

ANÁLISIS ESPACIAL, HACIA UNA UTILIZACIÓN MEJORADA DE LA INFORMACIÓN MEDIDA EN CAMPO Y POR SATÉLITES, PARA APOYAR LA TOMA DE DECISIONES EN MATERIA HÍDRICA Y AMBIENTAL

• Felipe Omar Tapia-Silva •

Universidad Autónoma Metropolitana, México

Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo" A.C., México

• José Luis Silván-Cárdenas • Edgar Rosales-Arriaga •

Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo" A.C., México

Resumen

La información sobre clima, suelos y agua disponible en México —medida de forma puntual en estaciones terrestres— no cumple con los criterios de representatividad adecuados que permitan la generación de datos, información y conocimiento, para apoyar una mejor toma de decisiones. Los sensores remotos (aerotransportados y plataformas satelitales) constituyen de manera potencial un excelente recurso de información para complementar las mediciones puntuales en tierra. Sin embargo, la información obtenida debe ser validada y complementada por mediciones en campo. Además, entre las limitaciones de este recurso de información se pueden mencionar aspectos relacionados con su resolución temporal y espacial, y su alta susceptibilidad a condiciones climáticas del momento de la toma para el caso de los satélites de mayor uso y disponibilidad (los ópticos). Las técnicas de análisis espacial permiten mejorar la representatividad espacial de los datos provenientes de campo y de satélites, al posibilitar la generación de esquemas de obtención de superficies (mapeo) de la información disponible, incluso mediante el uso de información adicional. Las técnicas expuestas en este artículo y aplicadas a la interpolación de datos para casos prácticos ejemplifican lo mencionado. Estos métodos no se han investigado de forma suficiente y menos puestos en operación en el proceso de disseminación de la información geoespacial básica en México. Resulta fundamental demostrar la utilidad práctica de este tipo de estudios, a fin de generar mejor información y conocimientos básicos para la toma de decisiones en materia de agua y de medio ambiente en general.

Palabras clave: análisis espacial, interpolación, kriging, percepción remota, mediciones en campo, agua.

Introducción

Las actividades de modelación hidrológica y las de previsión de acciones relacionadas con el manejo del agua y demás recursos conectados requieren de información confiable sobre clima, suelos y agua. Considerando que las tendencias climáticas regionales no pueden deducirse de los registros de un solo sitio, incluso en un terreno relativamente homogéneo (Pielke *et al.*, 2000), es

necesario considerar más puntos de medición de campo, a fin de capturar la variabilidad espacial de este parámetro a escala regional. Según WMO (2006), la representatividad de una observación es el grado en el que describe de manera adecuada el valor de la variable necesaria para un propósito específico. WMO (2006) también indica que lograr una buena ubicación de las estaciones para toma de datos es difícil. Sin embargo, esto no siempre es posible por

aspectos de corte técnico y económico. Acerca de este tema, Daly (2006) señala que el espacio de 100 km entre estaciones probablemente sea insuficiente para representar los patrones climáticos causados por los factores que influyen el clima (como la altitud y las zonas costeras). Este mismo autor indica que una estación en una zona montañosa o costera es probable que sólo sea representativa en la escala local (menos de 3 km).

Desafortunadamente, la situación en México en cuanto a representatividad de las estaciones en campo es desfavorable. La información medida en forma puntual en estaciones terrestres no cumple con los criterios de representatividad adecuados que permitan la generación de datos, información y conocimiento para apoyar una mejor toma de decisiones. Como ejemplo tenemos el caso de las estaciones del Servicio Meteorológico Nacional (SMN), que constituyen aproximadamente 3 300 sitios de observación, entre los que se encuentran observatorios, estaciones climatológicas convencionales y automáticas. Considerando que cada dato sea representativo de un área de 100 km², tal como se asume en zonas planas, en total se cubre un área de 330 000 km², que de acuerdo con datos de INEGI sobre la extensión territorial de México representan tan sólo un 17%.

Los sensores remotos (aerotransportados y plataformas satelitales) potencialmente constituyen un excelente recurso de información para complementar las mediciones puntuales en tierra. Sin embargo, la información obtenida primero debe ser validada e igualmente complementada mediante mediciones en campo. Además, entre las limitaciones de este recurso de información se pueden mencionar aspectos relacionados con su resolución temporal (el tiempo de obtención de una imagen siguiente de la misma zona), su resolución espacial y la alta susceptibilidad a las condiciones climáticas del momento de la toma para el caso de los satélites de mayor uso (los ópticos, como LANDSAT, SPOT y MODIS). A estos aspectos se debe que no sea poco frecuente observar

una limitada disponibilidad de imágenes para una zona determinada en un periodo de análisis específico y que se requiera recurrir a información obtenida en campo.

El presente artículo tiene la intención de mostrar cómo el análisis espacial puede proveer de valiosas herramientas conceptuales y prácticas para lograr una utilización mejorada de la información medida en campo y por satélites para apoyar la toma de decisiones en materia de agua. El documento está compuesto de dos secciones fundamentales: la primera refiere a aspectos teóricos y la segunda a una discusión de los resultados correspondientes a aplicaciones de las técnicas de análisis espacial seleccionadas en temas específicos de relevancia hidrológica. El resultado primordial de cada una de estas técnicas son mapas con superficies continuas de cada uno de los temas o variables de interés, tales como precipitación, permeabilidad de suelos, temperatura del aire y niveles estáticos de un acuífero. Entre las variables auxiliares utilizadas están modelos de elevación provenientes de misiones como SRTM (*Shuttle Radar Topography Mission*) y estimaciones de temperatura superficial, e índices de vegetación del satélite MODIS. El artículo termina con una sección de conclusiones respecto a las facilidades y limitaciones que el análisis espacial ofrece para mejorar la calidad de la información proveniente de mediciones en campo y de satélite.

Marco teórico

Análisis espacial puede ser definido en forma generalizada como el estudio cuantitativo de fenómenos localizados en el espacio (Bailey y Gatrell, 1995). Estos mismos autores aportan una definición más extensa y precisa del ámbito de acción del análisis espacial: situaciones en donde existen datos observacionales de algún proceso operando en el espacio, y en las que se requieren métodos para describir o explicar el comportamiento del proceso y para relacionarlo con otros fenómenos espaciales: el objetivo consiste en incrementar el conocimiento

básico del proceso, establecer evidencias que favorezcan hipótesis o posibilitar la predicción de valores en áreas donde no se han realizado observaciones.

De acuerdo con lo anterior, las técnicas de análisis espacial permiten incrementar el potencial de uso en cuanto a representatividad espacial y contenido de información de los datos provenientes de campo y de satélites. Estas técnicas posibilitan la generación de esquemas de generación de superficies (mapeo) de la información disponible, incluso mediante el uso de información adicional, que aportan valor a la misma información original y mejoran sus cualidades de representatividad espacial. En este proceso, la primera pregunta que puede surgir es qué herramienta seleccionar de las muchas que hay en el ámbito del análisis espacial. Al respecto, Hengl (2009) aporta un árbol de decisiones que puede ser utilizado para tal efecto y expone la manera de interpretarlo: es necesario definir si existe un modelo determinístico de la variable de interés; en caso negativo, se requiere correlacionar las variables definidas con factores medioambientales. Si éstos están significativamente correlacionados, es posible ajustar un modelo de regresión lineal múltiple y analizar si los residuos muestran autocorrelación espacial. Si no es así, es posible hacer una estimación tipo OLS (*Ordinary Least Squares*) de los coeficientes. En caso contrario, se puede llevar a cabo regresión kriging (RK). Si los datos no muestran correlación con factores medioambientales, es posible analizar el variograma de la variable de interés. Si es posible ajustar un variograma, entonces se puede usar kriging ordinario (KO). En caso contrario, se puede utilizar un interpolador mecánico como el inverso de la distancia. Si el variograma de la variable de interés no muestra autocorrelación espacial y no hay correlación con variables medioambientales, entonces sólo es posible estimar una media global para toda el área, que sería poco útil para mostrar la variabilidad espacial del parámetro de interés.

Los casos que se ejemplifican en este artículo y que no están mostrados en el árbol de decisiones antes mencionado son kriging con deriva externa (KED) y regresión geográficamente pesada (GWR). Como Hengl (2009) lo menciona, estos métodos pueden ser generalizados por RK y por tanto pueden aplicarse para mapear variables en los casos que esta técnica de análisis espacial sea pertinente de acuerdo con lo expuesto anteriormente. Los que sí están incluidos son RK y KO. En general, la mayor parte de los métodos (incluidos los analizados en este artículo) no han sido lo suficientemente investigados y menos puestos en operación en el proceso de diseminación de la información geoespacial básica en México. Resulta entonces fundamental demostrar la utilidad práctica de este tipo de estudios para generar mejor información y conocimiento básicos para la toma de decisiones en materia de agua y de medio ambiente en general.

Kriging ordinario (KO)

La mayoría de los procesos hidrológicos presentan variaciones espaciales, algunas más estructuradas que otras. Las variaciones no estructuradas son modeladas mediante procesos estocásticos no estacionarios, *e.g.*, con media variable. La estimación mediante KO permite tomar en cuenta las variaciones locales de la media, limitando el dominio de su estacionaridad a un ámbito local W alrededor de la posición, donde se pretende estimar la variable.

Sea $Z(x) = Y(x) + m(x)$ un proceso estocástico con media variable determinada por $m(x)$ y función de covarianza $C(h)$. Como tal, $Y(x)$ es un proceso estocástico con media nula. Un estimador lineal es una combinación lineal de mediciones $Z(x_1), Z(x_2), \dots, Z(x_n)$ en las posiciones $x_1, x_2, \dots, x_n \in \Omega$. Específicamente:

$$\hat{Y}(x) = \sum_{k=1}^n \lambda_k(x) Y(x_k)$$

o bien:

$$\hat{Z}(x) = m(x) + \sum_{k=1}^n \lambda_k(x) [Z(x_k) - m(x_k)]$$

Si la media es constante en el dominio Ω , entonces puede eliminarse de la ecuación anterior, forzando que los pesos kriging λ_k sumen la unidad. En tal caso, el estimador se denomina KO y se expresa como:

$$\hat{Z}_{KO}(x) = \sum_{k=1}^n \lambda_k(x) Z(x_k)$$

con:

$$\sum_{k=1}^n \lambda_k(x) = 1$$

Los pesos óptimos que minimizan la varianza del error de estimación se obtienen mediante el método de multiplicadores de Lagrange (Goovaerts, 1997), el cual resulta en el siguiente sistema de ecuaciones:

$$\begin{cases} \sum_{k=1}^n \lambda_k(x) C(x_j - x_k) + \mu(x) = C(x_j - x), & j = 1, \dots, n \\ \sum_{k=1}^n \lambda_k(x) = 1 \end{cases}$$

donde μ denota el multiplicador de Lagrange. Alternativamente, si se considera la relación entre la función de covarianza y la función de semivariograma $\gamma(h)$, i.e., $C(h) = C(0) - \gamma(h)$, el mismo sistema se puede escribir como:

$$\begin{cases} \sum_{k=1}^n \lambda_k(x) \gamma(x_j - x_k) + \mu(x) = \gamma(x_j - x), & j = 1, \dots, n \\ \sum_{k=1}^n \lambda_k(x) = 1 \end{cases}$$

Kriging con deriva externa (KED)

La estimación KO presentada anteriormente no incorpora información externa y sólo explota la correlación espacial. KED es una extensión de kriging con modelo de tendencia o (universal), que integra condiciones de

universalidad suplementarias relativas a una o varias variables externas $W_j(x)$, para $j = 1, \dots, N$, medidas de forma exhaustiva en el dominio donde se desea estimar variable de interés. De esta forma, el estimador KED proporciona una forma relativamente simple para modelar un proceso estocástico como función de otras variables secundarias (externas) conocidas en todo el espacio de estimación. La(s) variable(s) secundaria(s) $W_j(x)$ debe(n) mostrar un comportamiento lineal respecto a la variable de interés o primaria $Z(x)$.

En este caso, la media se modela como una función lineal de variables $W_1, W_2, \dots, W_m$, que varían suavemente en el dominio de interés Ω . Formalmente:

$$Z(x) = Y(x) + \sum_{j=1}^m a_j W_j(x)$$

Para el caso de un modelo lineal con un término independiente, se puede asumir en lo sucesivo $W_m = 1$. Usando el modelo anterior, el estimador lineal basado en las observaciones $Z(x_k)$, $k = 1, \dots, n$, se expresa como:

$$\hat{Z}(x) = \sum_{j=1}^m a_j W_j(x) + \sum_{k=1}^n \lambda_k(x) \left[Z(x_k) - \sum_{j=1}^m a_j W_j(x_k) \right]$$

Las constantes $a_1, a_2, \dots, a_m$ del modelo de deriva se eliminan de la ecuación al imponer la siguiente condición en los pesos kriging:

$$\sum_{k=1}^n \lambda_k(x) W_j(x_k) = W_j(x)$$

Considerando esta condición, el estimador KED se expresa simplemente como una combinación lineal de las observaciones, es decir:

$$\hat{Z}_{KED}(x) = \sum_{k=1}^n \lambda_k(x) Z(x_k)$$

En este caso, los pesos óptimos se obtienen al resolver el siguiente sistema de ecuaciones lineales (Goovaerts, 1997):

$$\left\{ \begin{array}{l} \sum_{k=1}^n \lambda_k(x) \gamma(x_i - x_k) + \sum_{j=1}^m \mu_j(x) W_j(x_k) = \gamma(x_j - x), \\ i = 1, \dots, n \\ \sum_{k=1}^n \lambda_k(x) = 1 \\ \sum_{k=1}^n \lambda_k(x) W_j(x_k) = W_j(x) \end{array} \right.$$

Donde $\mu_1, \mu_2, \dots, \mu_m$ son los multiplicadores de Lagrange, que resultan de minimizar $\text{Var}\{\hat{Z}_{KDE}(x) - Z(x)\}$, la varianza de los errores de predicción.

Regresión geográficamente pesada (GWR)

En los análisis de regresión, la influencia de algunas variables se ven afectadas por su ubicación geográfica y por lo tanto es necesario reemplazar el modelo de regresión espacial estacionario con uno no estacionario (Atkinson et al., 2003). Los efectos de ignorar el espacio en el uso de modelos de regresión han sido estudiados por algunos autores como Lichstein et al. (2002). Entre estos efectos, se puede mencionar que los residuales de la regresión por mínimos cuadrados ordinarios (OLS) pueden dar lugar a modelos de autocorrelación, lo que puede indicar que el supuesto básico de los errores independientes se viola. En este caso, el modelo OLS no es adecuado para generar las conclusiones correctas acerca del proceso estudiado.

GWR (Brunsdon et al., 1998; Fotheringham et al., 2002) permite usar tanto parámetros locales como globales, por lo que el modelo original OLS puede ser rescrito como:

$$y_i = \beta_{i0} + \sum_{k=1}^p \beta_{ik} X_{ik} + \varepsilon_i, \quad i = 1, \dots, n$$

donde y_i representa la variable dependiente en la posición i ; β_{i0} representa el parámetro del intersección en la posición i ; p representa el

número de variables independientes (predictivas o explicativas); β_{ik} representa el coeficiente de regresión local para la k -ésima variable independiente en la posición i ; X_{ik} es el valor de la k -ésima variable independiente en la posición i , y ε_i representa un término de error (también conocido como perturbación aleatoria) en la posición i .

En la calibración se asume que los datos observados cerca del punto i tienen más influencia en la estimación de β_{ik} que los datos más lejanos del punto i . En lugar de calibrar una sola ecuación de regresión, GWR genera una ecuación de regresión por separado para cada observación. En esencia, la ecuación mide la relación inherente en el modelo alrededor de cada punto i (Fotheringham, 2000).

El cálculo de la matriz para la estimación de los coeficientes de regresión es:

$$\hat{\beta}_i = [X^T \cdot W(i) \cdot X]^{-1} X^T \cdot W(i) \cdot y$$

donde $W(i) = \text{diag}[w_1(i), \dots, w_n(i)]$ es la diagonal de pesos de la matriz que varía en función de cualquier calibración o predicción localizada en i ; X es la matriz de variables exógenas con una primera columna para los primeros interceptos; y es el vector de variables dependientes y $\hat{\beta}_i = (\hat{\beta}_{i0}, \hat{\beta}_{i1}, \dots, \hat{\beta}_{ip})^T$ es el vector $p + 1$ del coeficiente de regresión local en la posición i .

Una función de kernel local se utiliza para especificar la matriz de pesos. En el modelo, esta función tiene el efecto de acortar las distancias de la calibración n de la predicción localizada en i , de acuerdo con Wheeler y Tiefelsdorf (2005); una de las funciones de kernel más comúnmente usada es la denominada *bi-square nearest neighbor function* (Fotheringham et al., 2002):

$$w(i) = \begin{cases} \left[1 - (d_{ij}/b)^2\right]^2 & \text{si } j \in \{N_i\} \\ 0 & \text{en caso contrario} \end{cases}$$

donde d_{ij} representa la distancia entre la calibración de la posición j y la localización de la predicción i ; b representa el umbral de la

distancia para el n -ésimo vecino más cercano y $\{N_i\}$ representa las observaciones que están dentro del rango de distancia del n -ésimo vecino más cercano.

La identificación del mejor modelo (o la mejor combinación de predictores) puede mejorarse considerando los coeficientes de determinación ajustados (R^2) y el coeficiente de correlación R , que son calculados localmente en GWR.

Regresión kriging (RK)

RK es una técnica genérica de mapeo que puede ser considerada como la mejor predicción lineal no desviada (BLUP: *Best Linear Unbiased Prediction*) para datos espaciales y de la cual todas las demás técnicas lineales pueden ser derivadas (Hengl, 2009). RK puede ser escrita en notación de matrices como:

$$\hat{Z}_{RK}(x_o) = q_o^T \cdot \hat{\beta}_{GLS} + \lambda_o^T \cdot (Z - q \cdot \hat{\beta}_{GLS})$$

Donde $\hat{Z}_{RK}(x_o)$ es el valor en el sitio x_o ; q_o es el vector de $p + 1$ predictores, $\hat{\beta}_{GLS}$ son los coeficientes de la regresión que pueden ser estimados mediante OLS (*Ordinary Least Squares*) u óptimamente utilizando GLS (*Generalized Least Squares*) y λ_o es el vector de n pesos (*weights*) de kriging utilizados para interpolar los residuos. La varianza de predicción refleja la posición de los nuevos sitios (los extrapolados) en el espacio geográfico (Hengl, 2009):

$$\hat{\sigma}_{RK}^2(x_o) = (C_o - C_1) - C_o^T \cdot C^{-1} \cdot c_o + (q_o - q^t \cdot C^{-1} \cdot c_o)^T \cdot (q^t \cdot C^{-1} \cdot q)^{-1} \cdot (q_o - q^t \cdot C^{-1} \cdot c_o)$$

Donde $C_o + C_1$ es la variación debida al *sill* y c_o es el vector de covarianzas de residuos en el sitio a estimar la variable de interés (*unvisited*). Hengl (2009) indica que si los residuos no muestran auto correlación espacial (efecto *nugget* puro), RK converge a regresión lineal múltiple debido a que la matriz de covarianzas

se vuelve la matriz identidad; de forma similar, si la variable no muestra correlación con los predictores auxiliares, el modelo RK se reduce a KO debido a que la parte determinística es igual al valor medio.

Resultados y discusión

En los siguientes apartados se presentan y discuten los resultados obtenidos de la aplicación de los métodos revisados en el apartado anterior.

Generación de una superficie de permeabilidad de suelos mediante kriging ordinario (KO)

Se llevaron a cabo mediciones de permeabilidad de suelos (en términos de infiltración hortoniiana simulada) en el denominado suelo de conservación del DF por INIFAP y PAOT, en el marco de un proyecto financiado por este último organismo. A fin de utilizar esta información para análisis posteriores, tales como la valoración del territorio en términos de las características que los hacen aptos para recarga del acuífero, es necesario contar con una superficie continua de este parámetro. El método de interpolación seleccionado fue KO debido a que no se encontró relación con otra variable de tipo medioambiental y a que fue posible el ajuste de un semivariograma a la variable de interés (de acuerdo con lo indicado por Hengl, 2009). La implementación se hizo en el *software* libre de estadística R. Los resultados se muestran en la figura 1. Los resultados son comparados contra una interpolación mecánica (IDW) utilizando una validación cruzada para cada caso; con el fin de asegurar que existe una diferencia estadística entre los dos modelos obtenidos, se corrió un test t (asumiendo distribución normal de los errores) sobre los residuales de cada uno de los métodos. Los estadísticos obtenidos de este test fueron $t = 0.1396$, $df = 95.892$, $p\text{-value} = 0.8893$. De acuerdo con este test, la hipótesis nula de que la diferencia entre los dos ejemplos es igual a 0 no se acepta, por lo que se define que existe

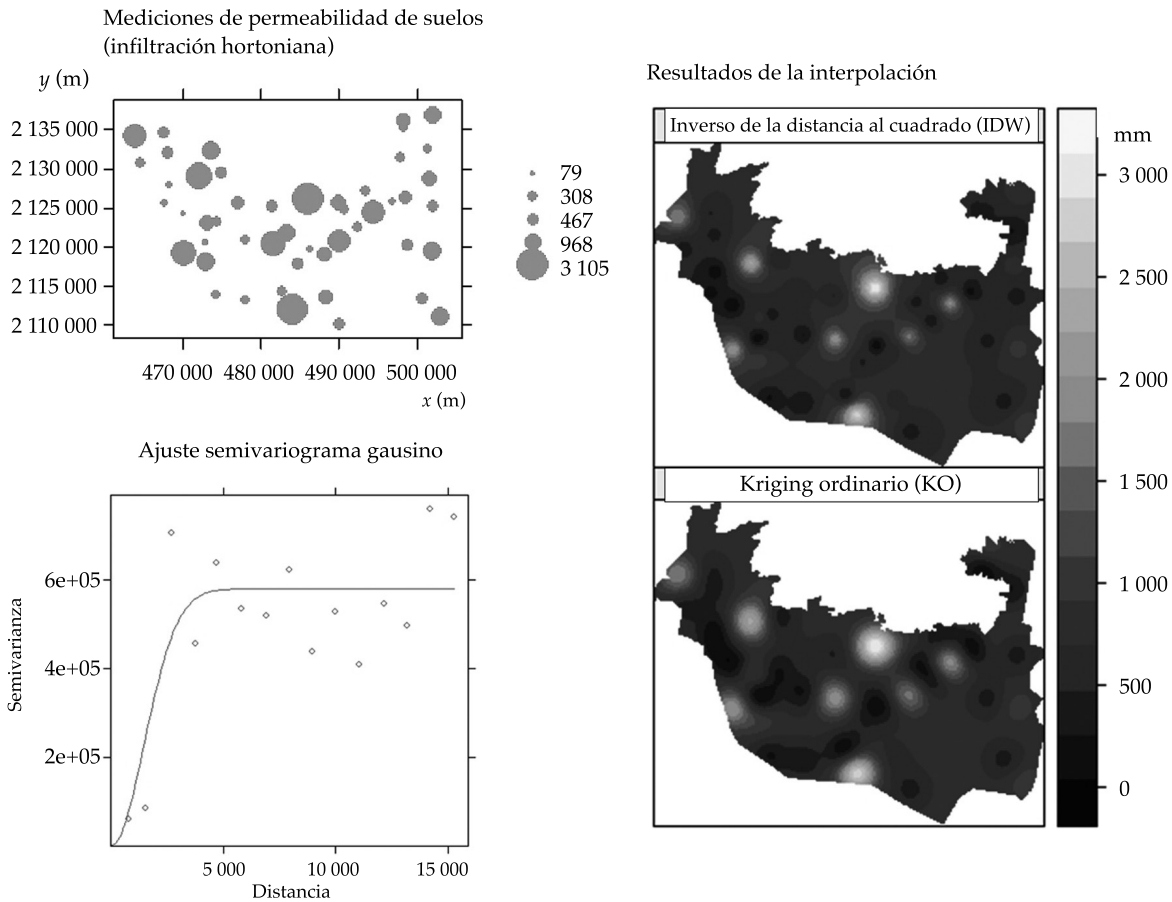


Figura 1. Interpolación mediante kriging ordinario (KO) de mediciones de permeabilidad de suelos (infiltración hortoniana simulada) en el suelo de conservación del DF y comparación contra interpolación usando inverso de la distancia al cuadrado (IDW).

una diferencia significativa entre el poder de predicción de los métodos comparados.

Kriging con deriva externa (KED)

Para el caso de esta técnica de análisis espacial de tipo geoestadístico se analizan dos casos de aplicación: la interpolación de niveles piezométricos del acuífero de valle de Santiago en Guanajuato, utilizando como variable auxiliar un modelo de elevación proveniente de SRTM, y la obtención de una superficie continua de precipitación multianual para los meses de abril y septiembre en la ciudad de México, utilizando el mismo insumo como variable auxiliar.

Niveles piezométricos del acuífero de valle de Santiago, Guanajuato

La cuenca Lerma-Chapala sufre procesos de intensa sobreexplotación de sus acuíferos, sobre todo para uso agrícola. Por ello, contar con superficies de parámetros indicativos de la condición de dichos acuíferos resulta de gran utilidad para fines de manejo y gestión del recurso hídrico. Con esta determinación se tiene el objetivo de explorar la utilización de KED en la generación de mapas de niveles piezométricos de acuíferos, ejemplificando con mediciones de niveles estáticos de la zona de valle de Santiago, Guanajuato. Se generó un script en Matlab para utilizar KED e interpolar

niveles estáticos en 1982 y 1985. La deriva fue ajustada a un modelo lineal entre elevación y nivel piezométrico (correlación lineal medida mediante coeficiente de correlación de Pearson). Los datos de elevación provienen de la misión *Shuttle Radar Topography Mission* (SRTM, 2003) de la NASA. Los resultados del método son valorados en función del cálculo de errores cuadráticos medios y mediante un ejercicio de validación cruzada, el cual consistió en interpolar cada valor de precipitación usando el resto de las mediciones. Los resultados de este procedimiento se muestran en la figura 2. Con un porcentaje de puntos bien estimados (indicados por los coeficientes de correlación) de 70% para 1982 y 86% para 1985, se considera que KED es un método adecuado para la estimación de superficies de niveles piezométricos. El error estándar medio de la estimación es menor a 15% de los valores manejados de elevación del nivel piezométrico.

Kriging con deriva externa: estimación de precipitación con elevaciones

En las regiones con mucho relieve, como la ciudad de México, la topografía actúa como un factor importante en la generación de lluvias de mayor intensidad. En muchos casos, se puede establecer una relación lineal entre la elevación y la precipitación medida por estaciones meteorológicas (Michauda et al., 1995; Weisse y Bois, 2001). Esta relación se justifica considerando que las lluvias son causadas por las corrientes de aire que se enfrían bruscamente por causas orográficas, produciendo un efecto que puede generalizarse, como: a mayor elevación, mayor precipitación.

En este estudio se usó la elevación como variable externa para modelar la deriva y se aplicó KED para interpolar las mediciones de precipitación en cincuenta estaciones meteorológicas (figura 3, arriba a la izquierda). Dicha interpolación se hizo en dos meses, que representan los extremos de un mes seco (abril) y uno lluvioso (septiembre). En este caso se evaluó la correlación lineal (coeficiente

de correlación de Pearson) y la correlación no lineal (coeficiente de correlación de Spearman) entre la precipitación medida y la elevación promedio en una vecindad alrededor de la localización de las estaciones. El objetivo, en este caso, fue identificar si el nivel de agregación espacial de la variable externa tiene alguna influencia en las estimaciones. Para la agregación de los datos de elevación se consideraron pesos gaussianos con desviación estándar (DE) incrementada en potencias de 2, i.e., $\sigma = 1, 2, 4, 8, 16, 32$ y 64. El modelo digital de elevaciones empleado fue proporcionado por el INEGI y presenta una resolución espacial de cincuenta metros.

En general, se observó una mejor correlación para los datos de septiembre (0.77-0.79 lineal y 0.73-0.79 no lineal) que para los del mes abril (0.38 lineal y 0.36-0.50 no lineal). Esto implica que las lluvias fuertes (2.65-9.32 mm) registradas en septiembre son efectivamente influenciadas por la orografía, mientras que las ligeras lluvias de abril (0.15-0.70 mm) no lo son tanto. En todo caso, la relación entre elevación y precipitación en abril parece ser fuertemente no lineal (figura 3 arriba a la derecha). Es notable también que la correlación lineal sea casi independiente del nivel de agregación, mientras que la correlación no lineal tiende a caer con el nivel de agregación de los datos.

La interpolación KED empleó un modelo de variograma exponencial para ambos casos. Dicha elección se basó únicamente en el error de ajuste entre el variograma estimado y el modelo. El modelo exponencial resultó ser el mejor de cinco modelos evaluados (exponencial, racional cuadrático, esférico, gaussiano y cúbico). Con el fin de evaluar la confiabilidad de los mapas de precipitación (figura 3, parte media), se aplicó el proceso de validación cruzada (figura 3, abajo). En términos absolutos se encontró que el error medio cuadrático fue menor para el caso de abril (RMSE = 0.1134 mm) que para septiembre (RMSE = 0.8124 mm). Esto se debe a la poca variabilidad de la precipitación en abril (DE = 0.1 mm) respecto a septiembre (DE = 1.43).

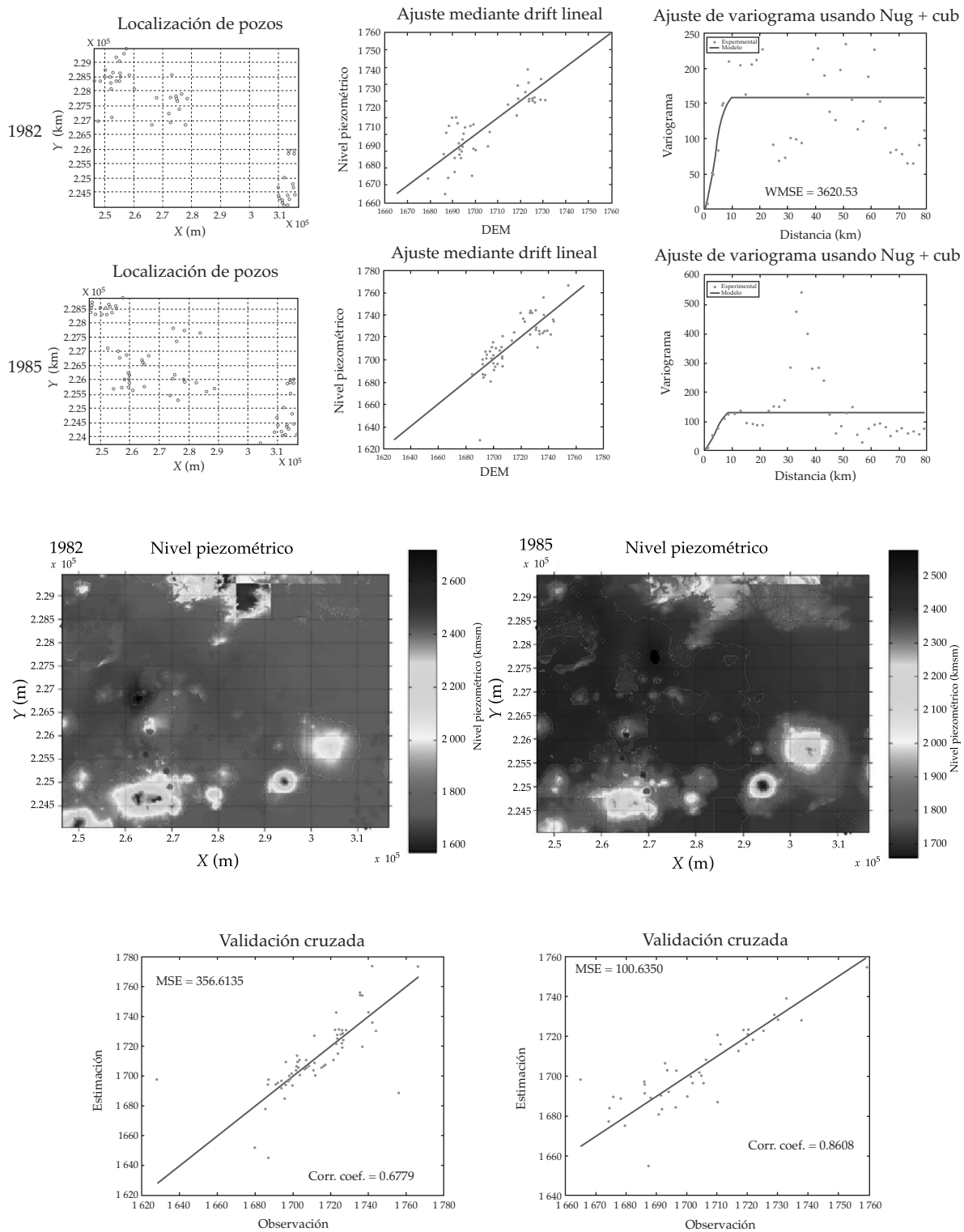


Figura 2. Interpolación mediante kriging con deriva externa (KED) de niveles piezométricos del acuífero en valle de Santiago, Guanajuato, para los años 1982 y 1985. La efectividad del método es verificada mediante validación cruzada (parte inferior de la figura).

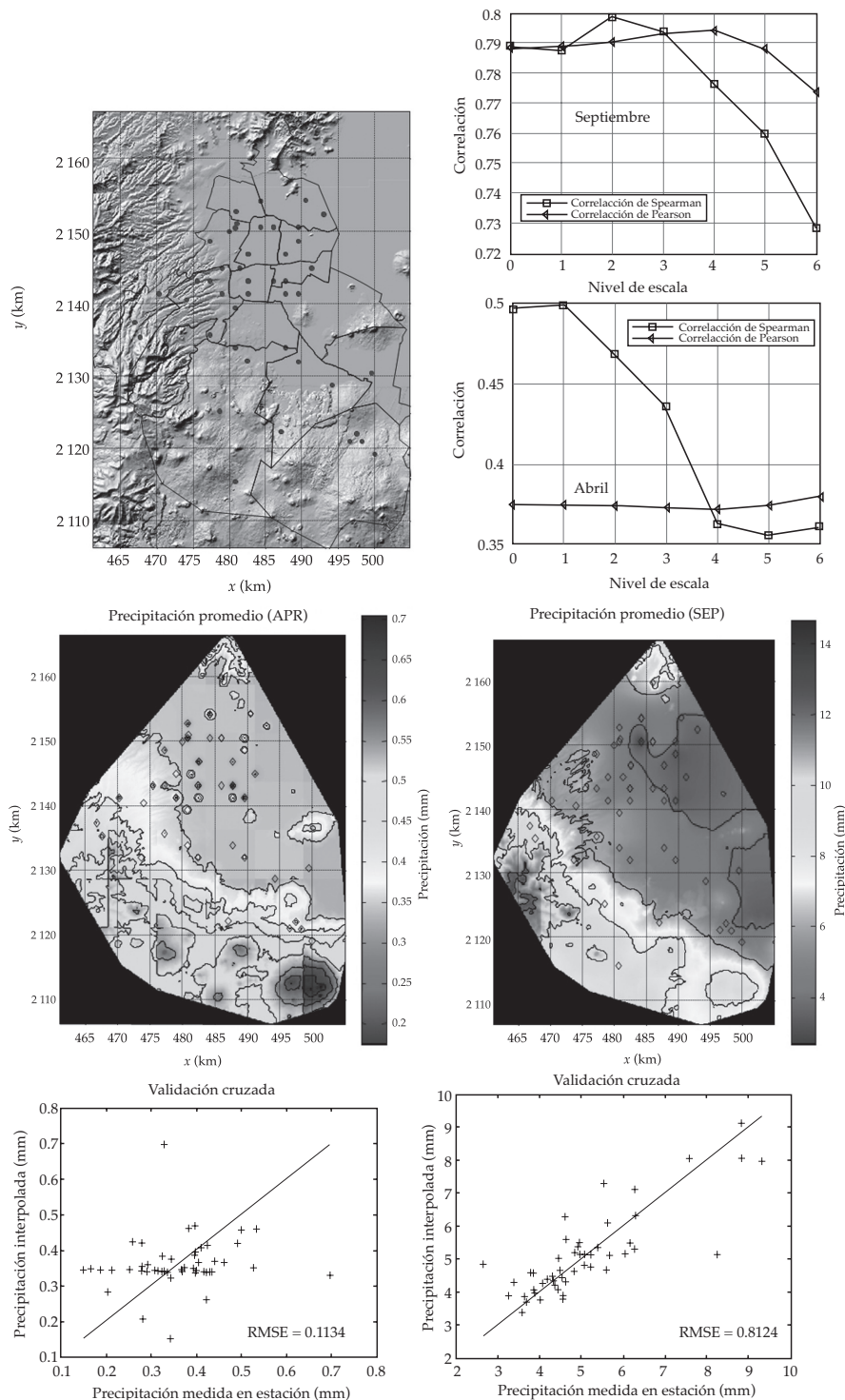


Figura 3. Arriba izquierda: localización de las cincuenta estaciones meteorológicas sobre la imagen de relieve y límites de delegaciones políticas del Distrito Federal. Arriba derecha: variación de la correlación lineal (Pearson) y no lineal (Spearman) entre la precipitación y la elevación, en función del nivel de agregación ($\log_2(\sigma)$) de la elevación para los meses de abril (panel superior) y septiembre (panel inferior). En medio: precipitación media interpolada mediante KED para los meses de abril (panel izquierdo) y septiembre (panel derecho). Abajo: diagrama de dispersión entre valores observados y valores interpolados de una validación cruzada para datos de abril y septiembre.

En términos relativos, sin embargo, el error fue mucho menor para septiembre (8%) que para abril (812%).

**Regresión geográficamente pesada (GWR).
Obtención de temperaturas máximas del aire
a partir de imágenes MODIS de temperatura
superficial, índice de vegetación, distancia al
océano y altitud.**

La aplicación de GWR presentada en este apartado tiene la intención de obtener un modelo adecuado de estimación de temperaturas máximas del aire (T_a) a partir de temperatura superficial (LST) e índice normalizado de vegetación (NDVI) obtenidos de imágenes MODIS, elevación (obtenido mediante un modelo de SRTM) y distancias al océano, en términos de lo investigado por Hengl (2009). Los datos fueron procesados para el periodo de ocho días que inicia el día juliano 257 del año 2008 (del 13 al 20 de septiembre de 2008).

Los errores producidos por toma de datos de estaciones poco representativas pueden ser mucho mayores que los que pueden esperarse de una estación aislada (WMO, 2006). Esa es la razón por la que se eligió en este caso un procedimiento que compara el valor de píxeles procedentes del sensor MODIS (con una resolución de 1.0 km), el valor de la elevación y el de la distancia al océano con el valor de la estación situada en la zona del píxel. Eso significa una comparación a escala local de acuerdo con la definición de la WMO (2006).

Los datos utilizados para la aplicación mostrada en este apartado y para la que se presenta en la sección "Regresión kriging (RK). Obtención de temperatura máxima del aire a partir de imágenes MODIS de temperatura superficial e índice de vegetación, distancia al océano y altitud", se obtuvieron de acuerdo con lo descrito a continuación: se generó una base de datos con valores de los píxeles de las imágenes MODIS LST (figura 4a), NDVI (figura 4b), distancia al océano (figura 4c) y modelo digital de elevación STRM a 1 km² para

obtener la altitud del dato observado (figura 4d). Los productos analizados corresponden a imágenes MODIS: MOD11A2.V5: LST y emisividad y MOD13A2.V5: NDVI. Ambos productos son presentados como listos para su uso en publicaciones científicas (USGS LP DACC, 2009). Tienen una resolución espacial de 1 km² (0.93 km²) por píxel. Para LST son el resultado de una composición de ocho días (*daytime and nighttime*), con una confiabilidad del 95% sobre tierra y un 66% sobre cuerpos de agua (USGS LP DACC, 2009); para NDVI son el resultado de una composición de 16 días. Este producto fue validado y tiene correcciones atmosféricas de reflectancias de superficie bidireccional que han sido enmascarados por el agua, las nubes, los aerosoles pesados y las sombras de las nubes (USGS LP DACC, 2009). Para cubrir el territorio nacional se requieren 6 *tiles* (unidades en que se segmentan y distribuyen las imágenes MODIS), que se descargaron de la siguiente dirección URL: https://lpdaac.usgs.gov/lpdaac/get_data. Los *tiles* tienen una proyección sinusoidal de origen y fueron reproyectados mediante la herramienta MRT MODIS *Reprojection Tool*, que también se puede descargar de la dirección URL antes mencionada a Cónica Conforme de Lambert con datum NAD27.

La base de datos de T_a proporcionada por el SMN sólo toma en cuenta datos validados por su personal. En este trabajo, los datos faltantes, estimados, dudosos o no confirmados son eliminados. Esto permite obtener una mejor base de comparación entre T_a y LST en el territorio nacional. Los datos de T_a fueron agrupados y promediados cada ocho días de acuerdo con el día juliano y sólo se utilizaron los que reportan una frecuencia mayor o igual a seis días. En la figura 4e se muestra la distribución espacial de las estaciones del SMN.

Los criterios utilizados para definir el nivel de la correlación son los propuestos por Downie (1973): $R < 0.3$ baja, R entre 0.3 y 0.6 mediana o moderada, y buena para $R > 0.8$. No obstante, se reconoce que la calificación de los

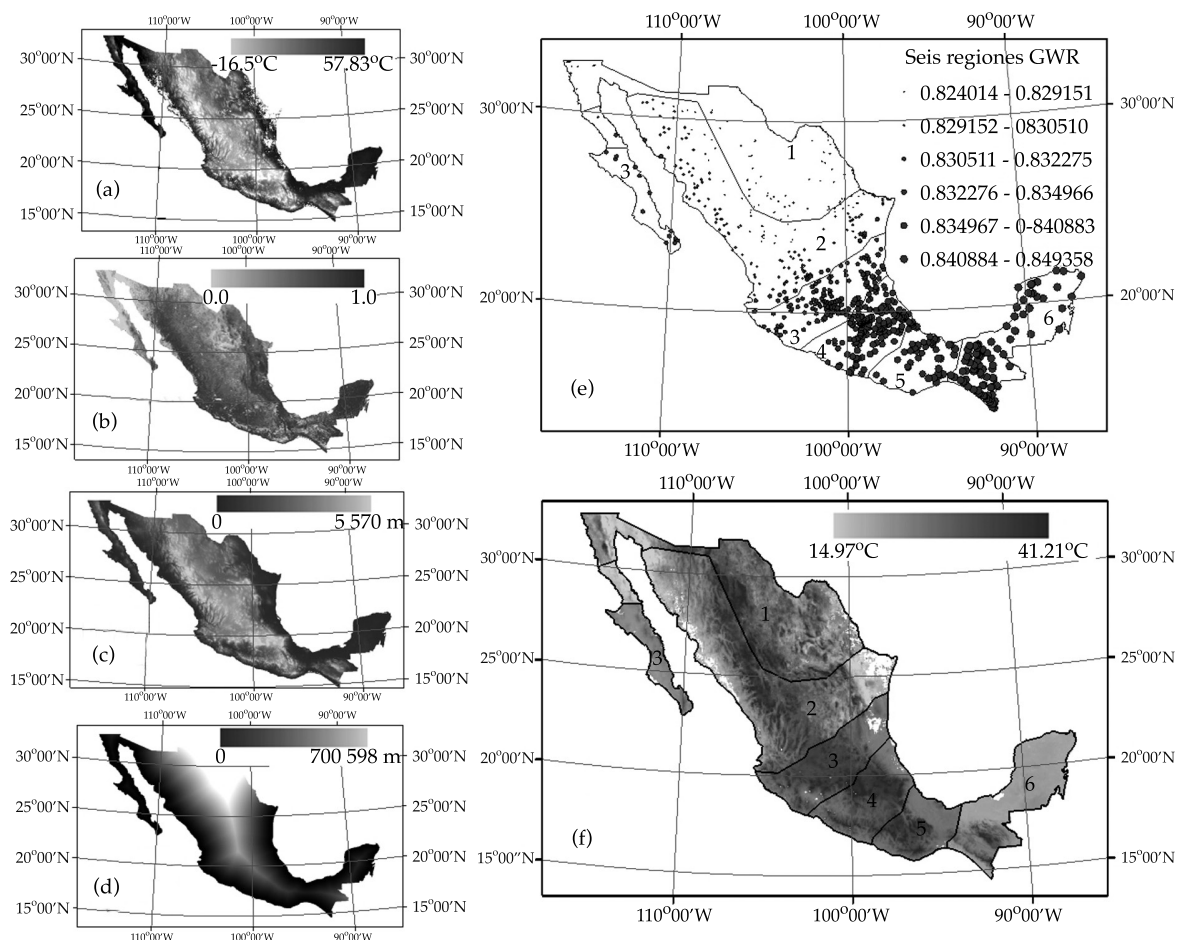


Figura 4. Variables de entrada y resultado de la aplicación de regresión geográficamente pesada (GWR) para obtención de temperaturas máxima del aire (T_a) en México (toda la información se obtuvo para el periodo de ocho días que inicia el día juliano 257 del año 2008 a una resolución de un 1 km^2 : a) temperatura de superficie (LST), composición de seis *tiles* de imágenes MODIS producto MOD11A2; b) Índice Normalizado de Vegetación (NDVI), composición de seis *tiles* de imágenes MODIS, producto MOD13A2; c) distancia al mar en metros (Distm); d) modelo digital de elevación (DEM) generado por SRTM; e) seis regiones obtenidas por GWR mediante la función $T_a = (LST, NDVI, DEM, Distm)$, T_a es medida en campo por el SMN; f) composición de seis regiones calculadas por GWR, pero correlacionadas individualmente por OLS para generar un continuo de T_a , mediante $T_a = (LST, NDVI, DEM, Distm)$.

niveles de correlación no es única ni invariable, sino como lo indica la autora antes mencionada, depende de la naturaleza del problema que se está modelando. A partir de los valores de R obtenidos mediante GWR fue posible definir seis regiones (niveles de agregación espacial). No se encontraron correlaciones estadísticamente no significativas en ninguna de las seis regiones o clases generadas por el modelo GWR. El modelo con los indicadores

más bajos de correlación fue el correspondiente a la clase 1, con un $R = 0.8240$ y $R^2 = 0.7105$, y el que tuvo los más altos fue el correspondiente a la clase 6, con un $R = 0.8494$ y $R^2 = 0.7214$. Estos datos demuestran que existe una alta correlación a nivel nacional, y que es posible utilizar datos de LST para estimar un continuo de T_a en el país.

Considerando la dificultad de aplicar aproximadamente mil modelos diferentes para

generar una superficie continua (mapa) de T_a , se optó por estimar un modelo de mínimos cuadrados ordinarios (OLS) para cada una de las seis regiones obtenidas (figura 4e) y aplicar estos modelos a todas las celdas de las regiones definidas, a efecto de obtener la variable de interés como superficie continua. Las variables dependientes y la independiente fueron las mismas en este análisis que para el caso de GWR. En el cuadro 1 se muestran los modelos obtenidos para cada región y sus estadísticos.

Como puede observarse, se obtuvieron regresiones estadísticamente significativas con R^2 superiores a 0.5 ($R > 0.71$). El único predictor que resultó estadísticamente no significativo es la distancia al mar, por lo que no fue considerado en los cálculos finales para obtener el continuo de T_a , mismo que se muestra en la figura 4f. El procedimiento implementado presentó el problema de generar diferencias notables en el valor estimado de T_a en las zonas límites de las regiones definidas.

En términos generales, el uso del modelo GWR tiene la desventaja de generar una gran cantidad de modelos (uno por punto o localidad analizada). Este modelo es adecuado

para cuando se tiene que trabajar con datos espaciales a escala fina. Sin embargo, cuando se requieren continuos de la variable de interés a nivel regional resulta difícil su aplicación, tomando en cuenta que hay muchos modelos por aplicar y debe buscarse un criterio adecuado para regionalizar las áreas de influencia de cada uno de ellos. Este aspecto puede ser sujeto de investigación posterior.

Regresión kriging (RK). Obtención de temperatura máxima del aire a partir de imágenes MODIS de temperatura superficial e índice de vegetación, distancia al océano y altitud

Temperatura máxima del aire (T_a) para el periodo de ocho días que inicia el día juliano 257 de 2008, medida en 796 estaciones meteorológicas del país, fue interpolada mediante RK. Un modelo lineal de cuatro variables independientes: elevación (SRTM, 1 km), temperatura superficial (LST medida por el sensor MODIS), distancia a la costa más cercana y NDVI (calculado a partir de datos de MODIS) se utilizó como deriva. Dichas

Cuadro 1. Resultados de la regresión OLS para las seis regiones definidas mediante GWR.

Región	1	2	3	4	5	6
Coefficientes de la regresión						
Intercepto	14.74	20.52	27.48	30.51	28.59	29.37
DEM	0.00	0.00	0.00	-0.01	0.00	0.00
NEAR_DIST	0.00	0.00	0.00	0.00	0.00	0.00
NDVI08257T	5.31	4.32	0.82	-1.92	1.35	-0.01
LST08257T	0.44	0.36	0.12	0.06	0.06	0.16
Estadísticos generales de la regresión OLS						
Aikake Criterium	391.65	1 047.87	651.34	868.97	218.87	469.38
R^2	0.58	0.72	0.73	0.66	0.54	0.66
Adjusted R^2	0.56	0.72	0.72	0.65	0.49	0.65
F-Statistic	25.70	136.67	100.43	83.96	12.96	55.34
F-Probability	0.00	0.00	0.00	0.00	0.00	0.00
Sigma2	8.34	7.49	3.59	7.94	4.24	2.90

variables presentaron una R^2 con respecto a la temperatura superficial de 0.676, 0.274, 0.260 y 0.023, respectivamente. La aplicación del método RK consistió de dos pasos. En el primer paso se realizó una regresión lineal múltiple ($R^2 = 0.7$, $R = 0.84$), donde el ajuste del modelo lineal se basó en el método de mínimos cuadrado (MC). En este caso, la varianza del error óptimo fue de 7.34. La fórmula de regresión múltiple ajustada se aplicó a todos los sitios no medidos por las estaciones, donde se conoce la elevación, temperatura superficial, distancia y NDVI (figura 5, arriba). El segundo paso consistió en modelar los residuos de la regresión como un proceso estocástico, con el fin de interpolarlos en los sitios no medidos. Para ello, primero se estimó el variograma y se determinó que el modelo esférico dio el mejor ajuste (figura 5, esquina superior derecha del panel inferior). El modelo de variograma esférico está determinado por la ecuación:

$$\gamma(h) = \begin{cases} n + \frac{s}{2} \left(\frac{3h}{r} - \frac{h^3}{r^3} \right), & 0 < h < r \\ n + s, & h \geq r \end{cases}$$

donde n , s y r son parámetros del modelo (*nugget*, *sill* y *range*) y h es la distancia entre observaciones. El ajuste por mínimos cuadrados ponderados por el número de pares (figura 5, panel inferior a la derecha arriba) dio como resultado valores de:

$$n = 3.7659, s = 3.2595 \text{ y } r = 256.0875 \text{ km}$$

Una vez ajustado el variograma, se procedió a la interpolación de los residuos, aplicando las fórmulas del método de interpolación KO. El resultado final se obtuvo agregando al mapa de la regresión los residuos interpolados (figura 5, abajo). La ventaja de considerar la componente espacial mediante KO consistió en un menor error para valores extremos de T_a .

Conclusiones

El método KO fue favorablemente aplicado para interpolar mediciones de permeabilidad de suelos (en términos de una infiltración horizontal simulada). Como está indicado teóricamente, este método fue útil debido a que no se encontró correlación con alguna variable ambiental y a que fue posible ajustar un semivariograma de la variable de interés. El test t realizado permitió establecer que existe una diferencia estadística entre los residuos de la interpolación mediante KO y una interpolación mecánica (IDW) que se realizó para comparar los resultados obtenidos mediante KO.

En el caso de la interpolación de niveles piezométricos del acuífero de valle de Santiago, en Guanajuato, utilizando como variable auxiliar un modelo de elevación proveniente de SRTM, KED se mostró como una técnica factible de ser aplicada con éxito. Las superficies de la variable de interés resultantes fueron adecuadamente verificadas a través de un ejercicio de validación cruzada. Mediante esta técnica se obtuvo un porcentaje de puntos bien estimados de 70% para 1982 y 86% para 1985, y un error estándar medio menor al 15%.

Para el caso de interpolación mediante KED de precipitación multianual para los meses de abril y septiembre en la ciudad de México, tomando como variable auxiliar a la elevación, se encontró que el modelo fue útil para la interpolación del mes de septiembre, que es cuando la época de lluvias está bien establecida y los procesos de precipitación favorecidos por la altitud pueden ser verificados. En este caso se obtuvo un error estándar medio de 8%, comparado con el de abril, que presentó un error muy elevado de 880%. En este análisis se encontró también que la correlación lineal entre las variables analizadas para ambos meses es prácticamente independiente del nivel de agregación espacial de los datos, mientras que la correlación no lineal tiende a caer con este nivel de agregación.

El método GWR se empleó para interpolar la temperatura máxima del aire (T_a , para

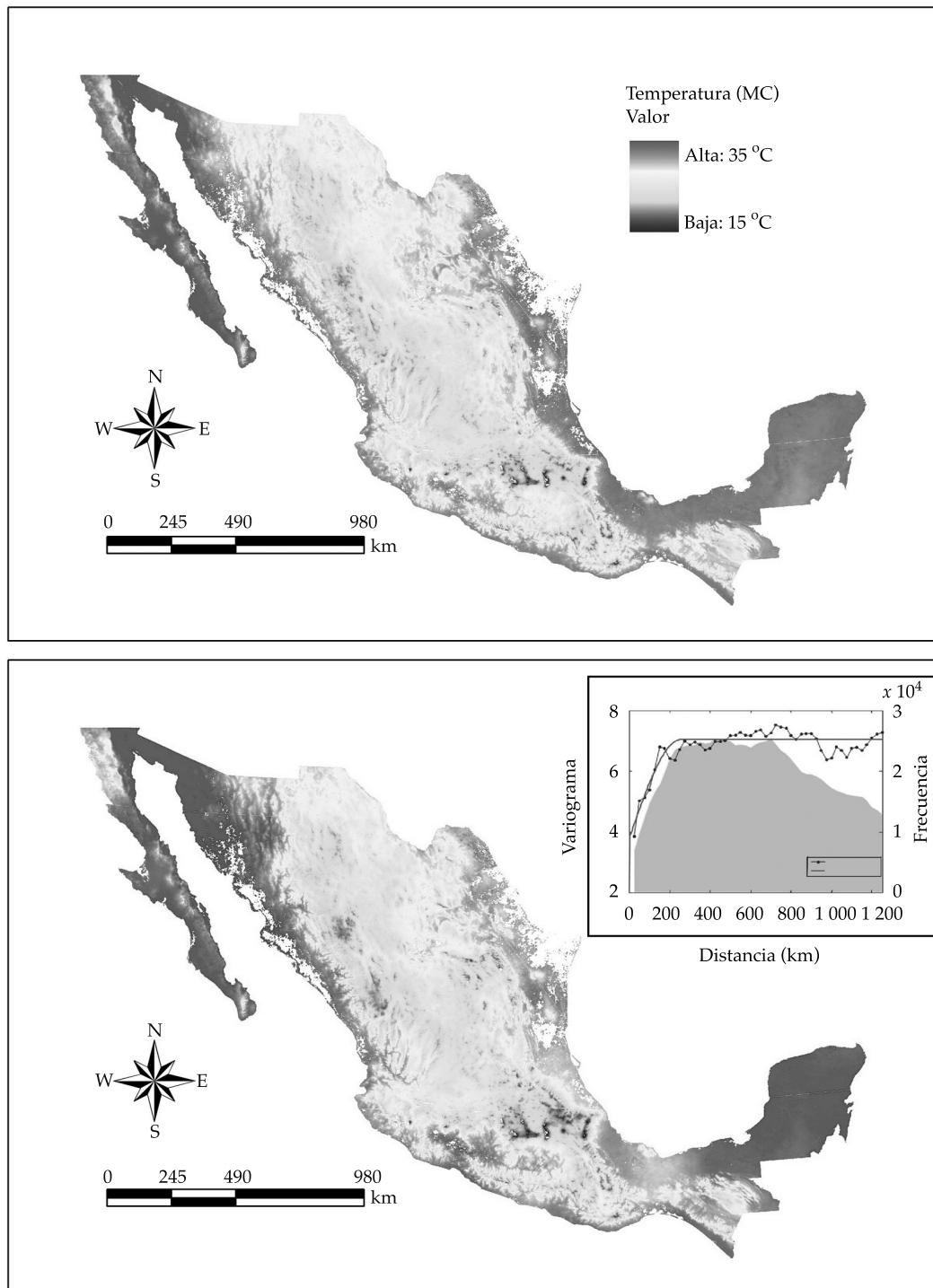


Figura 5. Mapas de temperatura máxima del aire para el periodo de ocho días, que inicia el día juliano 257 de 2008, estimada mediante regresión lineal múltiple (panel superior) y regresión kriging (panel inferior). En el panel inferior se muestra el ajuste del variograma para temperatura superficial medida en 796 estaciones en toda la república mexicana. El modelo de variograma ajustado corresponde a un modelo esférico. El área sombreada representa el número de pares de mediciones en cada clase de distancia (espaciamiento de 25 km).

un periodo de ocho días que inicia el día juliano 257 de 2008) medida en 796 estaciones meteorológicas del país, usando como deriva un modelo lineal de cuatro variables independientes: elevación (SRTM, 1 km), temperatura superficial (medida por el sensor MODIS), distancia a la costa más cercana, y NDVI (calculado a partir de datos de MODIS). Una ventaja primordial de GWR consiste en que posibilita la consideración en los análisis de regresión de la influencia de algunas variables que se ven afectadas por su ubicación geográfica, permitiendo definir al modelo de regresión espacial como no estacionario. Esto asegura que se obtengan residuales no correlacionados y que se respete el supuesto básico del análisis de regresión, relacionado con la independencia de los errores de la estimación. Una importante desventaja del uso del modelo GWR es la gran cantidad de modelos generados en cada análisis, ya que genera uno por dato correlacionado. Este modelo es adecuado para cuando se tiene que trabajar con datos espaciales a escala fina, pero cuando se requieren continuos de la variable de interés en el ámbito regional resulta difícil su aplicación. Lo anterior se debe a que hay muchos modelos que aplicar y debe buscarse un criterio adecuado para regionalizar sus áreas de influencia.

RK fue usado para interpolar T_a utilizando las mismas variables que para el caso de aplicación de WGR. La regresión lineal múltiple entre las variables independientes y la de interés se caracterizó por un nivel de ajuste adecuado ($R^2 = 0.7$, $R = 0.84$). Se ajustó un variograma de tipo esférico y los residuos se interpolaron mediante KO. Este enfoque permitió obtener un menor error para la estimación de superficies de valores extremos de temperatura.

De lo anterior es posible afirmar que las técnicas de análisis espacial permiten mejorar la efectividad del uso de datos provenientes de campo y de satélites, al mejorar su representatividad espacial y posibilitar la generación de esquemas de generación de superficies (mapeo) de la información

disponible, incluso mediante la utilización de información adicional que aportan valor a la misma información original. La principal ventaja de utilizar estos métodos es que aumentan la representatividad espacial de las variables mapeadas y es por ello que debe fomentarse su uso, en especial por los organismos encargados de diseminar la información base para la toma de decisiones en materia de agua y, en general, para el medio ambiente.

Recibido: 07/06/2011

Aceptado: 22/05/2012

Referencias

- ATKINSON, P.M., GERMAN, S., SEAR, D., and CLARK, M. Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. *Geographical Analysis*. Vol. 35, No. 2, January, 2003, pp. 58-82.
- BAILEY, T. and GATRELL, A. Interactive Spatial Data Analysis. Harlow, UK: Longman, 1995, p. 7.
- BRUNSDON, C., FOTHERINGHAM, S., and CHARLTON, M., Geographically Weighted Regression-Modelling Spatial Non-Stationarity. *Journal of the Royal Statistical Society. Series D (The Statistician)* Vol. 47, No. 3, 1998, pp. 431-443.
- DALY, C. Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology*. Vol. 26, March, 2006, pp. 707-721.
- FOTHERINGHAM, A. Context-dependent spatial analysis: A role for GIS? *Journal of Geographic Systems*. Vol. 2, No. 1, March, 2000, pp. 71-76.
- FOTHERINGHAM, A.S., BRUNSDON, C., and CHARLTON, M. Geographically weighted regression: the analysis of spatially varying relationships. Chichester, UK: Wiley, 2002, 269 pp.
- GOOVAERTS, P. Geostatistics for Natural Resources Evaluation. New York: Oxford University Press, September, 1997, pp. 133, 195.
- HENGL, T. A Practical Guide to Geostatistical Mapping (Vol. Second extended edition of the EUR 22904). Luxembourg: Office for Official Publications of the European Communities, 2009, pp. 27, 29, 32, 241.
- LICHSTEIN, J.W., SIMONS, T., SHRINER, S.A., and FRANZREB, K.E. Spatial Autocorrelation and Autoregressive Models in Ecology. *Ecological Monographs*. Vol. 72, No. 3, 2002, pp. 445-463.
- MICHAUDA, J., AUVINEB, B., and PENALBAB, O. Spatial and elevational variations of summer rainfall in the southwestern United States. *Journal of Applied Meteorology*. Vol. 34, 1995, pp. 2689-703.

- PIELKE, R.A., STOHLGREN, T., PARTON, W., DOESKEN, N., MONEY, J., and SCHELL, L. Spatial representativeness of temperature measurements from a single site. *Bulletin of the American Meteorological Society*. Vol. 81, No. 4, 2000, pp. 826-830.
- WEISSE, A. and BOIS, P. Topographic effects on statistical characteristics of heavy rainfall and mapping in the French Alps. *Journal of Applied Meteorology*. Vol. 40. No. 4, 2001, pp. 720-740.
- WHEELER, D. and TIEFELSDORF, M. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographic Systems*. Vol. 7, 2005, pp. 161-187.
- WMO. Guide to Meteorological Instruments and Methods of Observation. Preliminary seventh edition. Geneva: World Meteorological Organization, 2006.
- SRTM. The Shuttle Radar Topography Mission SRTM-30, Global 1km Digital Elevation Model. Shuttle Radar Topography Mission, 2003 [en línea] [citado el 19 de diciembre de 2008]. Disponible para *World Wide Web*: <ftp://e0srp01u.ocs.nasa.gov/srtm/version1/SRTM30/SRTM30_Documentation>.
- USGS LP DAAC. Land processes active archive center. Data set name: Land surface temperature & emissivity 8-Day L3 Global 1 km 2009 [en línea] [citado el 3 de abril de 2009]. Disponible para *World Wide Web*: https://lpdaac.usgs.gov/lpdaac/products/modis_products_table/land_surface_temperature_emissivity/8_day_l3_global_1km/v5/terra.

Abstract

TAPIA-SILVA, F.O., SILVÁN-CÁRDENAS, J.L. & ROSALES-ARRIAGA, E. *Spatial analysis to improve the use of field and satellite data to support better decision-making regarding water and the environment. Water Technology and Sciences (in Spanish). Vol. IV, No. 1, January-March, 2013, pp. 149-166.*

Climate, soil and water availability information in Mexico —determined through point measurements at ground stations— do not adequately meet the criteria for representativeness for the purpose of generating data, information and knowledge to improve decision-making. Remote sensors (airborne and satellite) are potentially an excellent resource of information to complement ground point measurements. Nevertheless, the information obtained must be validated and complemented by field measurements. In addition, worth mentioning among the limitations of this information resource is its temporal and spatial resolution and the high sensitivity to weather conditions at the moment the data are taken, in the case of the more commonly used and available satellites (optical). Spatial analysis techniques enable improving the spatial representativity of data derived from field and satellites by making mapping possible based on the information available as well as additional information. The techniques described in this article, which are applied to the interpolation of data for practical cases exemplify this. Research about these methods and their implementation have not been sufficient in terms of the process to disseminate basic geospatial information in Mexico. It is therefore crucial to demonstrate the practical usefulness of this type of study in order to generate better information and basic knowledge for decision-making about water and the environment in general.

Keywords: spatial analysis, interpolation, kriging, remote sensing, field measurements, water.

Dirección institucional de los autores

Dr. Felipe Omar Tapia Silva

Universidad Autónoma Metropolitana Campus Iztapalapa
Departamento de Hidrobiología
Centro de Investigación en Geografía y Geomática Ing.
Jorge L. Tamayo A.C. (CentroGeo)
Av. San Rafael Atlixco núm.186, Col. Vicentina
Delegación Iztapalapa
09340 México, D.F., MÉXICO
Teléfono: +52 (55) 5804 4600, extensión 3056
feomtasi@yahoo.com.mx

Dr. José Luis Silván Cárdenas
M. en G. Edgar Rosales Arriaga

Centro de Investigación en Geografía y Geomática "Ing.
Jorge L. Tamayo" A.C. (CentroGeo)
Contoy núm.137, Col. Lomas de Padierna
Delegación Tlalpan
14240 México, D.F., MÉXICO
Teléfono: +52 (55) 2615 2820
jlsilvan@centrogeo.org.mx
erosales@centrogeo.org.mx