DOI:10.24850/j-tyca-2018-06-10

Nota

Ajuste de un variograma esférico de la precipitación anual de las normales climatológicas 1951-2010

Adjustment of a spherical variogram of the annual precipitation of the Climatological Normals 1951-2010

José Antonio Pedraza Oropeza[1]

Enrique Palacios Vélez[2]

Oscar Palacios Vélez[3]

[1]Colegio de Posgraduados: fpedraza@colpos.mx

[2] Colegio de Posgraduados: epalacio@colpos.mx

[3] Colegio de Posgraduados: opalacio@colpos.mx

Autor para correspondencia: José Antonio Pedraza Oropeza, fpedraza@colpos.mx

Resumen

En este trabajo se analizó la precipitación media anual de las normales climatológicas de 1951-2010, para determinar su estructura espacial y para que pueda ser utilizada en el método de interpolación de kriging, que permite no sólo estimar los valores de esta variable, sino también la varianza de estimación. Para ello se generó un semivariograma (que en el presente trabajo denominaremos "variograma") experimental con la precipitación media anual de 2 003 estaciones con 30 o más años de información de precipitación, tomadas de un total de 4 385 estaciones. Al incio, el variograma se construyó con 2 005 003 pares de estaciones (combinaciones de 2 003 estaciones tomadas de dos en dos), que fue preciso depurar. Para este propósito se calculó un “índice de variación espacial”, igual a la diferencia de precipitación dividida entre la distancia que separa las estaciones, índice que permitió identificar 20 046 pares atípicos, como aquellos que tenían un valor superior a 9.85. Además se consideraron sólo los pares de valores con distancia entre estaciones no mayores de 1 900 km. El variograma experimental finalmente obtenido consta de 380 clases a intervalos de 5 km y fue ajustado a un modelo esférico. En este proceso se utilizó el método de los multiplicadores de Lagrange, para asegurar continuidad del variograma en el punto donde la parte curva del variograma se encuentra con la parte horizontal del mismo. El mejor variograma esférico tiene un radio de influencia a = 1 455 km, un efecto pepita C_0 = 18 082 y umbral C_1 = 602 452.

Palabras clave: geostadística, kriging, variograma, optimización, normales climatológicas.

Abstract

In this work the average annual rainfall of the Climatological Normals of 1951-2010 was analyzed to determine its spatial structure, which can be used in the Kriging interpolation method, which allows not only to estimate the values of this variable, but also the variance of estimation. For this purpose, an experimental semivariogram (which in the present work we will call "variogram") was generated with the average annual precipitation of 2 003 stations with 30 or more years of precipitation information, taken from a total of 4,385 stations. Initially the variogram was built with 2 005 003 pairs of stations (combinations of 2 003 stations taken in pairs), which had to be depurated. For this purpose, a "spatial variation index" was calculated, equal to the difference in precipitation divided by the distance separating the stations, an index that allowed identifying 20 046 atypical pairs, such as those with a value greater than 9.85. Additionally, only pairs of values with distance between stations no greater than 1 900 km were considered. The finally obtained experimental variogram consists of 380 classes at intervals of 5 km. The experimental variogram obtained was fitted to a spherical model. In this process, the Lagrange multipliers method was used to ensure continuity of the variogram at the point where the curved part of the variogram meets the horizontal part of the variogram. The best spherical variogram has a radius of influence a = 1 455 km, nugget effect C_0 = 18 082 and sill C_1 = 602 452.

Keywords: geostatistics, kriging, variogram, optimization, climatological normals.

Fecha de recibido dd/mm/aaaa (por la coordinación editorial)

Fecha de aceptado dd/mm/aaaa (por la coordinación editorial)

Introducción

Uno de los problemas que más se presentan en los estudios de la precipitación es su interpolación, cuando se utilizan los datos de las estaciones vecinas con el fin de calcular y dibujar isoyetas. Tanto para la interpolación de datos de lluvia, como de otras muchas variables geográficas, uno de los métodos utilizados con mayor frecuencia es el conocido con el nombre de kriging (Oliver & Webster, 2015).

El método de interpolación de kriging consiste en obtener un estimador insesgado y de mínima varianza, lo que permite conocer no sólo el comportamiento de la variable bajo estudio, sino también el error de estimación de la misma (incertidumbre), lo cual es de gran ayuda en la interpretación de los resultados, así como en la determinación de zonas de alta varianza de estimación, en las que se requiere de un mayor número de muestras (de estaciones pluviométricas en el caso de la precipitación). La utilización del método de kriging tiene como principal requisito la definición de un modelo teórico de variograma que describa la correlación espacial de los datos.

Objetivo

Caracterizar la estructura espacial de los datos de precipitación de las normales climatológicas 1951-2010, mediante el desarrollo de un variograma experimental ajustado a un modelo teórico del tipo conocido como “esférico”, que permita la utilización del método de interpolación kriging.

Revisión bibliográfica

Actualmente, se cuenta con una gran cantidad de métodos para la interpolación de variables geográficas en general, y de precipitación en particular, los cuales van desde métodos muy simples, como los polígonos de Thiessen, métodos más elaborados como el de las distancias inversas elevadas al cuadrado, hasta métodos analíticos como el de mínima curvatura, así como diferentes variantes del método de kriging. Una revisión concisa de los principales métodos de interpolación puede ser vista en el manual de usuario del programa Surfer (Golden Software, 2002).

Los métodos avanzados de interpolación se pueden subdividir convencionalmente en métodos determinísticos y estadísticos. Los primeros se basan en modelos matemáticos que representan de diferente manera la distribución de las variables sobre el área de estudio. Entre estos métodos destacan las funciones spline y la analogía de la sábana delgada o thin-plate splines (Wang & Wang, 2012); en los métodos estadísticos las variables de estudio se consideran realizaciones de un proceso aleatorio, entre éstos se encuentra el método de interpolación óptima (Thomson & Emery, 2014), el método de kriging (Matheron, 1971) y las funciones ortogonales empíricas (Hannachi, 2004), también conocidas como análisis armónico generalizado de procesos aleatorios.

El variograma

El interpolador kriging está basado en una medida de la correlación espacial de los datos a interpolar, llamada variograma, que matemáticamente se define como: "la mitad de la esperanza matemática del cuadrado de la diferencia de dos valores de la variable de estudio Z, correspondiente a sitios separados una distancia h". En matemáticas esto se escribe como:

$$\gamma(h) = \frac{1}{2}E[Z(X) - Z(X+h)]^2 \qquad (1)$$

Donde $E[Z(X)]$ es el operador "esperanza matemática"; X es un vector de coordenadas y h es la distancia que separa a dos puntos donde se mide la variable Z.

Existe una importante relación entre los conceptos varianza, covarianza y variograma, la cual se muestra en la siguiente ecuación (2):

$$\gamma(h) = \mathrm{Var}[Z(X)] - \mathrm{Cov}[Z(X), Z(X+h)] \qquad (2)$$

El variograma experimental

Las variables geográficas se distinguen entre sí de acuerdo con su variograma. Una estimación del mismo la da el llamado variograma experimental, que se define como (Freeden, Nashed, & Sonar 2015):

$$\gamma_e(h) = \frac{1}{2N(h)}\sum_{k=1}^{N(h)}(Z(x) - Z(x+h))^2 \qquad (3)$$

Donde *N(h)* es el número de pares de valores que se comparan.

Modelos teóricos de variograma

Para su utilización en el método de interpolación de kriging se requiere un modelo continuo de variograma, en lugar de la nube de puntos que presenta el variograma experimental. Estos puntos pueden ajustarse a uno de varios modelos (Oliver & Webster, 2015); sin embargo en el presente trabajo se utilizará el llamado "modelo esférico" que probablemente es el más utilizado, dado que sus tres parámetros corresponden a las propiedades deseables en un variograma teórico, que son:

a) Umbral o meseta. Si un proceso presenta estacionaridad de segundo orden, su variograma debe alcanzar un límite superior o varianza umbral, la cual es la varianza del proceso (Oliver & Webster, 2015).
b) Rango. La distancia a la cual se alcanza el umbral se conoce como rango y se puede interpretar como la distancia más allá de la cual las observaciones no presentan ninguna correlación, y por lo tanto, es una medida del radio de influencia de las observaciones. Los variogramas que se aproximan al umbral de manera asintótica no tienen un rango estricto, pero de manera práctica se considera como rango la distancia a la cual se alcanza el 0.95 del umbral (Oliver & Webster, 2015).
c) Ordenada al origen o efecto de pepita. Teóricamente cuando h = 0 la semivarianza debe también ser 0, de presentarse una ordenada positiva en el origen, se conoce como efecto de pepita y se puede deber a errores en las observaciones o al hecho de que exista variabilidad espacial a distancias menores de las utilizadas en el análisis (Oliver & Webster, 2015).

El modelo esférico se define como:

$$\left.\begin{array}{l}\gamma(h) = C_0 + \frac{C_1}{2}\frac{h}{a}\left(3 - \left(\frac{h}{a}\right)^2\right); 0 \leq h < a \\ \gamma(h) = C_0 + C_1 = \mathrm{Var}(h)\end{array}\right\} \quad (4)$$

Materiales y métodos

Las normales climatológicas se descargaron del sitio web del Servicio Meteorológico Nacional, de la Comisión Nacional del Agua, http://smn.cna.gob.mx/es/component/content/article?id=42, en el cual se encuentran las Normales Climatológicas 1971-2000, 1981-2010 y 1951-2010, y se seleccionaron estas últimas por ser las que cuentan con el mayor número de años de registro, 60 años, y el mayor número de estaciones, 4 380.

Cada uno de los 4 380 archivos descargados contiene los datos generales de cada estación, como son clave, nombre, latitud, longitud, altitud y promedios mensuales y anuales de las variables meteorológicas temperatura máxima, temperatura media, temperatura mínima, precipitación y evaporación, a partir de los cuales se generó una base de datos cuya estructura se indica en la Tabla 1.

Tabla 1. Campos de la base de datos de estaciones

CAMPO	TIPO	LON	DEC	DESCRIPCIÓN
Clave	Texto	5		Clave de la estación
Nombre	Texto	32		Nombre de la estación
Latitud	Real	11	8	Latitud de la estación en grados
Longitud	Real	12	8	Longitud de la estación en grados
Altitud	Real	6	1	Altitud de la estación en metros
Precipitación	Real	6	1	Precipitación anual en milímetros
Años	Entero	2		Número de años de registro
Xccl	Real	10	2	Coordenada X en PCCL*
Yccl	Real	10	2	Coordenada Y en PCCL*
* Proyección cartográfica Cónica Conforme de Lambert				

Depuración de los datos

Previo a la generación del variograma experimental se llevó a cabo una depuración de los datos recabados, que incluyó la eliminación de 5 estaciones localizadas en islas, por su fuerte efecto marítimo; igualmente se eliminaron tres estaciones sin datos de precipitación; de 14 pares de estaciones con las mismas coordenadas se tomó la de mayor número de años de registro; como se considera normal climatológica a una estación con 30 o más años de registro. Se procedió a eliminar las 2 363 estaciones con menos de 30 años de registro, quedando una base de datos de las 2 003 estaciones, las cuales se pueden observar en la Figura 1.

Figura 1. Ubicación de las 2 003 estaciones utilizadas.

Elaboración del variograma experimental

De acuerdo con la ecuación (3) para calcular las ordenadas del variograma es preciso calcular las diferencias entre la precipitación de dos estaciones separadas por una distancia *h*. Cuando las estaciones están irregularmente distribuidas, ante la casi certeza de no encontrar pares de estaciones separados por la misma distancia, es indispensable establecer intervalos de clase de distancias para promediar los cuadrados de las diferencias de precipitación, tal como lo indica la ecuación (3).

Se obtuvieron 2 005 003 de pares de estaciones (combinaciones de 2 003 estaciones tomadas de dos en dos), se calcularon las diferencias de precipitación y la distancia que las separaba. Ante la sospecha de que hubiera errores en las coordenadas de algunas estaciones, resultando en diferencias de precipitación muy grandes para distancias muy cortas entre las mismas, se decidió calcular un "índice de variación espacial", como el cociente de la diferencia de la precipitación en milímetros, entre la distancia entre estaciones, expresada en kilómetros. Valores muy elevados de este índice (mayores de 9.85 mm/km) se consideran datos atípicos que es conveniente eliminar para reducir los errores de cálculo.

Como se aprecia en la Figura 2, y se confirma con el cálculo de los percentiles, el 99% de los valores del índice se encuentran por debajo de 9.85, por lo que su eliminación reduce tan sólo en 1% el número de pares utilizados, pero incrementa la consistencia estadística del análisis.

Figura 2. Distribución del índice de variación espacial.

Después de eliminar los 20 046 pares con índice mayor a 9.85, los 1 984 957 restantes se agruparon en 652 clases con intervalos de 5 km, obteniéndose una frecuencia mínima de 1, una frecuencia máxima de 9 801, y una mediana de 1 692. Un variograma construido con estos datos cubriría distancias hasta de 3 260 km (652*5), distancia a la cual no es de esperar una correlación en los datos de lluvia. Con este motivo se elaboró un variograma de 380 clases de 5 km, cubriendo una distancia máxima de 1 900 km. Todos estos intervalos tuvieron una frecuencia superior a la mediana de las frecuencias de 1 692 pares de datos por intervalo.

Ajuste del variograma experimental a un modelo esférico

Para calcular los parámetros óptimos de un modelo esférico, es necesario resolver un problema de optimización (minimización), con la restricción (C_0 + C_1 = γ (h), para $h \geq a$, ver ecuación (4)), el cual puede resolverse por el método de mínimos cuadrados, modificado por los multiplicadores de Lagrange (Luenberger & Ye, 2016). Este método básicamente consiste en minimizar la siguiente función:

$$L = \sum_{j=1}^{n}\left(\gamma_e\left(h_j\right) - C_0 - C_1 X_j\right)^2 - 2\lambda(C_0 + C_1 - W) \tag{5}$$

$$Donde:$$

$$\gamma_e\left(h_j\right) = Valores\ del\ variograma\ experimental$$

$$X_j = \frac{1}{2}\frac{h_j}{a}\left(3 - \left(\frac{h_j}{a}\right)^2\right)$$

$$\lambda = Multiplicador\ de\ Lagrange$$

$$W = Promedio\ de\ \gamma_e\left(h_j\right)\ para\ h_j > a$$

Para calcular los parámetros óptimos de un modelo esférico, es necesario derivar la función L con respecto a C_0, C_1 y λ, igualar a cero las derivadas, y resolver el sistema de ecuaciones simultáneas, con lo cual, después de realizar simplificaciones algebráicas, se obtienen los siguientes parámetros del modelo esférico:

$$C_0 = \frac{\sum_{j=1}^{n}(X_j - 1)(WX_j - \gamma_e[h_j])}{\sum_{j=1}^{n}(X_j - 1)^2} \tag{6}$$

$$C_1 = \frac{\sum_{j=1}^{n}(X_j - 1)(\gamma_e[h_j] - W)}{\sum_{j=1}^{n}(X_j - 1)^2} \tag{7}$$

Tecnología y ciencias del agua, 9(5), 209-222, DOI:10.24850/j-tyca-2018-06-10

El cálculo de los parámetros del modelo esférico C_0, C_1 y a se tiene que realizar con un procedimiento iterativo como el que se indica a continuación:

1) Supóngase un valor para el parámetro a. Éste permitirá subdividir a los valores del variograma experimental en dos grupos: los valores donde la h es igual o menor que a y los valores con h mayor que a.
2) Se calcula el parámetro W igual al promedio de los valores del variograma para distancias superiores a a.
3) Se calculan los parámetros C_0 y C_1 mediante las fórmulas (6) y (7).
4) Se calcula el error asociado al valor supuesto de a mediante la fórmula (8).

$$E(a) = \sum_{j=1}^{n} (\gamma_e(h) - \gamma(h))^2 \qquad (8)$$

5) Se repiten los pasos 1-4 para diferentes valores de a y se seleccionan los valores de a, C_0 y C_1 con el mínimo error asociado.

El paso 5 puede automatizarse con diferentes algoritmos de optimización; sin embargo los cálculos realizados de forma interactiva, pueden llevar a un resultado equivalente.

Resultados

Se construyó un variograma experimental con la información de 2 003 estaciones con más de 30 años de registro, lo cual generó en primera instancia 2 005 003 pares de estaciones que se agruparon en 652 clases con intervalos de 5 km por clase. Para una mejor valoración de las diferencias en la precipitación y la distancia que separa a cada par de estaciones, como se señaló en la sección anterior, se calculó un índice de variación espacial como el cociente de la diferencia de precipitación de cada par de estaciones, dividido entre la distancia correspondiente. Este índice permitió identificar pares de estaciones atípicas, como aquellas que tenían una gran diferencia de precipitación expresada en mm/año y/o una distancia muy pequeña expresada en km. Dada la

Tecnología y ciencias del agua, 9(5), 209-222, DOI:10.24850/j-tyca-2018-06-10

distribución estadística de este parámetro, como se puede apreciar en la Figura 2, se decidió eliminar todos los pares (20 046) con índice superior a 9.85, los cuales representan menos del 1% del total de pares de estaciones.

Un variograma construido con estos datos cubriría distancias hasta de 3 260 km (652*5), distancia a la cual no es de esperar una correlación en los datos de lluvia, por lo que se eliminaron los pares de estaciones con distancias superiores a 1 900 km, quedando 380 clases, todas ellas con frecuencias mayores a la mediana de 1 ,692. En la Figura 3 se muestran los puntos del variograma experimental obtenido.

Figura 3. Semivariograma esférico ajustado.

Mediante el procedimiento explicado en la sección de Materiales y métodos se calcularon los parámetros óptimos del modelo esférico de variograma: rango (a = 1 455), efecto de pepita (C_0 = 18 082), y umbral (C_1 = 602 452). Con estos valores se calculó el variograma mostrado en la Figura 3. En esta figura se muestra también la suma de cuadrados del error para diferentes valores del radio de influencia, cuyo valor mínimo se encuentra en 1 455 km.

El variograma de la Figura 3 mostró con claridad que existe una estructura espacial en los datos de precipitación anual de las normales climatológicas 1951-2010, la cual es modelada por el variograma esférico. Este variograma puede ser utilizado para realizar interpolaciones por el método de kriging, para estimar valores de precipitación anual, pudiéndose además, calcular la varianza de estimación de los valores interpolados.

Discusión

Como se sabe, antes de iniciar un análisis de grandes cantidades de información, como se realizó en este trabajo donde se tienen 2 003 005 pares de estaciones, es indispensable depurar la información previamente a su análisis. Por esta razón, para cada combinación de dos estaciones se calculó la diferencia de precipitación en milímetros, la separación en kilómetros y el cociente de estos dos valores, al que se denominó "índice de variación espacial" de la precipitación, cuya distribución de frecuencias se muestra en la Figura 2, y se confirma con el cálculo de los percentiles. El 99% de los valores se encuentran por debajo de 9.85, por lo que la eliminación de los valores extremos reduce tan sólo en 1% el tamaño de la muestra, pero incrementa la consistencia estadística del análisis. Un valor elevado en el índice entre dos estaciones significa que hay inconsistencias en los valores de precipitación, la distancia entre ellas o ambos. La inconsistencia en los valores de precipitación se puede deber a diferencias en los rangos temporales de registro; por ejemplo, en el caso de que una estación tenga registros de los años 1951 a 1980, y la otra de los años 1981 a 2010, o cuando las estaciones que se comparan están en lados opuestos de un parteaguas. La inconsistencia en la distancia entre dos estaciones indica errores en las coordenadas de las mismas.

Por otro lado, el método desarrollado para calcular los parámetros óptimos del modelo esférico permite darle continuidad al variograma en el punto donde la curva se encuentra con la parte horizontal, cuya ordenada es igual a la varianza de los datos. Aparentemente, esta continuidad sólo se logra con la utilización de los multiplicadores de Lagrange.

En conclusión, para muchos climatólogos, el radio de influencia encontrado de 1 455 km pudiera parecer muy elevado, ya que esto significaría que dos estaciones separadas por esta distancia tienen valores de precipitación espacialmente correlacionados. Por esta razón, sería deseable subdividir el área de la República Mexicana en regiones más pequeñas, quizás a nivel de cuenca hidrológica, lo que daría lugar a un variograma con un radio de influencia menor y un menor número de datos a considerar en el proceso de interpolación.

Conclusiones

Se construyó un variograma experimental para los datos de precipitación de las normales climatológicas 1951-2010. En este proceso

Tecnología y ciencias del agua, 9(5), 209-222, DOI:10.24850/j-tyca-2018-06-10

se desarrolló un índice de variación espacial (como el cociente de dividir la diferencia de precipitación entre la distancia que separa las estaciones). Este índice cuya distribución estadística se muestra en la Figura 2 permitió identificar y eliminar 20 046 pares de estaciones con valores del índice superiores a 9.85, lográndose una mejor configuración del variograma experimental.

El variograma experimental se ajustó a un modelo esférico con parámetros: rango (a = 1 455), efecto de pepita (C_0 = 18 082), y umbral (C_1 = 602 452). Estos parámetros fueron calculados mediante un procedimiento iterativo desarrollado por los autores, utilizando el método de los multiplicadores de Lagrange.

El modelo esférico encontrado, puede servir de base para la interpolación mediante el método de kriging, con la gran ventaja que ofrece este método que permite calcular, no sólo los valores de la precipitación, sino también puede interpolar la varianza de estimación, característica que no tiene ningún otro método de interpolación.

Referencias

Freeden, W., Nashed, M. Z., & Sonar, T. (2015). *Handbook of Geomathematics*, (2nd ed.). Germany: Springer.

Golden Software, Inc. (2002). Surfer User's Guide. 809 14th Street, Golden, Colorado 80401-1866, U.S.A.

Hannachi, A. (2004). *A Primer for Empirical Orthogonal Functions Analysis of Climate Data*. Reading, UK: University of Reading.

Luenberger, D. G., & Ye, Y. (2016). *Linear and Nonlinear Programming*, (4th ed.). London, UK: Springer International Publishing.

Matheron, G. (1971). *The theory of regionalized variables and its applications*. París, Francia: École National supérieure des mines.

Oliver, M. A., & Webster, R. (2015). *Basic Steps in Geostatistics: The Variogram and Kriging*. London, UK: Springer International Publishing.

Thomson, R. E., & Emery, W. (2014). *Data Analysis Methods in Physical Oceanography* (3rd ed.). Amsterdam, Netherlands: Elsevier Science.
.

Wang, H. & Wang, Y. (2012). Analysis of Spatial Interpolation Methods. En: Quian, A., Cao, L., Su, W., Wang, T., Yang, H., (eds.) *Recent Advances in Computer Science and Information Engineering* (pp. 507-512). Berlin, Germany: Springer-Verlag.