INTERPOLACIÓN ESPACIAL DE DATOS METEOROLÓGICOS MEDIANTE TÉCNICAS DE *COKRIGING*. APLICACIÓN A LAS PRECIPITACIONES DE CASTILLA-LA MANCHA

Juan Andrés GARCÍA
Dpto. de Termodinámica. Facultad de Física. Universidad de Valencia.

RESUMEN
Este trabajo explica una forma de realizar el estudio de la variabilidad espacial de una variable climatológica. Se describe el método de interpolación basado en la técnica *cokriging* para la obtención de mapas, aplicándolo al estudio de las precipitaciones en Castilla-La Mancha.

Palabras clave: *Cokriging*, interpolación espacial, *kriging*, variograma

ABSTRACT
This work explains how to study the spatial variability of a climatological quantity. It describes the interpolation method based on the *cokriging* technic in order to be able to obtain maps. An application is made to the study of precipitation in Castilla-La Mancha (Spain)

Key words: Cokriging, kriging, spatial interpolation, variogram.

1. INTRODUCCIÓN

La interpolación de datos climatológicos consiste en la estimación del valor de la precipitación, temperatura o cualquier otro parámetro de estas características basándonos en observaciones de esa variable en la vecindad del punto espacial donde deseemos estimarla. Es una herramienta útil para la elaboración de mapas, dónde podemos visualizar la variabilidad espacial de la variable considerada.

No siempre va a ser posible la interpolación espacial de una variable, pues puede suceder que esa variable responda a un comportamiento estocástico, y por ello presente una alta variabilidad espacial difícil de entender para poder llevar a cabo la interpolación.

No todas las variables se interpolan del mismo modo. No es lo mismo, por ejemplo, estudiar la precipitación que la temperatura. Incluso en el caso de una misma variable, el procedimiento de interpolación puede cambiar según las condiciones de dicha variable. La precipitación, por ejemplo, se debe a distintos patrones de circulación, puede presentarse como consecuencia del paso de un frente u originarse por un proceso convectivo que dé lugar a un cumulonimbus local. Así pues, según sea la causa de la precipitación, la distribución espacial cambia mucho, y una precipitación tormentosa de carácter local, por ejemplo, puede caer en un área de 5 a 10 km^2 , mientras que un frente puede producir normalmente precipitaciones en un área mucho más extensa. Debido a esta alta variabilidad, no es recomendable obtener mapas con datos interpolados para cortos periodos de tiempo, por lo que estos mapas suelen hacerse, por ejemplo, para periodos estacionales. La temperatura tiene una menor variabilidad espacial, aunque puede presentar una alta variabilidad para situaciones particulares de carácter local, como puede ser el caso de una niebla muy localizada en un determinado lugar, o el efecto fohen, que produce variaciones muy importantes de la temperatura a uno y otro lado de una cordillera.

Vemos por tanto que es muy importante conocer la variabilidad espacial para poder desarrollar la interpolación de una variable, así como las causas que originan esa variabilidad, ya que éstas nos ayudarán a determinar el periodo de tiempo en que debe llevarse a cabo la toma de medidas, las cuales serán luego utilizadas en la interpolación.

2. PROCESO DE INTERPOLACIÓN

Todo modelo de interpolación espacial consta de tres etapas:

2.1. Análisis de la correlación de los datos

Este primer paso permite analizar la dependencia espacial existente en una variable consigo misma ó con otras, para ver si es posible la interpolación.

Es importante antes de llevar a cabo la interpolación buscar la posible existencia de correlaciones entre los datos de las variables que poseemos, y no sólo entre los datos de una misma variable, sino también entre variables distintas, ya que esto nos permite desarrollar posteriormente un estimador más fiable que haga que la estimación final de los valores interpolados tenga más fundamentos en los que apoyarse, de forma, que el valor estimado mediante el estudio de correlaciones entre distintas variables debe ser más fiable que el que se obtenga de una simple autocorrelación.

Para estudiar las correlaciones se utiliza el variograma o covariograma (MATERN,1986), según queramos estudiar la correlación de una variable consigo misma o con otras variables. El variograma es una función estadística que se basa en el estudio de la varianza de un conjunto de datos (de una misma variable), y que analiza la forma en que va cambiando ésta, según aumentemos ó disminuyamos la escala espacial, y con ello el número de datos. Esta función nos ayuda a analizar la desemejanza de los datos cuando vamos aumentando el área sobre la que se distribuyen. Así, cuando la dependencia entre los valores de la variable decrezca al aumentar el área, la varianza de ese conjunto de valores aumentará y el variograma presentará una tendencia creciente con el área, ó lo que es lo mismo, una tendencia creciente cuando la distancia media entre los puntos que estamos analizando aumenta. El variograma es una función siempre de valores positivos ó nulos, pues la varianza siempre es positiva ó nula. Un variograma creciente indicaría una buena correlación con la distancia, mientras que un variograma constante indicaría falta de esa correlación. El covariograma es similar al variograma pero con él se analiza cómo cambia la covarianza. El covariograma puede ser positivo, cuando los valores de las variables tienden a variar conjuntamente, negativo, si tienden a variar en direcciones contrarias, ó puede ser nulo, indicando que las variables no están correlacionadas.

La validez de los variogramas y covariogramas depende de dos factores principalmente:

2.1.a. *Estacionariedad de la media.*
Los variogramas y covariogramas son funciones arbitrarias, dependiendo de la distribución espacial y los valores de los datos, pero estas funciones deben ser estacionarias, es decir, si los valores de dichas funciones fluctúan de un punto a otro del espacio, éstos toman siempre un valor promedio idéntico en cada área.

Para evitar tener algún problema de no estacionariedad de la media, es aconsejable llevar a cabo una visualización de la distribución espacial de los datos que poseemos, para ver si la distribución es homogénea, o por el contrario, es heterogénea. Esto último sería un inconveniente, ya que impediría que este requisito se cumpliera y los variogramas no resultarían válidos. Es importante pues que los datos en los puntos conocidos estén regularmente distribuidos para llevar a cabo un método de interpolación basado en el variograma.

2.1.b. *Elección de las clases de distancia y número de parejas*
Tanto los variogramas como los covariogramas son funciones de la distancia, por lo que para su determinación es necesario establecer unas clases de distancias para el cálculo de la varianza en cada una de estas distancias. Para establecer las clases de distancias que deben tomarse se elige el criterio de que el número de clases de distancias debe ser suficiente para poder visualizar la tendencia del variograma, y que el mayor número de parejas de puntos usados en el cálculo hace más real el variograma.

Para que el variograma sea lo más fiable posible, es aconsejable retener un número de parejas de puntos igual ó mayor que 50 para cada clase de distancia, y un número de clases de distancias no inferior a 10. Generalmente, los valores más prácticos para llevar a cabo la interpolación son aquellos tomados desde cortas distancias, y por ello, las clases se eligen tomando la distancia nula como límite inferior de la primera clase, y una distancia igual a la mitad de la dimensión más grande del área estudiada como límite superior de la última clase, escogiendo las clases intermedias a intervalos iguales de amplitud.

2.2. Generación del estimador

Un estimador es una función que permite obtener un valor interpolado en un punto del espacio, a partir de las correlaciones existentes entre los valores de la variable o variables en puntos vecinos. Una vez analizadas las correlaciones y obtenidos los variogramas y covariogramas, hay que llevar a cabo un ajuste de estas funciones utilizando otras (modelos básicos) conocidas, de manera que podamos expresar los variogramas y covariogramas como combinación lineal de modelos básicos. A este ajuste se conoce por el nombre de L.M.C. (modelo lineal de corregionalización). Los modelos básicos más frecuentes que se utilizan son:

2.2.a. *Modelo exponencial*
Este modelo se aplica cuando la dependencia espacial decrece exponencialmente al aumentar la distancia, llegando a desaparecer por completo a una distancia infinita. En el variograma se obser-

varía un crecimiento logarítmico de la varianza cuando la distancia es pequeña, pero a medida que la distancia aumenta, la varianza tendería a un valor asintótico, haciéndose constante el variograma e indicando por tanto la falta de dependencia a grandes distancias. A la distancia a la cual la varianza alcanza un 95% de su valor asintótico se le conoce como distancia umbral, y es necesaria considerarla en el modelo.

2.2.b. *Modelo esférico*
Es un modelo muy similar al exponencial, se aplica cuando la dependencia decrece con la distancia, con la particularidad de que la dependencia desaparece para una distancia finita. También posee distancia umbral.

2.2.c. *Modelo gaussiano*
Igual que en el modelo exponencial, la dependencia espacial desaparece sólo a una distancia infinita, pero su característica principal es su forma parabólica en el origen, lo cual significa una variación espacial bastante suave de los valores de la variable para puntos vecinos. Posee una distancia umbral

2.2.d. *Modelo lineal*
La particularidad de este modelo es que no posee una distancia umbral, la dependencia decrece linealmente con la distancia, y el variograma aparece como una línea recta creciente desde el origen.

2.2.e. *Modelo nugget*
Este es un modelo totalmente diferente de los demás, y que se utiliza cuando hay una falta total de dependencia espacial de una variable, siendo la distancia diferente de cero. La función es discontinua en el origen donde salta desde un valor nulo hasta un valor constante.

Estos modelos son los más comúnmente utilizados para obtener el L.M.C., pero existen otros que por tener un uso menos frecuente no los mencionaremos ahora.

La elección de los modelos básicos se lleva a cabo según sea la forma de los variogramas y covariogramas. Es necesario que el L.M.C. "encierre" adecuadamente a estas funciones, y para ello se usa el criterio de minimización de la diferencia cuadrada de las medias de los valores de los variogramas y covariogramas modelizados. De esta forma, si tenemos varios L.M.C.s, nos quedaremos con aquel que proporcione un valor mínimo de la diferencia cuadrada de las medias (varianza).

En cuanto al número de modelos básicos que deben integrar el L.M.C es preferible usar el menor número posible de modelos (principio de parsimonia). Es muy frecuente utilizar dos modelos, correspondiendo uno de ellos al modelo *nugget,* pues es muy importante que el origen de los variogramas esté muy bien definido.

La construcción del L.M.C. es la etapa clave para llegar al estimador. Por ello es recomendable proceder detenidamente, comparando distintos L.M.C.s y observando cuál de ellos se ajusta mejor a los variogramas, dando importancia al ajuste para pequeñas distancias, ya que al final, los valores que tendrán mayor peso en la interpolación serán aquellos más próximos al punto que se desee, dependiendo este factor de peso del L.M.C.

Una expresión genérica del estimador es:

$$Z(x) = \sum \lambda_i Z(x_i)$$

con

$$\lambda_i = F(\sum b_i^n \text{(modelos básicos)}_n)$$

siendo,

$Z(x) \equiv$ valor estimado en el punto x,
$\sum Z(x_i) \equiv$ vecindad,
$\lambda_i \equiv$ peso del vecino $Z(x_i)$, y
$b_i^n \equiv$ contribución del (modelo básico)$_n$ en el L.M.C.

Los coeficientes se obtienen mediante un criterio de minimización de la varianza (MYERS,1982 y 1983). Una vez obtenido el estimador a partir del L.M.C, sólo queda la estimación de los valores en puntos no conocidos.

2.3. Estimación de los valores de la variable en puntos desconocidos

Para determinar el valor de la variable en un punto debemos saber qué coordenadas tiene ese punto, y qué vamos a considerar como la vecindad de ese punto. Para definir la vecindad de un punto, debemos hablar del tamaño y del número máximo de valores conocidos que determinan esa vecindad.

2.3.a. *Tamaño de la vecindad*

Se define como una distancia umbral entre el punto donde se lleva a cabo la estimación y los puntos donde son conocidos los valores.

El propósito principal de definir una vecindad es ahorrar tiempo de cálculo, sobre todo cuando tenemos ficheros muy voluminosos de datos conocidos, ya que de esta manera, excluimos valores conocidos que se encuentran muy lejos del valor a estimar y que contribuyen por tanto con un peso insignificante en el valor interpolado.

2.3.b. *Número máximo de valores conocidos*

También con el motivo de reducir el tiempo de cálculo, se designa un número máximo de valores conocidos dentro de la vecindad, de manera que sólo contribuirán en la estimación los puntos más cercanos.

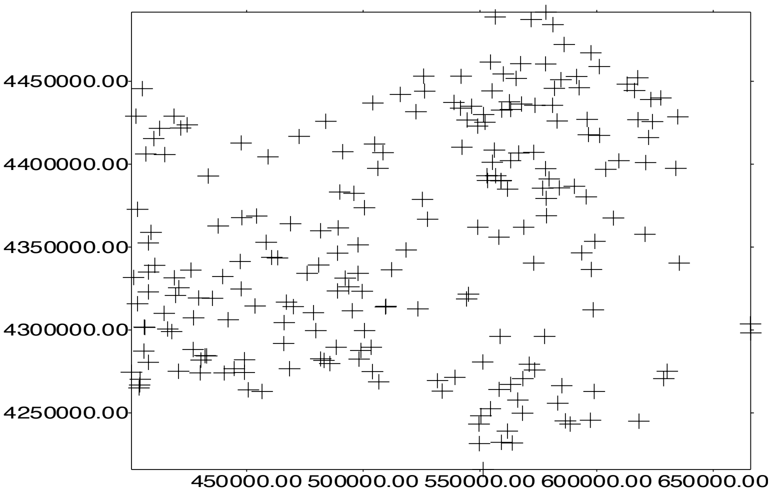
La elección del número óptimo de puntos considerados depende de la densidad de valores y de la forma de los variogramas y covariogramas. Si el número de puntos considerados es insuficiente el valor estimado no será óptimo, de manera que habrá que escoger un número mayor hasta conseguir que la varianza de las estimaciones no disminuya marcadamente cuando añadimos más puntos a la vecindad

Una forma de comprobar que nuestro estimador funciona correctamente es dándole las coordenadas de un punto del que conocemos el valor de la variable y hacer que estime dicho valor.

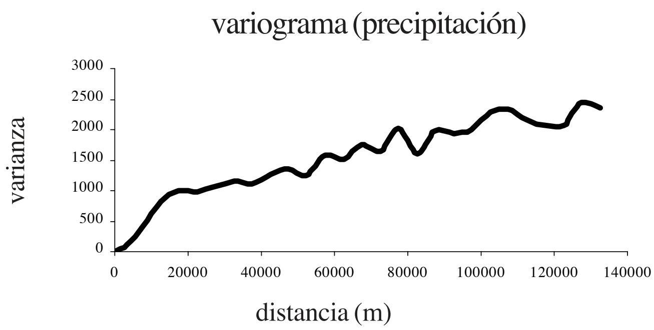
Esta forma de llevar a cabo la interpolación mediante el uso de variogramas y covariogramas, su ajuste obteniendo el L.M.C. usando el criterio de minimización de la varianza para las estimaciones se denomina *cokriging*. Un caso particular de esta técnica es el *kriging*, que se diferencia del *cokriging* en que tan sólo utiliza una variable para el estudio de la correlación, de manera que sólo se basa en la autocorrelación de esa variable para llevar a cabo la interpolación.

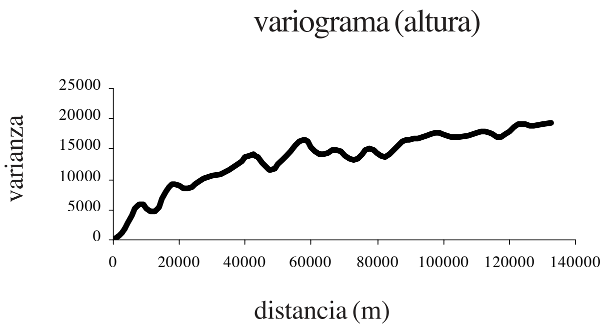
3. ESTUDIO PARTICULAR

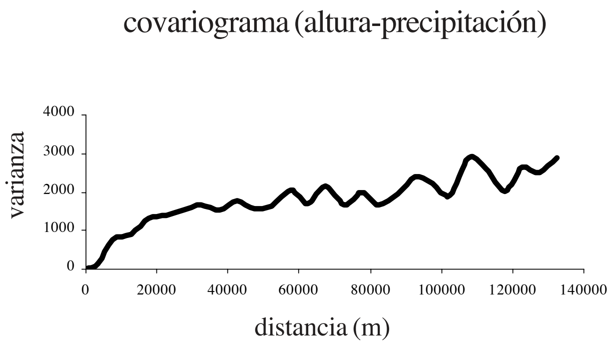
Este estudio trata sobre la distribución espacial de la precipitación en la zona central de Castilla-La Mancha incluyendo parte de las provincias de Albacete, Ciudad Real y Cuenca. Se han utilizado datos de precipitación registrada en invierno de 1990-1991, así como datos de longitud, latitud y altitud de las estaciones de medida. La siguiente gráfica muestra la distribución de esas estaciones:



Los variogramas obtenidos para el análisis de la correlación son:







Vemos que en todos ellos la dependencia espacial decrece de forma exponencial, por lo que utilizamos el modelo básico exponencial integrado con *nugget* para la obtención del L.M.C.

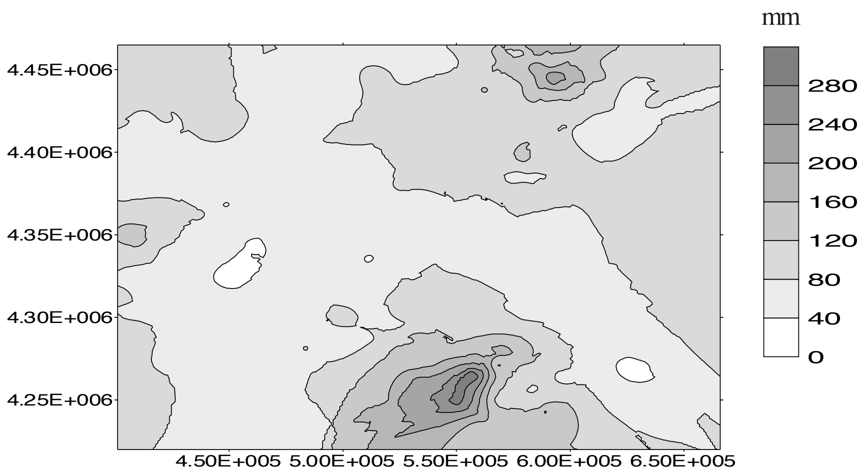
La interpolación la hemos llevado a cabo en los nodos de una malla que cubre aproximadamente la mitad del área de estudio, haciéndose así para evitar problemas de contorno en la interpolación, la resolución de un píxel de esta malla es de 1 km^2.

El mapa obtenido mediante el uso del programa Surfer y que expresa la variabilidad espacial de la precipitación (mm) se presenta en la página siguiente.

Para la estimación se ha escogido un tamaño de vecindad de 120 km, y 5 valores conocidos dentro de la vecindad, escogiendo estos valores, como ya dijimos, en función de la densidad de estaciones y la potencia de nuestro ordenador.

Vemos que aparecen sobre el mapa dos zonas de máxima precipitación, al sur y al norte, y otra zona intermedia de menor precipitación; las zonas de mayor precipitación se corresponden con lugares de mayor altitud.

Para comprobar la validez de los valores estimados, se han omitido algunas estaciones conocidas (elegidas al azar), siendo la diferencia media entre el valor real y el valor estimado de 10 mm.



REFERENCIAS BIBLIOGRÁFICAS

MATERN, B. (1986): *Spatial variation*, 2nd edition , Lecture Notes in Statistics, 36pp.

MYERS, D.E.(1982): *Matrix formulation of Cokriging*, Math, Gology 14, 249pp.

MYERS, D.E.(1983): *Estimation of linear combinations and Cokriging*, Math, Geology 15, 633pp.