

Estimación Espacial de la Temperatura del Aire mediante *Kriging* con Deriva Externa

Roberto Hernández Varela
Dep. Geografía. Universidad del País Vasco
e-mail: Inves-Meteo@ej-gv.es

1. Introducción

En la actualidad, la correcta evaluación de los distintos elementos del tiempo atmosférico y del clima requiere cada día mayor precisión. Además del propio interés científico, el contar con información meteorológica depurada y representada espacialmente, teniendo en cuenta las peculiaridades morfológicas de la región en la que incide, interviene positivamente en numerosas tareas de planificación hidrológica, agrícola, urbanística, etc.

Dicha información procede en gran medida de redes meteorológicas de superficie, de densidad frecuentemente baja y ubicación preferente en fondos de valle, lo que obliga a emplear algoritmos de estimación espacial para acercarse a la realidad meteorológica en aquellas zonas no muestreadas.

El objetivo del presente estudio¹ es la puesta a punto de un algoritmo perteneciente al campo de la Geoestadística, denominado *kriging con deriva externa* KDE. Se trata de un estimador que incorpora datos secundarios exhaustivamente muestreados para la caracterización de la tendencia espacial del atributo primario. Ha sido desarrollado por el autor en código IDL¹, y forma parte de un programa más extenso, cuyo fin es la visualización de datos meteorológicos en tiempo real.

En este caso, se ha ensayado la estimación de la temperatura del aire diezminutaria introduciendo la elevación del terreno como información secundaria. Teniendo como perspectiva la ejecución en tiempo real, se han establecido umbrales en el grado de correlación lineal entre ambas variables, que determinarán el uso de algoritmos univariados (*kriging ordinario* y *kriging con una tendencia*) o multivariados (*kriging con una deriva externa*).

2. Datos empleados

Los datos de partida corresponden a la red hidrometeorológica automática del Servicio Vasco de Meteorología (S.V.M.) -Dep. Transportes y Obras Públicas del Gobierno Vasco-. El ámbito de aplicación se ciñe, por tanto, a la C.A. del País Vasco. Con una resolución temporal de diez minutos (144 tomas diarias), cada estación envía en modo radio a la sede central del S.V.M. los valores sensorizados por los higrómetrotransmisores, además de otros atributos hidrometeorológicos (figura 1).

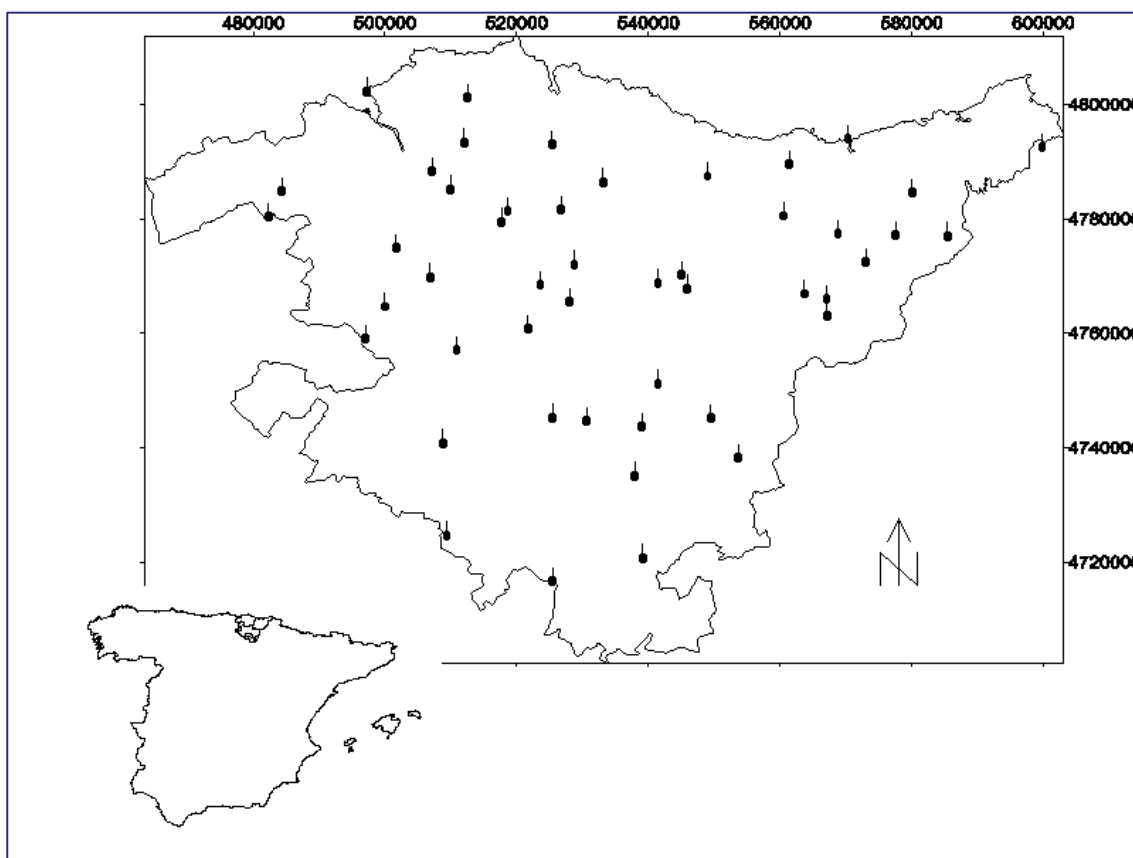


Fig. 1: Localización de las estaciones hidrometeorológicas automáticas del Servicio Vasco de Meteorología que transmiten vía radio.

Se han seleccionado los registros diezminutarios correspondientes a un determinado día. El requisito previo es que, durante el mismo, se encuentre presente una tendencia de la temperatura con la elevación del terreno. No es necesario buscar demasiado, no en vano dicho parámetro geoclimático es uno de los más determinantes en la distribución espacial de la temperatura. En concreto, el día elegido es el 24/04/99, durante el cual se observa además otra deriva en la dirección N-S (figura 2).

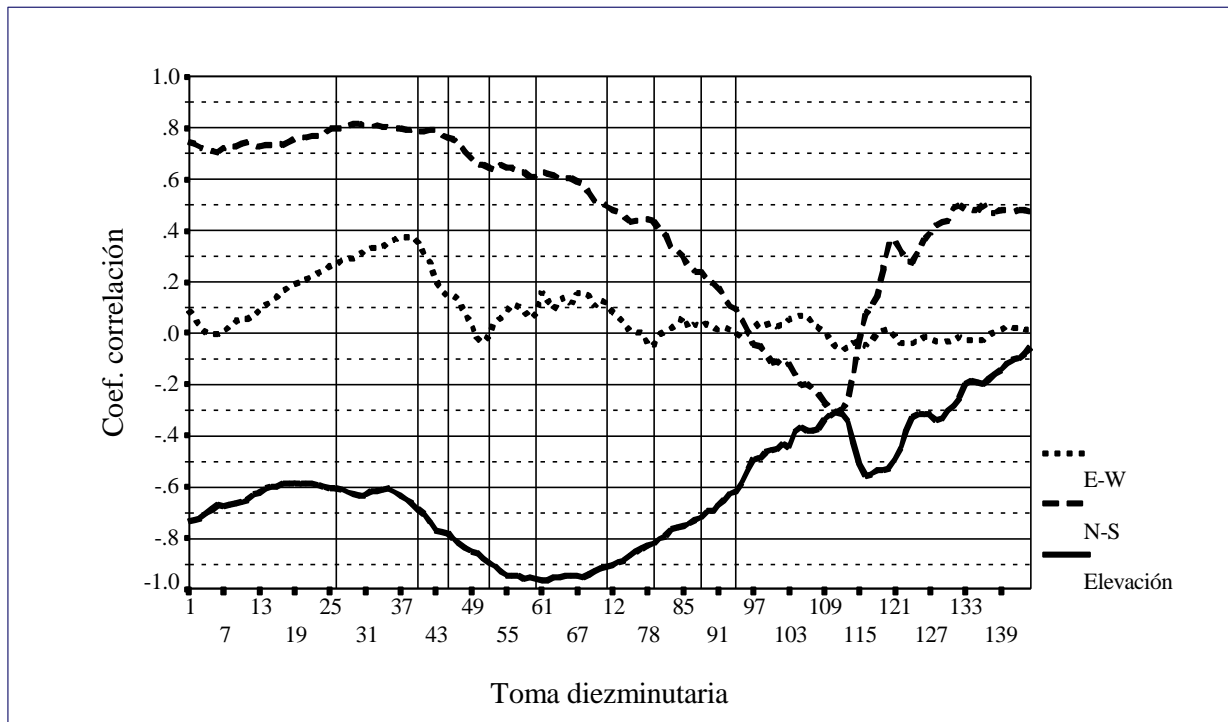


Fig. 2: Correlación de la temperatura con la elevación y las direcciones N-S y E-W durante el día 24/04/99, y tomas diezminutarias seleccionadas.

3. Análisis estructural

En lugar de la distancia euclidiana, el *kriging* emplea el semivariograma como medida de disimilitud entre observaciones. El semivariograma experimental se calcula como la mitad del promedio de las diferencias al cuadrado entre componentes de parejas de datos (Goovaerts, 1997, p. 28):

$$\mathbf{g}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{a=1}^{N(\mathbf{h})} [z(\mathbf{u}_a) - z(\mathbf{u}_a + \mathbf{h})]^2$$

donde $N(\mathbf{h})$ es el número de parejas de puntos muestrales un vector \mathbf{h} aparte. 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.

Como paso previo a la aplicación del *kriging* se debe elegir un modelo teórico de semivariograma y estimar sus parámetros. Aunque no se recomienda el uso de procedimientos de modelado automático (*black-box*), el trabajo en tiempo real no deja otra alternativa. El ajuste se ha realizado mediante una función iterativa de mínimos cuadrados no lineales, a la que se le suministra un modelo teórico de semivariograma expresado en derivadas parciales.

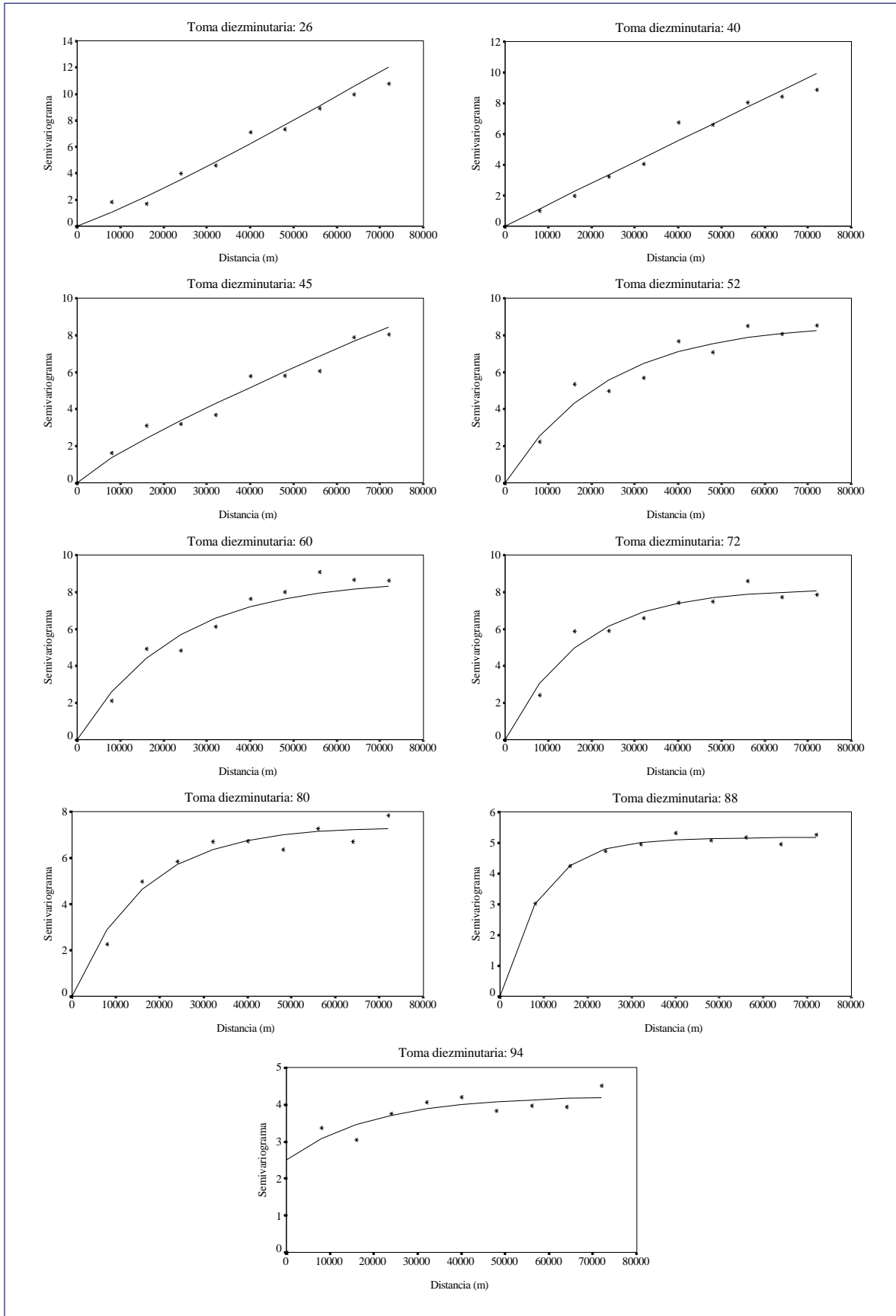


Fig. 3: Semivariogramas omnidireccionales y modelos teóricos ajustados.

Del análisis de los semivariogramas (figura 3) se deduce que no puede aplicarse a lo largo de todo el día la misma hipótesis sobre la variabilidad espacial de la temperatura. Los semivariogramas no acotados de las primeras tomas diezminutarias consideradas indican 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. Para estos casos, lo correcto es considerar la hipótesis no intrínseca, que implica que la esperanza matemática de la función aleatoria no es constante, ni sus incrementos de primer orden $[z(\mathbf{u}_a) - z(\mathbf{u}_a + \mathbf{h})]$ estacionarios. Una alternativa para su tratamiento es el *kriging con un modelo de tendencia*¹ KT. A medida que la tendencia con dicho eje desaparece, los semivariogramas van presentando mesetas y alcances mejor definidos. Parece adecuado entonces aplicar la hipótesis intrínseca, que supone que los incrementos tienen esperanza matemática y varianzas definidas e independientes de \mathbf{u} para todo vector \mathbf{h} , y tratar los datos mediante *kriging ordinario* KO (Samper y Carrera, 1990, p. 17-30).

4. El paradigma del *kriging*

Considérese el problema de estimar el valor de la temperatura z en cualquier punto no muestreado \mathbf{u} usando únicamente los datos z disponibles sobre el área de estudio A , es decir, los n datos $\{z(\mathbf{u}_a), a = 1, \dots, n\}$. En la práctica, sólo se retienen los $n(\mathbf{u})$ datos dentro de un vecindario dado o ventana $W(\mathbf{u})$ centrado en \mathbf{u} .

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 dos componentes: un residuo $R(\mathbf{u})$ y una tendencia $m(\mathbf{u})$. La componente residual se modela como una FA estacionaria con media cero y covarianza $C_R(\mathbf{h})$:

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

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

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

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

Los distintos tipos de *kriging* aplican para la tendencia $m(\mathbf{u})$ modelos diferentes. Así, por ejemplo, el KO explica las fluctuaciones locales de la media $m(\mathbf{u})$ limitando el dominio de la estacionariedad al vecindario local $W(\mathbf{u})$, en donde la media permanece constante pero desconocida. Para el KT, sin embargo, la media local desconocida $m(\mathbf{u})$ varía suavemente dentro de cada vecindario local $W(\mathbf{u})$, de forma que, la componente de tendencia es modelada como una combinación lineal de funciones $f_k(\mathbf{u})$ de las coordenadas.

¹ Se ha preferido emplear aquí esa denominación, en lugar de la también conocida *kriging universal*.

5. Kriging con una deriva externa

En el *kriging con una deriva externa* KDE la tendencia $m(\mathbf{u})$ se modela como una función lineal de una variable secundaria $y(\mathbf{u})$ (Goovaerts, 1997, p. 194-195):

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

Ambos coeficientes $a_0(\mathbf{u})$ y $a_1(\mathbf{u})$ se estiman implícitamente mediante el sistema de *kriging* dentro de cada vecindario $W(\mathbf{u})$. De esta manera, la relación entre la elevación y la temperatura se evalúa localmente, lo que permite tener en cuenta los posibles cambios en la correlación a lo largo del área de estudio (Goovaerts, 2000, p. 121).

El estimador KDE se expresa de la siguiente manera:

$$Z_{KDE}^*(\mathbf{u}) = \sum_{a=1}^{n(\mathbf{u})} I_a^{KDE}(\mathbf{u}) Z(\mathbf{u}_a)$$

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

$$\left\{ \begin{array}{l} \sum_{b=1}^{n(\mathbf{u})} I_b^{KDE}(\mathbf{u}) \mathbf{g}_R(\mathbf{u}_a - \mathbf{u}_b) + \mathbf{m}_0^{KDE}(\mathbf{u}) + \mathbf{m}_1^{KDE}(\mathbf{u}) y(\mathbf{u}_a) = \mathbf{g}_R(\mathbf{u}_a - \mathbf{u}); \quad \mathbf{a} = 1, \dots, n(\mathbf{u}) \\ \sum_{b=1}^{n(\mathbf{u})} I_b^{KDE}(\mathbf{u}) = 1 \\ \sum_{b=1}^{n(\mathbf{u})} I_b^{KDE}(\mathbf{u}) y(\mathbf{u}_b) = y(\mathbf{u}) \end{array} \right.$$

donde $\mathbf{m}_0^{KDE}(\mathbf{u})$ y $\mathbf{m}_1^{KDE}(\mathbf{u})$ son los dos parámetros de Lagrange que atienden a las restricciones impuestas a los pesos. Nótese que los datos de elevación $y(\mathbf{u}_a)$ y $y(\mathbf{u})$ no están incluidos directamente en el estimador, sino que intervienen en el sistema de *kriging*.

El KDE requiere la determinación a priori del semivariograma de los residuos $\mathbf{g}_R(\mathbf{u})$. Sin embargo, su cálculo no es directo, ya que los datos disponibles son los de z , no los residuos. En la práctica, $\mathbf{g}_R(\mathbf{u})$ es substituido por el semivariograma de z , intentando incluir pares de valores no afectados o levemente afectados por la tendencia. No obstante, esta solución tiende a sobrevalorar la varianza del error de estimación (Samper y Carrera, 1990, p. 218).

Para la óptima aplicación de este método la variable secundaria debe cumplir además una serie de premisas:

- i. Estar altamente correlacionada linealmente con la tendencia primaria, en caso contrario no debería incorporarse como una deriva externa (Hudson y Wackernagel, 1994, p. 85).
- ii. Conocerse tanto en los puntos primarios u_a como en todos los puntos u a ser estimados.
- iii. Variar de forma suave espacialmente con objeto de evitar inestabilidades en el sistema de KDE.

6. Resultados de la estimación

Se ha aplicado KO, KT y KDE a las tomas diezminutarias seleccionadas, repitiendo el proceso para dos vecindarios de diferente tamaño, compuestos por 8 y 16 estaciones. La bondad de sus resultados 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 medio cuadrático (EMC), el cual mide el promedio de las diferencias al cuadrado entre el valor real de temperatura $z(\mathbf{u}_a)$ y su estimación $z^*(\mathbf{u}_a)$:

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

donde $n = 48$ para el conjunto de la red hidrometeorológica del S.V.M. que emite vía radio. El resultado de la estimación es bueno si este criterio es cercano a cero.

Otorgando al estimador un carácter más local ($W(\mathbf{u}) = 8$), los métodos univariados ofrecen mejores resultados que el KDE siempre que el coeficiente de correlación entre la elevación y la temperatura sea inferior a 0.9 (tabla 1).

Toma 10'	Cf. Correlación $y(\mathbf{u}) / z(\mathbf{u})$	$W(\mathbf{u}) = 8$		$W(\mathbf{u}) = 16$	
		KO/KT*	KDE	KO/KT*	KDE
26	-0.61	1.06*	1.18	1.82*	2.0
40	-0.71	0.47	1.07	1.12*	1.04
45	-0.80	0.78	0.89	1.48	1.17
52	-0.91	1.04	0.66	2.36	0.77
60	-0.96	1.15	0.32	3.14	0.67
72	-0.91	1.24	0.7	3.08	1.1
80	-0.80	0.74	0.82	2.43	1.46
88	-0.70	0.89	1.1	2.72	1.68
94	-0.58	0.73	0.78	2.55*	1.64

Tabla 1: Error medio cuadrático para vecindarios de 8 y 16 estaciones y tipos de *kriging*. (*) indica que se ha empleado KT y que mejora los resultados del KO.

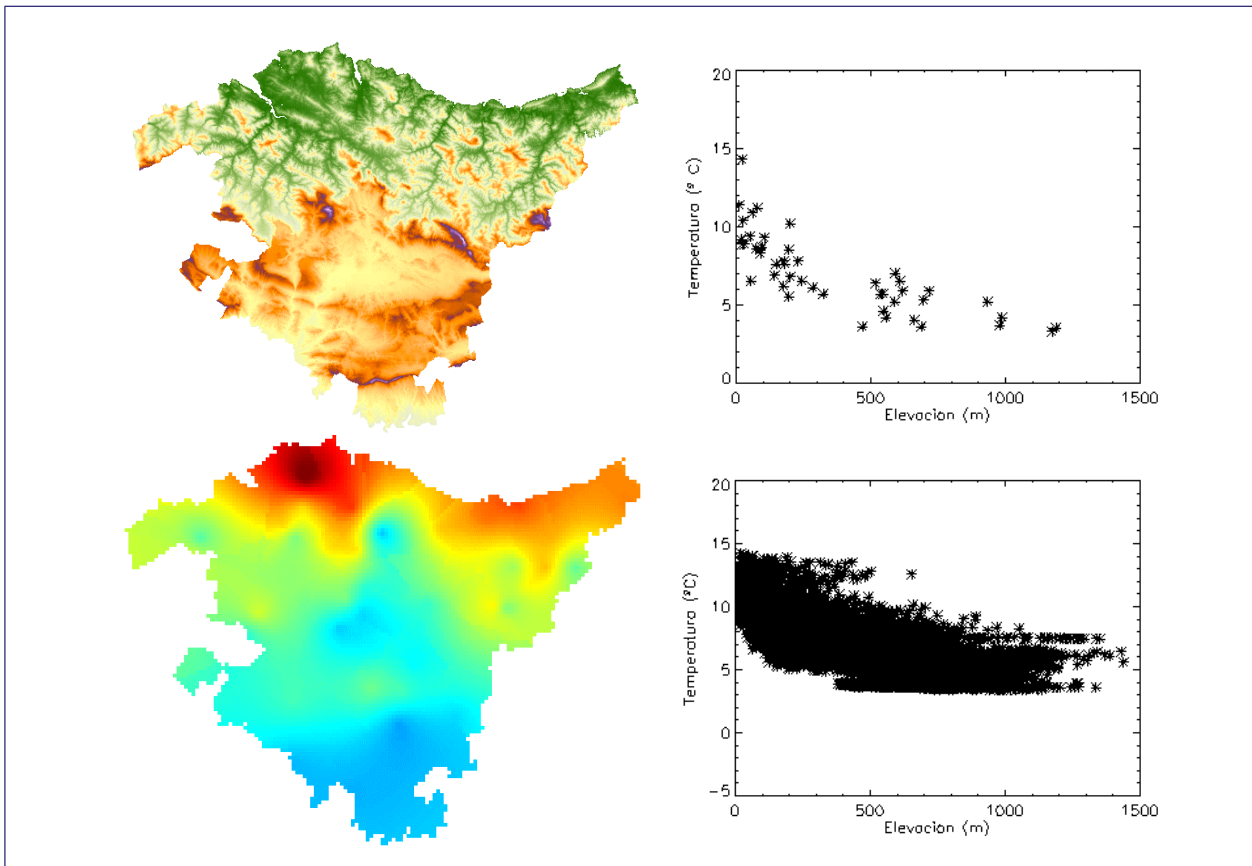


Fig. 5: a) Modelo digital de elevaciones de la C.A. del País Vasco; b) gráfica de dispersión elevación-temperatura muestral; c) mapa de temperatura estimada por KO ($W(u) = 8$); d) gráfica de dispersión elevación-temperatura estimada.

Esta primera conclusión, sin embargo, puede llevar a engaño a la hora de cartografiar la temperatura. Los mapas² producidos mediante KO o KT no son satisfactorios en las zonas de montaña, donde la densidad de las estaciones no es suficiente para reproducir los complejos patrones espaciales que se dan entre ambas variables (figura 5c). Como muestra la figura 5b, lo más lógico es asumir que el comportamiento lineal se extiende por encima de los 1200 m, altitud máxima de la que se dispone información. Las estimaciones no superan, sin embargo, el valor térmico alcanzado en dicha cota (figura 5d).

² 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 5a). Los datos corresponden a la toma diezminutaria 45. La escala de colores es única y su rango es [-1 a 15 °C].

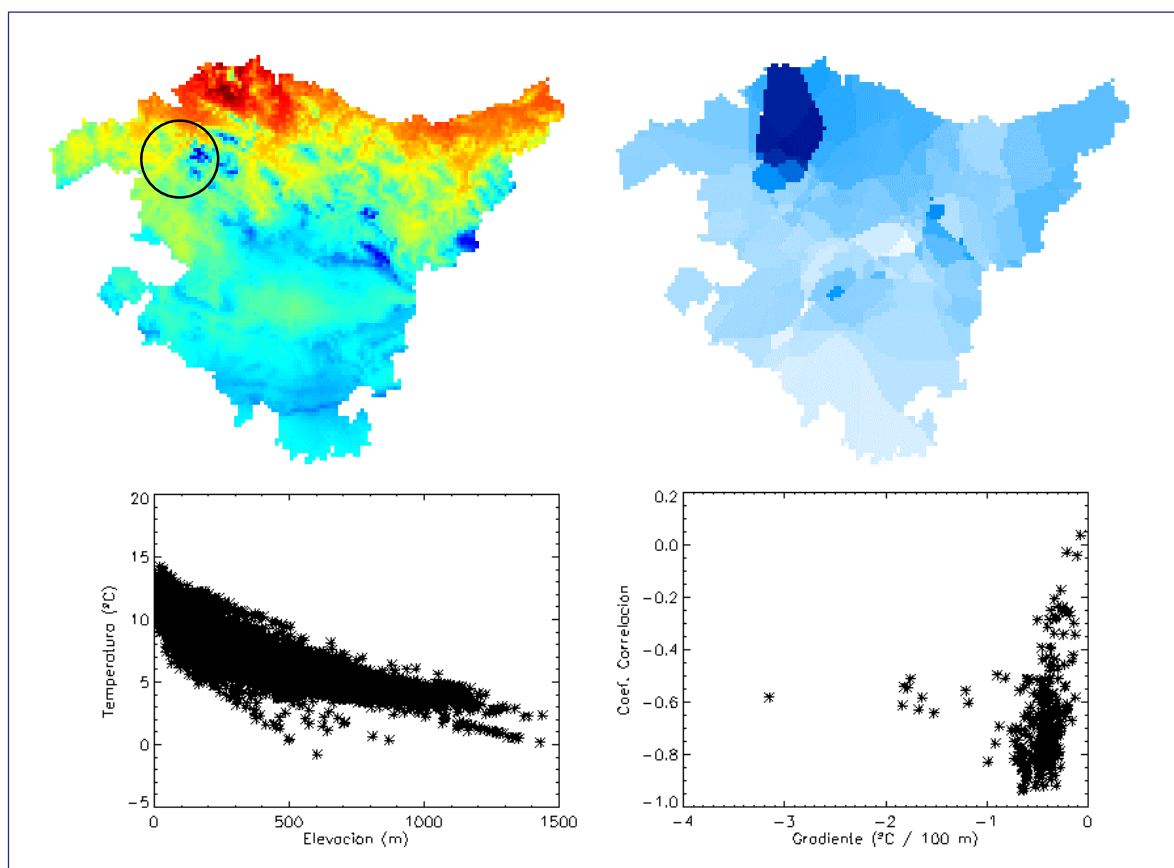


Fig. 6: Para $W(\mathbf{u}) = 8$. a) mapa de temperatura estimada por KDE; b) mapa de coeficientes $a_1^*(\mathbf{u})$, (los tonos más oscuros indican gradientes más elevados); c) gráfica de dispersión elevación–temperatura estimada; d) gráfica de dispersión coeficientes $a_1^*(\mathbf{u})$ -correlación elevación y temperatura.

Algo similar le sucede al KDE. Sus resultados son precisos si la tendencia local está bien representada, 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 preponderante y conducir a extrapolaciones incorrectas. A título de ejemplo, obsérvese la zona seleccionada en la figura 6a. El espacio dibujado entre los ríos Nervión y Kadagua presenta altitudes en torno a los 500 m, con picos cercanos a los 1000 m (Ganekogorta). Sin embargo, las estaciones que intervienen en el cálculo de la temperatura no superan los 200 m y dan lugar a un gradiente nada verosímil de 1 °C/100 m.

Además, con vecindarios mal representados el modelo de tendencia $m(\mathbf{u})$ no es capaz de poner de manifiesto la relación lineal entre la temperatura y la elevación en el punto \mathbf{u} (figura 6d).

La solución pasa por dotar al estimador de un carácter más global, es decir, incrementar el número de puntos en el vecindario. Con $W(\mathbf{u}) = 16$, tanto el gradiente como la asociación lineal no presentan valores fuera de rango (figura 7a). Se ha conseguido, incluso, que la precisión del KDE sea superior a la correspondiente al KO y KT, excepto en un solo caso. Con todo, esta solución no es del todo óptima: i) el EMC ha aumentado en los tres estimadores, muy especialmente en los univariados; ii) el tiempo necesario para resolver las ecuaciones aumenta considerablemente; iii) las estimaciones, al ser de carácter más global, pueden perder detalles interesantes.

Se hace necesario, por tanto, llegar a un compromiso entre la robustez de los vecindarios grandes y la precisión de los pequeños.

Calibración de la pendiente $a_1(\mathbf{u})$

La selección del vecindario es un punto crítico del proceso de estimación (Hudson y Wackernagel, 1994, p. 89). En este sentido, el cálculo de la pendiente $a_1^*(\mathbf{u})$ de la componente tendencia $m(\mathbf{u})$ empleada en la expresión del estimador KDE resulta de mucha utilidad. El estimador de $a_1^*(\mathbf{u})$ se escribe como sigue (Goovaerts, 1997, p. 196):

$$a_1^*(\mathbf{u}) = \sum_{a=1}^{n(\mathbf{u})} I_{a1}^{KDE}(\mathbf{u}) Z(\mathbf{u}_a)$$

donde $I_{a1}^{KDE}(\mathbf{u})$ es el peso asignado al dato $Z(\mathbf{u}_a)$, que se calcula resolviendo el sistema de $(n(\mathbf{u})+2)$ ecuaciones lineales:

$$\begin{cases} \sum_{b=1}^{n(\mathbf{u})} I_{b1}^{KDE}(\mathbf{u}) C_R(\mathbf{u}_a - \mathbf{u}_b) + m_{01}^{KDE}(\mathbf{u}) + m_{11}^{KDE}(\mathbf{u}) y(\mathbf{u}_a) = 0; & a = 1, \dots, n(\mathbf{u}) \\ \sum_{b=1}^{n(\mathbf{u})} I_{b1}^{KDE}(\mathbf{u}) = 0 \\ \sum_{b=1}^{n(\mathbf{u})} I_{b1}^{KDE}(\mathbf{u}) y(\mathbf{u}) = 1 \end{cases}$$

Nótese que el punto \mathbf{u} a estimar no aparece en este sistema de *kriging*, por lo que los coeficientes de la tendencia estimados $a_0^*(\mathbf{u})$ y $a_1^*(\mathbf{u})$ dependen sólo del vecindario empleado.

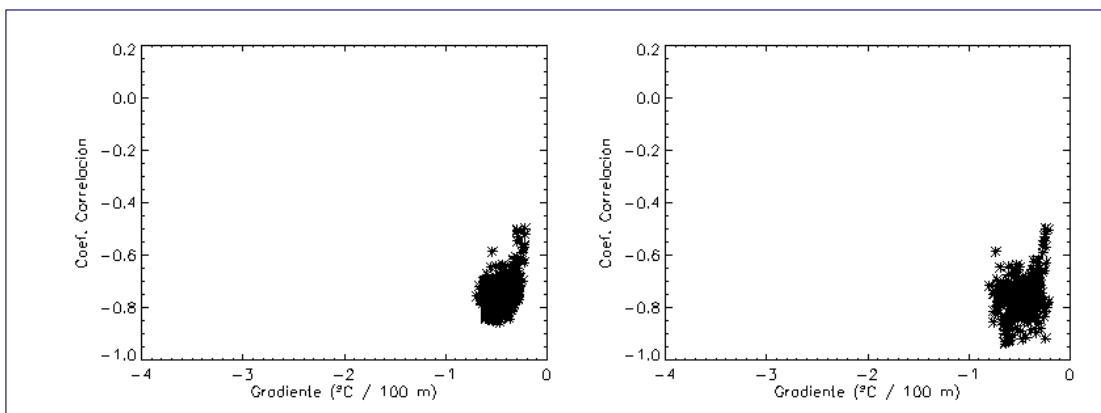


Fig. 7: Gráficas de dispersión coeficientes $a_1^*(\mathbf{u})$ -correlación elevación y temperatura. a) para ($W(\mathbf{u}) = 16$); b) para ($W(\mathbf{u}) = 8$ ó 16).

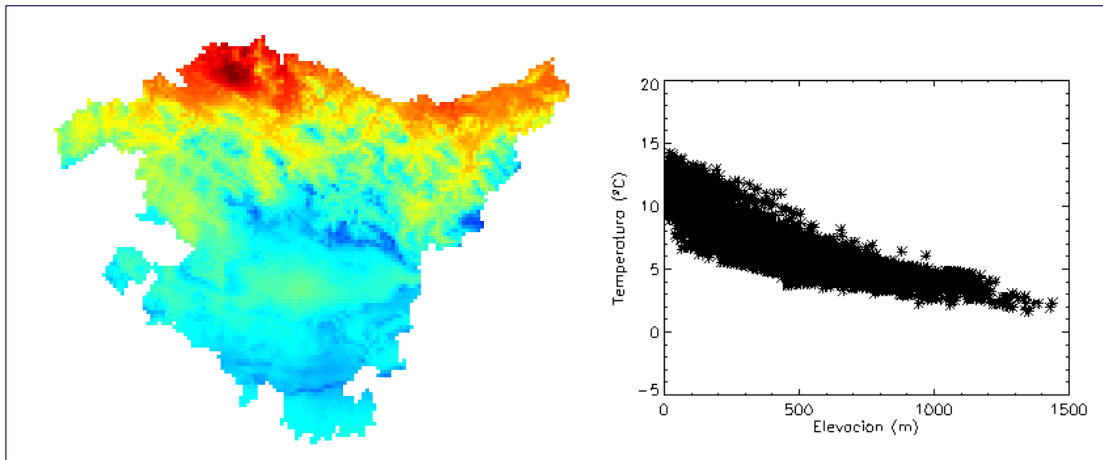


Fig. 8: Para ($W(u) = 8$ ó 16). a) mapa de temperatura estimada por KDE; b) gráfica de dispersión elevación-temperatura estimada.

Una vez conocidas las pendientes locales pueden filtrarse aquellos gradientes no realistas.

Para la toma diezminutaria presentada como ejemplo, se ha determinado que, siempre que el gradiente sea inferior a -0.8 °C/100m o que el coeficiente de correlación no supere 0.7, el vecindario constará de 16 estaciones, en caso contrario serán 8 las estaciones seleccionadas.

7. Conclusiones

La inclusión de la elevación del terreno mejora significativamente la representación espacial de la temperatura sensorizada a intervalos diezminutarios. Al ser un atributo secundario altamente muestreado, introducido en forma de MDE, complementa las lagunas muestrales de la red termométrica.

Para resolver el problema se ha hecho uso del *kriging con deriva externa*, un algoritmo que no ignora los patrones de dependencia espacial observados entre estaciones y que introduce en la predicción información secundaria correlacionada.

En el presente estudio se ha optado por hacer variable el número de estaciones implicadas en la estimación de cada uno de los puntos no muestreados \mathbf{u} ($W(\mathbf{u}) = 8$ y $W(\mathbf{u}) = 16$), con objeto de garantizar gradientes y tendencias locales realistas. De esta manera, se consigue la rapidez exigida por la ejecución en tiempo real, además de dotar al algoritmo de un carácter más local.

Como tareas para el futuro, señalar que este estudio debe extenderse hacia otras variables geoclimáticas, teniendo, no obstante, presente que no se debe primar la inclusión de multitud de variables secundarias si éstas realmente no aportan una mejora sustancial en la explicación de la estructura de distribución espacial de la variable primaria. De procederse así, la solución del método resulta más costosa y los resultados de la estimación no son significativamente más precisos (Balairón, 1995, p. 251). Cabe esperar, a priori, que las variables geoclimáticas involucradas sean escasas, puesto que con una resolución temporal de diez minutos se pierde el sentido de la relación mostrado en periodos de tiempo más largos (mensual y anual).

8. Referencias

- 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.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. New York, Oxford University Press.
- Goovaerts, P. (2000): Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, 228: 113-129.
- Hudson, G. y Wackernagel, H. (1994): Mapping temperature using kriging with external drift: theory and an example from Scotland. *International Journal of Climatology*, 14: 77-91.
- Samper, F.J. y Carrera, J. (1990). *Geoestadística. Aplicaciones a la hidrología subterránea*. Barcelona, Univ. Politécnica de Cataluña.

