

COMPARACIÓN DE MÉTODOS DE KRIGING MULTIVARIANTE PARA LA ESTIMACIÓN ESPACIAL DE LA TEMPERATURA DEL AIRE

POR

ROBERTO HERNÁNDEZ VARELA

Introducción

Una de las principales preocupaciones de cualquier servicio meteorológico es contar con información depurada y representada espacialmente, teniendo en cuenta las peculiaridades morfológicas de la región en la que incide. En este sentido, el carácter puntual de las observaciones obtenidas mediante la red de estaciones constituye una limitación para la correcta vigilancia y asistencia en las diversas tareas de planificación (hidrológica, agrícola, urbanística, etc.). Para acercarse a la realidad meteorológica en puntos no muestreados es entonces necesario aplicar técnicas de estimación espacial.

El contexto en el que se ha elaborado el presente trabajo se corresponde con el desarrollo de una herramienta de visualización de datos meteorológicos en tiempo real¹. El armazón metodológico elegido es la

¹ Este trabajo forma parte de un proyecto de investigación más amplio, denominado «Climatología del País Vasco», fruto de la colaboración entre el Dep. de Geografía de la Universidad del País Vasco y el Servicio Vasco de Meteorología (Dep. Transportes y Obras Públicas).

Roberto Hernández Varela. Dpto. Geografía. Universidad del País Vasco.

Estudios Geográficos, LXII, 243, 2001

Geoestadística, concretamente los algoritmos de *kriging*. Usados para hacer mapas, éstos no son significativamente mejores que otras técnicas de interpolación determinísticas preparadas para tener en cuenta la anisotropía y otras características espaciales de la variable objeto de estudio (Deutsch y Journel, 1998). Esta razón, unida a la generalmente mayor carga operacional, lleva a algunos autores a no justificar su empleo (Collins y Bolstad, 2000).

Con todo, las siguientes características de la Geoestadística avallan nuestra decisión: i) las medidas de variabilidad espacial (nube del semivariograma, etc.) son de utilidad en la validación de datos meteorológicos; ii) no ignora los patrones de dependencia espacial que normalmente se observan en las variables meteorológicas; iii) admite la incorporación de información secundaria correlacionada (parámetros geoclimáticos, otros sistemas de adquisición —satélites y radares meteorológicos—); iv) puede emplearse en la construcción de modelos de incertidumbre y en simulación estocástica; v) se adapta bien al diseño de redes de observación.

Bajo el supuesto de que la inclusión de la elevación del terreno mejora la representación espacial de la temperatura diezminutaria en determinados momentos del día, el presente artículo se detiene en tres métodos geoestadísticos multivariantes: *kriging simple con diversas medias locales*, *kriging con deriva externa* y *cokriging colocado*. Sus resultados son comparados mediante validación cruzada con los proporcionados por el caso univariado *kriging ordinario*². Todos ellos han sido programados por el autor en código IDL³.

Datos empleados

El ensayo de aplicación se ha realizado con datos registrados por la red hidrometeorológica automática del Servicio Vasco de Meteorología (S.V.M.) —Dep. Transportes y Obras Públicas del Gobierno Vasco—, y tiene como ámbito de estudio la C.A. del País Vasco. Con una resolución

² En aras de una mayor claridad, se ha seguido la notación matemática presentada por Goovaerts (1997) en su manual sobre Geoestadística aplicada a la evaluación de recursos naturales.

³ IDL es una marca registrada por Research Systems, Inc.

temporal de diez minutos (144 tomas diarias), cada estación envía en tiempo real a la sede central del S.V.M. los valores sensorizados por los higrotermotransmisores, además de otros atributos hidrometeorológicos (figura 1).

Para el ensayo de aplicación se han seleccionado las temperaturas registradas durante el día 24/04/99. Cumple con el requisito previo de observarse una tendencia con la elevación del terreno, algo por otra parte totalmente frecuente, no en vano dicho parámetro geoclimático es uno de los más determinantes en la distribución espacial de la temperatura. Se han extraído nueve registros diezminutarios con coeficientes de correlación próximos a -0.6, -0.7, -0.8, -0.9 y -0.95, afectados además de manera desigual por otra deriva en la dirección norte-sur (figura 2).

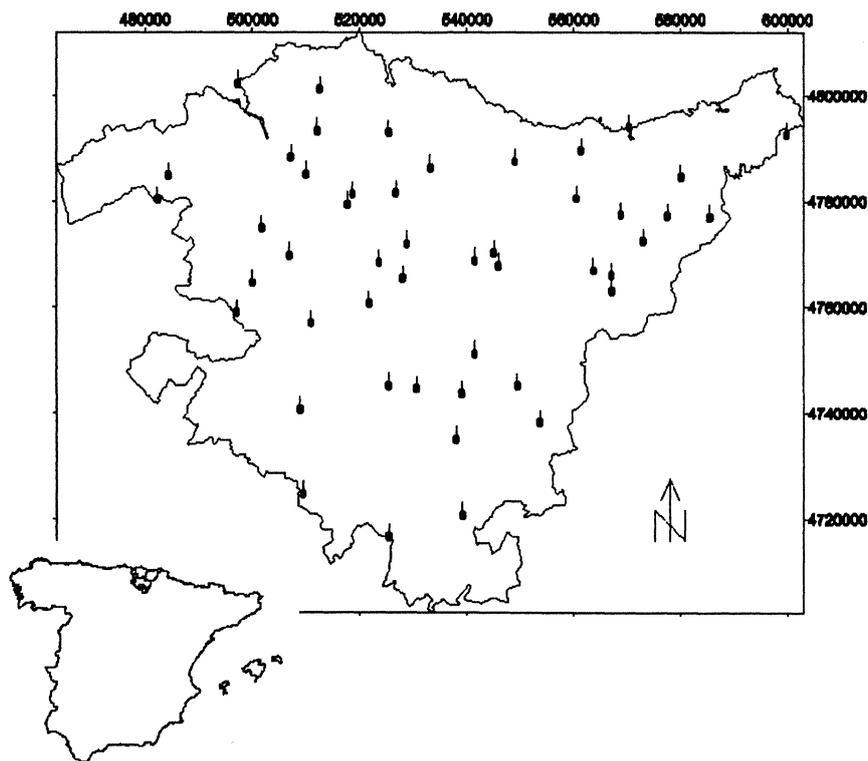


FIG. 1.—Localización de las estaciones hidrometeorológicas automáticas del Servicio Vasco de Meteorología disponibles vía radio durante el ensayo de aplicación.

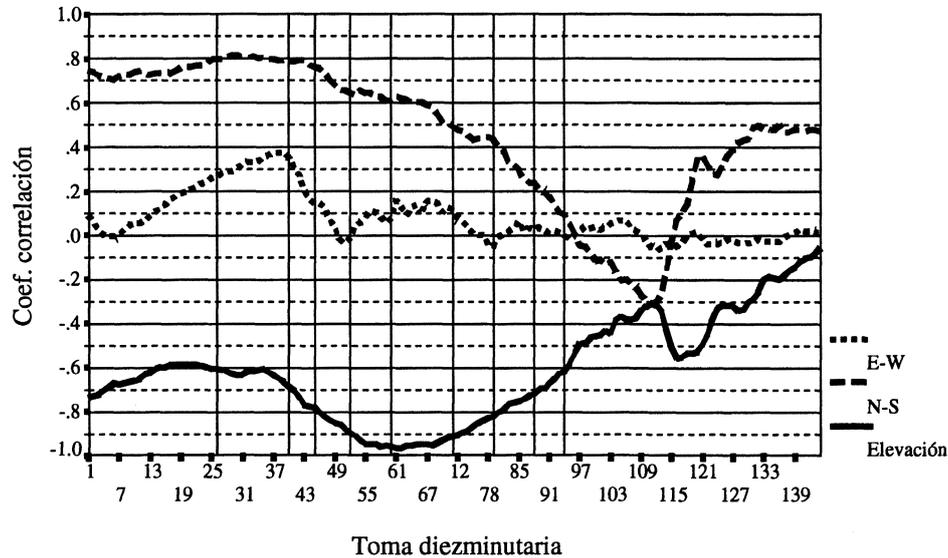


FIG. 2.—Correlación de la temperatura del aire con la elevación del terreno y las direcciones N-S, E-W durante el día 24/04/99, y tomas diezminutarias seleccionadas.

El concepto de función aleatoria y el paradigma del kriging

Nuestro problema consiste esencialmente en estimar el valor de la temperatura z en cualquier punto usando únicamente los datos z disponibles sobre el área de estudio A , es decir, los n datos $\{z(\mathbf{u}_a), a = 1, \dots, n\}$.

La Geoestadística interpreta los valores desconocidos $z(\mathbf{u})$ como realizaciones de una función aleatoria (FA). A través de este concepto, tales valores son tratados como un conjunto de variables aleatorias espacialmente dependientes.

Normalmente, una FA $Z(\mathbf{u})$ es descompuesta en un residuo $R(\mathbf{u})$ y una tendencia $m(\mathbf{u})$. La componente residual se modela como una FA estacionaria de media cero y covarianza $C_R(\mathbf{h})$ (Goovaerts, 1997, pp. 126):

$$E\{R(\mathbf{u})\} = 0$$

$$\text{Cov}\{R(\mathbf{u}), R(\mathbf{u} + \mathbf{h})\} = E\{R(\mathbf{u}) (R(\mathbf{u} + \mathbf{h}))\} = C_R(\mathbf{h})$$

El valor esperado o media de la FA Z en el punto \mathbf{u} es así el valor de la tendencia sobre ese punto:

$$E\{Z(\mathbf{u})\} = m(\mathbf{u})$$

Como se verá posteriormente, los distintos tipos de *kriging* analizados aplican modelos diferentes para caracterizar dicha tendencia $m(\mathbf{u})$. Todos ellos resultan variantes de una combinación lineal de los valores $z(\mathbf{u})$ y sus pesos asociados $\lambda(\mathbf{u})$, $\{\lambda_\alpha(\mathbf{u}), \alpha = 1, \dots, n\}$:

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) [Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)]$$

La función semivariograma

Como paso previo a la aplicación de los algoritmos de estimación espacial se debe calcular el semivariograma experimental, ajustarle un modelo teórico y estimar sus parámetros (figura 3). Se trata de una etapa importante dentro de este proceso, puesto que determina los pesos asignados a las observaciones. En ella, el investigador puede volcar sus conocimientos sobre el fenómeno de estudio. Este carácter subjetivo, precisamente, le ha costado a los métodos de *kriging* no pocas críticas (Philip y Watson, 1986).

El semivariograma experimental se calcula como la mitad del promedio de las diferencias al cuadrado entre componentes de parejas de datos (Goovaerts, 1997, pp. 28):

$$\gamma^*(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h})]^2$$

donde $N(\mathbf{h})$ es el número de parejas de puntos muestrales. Este vector da cuenta tanto de la distancia como de la dirección, por lo que la función semivariograma puede tomar en consideración direcciones de variabilidad dependientes.

Es frecuente que la disimilitud media incrementa también al hacerlo el espaciado entre las parejas de puntos muestrales, hasta un determinado rango. A partir de dicha distancia el semivariograma fluctúa al-

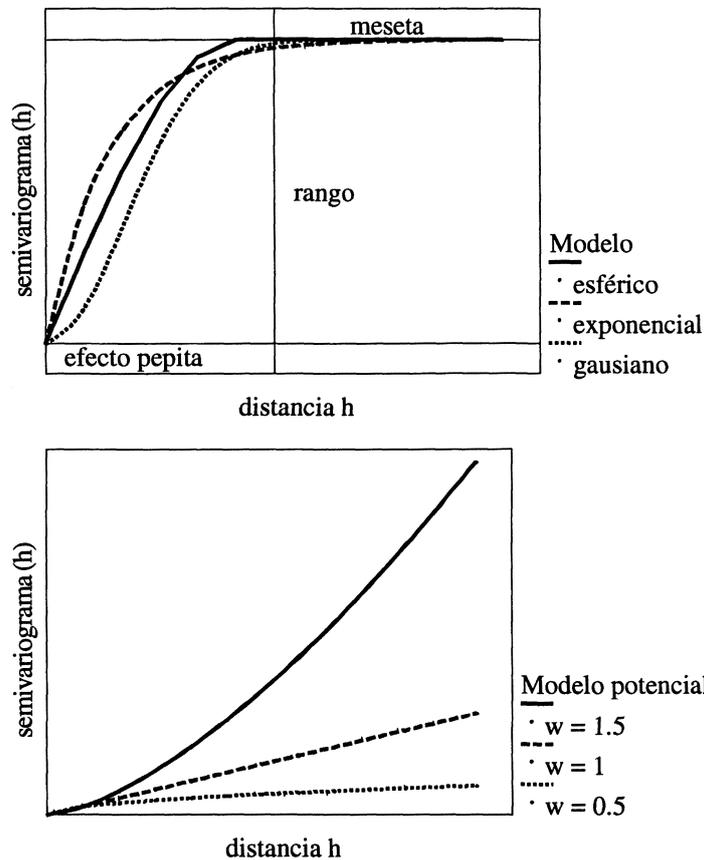


FIG. 3.—Modelos básicos de semivariogramas y sus parámetros. Gráf. sup.: modelos con meseta definida, para un mismo rango efectivo y efecto pepita. Gráf. inf.: modelo potencial, para distintos exponentes w .

rededor de una meseta y se considera que no existe correlación entre las observaciones (variable estacionaria, intrínseca). En otras ocasiones, el semivariograma no llega a estabilizarse dentro del dominio de estudio y crece al aumentar h (variable no estacionaria).

El comportamiento cerca del origen del semivariograma indica el tipo de continuidad de la variable regionalizada. Si es discontinuo en el origen existe efecto pepita, lo que implica que hay, bien fuentes de variación espacial a distancias inferiores al menor intervalo de muestreo, bien errores de medida, o una combinación de ambos efectos.

En cuanto al ajuste del semivariograma experimental, aunque no se recomienda el uso de procedimientos de modelado automático, el trabajo en tiempo real no deja otra alternativa. Aquí, se ha empleado una función iterativa no lineal de mínimos cuadrados, a la que se le suministra un modelo teórico de semivariograma expresado en derivadas parciales (función CURVEFIT de IDL). En la práctica, sin embargo, nos estamos encontrando con problemas de convergencia cuando el semivariograma experimental está compuesto por valores muy dispersos. Este efecto podría paliarse usando versiones más robustas del mismo.

Actualmente estamos contemplando como alternativa el uso del *semivariograma climatológico*, el cual ha demostrado su operatividad en aplicaciones en tiempo real para la visualización de campos de lluvias (Bastin et al., 1984).

Análisis estructural

Los semivariogramas presentados en la figura 4 capturan el grado de disimilitud media entre observaciones correspondientes a una realidad, la temperatura del aire, que es cambiante también en el tiempo. Queda patente que no puede aplicarse la misma hipótesis sobre su variabilidad espacial a lo largo del día. Para las primeras tomas diezminutarias el semivariograma no presenta meseta, indicativo de que la variabilidad espacial tiene lugar de acuerdo con una tendencia general. Efectivamente, la temperatura del aire va disminuyendo de forma progresiva de norte a sur. Lo más acertado es aplicar la hipótesis no intrínseca, que implica que el valor esperado de la función aleatoria depende de \mathbf{u} , y sus incrementos de primer orden $[z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h})]$ no son estacionarios.

A medida que la tendencia con dicho eje se debilita, los semivariogramas van presentando mesetas y alcances mejor definidos. Parece adecuado entonces aplicar la hipótesis intrínseca, que supone que los incrementos tienen valor esperado y varianzas definidas e independientes de \mathbf{u} para todo vector \mathbf{h} .

El contraste térmico entre costa e interior es, efectivamente, uno de los factores de diferenciación climática más fácilmente perceptible. Este hecho aparece reflejado en los semivariogramas direccionales en forma de anisotropía zonal (figura 5). La dirección de mayor continuidad,

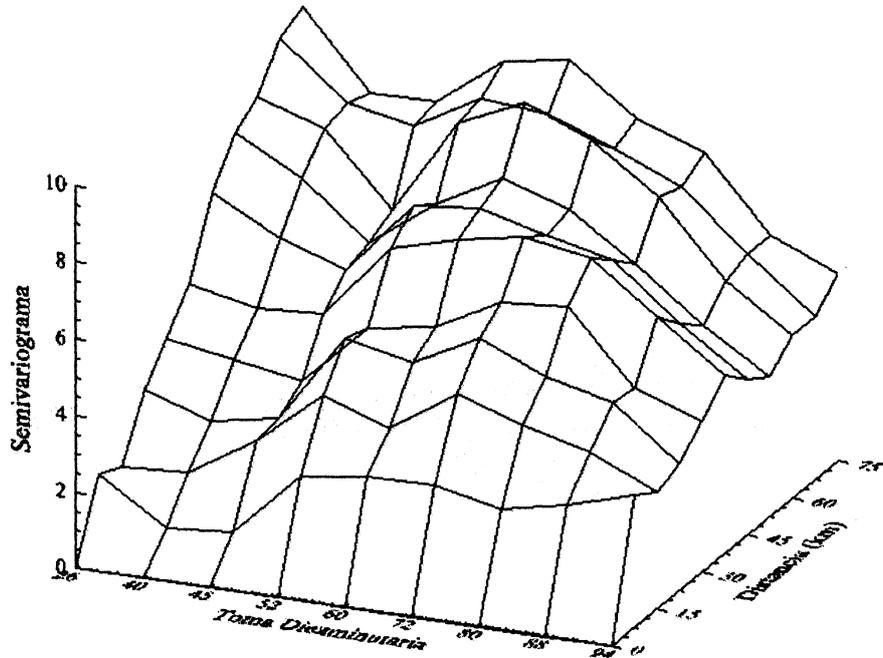


FIG. 4.—Semivariogramas experimentales omnidireccionales de las 9 tomas seleccionadas.

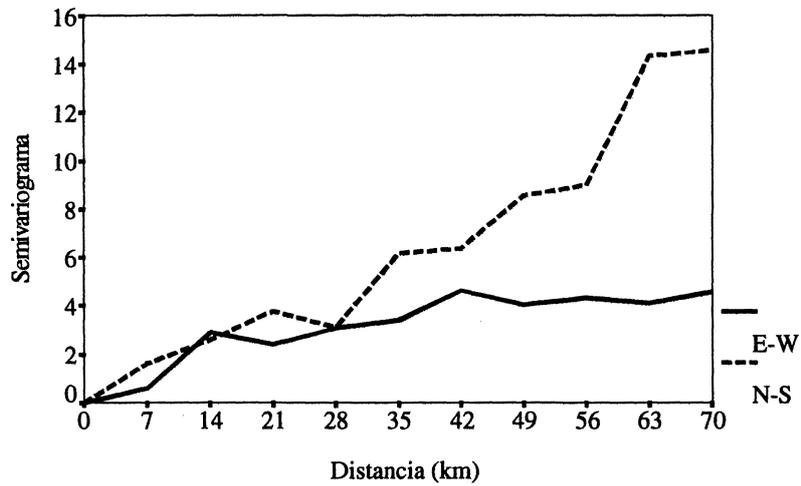


FIG. 5.—Semivariogramas experimentales direccionales (toma diezminutaria 45).

allí donde la hipótesis de estacionariedad se encuentra más cercana a cumplirse, es la este-oeste. Los semivariogramas calculados para dicho eje serán los incluidos en la fase del *kriging*.

Finalmente se aplicará *kriging* en un vecindario $W(\mathbf{u})$ para todas las realizaciones seleccionadas de $Z(\mathbf{u})$, bajo el supuesto de que tal FA es localmente intrínseca (cuasi-estacionaria), es decir, que la media $m(\mathbf{u})$ es constante o su rango de variación pequeño dentro del vecindario de estimación.

Se ha determinado que el vecindario $W(\mathbf{u})$ esté formado por las 8 estaciones más próximas al punto a estimar \mathbf{u} . Para ello no se ha empleado la distancia euclidiana, sino la distancia de semivariograma, de forma que la selección se realiza a lo largo de la dirección de máxima continuidad. La selección de un vecindario a priori tan pequeño obedece a que se ha pretendido ahorrar tiempo de cálculo, necesario en una aplicación en tiempo real, además de primar el carácter local de la estimación.

En cuanto al ajuste de los semivariogramas, el modelo más frecuentemente utilizado ha correspondido al *exponencial*, descrito como sigue (Deutsch y Journel, 1998, pp. 25):

$$\gamma(h) = c \cdot \left[1 - \exp\left(-\frac{3h}{a}\right) \right]$$

donde a es el rango efectivo y c la meseta.

Una vez realizado el modelado de los semivariogramas, los parámetros se introducen en los sistemas de *kriging*. Por razones de eficiencia computacional, dichos sistemas suelen resolverse en términos de covarianzas. Se atiende para ello a la relación $C(\mathbf{h}) = C(0) - g(\mathbf{h})$.

Estimación local univariada

Para el *kriging simple* (KS) la media $m(\mathbf{u})$ es constante y conocida a lo largo del área de estudio A . Su sistema se escribe en términos de covarianzas como sigue:

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_{\beta}^{KS}(\mathbf{u}) C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) = C(\mathbf{u}_{\alpha} - \mathbf{u}) \quad \alpha = 1, \dots, n(\mathbf{u})$$

El *kriging ordinario* (KO) considera la media $m(\mathbf{u})$ constante únicamente en el vecindario local $W(\mathbf{u})$, pero desconocida. Este estimador se escribe como una combinación de sólo los $n(\mathbf{u})$ datos observados $z(\mathbf{u}_\alpha)$:

$$z_{KO}^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{KO}(\mathbf{u}) z(\mathbf{u}_\alpha) \quad \text{con} \quad \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{KO}(\mathbf{u}) = 1$$

Como en todas las variantes del *kriging*, los pesos del KO $\lambda_{\alpha}^{KO}(\mathbf{u})$ se determinan de forma que se minimice la varianza de estimación, $\text{Var}\{Z_{KO}^*(\mathbf{u}) - Z(\mathbf{u})\}$, asegurando la insesgadez del estimador, $E\{Z_{KO}^*(\mathbf{u}) - Z(\mathbf{u})\} = 0$. Su sistema está formado por $(n(\mathbf{u}) + 1)$ ecuaciones lineales con $(n(\mathbf{u}) + 1)$ incógnitas, los $n(\mathbf{u})$ pesos $\lambda_{\alpha}^{KO}(\mathbf{u})$, más el parámetro de Lagrange $\mu_{KO}(\mathbf{u})$ incluido para la restricción de los pesos:

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_{\beta}^{KO}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu_{KO}(\mathbf{u}) = C(\mathbf{u}_\alpha - \mathbf{u}); \quad \alpha = 1, \dots, n(\mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_{\beta}^{KO}(\mathbf{u}) = 1 \end{cases}$$

Estimación local introduciendo información secundaria

Considérese ahora la inclusión de la elevación del terreno en la estimación del valor de la temperatura z en cualquier punto no muestreado \mathbf{u} . La información secundaria $y(\mathbf{u})$ se introduce como Modelo Digital de Elevaciones (MDE), de forma que es conocida tanto en los puntos de observación como en los puntos a estimar (figura 6).

Kriging simple con diversas medias locales.—Mediante el *kriging simple con diversas medias locales* (KSML) la variable Z en el punto \mathbf{u} se estima como la suma de una media local o tendencia primaria $m(\mathbf{u})$ y de un residuo $r(\mathbf{u})$ (Goovaerts, 1997, pp. 190):

$$z_{KSML}^*(\mathbf{u}) = m(\mathbf{u}) + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{KO}(\mathbf{u}) r(\mathbf{u}_\alpha)$$

donde los pesos $\lambda_{\alpha}^{KS}(\mathbf{u})$ se obtienen mediante *kriging simple*, introduciendo la función de covarianza $C_R(\mathbf{h})$ de la FA residual $R(\mathbf{u}) = Z(\mathbf{u})$

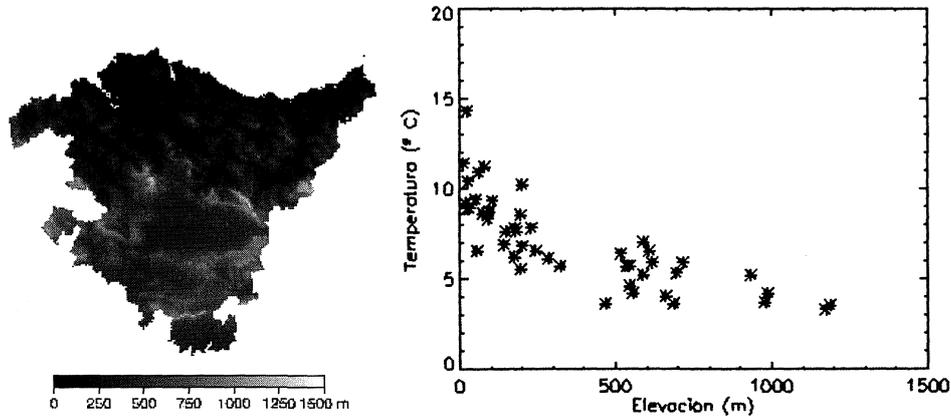


FIG. 6.—a) Modelo Digital de Elevaciones de la CAPV; b) gráfica de dispersión elevación-temperatura muestral (toma diezminutaria 45).

– $m(\mathbf{u})$, no la de $Z(\mathbf{u})$. En la figura 7 puede observarse el semivariograma de los residuos para la toma diezminutaria 45.

En el caso que nos ocupa, la media local $m(\mathbf{u})$ se calcula mediante la función lineal $f(y(\mathbf{u})) = a_0^* + a_1^* y(\mathbf{u})$, donde ambos coeficientes de regresión, a_0^* y a_1^* , se estiman a partir del conjunto de datos de temperatura y elevación coincidentes espacialmente $\{(z(\mathbf{u}_\alpha), y(\mathbf{u}_\alpha)), \alpha = 1, \dots, n\}$.

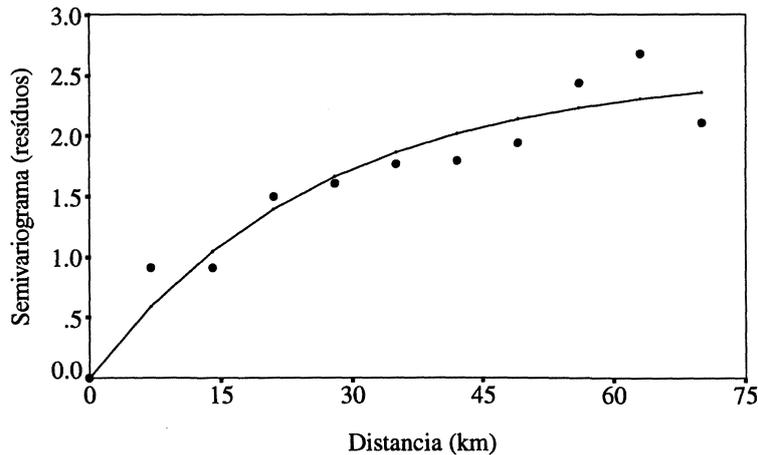


FIG. 7.—Semivariograma experimental de los residuos y modelo ajustado (toma diezminutaria 45).

Odeh *et al.* (1995), sin embargo, recomiendan aplicar *kriging ordinario* a los valores ajustados, de forma que se emplee más información local de las covariables.

Por la sencillez del KS, se trata quizá del método geoestadístico multivariante más frecuentemente utilizado, aunque bajo nomenclaturas distintas —*kriging residual modificado* (Samper y Carrera, 1990, Martínez-Cob, 1996), *regression-kriging* (Odeh *et al.*, 1995), *detrended kriging* (Phillips *et al.*, 1992).

Además, el KSML es fácilmente extensible a más de dos variables, no hay más que emplear una función de regresión múltiple para explicar $m(\mathbf{u})$. Así lo hacen, por ejemplo, Benichou y Le Breton (1987) en su conocido método *Aurelhy* (Analyse Utilisant le Relief pour l'Hydrométéorologie). En él sintetizan la morfología del terreno mediante un análisis de componentes principales, introducidas posteriormente en una función de regresión para calcular los campos de lluvias a distintas escalas temporales. Siguiendo este mismo criterio, Cornford (1998) selecciona 10 variables diferentes del terreno y las sintetiza en 5 componentes principales para interpolar la temperatura mínima diaria.

La generalización de los Sistemas de Información Geográfica y su integración con técnicas estadísticas está asistiendo todo este proceso (Ninyerola *et al.*, 2000).

Ahora bien, de acuerdo con Balairón (1995), la inclusión de variables secundarias debe hacerse si realmente aportan una mejora sustancial en la explicación de la estructura de distribución espacial del atributo primario. En caso contrario, la solución del método resulta más costosa y los resultados de la estimación no son significativamente más precisos.

Kriging con una deriva externa.—Es un método también documentado adecuadamente en la literatura al respecto, y con aplicaciones en ámbitos distintos. Hudson y Wackernagel (1994) preparan el mapa de la temperatura de enero en Escocia introduciendo la elevación como deriva externa. Encuentran que la bondad del método es superior a la del *kriging universal*. Armstrong *et al.* (1993) al no encontrar relación entre la elevación y la precipitación mensual registrada en Lesoto, optan por introducir los mapas de precipitación media estival o invernal.

Al igual que el KSML, el *kriging con una deriva externa* (KDE) define la tendencia primaria $m(\mathbf{u})$ como una función lineal de la variable secundaria $y(\mathbf{u})$:

$$m(\mathbf{u}) = a_0^* + a_1^* y(\mathbf{u})$$

Sin embargo, mientras que en el primer método los coeficientes de tendencia a_0^* y a_1^* se calculan una sola vez e independientemente del sistema de *kriging*, en el KDE se obtienen implícitamente mediante el sistema de *kriging* dentro de cada vecindario $W(\mathbf{u})$.

El estimador KDE se expresa de la siguiente manera:

$$z_{KDE}^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{KDE}(\mathbf{u}) z(\mathbf{u}_{\alpha})$$

donde $\lambda_{\alpha}^{KDE}(\mathbf{u})$ es el peso asignado al valor $z(\mathbf{u}_{\alpha})$. Su cálculo pasa por la resolución del siguiente sistema de $(n(\mathbf{u}) + 2)$ ecuaciones lineales:

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_{\beta}^{KDE}(\mathbf{u}) C_R(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) + \mu_0^{KDE}(\mathbf{u}) + \mu_1^{KDE}(\mathbf{u}) y(\mathbf{u}_{\alpha}) = C_R(\mathbf{u}_{\alpha} - \mathbf{u}); \alpha = 1, \dots, n(\mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_{\beta}^{KDE}(\mathbf{u}) = 1 \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_{\beta}^{KDE}(\mathbf{u}) y(\mathbf{u}_{\beta}) = y(\mathbf{u}) \end{cases}$$

donde $\mu_0^{KDE}(\mathbf{u})$ y $\mu_1^{KDE}(\mathbf{u})$ son los dos parámetros de Lagrange que atienden a las restricciones impuestas a los pesos.

El KDE requiere la determinación a priori del semivariograma de los residuos $\gamma_R(\mathbf{u})$. Sin embargo, su cálculo no es directo, ya que los datos disponibles son los de z , no los residuos. En la práctica, $\gamma_R(\mathbf{u})$ es sustituido por el semivariograma de z , aproximación que parece no afectar significativamente a los resultados de la estimación (Samper y Carrera, 1990, pp. 218).

El hecho de evaluar localmente la relación entre la elevación y la temperatura permite tener en cuenta posibles cambios en el gradiente térmico altitudinal a lo largo del área de estudio. Ahora bien, los resultados del KDE son precisos si la tendencia local está representada adecuadamente, lo cual se consigue siempre que las estaciones incluidas dentro del vecindario de estimación cubran un amplio rango de elevaciones. En caso contrario, la pendiente de la tendencia puede hacerse prepon-

derante y conducir a extrapolaciones incorrectas. La solución pasa por hacer más robusto el método, incrementando el número de puntos en el vecindario, de forma general (Hudson y Wackernagel, 1994), o sólo en aquellos casos donde la pendiente supere un determinado umbral prefijado (Hernández, 2000). Este último caso ya ha sido apuntado por Daly et al. (1994) en su modelo PRISM (Precipitation-elevation Regressions on Independent Slopes Model), diseñado para la estimación de la precipitación media mensual y anual en áreas montañosas.

Cokriging colocado.—El *cokriging colocado* (CKC) es una forma reducida del *cokriging*, otra de las extensiones multivariantes del *kriging*. Se trata de un algoritmo que tiene en cuenta explícitamente la correlación espacial cruzada entre las variables primaria y secundaria, expresada en forma de semivariograma cruzado:

$$\gamma_{ij}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_{\alpha}) - z(\mathbf{u}_{\alpha} + \mathbf{h})] \cdot [y(\mathbf{u}_{\alpha}) - y(\mathbf{u}_{\alpha} + \mathbf{h})]$$

Surge con la intención de evitar la tediosa inferencia y modelado del *cokriging*, método que presenta además otros inconvenientes derivados de la frecuentemente mayor densidad de muestreo de los valores secundarios. Los datos secundarios cercanos o coincidentes espacialmente con los valores primarios a estimar tienden a amortiguar la influencia de los datos secundarios más alejados (Goovaerts, 1997, pp. 235; Deutsch y Journel, 1998, pp. 73).

La solución aportada por el *cokriging colocado* consiste en retener sólo el dato secundario más próximo a cada uno de los puntos a estimar.

La variante desarrollada aquí es la *ordinaria*, pero pueden construirse otras dependiendo del modelo de tendencia adoptado (véase por ej. Almeida y Journel, 1994, para el caso *simple*).

El estimador *cokriging colocado ordinario* se escribe como sigue:

$$z_{CKC}^*(\mathbf{u}) = \sum_{\alpha_1=1}^{n_1(\mathbf{u})} \lambda_{\alpha_1}^{CKC}(\mathbf{u}) z(\mathbf{u}_{\alpha_1}) + \lambda_2^{CKC}(\mathbf{u}) [y(\mathbf{u}) - m_2 + m_1]$$

donde m_1 y m_2 son la media de los datos primarios y secundarios, respectivamente. Como indica la expresión, el valor de la elevación $y(\mathbf{u})$ no proporciona información sobre la tendencia primaria en el punto \mathbf{u} ,

aproximación que realizaban los dos algoritmos vistos anteriormente —KSML y KDE—, sino que interviene directamente en la estimación de la temperatura $z(\mathbf{u})$.

Los pesos se obtienen resolviendo el siguiente sistema de $(n_1(\mathbf{u}) + 2)$ ecuaciones lineales:

$$\begin{cases} \sum_{\beta_1=1}^{n_1(\mathbf{u})} \lambda_{\beta_1}^{CKO}(\mathbf{u}) C_{11}(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_1}) + \lambda_2^{CKO}(\mathbf{u}) C_{12}(\mathbf{u}_{\alpha_1} - \mathbf{u}) + \mu^{CKO}(\mathbf{u}) = C_{11}(\mathbf{u}_{\alpha_1} - \mathbf{u}) & \alpha_1 = 1, \dots, n_1(\mathbf{u}) \\ \sum_{\beta_1=1}^{n_1(\mathbf{u})} \lambda_{\beta_1}^{CKO}(\mathbf{u}) C_{21}(\mathbf{u} - \mathbf{u}_{\beta_1}) + \lambda_2^{CKO}(\mathbf{u}) C_{22}(0) + \mu^{CKO}(\mathbf{u}) = C_{21}(0) \\ \sum_{\beta_1=1}^{n_1(\mathbf{u})} \lambda_{\beta_1}^{CKO}(\mathbf{u}) + \lambda_2^{CKO}(\mathbf{u}) \end{cases}$$

Este sistema requiere conocer la función de covarianza primaria $C_{11}(\mathbf{h})$, la función de covarianza cruzada $C_{12}(\mathbf{h})$, así como la covarianza secundaria para una distancia cero $C_{22}(0)$. Esta superior carga de modelización puede aliviarse mediante la aproximación markoviana (Almeida y Journel, 1994, pp. 568; Goovaerts, 1997, pp. 237):

$$C_{12}(\mathbf{h}) \cong \frac{C_{12}(0)}{C_{11}(0)} C_{11}(\mathbf{h})$$

Cuando ambas variables tengan varianzas significativamente diferentes es recomendable estandarizarlas a media cero y varianza uno. Así se ha hecho en el presente estudio. A título de ejemplo, en la figura 8 puede observarse el modelo lineal de correogionalización ajustado para la toma diezminutaria 45. Los tres semivariogramas se ajustaron con una función exponencial, comprobando la legitimidad del modelo.

Resultados de la estimación

El funcionamiento de los algoritmos vistos aquí se ha comparado mediante validación cruzada, procedimiento que consiste en considerar temporalmente cada una de las estaciones como un punto no muestreado y estimar su valor a partir del resto. El criterio de comparación es el error cuadrático medio (EMC), el cual mide el promedio de las dife-

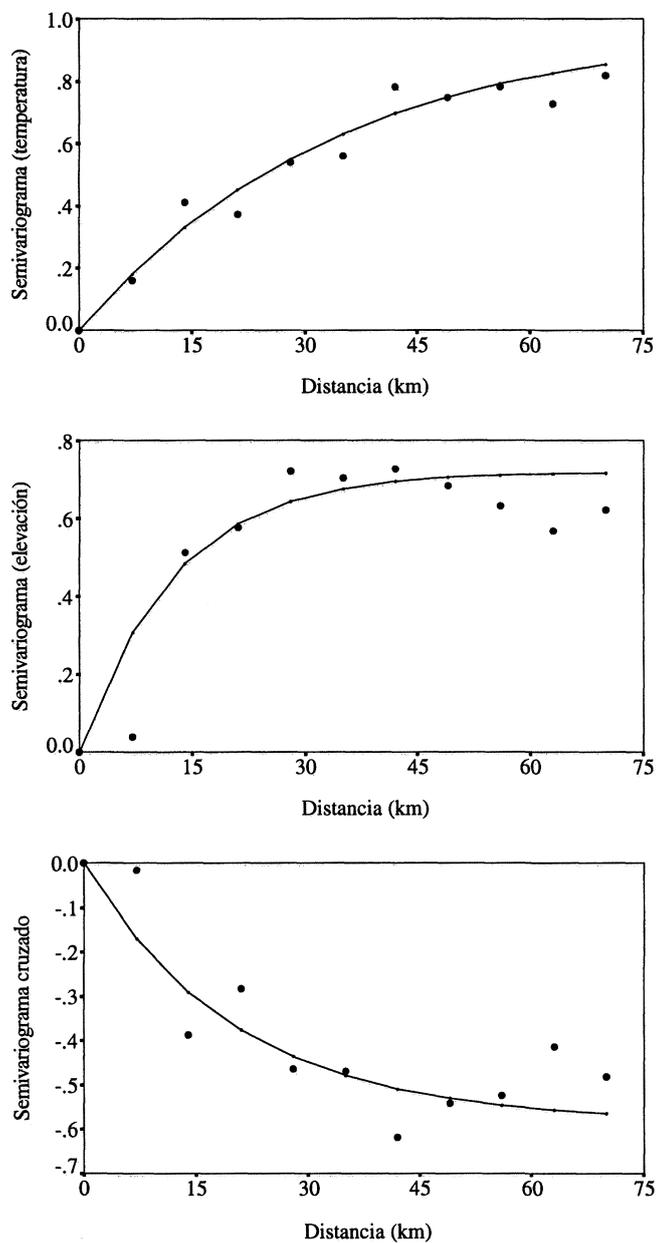


FIG. 8.—Semivariogramas experimentales estandarizados de la temperatura y de la elevación, y su semivariograma cruzado, con el modelo lineal de correlogramación ajustado (toma diezminutaria 45).

rencias al cuadrado entre el valor real de temperatura $z(\mathbf{u}_\alpha)$ y su estimación $z^*(\mathbf{u}_\alpha)$:

$$EMC = \frac{1}{n} \sum_{\alpha=1}^n [z(\mathbf{u}_\alpha) - z^*(\mathbf{u}_\alpha)]^2$$

donde $n = 48$ para el conjunto de estaciones disponibles durante el día de aplicación. El resultado de la estimación es bueno si este criterio es cercano a cero (tabla 1).

CUADRO I

ERROR MEDIO CUADRÁTICO SEGÚN TIPOS DE KRIGING. $W(\mathbf{u}) = 8$.

Toma 10'	KO	KSML	KDE	CKC
26	1,08	1,46	1,18	1,09
40	0,47	0,5	1,07	0,48
45	0,78	0,72	0,89	0,8
52	1,04	0,36	0,66	0,61
60	1,15	0,2	0,32	0,48
72	1,24	0,71	0,7	0,76
80	0,74	0,83	0,82	0,66
88	0,89	0,97	1,1	0,72
94	0,73	0,73	0,78	0,55

La suposición de que $Z(\mathbf{u})$ es localmente intrínseca y la selección del vecindario $W(\mathbf{u})$ según la dirección de mayor continuidad han resultado ser unas alternativas adecuadas para el tratamiento de aquellas tomas diezminutarias afectadas por la deriva en la dirección norte-sur.

El algoritmo que ignora la elevación, el KO, ofrece a primera vista resultados globales aceptables. No obstante, se comporta como un filtro de paso bajo, suavizando en exceso los detalles locales de la variación espacial de la temperatura. La gráfica de los valores observados frente a los errores de estimación revela un sesgo positivo (figura 9). Los valores más bajos se sobreestiman, mientras que los valores más altos se subestiman. La estimación es especialmente pobre en las zonas más abruptas, insuficientemente muestreadas por la red.

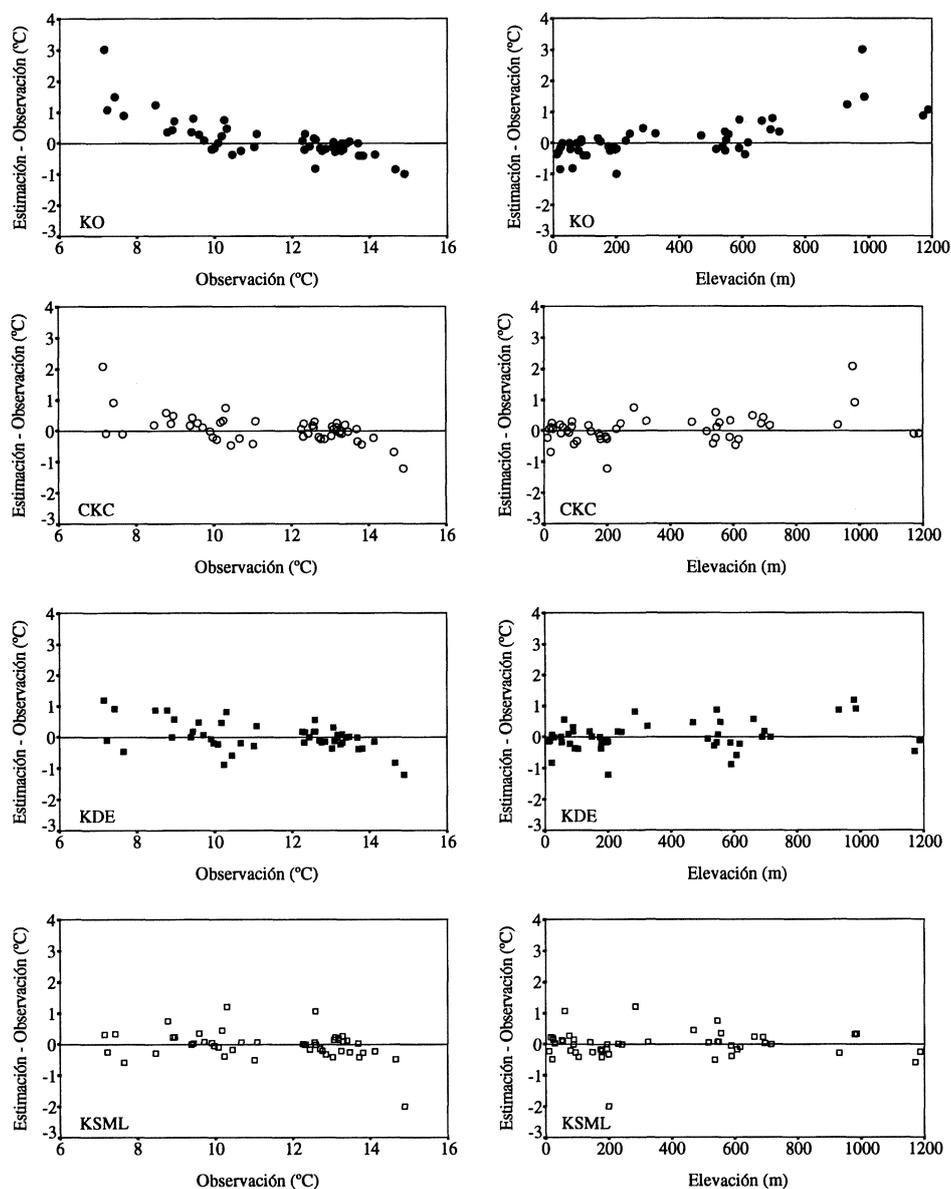


FIG. 9.—Gráfica de dispersión entre temperatura observada y error de estimación, y entre elevación del terreno y error de estimación (promedio de las 9 tomas diezminutarias seleccionadas).

Estas características tienen su reflejo en la cartografía resultante⁴. Tomando como ejemplo la toma diezminutaria 45, parece claro asumir que el comportamiento lineal se extiende por encima de los 1200 m, altitud máxima de la que se dispone información (figura 6b). Las estimaciones no superan, sin embargo, el valor termométrico alcanzado en dicha cota (figura 10).

Los problemas de sesgo del KO desaparecen en mayor o menor medida al emplear información secundaria correlacionada con el atributo

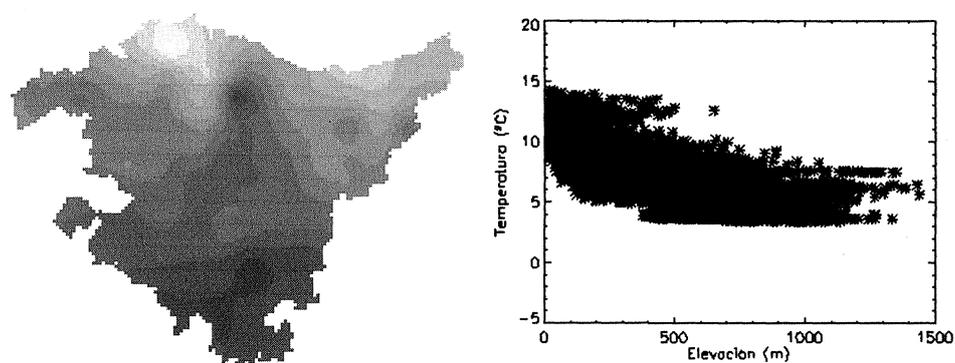


FIG. 10.—a) mapa de temperatura estimada por KO ($W(u) = 8$); b) gráfica de dispersión elevación-temperatura estimada por KO ($W(u) = 8$).

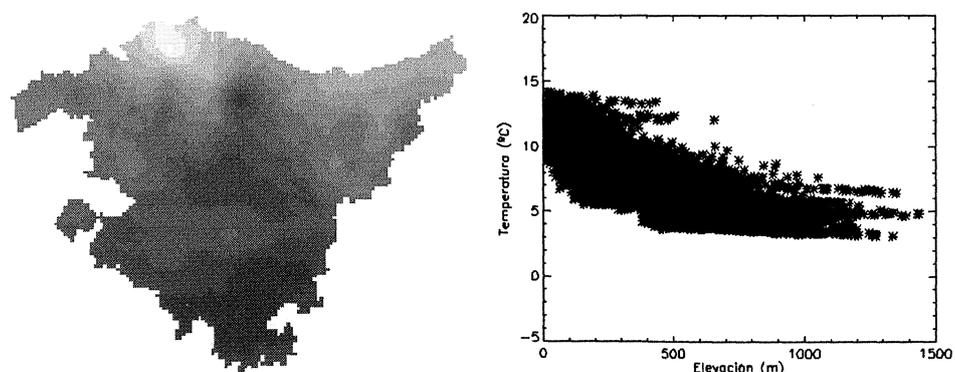


FIG. 11.—a) mapa de temperatura estimada por CKC ($W(u) = 8$); b) gráfica de dispersión elevación-temperatura estimada por CKC ($W(u) = 8$).

⁴ Los mapas de temperatura presentados son producto de la aplicación del *kriging* sobre los nodos de una malla regular de 1 km², coincidentes con los puntos de un Modelo Digital de Elevaciones, obtenido a su vez por generalización de un modelo de 25 m (figura a). La escala de grises es unitaria y su rango es [-1 a 15 °C].

objeto de estudio. Como era de esperar, la ventaja de emplear algoritmos multivariantes aumenta a medida que la relación entre la temperatura del aire y la elevación del terreno es más fuerte. Con un coeficiente de correlación superior a 0.9, KSML, KDE y CKC superan en todo momento al KO. En estos casos, sorprende el hecho de que la mejor alternativa sea interpolar los residuos de la regresión global entre ambas variables y añadirlos al ajuste, es decir el KSML, por encima incluso de un método más sofisticado como el KDE, que tiene en cuenta los posibles cambios en el gradiente altitudinal de la temperatura dentro de cada vecindario de estimación.

La información sobre la elevación del terreno tiende a influenciar mucho la apariencia de los mapas resultantes del KDE y del KSML, espe-

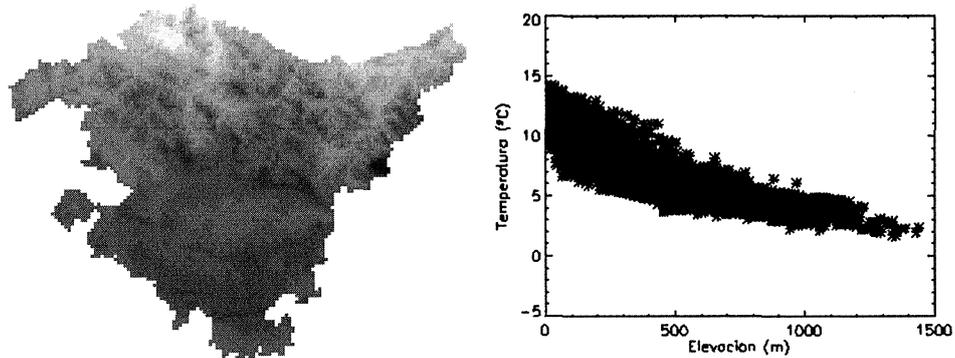


FIG. 12.—a) mapa de temperatura estimada por KDE ($W(u) = 8$ ó 16); b) gráfica de dispersión elevación-temperatura estimada por KDE ($W(u) = 8$ ó 16).

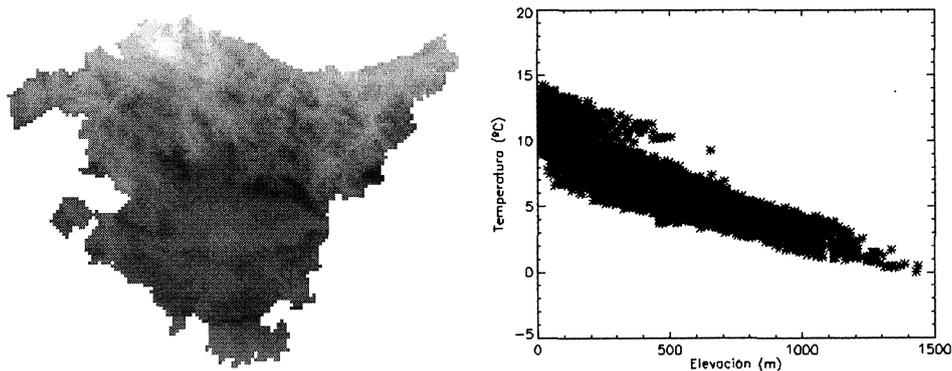


FIG. 13.—a) mapa de temperatura estimada por KSML ($W(u) = 8$); b) gráfica de dispersión elevación-temperatura estimada por KSML ($W(u) = 8$).

cialmente cuando el gradiente térmico altitudinal es fuerte. Ambos métodos tienen un comportamiento bastante robusto (figura 12 y 13).

Cuando la relación se debilita, el CKC saca partido de la introducción en su sistema de la dependencia entre las dos variables según se ha capturado por el semivariograma cruzado. Proporciona mapas no tan influenciados por la elevación del terreno, más cercanos a la apariencia de los realizados a partir de KO (figura 11). Si bien consigue reducir el mencionado sesgo positivo del KO, las estimaciones son menos robustas. Con todo, aunque exigente en cuanto a modelización se refiere, parece constituirse como la aproximación más acertada con correlaciones moderadas.

Conclusiones

La monitorización de la temperatura del aire en tiempo real, con intervalos cortos de tiempo, no puede obviar la correlación con parámetros geoclimáticos como la elevación del terreno.

En este sentido, la aplicación de métodos geoestadísticos multivariantes puede proporcionar mapas más efectistas, incluso con relaciones entre temperatura y elevación no demasiado fuertes.

La metodología empleada aquí, evidentemente, debe extenderse a tomas diezminutarias correspondientes a otros días, con patrones de continuidad espacial de la temperatura diferentes.

De igual modo, resultaría interesante incluir nuevas variables explicativas, como la orientación de ladera. Aunque los datos no nos lo expresaban así, en un día casi despejado como el elegido, parece lógico pensar en una cierta correlación según la exposición, fruto del distinto balance radiativo.

El proceso de monitorización meteorológica no se detiene con la obtención de una imagen digital, en este caso, representación espacial de la temperatura del aire. Ésta es susceptible de ser empleada como información secundaria exhaustiva en la predicción de otras variables. Piénsese, por ejemplo, en la humedad atmosférica. La capacidad del aire para contener vapor de agua depende directamente de la temperatura.

BIBLIOGRAFÍA

- ALMEIDA, A. S. y JOURNEL, A. G., 1994. "Joint simulation of multiple variables with a markov-type coregionalization model". *Math. Geol.*, 26, n° 5, 565-588.
- ARMSTRONG, M., CHETBOUN, G. y HUBERT, P., 1993. "Kriging the rainfall in Lesotho". A. Soares (ed.), *Geostatistics Troia '92*, Kluwer, 661-672.
- BALAIRÓN, L., 1995. "Análisis de la variabilidad espacial y temporal de los procesos meteorológicos". J.J. Ibáñez y C. Machado (ed.) *Análisis de la variabilidad espacio-temporal y procesos caóticos en ciencias medioambientales*, Logroño, Geoforma Ediciones, 243-258.
- BASTIN, G., LORENT, B., DUQUÉ, C. y GEVERS, M., 1984. "Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations". *Water Resour. Res.*, 20, n.º 4, 463-470.
- BENICHO, P. y LE BRETON, O., 1987. "Prise en compte de la topographie pour la cartographie de champs pluviométriques statistiques: la méthode Aurelhy". Ed. INRA, *Agrométéorologie des Régions de Moyenne Montagne*, Toulouse, 16-17 abril 1986, París, 51-69.
- CARREGA, P., 1995. "A method for the reconstruction of mountain air temperatures with automatic cartographie applications". *Theor. and Appl. Clim.*, Viena, Austria, 52 (1-2), 69-84.
- COLLINS, F. C. y BOLSTAD, P. V., 2000. "A comparison of spatial interpolation techniques in temperature estimation". *III International Conference/Workshop on Integrating GIS and Environmental Modeling*, Santa Fe, New Mexico, USA, 21-25 Jan 2000.
- CORNFORD, D., 1998. "Incorporating terrain information into spatialisation with particular reference to daily minimum air temperatures". En B. Gozzini y M. Hims (eds), *Workshop on Dealing on Spatialisation*, COST ACTION 79, Integration of Data and Methods in Agroclimatology, Toulouse, 24-25 Sep 1996, 69-96.
- DALY, C.; NEILSON, R. P. y PHILLIPS, D. L., 1994. "A statistical-topographic model for mapping climatological precipitation over mountainous terrain". *J. Appl. Meteor.*, 33, 140-158.
- DEUTSCH, C. V. y JOURNEL, A. G., 1998. *GSLIB. Geostatistical software library and user's guide*. Oxford University Press, New York.
- GOOVAERTS, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.
- GOOVAERTS, P., 2000. "Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall". *J. Hydrol.*, 228, 113-129.
- HERNÁNDEZ, R., 2000. "Estimación espacial de la temperatura del aire mediante kriging con deriva externa". En I. Aguado y M. Gómez (ed.) *Tecnologías geográficas para el desarrollo sostenible*, Alcalá de Henares, 20-22 septiembre 2000, 39-52.
- HUDSON, G. y WACKERNAGEL, H., 1994. "Mapping temperature using kriging with external drift: theory and an example from Scotland". *Int. J. Clim.*, 14, 77-91.
- MARTÍNEZ-COB, A., 1996. "Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain". *J. Hydrol.*, 174 (1-2), 19-35.
- NINYEROLA, M., PONS, X. y ROURE, J. M., 2000. "A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques". *Int. J. Clim.*, en prensa.

- ODEH, I. O. A., McBRATNEY, A. B. y CHITTLEBOROUGH, D. J., 1995. "Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging". *Geoderma*, 67, 215-226.
- PHILIP, G. M. y WATSON, D. F., 1986. "Matheronian geostatistics – Quo Vadis?". *Math. Geol.*, 18 (1), 93-117.
- PHILLIPS, D. L., DOLPH, J. y MARKS, D., 1992. "A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain". *Agr. Forest Meteor.*, 58, 119-141.
- SAMPER, F. J. y CARRERA, J., 1990. *Geoestadística. Aplicaciones a la hidrología subterránea*. Univ. Politécnica de Cataluña, Barcelona.

RESUMEN: *Comparación de métodos de kriging multivariante para la estimación espacial de la temperatura del aire.* En el presente artículo se desarrollan tres algoritmos geoestadísticos que incorporan información auxiliar exhaustiva en la estimación de un atributo primario: por una parte, *kriging simple con diversas medias locales* y *kriging con deriva externa*, que usan datos secundarios para caracterizar la tendencia espacial primaria; por otra, *cokriging colocado*, que atiende explícitamente a la correlación espacial cruzada entre ambos atributos.

Los distintos métodos se han ensayado con datos de temperatura del aire registrada por la red automática del Servicio Vasco de Meteorología, introduciendo la elevación del terreno como variable explicativa.

Los resultados se han comparado mediante validación cruzada. Se llega a la conclusión de que la aplicación de métodos geoestadísticos multivariantes puede proporcionar mapas más efectivos que el caso univariado, incluso con relaciones entre temperatura y elevación no demasiado fuertes.

PALABRAS CLAVE: geoestadística, semivariograma, kriging, temperatura, topografía, coregionalización.

ABSTRACT: *Comparing multivariate kriging methods for spatial estimation of air temperature.* In this paper three geostatistical algorithms are developed to incorporate exhaustive auxiliary information for the estimation of a primary attribute : *kriging with varying local means* and *kriging with external drift*, that use secondary data in order to characterize the primary spatial trend; and *collocated cokriging*, that explicitly account for the spatial cross-correlation between both attributes.

The different methods are tested by means of air temperature data sets from the automatic network of the Meteorological Basque Service, incorporating elevation as explicative variable.

Cross validation is used to compare the estimation performances. The main conclusion is that the application of multivariate geostatistical methods can improve the effectiveness of meteorological maps, even when relationship between temperature and elevation weakens.

KEYWORDS: geostatistics, semivariogram, kriging, temperature, topography, coregionalization.

RÉSUMÉ: *Comparaison des méthodes de krigeage multivariante pour l'estimation espacial de la température de l'air.* Dans cet article, on développe trois algorithmes géostatistiques qu'incorporent information auxiliaire exhaustive pour l'estimation d'un attribut primaire: *krigeage simple avec diverses mesures locales* et *krigeage avec dérivée externe*, qui emploient des données secondaires pour caractériser la tendance spatiale primaire; et *cokrigeage co-localisé*, qui tient compte de la corrélation spatiale croisée entre les deux attributs.

Les différentes méthodes sont essayées avec des données température de l'air enregistré par le réseau automatique du Service basque de Météorologie, en introduisant l'élévation du terrain comme variable explicative.

Les résultats sont comparés par validation croisée. La conclusion à laquelle on arrive est que l'application des méthodes géostatistiques multivariantes peut donner des cartes qui permettent d'interprétation plus claire que les cartes dérivées des cas univariants, même avec des relations température - élévation pas très fortes.

MOTS CLÉS: géostatistique, demi-variogramme, krigeage, température, topographie, corégionalisation.