

Artículo publicado en el Repositorio Institucional del IMTA

<i>Título</i>	Estimación multivariada espacio-tiempo de la carga hidráulica en el valle de Querétaro-Obrajuelo.
<i>Autor / Adscripción</i>	Edgar Yuri Mendoza Cázares Universidad Nacional Autónoma de México Graciela Herrera Zamarrón Instituto Mexicano de Tecnología del Agua
<i>Publicación</i>	Ingeniería Hidráulica en México, 22(1): 63-80
<i>Fecha de publicación</i>	2007
<i>Resumen</i>	En el presente trabajo se utilizó un método de estimación multivariada espacio-temporal, considerando la carga hidráulica en diferentes años como variables diferentes, pero correlacionadas temporalmente. El método se aplicó para estimar la carga hidráulica en tres años diferentes en el acuífero del valle de Querétaro-Obrajuelo, México.
<i>Identificador</i>	http://hdl.handle.net/123456789/848

Estimación multivariada espacio-tiempo de la carga hidráulica en el valle de Querétaro-Obrajuelo

Edgar Yuri Mendoza-Cázares

Universidad Nacional Autónoma de México

Graciela Herrera-Zamarrón

Instituto Mexicano de Tecnología del Agua

En el presente trabajo se utilizó un método de estimación multivariada espacio-temporal, considerando la carga hidráulica en diferentes años como variables diferentes, pero correlacionadas temporalmente. Este tipo de método permite aumentar la información utilizada en la estimación de la carga hidráulica para cada año. El método se aplicó para estimar la carga hidráulica en tres años diferentes en el acuífero del valle de Querétaro-Obrajuelo, México, obteniéndose buenos resultados en la validación cruzada y en el sentido de que al utilizar la información de los tres años conjuntamente se pueden observar claramente características de las configuraciones de las cargas que no se obtienen cuando únicamente se utilizan los datos del año estimado.

Palabras clave: análisis multivariado, geoestadística, cokrigado, espacio-tiempo, estimación carga hidráulica.

Introducción

La carga hidráulica en los acuíferos se puede representar como una variable que es función tanto de la posición como del tiempo. Con frecuencia, al llevar a cabo el análisis de la evolución de las cargas hidráulicas, se encuentra que los datos son escasos y que no para todos los años se cuenta con datos en los mismos pozos. Esto dificulta dicho análisis debido a que, por un lado, la estimación para cada año puede contar con poca información y hacer que su incertidumbre sea grande y, por otro lado, en los diferentes años la estimación tendrá un nivel de incertidumbre diferente en zonas distintas, lo que hace difícil compararlas.

En el presente trabajo se utilizó un método de estimación multivariado en el que se consideran las distribuciones espaciales de las cargas hidráulicas en distintos años como variables diferentes, pero correlacionadas temporalmente para estimar las cargas en diferentes fechas en el acuífero del valle de Querétaro-Obrajuelo, México. Este tipo de método permite aumentar la información utilizada en la estimación de la carga hidráulica para cada año y disminuir las diferencias en cuanto al

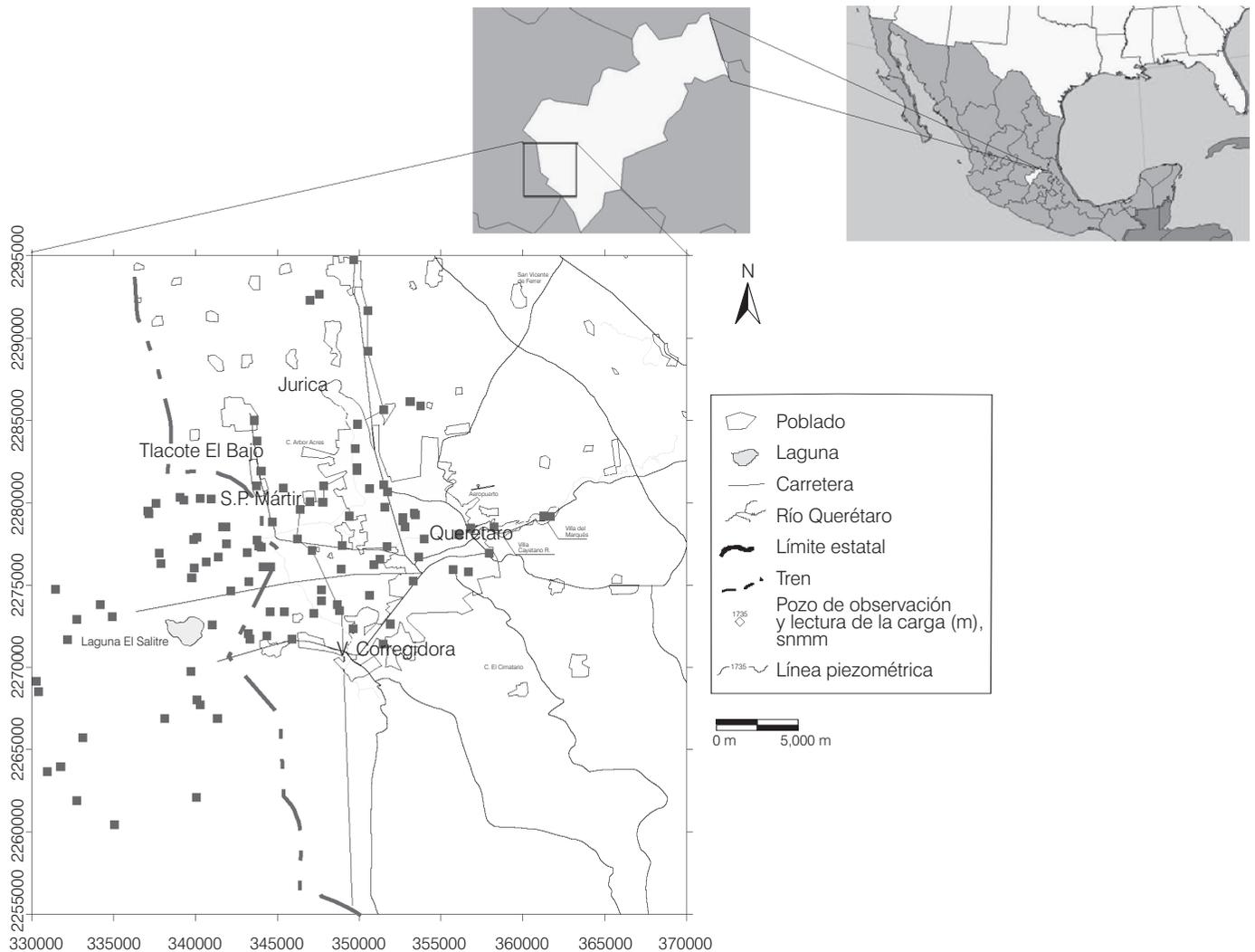
número de datos utilizados en las diferentes zonas para cada año.

Rouhani y Hall (1989) estudiaron por primera vez un problema en el que se consideraron tanto la variación espacial como temporal de las variables en el campo de la hidrogeología. Ellos aplicaron un método en el que el tiempo se consideró como una coordenada más añadida a las coordenadas espaciales.

Un segundo punto de vista para la modelación espacio-tiempo es considerar el proceso como un grupo espacial, temporalmente correlacionado (Papritz y Flühler, 1994; Kyriakidis y Journel, 1999) o un grupo de observaciones de series de tiempo espacialmente correlacionadas (Solow y Gorelick, 1986; Wackernagel, 1995). Esto dependerá de qué dominio (espacio o tiempo) esté mayormente muestreado.

En este trabajo se analizaron las observaciones espaciales de las cargas hidráulicas en las diferentes fechas, considerándolas temporalmente correlacionadas, aplicando la técnica multivariada de funciones aleatorias. Esta técnica no ha sido implementada en el campo de la hidrogeología; Papritz y Flühler (1994) la aplicaron en el campo de la edafología, pero a diferencia de lo realizado en este

Ilustración 1. Área de estudio.



artículo, incluyeron dentro de la formulación de *cokrigead* la función de pseudo semivariograma y la estimación de *cokrigead* en bloques. Para modelar la correlación entre las variables y sus respectivos elementos cruzados utilizamos un modelo lineal de correogionalización (MLC), basado en las funciones de semivariograma y semivariograma cruzado. La estimación se efectuó mediante el algoritmo de *cokrigead* puntual. Las estimaciones obtenidas con los modelos univariado y bivariado se compararon mediante el método de validación cruzada.

Este método no había sido implementado en el campo de la hidrogeología. Papritz y Flühler (1994) implementaron un método similar en el campo de la edafología, pero a diferencia de lo hecho en el presente trabajo, utilizaron *cokrigead* en bloques y la función de pseudo semivariograma para medir las correlaciones.

Área de estudio

La zona de estudio comprende una superficie de 426 km², incluye al acuífero del valle de Querétaro, ubicado en la porción sur oriental del estado de Querétaro, y al acuífero del valle de Obrajuelo, ubicado en el estado de Guanajuato, colindando con el acuífero del valle de Querétaro (ilustración 1).

En el área se encuentran dos corrientes poco caudalosas: los ríos Querétaro y El Pueblito. La corriente principal es el río Querétaro, que corre desde La Cañada hasta interceptar con el río El Pueblito. La información geológica y estructural indican que el valle de Querétaro-Obrajuelo se aloja en una secuencia de fosas tectónicas (originada por el tectonismo ocurrido en el Terciario), por lo que el acuífero está limitado por un sistema de fallas,

sepultadas por sedimentos aluviales, flujos de lava y depósitos volcanoclásticos lacustres potentes. La disposición natural de las unidades litológicas agrupadas en los medios poroso y fracturado conforman el sistema que define el flujo del agua subterránea en la zona de estudio, lo que constituye un acuífero del tipo libre, comportándose como confinado localmente en algunas localidades por la presencia de arcilla (Guysa, 1996).

El modelo conceptual infiere que las principales zonas de recarga al acuífero se presentan al este, por La Cañada; al noreste, proveniente de las montañas; al norte, en la zona de Jurica; al sur se tiene una entrada en el lineamiento del río El Pueblito al suroeste, y otra más en el lado norte de Obrajuelo.

Se considera dos sistemas de flujo, uno local que está formado por zonas con elevaciones del nivel freático de poca profundidad en la zona central del valle y un sistema de flujo intermedio que sustenta casi el total de extracción en el valle y que está formado también por medio poroso y fracturado. Por otro lado, en la zona entre los límites del estado de Querétaro y Guanajuato se tiene la presencia de abatimientos importantes, además de que no es posible definir la presencia de un partaguas subterráneo o alguna frontera física que permita separar a ambos acuíferos (Simuta-Champo, 2005). Por lo anterior, no hay evidencias de algún elemento que permita estudiar y, en específico, separar a ambos acuíferos, a pesar de que en estudios anteriores se ha hecho así.

Marco teórico

Una variable regionalizada, definida como una variable distribuida en el espacio-tiempo de forma que presenta una estructura de correlación, desde el punto de vista probabilístico es una realización de una función aleatoria en espacio-tiempo. El proceso de modelar probabilísticamente el espacio-tiempo puede ser simplificado en algunas circunstancias, dependiendo de la estructura que muestren los datos. Si los datos consisten de series de tiempo tomadas en pocos puntos de medición, entonces el proceso espacio-temporal puede ser modelado como series de tiempo multivariadas, es decir, con una sucesión de funciones que dependen únicamente del tiempo (Solow y Gorelick, 1986; Rouhani y Wackernagel, 1990). De igual forma, si se cuenta con varias observaciones, distribuidas en el espacio, que fueron medidas en diferentes tiempos, entonces el proceso espacio-tiempo puede ser modelado como un proceso aleatorio espacial multivariado, esto es, una sucesión de funciones que dependen únicamente del espacio (Papritz y Flühler, 1994).

En el caso multivariado espacial, la técnica utilizada para efectuar la estimación de la variable de interés es conocida como *cokrigado*, que es la extensión a dos o más variables de la técnica univariada conocida como *krigado*. Ambas técnicas están basadas en la representación de la variable regionalizada con una función aleatoria $Z(x)$ compuesta por dos componentes, como se muestra en la ecuación (1):

$$Z(x) = m(x) + R(x) \quad (1)$$

en donde x representa la posición en un sistema geográfico y puede tener una, dos o tres componentes. La componente $m(x)$ es conocida como la tendencia o deriva, y es habitual suponer que sea cero (Aboufirassi y Mariño, 1984; Samper y Carrera, 1990). La variable aleatoria $R(x)$, a la que llamaremos residual, cumple con lo siguiente:

- Su esperanza matemática existe, no depende del punto de observación y es nula.
- Para todos los incrementos (h) tienen una varianza finita.

Estas suposiciones son conocidas como hipótesis intrínseca (Journel y Huijbregts, 1978). En geostatística, debido a la hipótesis intrínseca, es suficiente con conocer los dos primeros momentos de la distribución para caracterizarla. El primero de los momentos es la esperanza matemática, seguido de los momentos de segundo orden, conocidos como varianza, covarianza y semivariograma.

El semivariograma de una variable se define como:

$$\gamma(h) = \frac{1}{2} E \left[\left\{ Z^i(x) - Z^i(x+h) \right\}^2 \right] \quad (2)$$

y para dos o más variables se define como:

$$\gamma_{ij}(h) = \frac{1}{2} E \left[\left\{ Z^i(x+h) - Z^i(x) \right\} \left\{ Z^j(x+h) - Z^j(x) \right\} \right] \quad (3)$$

en donde $Z^i(x)$ y $Z^j(x)$ son dos variables aleatorias con función de probabilidad conjunta que satisface la hipótesis intrínseca. Bajo la hipótesis intrínseca, la función de semivariograma muestral se calcula de acuerdo con la siguiente ecuación:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{k=1}^{N(h)} \left[Z^i(x_k+h) - Z^i(x_k) \right]^2 \quad (4)$$

en donde el valor de $N(h)$ representa el número de pares, $Z^i(x_k+h)$ es el valor de la variable en la k -ésima posición más un vector h y $Z^i(x_k)$ es el valor de la variable en la k -ésima posición.

Para el caso del semivariograma cruzado muestral, la ecuación que lo define es la siguiente:

$$\hat{\gamma}_{ij}(h) = \frac{1}{2N(h)} \sum_{k=1}^{N(h)} \left[\{Z^i(x_k+h) - Z^i(x_k)\} \{Z^j(x_k+h) - Z^j(x_k)\} \right] \quad (5)$$

Nuestro interés es estimar en el punto x_0 el valor desconocido de la variable principal $Z^i(x_0)$, tomando en cuenta los valores de las variables Z^{ki} y Z^{kj} . Para ello consideramos un estimador lineal Z^* de la forma:

$$Z^* = \sum_{ki=1}^{ni} \lambda_{ki}^{cok} Z^i(x_{ki}) + \sum_{kj=1}^{nj} \lambda_{kj}^{cok} Z^j(x_{kj}) \quad (6)$$

en donde λ_{ki}^{cok} es el valor del peso asignado a la variable primaria, Z^i , con posición x_{ki} y λ_{kj}^{cok} es el valor del peso asignado a la variable secundaria, Z^j , con posición x_{kj} . El estimador Z^* debe ser insesgado, por lo que los pesos deben cumplir:

$$\sum_{ki=1}^{ni} \lambda_{ki}^{cok} = 1, \quad \sum_{kj=1}^{nj} \lambda_{kj}^{cok} = 0 \quad (7)$$

y de mínima varianza (Myers, 1982). La minimización del error de la varianza bajo estas dos condiciones da como resultado el sistema lineal de $n_1(x) + n_2(x) + 2$, ecuaciones:

$$\begin{aligned} & \sum_{\beta i=1}^{n_1} \lambda_{\beta i}^{cok} \gamma_{ii}(x_{ki} - x'_{\beta i}) + \sum_{\beta j=1}^{n_2} \lambda_{\beta j}^{cok} \gamma_{ij}(x_{ki} - x'_{\beta j}) \\ & + \mu_i^{cok} = \gamma_{ii}(x_{ki} - x_0), \quad ki = 1, \dots, n_i \\ & \sum_{\beta i=1}^{n_1} \lambda_{\beta i}^{cok} \gamma_{ji}(x_{kj} - x'_{\beta i}) + \sum_{\beta j=1}^{n_2} \lambda_{\beta j}^{cok} \gamma_{jj}(x_{kj} - x'_{\beta j}) \\ & + \mu_j^{cok} = \gamma_{ji}(x_{kj} - x_0), \quad kj = 1, \dots, n_j \\ & \sum_{\beta i=1}^{n_1} \lambda_{\beta i}^{cok} = 1, \quad \sum_{\beta j=1}^{n_2} \lambda_{\beta j}^{cok} = 0 \end{aligned} \quad (8)$$

en donde los parámetros $\lambda_{\beta i}^{cok}$ y $\lambda_{\beta j}^{cok}$ son los pesos asignados a cada observación, condicionados para cumplir la condición de insesgo como indica la ecuación (7), los μ_i^{cok} , μ_j^{cok} son los parámetros de Lagrange (Goovaerts, 1998).

La varianza del cokrigado ordinario se obtiene de:

$$\begin{aligned} \text{Var}(Z^* - Z^i(x_0)) &= \sigma^2 + \mu_i^{cok} - \sum_{\alpha j=1}^{n_1} \lambda_{\alpha j}^{cok} \gamma_{ij}(x_i - x_0) \\ & - \sum_{\alpha j=1}^{n_2} \lambda_{\alpha j}^{cok} \gamma_{ij}(x_i - x_0) \end{aligned} \quad (9)$$

en donde σ^2 es la varianza de la variable primaria. Cabe comentar que las ecuaciones de cokrigado mostradas anteriormente corresponden al caso bivariado; sin embargo, la forma general de este sistema para el caso en que se analizan tres o más variables puede ser verificada en Goovaerts (1997, 1999).

Metodología y resultados

Los pasos que se utilizaron en el análisis geoestadístico de los datos fueron el análisis preliminar y exploratorio de los datos, el análisis estructural, la estimación de la variable y la validación cruzada; se describen a continuación.

Análisis preliminar

Este procedimiento consiste en clasificar, ordenar e integrar la información de la carga hidráulica en el acuífero, para los años 1993, 1995 y 1999, por lo que se creó una base de datos que comprende un total de 185 pozos, distribuidos en todo el acuífero, de los cuales se cuenta con la cota del brocal y el valor de elevación de la carga hidráulica. Reportando un total de 89 lecturas de pozos para el año de 1993, 112 para el año de 1995 y 75 pozos para 1999.

Análisis exploratorio

El análisis exploratorio se basa en técnicas estadísticas convencionales que nos permiten caracterizar los datos. Se realizó un análisis exploratorio de cada uno de los parámetros. Éste inició con la determinación de la posición y magnitud de los valores de los datos. Para ello se graficaron los puntos de medición con su respectivos valores. Posteriormente se realizó un análisis estadístico y se obtuvieron las elevaciones de la carga hidráulica. También se verificó que la distribución de la muestra fuera aproximadamente normal (véase cuadro 1).

Debido a que las distribuciones anteriores no son normales, se procedió a hacer una transformación de los datos que consistió en restarles el valor de la tendencia, misma que se calculó a partir de un análisis de regresión por mínimos cuadrados ordinarios, ajustando a un polinomio de segundo orden (véase cuadro 1).

Cuadro 1. Estadígrafos de la carga hidráulica.

Año	1993	1995	1999	1993 residual	1995 residual	1999 residual
Número de pozos	85	112	75	85	112	75
Media	1,737.35	1,728.37	1,720.65	-2.0E-6	-3.0E-6	4.74E-17
Mediana	1,713.48	1,706.89	1,693.65	2.82	0.54	0.75
Desviación estándar	58.12	54.46	62.15	28.64	27.42	30.71
Mínimo	1,640.42	1,629.24	1,623.33	-90.29	-92.66	-89.44
Máximo	1,894.43	1,892.48	1,890.40	79.13	89.45	83.74
Curtosis	1.64	2.72	1.55	1.57	2.02	1.37
Coefficiente de simetría	1.60	1.84	1.55	-0.39	-0.05	-0.11
	asimetría grave	asimetría grave	asimetría grave	asimetría leve	hay simetría	hay simetría

Análisis estructural

El propósito de este paso es determinar la estructura espacial de los datos, para este fin se determina el semivariograma muestral y el semivariograma cruzado muestral y, posteriormente, mediante un proceso de ajuste, se propone un modelo teórico para ambos semivariogramas. Estos modelos teóricos son conocidos en la geoestadística como modelos autorizados (Olea, 1991) y cumplen con la condición de ser condicionalmente negativos definidos (Christakos, 1984; Cressie, 1991).

En este trabajo el semivariograma muestral y el semivariograma cruzado muestral se calcularon con intervalos de 10 clases y un vector de separación h igual a un medio la distancia máxima de las observaciones, utilizando los algoritmos *GAMV.exe* y *VMODEL.exe* de *GSLIB* (Deutsch y Journel, 1998).

Ajuste de un modelo teórico al semivariograma muestral

Una vez propuesto el semivariograma muestral se le ajusta un modelo teórico. Para el caso univariado, se procede comparando con todos los modelos autorizado (e.g. lineal, exponencial, esférico, cuadrático, guassiano, agujero), con la finalidad de observar con cuál de ellos se tiene un mejor ajuste. Entendiéndose por ajuste el proceso mediante el cual se compara el modelo teórico con el muestral, modificando los parámetros del primero hasta que reproduzca de mejor manera el comportamiento del variograma muestral. En el proceso del ajuste se tienen varias metodologías, que van desde el ajuste visual, que es meramente cualitativo, utilizando el método de validación cruzada (Gambolati y Volpi, 1979; Volpi y Gambolati, 1979) o los ajustes por aproximación de mínimos cuadrados (Cressie, 1985; Tough y Leyshon, 1985), siendo este último el utilizado en este trabajo.

En el caso del semivariograma cruzado, el análisis es más complejo, debido a que no existen semivariogramas cruzados autorizados, por lo que el semivariograma cruzado muestral se modela ajustando un modelo lineal de correogionalización (MLC) de la forma:

$$\gamma_{ij}(h) = \sum_{u=0}^{N_s} b_{ij}^u g_u(h) \tag{10}$$

en donde u es el indicador de las diferentes estructuras espaciales, N_s es el número total de estructuras espaciales, $g_u(h)$ es un semivariograma con un alcance (a) específico y b_{ij}^u es una matriz positiva definida que contiene los coeficientes del semivariograma.

Journel y Huijbregts (1978) proponen que para el caso de tres variables, el MLC, necesariamente se debe comprobar que las estructuras de cada semivariograma cumpla con:

$$b_{ii}^u > 0; \quad |b_{ij}^u| = |b_{ji}^u| \leq \sqrt{b_{ii}^u b_{jj}^u} \text{ desigualdad}$$

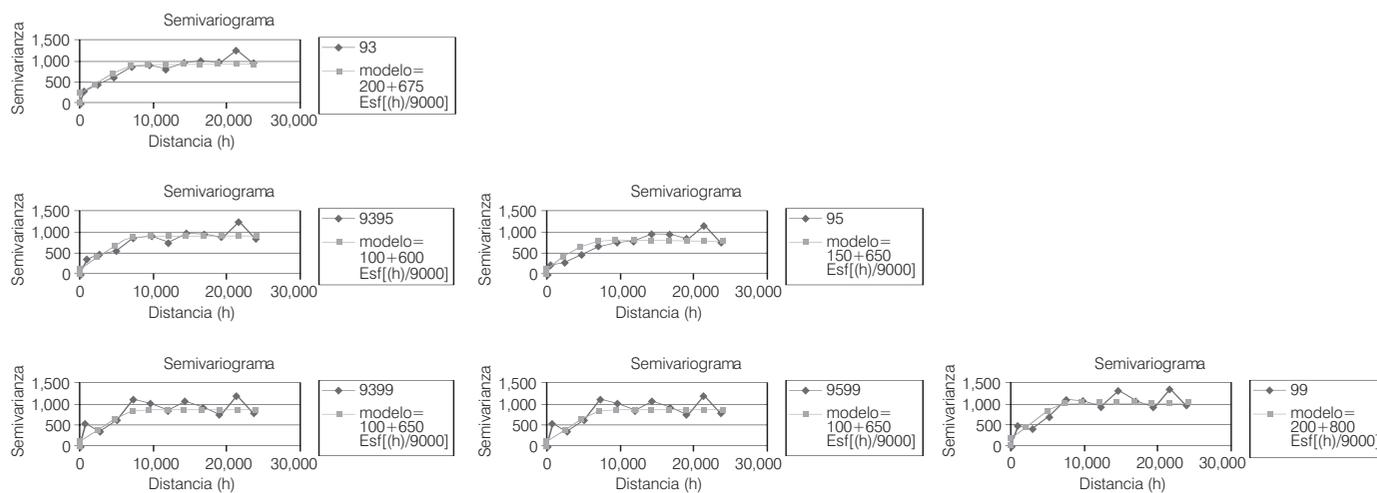
$$\text{“Cauchy-Schwartz”}; \quad |b_{kk}^u| \geq 0$$

Wackernagel (1989) propone resolver estas mismas condiciones mediante un algoritmo denominado *LINMOD*, que de forma automática comprueba que las matrices de los coeficientes cumplen con las desigualdades comentadas.

El modelo de *cokrigado* del *GSLIB*, el utilizado en este trabajo, comprueba de forma interna que la matriz del MLC sea positiva definida, de no serlo envía una advertencia avisando que la matriz de *cokrigado* es singular.

Goulard y Voltz (1992) proponen un método distinto, donde el MLC se ajusta por un método automático, utilizando la técnica de mínimos cuadrados. En nuestro caso se procedió con el criterio establecido por Journel

Ilustración 2. Semivariogramas muestral y modelo (esférico) para los años 1993, 1995 y 1999, con sus respectivos semivariogramas cruzados. Esf = esférico, 93 indica que es el semivariograma muestral del año 1993 y sucesivamente; 9395 indica que es el semivariograma cruzado muestral de los años 1993-1995 y sucesivamente, /9000 indica que es el alcance.



y Huijbregts (1978), seleccionando aquel modelo que mejor ajustara el proceso y fuera más sencillo (principio de parsimonia). Los resultados del análisis se resumen en la ilustración 2.

Estimación

La estimación de la elevación de la carga hidráulica y varianza se realizó por medio de *krigeado*, en el caso univariado (Delhomme, 1978), y de *cokrigeado* en el caso multivariado. La estimación de la carga hidráulica y varianza en el acuífero de Querétaro se efectuó en una malla de 41 columnas y 51 renglones, con una distancia entre renglón y columna de 1,100 m para ambos casos. Se utilizó un radio de búsqueda de 9,000 m y un número mínimo de cuatro datos en la estimación de la carga hidráulica; esto implica que los puntos en la malla que no cuenten con cuatro datos a una distancia de 9,000 m no son estimados, por lo que la región espacial que se estima en cada año depende de esto. En las diferentes zonas, la validez de los resultados debe ser analizada con un criterio geohidrológico y tomando en cuenta la varianza del error de la estimación en las diferentes zonas del área de estudio.

Estimación univariada, krigado ordinario

La elevación de la carga hidráulica en el año de 1993 (ilustración 3) presenta un marcado descenso en la porción central del valle, con una elevación mínima de 1,640.42 m. Los valores que indican una mayor carga

hidráulica se distribuyen en los márgenes del acuífero. En esta zona se encuentra el valor máximo, que es de 1,894.43 m; dicho valor se considera una sobreestimación, ya que la mayor concentración de pozos con datos se ubica en el centro del valle y en los márgenes hay pocos datos. La estimación en este año muestra un gradiente de forma casi concéntrica, alargándose en dirección E-W.

El plano de varianza puede ser interpretado como un mapa de confiabilidad de las estimaciones de *krigeado* o *cokrigeado*. Tomando en cuenta lo anterior, la ilustración 4 muestra que los valores de menor varianza se encuentran en las áreas donde existe mayor concentración de lecturas, en nuestro caso, en la porción central del acuífero del valle de Querétaro y Obrajuelo; en contraste, los valores altos de varianza se localizan en las zonas donde no existen lecturas y los bordes del acuífero.

En cuanto a la configuración del año 1995 (ilustración 5), se tiene un marcado descenso del carga hidráulica porción central del valle, con una elevación mínima de 1,629.24 m y un valor máximo de 1,892.48 metros.

De manera general, los valores de varianza (ilustración 6) se mantienen en la misma distribución que en 1993, presentando diferencias en la zona Laguna El Salitre, en donde existe un mayor número de lecturas, en comparación con el año de 1993.

Para el año de 1999 (ilustración 7) se mantiene el esquema de flujo que indica una confluencia de las líneas potenciales hacia la parte central del valle, con una elevación mínima de 1,623.33 m y un valor máximo de 1,890.4 metros.

Ilustración 3. Valores de la carga hidráulica *krigeado* ordinario, 1993.

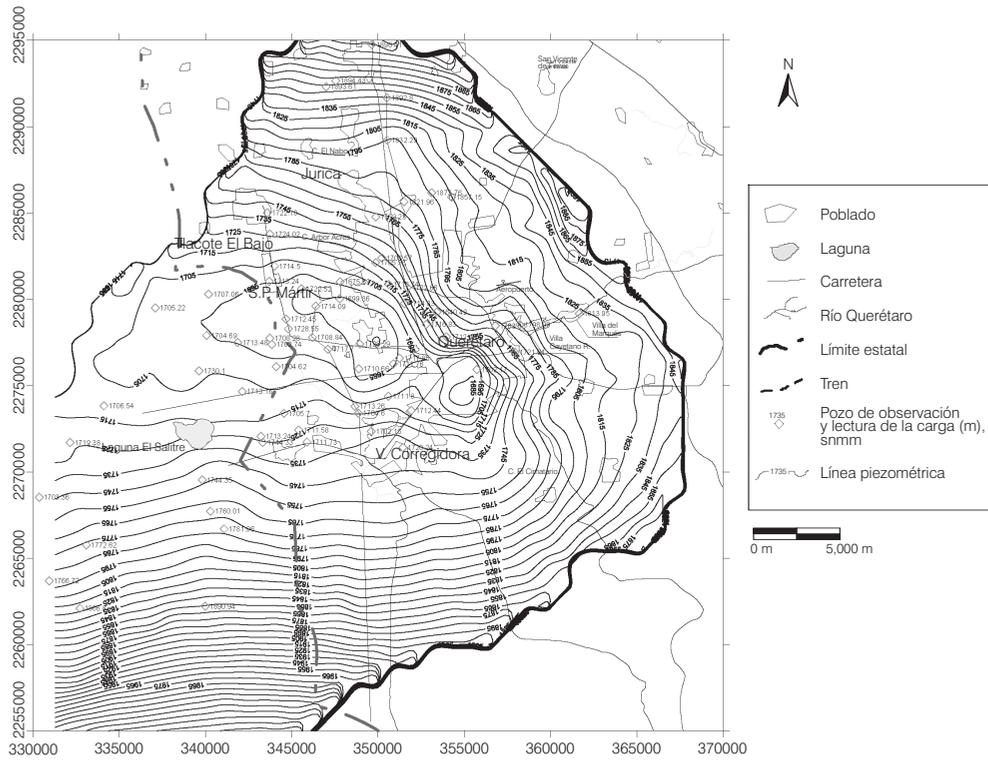


Ilustración 4. Valores de la varianza *krigeado* ordinario, 1993.

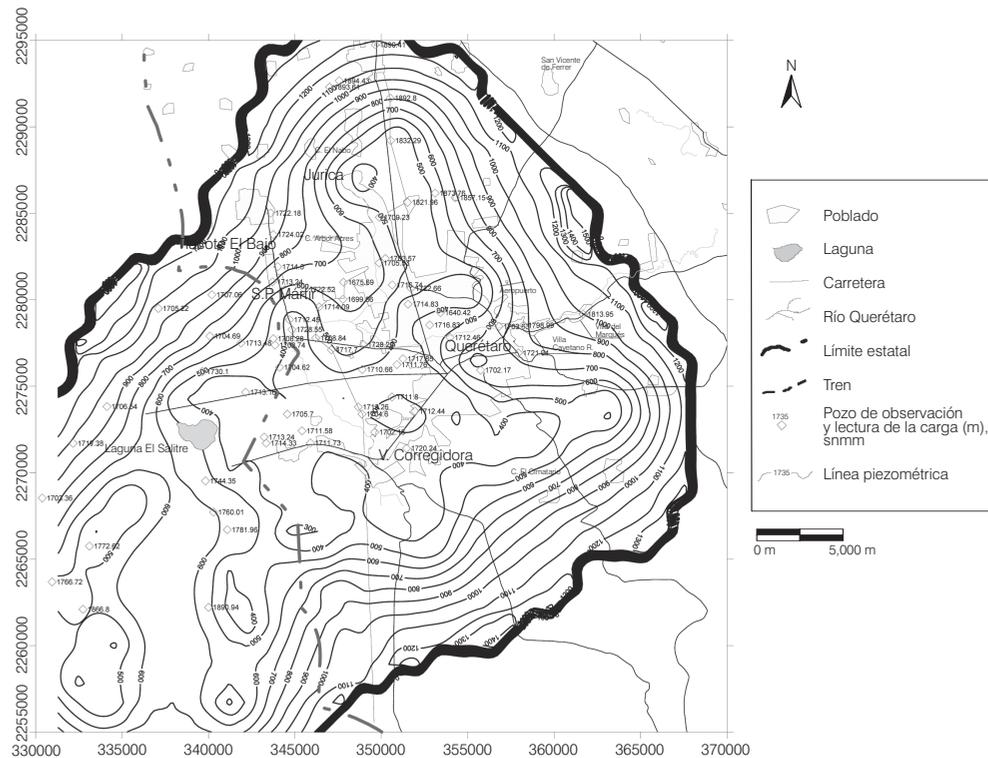


Ilustración 5. Valores de la carga hidráulica *krigeado* ordinario, 1995.

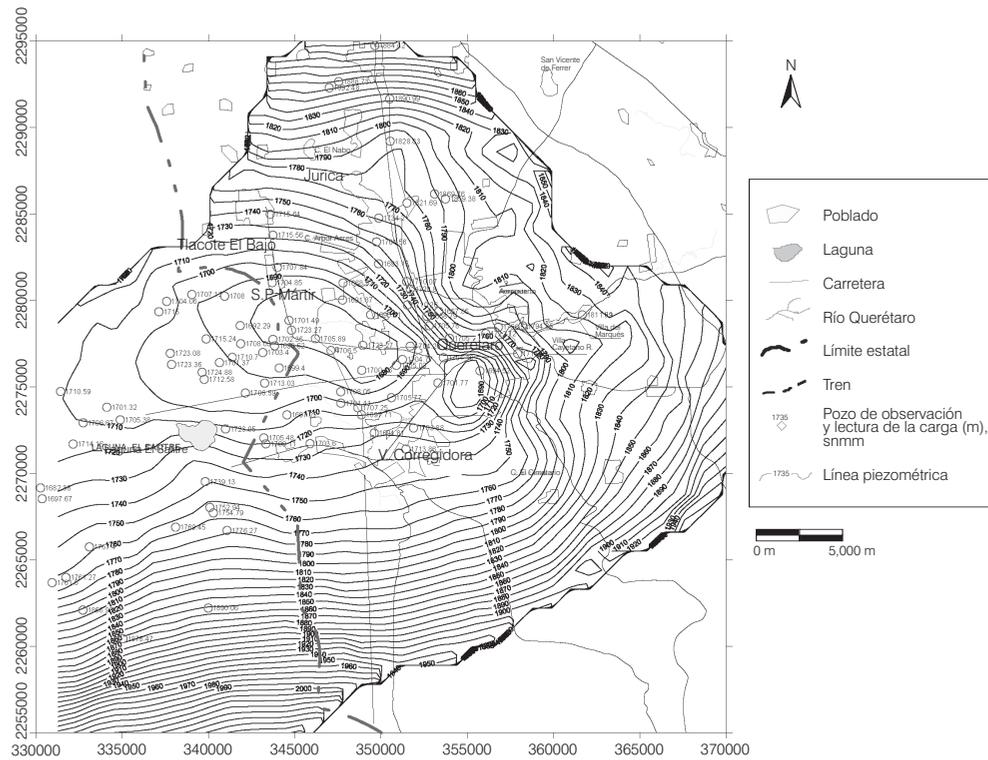


Ilustración 6. Valores de la varianza *krigeado* ordinario, 1995.

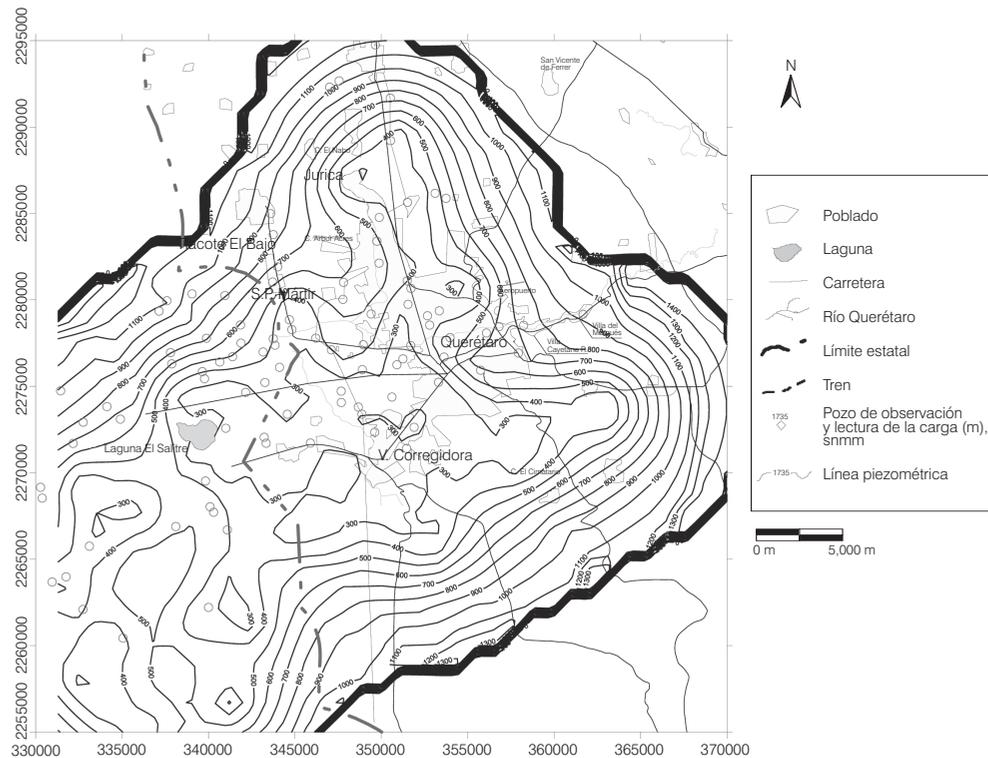


Ilustración 7. Valores de la carga hidráulica *krigeado* ordinario, 1999.

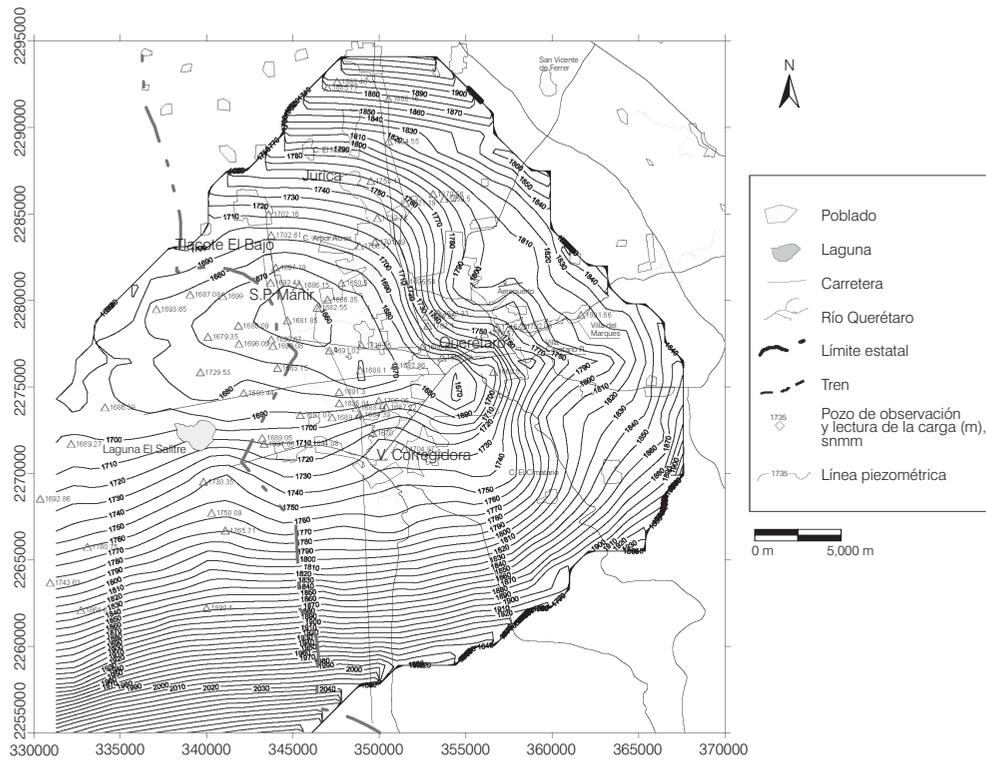
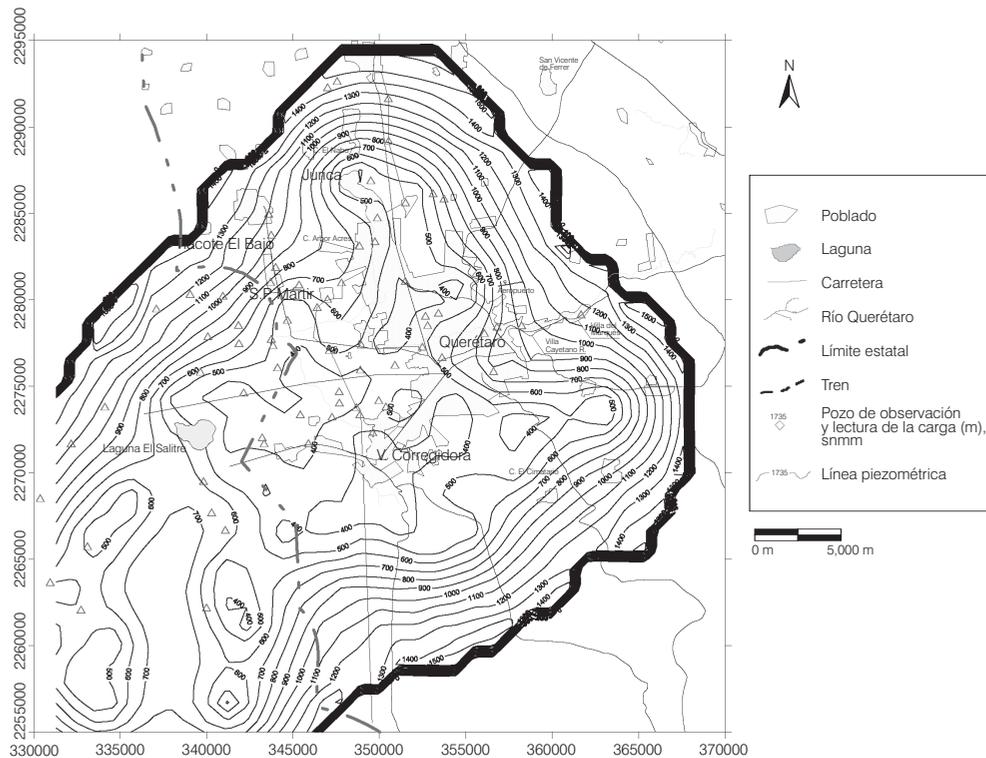


Ilustración 8. Valores de la varianza *krigeado* ordinario, 1999.



La distribución de la varianza para este año se mantiene de forma similar a los años 1993 y 1995, aunque en este año existe un menor número de lecturas (75), la distribución espacial de las observaciones se mantiene de forma homogénea; no existe un agrupamiento de los datos en zonas específicas, por lo que la varianza es parecida en todo el acuífero.

Estimación bivariada, cokrigado

En las ilustraciones 9, 11 y 13 es notable cómo cambia la estimación, manteniéndose las zonas de mayor abatimiento en la zona central, pero se observa una marcada zona que divide el acuífero del valle de Querétaro y el de Obrajuelo, principalmente en el área Laguna el Salitre, donde se tiene un parteaguas subterráneo con dos componentes que fluyen uno al centro del valle y el otro en dirección contraria.

El parteaguas se identifica en los años subsecuentes sin presentarse variaciones considerables.

Los planos de varianza (ilustraciones 10, 12 y 14) muestran cómo ésta ha disminuido en las zonas de Laguna El Salitre, Jurica y Tlacote El Bajo; esto es originado por el incremento de información al hacer la estimación. Lo anterior se discute de forma gráfica en la

siguiente sección. Lo importante es que la varianza calculada con el método de *cokrigado* indica una mejora en la estimación de las zonas de mayor interés y se mantienen los valores altos de varianza en las zonas donde no hay lecturas, que corresponden a los bordes del acuífero.

El método nos permite realizar una estimación con mayor certidumbre en las áreas donde se está interpolando, pero no resulta de la misma manera en las áreas donde se extrapola. Esta es sin duda una limitación del mismo.

Validación cruzada

Con el objetivo de comparar los resultados obtenidos en la estimación univariada y multivariada se procedió a realizar el método de validación cruzada o covalidación para calcular los errores medio (EM); cuadrático medio (ECM), y cuadrático medio estándar (ECME), como proponen Ahmed y De Marsily (1987).

El procedimiento de validación cruzada consiste en retirar un elemento de la muestra y estimar el valor en ese punto usando *krigeado* o *cokrigado*, con el modelo del semivariograma obtenido, para cada uno de los casos. La técnica de validación cruzada nos muestra el valor de

Ilustración 9. Valores de la carga hidráulica *cokrigado* 1993.

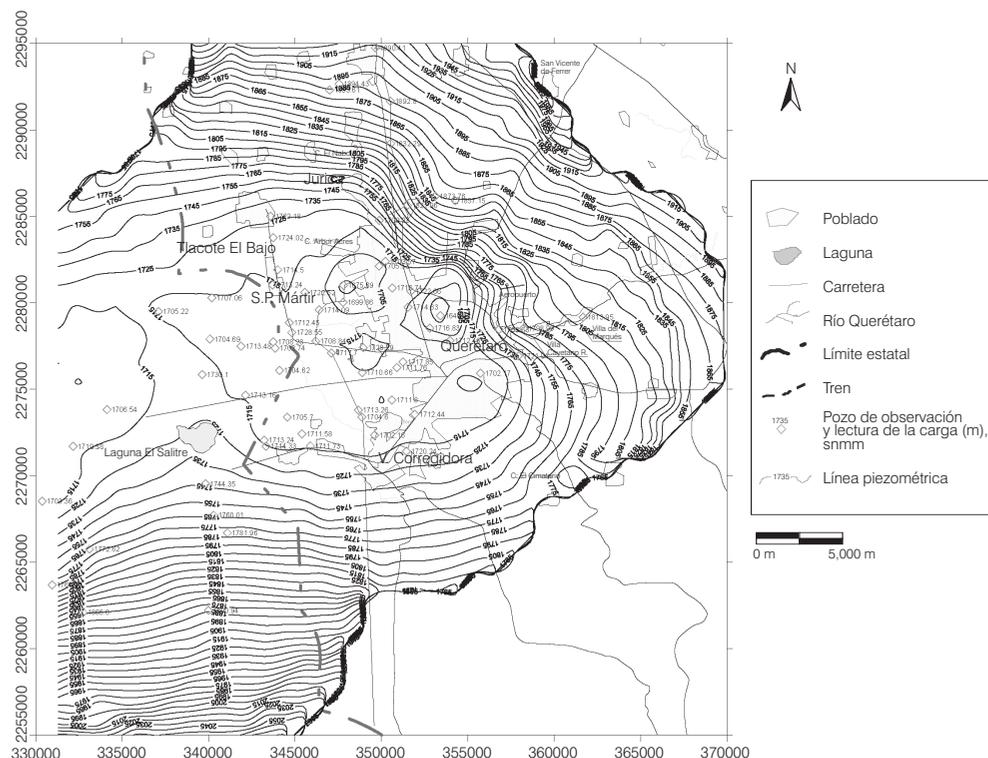


Ilustración 10. Valores de la varianza *cokrigado* 1993.

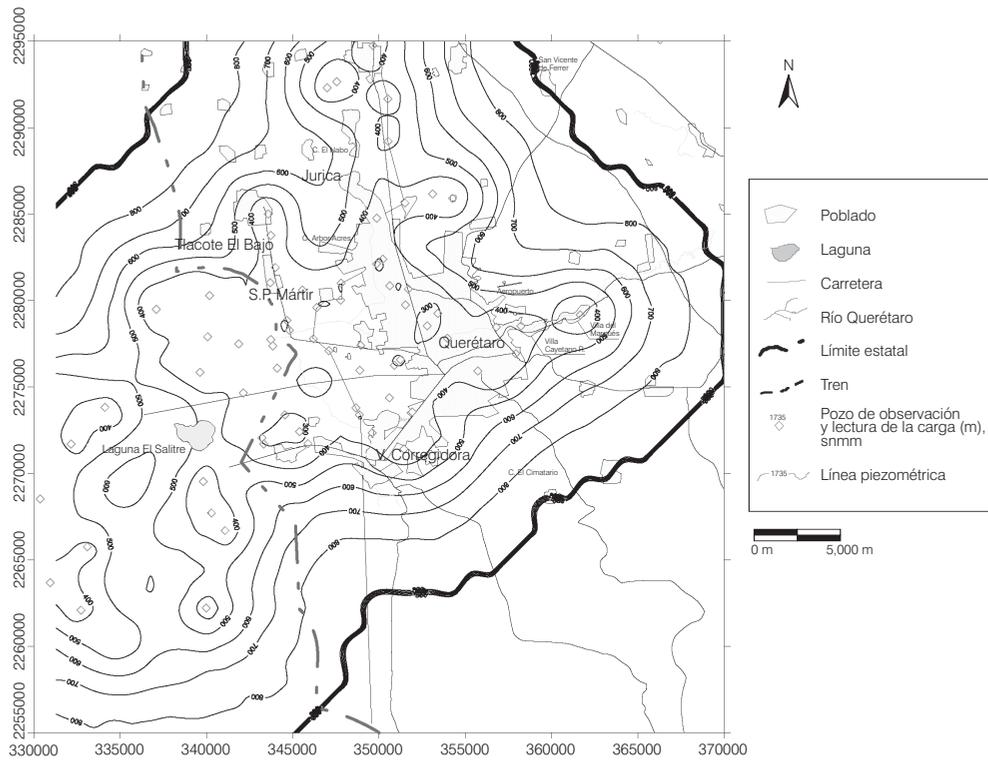


Ilustración 11. Valores de la carga hidráulica *cokrigado* 1995.

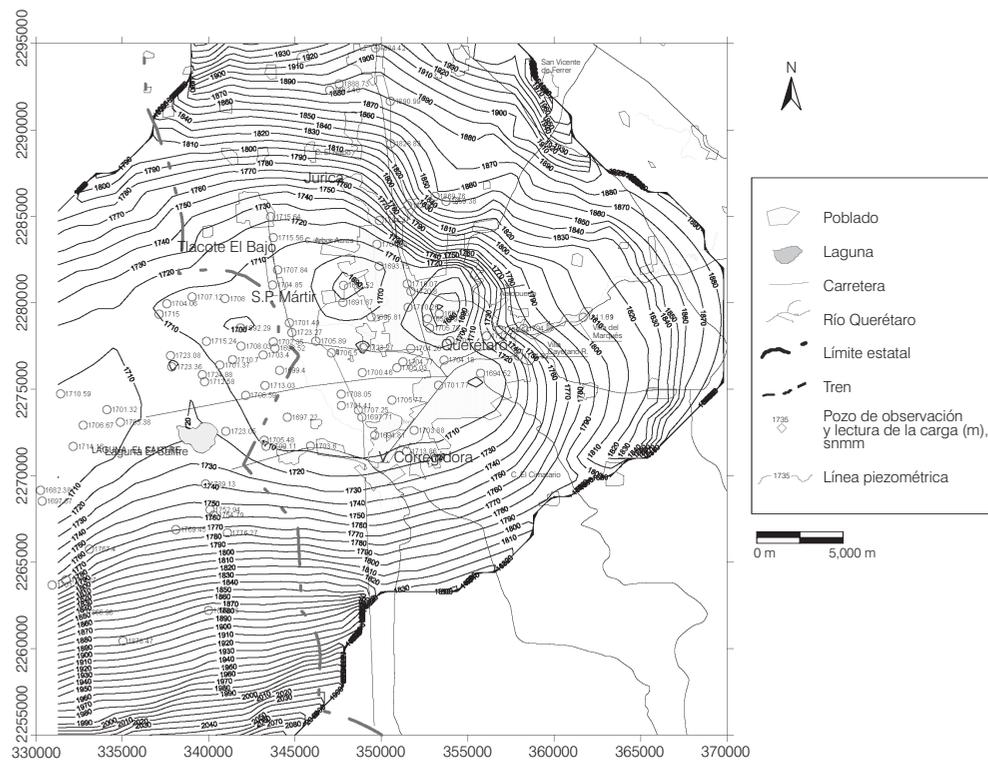


Ilustración 12. Valores de la varianza *cokrigado* 1995.

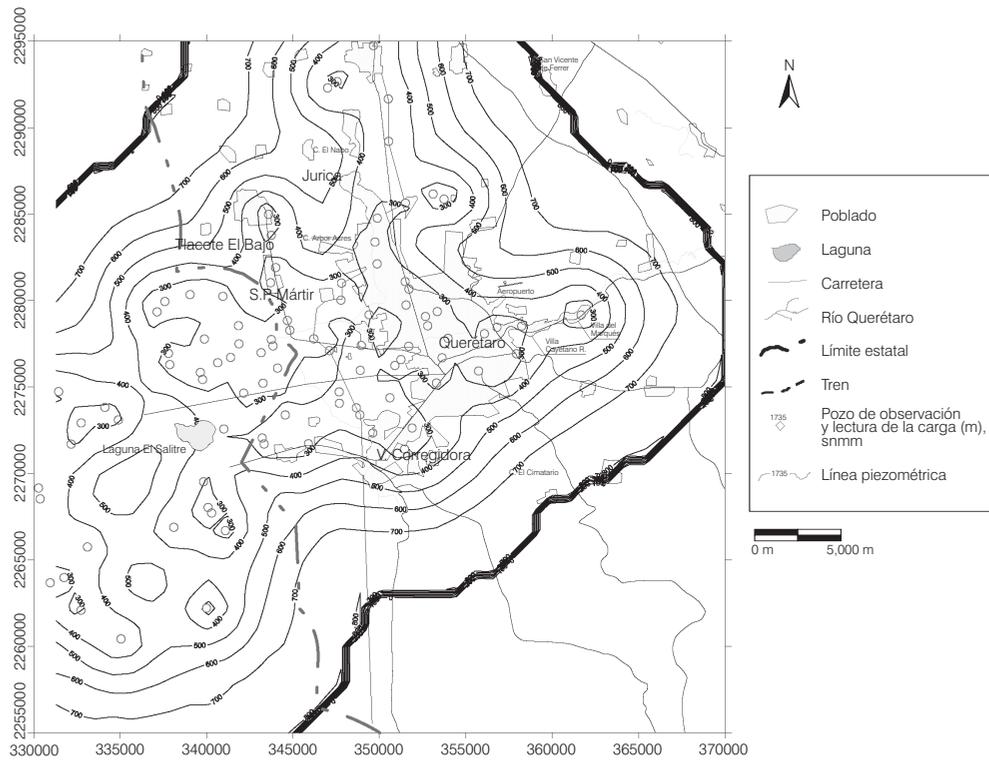


Ilustración 13. Valores de la carga hidráulica *cokrigado* 1999.

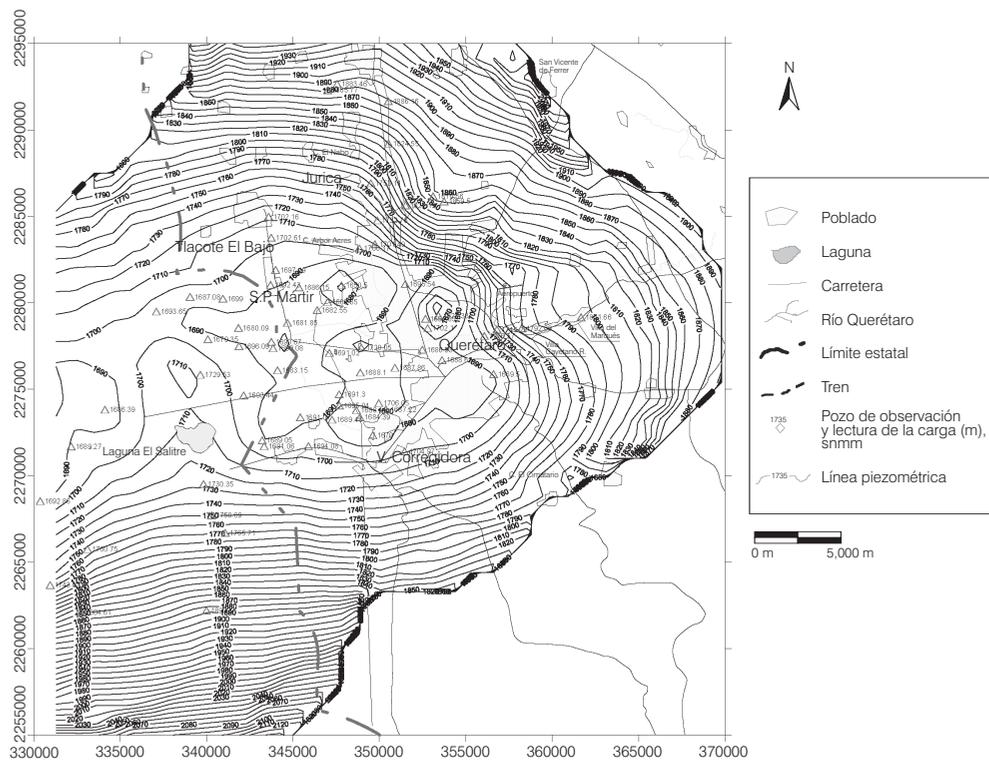
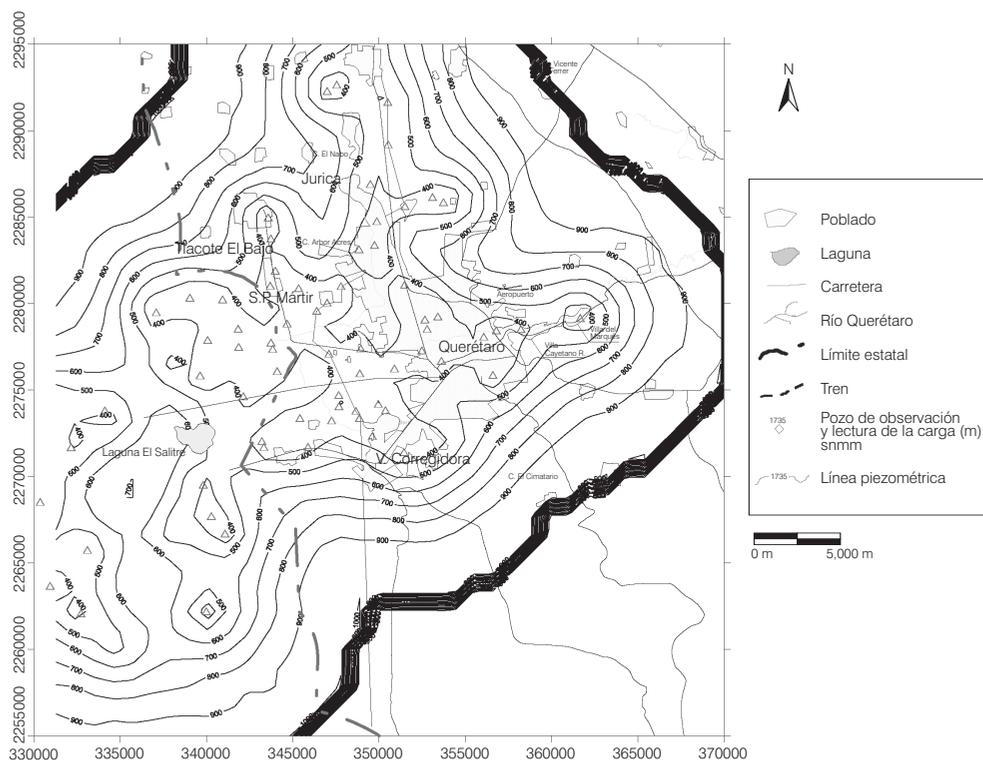


Ilustración 14. Valores de la varianza *cokrigead*o 1999.



la variable real y la estimada con *krigeado* y *cokrigead*o; con estos resultados se obtienen sus diferencias y se calculan los errores mencionados (EM, ECM y ECME). Los criterios para evaluar los modelos son que el EM sea cercano a cero; el ECM sea el menor, y el ECME sea cercano a uno. Los resultados se muestran en el cuadro 2.

Discusión

Existen diferencias significativas en la estimación de la carga hidráulica y el error de estimación utilizando ambas técnicas. De forma gráfica se observa una mejor relación entre las curvas equipotenciales y el valor observado en las ilustraciones del *cokrigead*o (ilustraciones 9, 11 y 13) que en las gráficas del *krigeado* (ilustraciones 3, 5 y 7). Asimismo, el auxiliarse de información en diferentes tiempos permite una representación más suavizada en aquellas áreas donde se carece de información. Las diferencias son notables, sobre todo en la zona Laguna el Salitre, donde se presenta un parteaguas subterráneo que se mantiene en los diferentes años.

En el cuadro 2 se muestran los resultados de la validación cruzada con los métodos de *krigeado* y *cokrigead*o. Respecto a los valores estimados de los residuales utilizando el método de *krigeado*, los valores estimados

se ubican en un rango más pequeño, basta con observar los valores máximos y mínimos para cada caso; de igual forma, el valor medio es cercano a cero y la varianza de los datos no es tan grande como la varianza de los datos utilizando *cokrigead*o.

Observando los valores de los residuales estimados con el método de *cokrigead*o se muestra que las estimaciones se ubican en un rango más amplio y sus varianzas son mayores, comparados con los resultados obtenidos con el método de *krigeado*. Esto indica que los valores estimados con *krigeado* son más suavizados, lo cual está relacionado con los pesos que se asignan a cada observación. Para este método, los pesos asignados a cada observación son parecidos, por lo que el efecto es una estimación suavizada.

Respecto a la estimación con *cokrigead*o se tiene que las estimaciones presentan una mayor variación, por lo que inferimos que los pesos asignados a cada observación son diferentes, reflejándose en la estimación que en su conjunto representa de mejor manera el fenómeno.

Tomando en cuenta los anteriores comentarios se considera pertinente verificar el valor asignado a los pesos. Para llevar esto a efecto se seleccionó el área donde de forma visual se observa un cambio, esta área corresponde a la Laguna El Salitre y se seleccionaron tres

Cuadro 2. Resultados de la validación cruzada.

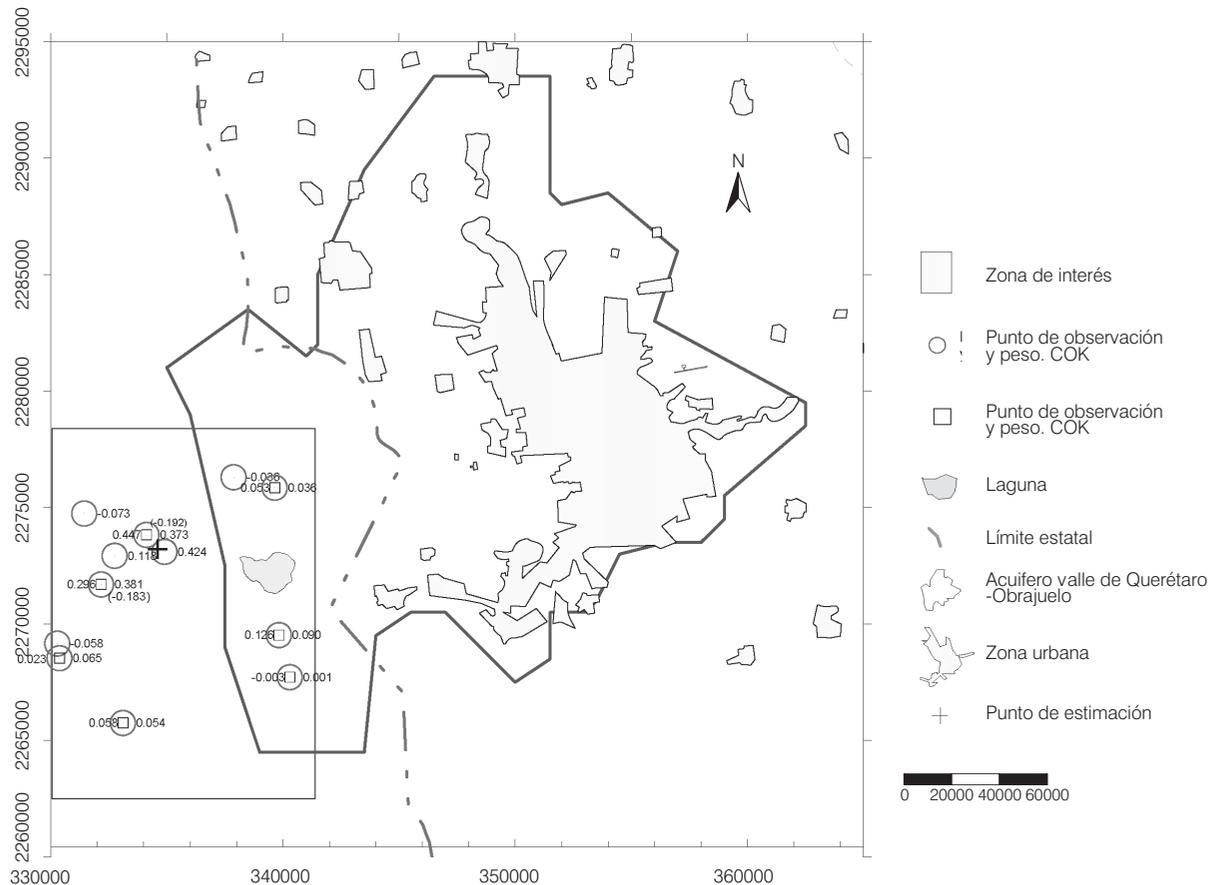
Método y año	Valores estimados de los residuales, con el método de validación cruzada.				Error		
	Mín.	Máx.	μ	σ^2	EM	ECM	ECME
KO, 1993	60.192	60.829	-0.192	418.36	-0.19	390.55	0.82
KO, 1995	-66.762	72.764	-0.098	463.75	-0.10	252.34	0.80
KO, 1999	-57.691	59.654	-0.396	446.104	-0.40	426.3	0.80
Cok., 1993	-89.733	69.223	-0.815	599.513	-0.79	349.51	0.92
Cok., 1995	-89.081	78.396	-0.620	611.010	-1.63	284.31	0.99
Cok., 1999	-85.555	71.300	0.071	723.890	-1.96	309.03	0.95

KO = *krigeado* ordinario; Cok = *cokrigeado*; Mín. = mínimo; Máx. = máximo; μ = media; σ^2 = desviación estándar; EM = error medio,

$$\frac{1}{n} \sum_{i=1}^n (Z(x_i) - Z^*(x_i)), \text{ donde } Z(x) \text{ es el valor observado, } Z^*(x) \text{ es el valor estimado; ECM = error cuadrático medio, } \frac{1}{n} \sum_{i=1}^n (Z(x_i) - Z^*(x_i))^2;$$

$$\text{ECME = error cuadrático medio estándar, } \frac{1}{n} \sum_{i=1}^n (Z(x_i) - Z^*(x_i))^2 / \sigma^2.$$

Ilustración 15. Puntos de observación utilizados con *krigeado* y *cokrigeado*.



puntos de estimación correspondientes a la malla, a los que nos referiremos como nodo (4, 17), (2, 17) y (1, 18), que están próximos a puntos de observación. La verificación se llevó a cabo para ambos métodos, *krigeado* y *cokrigeado*, para el año de 1993.

En la ilustración 15 se muestra el número de puntos de observación que se toman en cuenta al momento de hacer la estimación en los nodos mencionados. El símbolo cuadrado muestra los puntos utilizados con el método de *krigeado*. Observe que estos son siete y el valor

del peso asignado está indicado en su costado izquierdo. Como es de suponer, el valor de mayor peso se encuentra en el punto de observación más cercano (0.447) y disminuye conforme la distancia se incrementa (recordemos que el alcance de la estimación es de 9,000 m). Es fácil notar que los mayores pesos se asignan a los tres puntos más cercanos al punto que se está estimando (al nodo 4, 17).

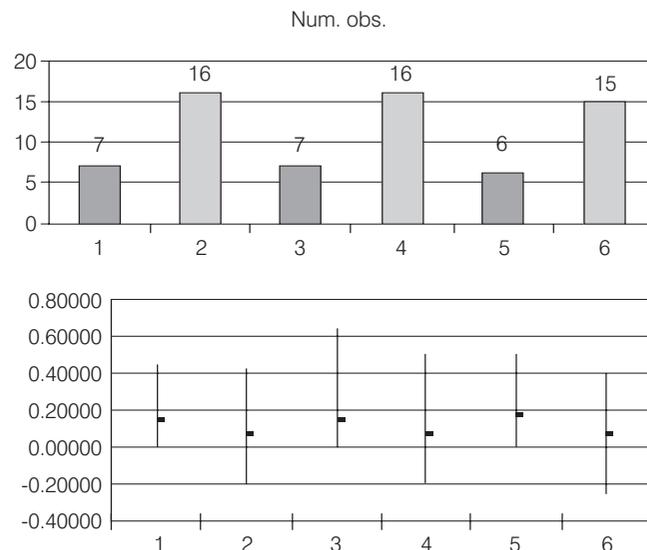
Para el caso del *cokrigado*, observando la ilustración 15 se tiene que el número de puntos de observación es mayor (círculos), de la misma forma que en el *krigeado* el valor de mayor peso es asignado a los puntos más cercanos (los pesos para el *cokrigado* se pueden observar del lado derecho del punto). En la ilustración 16 se nota cómo se han modificado los pesos tanto en valor como en número.

En la ilustración 16a se muestra el número de observaciones utilizadas para estimar el valor en los nodos (4, 17), (2, 17) y (1, 18) utilizando el método de *krigeado* y *cokrigado*. Obsérvese cómo el *krigeado* utiliza menor cantidad de puntos para hacer la estimación; en cuanto al *cokrigado*, se duplican los puntos de observación.

Considerando sólo la información de los pesos se generó la ilustración 16b, en donde se muestran el valor mínimo, máximo y media. Aquí es de destacar que los pesos asignados con el método de *krigeado* se presentan en un menor rango y los pesos asignados con *cokrigado* son mayores. En todos los casos, el valor de la desviación estándar es siempre mayor en el caso de *cokrigado*. Con lo anterior, de forma gráfica se muestra cómo el método de *cokrigado* mejora la estimación, ya que toma en cuenta la información en los diferentes tiempos, lo que se refleja en un mayor número de puntos al momento de efectuar la estimación. Por otro lado, al incrementar las observaciones, los pesos son reasignados manteniendo que el mayor peso es asignado a los puntos más cercanos.

En cuanto al criterio para verificar qué modelo es mejor, se observa en el cuadro 2, en la columna de error, que los estadígrafos indican para ambos casos que los residuos en promedio se subestiman ligeramente, ya que los valores de EM son negativos y cercanos a cero, aunque el *cokrigado* los subestima un poco más que el *krigeado* para los tres años. Sin embargo, el valor de ECM es menor para el método de *cokrigado*, excepto para el año 1995, para el que el ECM aumenta ligeramente, lo que indica que las magnitudes de los residuos tienden a ser menores utilizando *cokrigado*. El índice ECME mide la razón entre errores de medición cuadráticos y la varianza del modelo, por lo que un valor cercano a la unidad indica que la varianza obtenida por el modelo proporciona una buena medida de los errores de la

Ilustración 16. a) Histograma puntos de observación, el orden corresponde a los nodos (4,17), (2,17) y (1,18), las columnas con textura cuadrícula corresponden al método de *krigeado*; b) Grafico de caja "Box-whisker" valores mínimo, máximo y media de los pesos para los nodos (4,17), (2,17) y (1,18).



estimación. De esta forma, en el caso del *krigeado*, la varianza del modelo tiende a sobreestimar a los errores y en el caso del *cokrigado* existe una buena correspondencia entre ambos, por lo que en este último caso la medida del error en la estimación proporcionada por la varianza es confiable.

De los estadígrafos mostrados en el cuadro 2, es claro que los resultados de *cokrigado* indican que los modelos están mejor ajustados en términos de error y la varianza del modelo. El uso de la información de la carga hidráulica en los diferentes tiempos ayuda en la mejora de la estimación de la carga hidráulica. Esto es explicado desde el punto de vista de un aumento del número de observaciones, ya que de forma integrada se toma en cuenta la información obtenida en todos los años. Como referencia se comenta que el número de observaciones utilizadas con el método de *krigeado*, específicamente para la zona Laguna Salitre, es de siete puntos. Utilizando el método de *cokrigado*, el número de puntos para la estimación se duplica (16), como se muestra en la ilustración 16; desde luego que el número de observaciones varía conforme cambia la posición de la estimación.

En aquellas zonas donde se carece de información, como es el caso de la Laguna el Salitre, el *cokrigado* incrementa el número de puntos al momento de hacer la estimación, ya que integra la información de los diferentes tiempos.

Estos resultados indican claramente que el método multivariado ofrece los mejores resultados. También es importante destacar que al estimar la carga hidráulica utilizando toda la información, la estimación es vista como un proceso espacio-temporal que marca dichos cambios, por lo que se considera como un beneficio estimar de forma conjunta las observaciones para los tres años, lo que se observa en aquellas áreas donde se tiene menor número de pozos de observación, como son en los años 1993 y 1999.

Conclusión

Del análisis anterior se concluye que en el caso del acuífero valle de Querétaro-Obrajuelo puede ser muy útil estimar la carga hidráulica en un año dado, utilizando como variables auxiliares los niveles del agua de los años más cercanos en tiempo que se estén analizando, lo cual ha sido posible al considerar a las cargas hidráulicas como variables espaciales que están correlacionadas en el tiempo. Este punto de vista es potencialmente útil para hacer las estimaciones en diferentes tiempos.

Una desventaja es que el aumento del número de variables correlacionadas en el tiempo requiere de calcular $n(n+1)/2$ semivariogramas simples y cruzados, lo que vuelve que la técnica se torne tediosa y, por el número de variables consideradas, en algunos casos será poco práctica.

Es importante hacer notar que al utilizar el punto de vista multivariado, la continuidad espacio-tiempo es modelada vía un modelo lineal de correogionalización (MLC) y sólo se modelizan planos en tiempo determinados. Sin embargo, a pesar de que este punto de vista no proporciona la posibilidad de utilizar los datos disponibles en todos los tiempos, los resultados presentados muestran claramente cómo el utilizar la autocorrelación en forma conjunta con la correlación cruzada beneficia la estimación en aquellos puntos donde no se tiene información de la variable principal, ya que la variable secundaria o variables secundarias aportan la información que no es aportada por la variable principal.

Este tipo de métodos pueden ser muy útiles en aplicaciones prácticas, sin embargo el juicio de un hidrogeólogo es indispensable al analizar la pertinencia de los resultados, ya que aunque los datos en los diferentes años presenten una buena correlación, las condiciones hidrodinámicas del acuífero pudieran haber sido modificadas en el tiempo en algunas zonas, por lo que las estimaciones en las mismas pudieran no corresponder con la realidad.

Agradecimientos

A las comisiones Estatal de Aguas del Estado de Querétaro, Estatal del Agua del Estado de Guanajuato y Nacional del Agua, que proporcionaron los datos de piezometría del valle de Querétaro-Obrajuelo.

Al Conacyt, por otorgar la beca crédito para los estudios de doctorado del primer autor.

Recibido: 30/08/2005
Aprobado: 23/05/2006

Referencias

- ABOUFIRASSI, M. y MARIÑO, M.A. Cokriging of aquifer transmissivity and specific capacity. *Mathematical Geology*. Vol. 16, núm. 1, 1984, pp.19-35.
- AHMED, S. y DE MARSILY, G., Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. *Water Resources Research*. Vol. 23, núm. 9, 1987, pp.1717-1737.
- CHRISTAKOS, G. On the problem of permissible covariance and variogram models. *Water Resources Research*. Vol. 20, núm. 2, 1984, pp. 251-265.
- CRESSIE, N.A. Fitting variogram models by weighted least squares. *Mathematical Geology*. Vol. 17, núm. 5, 1985, pp. 563-581.
- CRESSIE, N.A. Statistics for spatial data. New York: Wiley, 1991, 900 pp.
- DELHOMME, J.P. Kriging in the hydrosociences. *Advances in Water Resources*. Vol. 1, núm. 5, 1978, pp. 251-266.
- DEUTSCH, C.V. y JOURNEL, A.G. GSLIB Geostatistical Software Library and User's Guide. 2nd ed. New York: Oxford University Press, 1998, 369 pp.
- GAMBOLATI, G. y VOLPI, G. Groundwater contour mapping in venice by stochastic interpolators. Part 1. Theory. *Water Resources Research*. Vol. 15, núm. 2, 1979, pp. 281-297.
- GOOVAERTS, P. Geostatistics natural resources evaluation. New York: Oxford University Press, 1997, 483 pp.
- GOOVAERTS, P. Ordinary cokriging revisited. *Mathematical Geology*. Vol. 30, núm. 1, 1998, pp. 21-42.
- GOOVAERTS, P. Geostatistics in soil science: state of the art and perspectives. *Water Resources Research*. Vol. 22, núm. 2, 1999, pp. 95-108.
- GOULARD, M. y VOLTZ, M. Linear coreogionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*. Vol. 24, núm. 3, 1992, pp. 269-286.
- GUYSA. *Geofísica de exploraciones, estudio de simulación hidrodinámica y diseño óptimo de la red de observación en los acuíferos de Aguascalientes y Querétaro*. Tomo III. Querétaro, México: Comisión Nacional del Agua, Gerencia de Aguas Subterráneas, 1996.

- JOURNEL, A.G. y HUIJBREGTS, CH.J. *Mining geostatistics, seventh printing*. Londres: Academic Press, 1978, 600 pp.
- KYRIAKIDIS, C. y JOURNEL G. Geostatistical space-time models: a review. *Mathematical Geology*. Vol. 31, núm. 6, 1999, pp. 651-684.
- MYERS, D.E. Matrix formulation of co-kriging. *Mathematical Geology*. Vol. 14, núm. 3, 1982, pp. 249-257.
- OLEA, R.A. (editor). Geostatistical glossary and multilingual dictionary (auspices of International Association for Mathematical Geology, 1984-1989 committee on Geostatistics. (CHRISTAKOS, G., CLARK, I. DAVID, M., JOURNEL, A.G., KRIGE, D.G., OLEA, R.A.). Oxford University Press. *Studies in Mathematical Geology*. Núm. 3, 1991, 177 pp.
- PAPRITZ, A. y FLÜHLER, H. Temporal change of spatial auto correlated soil properties: optimal estimation by cokriging. *Geoderma*. Vol. 62, 1994, pp. 29-43.
- ROUHANI, S., y HALL, T., Space-time kriging of groundwater data. En Armstrong, M. (editor). *Geostatistics*. Vol. 2. Kluwer Academic Publ., Dordrecht, 1989, pp. 639-650.
- ROUHANI, S. y WACKERNAGEL, H. Multivariate geostatistical approach to space-time data analysis. *Water Resources Research*. Vol. 26, núm. 4, 1990, pp. 585-591.
- SAMPER, F.J. y CARRERA, J. *Geoestadística, aplicaciones a la hidrogeología subterránea*. Barcelona: Centro Internacional de Métodos Numéricos en Ingeniería, 1990, 484 pp.
- SIMUTA-CHAMPO, R. *Modelo en elemento finito para el flujo del acuífero del valle de Querétaro*. Tesis de maestría. Universidad Nacional Autónoma de México, 2005.
- SOLOW, R. y GORELICK, M. Estimating monthly stream flow values by cokriging. *Mathematical Geology*. Vol. 18, núm. 8, 1986, pp. 785-809.
- TOUGH, J.G. y LEYSHON, P.R. SPHINX a program to fit the spherical and exponential models to experimental semi-variograms. *Computers & Geosciences*. Vol. 11, núm. 1, 1985, pp. 95-99.
- VOLPI, G. y GAMBOLATI, G. Groundwater contour mapping in venice by stochastic interpolators 2. *Results, Water Resources Research*. vol. 15, núm. 2, 1979, pp. 291-297.
- WACKERNAGEL, H. Description of a computer program for analyzing multivariate spatially distributed data, *Computers & Geosciences*. Vol. 15, núm. 4, 1989, pp.593-598.
- WACKERNAGEL, H. *Multivariate Geostatistics*. Second ed. Berlín: Springer, 1995, 257 pp.

Abstract

MENDOZA-CÁZARES, E.Y. & HERRERA-ZAMARRÓN, G. *Multivariate space-time estimation of hydraulic head in Querétaro-Obrajuelo Valley*. Hydraulic engineering in Mexico (in Spanish). Vol. XXII, no. 1, January-March, 2007, pp. 63-80.

In the present work a space-time method for multivariate estimation was used, considering the hydraulic head in different years as different variables but correlated in time. This type of method allows to increase the information used in the estimation of the hydraulic head for every year. The method was applied to estimate the hydraulic head in three different years in the aquifer Valle de Querétaro-Obrajuelo, Mexico, obtaining good results for cross validation and in the sense that when using the information of the three years jointly, some characteristics of the piezometric head configurations are clearly observed that are not when only the data of the considered year are used.

Keywords: *multivariate analysis, geostatistics, cokriging, space-time, estimation hydraulic head.*

Dirección institucional de los autores:

M.C. Edgar Yuri Mendoza-Cázares

División de Estudios de Posgrado de la Facultad de Ingeniería,
UNAM campus Morelos,
Paseo Cuauháhuac 8532, col. Progreso, Jiutepec, Morelos,
teléfono: + (52) (777) 329 3600, ext. 339,
fax: + (52) (777) 329 3682,
edgaryuri@hotmail.com

Dra. Graciela Herrera-Zamarrón

Especialista en Hidráulica.
Instituto Mexicano de Tecnología del Agua,
Paseo Cuauháhuac 8532, col. Progreso, Jiutepec, Morelos,
teléfono: + (52) (777) 329 3600, ext. 339,
fax: + (52) (777) 329 3682,
gherrera@tlaloc.imta.mx