

Curso de Geoestadística

Posgrado UNAM - IMP

Tarea 3: Estimación Espacial

**ESTIMACIÓN ESPACIAL DE DATOS
DE POZOS DE AGUA
VARIABLES: LITIO y FLUOR**

México D.F. 13/12/2005.

CONTENIDO:

1. INTRODUCCIÓN

2. ANÁLISIS EXPLORATORIO DE DATOS

3. ANÁLISIS ESTRUCTURAL

3.1. LITIO

3.2. FLUORURO

3.3. RESIDUOS

4. ESTIMACIÓN CON KRIGING

4.1. LITIO

4.2. FLUORURO

5. ESTIMACIÓN CON COKRIGING PARA LA VARIABLE LI

6. CONCLUSIONES

1. INTRODUCCIÓN

Los datos de Fluoruro (F) y de Litio (Li) fueron colectados en 50 pozos de agua potable y tres manantiales de las baterías que suministran agua a la Ciudad de León Guanajuato, localizados en un área de 40 km x 40 km (figura 1). En total son 53 pares de valores, los cuales se encuentran listados en la Tabla 1. Las concentraciones de Li y F están dadas en ppm y las coordenadas en UTM.

Seguidamente se presenta un resumen de los análisis exploratorios y estructurales de ambas variables. Se exponen los pasos seguidos para ajustar el modelo de variograma propuesto, el cual es usado para efectuar las estimaciones por Kriging. Posteriormente se estima por CoKriging la variable LI. Finalmente se realiza el análisis de los resultados de la estimación con Kriging y Cokriging para dicha variable, en función de los mapas del valor estimado y la varianza del error de dicha estimación.

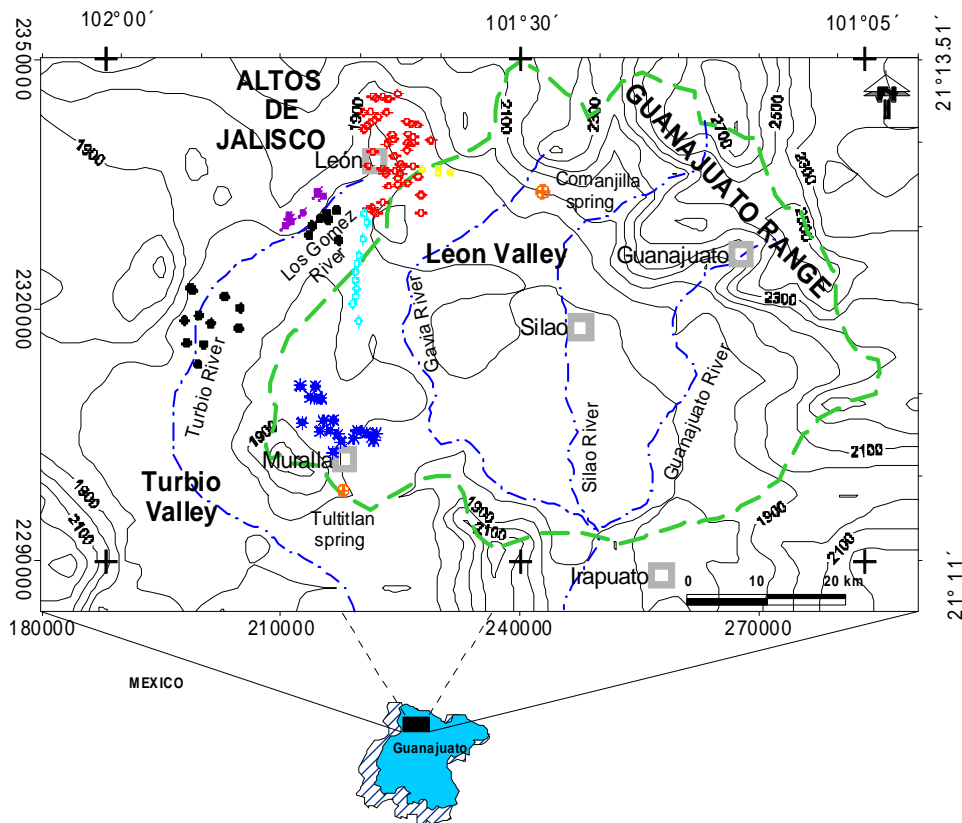


Figura 1. Distribución geográfica de baterías de pozos que suministran agua potable a la ciudad de León, Guanajuato. En Azul, la batería Muralla; en negro, la batería Turbio; en azul claro, la batería Sur; en violeta la batería Poniente y en amarillo, la batería Oriente.

X	Y	LI	F
UTM	UTM	(0.01 mg/L)	(0.1 mg/L)
216827	2332211	0.030	0.17
215789	2330968	0.088	0.45
215590	2331713	0.077	0.41
204944	2320603	0.112	0.94
204763	2318212	0.140	1.16
201286	2318812	0.103	0.71
199674	2313943	0.103	0.75
198325	2316422	0.153	0.91
199808	2319701	0.149	0.91
195500	2323100	0.069	1.27
198787	2323017	0.148	2.2
216032	2306264	0.222	1.45
216369	2303620	0.098	0.8
221304	2303100	0.154	0.99
221023	2306618	0.140	0.96
221503	2304997	0.147	1.09
219395	2306122	0.135	1.02
216448	2307422	0.230	1.38
212606	2307118	0.064	0.6
214984	2309982	0.180	1.41
218842	2321068	0.042	0.26
220134	2328685	0.035	0.21
220272	2331758	0.030	0.22
221690	2331808	0.017	0.1
221335	2332434	0.024	0.17
220829	2332390	0.016	0.08
222917	2336408	0.014	0.39
221621	2342744	0.019	0.31
224419	2345857	0.146	0.68
222534	2345432	0.079	0.39
221697	2345353	0.023	0.41
220279	2341649	0.038	0.43
218089	2338213	0.036	0.65
214566	2334357	0.026	0.48
214412	2333960	0.040	0.75
212690	2331786	0.029	0.71
211318	2330717	0.098	0.13
210754	2330686	0.171	0.57
215152	2333703	0.029	0.16
225531	2342541	0.023	0.24
224851	2344503	0.039	0.32
220791	2337335	0.044	0.34
221341	2338969	0.045	0.26
223774	2340944	0.058	0.23
228533	2340371	0.040	0
223937	2338460	0.019	0.22
226321	2337255	0.029	0.1
227467	2336873	0.030	0.03
230988	2336411	0.023	0.13
227255	2331810	0.015	0.07
215194	2340128	0.022	0.71
243068	2333454	1.077	9.8
228130	2345234	0.018	0.38

Tabla 1. Valores y coordenadas de datos de LI y F.

1. ANÁLISIS EXPLORATORIO DE DATOS

El AED implica realizar el examen gráfico cuantitativo y cualitativo de las variables estudiadas. Para ello se incluyen en este apartado los análisis de distribución espacial de los datos, normalidad y de valores atípicos (distribucionales y/o espaciales). A continuación se presentan los resultados para cada uno de los conjuntos de datos.

El histograma de frecuencia para la variable LI mostró una distribución leptocurtica de simetría positiva. Se observa un valor atípico. Al costado derecho de la figura 2 se presentan los valores estadísticos, donde resalta que la mediana es menor que el valor medio y éste a su vez es menor que la desviación estándar de la muestra, que es en realidad muy alta. La figura 3 muestra la distribución espacial de los datos LI. Casualmente el valor atípico se encuentra lago alejado del resto, lo cual podría implicar la expresión de un fenómeno distinto al manifestado por el común de los datos.

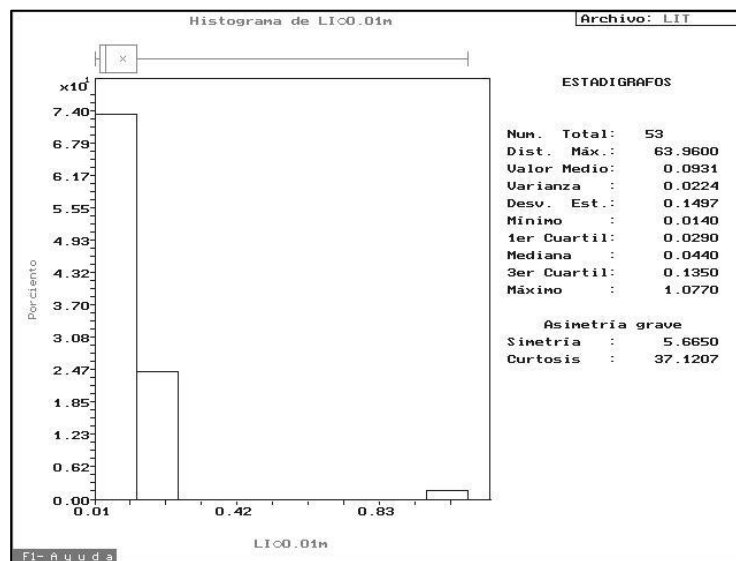


Figura 2. Histograma de frecuencia y estadígrafos de datos de Litio.

Por otra parte, el comportamiento de la variable F resultó similar al de LI. El histograma de frecuencia muestra una distribución leptocurtica con asimetría positiva. Similar al caso anterior, se observa un valor atípico principal (de 2 identificados), el cual resalta aún más en el diagrama Box Plot, pues está considerablemente alejado del intervalo intercuartil.

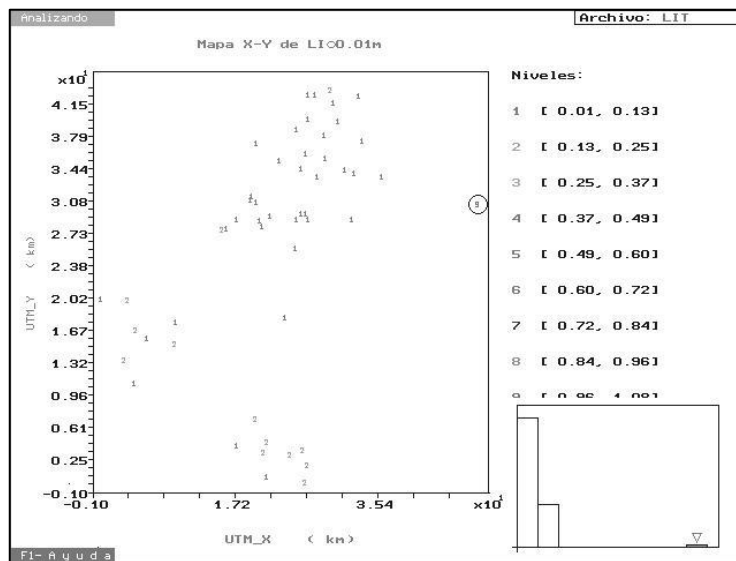


Figura 3. Distribución espacial de la variable LI. Se destaca la ubicación del valor atípico separado de los 3 subgrupos de datos.

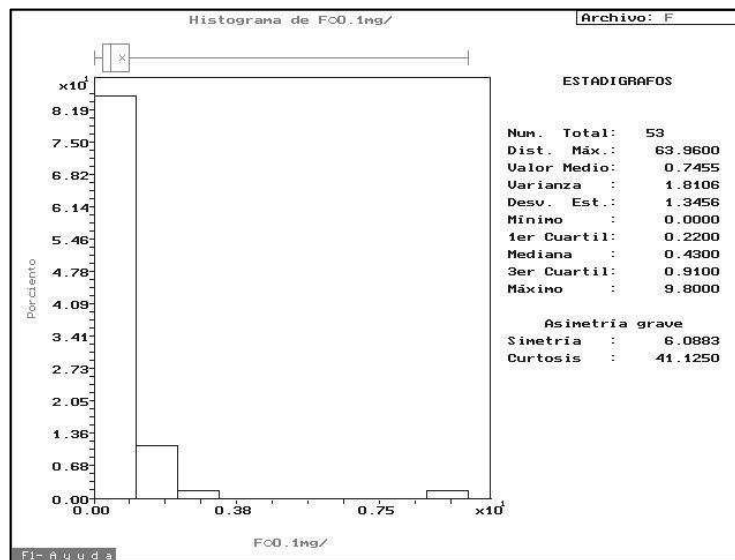


Figura 4. Histograma de frecuencia y estadígrafos de datos de Fluor.

Como se observa en los valores estadísticos, la mediana es casi la mitad del valor medio y este a su vez lo es de la desviación estándar de la muestra de F. Lo cual resulta de la marcada influencia generada por el mencionado valor anómalo. Luego, en la figura 5, se observa la distribución espacial de los datos, y al igual que en el caso de la variable LI, resalta el valor atípico principal aislado de los restantes subgrupos de la muestra.

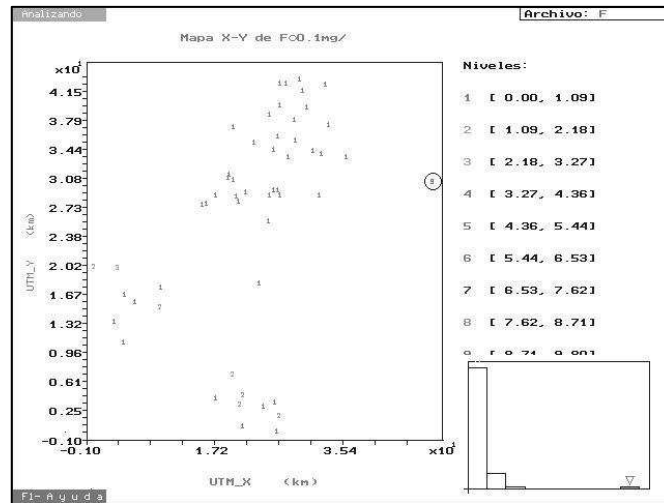


Figura 5. Distribución espacial de la variable F. Se destaca la ubicación del valor atípico separado de los 3 subgrupos de datos.

2. ANÁLISIS ESTRUCTURAL

Para tener una idea general de la estacionaridad del par de variables estudiadas se procedió a estimar el variograma adireccional. Para ello se incluyeron todos los datos de F y LI. El principal objetivo es apreciar el comportamiento de dicho gráfico para establecer intuitivamente la posible existencia de tendencia.

3.1. LITIO

Debido a las características del valor atípico presente en el grupo de datos LI, se decidió estimar el variograma si incluirlo. La figura 6 muestra el resultado, el cual se realizó aplicando los valores reportados por el programa estadístico GEOESTAD, a saber: Distancia máx=55.5km, lag=2.7km, dirección=0°, tolerancia=±90°. Como era de esperarse, se observa una notable reducción de la varianza y del valor medio, sin embargo, se manifiesta un rápido incremento de los estimados que hacen suponer la posible existencia de tendencia. En este caso el ajuste automático del programa identifica al modelo lineal del variograma como el de mejor AIC. La figura 7 muestra el variograma estimado en las cuatro direcciones solicitadas (0°, 45°, 90° y 135°). Se tienen 52 datos con una distribución espacial muy heterogénea representada por tres subconjuntos principales, factor que afecta fuertemente la estimación de los variogramas orientados.

Además, como ya se comprobó en el apartado anterior, el valor atípico principal también genera una influencia decisiva en estos estimados. De manera que en esta oportunidad no existen suficientes argumentos para soportar la hipótesis de anisotropía.

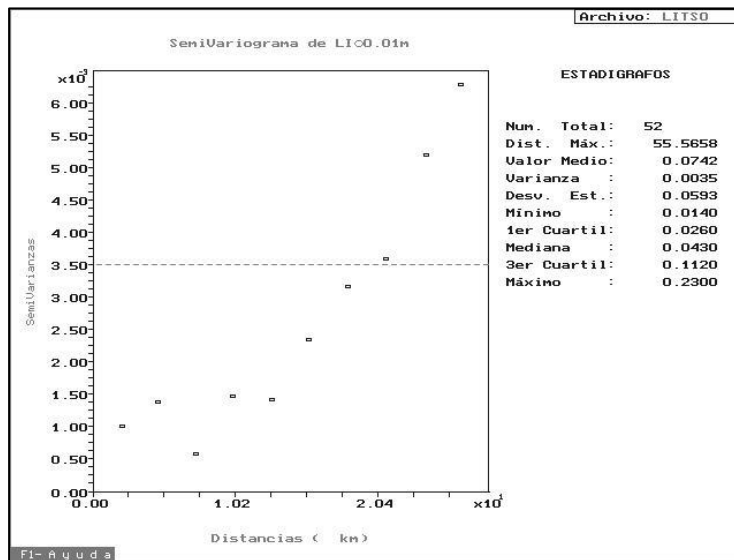


Figura 6. Variograma adireccional de datos LI sin outlier.

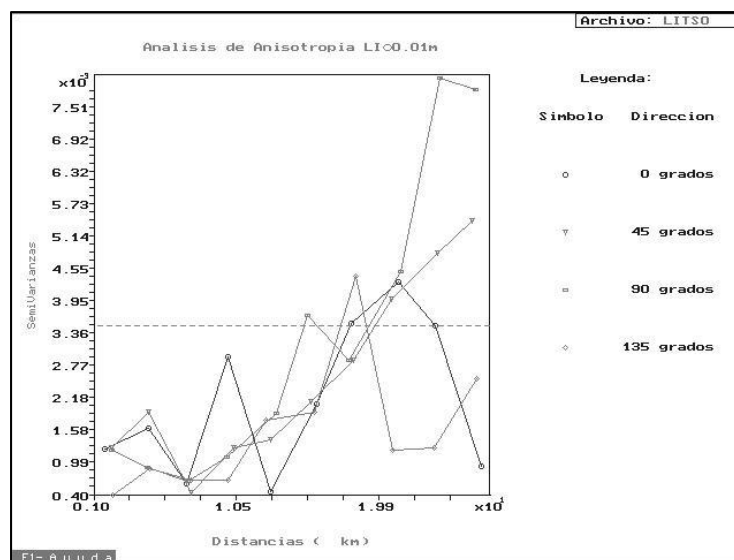


Figura 7. Variograma direccional de datos LI sin valor atípico.

Adicionalmente, la figura 8 muestra las proyecciones en los ejes X e Y de la variable LI sin outlier. En ambas se manifiesta una disposición "inclinada" de los datos con pendiente negativa, lo cual podría estar influyendo también en la estimación del variograma.

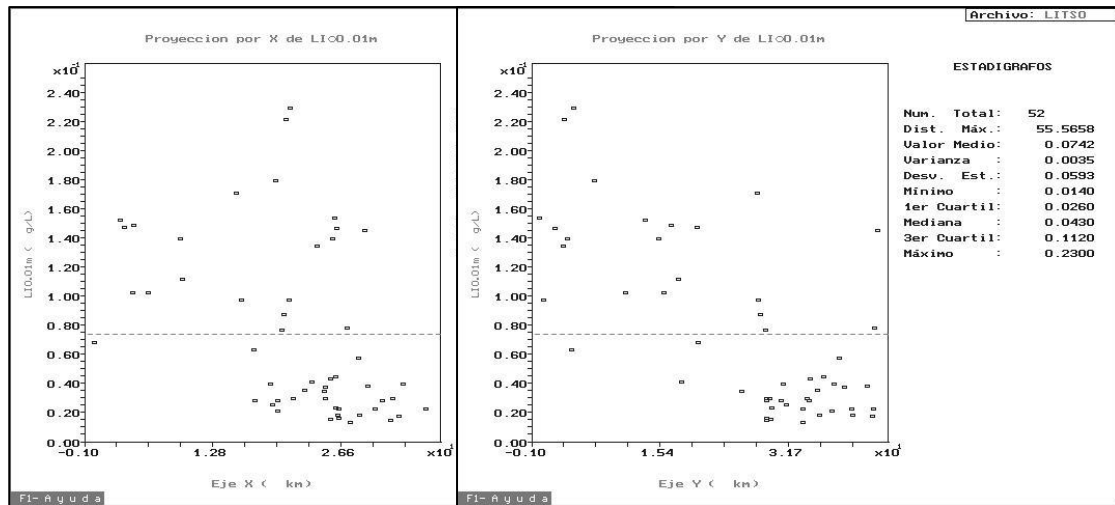


Figura 8. Proyecciones en ejes X e Y de datos LI sin valor atípico.

3.2. FLUORURO

El análisis de estacionaridad se hizo siguiendo el mismo procedimiento empleado para la variable LI. La figura 9 muestra el variograma estimado sin incluir el valor atípico principal. Igual que en caso anterior, se realizó aplicando los valores reportados por el programa estadístico GEOESTAD, a saber: Distancia máx.=55.5km, lag=2.7km, dirección=0°, tolerancia=±90°. Como se observa, el variograma crece rápidamente, no acotado por la varianza, describiendo una forma semihiperbólica, decayendo hacia el último punto estimado.

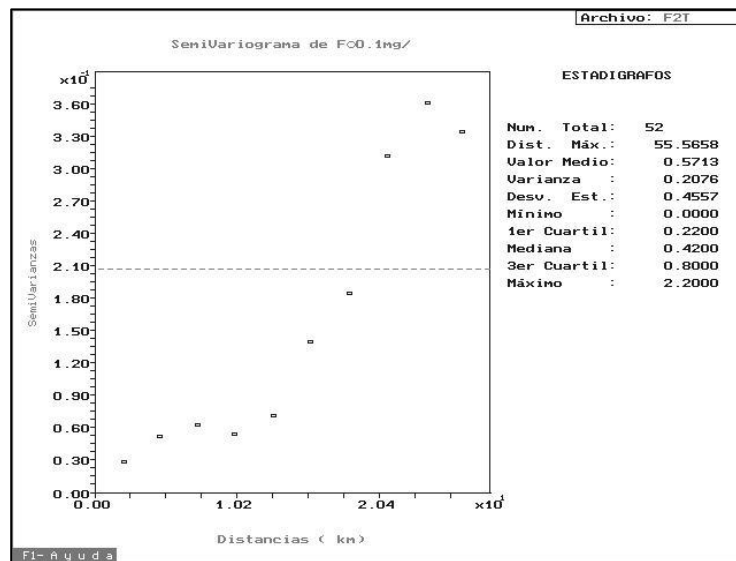


Figura 9. Variograma adireccional de datos F sin outlier.

Por otra parte, el variograma estimado en las cuatro direcciones solicitadas (0°, 45°, 90° y 135°) se muestra en la figura 10. Se observa un mejor comportamiento excepto para la dirección de 0°. Sin embargo, los comentarios en torno a la anisotropía son igualmente aplicables en ésta oportunidad. Finalmente, la figura 11 muestra los gráficos correspondientes a las proyecciones en los ejes X e Y de la variable F si el valor atípico. Análogo al caso de LI, se manifiesta una disposición "inclinada" de los datos con pendiente negativa, lo cual podría estar influyendo también en la estimación del variograma.

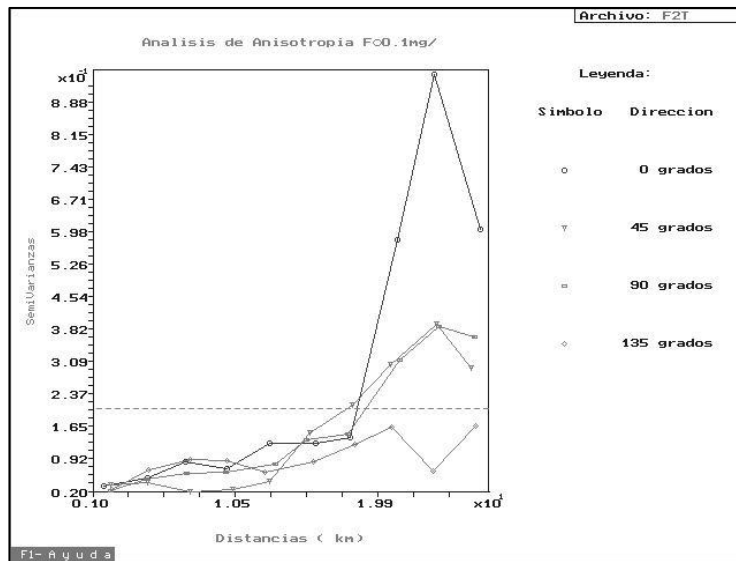


Figura 10. Variograma direccional de datos F sin valor atípico.

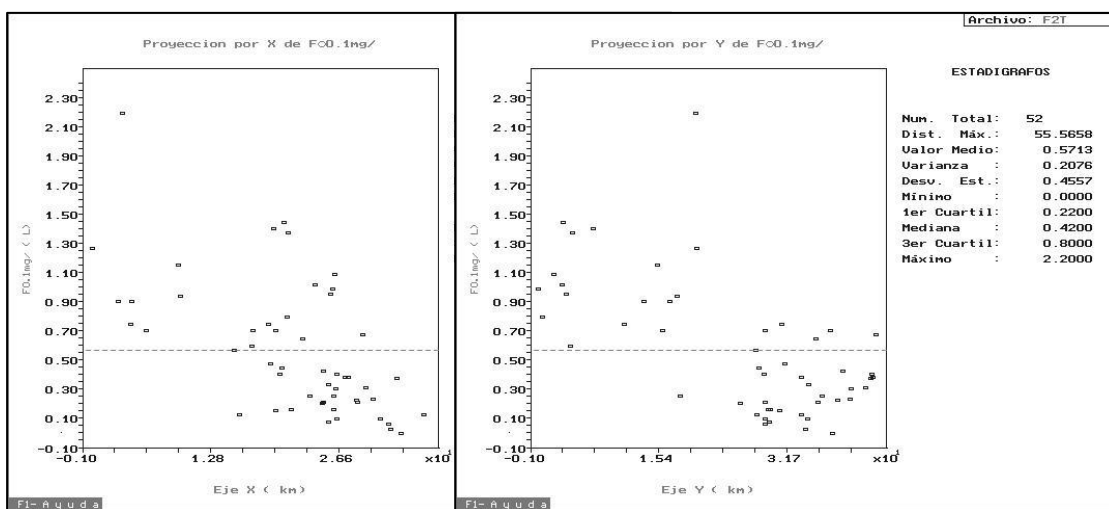


Figura 11. Proyecciones en ejes X e Y de datos F sin valor atípico.

Los modelos de variogramas para ambas variables (F y LI) no son totalmente satisfactorios, pues no describen completamente los fenómenos expresados por los mismos. Recordemos que una vez eliminados los valores atípicos, los ajustes lineales propuestos por el programa comportamientos anómalos (parabólicos), y no acotados por la varianza. Lo anterior nos permite concluir que es necesario estimar los residuos para establecer la posibilidad de expresar dichos fenómenos de la forma: $Z(x)=m(x)+R(x)$.

3.3 RESIDUOS

Los 52 datos de LI se ajustaron a un polinomio de grado 1. Cuyos coeficientes resultaron: 1-) 1.7336362614E-01, 2-) 1.1959557835E-03 y 3-) 2.8832165125E-03. La figura 12 muestra el gráfico de distribución de los residuos correspondientes al ajuste, con el respectivo histograma de frecuencias y estadígrafos. La mediana está muy cerca del valor medio, y ambos datos están próximos a cero, que es lo ideal. Con estos datos se procedió a estimar el variograma, mostrado en la figura 13 (izquierda). A pesar de que el modelo de variograma generado a partir del ajuste visual no se encuentra acotado por la varianza de los residuos, resulta bastante aceptable y aparentemente suficiente para generar un estimado del fenómeno expresado por la variable LI de la forma $Z(x)=m(x)+R(x)$.

Igual que para el caso de LI, los 52 datos de F se ajustaron a un polinomio de grado 1. Se realizó la estimación del variograma de los residuos y se obtuvo que aún era bastante inestable debido precisamente a la presencia del segundo valor atípico, el cual se eliminó también de los datos. Los nuevos coeficientes resultantes fueron: 1-) 1.2988220649E+00, 2-) 2 -1.4112985050E-02 y 3-) -1.7604455016E-02. La figura 14 muestra el gráfico de distribución de estos residuos. Es evidente que esta distribución posee un comportamiento normal. El variograma de estos residuos se presenta en la figura 15 (izquierda). A través de la prueba y error, se mejoró ligeramente el ajuste automático ofrecido por el programa. Las características finales del modelo de variograma se aprecian en la figura 15 (derecha). Este ajuste resulta más aceptable y aparentemente suficiente para generar un estimado del fenómeno expresado por la variable F de la forma $Z(x)=m(x)+R(x)$. Sólo resta realizar la respectiva validación cruzada y el análisis de las diferencias ($Z-Z^*$).

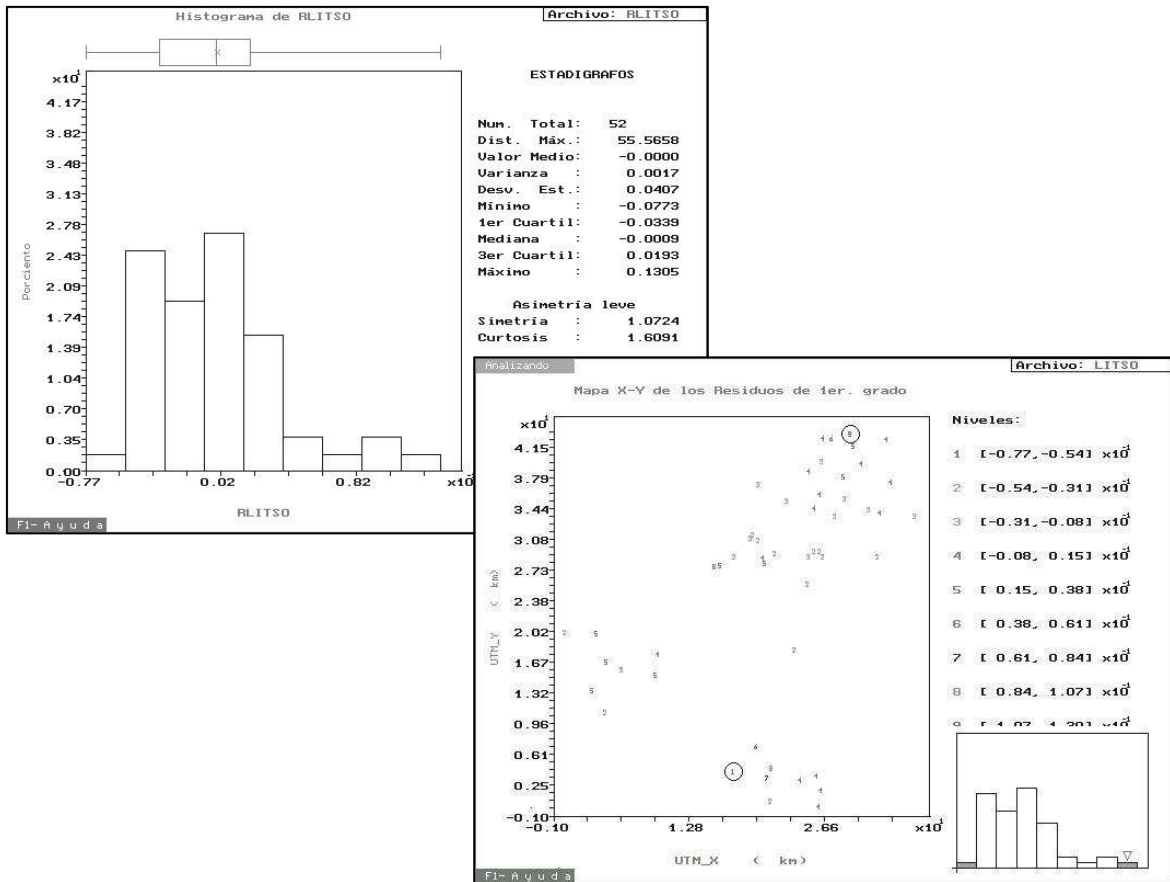


Figura 12. Distribucion espacial, Histograma y estadgrafos de datos LI (si valor atpico)

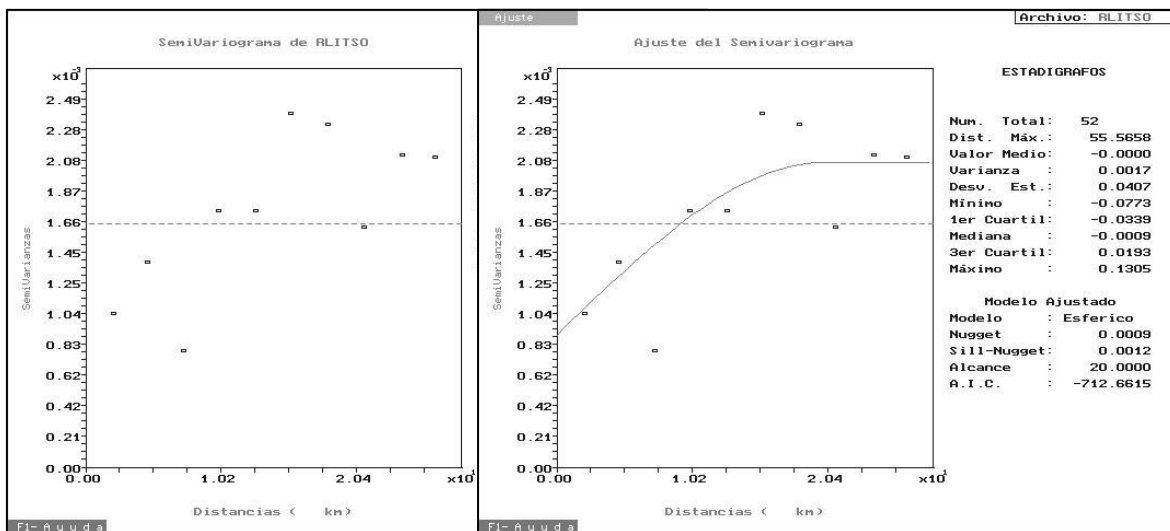


Figura 13. Variograma y modelo ajustado (prueba y error) de los residuos de 1er grado de la variable LI sin valor atpico.

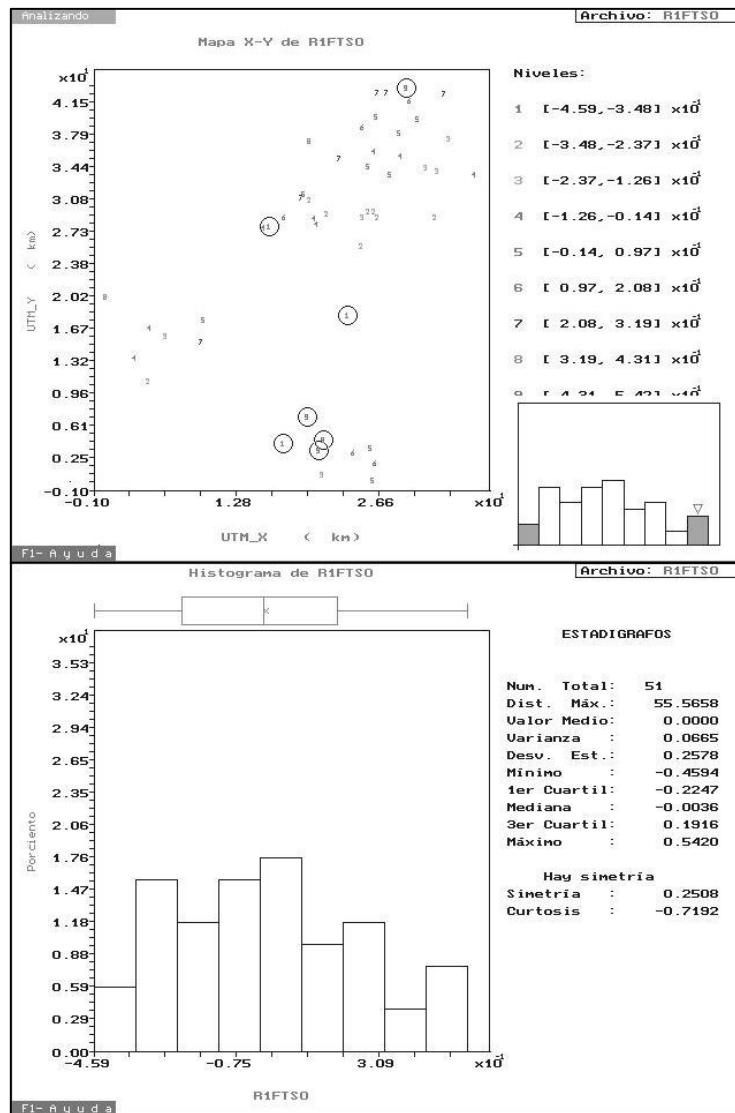


Figura 14. Distribuci3n espacial, Histograma y estadıgrafos de datos F (si valores atıpicos)

Para realizar la validaci3n cruzada se utiliz3 el m3todo de "dejar uno fuera" propuesto en la "receta practica del analisis estructural". Se incluyeron los modelos ajustados de variogramas de cada variable. Para LI result3: **Valor Medio de Z-Z*=-4.97E-03 / Varianza de Z-Z*=1.49E-03**. Mientras que para F result3: **Valor Medio de Z-Z*=-5.92E-04 / Varianza de Z-Z*=5.07E-02**. En ambos casos, los valores medios de las diferencias estan pr3ximos a cero y poseen varianzas pequenas, con lo cual resulta aceptable la validaci3n cruzada. Seguidamente pasaremos a la estimaci3n con Kriging.

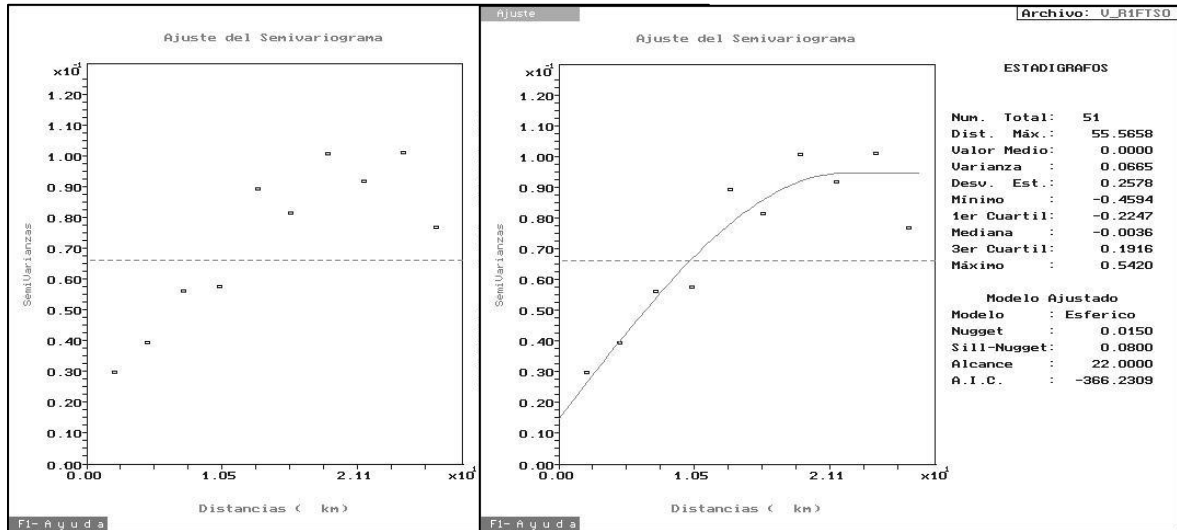


Figura 15. Variograma y modelo ajustado de los residuos de 1er grado de la variable F sin valores atípicos.

4. ESTIMACIÓN CON KRIGING

Los datos manejados en el presente estudio evidencian la presencia de tendencia. Luego de ajustar las funciones aleatorias a un polinomio de primer grado, se obtuvo un modelo de variograma para los residuos de cada variable. De esta manera se cumplen los requisitos necesarios para realizar la estimación espacial con *kriging Universal*.

El kriging universal asume que la función aleatoria Z se puede descomponer en la forma:

$$Z(u) = R(u) + m(u)$$

Donde $R(u)$ es la función aleatoria estacionaria de los residuos de $Z(u)$, tal que $E(R(u))=0$ y $m(u)$ es la función no aleatoria dependiente de la localización u . Los coeficientes que definen a la función $m(u)$ fueron estimados en la sección anterior, y poseen la siguiente forma: $m_0+m_1x+m_2y$. Dicho esto, a continuación se muestran los resultados obtenidos para cada variable.

4.1. LITIO

El kriging para los residuos de LI se hizo conforme a las siguientes especificaciones: puntual de 8 vecinos, con radio de búsqueda igual al alcance del variograma. Recordemos que el modelo de variograma es esférico, con efecto Nugget=0,0009 y S+N=0,0021. Se

definió una malla regular de 0.65 Km x 0.65 Km, para un total de 65 espacios en el eje Y y 54 en el eje X. A continuación se muestran los mapas resultantes de la estimación y de error asociado, con sus correspondientes leyendas de niveles, identificado con las posiciones de los datos originales (pozos).

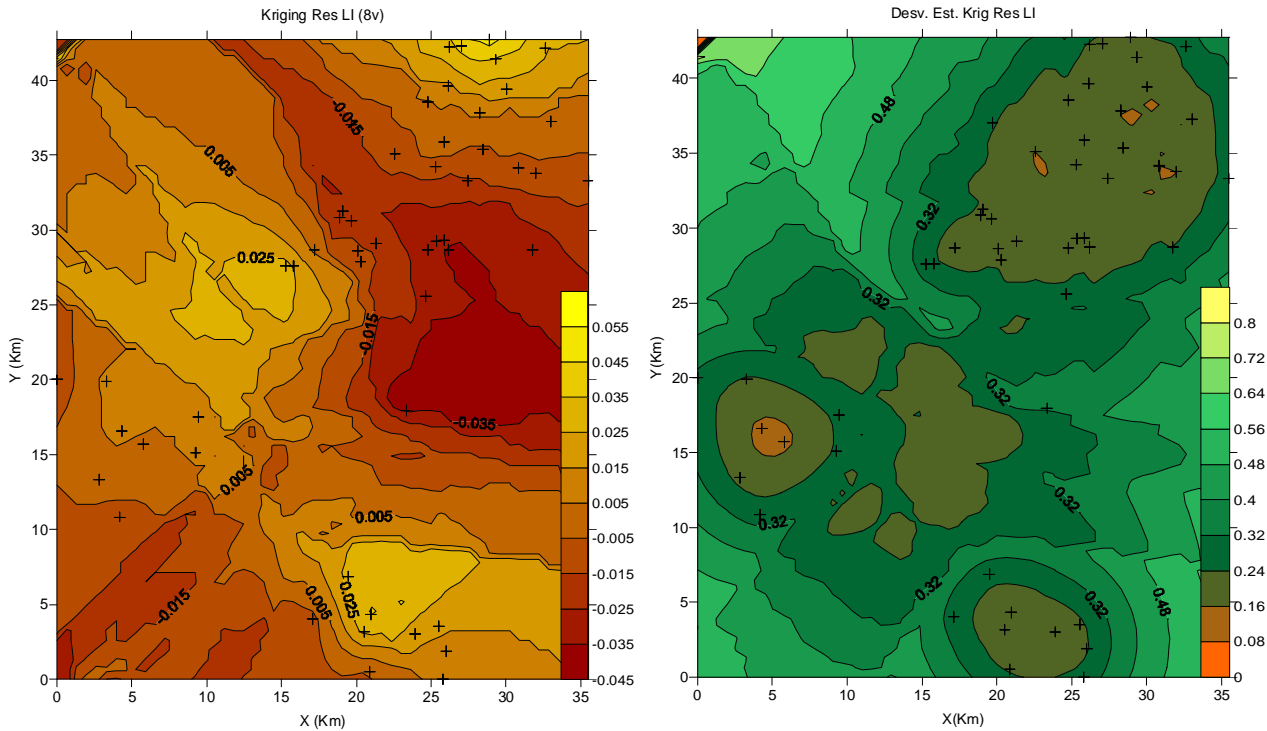


Figura 16. Kriging puntual de Residuos de LI (izq) y error de la estimación (der).

Finalmente, el estimado de la variable LI es el resultante de la suma de la media $M(x)$ + el mapa de kriging de Residuos $R(x)$. Ambos gráficos se aprecian en la figura 17.

4.2. FLUORURO

Análogo al caso anterior, el kriging para los residuos de F se hizo puntual de 8 vecinos, con un radio de búsqueda igual al alcance del variograma. En éste caso el modelo de variograma es también esférico, con efecto Nugget=0,025 y $S+N=0,075$. La malla regular es idéntica a la usada para LI. La figura 18 muestra los mapas resultantes de la estimación y del error asociado, con sus correspondientes leyendas de niveles. Como se aprecia, los errores asociados a la estimación por kriging para F resultaron ligeramente inferiores a los obtenidos para LI en las cercanías de los datos duros. La figura 19 muestra los valores totales estimados de la forma $Z(x)^*=M(x)+R^*(x)$.

Tarea 3: EE/ Grupo 3

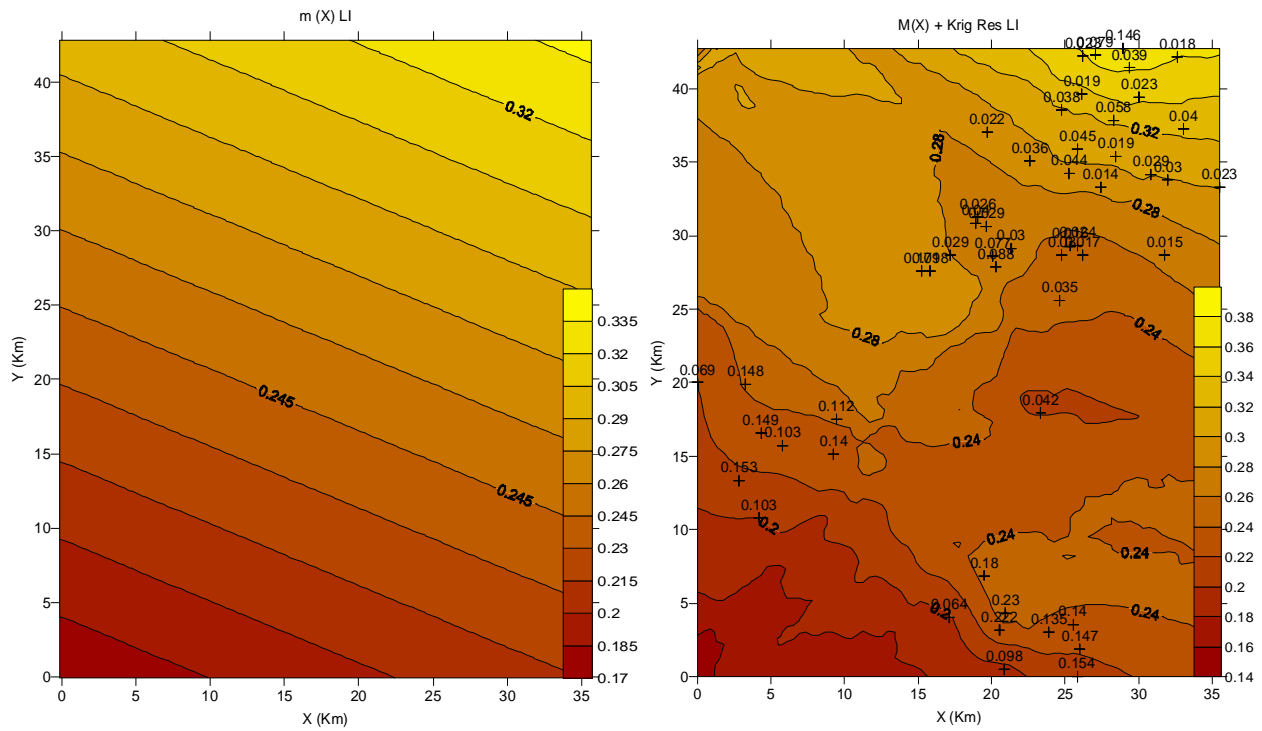


Figura 17. Media de LI = $M(X)$ (izq), y estimado total = $Z^*(X)=M(X)+R^*(X)$ (der).

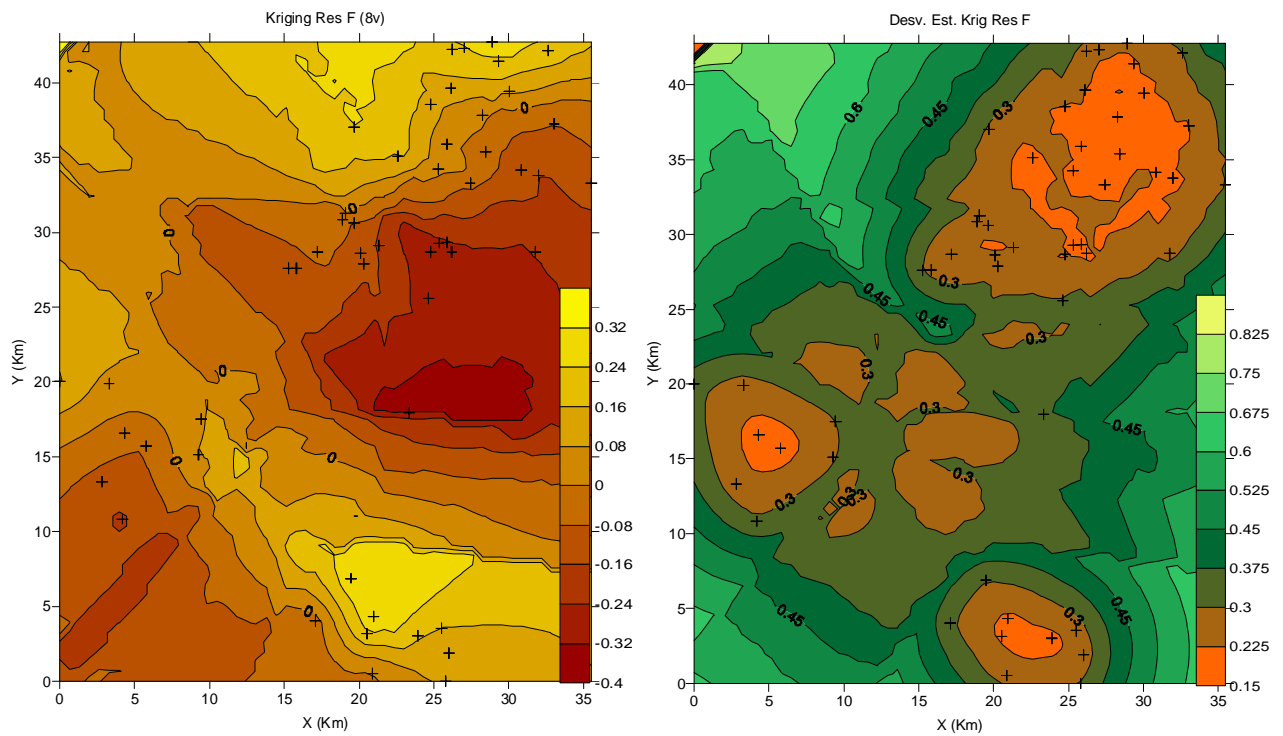


Figura 18. Kriging puntual de Residuos de F (izq) y error de la estimación (der).

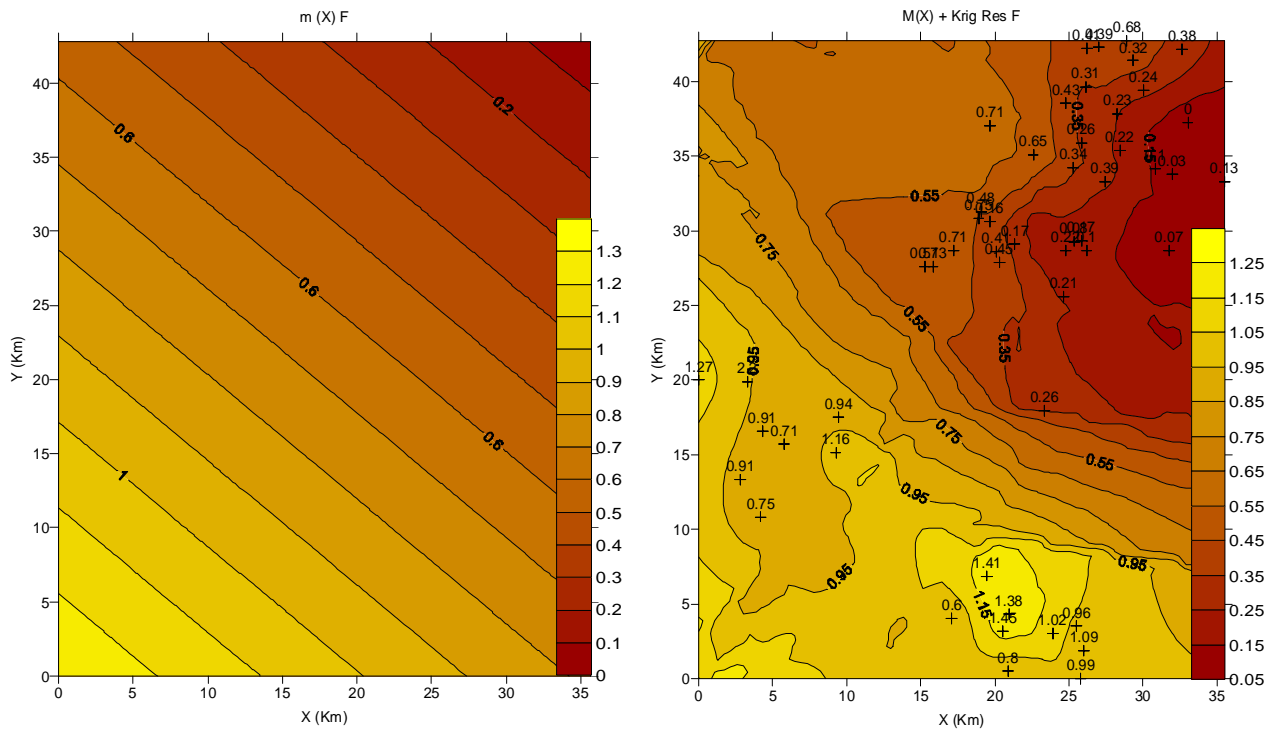


Figura 19. Media de F = M(X) (izq), y estimado total = Z*(X)=M(X)+R*(X) (der).

5. ESTIMACIÓN CON COKRIGING PARA LA VARIABLE LI

Para proceder a realizar la estimación con Cokriging es necesario verificar que el *modelo de correionalización lineal* sea válido. Para ello se requiere que los determinantes de las matrices que lo definen sean positivos. La ecuación del modelo tiene la siguiente forma:

$$\begin{bmatrix} \gamma_{LI}(h) & \gamma_{F-LI}(h) \\ \gamma_{LI-F}(h) & \gamma_F(h) \end{bmatrix} = \begin{bmatrix} \sigma_{LI}^0 & \sigma_{F-LI}^0 \\ \sigma_{LI-F}^0 & \sigma_F^0 \end{bmatrix} \gamma_0(h) + \begin{bmatrix} \sigma_{LI}^S & \sigma_{F-LI}^S \\ \sigma_{LI-F}^S & \sigma_F^S \end{bmatrix} \gamma_1(h)$$

Los parámetros de los variogramas de los residuos fijados anteriormente sufrieron ligeras modificaciones con el fin de ajustar éste modelo, los cuales quedaron definidos como se muestra en la siguiente tabla:

Variables	Modelo	Nugget	Sill-Nugget	Alcance (Km)	RSS
Res 1° LI	Esférico	0,0009	0,0012	20->21	3.658E-6
Res 1° F	Esférico	0,015->0,025	0,65->0,050	22->21	5.003E-3
Res LI – Res F	Esférico	0,0005	0,007	21	6.889E-5

Así todos los variogramas poseen modelo esférico de igual rango, variando sólo en el Sill. Sustituyendo en el anterior sistema de ecuaciones resulta:

$$\begin{bmatrix} \gamma_{LI}(h) & \gamma_{F-LI}(h) \\ \gamma_{LI-F}(h) & \gamma_F(h) \end{bmatrix} = \begin{bmatrix} 0,0009 & 0,0005 \\ 0,0005 & 0,025 \end{bmatrix} \gamma_0(h) + \begin{bmatrix} 0,0012 & 0,007 \\ 0,007 & 0,05 \end{bmatrix} \gamma_1(h)$$

Cuyos determinantes son respectivamente:

$$\begin{vmatrix} 0,0009 & 0,0005 \\ 0,0005 & 0,025 \end{vmatrix} = 0,00002475 > 0 \quad \text{y} \quad \begin{vmatrix} 0,0012 & 0,007 \\ 0,007 & 0,05 \end{vmatrix} = 0,000017 > 0$$

Ambos determinantes son positivos, de manera que el modelo de correlogramación es válido. Así las cosas, se procede a mostrar el resultado de la estimación por cokriging para la variable LI y el correspondiente error asociado.

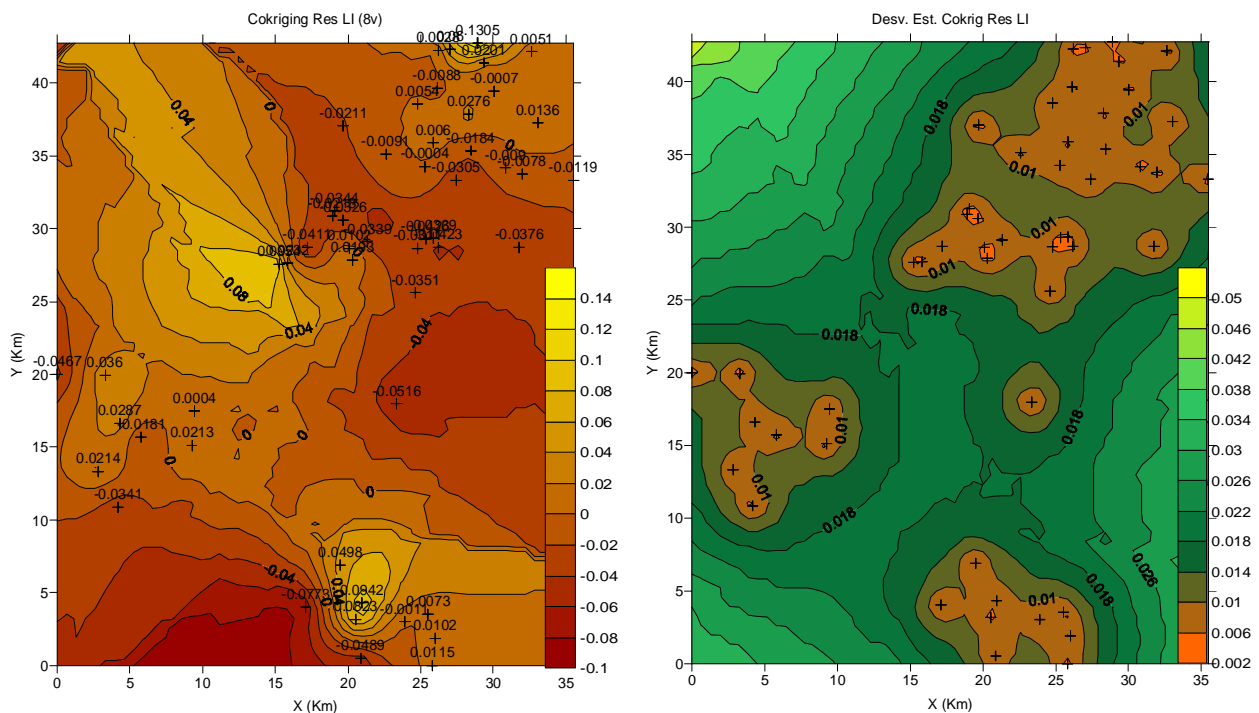


Figura 20. Cokriging puntual de Residuos de LI en función de los residuos de F (izq), y error de la estimación (der).

Es notable que en éste caso el error de la estimación de los residuos es menor que la obtenida mediante kriging Universal. Afirmación que resalta al comparar ambos mapas de error en la figura 22. Lo cual permite inferir que la estimación por cokriging resulta ser

Tarea 3: EE/ Grupo 3

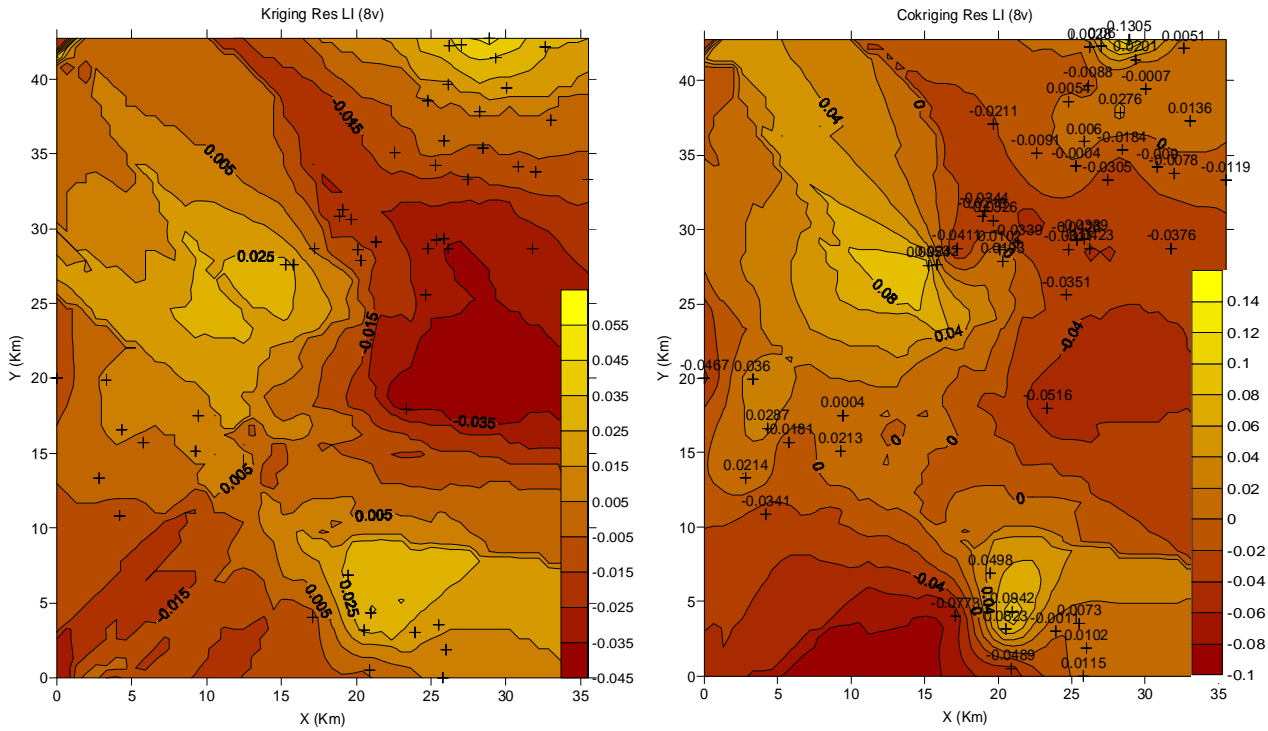


Figura 21. Comparación de la estimación de Residuos de LI por Kriging (izq) y Cokriging (der).

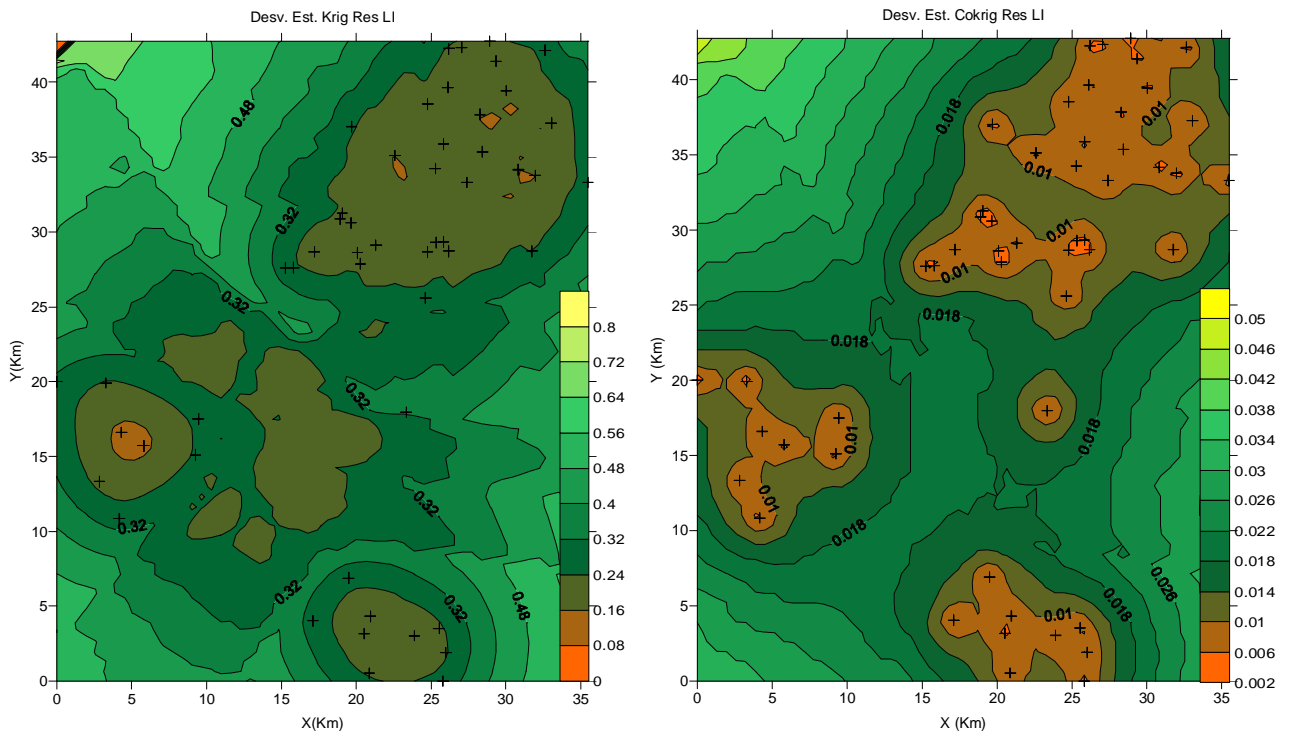


Figura 22. Comparación de los errores de la estimación de Residuos de LI por Kriging (izq) y Cokriging (der).

más exacta que la obtenida por kriging. Ésta es precisamente una de las conclusiones a la que se esperaba llegar en vista de la alta correlación existente entre ambos pares de datos de pozo. La inclusión de la variable F en la estimación de los residuos de LI minimiza la desviación estándar de los puntos donde originalmente no existen datos.

6. CONCLUSIONES

En el presente trabajo se realizó un análisis geoestadístico general a dos conjuntos de datos de pozo correlacionables: Litio y Fluoruro. El análisis exploratorio y de estacionaridad nos condujo a eliminar los valores atípicos que enmascaraban el comportamiento de la mayoría de los datos. La presencia de "tendencia" en ambas variables nos llevó a estimar los respectivos residuos de 1er grado. Por otra parte, no existieron suficientes argumentos para sostener la hipótesis de anisotropía, debido a la cantidad y distribución espacial de los valores.

Se realizó la estimación espacial por kriging puntual en una malla regular para los residuos de ambas variables. La teoría de Kriging Universal permitió generar mapas de los valores estimados y sus correspondientes errores asociados. Posteriormente, se estimó por cokriging la variable Litio, lográndose mejores resultados respecto a la varianza del error en comparación a la obtenida por kriging. Se argumenta que dicho fenómeno se debe a la correlación existente entre los datos de Litio y Fluoruro.