecto del relieve, las posiciones de las celdas ocuparán valores reales, que en general no serán números enteros. Pero para que la imagen pueda ser procesada, se requiere que cada celda corregida ocupe una posición entera de columna y línea.

En el tratamiento digital de imágenes se denomina remuestreo (*resampling*) al conjunto de operaciones necesarias para asignar nuevos valores a las celdas cuando se produzca alguna circunstancia como consecuencia de la cual sus posiciones no sean números enteros. Un ejemplo de tales circunstancias pueden ser el cambio de resolución espacial de una imagen, mediante el que se cambia el tamaño final de la celda, que no tiene que ser necesariamente múltiplo ni submúltiplo del tamaño original. En lo atañe a la corrección geométrica, el resultado de la transformación polinómica obliga a realizar un remuestreo final antes de hacer utilizable la imagen. Mediante el remuestreo, la imagen transformada se adapta a la rejilla del sistema.

Al tener que situar el centro de la celda en una posición intermedia a la que proporciona la transformación, los nuevos ND, en coordenadas de la imagen de salida, se recalculan en función de los de su entorno, existiendo diversos procedimientos para ello. Se persigue encontrar para la posición final un ND que exprese el valor radiométrico más fiel con el original. Los procedimientos más usuales son el del *vecino más próximo*, la *interpolación bilineal*, y la *convolución cúbica*.

En los párrafos siguientes se utilizará la denominación de *imagen transformada* para designar a la resultante de la transformación geométrica mediante las funciones de ajuste, e *imagen corregida* a la resultante del remuestreo de la transformada.

Vecino más próximo

El método del vecino más próximo (*nearest neighbour*) consiste en asignar a cada celda el ND correspondiente al de la posición entera más cercana. Algebraicamente se puede expresar:

$$f(m,n) = \begin{cases} f(x,y) & \Delta x < 0,5 \quad \Delta y < 0,5 \\ f(x+1,y) & \Delta x \geq 0,5 \quad \Delta y < 0,5 \\ f(x,y+1) & \Delta x < 0,5 \quad \Delta y \geq 0,5 \\ f(x+1,y+1) & \Delta x \geq 0,5 \quad \Delta y \geq 0,5 \end{cases} \quad (29)$$

siendo:

$$\left. \begin{aligned} m &= x + \Delta x \quad 0 \leq \Delta x \leq 1 \\ n &= y + \Delta y \quad 0 \leq \Delta y \leq 1 \end{aligned} \right\} \quad (30)$$

donde $f(x,y)$ es el ND de la imagen transformada en una posición intermedia determinada por un número real), $f(m,n)$ es el ND de la celda (m,n) en la imagen corregida y $\Delta x, \Delta y$ representa la fracción de celda entre las posiciones (x,y) y (m,n) .

Este procedimiento no altera los ND de la imagen original, sino que solamente traslada ligeramente la posición de la celda. Presenta el inconveniente de poder introducir fracturas en el trazado de los rasgos lineales de la imagen (*efecto escalera*). Sin embargo, es el único método utilizable en la corrección de imá-

genes con información cualitativa, como son las imágenes clasificadas, debido a que respeta el valor de la celda.

Interpolación bilineal

La interpolación bilineal asigna a la celda corregida la media ponderada de los ND correspondientes a las cuatro posiciones más cercanas de la imagen transformada. El peso asociado a cada nivel digital es proporcional a la proximidad entre ellos $(1-\Delta x, 1-\Delta y)$, medida entre centros de celdas.

En lugar de calcular distancias cuadráticas, que requieren mayor número de cálculos de proceso informático, se realiza una interpolación por filas, obteniéndose primeramente dos interpolantes con los valores provisionales en la fila anterior y posterior a la posición final. En un segundo paso, se realiza una nueva interpolación entre ambos interpolantes para deducir el ND que se atribuirá a la posición (m,n) .

Los interpolantes I_1 e I_2 , correspondientes a la fila anterior y posterior a la posición buscada, tiene los siguientes valores:

$$\begin{aligned} I_1 &= (1-\Delta x)f(x,y) + \Delta x f(x+1,y) \\ I_2 &= (1-\Delta x)f(x,y+1) + \Delta x f(x+1,y+1) \end{aligned} \tag{31}$$

Y el ND final de la celda corregida será:

$$\begin{aligned} f(m,n) &= (1-\Delta y)I_1 + \Delta y I_2 = \\ &= (1-\Delta x)(1-\Delta y)f(x,y) + \Delta x(1-\Delta y)f(x+1,y) + \\ &+ (1-\Delta x)\Delta y f(x,y+1) + \Delta x\Delta y f(x+1,y+1) \end{aligned} \tag{32}$$

Considerando coeficientes de ponderación c_i para cada una de las cuatro posiciones más cercanas a (m,n) , se puede escribir:

$$f(m,n) = c_1 f(x,y) + c_2 f(x+1,y) + c_3 f(x,y+1) + c_4 f(x+1,y+1) \tag{33}$$

siendo los factores de ponderación:

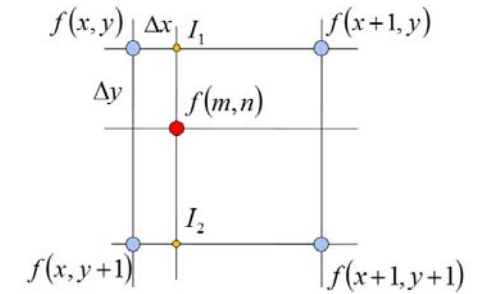
$$\left. \begin{aligned} c_1 &= (1-\Delta x)(1-\Delta y) \\ c_2 &= \Delta x(1-\Delta y) \\ c_3 &= (1-\Delta x)\Delta y \\ c_4 &= \Delta x\Delta y \end{aligned} \right\} \tag{34}$$

Este método produce resultados suavizados, pues las cuatro posiciones más próximas contribuyen al ND final. Es más apropiado cuando se trata de corregir imágenes con información cuantitativa.

Convolución cúbica

La convolución cúbica es un interpolador que emplea polinomios de tercer grado. Se involucran las dieciséis celdas más cercanas al punto considerado. Los ND se interpolan linealmente en grupos de cuatro líneas de cuatro píxeles cada una para formar cuatro interpolantes. Posteriormente se realiza otra interpolación lineal entre los cuatro valores obtenidos para asignar el resultante a la celda corregida.

Los cuatro interpolantes se calculan del siguiente modo:



Remuestreo por interpolación bilineal.

$$\begin{aligned}
 f(m) = & \left[(\Delta x)^3 - (\Delta x)^2 \right] f(x+2) - \\
 & - \left[(\Delta x)^3 - (\Delta x)^2 - (\Delta x) \right] f(x+1) + \\
 & + \left[(\Delta x)^3 - 2(\Delta x)^2 + 1 \right] f(x) - \\
 & - \left[(\Delta x)^3 - 2(\Delta x)^2 + (\Delta x) \right] f(x-1)
 \end{aligned} \quad (35)$$

Y el ND final

$$\begin{aligned}
 f(m, n) = & \left[(\Delta y)^3 - (\Delta y)^2 \right] f(m+2) - \\
 & - \left[(\Delta y)^3 - (\Delta y)^2 - (\Delta y) \right] f(m+1) + \\
 & + \left[(\Delta y)^3 - 2(\Delta y)^2 + 1 \right] f(m) - \\
 & - \left[(\Delta y)^3 - 2(\Delta y)^2 + (\Delta y) \right] f(m-1)
 \end{aligned} \quad (36)$$

Es un procedimiento muy respetuoso con los valores del entorno de la posición corregida, si bien es el que mayor capacidad de procesamiento requiere.

Si bien, para su mejor comprensión, los procesos explicados han seguido la dirección imagen transformada \rightarrow imagen corregida, metodológicamente se opera al contrario: al escribir una posición entera de columna y línea en la imagen de salida, hay que buscar el mejor ND o la forma de calcularlo a partir de las posiciones más próximas a la posición intermedia en la imagen transformada

Modelización matemática

Cuando las distorsiones sufridas por la imagen son sistemáticas es posible conocer la función que ha originado la deformación y, en consecuencia, la imagen puede ser corregida aplicando algunas transformaciones simples.

Traslación

$$\left. \begin{aligned} u &= x + a \\ v &= y + b \end{aligned} \right\} \quad (37)$$

Inclinación

$$\left. \begin{aligned} u &= x + ay \\ v &= y \end{aligned} \right\} \quad (38)$$

Cambio de escala

$$\left. \begin{aligned} u &= ax \\ v &= by \end{aligned} \right\} \quad (39)$$

Perspectiva

$$\left. \begin{aligned} u &= axy \\ v &= y \end{aligned} \right\} \quad (40)$$

Rotación

$$\left. \begin{aligned} u &= \cos \theta x + \text{sen } \theta y \\ v &= -\text{sen } \theta x + \cos \theta y \end{aligned} \right\} \quad (41)$$

Si bien las distorsiones introducidas en la imagen como consecuencia de la oscilación de la plataforma, al ser aleatorias suelen corregirse recurriendo a las transformaciones polinómicas, algunas otras pueden corregirse mediante modelización matemática aplicando operadores sencillos. Por ejemplo, para el efecto panorámico puede aplicarse la transformación:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} (h/x) \arctg(x/h) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (42)$$

donde h es la altura del satélite.

La inclinación de la órbita se corrige haciendo:

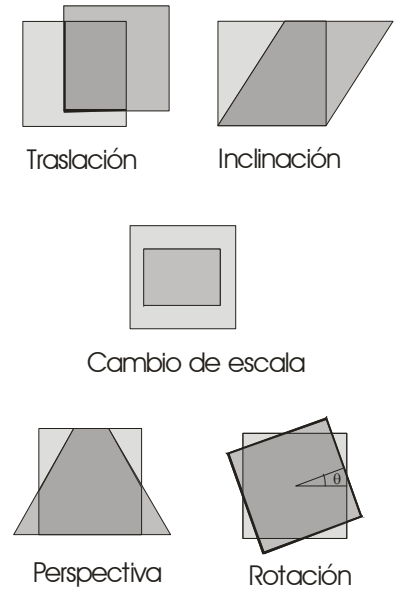
$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos \theta & \text{sen } \theta \\ -\text{sen } \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (41)$$

siendo θ el ángulo formado por la traza del satélite y el paralelo del lugar.

La ventaja de trabajar con operadores matriciales es que se pueden combinar varias correcciones simplemente multiplicando las matrices que caracterizan a cada una. Por ejemplo, la eliminación del desplazamiento de líneas debido a la rotación terrestre y la

inclinación de la órbita se corrigen:

$$\begin{aligned} \begin{bmatrix} u \\ v \end{bmatrix} &= \begin{bmatrix} 1 & -\alpha \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & \text{sen } \theta \\ -\text{sen } \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \\ &= \begin{bmatrix} \cos \theta + \alpha \text{sen } \theta & \text{sen } \theta - \alpha \cos \theta \\ -\text{sen } \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \end{aligned} \quad (43)$$



Transformaciones geométricas simples.

12

Álgebra de imágenes

Introducción

Con el objetivo de poner de manifiesto ciertos rasgos de la imagen que pueden quedar enmascarados con la abundancia de datos, o bien con el fin de destacar algún otro tipo de información poco evidente o incluso aparentemente oculta, se pueden realizar operaciones algebraicas entre bandas de una misma imagen o entre imágenes diferentes. Mediante el álgebra de bandas o de imágenes se sintetiza en un solo valor información procedente de varias o de todas las bandas de la imagen.

Las transformaciones globales, es decir, aquellas en las que *toda* la imagen resultante depende de *toda* la

imagen o imágenes originales, se basan en el álgebra. En general, el álgebra de imágenes proporciona nuevas imágenes con bandas obtenidas a partir combinaciones lineales de las bandas originales.

Existen multitud de operaciones que pueden realizarse con imágenes o con bandas de ellas, orientadas cada una a un objetivo concreto. En esta lección se estudiarán el análisis de componentes principales, la transformación *Tasseled Cap*, los cocientes de bandas y, como caso particular de ellos, los índices de vegetación.

Análisis de componentes principales

El análisis de componentes principales (ACP) o transformación a componentes principales es una técnica estadística multivariante muy utilizada para definir la mejor combinación de nuevas bandas capaz de conservar la mayor parte de la información de la imagen completa.

El ACP busca los ejes de mayor variabilidad e información de la imagen y es de gran utilidad para los estudios multitemporales, o para seleccionar tres nuevas bandas para una composición en color que concentren más información que cualesquiera de las originales.

La distribución de niveles digitales en una banda de una imagen multispectral puede considerarse una

variable aleatoria y, como tal, puede someterse a cualquiera de las técnicas estadísticas habituales

La función de densidad univariante de la distribución de los ND es en realidad el histograma de cada banda. Para facilitar el trabajo con los datos, dicha distribución puede asimilarse a una distribución normal. De esta forma, la distribución de los ND cada banda queda caracterizada completamente por sólo dos parámetros: la media y la varianza de sus datos.

A su vez, considerada la imagen multispectral completa, los ND multispectrales pueden considerarse como una variable aleatoria vectorial multivariante y cada banda constituye una componente del vector. La función de densidad multivariante es ahora el histograma multispectral de la imagen y el equivalente

multiespectral del valor promedio es el vector de medias.

La relación entre bandas está caracterizada por las covarianzas o los correspondientes coeficientes de correlación.

En una imagen multiespectral, el vector

$$\bar{z} = (z_1, z_2, \dots, z_n) \tag{1}$$

Se denomina *vector de características* y representa de forma vectorial los ND de las n bandas de la imagen.

Dentro de una determinada banda K , el valor medio es:

$$\mu_k = \sum_{i,j} \frac{z_{ij(k)}}{N} \tag{2}$$

en donde N es el número de celdas de la banda, igual al producto del número de columnas por el de líneas.

La varianza de esa banda es:

$$\sigma_k^2 = \sum_{i,j} \frac{(z_{ij(k)} - m_k)^2}{N} \tag{3}$$

La covarianza entre dos bandas K y L es:

$$\sigma_{k,l} = \sum_{i,j} \frac{[z(i,j)_k - m_k][z(i,j)_l - m_l]}{N} \tag{4}$$

Y conviene recordar que $\sigma_{k,l} = \sigma_{l,k}$ y que $\sigma_{k,k} = \sigma_k^2$,

esto es, la varianza de la propia banda.

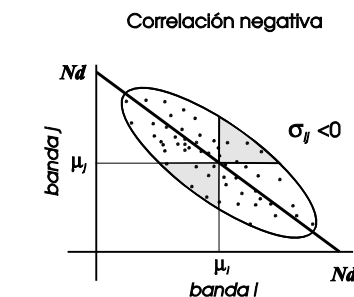
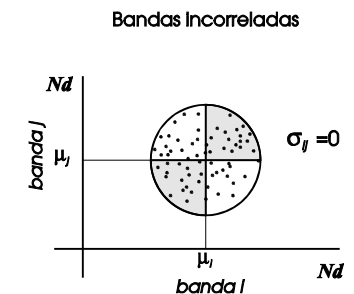
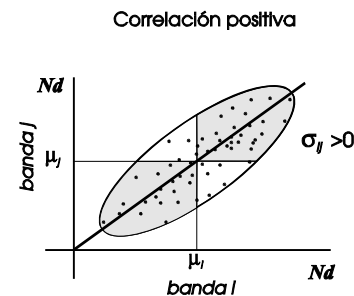
Finalmente, el coeficiente de correlación entre esas dos bandas es:

$$\rho_{k,l} = \frac{\sigma_{k,l}}{\sqrt{\sigma_{k,k}\sigma_{l,l}}} = \frac{\sigma_{k,l}}{\sqrt{\sigma_k^2\sigma_l^2}} = \frac{\sigma_{k,l}}{\sigma_k\sigma_l} \tag{5}$$

La varianza de una banda es una medida de la dispersión de sus ND en torno al valor medio y representa el contraste de la imagen en esa banda. Visto desde otro punto de vista, es una buena medida de la cantidad de información que proporciona esa banda. Por su parte, la covarianza y el coeficiente de correlación entre dos bandas indican el grado de relación que hay entre ambas, este último de forma normalizada entre -1 y 1. Las bandas correladas muestran distribuciones de sus hiperelipsoides de puntos en el espacio de características alargadas. En las incorreladas, sin embargo, las nubes presentan simetría con respecto a ejes paralelos a los de coordenadas, aunque no necesariamente tienen que tener forma esférica o hiperesférica.

Considerando las n bandas de la imagen, se pueden definir las matrices de covarianza (6) y de correlación (7), ambas de dimensión $n \times n$ y simétricas:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix} \tag{6}$$



Correlación entre bandas.

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{bmatrix} \quad (7)$$

La matriz de covarianzas de un conjunto de bandas linealmente independientes es una matriz diagonal, porque todas las covarianzas son nulas.

Por lo tanto, si se quiere obtener un conjunto de bandas linealmente independientes a partir de otro conjunto cuyas bandas no lo son, hay que buscar una transformación que consiga diagonalizar la matriz de covarianzas de este último.

Diagonalización de la matriz de covarianzas

Dada una matriz de transformación \mathbf{A} , se denominan vectores propios o autovectores de \mathbf{A} a los que son colineales con sus transformados:

$$\mathbf{A}\vec{\mathbf{X}} = \lambda\vec{\mathbf{X}} \quad (8)$$

También se puede escribir:

$$\mathbf{A}\vec{\mathbf{X}} = \lambda\mathbf{I}\vec{\mathbf{X}} \quad (9)$$

Siendo \mathbf{I} la matriz identidad:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad (10)$$

La expresión (9) se puede poner de la forma:

$$(\mathbf{A} - \lambda\mathbf{I})\vec{\mathbf{X}} = 0 \quad (11)$$

Y para que esta ecuación se verifique, se requiere resolver la ecuación característica que proporciona los valores propios de la matriz \mathbf{A} igualando a cero el determinante:

$$\phi(\lambda) \equiv |\mathbf{A} - \lambda\mathbf{I}| = 0 \quad (12)$$

Particularizando para la matriz de covarianzas, la ecuación característica sería:

$$\phi(\lambda) \equiv |\Sigma - \lambda\mathbf{I}| = 0 \quad (13)$$

Que, desarrollada, quedaría:

$$|\Sigma - \lambda\mathbf{I}| = \begin{vmatrix} \left(\begin{matrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{matrix} \right) & - & \left(\begin{matrix} \lambda & 0 & \dots & 0 \\ 0 & \lambda & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda \end{matrix} \right) \end{vmatrix} =$$

$$= \begin{vmatrix} \sigma_1^2 - \lambda & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 - \lambda & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 - \lambda \end{vmatrix} = 0 \quad (14)$$

Resolviendo la ecuación característica se obtienen n soluciones $\lambda_1, \lambda_2, \dots, \lambda_n$, que son los valores propios o autovalores de la citada matriz, los cuales habitualmente se ordenan de forma que $\lambda_1 > \lambda_2 > \dots > \lambda_n$.

Los vectores propios se obtienen resolviendo para cada valor propio la expresión:

$$\Sigma \bar{\mathbf{X}}_i = \lambda_i \bar{\mathbf{X}}_i \quad (15)$$

es decir:

$$\begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22}^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn}^2 \end{bmatrix} \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{bmatrix} = \begin{bmatrix} \lambda_i x_{i1} \\ \lambda_i x_{i2} \\ \vdots \\ \lambda_i x_{in} \end{bmatrix} \quad (16)$$

Y la transformación que diagonaliza la matriz de covarianzas es:

$$\mathbf{P}^{-1} \Sigma \mathbf{P} = \Lambda \quad (17)$$

donde \mathbf{P} es una matriz de paso constituida por los vectores propios puestos por columnas en el mismo orden que los valores propios asociados a cada uno de ellos:

$$\mathbf{P} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \dots & x_{nn} \end{bmatrix} \quad (18)$$

y Λ es la matriz diagonal conformada con los valores propios ordenados:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \quad (19)$$

La demostración de (17) es trivial, pero queda fuera del alcance de este texto.

Es interesante y, en el caso del análisis digital de imágenes, muy importante considerar que la transformación a componentes principales es una transformación ortogonal, es decir, la traza de Σ es igual a la de Λ :

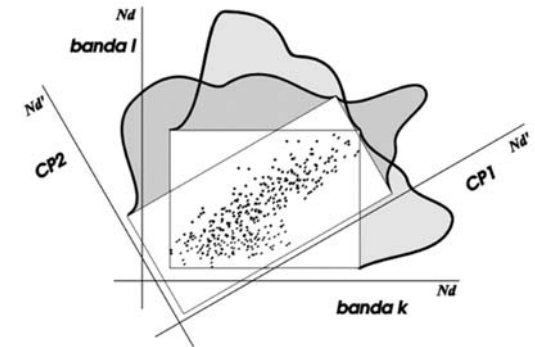
$$\sum_{i=1}^n \sigma_i^2 = \sum_{i=1}^n \lambda_i \quad (20)$$

lo que quiere decir que la varianza, es decir, la información, se redistribuye sin pérdidas entre los componentes principales. Además, a la primera componente le corresponde la mayor parte de la varianza de la imagen. De la varianza restante, su mayor parte queda explicada por la siguiente componente, y así sucesivamente. Dicho de otro modo, las nuevas bandas, llamadas ahora componentes principales contienen toda la información original, pero sin redundancia entre ellas, y están ordenadas de mayor a menor grado de información.

Utilidad en teledetección

El ACP es utilizado en el análisis digital de imágenes para mejorar el análisis visual al introducir prácticamente toda la información de la imagen en la composición en color, que solamente puede hacer uso de tres bandas, si bien hay que tener en cuenta que el peso de información que se le da a cada canal es diferente, puesto los componentes principales retienen cantidades diferentes de varianza.

En teledetección multitemporal es muy útil para estudiar de series temporales. El procedimiento que suele seguirse es el siguiente. Se extrae el primer componente principal de las bandas de la imagen de



Cambio de base en la transformación ACP. Dos bandas correladas dejan de estarlo si se las somete a un giro en el espacio de características de modo que los nuevos ejes, componentes principales, se orienten según los vectores propios de la distribución original.

cada una de las fechas. Posteriormente se realiza una clasificación con los CP. Quedarán al descubierto las áreas de la imagen que muestren diferencias entre unos componentes y otros, o lo que es lo mismo, evolución entre fechas diferentes.

Para la detección de cambios entre dos fechas, se transforman a CP el conjunto de todas las bandas de ambas fechas. Las primeras CP explicarán la mayor

parte de la varianza de la imagen, y por tanto incluirán todos los aspectos comunes a las dos fechas. Las componentes secundarias contendrán información sobre aquellos fenómenos que no puedan explicarse en las dos fechas simultáneamente, por lo que indicarán la evolución de los mismos. Las últimas componentes principales retienen la información espuria de la imagen, esto es, el ruido.

Transformación *tasseled cap*

Se denomina diagrama de dispersión a la representación de la nube de puntos espectral, esto es, de niveles digitales o de las magnitudes físicas de reflectancia o radiancia, en un espacio espectral o de características de dos o más dimensiones. Lógicamente, sobre un papel solamente pueden representarse diagramas de dispersión bidimensionales, pero un sistema informático no tiene el menor problema para procesar diagramas n -dimensionales.

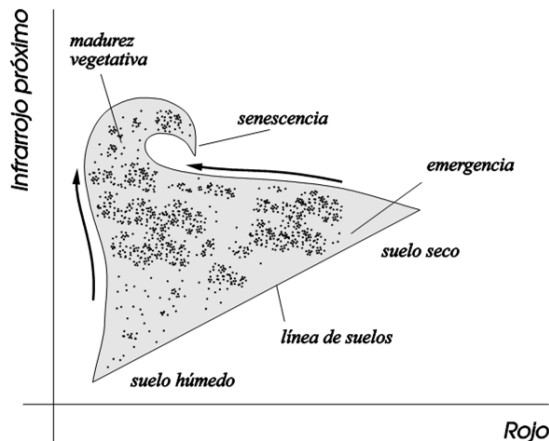
Evolución espectral de los terrenos cultivados

Se puede estudiar la posición espectral de las celdas de terreno cultivado en un diagrama de dispersión desde el estado de nacimiento hasta la madurez y el agostamiento del cultivo. A lo largo de su período vegetativo el rastro que va dejando la nube de puntos en su evolución espectral en un espacio tridimensional adquiere la forma de un gorro borlado (*tasseled cap*). Si solo se consideran las bandas del infrarrojo próximo y del rojo espectral, la base del gorro está definida por la línea de suelos, a lo largo de la cual se distribuyen los distintos tipos de suelos, de acuerdo a su estado (humedad e iluminación, fundamentalmente).

A medida que el cultivo se desarrolla y llega a su pleno vigor, los puntos tienden a aproximarse al eje del infrarrojo, alejándose de la línea de suelos. Cuando se inicia el período de senescencia, la nube de puntos va convergiendo hacia un punto único, que alcanza cuando el cultivo llega al estado de marchitez permanente, y cuya posición tiende a regresar de nuevo a la línea de suelos.

La transformación *tasseled cap* (TCT, *Tasseled Cap Transformation*) es una herramienta que permite comprimir la información espectral de una imagen en pocas bandas asociadas a ciertas características de la escena, sin perder información significativa. Consiste en un giro en el espacio espectral mediante el cual se pone de manifiesto la evolución espectral de los cultivos anuales o bienales, desde su nacimiento hasta la madurez, el agostamiento y la marchitez final.

Tras el giro, los nuevos ejes en un espacio espectral tridimensional representan las variables *brillo*, *verdor* y *humedad*. Los ejes brillo y verdor definen el plano de la *vegetación* y los ejes brillo y humedad, el de los *suelos*.



Evolución espectral de un suelo de cultivo

Esta transformación es empírica y se aplica con diferentes matrices a cada imagen de satélite, dependiendo del número de bandas y de la longitud de

onda de cada una de ellas. En los ejemplos siguientes se muestran las transformaciones sobre Landsat-TM y sobre Landsat 8-OLI:

TM:

$$\begin{bmatrix} \text{brillo} \\ \text{verdor} \\ \text{humedad} \end{bmatrix} = \begin{bmatrix} 0,3037 & 0,2793 & 0,4743 & 0,5585 & 0,5082 & 0,1863 \\ -0,2848 & -0,2435 & -0,5436 & 0,7243 & 0,0840 & -0,1800 \\ 0,1509 & 0,1973 & 0,3279 & 0,3406 & -0,7112 & -0,4572 \end{bmatrix} \cdot \begin{bmatrix} TM1 \\ TM2 \\ TM3 \\ TM4 \\ TM5 \\ TM7 \end{bmatrix}$$

OLI:

$$\begin{bmatrix} \text{Brillo} \\ \text{Verdor} \\ \text{Humedad} \\ \text{TCT4} \\ \text{TCT5} \\ \text{TCT6} \end{bmatrix} = \begin{bmatrix} 0,3029 & 0,2786 & 0,4733 & 0,5599 & 0,5080 & 0,1872 \\ -0,2941 & -0,2430 & -0,5424 & 0,7276 & 0,0713 & -0,1608 \\ 0,1511 & 0,1973 & 0,3283 & 0,3407 & -0,7117 & -0,4559 \\ -0,8239 & 0,0849 & 0,4393 & -0,0580 & 0,2013 & -0,2773 \\ -0,3294 & 0,0557 & 0,1056 & 0,1855 & -0,4349 & 0,8085 \\ 0,1079 & -0,9023 & 0,4119 & 0,0575 & -0,0259 & 0,0252 \end{bmatrix} \cdot \begin{bmatrix} OLI2 \\ OLI3 \\ OLI4 \\ OLI5 \\ OLI6 \\ OLI7 \end{bmatrix}$$

Como puede verse en ambas, los coeficientes de las matrices de giro buscan ponderar equilibradamente todas las bandas para buscar la componente brillo. En cambio, enfatizan el contraste que existe entre el infrarrojo próximo y el visible, particularmente el rojo, para deducir la componente verdor. Finalmente, la humedad queda expresada contraponiendo los canales del infrarrojo medio contra los demás.

La primera es obvia. El verdor se basa en la diferencia de reflectancia que existe entre el infrarrojo próximo y el rojo si la vegetación es vigorosa y el contenido en clorofila fotosintéticamente activa es alto. Por último, son muy características de la molécula de agua sus bandas de absorción en diversos intervalos del espectro en el infrarrojo medio.

Índices de vegetación

En la superficie sólida del planeta solamente caben encontrar unos cuantos tipos de superficies genéricas: o bien se trata de suelo desnudo, o bien de vegetación, o de la mezcla de ambos en cualquier proporción, o bien es agua en estado sólido o en masas líquidas. Los índices de vegetación son elementos adecuados para caracterizar el tipo de cubierta y sintetizar la información en una sola banda. Además, son útiles para deducir otras variables de gran trascendencia en los estudios físicos de la superficie, como es el caso de la emisividad.

Cocientes de bandas

Los cocientes son transformaciones consistentes en dividir celda a celda todos los ND pertenecientes a una banda por los de otra. El caso más general puede expresarse:

$$z' = C_0 + C_1 \frac{(z_k + a)}{(z_l + b)} \tag{21}$$

aunque no necesariamente hay que añadir constantes aditivas.

En ciertas ocasiones, los cocientes proporcionan ventajas en la interpretación de las imágenes. Cuando es difícil discriminar entre dos cubiertas cuyos ND en varias bandas son muy similares, el cociente entre ellas puede ofrecer un mejor criterio de separabilidad.

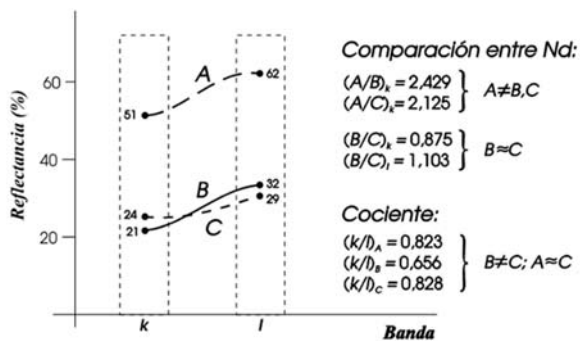
Para la misma irradiancia extraterrestre, el relieve origina gradientes de irradiancia local y, en consecuencia, niveles variables de radiancia que son independientes de la ocupación de suelo y que, además, afectan a todas las bandas de la imagen. Por el contrario, el cociente entre bandas independiza de la irradiancia la interpretación de los resultados. En consecuencia, es una transformación muy indicada para reducir el efecto radiométrico del relieve en la imagen final.

Por ejemplo, en igualdad de condiciones de iluminación, las coníferas siempre son más reflexivas que las caducifolias, tanto en el visible como en el infrarrojo próximo. En cambio, una conífera en sombra brilla menos que una caducifolia al sol. El cociente IRp/R soluciona esta cuestión. Este cociente es un índice de vegetación.

Índices de vegetación

Los índices de vegetación son transformaciones sencillas y rápidas cuyo objetivo es realzar las zonas cubiertas por material fotosintéticamente activo, maximizando la contribución espectral inherente a la cubierta vegetal, en tanto que minimizan la influencia de otros factores que pueden distorsionar la interpretación, como pueden ser el tipo de suelo, la irradiancia solar, etc.

Existen dos grandes grupos de índices de vegetación, los que se basan en transformaciones ortogonales y los basados en cocientes entre bandas.



Comparación directa de ND y de sus cocientes.

Los índices basados en transformaciones recorren a giros en el espacio espectral IRp/R. Para construir el Índice de Vegetación Perpendicular (PVI, Perpendicular Vegetation Index), uno de los nuevos ejes queda orientado según la dirección de la línea de suelos, y el otro perpendicularmente a ella. El PVI mide la distancia espectral entre el vector de características de la celda estudiada y la línea de suelos:

$$PVI = \sqrt{[\rho(IRp_v) - \rho(IRp_s)]^2 + [\rho(R_v) - \rho(R_s)]^2} \tag{22}$$

donde ρ es la reflectancia en los canales IRp y R y los subíndices v y s se refieren a las celdas con vegetación y a las de suelo desnudo. En general, las celdas ocupadas por suelo desnudo tendrán PVI=0, las ocupadas por agua podrán tener PVI<0 y las cubiertas por vegetación tendrán valores de PVI positivos y tanto mayores cuando más vigorosa sea la vegetación de la celda.

Es un rasgo muy característico de la vegetación a medida que cubre el suelo una disminución progresiva de la reflectancia en el visible, particularmente en el rojo espectral debido a la absorción de la luz por parte de los pigmentos fotosintéticos, especialmente la clorofila, y un aumento simultáneo de la reflectancia en el infrarrojo próximo. Además, las líneas de isocobertura, para distintos tipos de atmósfera se van inclinando a medida que el cultivo crece y madura.

Un índice que proporciona información acerca de la cantidad de masa foliar en un determinado punto es el índice de área foliar (LAI, Leaf Area Index) que es la proporción de suelo cubierto eficazmente por hojas u otro material vegetal. Conforme se desarrolla una cubierta vegetal, su índice foliar se incrementa hasta llegar a un máximo (próximo a la floración). A

partir de entonces decrece hasta alcanzar el punto de marchitez permanente. Durante el crecimiento del cultivo la contribución reflectora del suelo va disminuyendo. El LAI puede superar ampliamente la unidad.

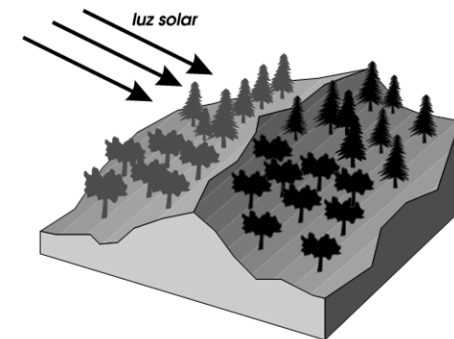
Entre los índices basados en cocientes, los más utilizados consisten en los que combinan bandas del infrarrojo próximo con las del visible y se basan, como se ha dicho en la alta reflectancia de la vegetación en el IRp y la baja en el VIS, especialmente en el rojo. El más simple de los índices VIS-IRp es el Índice de Vegetación de la Razón RVI (*Ratio Vegetation Index*):

$$RVI = \frac{\rho_{IRp}}{\rho_R} \tag{23}$$

elaborado con los cocientes celda a celda de las bandas IRp y R de la imagen. Si bien este índice muestra claramente la relación entre las relectancias de las bandas citadas tiene el inconveniente de no estar acotado y tender a valores muy altos cuando el denominador es pequeño, razón por la cual no son fáciles las comparaciones. Por esta razón se prefiere *normalizar* el índice y elaborar el denominado Índice de Vegetación de Diferencia Normalizada NDVI (*Normalized Difference Vegetation Index*):

$$NDVI = \frac{\rho_{IRp} - \rho_R}{\rho_{IRp} + \rho_R} \tag{24}$$

La ventaja de este índice es que está acotado entre -1 y 1. Superficies sin vegetación proporcionan valores de NDVI próximos a cero, tanto positivos como negativos. Valores altos de NDVI indican celdas con alto contenido en vegetación vigorosa y fotosintéticamente activa. Valores francamente negativos de



Nd		Caducifolias		Coníferas	
		sol	sombra	sol	sombra
Banda	R	89	49	96	53
	IRp	124	68	173	95
	IRp/R	1,393	1,388	1,802	1,792

Problemas en la interpretación de la cobertura de una imagen si coexisten laderas bien y mal iluminadas.

NDVI pueden indicar superficies de agua o cubiertas nubosas.

Es importante hacer notar que tanto éste como el resto de los índices están definidos y se calculan con las reflectancias, no con los niveles digitales directamente.

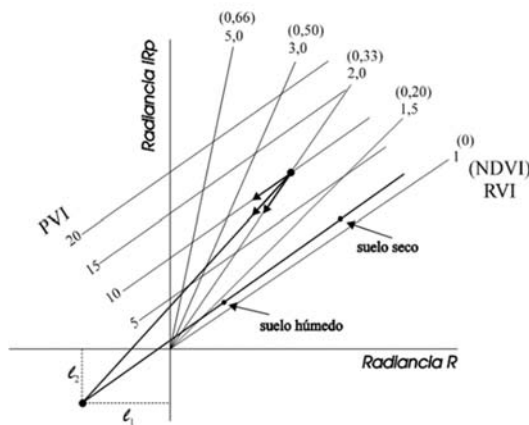
Directamente relacionado con el anterior está el índice de vegetación transformado TVI (*Transformed Vegetation Index*):

$$TVI = \sqrt{|NDVI + 0,5|} \quad (25)$$

que evita los valores negativos del NDVI. La introducción de la raíz cuadrada persigue asemejar la distribución de valores a una distribución normal. No hay grandes diferencias entre los resultados obtenidos con con NDVI y TVI. Se suele tomar el umbral $\sqrt{2}/2$ para delimitar las áreas de vegetación. Valores superiores a esta cifra indican cobertura vegetal.

Corrección de la influencia del suelo

En numerosas ocasiones la vegetación no tapiza completamente el suelo y el índice que se calcule puede no reflejar fielmente el estado vegetativo de la cubierta simplemente por el hecho de estar la radiancia contaminada por el efecto del suelo subyacente. Esto sucede tanto en cultivos herbáceos en hilera como en cultivos leñosos, pero también es frecuente en bosques del tipo mediterráneo donde la cubierta es escasa y entre árboles y arbustos se deja ver el suelo desnudo. En estos casos ha de recurrirse a procedimientos que minimicen ese efecto del suelo parcialmente cubierto.



Evolución espectral de una celda con la humedad.

La evolución espectral de una celda que paulatinamente aumenta su humedad en un diagrama IRp/R es contradictoria en función de qué tipo de índice se utilice: si se utiliza el PVI, el desplazamiento se realizará según líneas paralelas a la línea de suelo, puesto que el contenido de material vegetal es invariable. Sin embargo, si se utiliza el criterio del RVI o del NDVI, la celda tiende hacia el origen de coordenadas.

El índice de vegetación ajustado al suelo, SAVI (*Soil Adjusted Vegetation Index*) proporciona una solución de compromiso entre ambas situaciones y reúne características comunes a los anteriores. El punto de convergencia de la bisectriz de cada par de isolíneas con la línea de suelo está en $(-\ell, -\ell)$, y de valores experimentales se deduce que habitualmente $(\ell \approx \ell \approx 0,25)$ unidades de reflectancia. Para eliminar los valores negativos en el diagrama IRp/R se traslada el origen a la posición $(-\ell, -\ell)$, obteniéndose un NDVI desplazado:

$$NDVI_{desplazado} = \frac{(\rho_{IRp} + \ell_2) - (\rho_R + \ell_1)}{(\rho_{IRp} + \ell_2) + (\rho_R + \ell_1)}$$

Si se particulariza este índice para los valores de intersección de la prolongación de la línea de suelos y la bisectriz de los pares de isolíneas ($\ell \approx \ell \approx 0,25$) y se multiplica por una constante que homologue su intervalo de existencia al del NDVI se tiene finalmente:

$$SAVI = \frac{\rho_{IRp} - \rho_R}{\rho_{IRp} + \rho_R + L} \cdot (1 + L) \quad (26)$$

Donde L adquiere el valor 1 para baja proporción de vegetación, 0,5 para una cobertura intermedia y 0,25 para alto contenido de vegetación.

Otros índices de vegetación

Existen otros muchos índices de vegetación relacionados o derivados de los anteriores, apropiados cada uno de ellos para circunstancias defendidas por sus autores respectivos. A continuación se muestra una mínima parte de la que los investigadores han ido desarrollando a lo largo de los últimos años.

- SAVI modificado:

$$MSAVI = \frac{2\rho_{IRp} + 1 - \sqrt{[2\rho_{IRp} + 1]^2 - 8[\rho_{IRp} - \rho_R]}}{2} \quad (27)$$

- SAVI transformado:

$$TSAVI = \frac{a[\rho_{IRp} - a\rho_R - b]}{\rho_R + a\rho_{IRp} - ab + X(1 + a^2)} \quad (28)$$

donde a es la pendiente de la línea de suelos, b es la ordenada en el origen de dicha línea y $X \approx 0,08$ es un parámetro a determinar según el tipo de cultivo.

- GEMI (*Global Environment Monitoring*):

$$GEMI = \frac{\eta(1 - 0,25\eta) - (\rho_R - 0,125)}{1 - \rho_R} \quad (29)$$

Donde:

$$\eta = \frac{2(\rho_{IRp}^2 - \rho_R^2) + 1,5\rho_{IRp} + 0,5\rho_R}{\rho_{IRp} + \rho_R + 0,5}$$

- GESAVI (*Generalized SAVI*):

$$GESAVI = \frac{\rho_{IRp} - a\rho_R - b}{\rho_R + Z} \quad (30)$$

donde a y b son respectivamente la pendiente y la ordenada en el origen de la línea de suelos en un diagrama IRp/R:

$$(IRp_{suelo} = aR_{suelo} + b)$$

Z hace referencia al valor absoluto de la abscisa del punto de corte de las isolíneas de vegetación con la línea de suelo. Un valor empírico adecuado es $Z=0,35$.

- OSAVI (*Optimized SAVI*):

$$OSAVI = \frac{\rho_{IRp} - \rho_R}{\rho_{IRp} + \rho_R + Y}(1 + Y) \quad (31)$$

donde Y es un parámetro de ajuste de valor 0,16. Para bajos valores de LAI, el OSAVI independiza el resultado del suelo subyacente muy eficazmente. Se utiliza sobre todo para el cálculo de la biomasa sobre el suelo, nivel de nitrógeno en hoja y contenido de clorofila, Este índice equivale al TSAVI con $a=1$ y $b=0$.

- ARVI (*Atmosferically Resistant Vegetation Index*):

$$ARVI = \frac{\rho_{IRp} - [2\rho_R - \rho_A]}{\rho_{IRp} + [2\rho_R - \rho_A]} \quad (32)$$

en cuya expresión, ρ_A es la reflectancia en el azul espectral.

- *CARI (Chlorophyll Absorption Ratio Index):*

$$CARI = \frac{|a \cdot 670 + \rho_{670} + b| \rho_{700}}{\sqrt{a^2 + 1} \rho_{670}} \quad (33)$$

$$\text{con } a = \frac{\rho_{700} - \rho_{500}}{150}; \quad b = \rho_{550} - (a \cdot 550)$$

Este índice minimiza el efecto de los materiales no fotosintéticos en la cubierta vegetal.

- *MCARI (CARI modificado):*

$$MCARI = \frac{2,5(\rho_{800} - \rho_{670}) - 1,3(\rho_{800} - \rho_{550})}{\sqrt{(2\rho_{800} + 1)^2 - (6\rho_{800} - 5\rho_{670})} - 0,5} \cdot 1,5 \quad (34)$$

El *MCARI* es más sensible a la concentración de clorofila en las hojas.

- *TCARI (CARI transformado):*

$$TCARI = 3 \cdot \left[(\rho_{700} - \rho_{670}) - 0,2 \cdot (\rho_{700} - \rho_{550}) \frac{\rho_{700}}{\rho_{670}} \right] \quad (35)$$

Mejora la sensibilidad ante bajos contenidos en clorofila.

- Índice de suma verde *SGI (Sum Green Index):*

Es un índice hiperespectral muy simple sensible a pequeños cambios en la densidad del dosel vegetal y,

por tanto, válido para hacer el seguimiento de las cubiertas forestales. Consiste en calcular el promedio de la reflectancia en el intervalo espectral comprendido entre 500 y 600 nm.

$$SGI = \frac{\sum_{\lambda=500}^{600} \rho_{\lambda}}{N} \quad (36)$$

Con valores $0 \leq SGI \leq \sim 50\%$ en porcentaje de reflectancia. Para la vegetación verde el rango más común es del 10 al 25 %.

Índices de humedad

Al igual que mediante cocientes se puede determinar el contenido en vegetación fotosintéticamente activa en una celda de imagen, también existen índices similares para medir el contenido de humedad. Dos índices representativos de ello son los siguientes, para cuyo cálculo se necesita conocer la reflectancia en ciertas bandas muy estrechas del espectro, lo que obliga a utilizar información hiperespectral.

- *WBI (Water Band Index):*

$$WBI = \frac{\rho_{900}}{\rho_{970}} \quad (37)$$

La absorción alrededor de 970 nm aumenta en relación a la de 900 nm cuando aumenta el contenido de agua del dosel vegetal. Este índice se aplica al análisis de estrés hídrico de la vegetación, a la predicción de la producción, al análisis de riesgo de incendio, y en general en la gestión agrícola y en estudios medioambientales. Rangos habituales para la vegetación verde son de 0,8 a 1,2.

- NDWI (*Normalized difference Water Index*):

$$NDWI = \frac{\rho_{857} - \rho_{1241}}{\rho_{857} + \rho_{1241}} \quad (38)$$

El NDWI es sensible a los cambios en la humedad de la vegetación. La dispersión de la luz por el dosel vegetal favorece la absorción del agua en 1241 nm. Se aplica al análisis de estrés del dosel forestal, al estudio del LAI áreas de vegetación densa, a los modelos de productividad vegetal y al riesgo de incendios. El NDWI está acotado entre -1 y 1, y sus valores usuales para la vegetación verde son de -0,1 a 0,4.

- MSI (*Moisture Stress Index*):

$$MSI = \frac{\rho_{1599}}{\rho_{819}} \quad (39)$$

El MSI es sensible al incremento de la humedad foliar. La absorción en 1599 nm aumenta con el contenido de agua y en cambio la registrada en 819 nm permanece prácticamente invariable. Este índice se aplica al análisis de estrés vegetal, a la predicción de cosecha, al estudio de los riesgos de incendio y en estudios ambientales. La interpretación del MSI es inversa en relación con otros índices: valores más altos indican mayor estrés hídrico. El rango puede oscilar entre 0 y más de 3, siendo usuales para la vegetación verde valores entre 0,4 y 2.

- NDII (*Normalized Difference Infrared Index*):

$$NDII = \frac{\rho_{819} - \rho_{1649}}{\rho_{819} + \rho_{1649}} \quad (37)$$

El NDII es sensible a los cambios en el contenido de

agua de la cubierta vegetal. Sus valores aumentan con el contenido de agua. Sus aplicaciones principales son la gestión agrícola y forestal y la detección de estrés hídrico de la vegetación. Al igual que el NDWI, sus valores están acotados entre -1 y 1, y los más comunes para la vegetación verde están comprendidos entre 0,02 y 0,6.

Índice de área quemada

Se han desarrollado también índices basados en cocientes para calcular la severidad de los incendios forestales. Uno representativo es el siguiente.

- MSI (*Normalized Burn Ratio*):

$$NBR = \frac{\rho_{865} - \rho_{10895}}{\rho_{865} + \rho_{10895}} \quad (40)$$

Para una zona determinada, se calcula el NBR antes y después de un incendio forestal y se determina el incremento en el NBR:

$$DNBR = NBR_{pre} - NBR_{pos}$$

en donde los subíndices pre y pos indican las situaciones previas y posteriores al incendio. Valores inferiores a -0,25 indican una revegetación intensa posterior al incendio. Valores entre -0,25 y -0,1 indican baja revegetación. Valores entre -0,1 y 0,1 indican áreas no quemadas. Valores de 0,1 a 0,27 indican baja severidad del incendio. Valores de 0,27 a 0,44 indican severidad de baja a moderada de los efectos del fuego. Valores de 0,44 a 0,66 indican severidad de moderada a alta del fuego. Finalmente, valores superiores a 0,66 indican una alta severidad de los efectos del incendio.

Si no se dispone de información hiperespectral, el índice puede acomodarse a la anchura de banda del sensor TM o ETM+ de Landsat 5 o 7, o sus bandas equivalentes de OLI en el satélite Landsat 8:

$$NBR = \frac{TM\ 4 - TM\ 6}{TM\ 4 + TM\ 6}$$

13

Clasificación digital de imágenes

Introducción

Una imagen de satélite es un conjunto de datos con información cuantitativa. Tal como se definió, la imagen digital es un modelo de reflectancia del terreno, que bien se puede referir a un solo intervalo espectral, en el caso de las imágenes pancromáticas mono-banda, o bien puede estar constituida por un conjunto de bandas, si se trata imágenes multi o hiperespectrales. En cualquier caso, la información que proporciona la imagen está contenida exclusivamente en ese conjunto de datos. Ahora bien, la interpretación de tal información cuantitativa no es inmediata, especialmente para un usuario medio no familiarizado. Para este tipo de destinatario, es preferible etiquetar las áreas de la imagen con las características temáticas buscadas, para así facilitar su comprensión. Incluso el personal investigador puede requerir calificar la imagen para seleccionar áreas de interés o encontrar en la escena determinadas cualidades buscadas. La clasificación de imágenes se orienta en el campo de la teledetección a la ocupación de suelo o, si la resolución espectral de la imagen lo permite, a establecer categorías de las relacionadas en una leyenda previamente confeccionada.

La clasificación digital es un proceso de generalización temática que categoriza la información de la imagen, esto es, convierte la información cuantitativa en información cualitativa. Como en cualquier proceso de generalización cartográfica, siempre se pierde información con relación al conjunto original de datos, pero en el caso de la clasificación, esa merma está justificada por la mejora en la interpretación y en la usabilidad de la imagen.

El problema en el ámbito de la clasificación de imágenes espaciales queda planteado del siguiente modo: se trata de establecer a qué clase pertenece una determinada celda, de entre las i clases que integran la leyenda, en virtud de las componentes de su vector de características. Por ejemplo, se pretende establecer si una celda que presenta unos determinados niveles digitales en cada una de las distintas bandas de la imagen está ocupada por un suelo desnudo, un cultivo herbáceo, una zona urbana o un bosque de coníferas.

La imagen como variable aleatoria multivariante

En análisis digital de imágenes es muy común recurrir a métodos paramétricos de clasificación por la sencillez del tratamiento.

Los niveles digitales en una imagen de satélite pueden considerarse como una variable aleatoria. Será

univariante cuando la imagen tenga una sola banda y multivariante en el caso de las imágenes multi o hiperespectrales. En la mayoría de los casos, la distribución de los ND tiende a la distribución normal. Solamente cuando en la imagen existen zonas bien con-

trastadas y diferenciadas, la distribución puede par-tirse, aunque cada una de las zonas responde a su vez a la forma gaussiana de distribución.

Considerar ajustada a la distribución normal el conjunto de ND de la imagen tiene la ventaja de poder representar la distribución solamente con dos parámetros: su media y su varianza en el caso de imágenes monobanda y su vector de medias y la matriz de covarianzas, en el caso de multiespectrales.

La función de densidad de la distribución normal univariante es:

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} \quad (1)$$

y, de forma análoga, la de la normal multivariante es:

$$f(\vec{z}) = \frac{1}{\sqrt{|\Sigma|}(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\vec{z}-\vec{M})^T \Sigma^{-1}(\vec{z}-\vec{M})} \quad (2)$$

donde \vec{z} es el vector de características de una celda determinada, \vec{M} es el vector de medias y Σ es la matriz de covarianzas, ya utilizada en la transformación a componentes principales, explicada en el capítulo 12:

$$\vec{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}; \quad \vec{M} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}; \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}$$

Considerando solamente dos bandas, la distribución de ND adquiere la forma de una nube bidimensional, con una mayor densidad de puntos en su zona central, en torno a lo que se podría denominar su centro de gravedad y que muestra habitualmente una cierta alineación o tendencia en una determinada dirección.

Si la imagen tiene tres bandas, en condiciones ideales, la nube de puntos tendrá una forma que podría asemejarse a un elipsoide. Si, finalmente, contiene cuatro o más bandas, la nube podría considerarse un hiperelipsoide. En cualquiera de los casos, para distribuciones normales, el agrupamiento de puntos en el espacio de características queda definido por su vector de medias, que señala el centro de gravedad de la nube, y la matriz de covarianzas de la imagen, que determina su forma. El hiperelipsoide puede obtenerse haciendo constante la forma cuadrática del exponente de la función de densidad multivariante:

$$(\vec{z}-\vec{M})^T \Sigma^{-1}(\vec{z}-\vec{M}) = K \quad (3)$$

que es el cuadrado de la llamada distancia de Mahalanobis entre \vec{z} y \vec{M} en la geometría de Σ .

La orientación de los ejes está definida por los vectores propios de la matriz de covarianzas, y el grado de dispersión o alargamiento de la nube en cada dirección depende del valor propio asociado a cada vector propio.

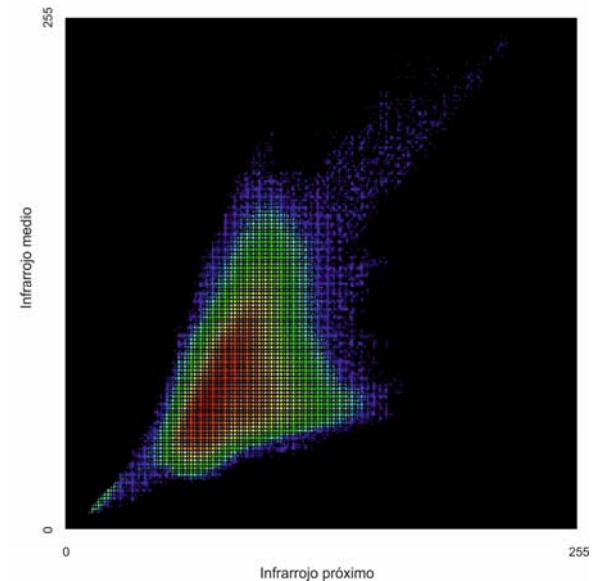
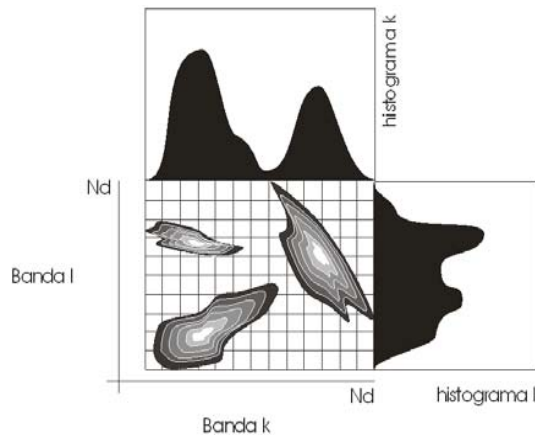


Imagen en color compuesto RGB=IRm-IRp-R y un diagrama de dispersión IRm/IRp con resolución radiométrica de 8 bits. Los colores cálidos indican mayor concentración de celdas.

Entrenamiento del clasificador

Es un hecho que grupos de celdas con un comportamiento espectral similar aparecerán en lugares próximos cuando se representan en un diagrama de dispersión. En el terreno, esas celdas tienen una misma naturaleza química, y, en definitiva, están ocupadas por la misma cobertura. En el proceso de identificar y discriminar categorías nominales en la imagen, ese conjunto de celdas constituirán una clase.



Tres agrupamientos de puntos pueden ser confundidos con dos si solamente se estudia la distribución de ND en una sola banda. El aumento de la espectralidad beneficia la identificación.

Significa esto que en el espacio espectral, las celdas se organizan como nubes de puntos próximos entre ellos y, a su vez, diferenciados de otras categorías de celdas. En una situación ideal, las nubes de puntos son independientes, alejadas unas de otras y claramente identificables y separables. En las imágenes reales, las nubes se mezclan, se solapan parcialmente y su discriminación no resulta tan fácil. Sin embargo, agrupamientos o racimos (*clusters*) de puntos que pueden ser confundidos como uno solo cuando se los estudia en un espacio espectral de tan solo unas cuantas bandas, pueden ser separables si se recurre a un mayor número de canales. La *espectralidad* de la imagen es un factor esencial a la hora de discriminar categorías.

Entendiendo *clasificador* como el algoritmo o el operador matemático de clasificación, el entrenamiento del clasificador es una tarea mediante la cual se definen los parámetros esenciales que identifican a cada clase, esto es, su vector de medias y su matriz de covarianzas.

El entrenamiento del clasificador puede ser de dos tipos: supervisado o no supervisado. En el primero, el operador delimita una muestra de celdas por cada

clase, a partir de la cual se calculan ambos estadísticos. En el entrenamiento no supervisado, a partir de unas instrucciones mínimas, se deja al sistema de tratamiento que busque agrupamientos *naturales* de valores en el espacio de características.

Entrenamiento supervisado

Un paso previo al entrenamiento supervisado es la elaboración de la leyenda de la clasificación. El analista deberá decidir qué clases buscará en la imagen, y una vez establecidas las categorías, habrá de definir en la escena los campos de entrenamiento (*training fields*), que son un conjunto de parcelas constituidas por celdas suficientemente representativas de cada clase.

De la correcta selección de los campos de entrenamiento depende estrechamente el resultado de la clasificación. En primer lugar, la pureza de los píxeles es un requisito necesario. Las celdas recolectadas en los campos de entrenamiento deben representar solamente a la cubierta que se pretende caracterizar, sin contaminación de otras. Dicho de otro modo, hay que descartar los píxeles mixtos, situados sobre dos clases diferentes o en las proximidades de bordes de parcela con las categorías que se pretenden representar.

Pero a su vez, para que la muestra sea representativa, hay que respetar y recoger en ella la variabilidad interna de la clase. El conjunto de campos de entrenamiento debe incluir la variación interna de cada clase para dar cabida a las posibles condiciones diferencia-

das de la cubierta, como su orientación, la iluminación, el tipo de suelo, su grado de humedad, amén de las diferencias inherentes a la propia cubierta.

En cuanto al número y distribución de los campos de entrenamiento, utilizando uno solo para representar cada clase se corre el riesgo de particularizar excesivamente sus características. Así, tras la clasificación podrían aparecer zonas que, aun estando ocupadas por la misma cubierta, quedarían sin atribuir a esa clase. Por ejemplo, a la hora de entrenar un cultivo como el olivar, conviene tomar varios campos de entrenamiento en los que se incluyan celdas ocupadas por distintos tipos y condiciones de cultivo: tradicionales, intensivos, en secano, en regadío, sobre suelos rojos, sobre albarizas, en ladera norte y sur, con poca y mucha pendiente, etc.

El número de celdas necesario para configurar la clase en el entrenamiento depende de la variabilidad interna de la clase. Considerando que los ND de cada clase siguieran una distribución normal, el número mínimo de celdas a tomar para estimar un ND promedio representativo es:

$$n \geq \frac{z_{\alpha/2}^2 \sigma^2}{\varepsilon^2} \quad (4)$$

donde $z_{\alpha/2}^2$ es el percentil de la distribución normal $N(0,1)$ correspondiente a la mitad del nivel de confianza exigido en la estimación, σ^2 es la varianza de la clase en la banda más desfavorable y ε es el error que se está dispuesto a admitir en la estimación del ND promedio.

Por ejemplo, para un intervalo de confianza del 95 % ($z_{0,05/2}=1,96$), una desviación típica de los ND de 15 unidades y un error máximo de 2 niveles digitales

para la estimación del ND promedio, se necesita una muestra de al menos $n \geq 1,96^2 \cdot 15^2 / 2^2 = 216$ celdas. Para un nivel de confianza del 99 %, en cambio, se necesitarían $n \geq 371$ celdas.

En conclusión, clases muy uniformes quedarán bien descritas con pocas celdas. Por el contrario, clases de reflectancia muy heterogénea necesitarán un mayor número de celdas para estar bien caracterizadas.

Las celdas de entrenamiento deben estar distribuidas en campos o polígonos de entrenamiento, siendo preferible un mayor número de ellos, aunque sean de menor tamaño, que un sólo campo de entrenamiento que contenga todas las celdas de la muestra. En cuanto a su situación, los campos de entrenamiento deben distribuirse espacialmente según la variabilidad de la escena.

Cuando no se disponga de otra información, como aproximación se pueden tomar entre 10 y 100 celdas por banda de la imagen multispectral para cada una de las clases. Así pues, para una imagen de 5 bandas, deberían recolectarse entre 50 y 500 celdas para configurar la muestra de entrenamiento de cada clase, distribuidas en varios polígonos.

Entrenamiento no supervisado

Los métodos no supervisados o automáticos buscan en el espacio de características de la imagen agrupaciones *espontáneas* de los datos, de acuerdo a alguna función de similaridad. En la literatura, esas nubes de puntos aparecen con diferentes denominaciones: agrupamientos, *clusters*, grupos o racimos, y pueden considerarse como regiones continuas del espacio de características que contienen densidades relativamente altas de puntos, y que se encuentran a su vez

separada de otras regiones de alta densidad por regiones cuya densidad de puntos es relativamente baja.

El entrenamiento no supervisado ordinariamente va unido al propio proceso de clasificación y recorre un triple proceso. En primer lugar, se determinan las características de las celdas de entrenamiento, a continuación, se calcula la similaridad o distancia estadística entre casos y, finalmente, se aplica algún criterio de agrupamiento de los casos similares o de escisión en clases diferentes.

La clasificación automática tiene el inconveniente de poder encontrar agrupamientos que realmente no tienen significado identificable en la leyenda de la clasificación. En realidad, desde un punto de vista teórico, siempre que no existan dos vectores de características exactamente iguales, es posible identificar tantas clases diferentes como celdas tenga la imagen, lo cual carece de sentido cartográfico. Por esta razón hay que informar al clasificador de numerosos detalles para que busque solamente los *clusters* que se necesitan, o un número próximo al deseado. En definitiva, el número de parámetros requeridos para que el algoritmo proporcione resultados consistentes, en numerosas veces desconocidos, hace dudar del propio apelativo de *no supervisado*.

Por el contrario, la clasificación supervisada puede adolecer del defecto contrario, esto es, disponer de una leyenda excesivamente ambiciosa que incorpora clases que espectralmente no son separables.

De todo esto se concluye que los métodos mixtos de clasificación, que incluyen a los dos anteriores, son deseables sobre cualquiera de ambos por separado, siempre que se disponga de los recursos necesarios para llevarlos a cabo. En la estrategia para combinar

los procedimientos supervisados y no supervisados pueden seguirse dos caminos: se puede realizar una clasificación inicial no supervisada cuyo resultado será posteriormente supervisado por el analista, agrupando *clusters* de igual significación temática y separando en nuevas clases las que mezclen dos categorías diferentes, o bien se pueden hacer las dos clasificaciones, supervisada y no supervisada, y posteriormente se sintetizan los resultados obtenidos. Para las clases obtenidas pueden darse tres casos:

- **Categorías que aparezcan en las dos.** La leyenda es correcta porque las clases cartográficas propuestas en el entrenamiento supervisado tienen además significación espectral.
- **Categorías recogidas en la leyenda, pero no identificables en la clasificación no supervisada.** En este caso se ha sido demasiado ambicioso en el diseño de las clases y la espectralidad de la imagen no permite su separación. Por lo tanto, deben re-estudiarse estratificando la imagen o, simplemente, simplificando la leyenda.
- **clases no supervisadas que aparecen sin estar en la leyenda.** La clasificación proporciona mayor número de clases espectrales que cartográficas y por ello el entrenamiento ha sido tímido. La información contenida en la imagen permite desagregar más la leyenda, adaptándola a la heterogeneidad de la escena.

Análisis de separabilidad

Con el fin de dimensionar adecuadamente la campaña de entrenamiento y tratar de conseguir que las clases cartográficas coincidan con las clases espectrales, es aconsejable realizar previamente un análisis de separabilidad de las clases. Esta información permitirá al analista seleccionar solamente las clases que el clasificador sea capaz de resolver como diferenciadas.

Si las clases son patentemente diferentes, el riesgo de error en la clasificación será bajo en general, pero probablemente la leyenda será también más pobre de lo que permitiría la riqueza espectral de la imagen. Por el contrario, si las clases son muy similares desde el punto de vista espectral, la clasificación entrañará un elevado riesgo de confusión que requeriría, en su caso, sustituir la leyenda por otra menos desagregada. Antes de la fase de clasificación conviene, pues, estudiar la separabilidad de las clases.

Existen dos grupos de métodos para ello:

- **Métodos gráficos**, que evalúan la separabilidad a partir de los diagramas de firmas.
- **Métodos estadísticos**, que evalúan la separabilidad de dos agrupamientos de datos midiendo la distancia entre ellos en el espacio de características.

Los primeros proporcionan una información preliminar para decidir *a priori* si las categorías serán o no discriminables. Un diagrama de firmas es una generalización de la firma espectral de las clases, sobre la que se remuestra el barrido espectral de acuerdo con la anchura de banda y la longitud de

onda central de los canales del sensor. Las gráficas incluyen el valor medio por clase en cada banda y alguna medida de dispersión alrededor de ella, como el rango de variación, la desviación estándar, etc. Para cada banda es posible estudiar más detalladamente el solapamiento estadístico de las clases superponiendo sus histogramas.

Los métodos estadísticos miden la distancia existente entre los centros de dos agrupamientos, es decir, la que hay entre sus vectores de medias. La distancia de Minkowski se define así:

$$D_p(M_A, M_B) = \left[\sum_{i=1}^n |a_i - b_i|^p \right]^{1/p} \quad (5)$$

donde $p \geq 1$, y a y b son los componentes de los vectores de medias de los agrupamientos A y B . Dando valores a p se obtienen las distintas distancias métricas, de entre las cuales la más familiar es la distancia euclídea para n bandas, que se obtiene cuando $p=2$:

$$D_2(M_A, M_B) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (6)$$

Escrito de otra forma, la distancia euclídea entre los centros de dos agrupamientos es:

$$\delta_{AB} = \sqrt{\sum_{i=1}^n (\overline{ND}_{A_i} - \overline{ND}_{B_i})^2} \quad (7)$$

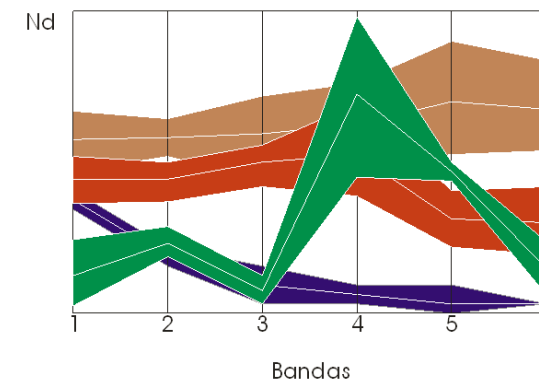


Diagrama de firma de cuatro clases, para un sensor de 7 canales.

siendo \overline{ND}_A y \overline{ND}_B los niveles digitales medios en cada banda i de los agrupamientos A y B .

Otra forma de medir la distinguibilidad de las clases es mediante la separabilidad estadística, que se define así:

$$d_{AB} = \frac{|\mu_A - \mu_B|}{\sigma_A + \sigma_B} \quad (8)$$

Lo cual es un método similar al anterior, pero proporciona valores normalizados al dividir por la suma de las varianzas de ambos agrupamientos, es decir, relativiza las distancias de acuerdo a la dispersión de los valores.

En distribuciones normales multivariantes puede adoptarse la divergencia estadística como parámetro de separabilidad:

$$D_{AB} = \frac{1}{2} Tr \left[(\Sigma_A - \Sigma_B) (\Sigma_B^{-1} - \Sigma_A^{-1}) \right] + \frac{1}{2} Tr \left[(\vec{M}_A - \vec{M}_B)^t (\Sigma_A^{-1} + \Sigma_B^{-1}) (\vec{M}_A - \vec{M}_B) \right] \quad (9)$$

Donde Tr indica la traza de las matrices resultantes.

Es también relativamente frecuente el uso de la divergencia transformada:

$$DT_{AB} = 2 \left(1 - \frac{1}{8} e^{-D_{AB}} \right) \quad (10)$$

donde D_{AB} es la divergencia estadística entre A y B y siempre $0 \leq DT_{AB} \leq 2$. El criterio de interpretación es el siguiente:

- Si $DT_{AB} \geq 1,9$, las clases A y B son separables.
- Si $1,7 < DT_{AB} < 1,9$, la separación entre A y B es aceptable.
- Si $DT_{AB} < 1,7$, las clases A y B son difícilmente separables.

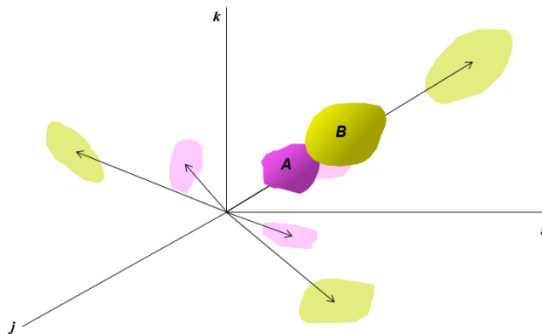
Un procedimiento muy utilizado para evaluar distancias entre clases en teledetección hiperespectral es medir el coseno del ángulo espectral, o el ángulo directamente, que da lugar a algoritmos como el mapeador de ángulo espectral (SAM). Dado que el producto escalar de los vectores de medias de dos agrupamientos A y B es:

$$\vec{M}_A^t \cdot \vec{M}_B = |\vec{M}_A| \cdot |\vec{M}_B| \cdot \cos \alpha$$

se puede escribir:

$$\alpha = \cos^{-1} \left[\frac{\vec{M}_A^t \vec{M}_B}{|\vec{M}_A| |\vec{M}_B|} \right] = \cos^{-1} \left[\frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}} \right] \quad (11)$$

aunque a primera vista pueda parecer que agrupamientos diferentes que estén alineados con el origen de coordenadas espectrales serían confundidos con un solo agrupamiento según este criterio, puesto que $\cos \alpha = 1$, esta alineación habitualmente no ocurre en todas las bandas, por lo cual siempre se encontrará un cierto ángulo espectral distinto de cero. Solamente en el caso de que el alineamiento se verifique en el conjunto de todas las bandas, es cuando habrá que considerar que la clase es única, aun mostrando distintos centros de alta densidad. Esto sucede cuando la cobertura estudiada muestra áreas mal iluminadas, que acercan las nubes de puntos hacia el



Aunque en el plano ik , los agrupamientos A y B están alineados y ello pudiera hacer pensar de que se trata de una misma categoría, el examen de sus proyecciones sobre los planos ij y jk revela que se trata de dos agrupamientos diferentes. Su ángulo espectral será distinto de cero.

origen, frente a otras correctamente iluminados. Un efecto similar produce el exceso de humedad, que siempre oscurece la clase en las bandas del espectro visible y de los distintos infrarrojos.

La distancia de Mahalanobis proporciona un método para comparar distancias dentro de la métrica generada por cada agrupamiento, derivada de la distribución y densidad de los puntos de la nube. Se expresa así:

$$d_m(\vec{z}, \vec{M}) = \sqrt{(\vec{z} - \vec{M})^t \Sigma^{-1} (\vec{z} - \vec{M})} \quad (12)$$

y como puede observarse es realmente la raíz de la forma cuadrática de la distribución normal multivariante. En una nube hiperelipsoidal, las superficies de isodistancia no son hiperesferas sino hiperelipsoides, que quedan mejor caracterizados con la distancia de Mahalanobis.

Fase de asignación

Durante la fase de asignación se asigna una etiqueta de clase a cada celda, es decir, se atribuye cada celda de la imagen a una de las categorías de la leyenda previamente elaborada.

Para ello es necesario establecer una función de similaridad que permita asegurar que las celdas que pertenecen a un agrupamiento son más parecidas entre sí que a cualesquiera otras que pertenezcan a otro diferente. Las funciones de similaridad más habituales son las siguientes:

Finalmente, otra distancia que suele emplearse con fines de separabilidad entre clases es la de Bhattacharyya:

$$D_B = \frac{1}{8} (\vec{M}_2 - \vec{M}_1)^t \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (\vec{M}_2 - \vec{M}_1) + \frac{1}{2} \ln \frac{|\Sigma_1 + \Sigma_2|}{\sqrt{|\Sigma_1| |\Sigma_2|}} \quad (13)$$

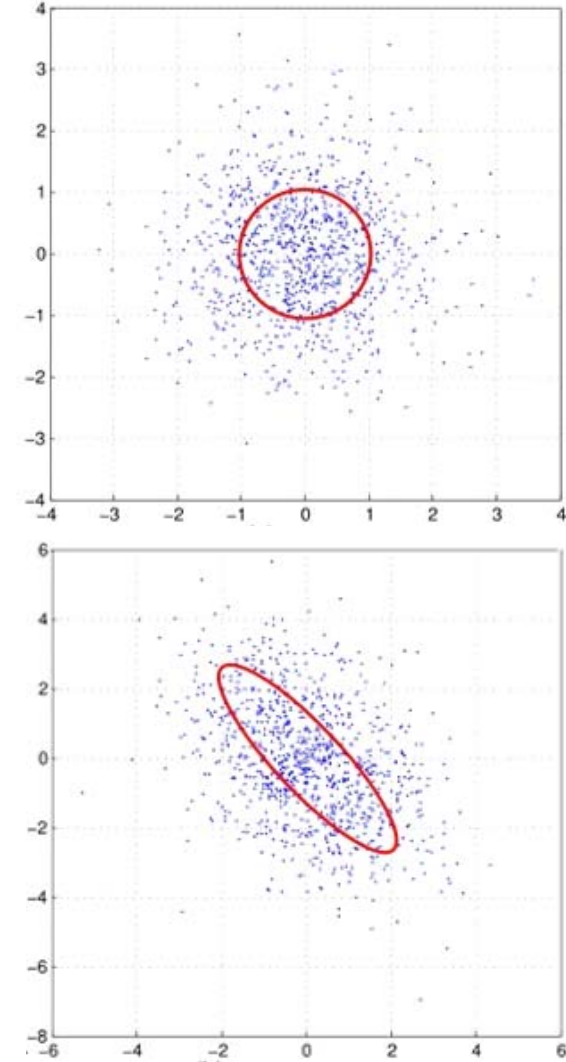
Se considera un estadístico más fiable que la distancia de Mahalanobis, dado que esta última es un caso particular de la de Bhattacharyya cuando las desviaciones estándar de las dos clases son iguales. Cuando dos clases tienen valores promedio muy similares pero diferentes desviaciones estándar, la distancia de Mahalanobis tiende a cero, mientras que la distancia Bhattacharyya crece a medida que lo hace la diferencia entre las desviaciones estándar.

Distancia mínima: $d_{\min}(Z_i, Z_j) = \min_{\substack{z \in Z_i \\ z' \in Z_j}} \|\vec{z} - \vec{z}'\| \quad (14)$

Distancia máxima: $d_{\max}(Z_i, Z_j) = \max_{\substack{z \in Z_i \\ z' \in Z_j}} \|\vec{z} - \vec{z}'\| \quad (15)$

Distancia media: $d_{\text{med}}(Z_i, Z_j) = \|\vec{M}_i - \vec{M}_j\| \quad (16)$

Y en todas ellas, la norma $\|X\|$ utilizada es la euclídea:



Uso de la distancia euclídea (arriba) y de Mahalanobis (abajo). Las líneas rojas representan los puntos de igual distancia al centro del agrupamiento.

$$\|X\| = \sqrt{\sum_{i=1}^n |X_i|^2}$$

Clasificadores paramétricos

Los clasificadores son los operadores encargados de etiquetar los vectores de características \vec{z} con alguna de las clases w_i comparando un conjunto de c funciones discriminantes $g_i(\vec{z})$, donde $i=1, 2, \dots, c$:

$$\vec{z} \in w_i \quad \text{si} \quad g_i(\vec{z}) > g_j(\vec{z}) \quad \forall j \neq i \quad (17)$$

El clasificador divide el espacio de características en c clases diferentes o regiones de decisión w_1, w_2, \dots, w_c y asigna la etiqueta de clase w_i a un determinado vector de características \vec{z} cuando la función discriminante que le corresponde para esa clase supera en valor a cualquiera del resto de las clases.

Entre los clasificadores más empleados se encuentran el de mínima distancia, el paralelepípedo, el isodato, el de k-Means, y el de máxima probabilidad.

- **Clasificador de mínima distancia.** El criterio más sencillo para asignar una celda a una categoría consiste en incluirlo en la espectralmente más cercana, dentro del espacio de características. La función de similitud utilizada es la distancia euclídea entre el vector de características de la celda y el centro de la clase, representado por su vector de medias:

$$d_c = \sqrt{\sum_{i=1}^n (ND_i - \overline{ND}_{C_i})^2} \quad (18)$$

La celda se asignará a la clase que presente una

menor distancia d_c .

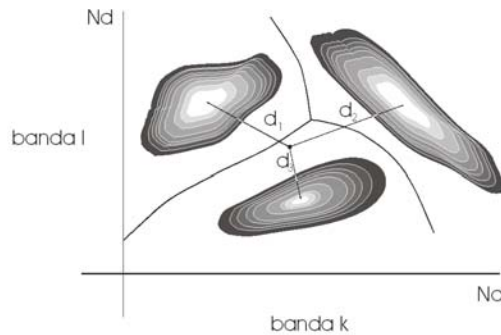
- **Clasificador paralelepípedo.** La clasificación paralelepípedica o de hipercubos fija áreas de dominio para cada categoría. Una celda de nivel digital ND^k en la banda k de una imagen de n bandas será atribuida a la clase C si se verifica:

$$\overline{ND}_C^{(k)} - R_C^{(k)} < ND^k < \overline{ND}_C^{(k)} + R_C^{(k)} \quad \forall 1 \leq k \leq n \quad (19)$$

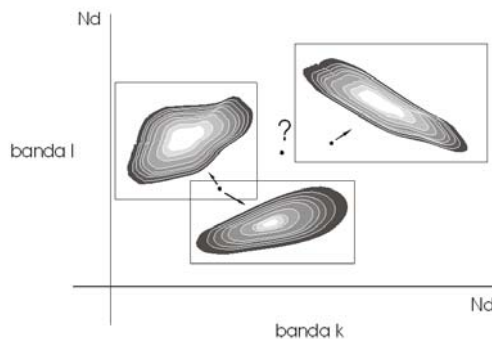
siendo $R_C^{(k)}$ un rango de variación concedido alrededor del valor medio de la clase C en la banda k . Cuanto mayor sea el rango, mayor será el número de celdas incluidas en la categoría, pero se incrementará el riesgo de error de comisión de la clasificación. Por el contrario, si el umbral es excesivamente pequeño, ese tipo de error se minimizará, pero también aumentarán las celdas sin clasificar, incrementándose así el error de omisión.

- **Clasificador isodato.** Es un algoritmo de clasificación automática iterativo basado en la función de mínima distancia euclídea. En cada ciclo de clasificación se asignan las celdas a una clase provisional. Durante el desarrollo del algoritmo se registran fugas de celdas de una clase a otra hasta que el proceso converge. El algoritmo de clasificación isodato sigue los siguientes pasos:

1. Se eligen los valores iniciales $\vec{m}_1, \vec{m}_2, \dots, \vec{m}_c$ que actuarán como semillas de los *clusters*. Si no hay valores previos, se toma el vector de características de la primera celda de la imagen.
2. Calcular la distancia entre la semilla y la segunda celda.



Criterio de asignación por mínima distancia.



Criterio de asignación por paralelepípedos, aplicado a un espacio espectral de sólo dos dimensiones.

- a. Si es inferior a un radio espectral r prefijado, se incluye en el mismo *cluster*.
 - b. En caso contrario se considera la segunda celda como segunda semilla, y así sucesivamente hasta c clases.
3. Se clasifica la imagen completa, asignando las celdas al agrupamiento más próximo.
 4. Se recalculan las medias de los agrupamientos considerando las nuevas asignaciones, pasando a tomar los valores $\bar{m}'_1, \bar{m}'_2, \dots, \bar{m}'_c$.
 5. Se reitera el ciclo de reclasificación hasta mejorar la convergencia de los valores, o hasta alcanzar un número prefijado de iteraciones.
 6. Se descarta cualquier agrupamiento que contenga menos datos que un determinado valor también prefijado.
 7. Se reagrupan o se escinden los agrupamientos, de acuerdo con las siguientes reglas:
 - a. Agrupar *clusters* si su número excediese de un valor prefijado.
 - b. Dividir los *clusters* si su número no alcanza un valor umbral.
 8. Se toma el centro de masas cada *cluster* como una nueva semilla y se reitera la clasificación como en la etapa 2.
 9. Se repiten los pasos 3, 4, 5, 6 y 7 hasta que el proceso converja o hasta alcanzar un número de iteraciones establecido con anterioridad.
- **Clasificador k-means.** Este algoritmo de clasificación automática opera de un modo parecido al anterior, pero si en aquél se establecían las regiones de decisión como círculos en un plano euclídeo espectral R^2 , en éste se disponen como polígonos de Thiessen. El algoritmo sigue los siguientes pasos:
1. Se especifica un número inicial de clases y se generan las clases por distancia euclídea. Se

toma una celda de cada clase como centroide del agrupamiento. Se definen los segmentos que unen todos los centroides entre sí y se trazan sus mediatrices, con las que se generarán tantos polígonos de Thiessen como clases se han establecido.

2. Se configuran las nuevas clases con las celdas que pertenecen a cada polígono. Es posible que celdas que anteriormente pertenecían a una categoría, tras la teselación del espacio espectral, ahora pertenezcan a otra.
3. Se calcula el centro de masas de las nuevas nubes de puntos, es decir, sus vectores de medias, y se erigen como nuevas semillas de los agrupamientos para la segunda iteración.
4. Se reiteran los pasos 2,3 y 4 hasta que el proceso sea convergente y no se registren más fugas de celdas entre categorías, o su número sea inferior a uno establecido con anterioridad.

Naturalmente, tanto este clasificador como el isodata, trabajarán con la dimensionalidad completa de la imagen en un espacio generalizado, y distribuirán las regiones de decisión en hiperplanos. R^n , siendo n en el número de bandas.

- **Clasificador de máxima probabilidad.** Este clasificador es de los más robustos y utilizados dentro de los procedimientos paramétricos de clasificación digital. Consiste en etiquetar las celdas con la categoría a la que más probabilidad tenga de pertenecer por su vector de características, asumiendo distribuciones normales para las clases. Para ello se establece la probabilidad de pertenencia a cada clase, utilizando la función de densidad de probabilidad enunciada en (2).

Clasificador de máxima probabilidad bayesiana. La aplicación del teorema de Bayes a la clasifica-

ción de máxima probabilidad mejora los resultados finales. La probabilidad de un determinado suceso B condicionada a otro suceso A es la probabilidad de que ocurra B , supuesto que también se verifica la ocurrencia de A :
$$P(B/A) = \frac{P(A \cap B)}{P(A)} \quad (20)$$

De igual forma:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad (21)$$

de modo que se puede escribir:

$$P(A \cap B) = P(A)P(B/A) = P(B)P(A/B) \quad (22)$$

Dos sucesos se consideran independientes cuando:

$$P(A \cap B) = P(A)P(B) \quad (23)$$

Lo cual quiere decir que $P(A/B) = P(A)$ y que $P(B/A) = P(B)$. De lo contrario, son dependientes entre sí.

Pues bien, si en un contexto de decisión se consideran n sucesos mutuamente excluyentes A_i que formen un sistema exhaustivo, es decir que la unión de todos ellos dé lugar al suceso seguro o universo muestral E :

$$\bigcup_{i=1}^n A_i = E \quad (24)$$

y de todos ellos se conocen las probabilidades $P(A_i)$ de que sucedan, la probabilidad del suceso B , del que se conocen todas las probabilidades condicionadas $P(B/A_i)$, será:

$$P(B) = P(E \cap B) = P\left[\left(\bigcup_{i=1}^n A_i\right) \cap B\right] = \sum_{i=1}^n P(A_i \cap B) \quad (25)$$

o sea:

$$P(B) = \sum_{i=1}^n P(A_i)P(B/A_i) \quad (26)$$

La expresión (22), aplicada a este caso, se puede poner:

$$P(A_i \cap B) = P(A_i)P(B/A_i) = P(B)P(A_i/B) \quad (27)$$

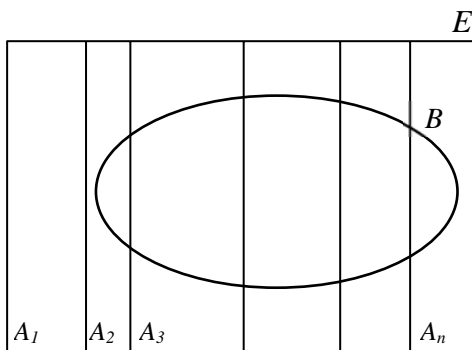
o sea:

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(B)} \quad (28)$$

Sustituyendo el denominador por su valor (26):

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^n P(A_i)P(B/A_i)} \quad (29)$$

que es la expresión del teorema de Bayes. En esta expresión, $P(A_i/B)$ es la probabilidad *a posteriori*, $P(A_i)$ es la probabilidad *a priori* y



Universo muestral E teselado por n sucesos A_i y, sobre ellos, el suceso B . Las probabilidades condicionadas $P(B/A_i)$ son las proporciones de elipse que interceptan cada rectángulo A_i en relación con la superficie de este último.

$P(B/A_i)$ se denomina *verosimilitud*.

En el contexto de decisión dentro de la clasificación digital de imágenes en teledetección, la variable aleatoria es el vector de características de una celda \vec{z} , y el teorema de Bayes puede escribirse así:

$$P(w_i/\vec{z}) = \frac{P(w_i)P(\vec{z}/w_i)}{\sum_{j=1}^s P(w_j)P(\vec{z}/w_j)} \quad (30)$$

Si w_1 y w_2 son dos posibles estados para una celda, es decir, dos clases a las que hay probabilidad de que pertenezcan, se asume la ocurrencia del suceso w_1 cuando:

$$P(w_1/\vec{z}) > P(w_2/\vec{z}) \quad (31)$$

Utilizando el clasificador bayesiano, dado que los denominadores de la expresión (30) son siempre los mismos, sea cual sea la clase considerada, la expresión (31) conduce a la siguiente operación de decisión:

$$\text{decidir } \begin{cases} w_1 & \text{si } P(w_1)P(\vec{z}/w_1) \geq P(w_2)P(\vec{z}/w_2) \\ w_2 & \text{si } P(w_1)P(\vec{z}/w_1) < P(w_2)P(\vec{z}/w_2) \end{cases} \quad (31)$$

en cuya expresión se han utilizado solamente los numeradores de la expresión de Bayes como función de decisión.

De modo general, la función discriminante en la decisión bayesiana es la probabilidad *a posteriori*:

$$g_i(\vec{z}) = P(w_i/\vec{z}) \quad (32)$$

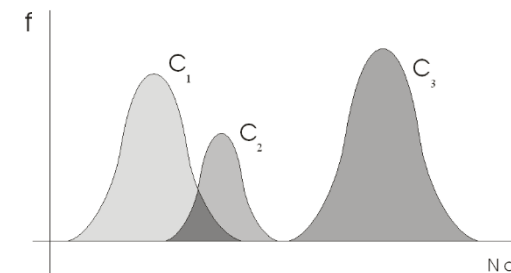
y la celda será atribuida a la clase para la que la probabilidad a posteriori sea más alta. Aplicando el teorema de Bayes:

$$g_i(\vec{z}) = P(\vec{z}/w_i)P(w_i) \quad (33)$$

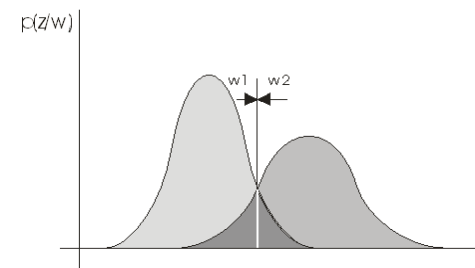
De modo que la celda pertenece a la clase w_i si $g_i(\vec{z}) > g_j(\vec{z})$, o lo que es lo mismo:

$$\begin{aligned} &\text{decidir } w_i \text{ si } P(w_i/\vec{z}) > P(w_j/\vec{z}) \\ &\text{es decir, si } P(w_i)P(\vec{z}/w_i) > P(w_j)P(\vec{z}/w_j) \end{aligned} \quad (32)$$

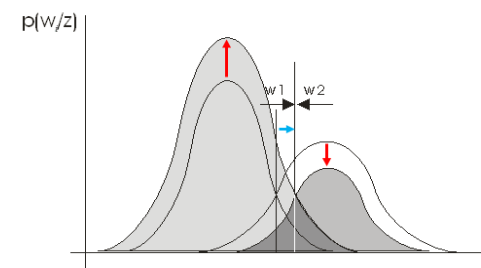
La función de decisión es, pues, la probabilidad *a posteriori* $P(w_i/\vec{z})$, pero no es fácil decidir a qué clase puede pertenecer una celda condicionada a que se verifiquen unos determinados niveles digitales, o lo que es lo mismo, a la vista de su vector de características. En cambio, resulta más fácil plantear el problema de la decisión en términos bayesianos. Por un lado, se conoce, o se puede conocer, la probabilidad *a priori* de cada clase $P(w_i)$, es decir, su abundancia dentro de la imagen que se va a clasificar o, lo que es lo mismo, la proporción que ocupa cada clase dentro de la escena, sea cual fuere la ubicación de las celdas que le corresponden. Esta información puede obtenerse a partir de clasificaciones anteriores, u otras fuentes auxiliares. En el caso de no contar con dicha información, se puede adoptar provisionalmente una probabilidad igual a todas las categorías para iniciar la clasificación.



Un clasificador de máxima verosimilitud clasificará las categorías C_1 , C_2 y C_3 como diferentes, aunque con errores de asignación entre la clase C_1 y C_2 , debido al solapamiento de sus distribuciones.



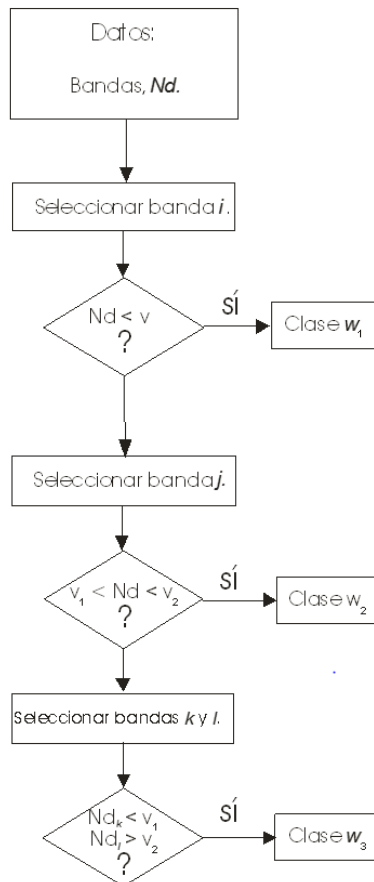
máxima verosimilitud



máxima probabilidad

El clasificador de máxima probabilidad bayesiano...

El otro factor pendiente de conocer para aplicar el teorema de Bayes es la verosimilitud $P(\bar{z}/w_i)$, que en este contexto es la probabilidad de que se cumpla un determinado nivel digital, o un vector de características si se habla de una imagen multiespectral, condicionado a cada clase de la leyenda. Si se asume que los niveles digitales de las clases siguen distribuciones normales multivariantes, la verosimilitud es simplemente la probabilidad leída en el eje de ordenadas de la gráfica de la distribución normal, o en las tablas estadísticas habituales.



Árbol de clasificación.

Otros criterios de asignación

Hay otros clasificadores que no se atienen a los parámetros de la distribución normal. Entre ellos pueden citarse los siguientes.

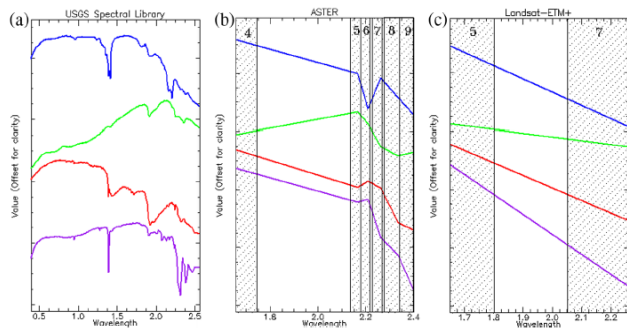
- **Clasificador en árbol.** Estos operadores identifican las categorías secuencialmente y en cada paso utilizan las bandas más apropiadas para la discriminación. Para configurar una determinada clase se seleccionan las celdas cuyo ND esté comprendido dentro de un cierto intervalo, o bien aquellas que cumplan una determinada condición, y se continúa el proceso sólo con ellas. Cuando, en una determinada banda, una celda exhiba un ND con el que pueda ser asignada a más de una categoría, el clasificador establecerá nuevas condiciones que involucren a varias bandas simultáneamente.

Estos clasificadores suelen estar constituidos por centenares de reglas y proporcionan resultados muy satisfactorios, siempre la base de conocimiento sea suficiente. En ellos se basan los sistemas expertos que, además, aprenden de los resultados previos para sucesivas clasificaciones.

- **Clasificadores texturales.** En análisis de imagen, la textura se refiere a las diferencias de reflectancia entre celdas adyacentes. Estos clasificadores utilizan determinadas características en la vecindad de un pixel, como la reflectancia media, la varianza, la autocorrelación, etc., con el fin de incorporar a la clasificación algunas de las propiedades cualitativas que suelen ser tenidas en cuenta en fotointerpretación. Puede realizarse un análisis de la textura de una escena aplicando determinados filtros e incorporando esta información como una fuente más de discriminación.

- **Clasificadores de contexto.** Todos los clasificadores enunciados hasta ahora clasifican la imagen celda a celda utilizando cualquiera de los criterios explicados. Sin embargo, en las imágenes reales, lo más habitual es que las clases se presenten en la escena en grupos de celdas más o menos grandes, y con mucha menor frecuencia como celdas aisladas. Los clasificadores de contexto clasifican las celdas por grupos, previamente reunidos por similitud de los ND, etiquetando cada una de ellas en relación con la asignación de las adyacentes.

- **Clasificación hiperespectral.** Además de las técnicas clásicas de clasificación paramétrica, la teledetección hiperespectral utiliza procedimientos específicos que aseguran una mayor discernibilidad entre categorías, pudiéndose con ello separar eficazmente docenas de clases. En ocasiones no se requiere clasificar completamente la escena sino solamente identificar la presencia de ciertas clases a lo largo de ella. Para ello se realizan cocientes y otras operaciones algebraicas con las bandas, basados en rasgos espectrales únicos y específicos de los componentes buscados.



Sucesiva degradación de la signature espectral: la obtenida mediante un espectrorradiómetro (a), la obtenida por el sensor Aster (b) y la de Landsat-ETM+ (c) de cuatro minerales de la arcilla: montmorillonita (azul) nontronita (verde), vermiculita (rojo) y talco (magenta). Fatima, K. *et al.* 2017. Minerals identification and mapping using satellite image. *Journal of Applied Remote Sensing*. SPIED.

De modo general, siempre es preferible un mayor número de bandas para establecer un mayor número de clases, tanto si se trabaja con operadores paramétricos como si se emplean otros procedimientos. La signature espectral queda mejor establecida realizando barridos espectrales minuciosos que con los diagramas de signature con los que trabajan los operadores multispectrales clásicos.

Control de calidad temático

La imagen clasificada no tiene una fiabilidad absoluta. Existen errores de asignación por similitud espectral entre celdas que, perteneciendo a categorías cartográficas diferentes, presentan signature espectrales muy parecidas.

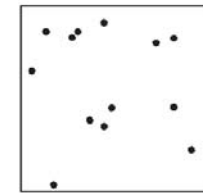
Aunque los errores de asignación se refieren siempre a un mal etiquetado de las celdas dentro de una categoría de clase determinada, en el control de calidad estas incorrecciones pueden estudiarse desde dos puntos de vista, que dan lugar a dos tipos de error temático en el proceso de clasificación:

→ **Error de omisión.** Se produce cuando, perteneciendo la celda a una determinada categoría, no

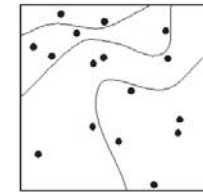
ha sido asignada a ella. Es el denominado en estadística error tipo I.

→ **Error de comisión.** Se produce cuando la celda es etiquetada dentro de una categoría a la que realmente no pertenece. Es el error tipo II.

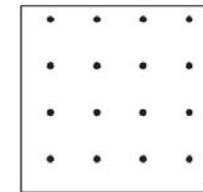
Con el fin de ofrecer al usuario la información completa del proceso de clasificación, es necesario realizar un control de calidad que establezca la fiabilidad del proceso. Para ello, tras la clasificación, se debe realizar un proceso de verificación. Si se dispone de una fuente de verdad-terreno exhaustiva, por ejemplo, procedente de fuentes cartográficas o de otras clasificaciones anteriores, la verificación puede abarcar la imagen completa. Sin embargo, lo más frecuente es no disponer de esa información, por lo cual



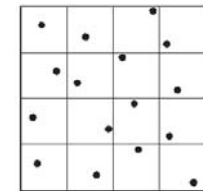
Aleatorio simple



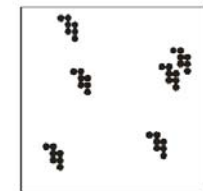
Aleatorio estratificado



Sistemático



Sistemático no alineado



Por conglomerados

Tipos de muestreo de verificación.

es necesario plantear una campaña de muestreo para la verificación.

Existen varios tipos de muestreo de verificación. Entre los más difundidas están los siguientes.

- **Aleatorio simple:** las celdas a verificar se designan al azar.
- **Aleatorio estratificado:** se divide el espacio muestral en diversas subpoblaciones o estratos sobre cada una de los cuales se aplica un muestreo aleatorio.
- **Sistemático:** a partir de un punto designado al azar, se confecciona la muestra mediante la selección de puntos equidistantes en las dos dimensiones de la imagen.
- **Sistemático no alineado:** se traza una cuadrícula sistemática, a partir de cuyos nudos se varían al azar las coordenadas de los elementos de la muestra.
- **Por conglomerados:** es un muestreo aleatorio en el que en lugar de una sola celda se selecciona un grupo de ellas para formar la muestra, siguiendo siempre el mismo patrón geométrico.

Lo que la verificación de la exactitud de la clasificación trata de establecer es el nivel de error en el etiquetado de las celdas. Para ello, juega un papel fundamental el tamaño de la muestra. El número de celdas que deben verificarse depende de dos factores. En primer lugar, del nivel de confianza que quiera otorgarse a la estimación. Es obvio que para tener el 100 % de confianza en la cifra de error debería verificarse el conjunto completo de celdas de la imagen. En segundo lugar, de la información previa que se posea sobre la población. El conocimiento previo de la dispersión de ND dentro de cada clase, es decir, su varianza, permite reducir el número de muestras sin disminuir el nivel de confianza de la estimación.

El suceso {celda bien clasificada-celda erróneamente clasificada} se distribuye como una variable aleatoria discreta binomial:

$$X \in b(n, p) \quad (33)$$

donde n es el tamaño de la muestra y p es la probabilidad de acierto en la clasificación. La función de densidad de probabilidad de la distribución binomial es:

$$f(x) = P(X=x) = \binom{n}{x} p^x q^{n-x} \quad (34)$$

con $x \in [0, 1, \dots, n]$ y $q=1-p$, es decir, la probabilidad de error en la clasificación.

Según el teorema de Moivre, cuando $n \rightarrow \infty$, la distribución binomial es asintóticamente normal $N(np, npq)$, es decir, converge en distribución a una normal de $\mu=np$ y $\sigma^2=npq$. En la práctica, el criterio de convergencia puede adoptarse cuando $n \geq 30$.

Con todo ello, el problema de determinar el tamaño muestral se plantea así: se trata de calcular el tamaño n de la muestra para estimar la probabilidad p de que una categoría esté bien clasificada, con un error de $\pm \varepsilon$ y un nivel de confianza en la estimación de $1-\alpha$, esto es, que el error tipo I (rechazar la hipótesis H_0 de asignación acertada, siendo ésta verdadera) sea inferior a α . Se trata, pues, de un problema de estimación por intervalo de confianza, donde éste es:

$$I_{1-\alpha} = \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{pq}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{pq}{n}} \right] \quad (35)$$

siendo $z_{\alpha/2}$ el percentil de la distribución normal $N(0,1)$ correspondiente a la mitad del nivel de confianza exigido en la estimación.

El error admitido en la asignación de etiquetas a la celda será:

$$\varepsilon = z_{\alpha/2} \sqrt{\frac{pq}{n}} \tag{36}$$

De donde puede deducirse el mínimo número de celdas que deben integrar la muestra de verificación:

$$n = \frac{z_{\alpha/2}^2 pq}{\varepsilon^2} \tag{37}$$

Si no se dispone de información más precisa sobre la población, en el peor de los casos puede adoptarse $p = q = 0,5$, de tal modo que pueda disponerse de un límite inferior para el tamaño muestral:

$$n \geq \frac{z_{\alpha/2}^2}{4\varepsilon^2} \tag{38}$$

Este valor es válido para determinar el número de celdas a muestrear para cada categoría. El número total de celdas de verificación será el producto de este valor por el número de clases de la leyenda. Si, por el contrario, se conoce *a priori* el intervalo de existencia de la probabilidad p de resultar la celda bien clasificada, se puede reducir el tamaño de la muestra, en la medida que se reduce el producto pq .

Es bastante frecuente verificar los resultados a partir de un muestreo sistemático no alineado o, en su defecto, uno aleatorio simple. Para cualquiera de estos casos, si se requiere un nivel de confianza del 99 %,

el percentil buscado es $1-\alpha = 0,99$; $\alpha = 0,01$ y $\alpha/2 = 0,005$. El valor de z que proporciona ese nivel de confianza es:

$$z_{\alpha/2} = 2,575; \quad z_{\alpha/2}^2 = 6,630625$$

Admitiendo un error máximo del $\pm 5\%$ en la asignación de la celda, el tamaño de la muestra de verificación deberá ser como mínimo:

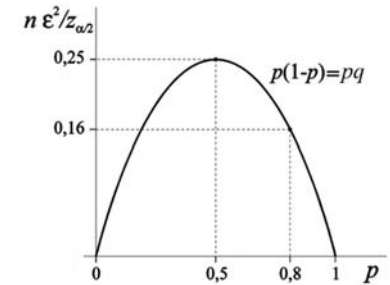
$$n = \frac{z_{\alpha/2}^2}{4\varepsilon^2}; \quad n = \frac{6,630625}{4 \cdot 0,05^2} = \frac{6,630625}{0,01} \approx 663$$

Matriz de confusión

Los resultados de la verificación suelen presentarse en una tabla de contingencia cuadrada en la que las filas recogen las categorías obtenidas en el proceso de clasificación y las columnas la verdad-terreno, es decir, las categorías verificadas.

Cada elemento de la matriz representa el número de celdas que, perteneciendo en la clasificación a la categoría que marca su fila, realmente la verificación ha demostrado su pertenencia a la clase que indica su columna. La diagonal principal de la matriz representa el número de celdas correctamente clasificadas para cada categoría de la leyenda y los elementos fuera de la diagonal principal indican errores de asignación, que, como ha quedado establecido, son de dos tipos:

- Los situados dentro de una misma columna representan el error de omisión.
- Los pertenecientes a una misma fila representan el error de comisión.



El peor escenario para determinar el mínimo número de celdas de la muestra de verificación es en el que no se conoce la probabilidad de ser bien clasificada la categoría. En ese caso, $p = q = 0,5$ y n es máximo.

		CATEGORÍAS VERIFICADAS										SUMA	Error absoluto de comisión	Riesgo del usuario (%)		
		1	2	3	4	5	6	7	8	9	10					
CATEGORÍAS EN LA IMAGEN CLASIFICADA	Agua 1	3										3	0	0		
	Suelos desnudos 2		2								1	3	1	33		
	Coníferas 3			27					3	2		32	5	16		
	Frondosas 4				1	7	1	2				11	4	36		
	Herbáceos extensivos 5						86		1	4		91	5	5		
	Regadío 6					3		19				22	3	15		
	Pradera 7							1	9	1		11	2	18		
	Monte bajo 8								1	2	31	36	5	24		
	Urbano cerrado 9											9	0	0		
	Urbano abierto 10											1	2	3	1	33
	SUMA		3	4	28	10	89	21	15	39	10	2	221	26		
Error absoluto de omisión		0	2	1	3	3	2	6	8	1	0	26				
Riesgo del productor (%)		0	50	4	30	3	10	40	21	10	0					

Un ejemplo de matriz de confusión para 10 categorías.

El error de omisión, también llamado riesgo del productor, se calcula como la proporción de los residuos por columnas en relación con el total:

$$EO_j = \frac{\sum_i x_{ij} - x_{jj}}{\sum_i x_{ij}} \quad i \neq j \quad (39)$$

Por su parte, el error de comisión, o riesgo del usuario, se calcula como la proporción de los residuos por

filas en relación con el total:

$$EC_i = \frac{\sum_j x_{ij} - x_{ii}}{\sum_j x_{ij}} \quad i \neq j \quad (40)$$

Los complementarios de los errores de omisión y comisión se llaman fiabilidades del productor y del usuario, respectivamente:

$$FP_j = \frac{x_{jj}}{\sum_i x_{ij}} \quad (41)$$

$$FU_i = \frac{x_{ii}}{\sum_j x_{ij}} \quad (42)$$

Cuando la clasificación solamente tiene dos categorías, resulta evidente comprobar que el error de omisión de una de ellas es justamente el de comisión de la otra. Sin embargo, cuando el número de clases es mayor, los errores se fugan entre las categorías y solamente se puede afirmar que la suma de los errores de omisión es igual a la suma de los de comisión, lo cual puede comprobarse en los ejemplos al margen.

Dentro del ámbito del análisis categórico multivariante, uno de los índices más empleados en la evaluación de la calidad de la clasificación es el porcentaje de acuerdo, que se refiere a la concordancia observada, esto es, la suma de los elementos situados en la diagonal de la matriz de confusión, que son los elementos correctamente identificados en la clasificación, en relación con el total de la muestra, expresada en tanto por ciento:

$$P_a = \frac{\sum_i x_{ii}}{n} \cdot 100 \quad (43)$$

Y otro índice muy empleado en la evaluación de la calidad de la clasificación es el estadístico *Kappa* κ de Cohen, que cuantifica el nivel de acierto atribuible al método de clasificación seguido, por encima del que se hubiese obtenido meramente por azar:

$$\kappa = \frac{P_a - P_e}{1 - P_e} \quad (44)$$

donde P_a es la concordancia observada (porcentaje de acuerdo) y P_e la concordancia esperada por azar.

Y teniendo en cuenta (100), y que:

$$P_e = \frac{\sum_i \left[\sum_j x_{ij} \sum_j x_{ij} \right]}{n^2} \cdot 100 \quad i \neq j \quad (45)$$

finalmente se tiene:

$$\kappa = \frac{n \sum_i x_{ii} - \sum_i \left[\sum_j x_{ij} \sum_j x_{ij} \right]}{n^2 - \sum_i \left[\sum_j x_{ij} \sum_j x_{ij} \right]} \quad (46)$$

El índice κ puede adoptar valores comprendidos entre 0 y 1. La interpretación de los resultados puede seguir el criterio recogido en la Tabla 1:

Tabla 1. Significado de los valores de Kappa.

Valor de κ	Concordancia
0,00	Pobre
0,01-0,20	Leve
0,21-0,40	Aceptable
0,41-0,60	Moderada
0,61-0,80	Considerable
0,81-1,00	Casi perfecta

		Verdad-terreno											
		1	2	3	4	5	6	7	8	9	10	E.com	
Clasificación	1	3										3	0
	2		2						1			3	1
	3			27				3	2			32	5
	4			1	7	1	2					11	4
	5					86		1	4			91	5
	6				3		19					22	3
	7					1		9	1			11	2
	8		2				1	2	31			36	8
	9									9		9	0
	10										1	2	3
E.Om		3	4	28	10	89	21	15	39	10	2	221	26

Forma de calcular el estadístico κ .

$Com(A)=Om(B)$

		verdad-terreno			
		A	B	Σ	
clasificación	A	2	3	5	3
	B	1	5	6	1
Σ		3	8	11	
omisión		1	3		

$\Sigma Com = \Sigma Om$

		verdad-terreno				
		A	B	C	Σ	comisión
clasificación	A	2	3	2	7	5
	B	1	5	1	7	2
	C	1	3	7	11	4
		4	11	10	25	11
omisión		2	6	3		11

Cuando el número de clases excede de dos, solo se puede asegurar que la suma de errores de omisión es igual a la de los de comisión

14

Teledetección multitemporal

Introducción

Una de las características más interesantes de la teledetección es la posibilidad que proporciona el hecho de disponer de información periódica de un mismo lugar a lo largo del tiempo a un coste razonable. Esta propiedad facilita significativamente la realización de estudios multitemporales sobre las zonas de interés, con el fin de poner de relieve la evolución de ciertos rasgos en el territorio. De hecho, el análisis de cambios se ha convertido en una pieza fundamental dentro de las aplicaciones de la teledetección.

La identificación de los cambios entre dos imágenes de fechas diferentes ha sido abordada desde hace décadas mediante el análisis visual. La fotointerpretación tenía en ello uno de sus objetivos primordiales. El intérprete analizaba las formas, los colores, las texturas, los matices, la sombras y cuantos atributos visuales percibía y comparaba las dos imágenes para, finalmente, establecer los elementos que habían aparecido, desaparecido o, simplemente, cambiado en la escena.

El análisis digital de imagen ha venido a mejorar y profundizar en la identificación de cambios que incluso al operador podían escapársele. Sin embargo, aún hoy, carece del poder de síntesis que la mente humana desarrolla. Pequeñas diferencias entre las formas presentes en las imágenes o sutiles cambios de tonalidad pueden ser interpretados automáticamente como cambios, cuando un analista, en cambio, puede pasarlos por alto y considerar que no constituyen cambio en sí mismos. La comparación automática de imágenes no es inteligente hoy por hoy, por lo cual, las fuentes a comparar y necesitan cumplir unos

requisitos muy estrictos con carácter previo.

Estos requisitos concretan en la invariabilidad de los elementos de imagen que realmente no hayan sufrido alteración con el paso del tiempo. Y dicha invariabilidad debe ser satisfecha tanto desde el punto de vista geométrico como radiométrico.

Como la comparación digital suele estar orientada básicamente a la celda de imagen y, por lo general, el análisis automatizado no tiene en consideración factores tales como el contexto, las formas, etc., en el análisis, pequeñas desviaciones de la posición de un mismo píxel serán interpretadas como un cambio. El ejemplo más paradigmático es el de una imagen suficientemente abigarrada que no ha cambiado en absoluto, pero que en una segunda fecha aparece desplazada una sola celda en cualquier dirección con respecto a la original. El intérprete humano concluirá rápidamente que durante el plazo estudiado la escena no ha sufrido alteraciones, mientras que el análisis digital advertirá de numerosos cambios. Por esta razón es de vital importancia la correcta y absoluta *superponibilidad* de las imágenes comparadas, lo cual requiere simplemente una eficaz corrección geométrica, incluida la ortorrectificación cuando sea necesaria.

De otra parte, se impone también un requisito adicional, a saber, que la radiometría de las imágenes comparadas sea equivalente, lo cual es necesario porque, de no serlo, cualquier desviación de los valores de luminancia, por ligera que fuera, podría ser interpretada como cambio.

Para identificar cambios temporales entre dos fechas con el objeto de actualizar una cartografía, por ejemplo, o estudiar la evolución de un determinado fenómeno, es necesario que los cambios reales en la radiancia medida de cada celda sean mayores que los debidos a factores externos como las diferencias atmosféricas o el ángulo solar. El ejemplo extremo sería el de dos imágenes de un territorio que no ha cambiado en absoluto, pero que en una segunda fecha ha resultado afectado por una ligera calima originada por plumas de polvo en suspensión. Nuevamente el sistema de tratamiento digital puede informar de un cambio en toda la escena, aunque el intérprete dirá que la segunda imagen es igual que la primera, sólo que perturbada por el efecto atmosférico. En cualquier caso, para neutralizar las consecuencias de la variación de la irradiancia entre las fechas, lo que repercute en los valores de radiancia registrados, es preferible siempre realizar todos los análisis sobre reflectancia.

Por todo lo expuesto, es obligada la corrección radiométrica rigurosa basada en modelos como LOWTRAN, MODTRAN, FLAASH, etc. Para imágenes

nes aéreas, con menor incidencia del efecto atmosférico, en muchos casos es suficiente realizar una normalización radiométrica lineal, alterando la reflectancia de una de las dos imágenes reescalándola y sumando o restando alguna constante, de modo que la imagen de referencia y la normalizada sean equivalente desde el punto de vista radiométrico:

$$\hat{\rho}_k = a_k \rho_k + b_k \quad (1)$$

donde ρ_k es la reflectancia en la banda y $\hat{\rho}_k$ es el valor normalizado de reflectancia normalizada que se pretende conseguir en esa banda. De este modo se altera la imagen objetivo con la ayuda de los parámetros a_k y b_k , que se extraen de la imagen de referencia.

Dado que la invariabilidad geométrica se confía a la corrección geométrica, no se insistirá más en ello en este capítulo. Quedan por establecer los procedimientos de homologación de las imágenes desde el punto de vista radiométrico, a lo cual se dedica el siguiente apartado, para pasar finalmente a exponer algunas de las técnicas más usuales en la detección de cambios.

Métodos de normalización radiométrica

Como se ha dicho, es un requisito previo que las imágenes a comparar sean equivalentes desde el punto de vista radiométrico, de tal manera que las variaciones en los ND de las celdas entre una y otra fecha, más allá de la variación inherente a la propia cubierta, sea debido a un cambio real en la ocupación de suelo. Aun habiendo realizado una normalización topográfica y la corrección atmosférica previa, suelen

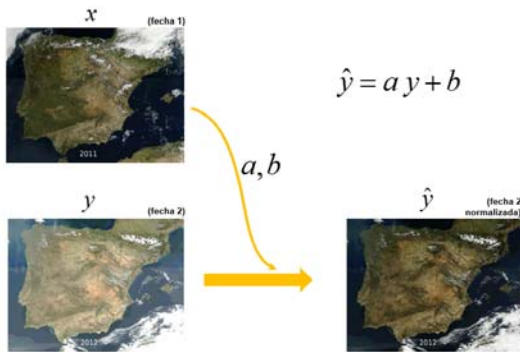
persistir diferencias ligeras en los ND de celdas supuestamente invariantes. Es por esto por lo que se impone la necesidad de normalizar una de las imágenes a los niveles radiométricos de la otra, la cual será considerada como referencia, siendo irrelevante cuál de las dos sea tomada como tal. En la exposición que sigue, se toma la imagen de la primera fecha como referencia, pero bien podría ser al contrario.

El procedimiento general a seguir consistirá en una transformación lineal, calculada banda a banda. Para una determinada banda k :

$$\hat{y} = a y + b \quad (2)$$

en cuya expresión se considera la siguiente terminología, que será la que se mantendrá en lo sucesivo:

- a y b son dos parámetros deducidos de ciertos valores de la variable x , que representa el ND , o la reflectancia en su caso, en la fecha 1, que se toma como imagen de referencia.
- y es el ND , o la reflectancia en su caso, en la fecha 2. Corresponde a la imagen objetivo.
- \hat{y} es el ND normalizado, o la reflectancia normalizada en su caso, en la fecha 2, con los valores de referencia de la fecha 1.



Esquema del proceso de normalización radiométrica.

Los distintos procedimientos de normalización sólo difieren en la forma de calcular a y b , a partir de los valores de la imagen de referencia x , que suelen deducirse de los estadísticos de esta imagen.

Normalización por regresión en zonas invariantes

Este método, conocido como *No-Change Regression Normalization*, se basa en la búsqueda de regiones de no-cambio de tierra firme y agua en el diagrama de dispersión multitemporal del IRp.

Los parámetros a y b se calculan del siguiente modo:

$$a = \frac{\sigma_{xy}^{(NC)}}{\sigma_{yy}^{(NC)}}$$

$$b = \bar{x}^{(NC)} - a \bar{y}^{(NC)}$$

de tal modo que la transformación es:

$$\hat{y} = \frac{\sigma_{xy}^{(NC)}}{\sigma_y^{2(NC)}} (y - \bar{y}^{(NC)}) + \bar{x}^{(NC)} \quad (3)$$

Nótese que esta expresión tiene la forma de (1). En todas las ecuaciones anteriores:

$\sigma_{xy}^{(NC)} = \frac{1}{N_{NC}} \sum_{NC} (x - \bar{x}^{(NC)}) (y - \bar{y}^{(NC)})$ es la covarianza entre las dos fechas para las zonas invariantes.

$\sigma_{yy}^{(NC)} = \frac{1}{N_{NC}} \sum_{NC} (y - \bar{y}^{(NC)})^2 = \sigma^{2(NC)}$ es la varianza de las zonas invariantes.

Normalización por regresión simple

Conocido como *Simple Regression Normalization*, éste es un método similar al anterior, pero en el que se extraen los estadísticos de toda la imagen, en lugar de hacerlo sobre una zona concreta de no-cambio.

Ahora:

$$a = \frac{\sigma_{xy}}{\sigma_y^2}$$

$$b = \bar{x} - a \bar{y}$$

y

$$\hat{y} = \frac{\sigma_{xy}}{\sigma_y^2} (y - \bar{y}) + \bar{x} \quad (4)$$

donde:

$\sigma_{xy} = \frac{1}{N} \sum_N (x - \bar{x})(x - \bar{y})$ es la covarianza [fecha 1- fecha 2] de las dos imágenes.

$\sigma_y^2 = \frac{1}{N} \sum_N (y - \bar{y})^2$ es la varianza de la imagen de la fecha objetivo (fecha 2).

Normalización por conjunto de control radiométrico

La *Radiometric Control Set Normalization* utiliza la media aritmética de un subconjunto de celdas oscuras y brillantes extraído de las imágenes referencia y objetivo mediante la transformación *tasseled cap*. Establecidos ambos conjuntos de datos, se calculan los parámetros con el siguiente procedimiento:

$$a = \frac{\bar{x}^{(B)} - \bar{x}^{(D)}}{\bar{y}^{(B)} - \bar{y}^{(D)}}$$

$$b = \frac{\bar{x}^{(D)} \bar{y}^{(B)} - \bar{x}^{(B)} \bar{y}^{(D)}}{\bar{y}^{(B)} - \bar{y}^{(D)}}$$

donde el superíndice (*B* indica las regiones de alto brillo (*bright*) y *D* indica las oscuras (*dark*). La normalización se lleva a cabo con:

$$\hat{y} = \frac{\bar{x}^{(B)} (y - \bar{y}^{(D)}) - \bar{x}^{(D)} (y + \bar{y}^{(B)})}{\bar{y}^{(B)} - \bar{y}^{(D)}} \quad (5)$$

Normalización por corrección de la neblina

En realidad, la *Haze Correction Normalization*, como se conoce en los textos ingleses, se refiere a cualquier tipo de efecto atmosférico dispersivo, bien

causado por aerosoles líquidos, como las nieblas, bien sólidos, como las calimas. Se basa en el siguiente postulado: sobre un área de reflectancia nula, la imagen debería aparecer oscura. De apreciarse alguna reflectancia aparente sobre este tipo de zonas, con de ND significativamente distintos de cero, debe ser la atmósfera la que está proporcionando ese nivel de brillo como consecuencia del efecto atmosférico, particularmente debido a fenómenos de dispersión aerosólica.

Esta corrección se basa en hacer el mínimo *ND* de la imagen objetivo igual al de la de referencia. Los parámetros son triviales:

$$a = 1$$

$$b = x_{\min} - y_{\min}$$

quedando:

$$\hat{y} = y - y_{\min} + x_{\min} \quad (6)$$

Solamente desplaza el histograma de la imagen objetivo, sin escalarlo, para hacer coincidir los valores mínimos de los dos histogramas.

Normalización por la media y desviación típica

Conocido como *Mean-Standard Deviation Normalization*, mediante este procedimiento se igualan las medias y las desviaciones estándar de la imagen objetivo a las de la imagen de referencia. La ganancia y el sesgo de la transformación son:

$$a = \frac{\sigma_x}{\sigma_y}$$

$$b = \bar{x} - a\bar{y}$$

y la normalización es:

$$\hat{y} = \frac{\sigma_x}{\sigma_y} (y - \bar{y}) + \bar{x} \quad (7)$$

Se desplaza y reescala el histograma de la imagen objetivo para asemejarlo al de la de referencia. En éste, como en todos los procedimientos similares que pretenden igualar (*matching*) dos histogramas, se sigue el esquema siguiente:

1. Se desplaza el histograma de la imagen objetivo hasta centrar su media en el origen $(y - \bar{y})$.
2. Se dividen los valores por la varianza o la desviación típica del histograma objetivo y se multiplica por la del histograma de referencia que se desea obtener σ_x/σ_y .
3. Se desplaza el histograma resultante hasta alcanzar su media el valor promedio de la imagen de referencia $(+\bar{x})$.

Normalización por mínimo y máximo

La *Minimum-Maximum Normalization* ajusta, banda a banda, el contraste de la imagen objetivo para conseguir histogramas con los mismos valores máximos y mínimos de las bandas homólogas de la imagen de referencia.

$$a = \frac{x_{m\acute{a}x} - x_{m\acute{i}n}}{y_{m\acute{a}x} - y_{m\acute{i}n}}$$

$$b = x_{m\acute{i}n} - a y_{m\acute{i}n}$$

La imagen normalizada se obtiene aplicando:

$$\hat{y} = \frac{x_{m\acute{a}x} - x_{m\acute{i}n}}{y_{m\acute{a}x} - y_{m\acute{i}n}} (y - y_{m\acute{i}n}) + x_{m\acute{i}n} \quad (8)$$

La transformación escala el histograma de la segunda fecha utilizando el criterio de valores mínimo y máximo, sin reparar en la varianza.

Normalización por regiones pseudo invariantes

En esta transformación, conocida como *Pseudo-Invariant Normalization*, se analizan elementos de reflectancia invariante (*Pseudo-Invariant Features, PIFs*). Su desarrollo es similar al visto en (7), pero en esta ocasión, los estadísticos se extraen solamente de un subconjunto de celdas de las que se tiene certeza de que su reflectancia no cambia con el tiempo, ni por la estación del año, ni por la hora, ni por el grado de humedad. Los parámetros de la transformación se basan en el cálculo de la media y varianza de esas zonas, igualando las de la imagen objetivo a las de la de referencia.

$$a = \frac{\sigma_x^{(pi)}}{\sigma_y^{(pi)}}$$

$$b = \bar{x}^{(pi)} - a \bar{y}^{(pi)}$$

$$\hat{y} = \frac{\sigma_x^{(pi)}}{\sigma_y^{(pi)}} (y - \bar{y}^{(pi)}) + \bar{x}^{(pi)} \quad (9)$$

donde el superíndice (pi hace referencia a las regiones pseudoinvariantes. Para la aplicación del método pueden utilizarse superficies como el hormigón, las

superficies de aglomerado asfáltico o ciertas cubiertas de edificios, en general, elementos exentos de vegetación y de materiales que cambien su contenido de humedad.

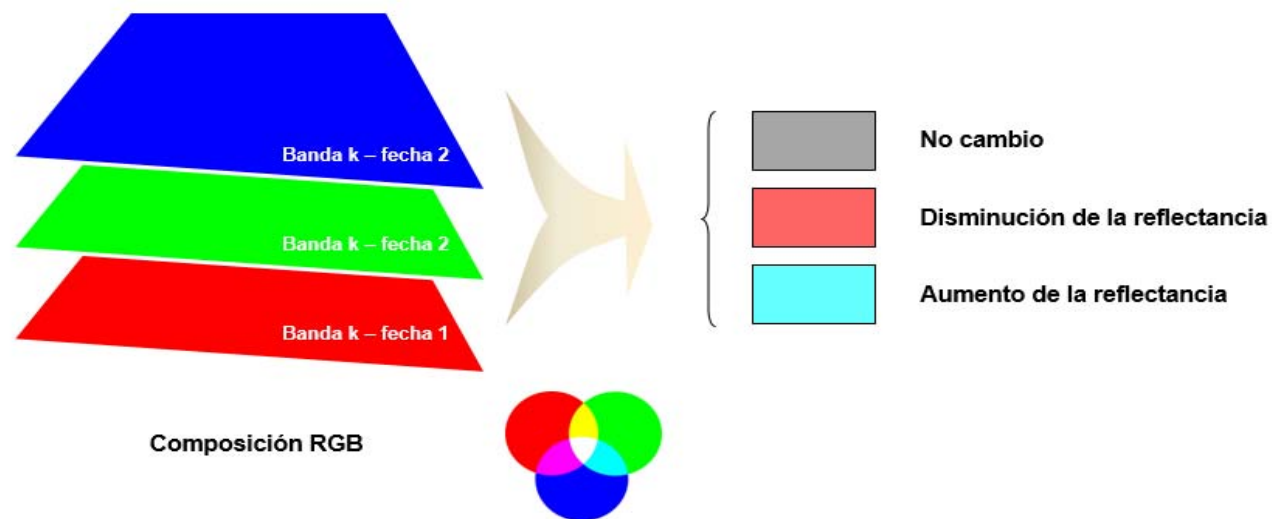
Técnicas de detección de cambios

En este apartado se recogen algunos de los procedimientos habituales para comparar dos imágenes adquiridas en fechas diferentes. El análisis multitemporal puede exigir un estudio más minucioso que recorra varias etapas entre el instante inicial y el final. De hecho, la evolución de una superficie puede requerir muchas observaciones a lo largo de un plazo amplio. Esto entraña cierta dificultad a la hora de representar la evolución, salvo que se emplee el tiempo en la propia representación bajo alguna modalidad de animación o se represente en forma de gráfica de dos ejes la evolución de alguna característica de una sola celda o de un conjunto de ellas, usando la variable tiempo como abscisa. Las explicaciones siguientes son válidas para la comparación de sólo dos momentos, que pueden ser el inicial y el final o bien cualquier otro intervalo temporal intermedio.

Composiciones en color multitemporal

Cuando se desea comparar mediante un análisis visual rápido la variación de una superficie entre dos fechas se puede recurrir a la composición en color RGB de una banda seleccionada de la imagen en cada una de las fechas. Como en este caso se dispone de tres canales, pero sólo de dos fuentes de información, una de las bandas ha de repetirse en la composición. Así pues, se selecciona la banda en la cual la reflectancia sea más variable entre las dos fechas.

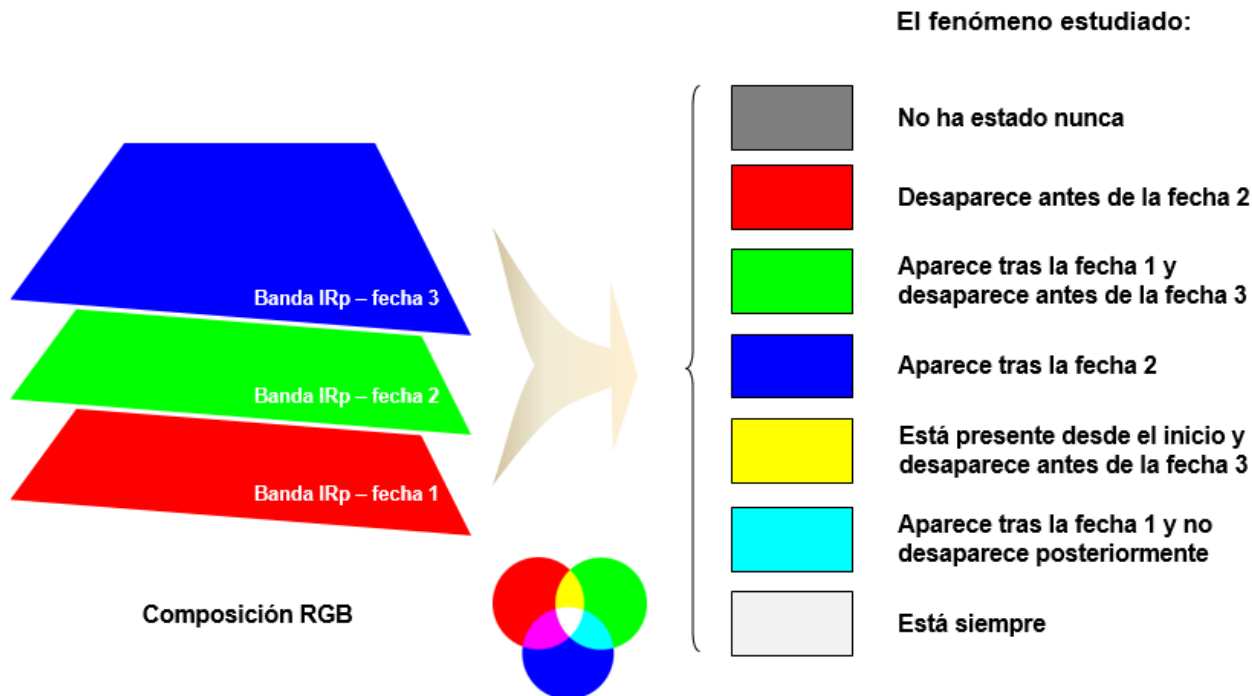
Por ejemplo, si se trata de identificar la presencia de un cierto cultivo y se desea conocer cuándo ha estado presente sobre el terreno, se puede seleccionar la banda del infrarrojo próximo, en la cual el vigor vegetal proporcionará crecientes niveles de reflectancia. Si se asigna la banda de la fecha inicial al color rojo (R) de la composición en falso color y la de la fecha final tanto al verde (G) como al azul (B), un incremento de la reflectancia entre ambas fechas



Composición en color multitemporal utilizando sólo dos fechas.

mostrará tonos en los que estará ausente la componente R y predominarán las componentes G y B. El resultado, pues, serán tonos próximos al cian (C). Esto viene a indicar que, si el cultivo buscado se desarrolla entre las dos fechas y aún está vigoroso durante la segunda, sólo las celdas de color cian serán las que cumplan esa condición.

Puede suceder que las parcelas con el cultivo buscado estuvieran en pleno desarrollo vegetativo en la primera fecha, pero en la segunda el cultivo o bien esté agostado o directamente haya sido cosechado y



Composición en color multitemporal utilizando tres fechas.

labrado. En ese caso predominará el color rojo sobre la mezcla de verde y azul, aunque esta última no desaparezca del todo. El color resultante será un rojizo más o menos puro.

El tercer caso que puede darse es cuando el cultivo haya estado presente en las dos fechas consideradas, y con un vigor vegetal similar, o bien no haya estado nunca presente. En ambos casos los niveles asignados a los canales R, G y B será similares y el resultado es que las parcelas que cumplan esa condición exhibirán una tonalidad gris-parda.

Sin embargo, dado que se dispone de tres canales, es posible introducir en el estudio multitemporal hasta tres fechas sucesivas y sintetizar el resultado en una sola imagen. El intérprete adiestrado identificará sin problemas la presencia, ausencia o evolución de un cierto cultivo, dentro de un conjunto reducido y conocido de ellos. En ese caso se asigna la banda del infrarrojo próximo de la fecha inicial, 1, al canal R, la de la fecha intermedia, 2, al G y la de la fecha final, 3, al B.

Recordando la teoría de la síntesis aditiva de formación del color en el sistema RGB, la interpretación de los posibles casos viene dada así:

- En las celdas con tonalidades gris o pardo oscuro el cultivo buscado no ha estado presente nunca. El suelo está desnudo.
- En las celdas rojizas, el cultivo está presente en la fecha 1 pero desaparece antes de la fecha 2.
- En las celdas verdosas, el cultivo aparece tras la fecha 1 y desaparece antes de la fecha 3.
- El color azulado indica que el cultivo sólo aparece tras la fecha 2, y sigue presente en la fecha 3.
- Si el color de la celda es amarillento, el cultivo está presente desde el inicio y durante la fecha 2,

- pero desaparece antes de la fecha 3.
- Si el color tiende al cian, el cultivo no está presente en el inicio del período estudiado, sino que nace tras esa fecha, pero ya está desarrollado en la fecha 2 y se mantiene sobre el terreno en la última fecha.
 - Finalmente, si el tono es muy claro, gris o pardo brillante, el cultivo ha estado sobre el terreno durante todo el intervalo estudiado.

Esta metodología se utiliza habitualmente en el seguimiento de la ocupación de suelo por determinados cultivos beneficiarios de subvenciones por las administraciones públicas. En su origen se realizaba un análisis visual y hoy día se lleva a cabo mediante análisis automatizado digital de imágenes.

Diferencia entre imágenes

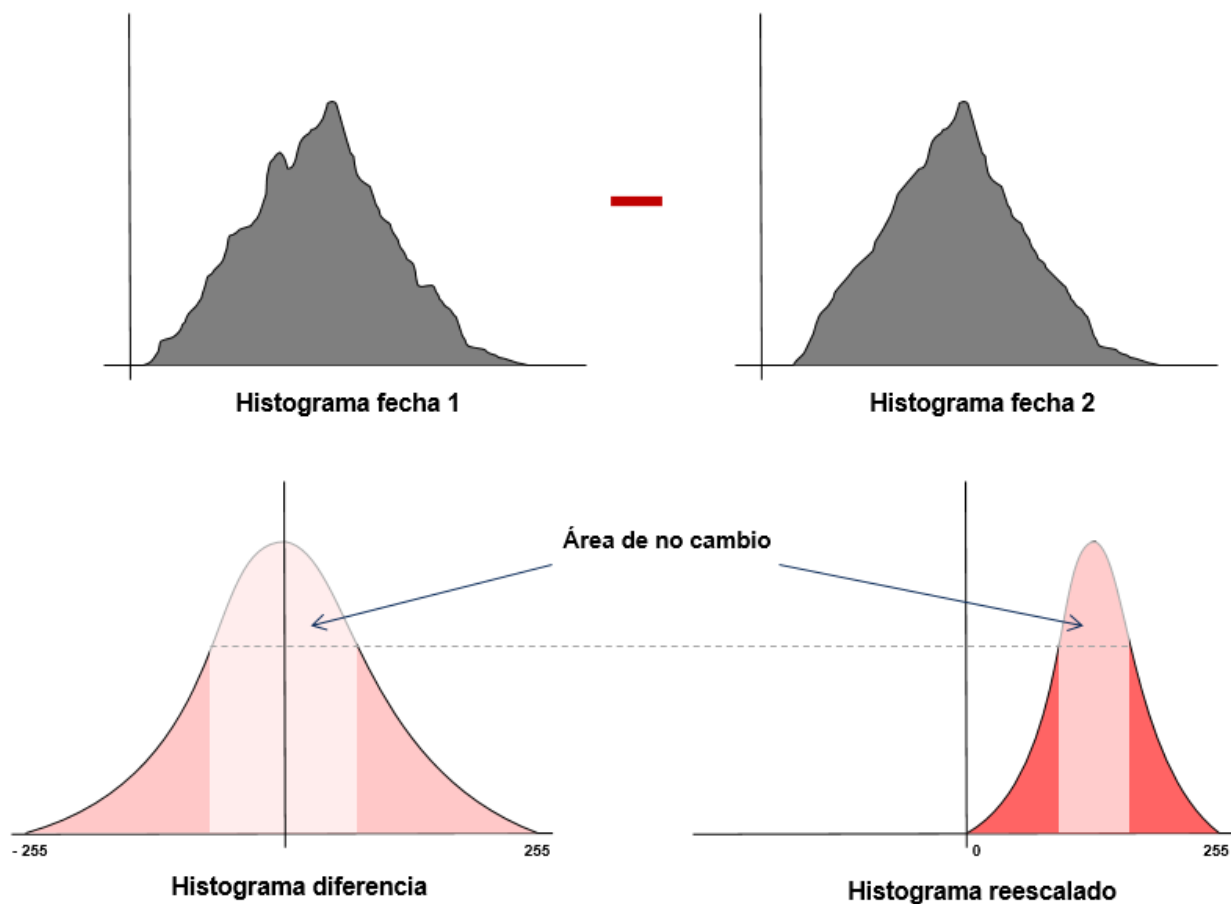
Una de las opciones más inmediatas para detectar cualquier cambio entre dos fechas en una escena es restar las imágenes de reflectancia de ambas:

$$DI = \rho_{t_2} - \rho_{t_1} + C$$

Es obvio que, admitiendo un mínimo grado de variación de la variable aun no habiendo cambiado la cobertura del suelo, tras la diferencia de las imágenes, las celdas que no hayan mutado ostentarán unos valores próximos a cero. Por el contrario, las áreas de cambio mostrarán resultados sensiblemente positivos o negativos.

Habitualmente, será necesario reescalar el contraste de la imagen diferencia, o trasladar su histograma, para hacerla apreciable en un sistema de visualización de 8 bits o superior. Ese histograma de los valores de la imagen diferencia tenderá a distribuirse gaussianamente. Para definir las zonas de cambio en

la imagen se deberá establecer un valor umbral para la frontera cambio-no cambio en el histograma atendiendo, si es posible, a algún rasgo peculiar evidente, a izquierda y derecha del valor modal. La elección de este umbral es determinante para obtener resultados demasiado restrictivos si el valor está próximo a la moda de la distribución, esto es, perdiendo cambios y sólo identificando los muy ostensibles, con gran



Identificación de las áreas de no cambio mediante la diferencia de imágenes.

error de omisión, o bien obtener resultados muy laxos, resultando cambios en celdas cuya reflectancia ha variado muy ligeramente, posiblemente como consecuencia de la propia variación interna de la clase que representa y cometiendo así numerosos errores de comisión.

Cocientes multitemporales

Otra opción sencilla es dividir celda a celda los valores de reflectancia de una determinada banda de las imágenes de las dos fechas.

$$CM = \frac{\rho_{t_2}}{\rho_{t_1}} \cdot C$$

El valor de CM se mantendrá próximo al valor de C para las zonas invariantes. Este procedimiento tiene la ventaja de reflejar el valor relativo del cambio, frente al valor absoluto que proporciona el análisis por diferencias de imágenes.

Para evitar valores fuera de rango y hacer comparables los cocientes, puede emplearse el índice del cociente multitemporal de diferencia normalizada, que aporta la cualidad de estar acotado entre 0 y 1. Para un sistema de visualización de 8 bits de profundidad:

$$NDRI = \left(\frac{\rho_{t_2} - \rho_{t_1}}{\rho_{t_2} + \rho_{t_1}} \right)$$

La interpretación del *NDRI* es la siguiente:

- Si ρ crece mucho, el *NDRI* tiende a 1.
- Si ρ , varía muy poco el *NDRI* está próximo a 0.
- Si ρ disminuye mucho, el *NDRI* tiende a -1.

También puede utilizarse una versión escalada del mismo índice:

$$NDRI' = \left(\frac{\rho_{t_2} - \rho_{t_1}}{\rho_{t_2} + \rho_{t_1}} + 1 \right) \cdot 127,5$$

Ahora:

- Si ρ crece mucho, el *NDRI'* tiende a 255.
- Si ρ varía muy poco, el *NDRI'* oscila sobre 127.
- Si ρ disminuye mucho, el *NDRI'* tiende a 0.

Diferencia entre los índices de vegetación

Al igual que se hace con las imágenes, se puede calcular el cambio que el índice de vegetación de diferencia normalizada ha sufrido entre las dos fechas sustrayendo uno de otro.

$$DIV = NDVI_{t_2} - NDVI_{t_1} + C$$

Puede usarse cualquier otro índice de vegetación, por ejemplo, el SAVI:

$$DIV' = SAVI_{t_2} - SAVI_{t_1} + C$$

De no añadir la constante C, ambos índices estarán acotados entre -2 y 2. Al igual que en la diferencia de imágenes, los valores de la diferencia de índices suelen presentar una distribución normal. Las zonas de no cambio tendrán valores próximos a 0 y las que hayan cambiado serán las de mayor valor absoluto, tanto positivo como negativo. Es conveniente buscar algún rasgo significativo en el histograma de la distribución para delimitar el umbral de valores de la zona de cambio.

También se puede recurrir a la diferencia normalizada de los índices de ambas fechas, obteniéndose el índice multitemporal de diferencia normalizada:

$$MNDVI = \frac{NDVI_{t_2} - NDVI_{t_1}}{NDVI_{t_2} + NDVI_{t_1}}$$

que proporciona valores normalizados entre -1 y 1.

Otra forma de enfocar el análisis multitemporal con los índices de vegetación es aplicando el color tal como se ha explicado arriba sintetizando una sola imagen analítica con la información del índice en hasta tres fechas diferentes. Como antes, se puede utilizar el NDVI o bien el SAVI, si la escena muestra gran proporción de suelo desnudo. Para la composición en color se asigna el cañón R al índice de la primera fecha, el G al de la segunda y el B al de la última fecha. La composición RGB resultante permite identificar las masas vegetales que aparecieron o desaparecieron a lo largo plazo considerado, considerando la interpretación de los colores descrita anteriormente.

Regresión

Se considera que en un contexto de no-cambio, la reflectancia de la celda en el instante t_2 es una función lineal de la correspondiente al instante t_1 para cada banda k :

$$\hat{\rho}_k(t_2) = a \cdot \rho_k(t_1) + b$$

siendo $\hat{\rho}_k(t_2)$ la reflectancia estimada en t_2 a partir de la del momento t_1 y a y b , los parámetros del ajuste. El cambio real sufrido entre ambas fechas se calcula como la diferencia entre el valor realmente

observado y el estimado por regresión, que equivaldría al no-cambio:

$$\hat{D} = \rho_k(t_2) - \hat{\rho}_k(t_2)$$

Análisis del vector de cambio

Cada celda de una imagen multispectral de k bandas está definida por su vector de características en un espacio vectorial k -dimensional. En el instante t_2 , el vector de características habrá variado con respecto al instante t_1 . El cambio se puede estimar midiendo la diferencia entre los vectores de características o el ángulo espectral entre ambos.

Atendiendo al primer criterio, la magnitud del cambio viene determinada por la distancia euclídea espectral entre los extremos de los dos vectores de características:

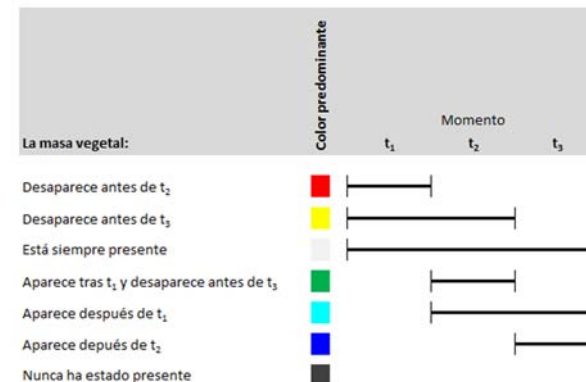
$$d_c = \left| \Delta \vec{z}_{(t_2-t_1)} \right| = \sqrt{(z_{1t_2} - z_{1t_1})^2 + (z_{2t_2} - z_{2t_1})^2 + \dots + (z_{kt_2} - z_{kt_1})^2}$$

midiendo el módulo del vector diferencia entre las dos posiciones espectrales de la celda en cuestión.

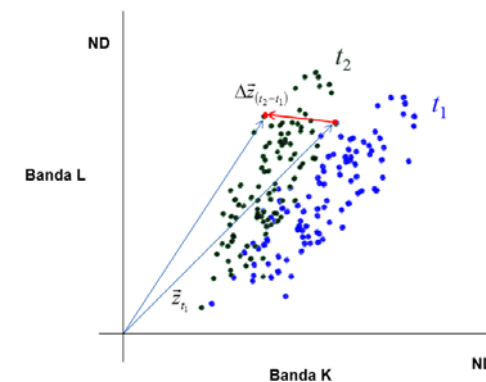
Ángulo espectral de cambio

Una forma alternativa de poner de manifiesto el cambio de la posición espectral de una celda es medir el ángulo formado por el vector de características de la celda entre los instantes inicial y final.

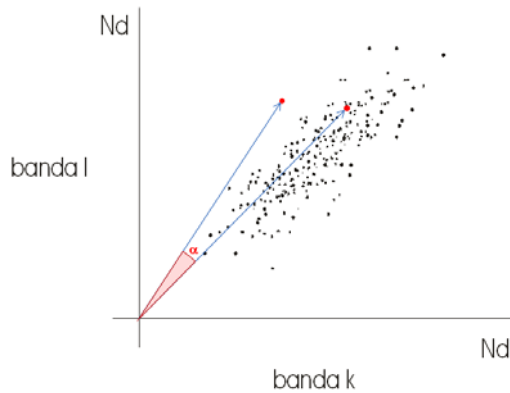
La abundancia de bandas asegura no detectar falsos negativos, ya que aunque el ángulo espectral sea próximo a 0 en algunas bandas, casi siempre se encontrará alguna otra en la que esto no se cumpla y, en



Interpretación del color en la composición de índices multitemporales.



Diferencia entre los vectores de características en dos momentos.



Ángulo espectral de cambio.

consecuencia, el ángulo espectral final será diferente de 0, indicando ello que habrá habido algún cambio. Solamente no se considerará cambio cuando el ángulo espectral continúe siendo 0 o muy próximo a este valor, tomado el conjunto completo de bandas. En este caso la celda habrá podido migrar de posición en el espacio espectral pero sólo en la dirección del origen de coordenadas espectrales. Lo cual indica una pérdida generalizada de la reflectancia, si es ésta la variable medida, que suele venir asociada a un aumento de la humedad. Si lo que se registra es una pérdida de radiancia, puede estar relacionada también con una peor iluminación de la ladera.

El ángulo espectral viene establecido:

$$\alpha = \arccos \left[\frac{\sum_{i=1}^k z_{i_1} z_{i_2}}{\sqrt{\sum_{i=1}^k z_{i_1}^2} \sqrt{\sum_{i=1}^k z_{i_2}^2}} \right]$$

donde z_{i_1} y z_{i_2} son las componentes del vector de características de la celda en el momento 1 y 2, respectivamente.

Transformación a componentes principales

La información contenida en una imagen multispectral está distribuida entre el conjunto de bandas, si bien no todas ellas proporcionan información independiente. Por el contrario, la información contenida en cualquier banda suele ser redundante en cierta medida con la contenida en otras bandas, sobre todo en las más cercanas en el espectro electromagnético. La redundancia en la información se manifiesta en una alta correlación entre los niveles digitales de ciertos canales.

La transformación o análisis de componentes principales (ACP) proporciona unas nuevas bandas linealmente independientes mediante la combinación lineal de las originales.

La aplicación de esta metodología al análisis multitemporal consiste en realizar la transformación con los dos juegos completos de bandas de las fechas a comparar. De esta manera, se extraen componentes principales, ordenados por varianza decreciente, del conjunto de bandas de las dos fechas.

La mayor varianza corresponde a los primeros componentes principales, que son los que contienen mayor cantidad de información. Han de contener, pues, aquellos rasgos presentes en ambas fechas, puesto que explican la mayor parte de información del doblete. En ellos habrá que buscar los elementos invariantes. Los componentes principales de ordinal superior muestran las superficies que hayan sufrido cambios. Finalmente, los últimos componentes principales suelen retener solamente el ruido de la imagen y no representan necesariamente los cambios netos reales.

Una técnica alternativa que utiliza el ACP consiste en restar los componentes principales extraídos separadamente de cada fecha.

Correlación cruzada

Se basa en la comparación de una imagen multispectral del momento t_2 con una imagen clasificada en el momento t_1 .

Se calcula el cuadrado de la diferencia entre el nivel digital de una celda en el instante t_2 y el promedio de todas las celdas que, en ese mismo instante, están en-

globadas bajo el perímetro de clase a la cual pertenecía en t_1 , según una clasificación realizada en esa primera fecha. Se divide el valor anterior por la varianza de cada área y se suman los valores obtenidos para todas las bandas:

$$Z_C = \sum_{i=1}^k \frac{(z_i - \bar{z}_{Ci})^2}{\sigma_{Ci}^2}$$

Valores altos de Z_C indican mayor probabilidad de cambio entre la clase inicial en t_1 y la posterior en t_2 .

Comparación posclasificación

Este procedimiento es uno de los que goza de mayor difusión, en parte por la robustez de sus resultados y en parte por la facilidad en su interpretación.

Como en todos los métodos, ha de comenzar con la georreferenciación rigurosa de las imágenes de las dos fechas. A partir de entonces, se clasifican ambas con el mismo algoritmo. Es conveniente que la leyenda sea idéntica, para lo cual, si es necesario, puede incorporarse la categoría *no clasificado*.

Tras la clasificación, se comparan celda a celda las dos imágenes y se confecciona una tabla de contingencia con los resultados. Pueden ponerse en columnas las clases en el momento t_1 y en filas las del momento t_2 , de tal forma que cada celda de la matriz indica el número de celdas de la imagen que en la fecha inicial pertenecían a la categoría señalada en el encabezado de su columna y en la fecha final lo están en la categoría marcada por su fila. Como es obvio, no habrán cambiado de categoría las cantidades de celdas recogidas en la diagonal principal de la matriz, cada una en la clase que le corresponda por columna o línea. La suma de las celdas fuera de la diagonal

principal indica el número de celdas de imagen que han sufrido algún tipo de cambio entre las fechas examinadas.

El estudio analítico de los cambios es sencillo. Sin embargo, es más compleja su presentación en forma de una imagen sinóptica que dé una idea fácilmente comprensible de las zonas de la escena que han permanecido inalteradas y aquellas otras que han sufrido cambio, y de qué naturaleza ha sido éste. Hay que tener en cuenta que, si el número de clases considerado es n , si bien algunas mutaciones en la mayoría de los casos son inusuales o incluso inviables, el posible número de cambios podría llegar a ser n^2 . Una manera rápida de confeccionar una leyenda de n^2 categorías sin confusión entre todas las posibles mutaciones de clase y que mantiene la identidad de las celdas que no han cambiado es la siguiente:

Si n_2 es el ordinal de la clase en t_2 y n_1 el correspondiente al momento t_1 , una asignación inequívoca podría consistir en etiquetar el cambio $n_1 \rightarrow n_2$ calculando:

$$\left[(n_2 + 1)^2 + 1 - n_1 \right]$$

Obtenida la etiqueta del cambio, sólo resta asignarle un color de la paleta cualitativa previamente confeccionada.

No obstante, es discutible que la imagen sintética de cambios obtenida pueda ser fácilmente interpretable sin recurrir continuamente a consultar la leyenda. Por encima de 4 clases no es práctico representar los cambios en un solo mapa cualitativo, ya que la lectura de los colores o de las tramas de la paleta puede resultar difícil. Por ello, es más práctico, aunque requiera presentar un mayor número de imágenes, realizar n mapas, uno por clase, siendo n el número

Imagen de 1200x800	t_1				
	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5
	1	2	3	4	5
t_2 Clase 1 5	4	3	2	1	0
Clase 2 10	9	8	7	6	5
Clase 3 17	16	15	14	13	12
Clase 4 26	25	24	23	22	21
Clase 5 37	36	35	34	33	32

Codificación de cambios para una imagen con 5 clases iniciales y finales. Se confeccionan primeramente los ordinales modificados de las filas (en blanco) a partir de los ordinarios de clase mediante $(n_2+1)^2$ y se le sustrae el ordinal de la columna (también en blanco). Con ello se obtienen índices biunívocos de cambio de clase para las 25 posibilidades, con los que confeccionar una leyenda.

de clases, cada uno de los cuales contiene un máximo de n categorías, en los cuales se presenta la evolución de cada una de las clases de la leyenda. Puede realizarse desde la perspectiva de las clases originales, es decir, la superficie que en t_1 era clase c_i ha pasado a ser c_1 , c_2 , etc. (de ahí que se originen hasta n categorías), o también desde la perspectiva de la ocupación actual del suelo, o sea, lo que en t_2 es realmente c_j , en t_1 fue c_1 , c_2 , etc. Las aplicaciones informáticas de uso común en teledetección proporcionan una imagen cualitativa multispectral con n bandas, una por clase final, cada una de las cuales tiene a su vez n categorías, una por clase inicial, y son susceptibles de invertir de igual forma la dirección temporal del análisis.

Clasificación multitemporal

Este procedimiento clasifica como un solo conjunto los dos juegos de bandas correspondientes a ambas fechas. En el espacio de características se formarán dos tipos de nubes de puntos: aquellos que corresponden a celdas invariantes, que verán *duplicados* sus vectores de características, al haberse mantenido estáticas en el espacio espectral, y aquellas otras que, al haber sufrido cambios, presentarán un desdoblamiento de tales vectores. En el proceso de clasificación no supervisada, será fácil identificar esos *clusters* mutantes.

15

Metodología en teledetección

Introducción

Una vez establecidos los principios físicos de la teledetección, presentados los elementos básicos de los sistemas de adquisición y tratamiento de imágenes y abordados los procedimientos más al uso de la explotación de la información contenida en ellas, es el momento de integrar estos conocimientos para dirigir adecuada y eficazmente las tareas que deben abordarse en proyectos que impliquen el uso de imágenes de satélite.

En realidad, no existe una metodología general que sea extensible a todos los trabajos basados en la teledetección. Cada proyecto exige emplear procedimientos específicos orientados a obtener de las imágenes empleadas la información que se busca y desentrañar de ellas las capacidades que encierran. Aun así, hay procesos más o menos recurrentes que son necesarios en la mayor parte de estos proyectos y serán el objeto de este capítulo.

Decisiones iniciales

El primer problema con el que va a enfrentarse el responsable del proyecto es evaluar si la teledetección cubre total o parcialmente el objetivo, y a qué coste. La Ingeniería, acomete sus objetivos desde un punto de vista en el que el criterio de la rentabilidad siempre debe estar presente. En este ámbito, la toma de decisiones debe integrar la eficiencia como parte esencial. Si se pueden obtener resultados similares a un menor coste, se preferirán las otras vías. No debe olvidarse que la teledetección en sí misma es una forma de adquisición de información, y que lo verdaderamente importante es el tratamiento posterior de los datos, a partir del cual podrán elevarse las conclusiones del estudio.

Es en la toma de datos en la que hay que juzgar la eventual ventaja que puede tener cada opción. Si el estudio se realiza sobre una gran área territorial, es muy posible que las imágenes de satélite aporten una información suficientemente extensa, aunque se haya

de sacrificar en parte el detalle. En otras ocasiones, la fotogrametría o la fotointerpretación, incluso asistida por ordenador, puede ser ventajosas cuando el coste de adquisición y el tratamiento de las imágenes aéreas sea inferior al de utilización de las de satélite y el objetivo propuesto no requiera de información de gran resolución espectral.

La imagen de satélite ofrece dos tipos de información, métrica y temática. En el primero, la resolución espacial que se ha conseguido en los últimos años en los productos de satélite rivaliza casi en pie de igualdad con las imágenes aéreas. Sin embargo, cuando el tamaño requerido de celda es del orden del décimetro, poder jugar con la altura de vuelo es claramente una ventaja de la que no gozan las imágenes de satélite. Por su parte, la dimensión temática —o semántica, como también se denomina— está muy relacionada con las capacidades espectrales de la imagen y, por esta razón, la cualidad más deseable es contar con un

número de canales tal que permita aplicar metodologías multi o hiperespectrales, o bien tener disponibles las bandas adecuadas para caracterizar el fenómeno de estudiado. Esta idoneidad se refiere tanto a la longitud de onda central del intervalo espectral observado como por a la propia anchura de la banda

La imagen debe seleccionarse según la precisión que

el trabajo requiera. No es lo mismo un proyecto de índole topográfica, en el que la exactitud geométrica es esencial, que otro de carácter eminentemente temático, donde pueda prevalecer la información espectral, textural, etc., aunque nunca haya que olvidar la calidad posicional. La cartografía puede ser *temática*, pero para serlo, antes debe ser *cartografía*.

Fases del proyecto de teledetección

Cuando el responsable de un proyecto en el que utilice la teledetección como herramienta acomete la tarea, ha de enfrentarse a una serie de decisiones que habrán de ir marcando los pasos que debe ir dando desde el principio, con las decisiones preliminares, hasta el final con el enunciado de las conclusiones del trabajo.

Las fases que a continuación se enumeran son las más habituales, aunque pueden o no estar todas ellas presentes en algunos proyectos específicos.

Definición de objetivos y propuesta metodológica

En primer lugar, se debe establecer con claridad el objetivo perseguido con el trabajo. Cuando se trata de un encargo, es de suponer que el promotor lo ha establecido con anterioridad. Sin embargo, es relativamente frecuente encontrarse con proyectos de teledetección cuyos pliegos de condiciones fijan estipulaciones difíciles o incluso imposibles de cumplir, que manifiestan el desconocimiento de la materia por parte de su redactor. En estos casos, es el propio responsable del trabajo el que debe reformular los objetivos acomodándolos al marco conceptual y técnico que la teledetección es capaz de ofrecer. Cuando se

trata de un trabajo de investigación, es el propio investigador principal el que se impone el objetivo, pero hasta en este caso resulta más eficaz enunciarlo claramente desde el principio, en lugar de ir reformulándolo a medida que se avanza en el estudio, a la luz de los resultados parciales que se vayan obteniendo. No es infrecuente saber de trabajos en pleno desarrollo pendientes de un título que se supeditarán finalmente a los logros conseguidos.

Establecido, pues el objetivo general del trabajo, y los objetivos secundarios que puedan desagregarse del principal, es necesario organizar la metodología más adecuada para alcanzarlos. En este caso está justificada la expresión *más adecuada* porque en determinados tipos de trabajos se puede llegar a unas conclusiones acordes con el objetivo desde varios enfoques diferentes. Obviamente en otros casos, sólo un itinerario conduce a cumplir el objetivo propuesto.

Desgranar la metodología supone definir el camino a seguir tras realizar un minucioso examen de todas las alternativas posibles en cada paso. Esto implica revisar la bibliografía, recopilar información auxiliar, decidir qué tipo de imágenes habrán de utilizarse, diseñar las características del trabajo o especificar las variables que se deben buscar y medir, planear los

trabajos de campo necesarios y suficientes, especificar las tareas de tratamiento y análisis digital de las imágenes, precisar la metodología de control de calidad, en el caso de no venir impuesta en las condiciones iniciales, especificar el alcance y la magnitud del informe final de conclusiones o diseñar los productos cartográficos, de haberlos.

Trabajos preliminares

Si bien es cierto que existen trabajos completamente pioneros e inéditos, la gran mayoría de los que se pueden acometer en la práctica cotidiana o mantienen similitudes con otros ya realizados o bien tienen antecedentes en los cuales basarse para proseguir hacia el objetivo. Para todos estos casos es, no sólo aconsejable, sino imprescindible realizar una revisión bibliográfica profunda. En el caso de trabajos de investigación, la doctrina impone esta fase como obligatoria para no repetir esfuerzos y optimizar así el tiempo, dedicándolo netamente a profundizar en el objeto investigado. En la práctica, una revisión bibliográfica rigurosa ahorra mucho tiempo al joven investigador. Jocosamente se dice de los que olvidan esta norma que *un par de años de laboratorio ahorran una buena semana de hemeroteca*.

Cuando se trata de trabajos de producción no necesariamente pioneros, consultar el estado de la cuestión —o estado del arte, una mala traducción de *state of art*—, es decir, el nivel de desarrollo de los dispositivos, procedimientos o técnicas alcanzado como resultado de la aplicación de métodos actuales en prácticas similares permite aprovechar el conocimiento adquirido sobre una materia para, basándose en él, aplicar la metodología más adecuada y eficiente.

Suele ser de gran utilidad familiarizarse con el área de trabajo visitándola, si es posible, y recopilando in-

formación auxiliar *in situ*. El conocimiento de las características y particularidades del terreno pueden facilitar encaminar los procesos de forma más eficiente, omitiendo pasos innecesarios, o allanando la elección de los métodos más eficaces.

Selección de la imagen

La correcta selección del sensor y de la imagen es determinante en el resultado del trabajo. Es crucial contar con los datos más adecuados de acuerdo al fin que se persigue, y cada satélite, y dentro de ellos, cada sensor, tiene un objetivo específico, para el cual ha cuidadosamente diseñado y desarrollado.

La selección de la imagen estará presidida por el tipo o los tipos de resolución que se requieran para el trabajo. Recordando los cuatro posibles, habrá que decidir si son limitantes o no la resolución espacial, radiométrica, espectral y temporal.

Si el proyecto tiene índole esencialmente geométrica, como puede ser una actualización cartográfica de escala media o grande, la resolución espacial no sólo es apreciada, sino requerida. Sólo con celdas pequeñas pueden reproducirse los detalles que marcan los elementos lineales y puntuales necesarios para este tipo de productos. Aunque con carácter general pudiera parecer que una resolución geométrica es siempre deseable, la práctica común lo desmiente. Algunas veces no es necesaria, como es el caso de la medida de la temperatura de la superficie marina, la cual es la misma en un determinado punto que unos cuantos metros más allá, y muchas más es contraproducente, porque el conjunto de datos a procesar es demasiado pesado para el beneficio que puede reportar. Cada tipo de trabajo tiene su escala y en función de ella debe optarse por un sensor u otro, en lo que se refiere a la resolución espacial. Los estudios regionales necesitan un tipo de imágenes y las actualizaciones

catastrales, otro.

La resolución radiométrica incrementa también el peso de los datos. Las imágenes con profundidad de color de 8 bits requieren un byte por celda. En cambio, las imágenes actuales, con resoluciones radiométricas de 10 o 12 bits requieren 2 bytes por celda. Si a eso se une que en muchas etapas del tratamiento digital se requiere operar en coma flotante, el procesamiento es mucho más voluminoso. Los equipos informáticos actuales no suelen ser limitantes en este sentido y sólo registran un ligero incremento en el tiempo de proceso, que no suele ser significativo.

El incremento de la resolución radiométrica está ofreciendo imágenes con márgenes dinámicos de miles de niveles digitales, lo cual ha posibilitado acometer estudios urbanos, donde las sombras dominan gran parte de las escenas y, aun así, dentro de ellas es posible discriminar objetos con una eficacia imposible anteriormente.

La resolución espectral es de las características más apreciadas en la teledetección, ya que es la que le otorga la capacidad de reproducir la signatura espectral de la superficie observada, y deducir a partir de ella su naturaleza, el estado en el que se encuentra e incluso la proporción de componentes que se encuentran mezclados en cada celda. Dentro de lo posible, pero siempre supeditado al objetivo del trabajo, se preferirán imágenes de mayor resolución espectral, si bien no necesariamente han de ser hiperespectrales.

Al carecer en la actualidad de sensores orbitales auténticamente hiperespectrales en estado operativo, se están utilizando cámaras similares instaladas a bordo de aeronaves. A su calidad hiperespectral, las imágenes generadas por estos sensores aúnan una destacada resolución espacial, que puede llegar a ser tan alta como se quiera, dado que depende de la altura

de la cámara sobre el terreno, y ésta obedece al plan de vuelo.

Una estrategia alternativa a la teledetección hiperespectral es la utilización de cámaras multispectrales más simples embarcadas en aviones o incluso en drones, pero que tienen la particularidad de tener sus pocos canales centrados en longitudes de onda escogidas de acuerdo al objetivo a observar, pero con una anchura espectral muy reducida, característica que comparten con sus semejantes hiperespectrales. Con estos instrumentos podrán caracterizarse muy eficazmente algunos tipos determinados de superficies y no serán válidos para otros proyectos, para los cuales habrán de montarse cámaras con otra configuración de canales, pero su bajo coste y la asequibilidad de las plataformas están marcando el presente y el futuro de muchas aplicaciones de la teledetección.

Queda por citar la resolución temporal. Esta característica, nacida al amparo de la repetitividad de la órbita de los satélites, sigue siendo de interés en determinado tipo de estudios y aplicaciones, particularmente las de índole meteorológico. Es de vital importancia para la predicción meteorológica alimentar los modelos matemáticos actuales con la mayor y más fiable fuente de datos, en lo posible en tiempo casi real. La resolución temporal de los satélites meteorológicos geoestacionarios ha alcanzado los pocos minutos y los datos que proporcionan permiten incrementar la exactitud del pronóstico a corto y medio plazo. En cuanto a los modelos climáticos, si bien se basan en los datos meteorológicos adquiridos por ese tipo de satélites, no requieren tanta resolución temporal porque solamente se nutren de datos promedios.

Evidentemente, para estudios de tipo cartográfico, geológico, agrícola, forestal, etc., la resolución temporal no es un requisito imprescindible.

Una vez que se ha seleccionado el satélite y el sensor que haya de utilizarse, hay otra cuestión que debe decidirse, que es la fecha de adquisición de la imagen. Es obvio que, si se trata de estudiar un fenómeno que sólo acontece sobre la superficie en determinadas épocas del año, la fecha de la imagen deberá pertenecer a ese intervalo temporal, probablemente contrastada con la información contenida en imágenes de fuera de esos plazos. En ausencia de requisitos o condicionamientos temporales, en general son preferibles imágenes de verano, debido a que el mayor nivel de radiancia asegura mayores rangos dinámicos en la imagen, y por ello la cantidad de información útil, medida por el contraste o la varianza, es mayor, facilitándose así el estudio y la discriminación de las superficies.

La resolución espacial del sensor debe estar en consonancia con la escala del trabajo, particularmente si la fase final incluye algún tipo de representación cartográfica. La celda de la imagen es la unidad de información geométrica y no cabe identificar detalles por debajo de su tamaño nominal. Para que las dimensiones de la celda queden por debajo del límite de la percepción visual, establecido en España por convenio en 0,2 mm, se requiere que su tamaño sea inferior a 0,04 mm². Este valor determina la escala de representación final. Si la escala viene impuesta, se deberá escoger el sensor que proporcione celdas que cumplan esa condición a dicha escala, y si, en cambio, es la imagen la que condiciona el producto, será la escala de representación la que habrá que acomodar. En cualquier caso, se deberá cumplir que:

$$\frac{p(m)}{e} \times 1000 \leq 0,2(mm)$$

donde p es el tamaño nominal de la celda en metros

y e es el denominador de la escala que debe adoptarse.

En la práctica, sin embargo, dado que en numerosos proyectos de teledetección priman las cualidades temáticas sobre las geométricas, las escalas recomendadas para cartografías realizadas a partir de los sensores comerciales más comunes suelen ser menos restrictivas, como las presentadas en la Tabla 1. Con todo, como ya se ha dicho, el mapa temático debe en primer lugar cumplir los requisitos mínimos de calidad posicional exigibles a toda cartografía. Sin embargo, de cara a la capacidad comunicativa del producto, es necesario recurrir al concepto de generalización temática. Uno de los objetivos de la generalización es incrementar la calidad interpretativa del mapa. En este contexto, la mínima unidad cartografiable es el mínimo tamaño de área que debe aparecer en un mapa cualitativo, a fin de no inducir a confusión. Por ello, la resolución de la información temática a menudo puede, y debe, ser generalizada.

Tabla 1. Resoluciones habituales en productos derivados de imágenes de satélite.

Sensor	Resolución espacial (m)	Tamaño de píxel (mm)	Escala final del mapa
NOAA-AVHRR	1100	0,55	1:2 000 000
LANDSAT 8-OLI (multiespectral)	30	0,3	1:100 000
LANDSAT 8-OLI (pancromático)	15	0,3	1:50 000
SPOT-HRV (XS)	10	0,4	1:25 000
SPOT-HRV (Pan)	5	0,2	1:25 000
IKONOS (Pan)	1	0,2	1:5 000
QuickBird (Pan)	0,6	0,3	1:2 000
GeoEye-2 (Pan)	0,34	0,23	1:1 500

Un ejemplo paradigmático de ello lo constituyen las imágenes clasificadas de ocupación de suelo. Si la escena presenta un área muy heterogénea y muy fragmentada en clases de poca extensión, el producto resultante será muy abigarrado. Es posible que presente toda la información posible, pero no es útil para un usuario medio que lo que desea es conocer la ocupación de suelo predominante en cada lugar, no los exotismos puntuales. Por esta razón, este tipo de imágenes suelen someterse a un filtrado de moda o mayoría que elimina zonas muy pequeñas pertenecientes a una clase y que están situadas en el seno de otras mayores que pertenecen a otra categoría. Se pierde información, pero se simplifica la interpretación y se gana en la legibilidad del mapa. Es un proceso de generalización temática.

Diseño de la leyenda

En numerosos trabajos basados en la teledetección se requiere clasificar automáticamente la escena. En una clasificación digital, la escala elegida, junto con

el adecuado nivel de generalización, condiciona el número de categorías a incluir en la leyenda, lo que a su vez determina el nivel de desagregación exigible a las categorías clasificadas.

La leyenda puede ser en sí una especificación del trabajo y, en este caso, el redactor del pliego de condiciones deberá conocer las posibilidades de las imágenes con las se ha de trabajar. El número de clases discriminables viene condicionado por la espectralidad de la imagen. Con imágenes de gran cantidad de bandas es posible construir leyendas muy extensas porque las clases podrán ser discriminadas eficazmente en virtud de pequeñas diferencias en sus firmas espectrales, bien caracterizadas por los canales disponibles. Si, por el contrario, se cuenta con un escaso número de bandas, y además de gran anchura espectral, la capacidad de diferenciación del clasificador estará muy limitada. Si en el primer caso se podía aspirar a construir una leyenda, por ejemplo, de pino carrasco, pino piñonero, pino radiata, pino negro o pino pinaster, en el segundo apenas se podrá aislar la categoría «pino» sin diferenciación intraclase, o incluso esperar ser capaces de discriminar simplemente la clase «monte alto». Para dimensionar adecuadamente el número de categorías previamente al entrenamiento del clasificador es muy conveniente realizar un análisis de separabilidad o incluso ejecutar una clasificación no supervisada con el fin de establecer la capacidad de la imagen para mostrar diferencias en la ocupación de suelo.

Trabajos de campo

En determinadas ocasiones es necesario realizar radiometría de campo o establecer con la máxima exactitud la posición de puntos de control geométrico. En esta fase es el momento de llevar ambas tareas a cabo. Algunos trabajos requieren calibrar los datos

proporcionados por el satélite con medidas espectrorradiométricas de la superficie. De este modo se puede, por ejemplo, definir la presencia de componentes puros —*endmembers*— a partir de los cuales en la etapa de proceso se podrá establecer la abundancia de cada uno de ellos en las celdas de la imagen de satélite. En otras ocasiones, simplemente se deseará contrastar las medidas de radiancia deducidas de los niveles digitales de la imagen con las medidas reales sobre el terreno.

Cuando la exactitud geométrica sea un requerimiento prioritario, será conveniente tomar las coordenadas de una serie de puntos de control con los métodos topográficos acordes con la precisión que se requiera para proceder posteriormente a la corrección geométrica de la imagen y a su ortorrectificación. En estas ocasiones se puede aprovechar la campaña de toma de datos para reservar una colección de puntos, no utilizados en la corrección, para el control final de la calidad posicional.

En cuanto a la componente temática, también esta fase es la adecuada para recolectar información sobre la ocupación de suelo o sobre la naturaleza de la superficie observada. Con ello se podrá asignar una etiqueta de clase a los campos de entrenamiento que se seleccionen en la imagen durante el procesamiento digital y también se dispondrá de información para el control de calidad temático si se planifica de antemano la campaña de verificación fijando las coordenadas de los puntos o áreas de comprobación y tomando nota de la ocupación real de ellas.

Muchas de estas tareas hoy día no se realizan necesariamente en campo. La riqueza de información gráfica disponible en distintos formatos, particularmente en forma de ortofotografías, permite en ciertos casos sustituir la visita al terreno por la consulta de tales fuentes. Sin embargo, es necesario precisar que el

analista deberá poner cuidado en la selección de las fuentes y asegurarse de que proporcionan una precisión igual o superior a la requerida para su trabajo. Idealmente, sería aconsejable incluso que esas fuentes gozaran de una precisión un orden de magnitud por encima del trabajo final. También se debe recordar que al contrastar una imagen de satélite con una fuente cartográfica y, sobre todo con una ortofotografía, la época de ambas coincida. Es más, es necesario que lo haga también la estación anual y, si es posible, el mes, a fin de que la radiometría sea similar y, lo que es muy importante cuando se trata de clasificaciones de usos del suelo, que el terreno esté ocupado por la misma cubierta. Cuando se clasifican imágenes orbitales antiguas, no siempre es posible contar con una ortofoto que cumpla esas condiciones. En estos casos, es necesario consultar otras fuentes bien o dejar a la experiencia del analista la interpretación final de los resultados.

Tratamiento digital

Esta etapa constituye el núcleo de los trabajos y en ella se extraerá la información buscada, oculta tras los niveles digitales de la imagen. En este punto no existe un itinerario fijado de antemano. El procedimiento a seguir dependerá del objeto del proyecto. En unos casos se cubrirá con operaciones de álgebra de bandas más o menos complejas, en otros se requerirá clasificación digital, etc. Pero sea cual sea el camino, previamente será necesario realizar las correcciones imprescindibles sobre la imagen, tanto en los aspectos radiométricos como en los geométricos.

Es preciso trabajar con imágenes corregidas del efecto atmosférico, en lo posible aplicando algoritmos potentes. Además, es muy conveniente, cuando no imprescindible, operar con niveles de reflectancia. Como se ha dicho anteriormente, lo que caracteriza a cada superficie es su reflectancia espectral. Salvo

que se esté estudiando la radiación solar mediante la medida de la radiancia procedente de la superficie, sólo la calibración a reflectancia proporcionará la información útil de la superficie.

Cuando se está trabajando con un conjunto de imágenes de fechas diferentes puede ser necesario también normalizar radiométricamente las imágenes. Así ha quedado establecido cuando se han estudiado la teledetección multitemporal. Entre los procedimientos disponibles, el analista podrá escoger el que sea más acorde con los datos disponibles, si bien goza de gran rigor el basado en las regiones de reflectancia pseudoinvariante.

Otra tarea obligatoria es la corrección geométrica de la imagen, especialmente si va a ser tratada en conjunto con otras. Las pequeñas deformaciones o diferencias posicionales de algunos elementos de las imágenes pueden ser suplidas sin problema en un análisis visual, pero cuando la tarea es confiada a un sistema de tratamiento digital, es fundamental la absoluta concordancia de los píxeles de todas las imágenes. De ahí la necesidad de acometer esta tarea. El tipo de corrección elegida debe asegurar la calidad posicional de las imágenes, no admitiéndose por lo general desviaciones posicionales que superen 2/3 del tamaño nominal de la celda. Para ello es muy aconsejable incorporar la corrección del efecto proyectivo del relieve. En cuanto al método de remuestreo elegido para escribir la imagen corregida, ha de tenerse en cuenta que, si se desea por encima de todo conservar el nivel digital o de reflectancia original, sólo el método del *vecino más próximo* lo garantiza. Por ello es el único útil cuando la información es de carácter cualitativo, como por ejemplo en las imágenes clasificadas. Si, por el contrario, la información es cuantitativa –caso de la reflectancia– puede venir utilizar procedimientos como el de *interpolación*

bilineal o el de *convolución cúbica* para conservar mejor la variabilidad del terreno, conseguir mayor suavidad en los resultados y evitar el *efecto escalera* en los rasgos lineales inclinados de la imagen. Finalmente, si la información obtenida ha de ser fusionada o superpuesta con otras imágenes o con fuentes vectoriales, conviene que en el mismo proceso de corrección geométrica se georreferencien en el sistema de coordenadas elegido.

En cuanto al momento de realizar las correcciones radiométricas y geométricas, conviene introducir algunas apreciaciones. Si la imagen es única, esto es, va a ser tratada independientemente sin vinculación a otras imágenes, aunque la corrección radiométrica es siempre necesaria, puede prescindirse de la normalización. Las operaciones de corrección geométrica pueden ser abordadas al inicio del tratamiento o en fases posteriores, tras acometer diversas operaciones algebraicas con las bandas, siempre que no se requiera por algún motivo la georreferenciación previa. Si, por ejemplo, han de establecerse campos de entrenamiento, es obvio que deberá hacerse sobre la imagen georreferenciada. Sin embargo, calcular un índice de vegetación no requiere la corrección geométrica anterior. Incluso, bajo determinadas circunstancias, puede ser indiferente el orden de los dos tipos de corrección, geométrica y radiométrica. Sin embargo, si tras el proceso de georreferenciación, la escena ha de ser recortada para obtener una subimagen y continuar el tratamiento con ella, conviene que la corrección atmosférica sea la primera, porque los algoritmos de tratamiento leen de la cabecera de la imagen los datos imprescindibles de fecha, altura solar, coordenadas del centro de la imagen, etc., información que puede ser alterada o incluso desaparecer con el recorte.

También debe tenerse muy en cuenta el nivel de corrección de la información de partida. Las imágenes

de satélite se distribuyen comercialmente con diferentes niveles de preprocesamiento que involucran correcciones radiométricas y geométricas. Cada gestor de las imágenes tiene los suyos, lo que hace necesaria la consulta a la información técnica correspondiente para conocer cuáles son las alteraciones introducidas. SPOT, por ejemplo, ofrece una gama de 5 niveles:

- Nivel 1A. Son corregidos los errores radiométricos originados por las diferencias de calibración entre los detectores del instrumento. No se aplica ninguna corrección geométrica. Tan sólo se proporcionan las coordenadas del centro y las esquinas de la escena para su localización sobre el terreno, sin tener en cuenta el desplazamiento debido al relieve. Es un nivel de preprocesamiento grosero solamente útil para usuarios experimentados que deseen operar con los datos originales.
- Nivel 1B. Además de la igualación radiométrica del nivel anterior, incluye el procesamiento geométrico. La imagen es corregida de las distorsiones causadas por las condiciones de captura como la variación de la actitud del satélite, el efecto panorámico, la curvatura y rotación terrestre, etc. La precisión en la localización es similar a la del nivel 1A y siguen sin tenerse en cuenta los efectos proyectivos del relieve.
- Nivel 2A. La escena se rectifica en una proyección cartográfica estándar, sin utilizar puntos de apoyo. Las correcciones geométricas tienen en cuenta las distorsiones debidas a las condiciones de captura de imagen e incluyen la georreferenciación. La imagen ya es superponible a otras fuentes cartográficas convencionales en el mismo sistema proyectivo, usualmente UTM-WGS84.
- Nivel 2B. Este nivel de corrección es denominado «precisión» y para conseguirlo se utilizan puntos de apoyo en el terreno. Este tipo de productos es

apto para su uso directo siempre que las deformaciones debidas al relieve no sean determinantes.

- Nivel 3. También denominado «orto», es un nivel de procesamiento geocodificado como el anterior, pero que además incluye las correcciones de paralaje debida al relieve, es decir, la ortorrectificación, de tal modo que el producto final es una ortoimagen.

Saber qué nivel de procesamiento incluye la imagen es importante de cara a establecer los pasos a seguir que siguen siendo necesarios en el preprocesamiento digital por parte del analista.

En cuanto a las tareas de clasificación, lo primero a tener en cuenta es el camino que se va a seguir. Entre los métodos no supervisados o automáticos y los supervisados, siempre son preferibles los segundos. Pero para conocer *a priori* las posibilidades que brinda la imagen con el fin de reajustar el número de categorías de la leyenda, especialmente cuando no se está familiarizado con la zona de actuación, puede ser muy útil una clasificación no supervisada previa, que puede ser, a su vez, coordinarse con un análisis de separabilidad de clases.

La clasificación no supervisada es capaz de detectar cualquier diferencia espectral entre celdas y encontrar agrupamientos *naturales* de valores que sugieren clases diferentes. Pero para obtener resultados coherentes y realmente útiles se le debe proporcionar al clasificador una cierta información previa de la cual es frecuente carecer. Por ejemplo, número de categorías deseado, porcentaje mínimo de celdas que debe ostentar cualquier clase, varianza de la clase a partir de la cual puede escindirse en dos, número de clases a fusionar en una sola categoría cuando la similitud sea alta, umbral de fugas de celdas entre categorías en las sucesivas iteraciones que provoque el final del ciclo de clasificación o número de máximo

iteraciones del proceso para producir la parada. Cuando se carece de este tipo de información es útil jugar con algunos de esos parámetros para detectar la mejor convergencia del proceso. Sin embargo, la solidez de estos clasificadores no produce otra cosa que clases *espectralmente* diferentes, pero no tienen la capacidad de etiquetarlas. Para ello será necesario la intervención del analista, que será quien identifique las categorías y decida si desagregarlas más aún o bien fusionar algunas de ellas.

Superada la fase de clasificación automática, conviene someter la imagen a una clasificación supervisada y para ello, de nuevo, habrá que fijar de antemano el número de clases de la leyenda, para lo cual será de gran utilidad la información obtenida en la primera. A continuación, se acomete la fase de entrenamiento del clasificador. Se trata de dotar al sistema de tratamiento del conocimiento necesario para establecer las clases. Dentro de los métodos paramétricos de clasificación, cada clase puede ser perfectamente definida por solo dos parámetros: el vector de medias y la matriz de covarianzas. La fase de entrenamiento consiste en alimentar el clasificador con información para que encuentre estos dos estadísticos para cada muestra de clase. Los campos de entrenamiento son polígonos que han de digitalizarse sobre la imagen a los cuales se les coloca la etiqueta de clase que corresponda. El número de campos de entrenamiento y la extensión que debe tener cada uno depende en gran medida de la variabilidad interna de las clases. Es preferible digitalizar varios polígonos por cada clase, tratando de que esa variabilidad quede bien representada que recolectar mayor cantidad de celdas bajo un solo polígono. En cualquier caso, el tamaño de la muestra debe regirse por las pautas indicadas en el capítulo de clasificación digital.

También conviene tener presente que cuando se clasifica una imagen de satélite el producto obtenido es

un mapa de la *ocupación* del suelo sólo en ese momento. Los mapas de *usos* del suelo se refieren a la ocupación habitual del terreno a lo largo de varios años. Para generar un producto así a partir de imágenes sería necesario repetir la clasificación al menos sobre un puñado de imágenes que abarquen un período de al menos dos años y aprecien los usos agrarios de la escena a lo largo del año. Sólo así podrán obviarse problemas como el de encontrar una parcela de cultivo en barbecho en el momento de la adquisición de la imagen, que sería clasificada como suelo desnudo, cuando lo habitual en ella sería obtener varias cosechas de cultivo intensivo al cabo del año.

Necesaria será también la habilidad del analista a la hora de etiquetar las categorías. En ocasiones es muy útil emplear denominaciones muy genéricas como *labor intensiva*, *labor extensiva* o *monte bajo* para designar zonas ocupadas respectivamente por cultivos de huerta o de frutales en regadío en alta densidad, tierras de secano dedicadas al cereal o leguminosas o, finalmente, asociaciones de matorral, arbolado muy disperso y pastizal, sin tener que particularizar excesivamente las clases, liberándose con ello de establecer una leyenda del mapa innecesariamente extensa.

Control de calidad

Dado que el proceso de clasificación digital está sometido a error, se hace necesario incluir en el trabajo el correspondiente control de calidad temático. El procedimiento implica comparar la imagen clasificada con una fuente de verdad-terreno que puede ser otra imagen clasificada previamente, una fuente cartográfica de veracidad contrastada o el propio terreno. Si las dos primeras se encuentran en formato digital, la comparación puede ser exhaustiva, esto es, examina las etiquetas de la imagen completa, celda a

celda. De esta manera los errores encontrados solamente serán imputables a un deficiente entrenamiento o al método de clasificación. Si, por el contrario, es necesario recurrir al terreno, aunque sea en un análisis visual sobre una ortofotografía aérea, obviamente habrá que elaborar una muestra de verificación. Entre los procedimientos vistos, uno de los más utilizados es el de muestreo sistemático no alineado. Hay que insistir en que lo que se determina ahora es el conjunto de lugares, celdas aisladas en este caso, a *visitar* para etiquetarlas de acuerdo con la leyenda del trabajo, según sea su ocupación real. Más que nunca importa que la fecha de la imagen y la de la fuente de comparación sean lo más cercanas posibles.

Establecida la verdad-terreno, su cruce con la imagen clasificada permite construir la matriz de confusión, a partir de la cual se extraerán los estadísticos de calidad, habitualmente porcentajes de acuerdo total y por clases, error del productor y del usuario y coeficiente kappa.

Si el trabajo tiene una vocación cartográfica clara, es muy conveniente someterlo además a pruebas de calidad posicional basadas en los estándares más extendidos: NMAS, EMAS, NSSDA, etc., de tal modo que el usuario pueda tener una idea de la exactitud geométrica del producto.

En cualquier caso, aún en el caso de no constituir una especificación en origen, es conveniente ajustar el control de calidad a los requerimientos establecidos en el Marco de Aseguramiento de la Calidad para la Observación de la Tierra (QA4EO), que pueden consultarse en <http://qa4eo.org/documentation/>.

Memoria

Parece superfluo añadir que todo trabajo de índole

cartográfico basado o no en la teledetección debería estar acompañado de un texto explicativo, pero la experiencia enseña que no pocos profesionales de la geomática y la topografía vuelcan su saber en realizar, y probablemente muy bien, su trabajo y, sin embargo, descuidan esta parte.

Todo trabajo debe adjuntar una memoria descriptiva que incluya los agentes intervinientes, los antecedentes, el objetivo del estudio, la información y/o los datos de partida, la metodología empleada, los resultados, con o sin discusión sobre ellos, y las conclusiones, si las hubiere. Cuando el procedimiento utilizado sea novedoso, puede ser conveniente precederlo de una pequeña revisión del estado de la cuestión y referirse a los autores que hayan preconizado dichos métodos. La inclusión o no de referencias es optativa en el caso de los trabajos profesionales y obligatoria cuando se trate de trabajos de investigación destinados a ser publicados en revistas científicas o técnicas. Finalmente, es deseable que la memoria incluya los metadatos a los que se refiere la norma ISO 19115-1.

Producción cartográfica

Si el trabajo tiene como fin la elaboración de un mapa, será necesario adoptar la metodología propia de la producción cartográfica. Desgranarla queda fuera del objeto de este texto, pero cabe recordar algunos aspectos importantes.

En primer lugar, deberá realizarse un diseño cartográfico que cumpla con el objeto del trabajo, a la par que reúna las cualidades básicas de la comunicación gráfica, como por ejemplo que sea legible y atractivo, y sencillo, pero a la vez completo. Por supuesto, el mapa debe contener todos los elementos cartográficos requeridos de acuerdo con su escala. La introducción de una escala gráfica previene de las posibles reproducciones *a posteriori* a diferente tamaño por

parte de los usuarios.

Si incluye una imagen de satélite tratada sobre la cual se pretende superponer información vectorial obtenida de fuentes ajenas, es vital que la corrección geométrica haya sido extremadamente rigurosa, so pena de quedar al descubierto desplazamientos traicioneros. Dicha imagen habrá sido realzada radiométrica y geoméricamente mediante la expansión del contraste y un filtrado de paso alto, para expresar del modo más eficaz todos los elementos presentes en la escena.

Cuando la información vectorial sea la prioritaria, la imagen de fondo no debe restar protagonismo ni dificultar su legibilidad. Para ello se degradará la imagen aclarándola y disminuyendo el contraste, todo ello en su justa medida para que, reconociéndola, no estorbe la identificación de rótulos o símbolos.

Si se ha realizado una clasificación digital, es evidente que lo importante es ella, y que la información vectorial es, en este caso, accesoria. Deberá incluirse la red hidrográfica y de vías de comunicación, porque son las que articulan el territorio. Se incluirán también los núcleos urbanos como vectores si no son lo suficientemente evidentes en la imagen. En cuanto a los colores adoptados para la leyenda de clases, deberán tener alguna relación con el tipo de superficie que representan: verde para la vegetación, azul para el agua, pardos o marrones para suelos desnudos, blanco para improductivos, etc. El blanco no hay que descartarlo y puede utilizarse perfectamente como un color más. Es muy importante advertir que los tonos empleados deberán ser *pastel*, esto es, muy poco saturados, y muy claros. Téngase en cuenta que, alejándose solo ligeramente del blanco, el ojo humano ya percibe el color como una clase diferente en el mapa. Lo contrario conduciría, se quiera o no, a la producción de un mapa con apariencia escolar. El

resto queda ya a la decisión y el gusto del diseñador. En mapas de ocupación de suelo, para enriquecer el fondo y evitar grandes zonas planas, hay quien fusiona la imagen clasificada con un sombreado ligero realizado a partir de un modelo digital de elevaciones, siempre iluminado desde la esquina superior izquierda, tan ligero que no llegue a desvirtuar la identificación de cada color de clase.

Cuando haya que representar las curvas de nivel del terreno sobre la imagen tratada, utilizar el color habitual *siena* produce resultados insatisfactorios. En estos casos una buena solución, aunque no admitida por todo el mundo, consiste en representar en blanco tanto las propias curvas como sus etiquetas, lo que asegura su identificación.

El texto de la toponimia es un aspecto delicado, puesto que no siempre es fácil que sea perfectamente perceptible. Si el mapa es oscuro, la opción de ofrecer los textos en negativo es una opción, pero sucede que frecuentemente las palabras cruzan espacios de diversos coloridos y claridades, de modo que mejor aún es dotar las letras de un filete del color opuesto al relleno, blanco sobre negro o negro sobre blanco. Otra solución acertada es adornar las palabras con un ligero halo de tonalidades contrarias al fondo del mapa.

El tamaño del documento final no debe estar sometido a un formato normalizado, salvo exigencias del editor o especificaciones de origen. Más bien, debe ser el adecuado para ofrecer, incluyendo un margen razonable, la información de la manera más clara y concisa posible, teniendo en cuenta a su vez el uso final. No es lo mismo un mapa destinado a utilizarse sobre una mesa que otro que deba ser consultado en un vehículo. En algunos casos, el tamaño excesivo o demasiado pequeño resulta inadecuado.

Finalmente, si el mapa va a ser plegado, su diseño deberá incluir este propósito, distribuyendo las masas cromáticas del documento adecuadamente, jerarquizando, ordenando y centrando los textos del reverso, de haberlo, en las columnas o cuarteles resultantes. Además, debe ajustarse el tamaño de los pliegues para que el producto final pueda caber, por ejemplo, en una carpeta o en un bolsillo.

Apéndice 1

Glosario

Absorbancia. Medida de la capacidad de un material para absorber energía electromagnética en una determinada longitud de onda. Es la razón existente entre el flujo absorbido y el incidente sobre una superficie.

ACP. Análisis de Componentes Principales. Transformación estadística generadora de nuevas bandas linealmente independientes a partir de otras parcialmente redundantes.

Albedo. Fracción de radiación que es reflejada por una superficie. Es sinónimo de reflectancia.

AVHRR. Acrónimo de *Advanced Very High Resolution Radiometer*, sensor montado sobre la plataforma NOAA.

Banda. Intervalo de longitud de onda dentro del espectro electromagnético. Por extensión, se denomina banda a cada uno de los canales de adquisición de datos de un sistema sensor.

Banda de absorción. Cada uno de los intervalos de longitud de onda en los que un cuerpo absorbe radiación electromagnética. En consecuencia, en un espectro de emisión/reflexión dichos intervalos presentarán valores nulos o proporcionalmente muy pequeños.

Bandeado de la imagen. Fenómeno por el cual la imagen espacial formada mediante un barredor de líneas (scanner) presenta bandas horizontales a intervalos regulares anormalmente claras u oscuras, como consecuencia del fallo de un detector del sensor durante la adquisición.

BIL. Formato de grabación de la imagen de bandas intercaladas por líneas, consistente en la disposición de una misma línea de barrido en todas las bandas sucesivamente, seguida de la siguiente línea en todas las bandas, y así sucesivamente.

Binarización de la imagen. Transformación consistente en la asignación exclusiva de valores 0

y 1 a la imagen transformada para las celdillas cuyos Nd en la imagen primitiva fuesen inferiores o superiores, respectivamente, a un determinado valor prefijado, mediante la aplicación de una función de transformación en escalón.

BIP. Formato de grabación de la imagen de bandas intercaladas por *pixel*, consistente en la disposición de una misma celda en todas las bandas sucesivamente, seguida de la siguiente celda de la misma línea de barrido en todas las bandas, y así sucesivamente hasta completar las líneas y la escena entera.

Bit. Abreviatura de *Binary Digit*, unidad de información binaria utilizada por los sistemas informáticos.

BRDF. Acrónimo de *Bidirectional Reflectance Distribution Function*, función de distribución de reflectancia bidireccional. Representa la proporción de radiación reflejada por una determinada superficie en cada dirección proyectada del rayo reflejado.

BSQ. Formato de grabación de la imagen de bandas secuenciales, consistente en la disposición de todas las celdillas de la imagen en una misma banda antes de pasar a la siguiente banda. El orden dentro de cada banda es el de completar líneas antes que columnas.

Byte. Grupo de 8 bit. Representa un carácter, y en teledetección suele ser la unidad para asignar el nivel digital a una celda, pudiendo oscilar éste entre 0 y 255.

Campo de entrenamiento. Conjunto de celdillas de la imagen utilizadas por el sistema para el cálculo de los límites de decisión en la clasificación, a partir de su vector de medias y de la matriz de covarianzas.

CCD. Acrónimo de *Charge-Coupled Device*, dispositivo generador de cargas proporcionales a la intensidad lumínica recibida, utilizado en los

exploradores de empuje y en las modernas cámaras de vídeo.

Clasificación no supervisada. Procedimiento de clasificación digital basado en el análisis *cluster*, mediante el cual el algoritmo de clasificación busca los agrupamientos *naturales* de los datos para establecer las correspondientes clases.

Clasificación supervisada. Procedimiento de clasificación digital mediante el cual se establecen una serie de campos de entrenamiento seleccionados por el analista en razón de su homogeneidad temática, en función de cuya similitud el sistema informático realiza la clasificación de la escena completa.

Clasificador. Operador que asigna el vector de características de una celda a una clase determinada, en función de ciertos criterios establecidos.

Clasificador de máxima probabilidad. Operador que utiliza el criterio de máxima probabilidad, basado en la teoría bayesiana de la decisión.

Clasificador de mínima distancia. Operador cuyo criterio de asignación es el de la distancia estadística entre el vector de características de la celda considerada y el vector de medias de cada clase, establecidas éstas a partir de los campos de entrenamiento.

Clasificador ISODATA. Operador que utiliza un procedimiento *clustering* no jerárquico.

Clasificador paralelepédico. Operador cuyo criterio de asignación es el establecimiento de unas áreas de dominio hipercúbicas alrededor del vector de medias de cada clase.

Cluster. Agrupamiento estadístico de datos en torno a características espectrales similares. Este agrupamiento es la consecuencia estadís-

tica de la existencia de distintas clases informacionales en la imagen.

Color compuesto. Proceso de formación de una imagen en color mediante la composición de tres bandas, a cada una de las cuales se les hace corresponder uno de los tonos primarios rojo, verde y azul, en una intensidad proporcional a la luminancia que muestre cada una.

Contraste. Razón existente entre la energía emitida o reflejada por un objeto y sus alrededores más inmediatos.

Convolución cúbica. Procedimiento de remuestreo consistente en asignar un nivel digital a la celda de la imagen geoméricamente corregida, a partir de la interpolación de los 16 *ND* más próximos de la imagen transformada mediante los polinomios de ajuste.

Corrección atmosférica. Procedimiento de ajuste del contraste para eliminar el efecto producido por la radiación difusa de la atmósfera.

Corrección geométrica. Transformación de la imagen consistente en el desplazamiento de las celdillas de su posición original para minimizar las distorsiones existentes en la misma, con el propósito de darle a aquella validez cartográfica.

Cuerpo negro. Es un emisor y receptor perfecto de energía electromagnética. Toda la radiación incidente es absorbida por él y a una determinada temperatura, la radiación que emite es la máxima posible, en virtud de la Ley de Stefan-Boltzmann.

Dispersión no selectiva. Dispersión de la radiación en el seno de la atmósfera producida por aerosoles, cuyo tamaño de coloides es suficientemente grande para afectar a todas las longitudes de onda. Es suma de efectos de reflexión, refracción y difracción.

Dispersión. Fenómeno por el cual la radiación electromagnética es desviada de su trayectoria original, como consecuencia de los choques entre los rayos luminosos y los aerosoles atmosféricos.

Distancia estadística. Es la existente entre los extremos de dos vectores de características en un espacio estadístico multivariante. Sirve para la medida de la similitud entre celdillas o entre celdillas y una determinada clase ya establecida.

Distorsión. Alteración de la geometría de la imagen en el proceso de adquisición, como consecuencia de los movimientos de oscilación accidentales de la plataforma, de la rotación terrestre, del tiempo de barrido, de la velocidad de basculación del espejo barredor, del alejamiento de la celda del nadir y de la curvatura terrestre. Alteración radiométrica debido al efecto atmosférico.

Dominio de la frecuencia. Espacio para la definición de una imagen, tratada ésta como una composición de ondas sinusoidales, generado a partir de los ejes amplitud, frecuencia y dirección.

Dominio espacial. Sistema de coordenadas de localización de los elementos de una imagen en forma vectorial o matricial.

Ecuación del histograma. Procedimiento de expansión del contraste consistente en dotar a las celdillas de un nuevo nivel digital proporcional a su frecuencia en la imagen primitiva.

Emisividad. Razón entre la energía emitida por un cuerpo y la que emitiría un cuerpo negro a la misma temperatura. La emisividad de los cuerpos naturales puede oscilar entre 0 y 1.

Emitancia. Densidad de flujo radiante saliente desde una superficie, denominada también excitancia. Se mide en vatios/m².

EOSAT. Acrónimo de *Earth Observation Satellite Company*, empresa privada contratada por el Gobierno USA para la comercialización de los datos y para el desarrollo de los futuros sistemas Landsat.

ESA. Acrónimo de *European Space Agency*, Agencia Espacial Europea, con sede en París.

Escena. Área terrestre grabada por un sistema de adquisición de imágenes.

Espectro electromagnético. Sucesión creciente de longitudes de onda de todas las radiaciones conocidas.

Expansión del contraste. Procedimiento por el cual se adecua el rango de luminancias de la imagen a la gama de grises que el sistema de visualización es capaz de proporcionar.

Exploradores de empuje. Sistemas formadores de imágenes cuyos sensores están dotados del suficiente número de detectores para poder explorar toda una línea de la imagen simultáneamente.

Falso color. Composición de tres bandas para la formación de una imagen en color compuesto, de tal modo que se establezca una relación de equivalencia entre banda espectral y tono primario asociado a ella que no corresponda a los intervalos espectrales naturales o al orden de éstos.

Fiabilidad. Razón existente entre el número de celdillas correctamente clasificadas y el total de celdillas muestreadas en un proceso de verificación de los resultados de una clasificación digital.

Filtro de media. Tipo de filtro de paso bajo que asigna un nuevo *ND* a la celda, resultado de calcular la media aritmética de los nueve *ND* originales centrados sobre aquélla.

Filtro de mediana. Tipo de filtro de paso bajo no

lineal que asigna un nuevo *ND* a la celda, resultado de calcular la mediana de los nueve *ND* originales centrados sobre aquélla.

Filtro. Operador local mediante el cual se obtiene una nueva imagen cuyos *ND* son función de los *ND* de las celdillas contiguas en la imagen inicial. Se utilizan para la eliminación de ruidos y para el realce de los rasgos lineales de una imagen.

Filtro de paso alto. Filtros destinados a reforzar los componentes de alta frecuencia de una imagen digital. El resultado de ello es el énfasis de los contrastes entre áreas adyacentes, por cuya razón tienden a realzar los rasgos lineales de la imagen.

Filtro de paso bajo. Filtros destinados al suavizado de la imagen mediante la retención de los componentes de alta frecuencia de ésta.

Filtro direccional. Filtro de paso alto que prima una determinada dirección, e incluso un sentido concreto.

Filtro en el dominio de la frecuencia. Producto de la función de filtrado por la función imagen, definida ésta en el dominio de la frecuencia.

Formato ráster. Forma de tratamiento y representación espacial de las entidades mediante la disposición de celdas o *píxeles* en forma de matriz numérica o red de *ND*.

Formato vectorial. Forma de tratamiento y representación espacial de las entidades mediante la asignación de un par de datos para cada punto, cuyo valor es el de sus propias coordenadas cartesianas. Cuando la entidad sea puntual solamente constará de un par de datos. Si se trata de una entidad lineal requerirá dos pares de valores por cada segmento. Por último, si la entidad es zonal, se representará una polilínea cuyo fin deberá coincidir con el principio de esta.

FOV. Acrónimo de *Field of View*, campo de visión de un instrumento sensor o ángulo de apertura. Se trata de la sección recta del ángulo sólido formado por el punto de vista (situación del sensor) y la escena completa observada.

Geoestacionario. Condición de un satélite cuya velocidad angular en la descripción de su órbita es coincidente con la de la Tierra. En consecuencia, el vector de posición del satélite cortará a la superficie del geoide siempre en el mismo punto.

Georreferenciación. Procedimiento mediante el cual se dota de validez cartográfica a una imagen digital corrigiendo geoméricamente la posición de las celdillas y atribuyéndoles coordenadas en algún sistema de referencia.

Histograma. Representación de la función de densidad de probabilidad de la luminancia de una imagen en forma de gráfico de barras.

IFOV. Acrónimo de *Instantaneous Field of View*, campo de visión instantáneo. Se trata de la sección recta del ángulo sólido que un sistema sensor es capaz de observar en un instante determinado, sin tener en cuenta la propia capacidad de barrido.

IHS. Acrónimo de *Intensity-Hue-Saturation*, sistema de especificación del color por sus coordenadas tridimensionales intensidad (valor)-tono-saturación. En teledetección se aplica esta denominación a la transformación consistente en la extracción de las tres nuevas bandas I, H y S a partir de las bandas correspondientes al rojo, al verde y al azul, con el objeto de paliar la frecuente escasez de contraste en una imagen en color natural compuesto. También se emplea esta transformación para componer imágenes por fusión de la información procedente de sensores de diferente resolución espacial.

Imagen analógica. Función continua de la luminancia de una determinada superficie.

Imagen digital. Función discretizada de la imagen analógica, tanto en sentido espacial, mediante la generación de celdillas, como en sentido radiométrico, mediante la asignación a éstas de valores enteros comprendidos entre 0 y 255 ó entre 0 y 1.023, según la resolución radiométrica del sistema.

Intensidad. Luminancia de una celda, nivel de gris de la misma. Hace referencia este concepto al grado de claridad de la celda, correspondiendo a la variable visual valor.

Interpolación bilineal. Procedimiento de remuestreo consistente en asignar un nivel digital a la celda de la imagen geoméricamente corregida, a partir de la doble interpolación lineal de los 4 *ND* más próximos de la imagen transformada mediante los polinomios de ajuste.

Irradiancia. Densidad de flujo radiante que penetra en una superficie. Se mide en vatios/m².

Láser. Acrónimo de *Light Artificially Stimulated Electromagnetic Radiation*, haz de radiación coherente con una longitud de onda única.

Ley de desplazamiento de Wien. Ecuación que gobierna el desplazamiento hacia valores más pequeños de la longitud de onda para la que se obtiene el máximo de radiación que emite un cuerpo negro, en función de las temperaturas crecientes.

Ley de Planck. Ecuación que pone en relación la energía emitida por un cuerpo en una determinada longitud de onda, en función de su temperatura absoluta.

LíDAR. Acrónimo de *Light Intensity Detection and Ranging*. Se trata de un sistema de pulsos láser, similar al radar que en las bandas del ul-

travioleta, el visible o el infrarrojo próximo, mediante el que, midiendo el retardo del pulso de retorno, se puede conocer la distancia al objetivo.

Longitud de onda. Distancia entre dos nodos o dos valles consecutivos de una onda. En el caso de la radiación electromagnética, es el recíproco de la frecuencia de dicha radiación multiplicada por la velocidad de la luz. De su longitud de onda derivan la mayor parte de las propiedades de la radiación electromagnética, y en base a intervalos de la cual pueden establecerse bandas espectrales con comportamientos electromagnéticos similares.

Luminancia. Intensidad luminosa de una imagen. En la imagen digital, la luminancia se refiere al nivel de gris o nivel digital (*ND*) de cada una de sus celdillas.

Matriz de confusión. Tabla de contingencia donde se muestran los resultados del muestreo de control de la calidad de una clasificación automática.

METEOSAT. Satélite meteorológico geoestacionario de la Agencia Espacial Europea, situado en órbita ecuatorial y a longitud 0°.

Mie. Tipo de dispersión atmosférica causado por partículas de tamaño similar a la longitud de onda considerada.

MSS. Acrónimo de *MultiSpectral Scanner*, sensor a bordo del satélite Landsat.

Nadir. Intersección de la vertical que pasa por el centro del sistema óptico-electrónico de adquisición con el terreno. Coincide con la intersección de la prolongación del eje óptico del sistema sensor de una plataforma con la superficie terrestre cuando la imagen es absolutamente vertical.

Nd. Abreviatura de nivel digital. Se trata del valor

numérico discreto asignado por el sistema formador de imágenes a cada celda en respuesta a la intensidad radiométrica recibida por el sensor. Se le conoce asimismo como nivel de gris, luminancia, número digital, valor de pixel, etc.

NDVI. Acrónimo de *Normalized Difference Vegetation Index*, índice de vegetación de diferencia normalizada.

NOAA. Acrónimo de *National Oceanic and Atmospheric Administration*, estamento USA encargado de la gestión y predicción atmosférica y oceanográfica. Es el responsable de los satélites de la serie TIROS-NOAA, de los cuales está en órbita el número catorce, en funcionamiento coordinado con los actualmente en servicio.

Número de onda. Inverso de la longitud de onda.

Órbita polar. Órbita que pasa por la vertical de los polos, permitiendo con ello al satélite pasar por encima de la mayor parte de la superficie terrestre, apenas su período orbital no esté sincronizado con el de revolución de la Tierra.

Píxel. Abreviatura de *Picture Element*, cada uno de los elementos que componen una imagen dispuestos matricialmente en filas y columnas.

Polinomios de ajuste. Funciones de transformación de coordenadas aplicables a todas las celdillas de una imagen, cuyos coeficientes se obtienen mediante el ajuste mínimo-cuadrático de las desviaciones de las coordenadas corregidas de una serie de puntos de control con respecto a las coordenadas de éstos en la imagen.

Proceso aditivo de formación del color. Fenómeno de formación del color mediante la suma de radiaciones electromagnéticas de bandas más estrechas correspondientes a cada uno de los tonos primarios. Según la teoría del color, es el proceso seguido en la formación del color

por transparencia o adición de haces cromáticos. Así, la suma de rojo más verde proporcionará un amarillo; la suma de rojo más azul, un magenta; la suma de verde más azul, un cian; y la suma de los tres simultáneamente, el blanco. Por el contrario, el proceso sustractivo es el realizado mediante mezcla de tintas. Una tinta cian más otra magenta proporcionará un color azul; una magenta más otra amarillo, un rojo; una tinta cian más otra amarillo, un verde; y la suma de las tres tintas simultáneamente, un color negro, que significa la absorción completa de la radiación incidente.

PSF. Acrónimo de *Point Spread Function*, función de dispersión puntual, imagen de una fuente puntual obtenida por un sistema formador de imágenes. Es una medida de la fidelidad espacial del sistema.

Puntos de control. Son una serie de puntos utilizados en la búsqueda de las funciones de transformación que permitan la corrección de las distorsiones de una imagen.

Radar. Abreviatura de *Radio Detection and Ranging*, sistema que utiliza pulsos de energía electromagnética de longitud de onda comprendida entre 1 mm y 1 m producida artificialmente, para localizar objetos mediante la detección de la radiación reflejada.

Radar altímetro. Dispositivo no formador de imágenes que utiliza el sistema de radar para medir la distancia vertical desde la plataforma portadora hasta la superficie terrestre, obteniéndose precisiones de centímetros.

Radiancia. Es la cantidad de energía radiada desde un objeto en la unidad de tiempo por cada unidad de ángulo sólido y de unidad de superficie del objeto perpendicular a la dirección de propagación. Es el parámetro físico medible con un radiómetro.

Radiometría. Es el conjunto de medidas efectuadas en campo con radiómetros, encaminadas a contrastar *in situ* la información adquirida mediante sensores situados en satélites.

Rayleigh. Tipo de dispersión atmosférica producida por partículas cuyo tamaño es pequeño en comparación con la longitud de onda de la radiación.

Realce de bordes. Operación de filtrado de paso alto mediante la cual se enfatizan los contrastes entre áreas contiguas, resultando una mejor definición de las fronteras existentes en la imagen.

Redundancia. Reiteración de la información proporcionada por dos o más bandas. Es consecuencia de la correlación que puede y suele existir entre las distintas bandas pertenecientes a zonas espectrales cercanas.

Reflectancia. Medida de la capacidad de una superficie para reflejar energía electromagnética en una determinada longitud de onda. Es la razón existente entre el flujo reflejado y el incidente sobre dicha superficie. Es sinónimo de albedo.

Remuestreo. Procedimiento de reubicación de los *ND* de una imagen que ha sido corregida espacialmente para adaptarlos a posiciones discretas enteras, que son las que el sistema informático puede procesar y presentar a su vez en pantalla.

Resolución. Capacidad de un sistema sensor para distinguir información de detalle en un objeto. En teledetección se habla de distintos tipos de resolución: espacial, espectral, radiométrica y temporal.

RGB. Acrónimo de rojo-verde-azul. Se trata de un sistema de especificación del color basado en la propiedad aditiva de los tonos primarios que

es el comúnmente utilizado en los sistemas informáticos y en la composición en verdadero o en falso color de las imágenes espaciales.

Ruido. Todos aquellos errores aleatorios introducidos en la imagen durante el proceso de adquisición, transmisión o grabación de esta. Producen una alteración de los niveles digitales que corresponderían a la escena real. El proceso de análisis de la señal debe ser capaz de distinguir y separar las alteraciones que son consecuencia del ruido sin eliminar con ello información sustancial de la imagen.

SAR. Acrónimo de *Synthetic Aperture Radar*, radar de apertura sintética.

Saturación. Proporción de tono *puro* que tiene una determinada tinta o color. Una saturación nula indicaría un gris del mismo valor que el color considerado. Una saturación igual a la unidad representaría el mencionado color completamente *puro*.

Scanner de línea. Sistema formador de imágenes mediante el barrido secuencial de líneas sucesivas, gracias a la oscilación de un espejo que envía a los detectores las celdillas barridas, cuyos tamaños son determinados por el IFOV del sistema.

Sensores activos. Aquellos que emiten energía electromagnética generada artificialmente en la plataforma, la cual será después detectada a bordo tras sufrir la reflexión en la superficie objeto de estudio.

Sensores pasivos. Aquellos otros que solamente registran la energía emitida por la superficie estudiada o la que, procedente del sol, es reflejada por ésta.

Signatura espectral. Forma característica del espectro de emisión/reflexión de una determinada superficie. Como cada material posee bandas de absorción diferentes de acuerdo a

su composición química, la energía solar reflejada será la complementaria a la incidente más la virtualmente emitida por aquélla. En consecuencia, el espectro electromagnético de un objeto detectado a bordo de un sistema sensor será una señal prácticamente inequívoca de la composición del mismo. La única restricción que se impone a la identificación de superficies mediante la teledetección deriva de la propia resolución espectral del sensor, el cual no es capaz de proporcionar un espectro continuo de la superficie observada sino a lo sumo unas cuantas bandas, dentro de cuyos intervalos espectrales no se hace distinción alguna.

Sistema formador de imágenes. Sistema sensor con el que van equipados las plataformas de teledetección en el que la entrada al mismo es la energía electromagnética procedente de la superficie observada y la salida la irradiancia de la imagen, esto es, el conjunto de *ND* conferidos a las celdillas integrantes de la imagen.

SLAR. Acrónimo de *side looking airborne radar*, radar lateral de aerotransportado.

SPOT. Acrónimo de *Système Probatoire d'Observation de la Terre*, satélite francés portador de dos instrumentos AVHRR.

Tasseled Cap. Transformación global de una imagen consistente en la extracción de tres nuevas bandas a partir de las originales, las cuales están relacionadas con determinados fenómenos fisiológicos de la materia vegetal.

Temperatura absoluta. Temperatura medida en grados Kelvin (K), cuya escala comienza el cero absoluto (-273 °C).

Textura. Frecuencia de cambio y disposición que tienen los niveles de gris en una imagen espacial o fotográfica.

TM. Acrónimo de *Thematic Mapper*, sensor de alta resolución espectral a bordo del satélite

Landsat.

Tono. Percepción diferencial por parte del ojo humano de las distintas longitudes de onda del espectro visible. Es lo que coloquialmente se designa por *color*.

Transmitancia. Medida de la capacidad de un material para transmitir energía electromagnética en una determinada longitud de onda hacia otras capas contiguas. Es la razón existente entre el flujo transmitido y el incidente sobre la superficie del material transmisor.

Valor. Grado de claridad u oscuridad de una imagen en general, o de una celda en particular.

Varianza. Medida de la dispersión de los valores de un conjunto de éstos a partir de una magnitud promedio. En teledetección se aplica este término a la dispersión de los niveles digitales de la imagen con respecto al *ND* medio. Por tanto se trata de una medida del contraste y, en consecuencia, de la cantidad de información que proporciona dicha imagen.

Vecino más próximo. Se aplica esta expresión a dos conceptos bien diferentes. Uno de ellos es uno de los métodos de remuestreo existentes, que consiste en asignar a una celda en la imagen corregida el *ND* que tiene más próximo en la imagen transformada, después de la aplicación de las funciones de ajuste. El otro se refiere a uno de los métodos de clasificación, en el que el clasificador adopta el criterio de vecindad estadística o de mínima distancia.

Ventana atmosférica. Cada uno de los intervalos espectrales a los que la atmósfera resulta diáfana.

WWW. Acrónimo de *World Weather Watch*, sistema de observación y predicción meteorológica mundial.