



Comparación de dos métodos de interpolación para la estimación de la temperatura del aire aplicando técnicas geo-estadísticas.

Comparison of two interpolation methods to estimate air temperature applying geostatistical techniques.

KARIM QUEVEDO ¹ & KEVIN SÁNCHEZ ¹

¹ Servicio Nacional de Meteorología e Hidrología, SENAMHI, Casilla 11 1308, Lima 11, Perú.
kquevedo@senamhi.gob.pe

Resumen

El presente trabajo compara dos métodos de interpolación para la variable temperatura. Los modelos evaluados fueron el Kriging, en donde no se incluye la variable topográfica; y el de regresión lineal múltiple, donde sí es incluida la variable topográfica. Las metodologías han sido evaluadas mes a mes, tanto para temperatura máxima como para la temperatura mínima, para los doce meses del año con datos pertenecientes a la base de datos del SENAMHI con un registro no menor a 20 años, desde 1971 al 2000. El indicador utilizado para el cálculo de la exactitud del modelo fue el coeficiente de determinación (R^2), y para la validación cruzada se utilizaron el Sesgo (BIAS), Error Medio Cuadrático (RSME), y la eficiencia del modelo (ME). Los resultados muestran la ventaja del método de regresión lineal múltiple frente al Kriging, en función a su representación espacial, de acuerdo al método de validación utilizado, con resultados de R^2 para la temperatura mínima que van de 0.9 a 0.98; y para la temperatura máxima de 0.69 a 0.92; el BIAS, en temperatura máxima van de -0.1 a 0.3; y para la mínima de -0.3 a 0.8; en el RMSE, se observan mejores resultados en la temperatura mínima, con valores de 1.3 a 2.9; mientras que para la temperatura máxima los resultados van de 2.0 a 2.4; sin embargo se observa que entre los meses de mayo a agosto, los resultados son mejores para la temperatura máxima; y en cuanto a los resultados de ME, se muestran valores muy similares en ambas temperaturas, con valores que van de 0.8 a 0.9 en las temperaturas máximas, y de 0.9 a 1.0 en la temperatura mínima. Mientras que los resultados del Kriging muestran resultados de R^2 de 0.83 a 0.9 para temperatura máxima, y de 0.9 a 0.96 para temperatura mínima, en el RMSE de 3.4 a 4.4 para máxima, y de 2.5 a 3.2 para mínima; en el BIAS los resultados mostraron ser mejor para la mínima, con valores de 0.3 a 0.9, mientras que para la máxima estuvieron entre 0.6 a 1.4; y el ME mostro mejores resultados en la mínima con valores de 0.8 a 0.9, frente a 0.5 a 0.7 en la máxima.

En conclusión se puede decir que el método de la regresión lineal múltiple resulta ser un método bastante adecuado para representación de la variable temperatura en zonas con montaña. Mostrando una gran ventaja del método frente a cualquier método de interpolación que no considere la variable altitud.

Palabras clave: Temperatura, Interpolación espacial, Kriging, Regresión lineal múltiple, algebra de mapas, Perú.

Abstract

This paper presents the results of a comparison of two interpolation methods for air temperature. Two models were evaluated: the Kriging model that includes the topographic variable and the multiple linear regression model without the topographic variable. Both methods were evaluated for monthly maximum and minimum temperature using data from the database of SENAMHI with record length of no less than 20 years in the period 1971-2000. To calculate the accuracy of the model coefficient of determination (R^2) was used as the indicator. For cross validation the bias (BIAS), the root square mean error (RSME) and the efficiency of the model (ME) were used. The results demonstrate the advantage of the multiple linear regression method over the Kriging method, due to its spatial representation, with R^2 values ranging from 0.9 to 0.98 for minimum temperature and from 0.69 to 0.92 for maximum temperature. For the BIAS values range from -0.1 to 0.3 for maximum and from -0.3 to 0.8 for minimum temperature. In the RMSE better results are obtained for minimum temperature, ranging from 1.3 to 2.9, while for maximum temperature values range from 2.0 to 2.4. However, between May and August the values are better for the maximum temperature. The results of ME show very similar values for both extremes, ranging from 0.8 to 0.9 for maximum and from 0.9 to 1.0 for minimum temperature. While the results of Kriging show values of R^2 ranging from 0.83 to 0.9 for maximum and from 0.9 to 0.96 for minimum temperature, in the RSME they range from 3.4 to 4.4 for maximum and 2.5 to 3.2 for minimum temperature. In the BIAS results were better for the minimum temperature, with values from 0.3 to 0.9, while for the maximum temperature values lie between 0.6 and 1.4. The ME showed best results for minimum temperature with values of 0.8 to 0.9, in comparison to 0.5 to 0.7 for the maximum temperature. In conclusion, the results indicate that the multiple linear regression method proves to be quite suitable for representation of the air temperature in mountain areas. It shows a great advantage over other interpolation methods that don't consider the altitude variable.

Keywords: Temperature, spatial interpolation, Kriging, multiple linear regression, algebra maps, Peru.

INTRODUCCIÓN

Los métodos de interpolación de las variables climáticas han sido ampliamente estudiados, sin embargo en los últimos años se han ido desarrollando nuevas metodologías, encontrándose metodologías específicas para cada zona dependiendo de la variable de interés. Con el avance de las tecnologías, como las herramientas computacionales y en el conocimiento de técnicas geoestadísticas, se ha logrado mejorar esta representación espacial, consiguiendo en ocasiones representar el comportamiento físico de la variable, como es el caso de la temperatura, el cual en la tropósfera libre disminuye a una razón aproximada de 6.5 °C/km (Cuadrat, 1997); este gradiente negativo también logra observarse conforme aumenta la altura topográfica, sin embargo éste debe ser calculado con los datos de campo debido a que es influenciado por otros factores como son la vegetación, presencia de obstáculos, entre otros.

Trabajos como el realizado por Pons (1996) y Ninyerola (2000), proponen una metodología que relaciona la información geográfica (variables independientes) con la información climática (variable dependiente), dando una buena alternativa para generar información en zonas con baja densidad de estaciones. Otra gran ventaja del método propuesto es que la interpolación es corregida mediante un mapa de errores, obtenido de la regresión lineal, y que luego es utilizado para ajustar el mapa resultante de la fórmula encontrada, con el fin de reproducir en el mapa aquel porcentaje del comportamiento de la variable que no es explicada mediante la fórmula.

El presente trabajo tiene por objetivo comparar el método de interpolación Kriging, normalmente utilizado para representar la variable temperatura, con el método de regresión lineal múltiple, y demostrar la ventaja comparativa de este último método para zonas de montaña. Otra de las ventajas del método propuesto que se intenta demostrar, es su aplicabilidad a áreas extensas, es así que la interpolación se ha realizado para todo el territorio peruano.

MATERIALES Y MÉTODOS

Data climática y topográfica

El Perú se encuentra ubicado entre las latitudes 0°01'48" S y 18°21'03" S, y entre las longitudes 68°39'27" O y 81°20'11" O. Tiene una extensión aproximada de 1 285 215 km². Su topografía es bastante accidentada debido a la presencia de la cadena montañosa "Los Andes", que atraviesa todo el país generando diversos microclimas. Su altitud varía desde los 0 hasta los 6768 msnm.

En el siguiente trabajo se utilizaron los datos de las estaciones meteorológicas pertenecientes a la red de estaciones del Servicio Nacional de Meteorología (SENAMHI-Perú), con un periodo de registro común desde 1971 al 2000. Los datos de temperatura mínima y máxima debían de tener un registro no menor de 20 años (ver Fig. 1 y Tabla 1).

A la data se le realizaron pruebas de control de calidad, como el eliminar los datos outliers, identificados por el método de cuartiles; el comprobar que no tuvieran saltos a través de la observación visual de la serie y utilizando las pruebas estadísticas Fisher y T-student. Asimismo fue de utilidad comparar los comportamientos temporales de las variables con las series de estaciones cercanas, obteniéndose coeficientes de correlación (R) y determinación (R²). De esta manera se obtuvieron las series homogéneas de 108 estaciones a nivel nacional.

Además, se requirió para la interpolación con el método de Regresión Lineal Múltiple el Modelo de elevación Digital (DEM), el cual se obtuvo del Global Land Cover Facility (glcf) de la NASA, el cual se encuentra a una resolución espacial de 90 metros; sin embargo por razones operacionales tuvo que ser remuestreado a un 1 km de resolución. Luego del DEM se extrajo la altura de estación que se utilizó en todo el proceso de interpolación, esto se realiza debido a que al remuestrear se origina un nuevo valor por pixel, resultado del promedio de las alturas encontradas dentro del km; esto es un factor que hay que tomar en cuenta en los resultados finales del modelo de regresión.

La información de referencia espacial de los archivos utilizados durante el proceso de elaboración de los mapas fueron: WGS 84, UTM zona 18 Sur.

COMPARACIÓN DE DOS MÉTODOS DE INTERPOLACIÓN



Fig. 1. Mapa de ubicación de estaciones.

Tabla 1. Ubicación de las estaciones utilizadas en la interpolación (periodo: 1971-2000).

Estación meteorológica	Periodo	Años	Código	Latitud	Longitud	Alt. DEM
A. Weberbauer	1971-2000	30	304	07°09'08"	78°29'29"	2661
Acomayo	1971-2000	26	687	13°55'01"	71°41'01"	3336
Alcantarilla	1971-2000	30	501	11°03'01"	77°33'01"	130
Andahuaylas	1971-2000	22	669	13°39'25"	73°22'15"	2928
Aplao	1971-2000	30	833	16°04'10"	72°29'26"	769
Arapa	1971-2000	30	783	15°08'08"	70°07'07"	3882
Ayabaca	1971-2000	30	237	04°38'38"	79°43'43"	2794
Ayaviri	1971-2000	27	776	14°52'52"	70°35'34"	3933
Azangaro	1971-2000	30	781	14°54'53"	70°11'11"	3853
Bagua Chica	1971-2000	26	253	05°39'38"	77°32'32"	394
Bambamarca	1971-2000	30	362	06°40'39"	78°31'31"	2516
Buena Vista	1971-2000	30	435	09°26'01"	78°12'01"	207
Cabalcocha	1971-2000	23	291	03°54'42"	70°30'44"	75
Cabanillas	1971-2000	30	780	15°38'38"	70°20'19"	3885
Cajabamba	1971-2000	26	373	07°37'37"	78°03'02"	2647
Calana	1971-2000	30	807	17°56'56"	70°11'10"	889
Camaná	1971-2000	29	832	16°36'24"	72°41'49"	23

Cañete	1971-2000	26	616	13°04'30''	76°19'50''	124
Caraveli	1971-2000	29	746	15°46'17''	73°21'42''	1767
Cayalti	1971-2000	29	320	06°53'52''	79°33'32''	76
Catcca	1971-2000	28	690	13°36'36''	71°33'37''	3709
Chivay	1971-2000	30	758	15°38'17''	71°35'49''	3716
Chulucanas	1971-2000	30	255	05°06'06''	80°10'10''	90
Chuquibambilla	1971-2000	30	764	14°47'47''	70°43'42''	3917
Chusis	1971-2000	27	231	05°31'31''	80°50'50''	7
Cochabamba	1971-2000	27	353	06°27'26''	78°53'53''	1825
El Espinal	1971-2000	23	340	06°49'32''	79°12'59''	438
El Frayle	1971-2000	29	849	16°05'05''	71°11'14''	4241
El Limón	1975-2000	26	241	05°55'55''	79°19'19''	1223
El Porvenir	1978-2000	20	310	06°35'35''	76°19'19''	259
El Salto	1971-2000	29	135	03°26'26''	80°19'19''	5
Ferreñafe	1971-2000	30	331	06°43'42''	79°46'45''	41
Genaro Herrera	1971-2000	30	281	04°54'01''	73°38'00''	121
Granja Kcayra	1976-2000	25	607	13°33'25''	71°52'31''	3201
Hacienda Bernales	1972-2000	28	650	13°45'45''	75°57'57''	312
Huamachuco	1971-2000	28	374	07°49'49''	78°03'03''	3225
Huamani	1971-2000	30	640	13°50'50''	75°35'35''	948
Huambos	1971-2000	24	343	06°27'27''	78°57'56''	2280
Huancabamba	1971-2000	30	239	05°15'01''	79°33'01''	3159
Huancané	1971-2000	30	786	15°12'12''	69°45'45''	3841
Huánuco	1971-2000	24	404	09°57'56''	76°14'13''	1961
Huaraya Moho	1971-2000	30	787	15°23'23''	69°29'29''	3824
Huarmaca	1973-2000	27	248	05°34'34''	79°31'31''	2281
Huayan	1971-2000	30	539	11°27'01''	77°07'01''	352
Huayao	1971-2000	29	635	12°02'01''	75°20'20''	3362
Ilave	1971-2000	30	879	16°05'05''	69°38'37''	3846
Imata	1971-2000	30	765	15°50'12''	71°05'16''	4444
Jaén	1971-2000	26	252	05°40'39''	78°46'46''	609
Jauja	1971-2000	25	503	11°46'45''	75°28'27''	3359
Jayanca	1971-2000	30	333	06°19'18''	79°46'46''	78
Juli	1971-2000	30	880	16°12'12''	69°27'26''	3830
La Angostura	1971-2000	30	754	15°10'47''	71°38'58''	4183
La Esperanza	1972-2000	29	230	04°55'55''	81°04'04''	5
La Haciendita	1971-2000	30	838	16°59'57''	71°35'19''	281
La Joya	1971-2000	30	804	16°35'33''	71°55'09''	1283
Lagunillas	1971-2000	30	763	15°46'46''	70°39'38''	4199
Lambayeque	1971-2000	30	301	06°42'13''	79°55'17''	14
Lampa	1971-2000	30	779	15°21'21''	70°22'22''	3875
Laredo	1971-2000	30	410	08°05'01''	78°51'01''	236
Lircay	1971-2000	27	657	12°58'57''	74°43'43''	3562
Llally	1971-2000	30	761	14°56'56''	70°53'53''	4189
Locumba	1971-2000	27	853	17°36'35''	70°45'44''	621
Lomas de Lachay	1971-2000	25	534	11°22'01''	77°22'01''	410
Mallares	1971-2000	30	208	04°51'51''	80°44'44''	47
Mazo Cruz	1971-2000	29	878	16°44'44''	69°42'42''	3976
Miraflores	1971-2000	30	207	05°10'10''	80°37'37''	37
Moquegua	1971-2000	26	806	17°10'09''	70°55'54''	1453
Morropón	1971-2000	29	235	05°11'01''	79°59'01''	121
Motupe	1971-2000	27	334	06°09'01''	79°44'01''	129
Muñani	1977-2000	21	785	14°46'46''	69°57'57''	3924
Navarro	1971-2000	30	386	06°21'21''	75°46'46''	158
Ocucaje	1971-2000	25	730	14°23'01''	75°40'01''	308
Olmos	1971-2000	27	236	05°45'44''	79°51'50''	121
Pacarán	1971-2000	30	638	12°51'43''	76°03'18''	854
Pampa Blanca	1971-2000	30	837	17°04'07''	71°43'22''	115
Pampa de Majes	1971-2000	30	805	16°19'40''	72°12'39''	1453

COMPARACIÓN DE DOS MÉTODOS DE INTERPOLACIÓN

Pampa De Villacuri	1971-2000	30	637	13°57'01''	75°48'01''	432
Pampahuta	1971-2000	30	762	15°29'29''	70°40'39''	4315
Papayal	1971-2000	29	134	03°34'34''	80°14'14''	48
Paramonga	1971-2000	29	528	10°40'01''	77°47'01''	123
Picoy	1971-2000	28	542	10°55'01''	76°44'01''	3130
Pilchaca	1971-2000	27	648	12°24'23''	75°05'05''	3690
Progreso	1971-2000	28	778	14°41'41''	70°21'20''	3923
Puerto Pizarro	1971-2000	30	132	3°30'30''	80°27'26''	14
Puno	1971-2000	30	708	15°49'49''	70°01'01''	3872
Punta Atico	1971-2000	29	830	16°13'43''	73°41'40''	40
Punta Coles	1971-2000	30	846	17°41'55''	71°22'24''	35
Quillabamba	1971-2000	21	606	12°51'22''	72°41'31''	1041
Recuay	1978-1999	21	441	09°43'46''	77°27'14''	3408
Reque	1971-2000	30	332	06°53'53''	79°50'50''	14
Sama Grande	1971-2000	30	875	17°47'2''	70°29'17''	562
San Camilo	1971-2000	30	700	14°04'04''	75°43'43''	407
San Juan	1971-2000	30	369	07°17'17''	78°29'28''	2605
San Marcos	1971-2000	29	370	07°19'19''	78°10'10''	2393
San Miguel	1973-2000	28	247	06°14'14''	80°41'41''	24
Santa Rosa	1971-2000	30	536	11°13'01''	77°23'01''	384
Sausal de Culucan	1971-2000	30	238	04°45'45''	79°46'46''	1155
Sibayo	1971-2000	25	755	15°29'08''	71°27'11''	3847
Tabaconas	1977-2000	23	240	05°19'19''	79°17'17''	1968
Tahuaco-Yunguyo	1971-2000	26	882	16°18'18''	69°04'03''	3871
Talla	1971-2000	30	325	07°16'16''	79°25'24''	113
Tambopata	1971-2000	30	790	14°13'12''	69°09'08''	1371
Tamshiyacu	1971-2000	29	172	04°00'12''	73°09'39''	88
Tarapoto	1971-1996	26	401	06°28'01''	76°22'01''	282
Tingo Maria	1971-2000	30	468	09°17'17''	75°59'59''	773
Tulumayo	1971-2000	20	469	09°06'01''	75°54'01''	1212
Ubinas	1971-2000	30	851	16°22'57''	70°51'24''	3641
Von Humboldt	1971-2000	30	610	12°05'01''	76°56'01''	253

MÉTODOS

Interpolación por Kriging

El método ordinario Kriging obtiene los pesos (o influencia de los valores), (Schloeder *et al.*, 2001).

$$\sum_{i=1}^n \lambda_i \gamma[d(S_i, S_j)] + m = \gamma[d(S_0, S_i)], i = 1, \dots, n; \sum_{i=1}^n \lambda_i = 1 \quad (1)$$

Donde n es el número de observaciones, m es el multiplicador de Lagrange usado para la minimización de las restricciones, λ es el peso dado a cada una de las observaciones y la suma de todos los λ es igual a uno. Los subíndices i y j, denotan los puntos muestreados, el subíndice 0, es el punto en estimación, s simboliza la medición efectuada (variable medida) y d(S_i, S₀) es la distancia entre S_i y S₀ a partir del semivariograma:

$$\gamma[d(S_i, S_0)] = var[Z(S_i) - Z(S_0)] \quad (2)$$

Esta semivarianza calculada es una medida para determinar la similitud entre observaciones, en donde a mayor similitud, menor semivarianza (Lozano *et al.*, 2004). Los pesos λ o las relevancias de los valores, son determinados con el fin de asegurar que el error promedio para el modelo sea cero y además la varianza del error es minimizada (Schloeder *et al.*, 2001), lo cual ofrece una predicción no sesgada. Pese a ello este método requiere de supuestos estadísticos muy fuertes, como la hipótesis intrínseca de estacionalidad sea aceptada, lo cual raramente se observa en la naturaleza (Villaroto *et al.* 2008).

Interpolación por regresión lineal múltiple

El método utilizado fue propuesto por Pons (1996) y Ninyerola *et al.* (2000), basado en técnicas de regresión lineal múltiple para la interpolación espacial de los datos provenientes

de las estaciones meteorológicas. Lo resaltante de este método es que al final los mapas resultantes son corregidos utilizando los residuales, producto de la regresión lineal múltiple.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + e_i, \text{ para } i = 1, 2, \dots, n \quad (3)$$

Y en forma matricial:

$$Y = X\beta + e$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \quad (4)$$

La fórmula de regresión lineal múltiple debe de cumplir ciertos supuestos para que pueda ser aplicable, los cuales se mencionan a continuación:

- 1) La variable a predecir tiene que tener un comportamiento normal
- 2) Homocedasticidad, es decir varianzas homogéneas
- 3) Autocorrelación, errores no correlacionados
- 4) Multicolinealidad, regresores independientes

El primer supuesto se cumple debido a que se trata de la variable temperatura, que de por sí tiene un comportamiento normal; sin embargo para comprobar esta condición se puede realizar el gráfico P-P plot, en donde los valores deben acercarse al comportamiento de una curva normal, línea patrón que atraviesa el gráfico; o analizarse por medio de un histograma.

La homocedasticidad se estudió utilizando las variables que proporciona el SPSS en la opción

de gráficos: Pronósticos tipificados (ZPRED) y residuos tipificados (ZRESID). Implica que la variación de los residuos sea uniforme en todo el rango de valores de los pronósticos.

Para comprobar si existe o no autocorrelación se hace uso de la prueba Durbin y Watson, la cual indica que para rechazar o no la hipótesis nula de que no hay autocorrelación de primer orden en las perturbaciones, se consideran los límites inferior d_L y superior d_U , dados en tablas, valores que dependen del número de muestras. Si $d < d_L$ se rechaza la hipótesis, y si es $d > d_U$ no se rechaza H_0 ; dicho en otras palabras, si no se rechaza la hipótesis H_0 significa que no existe autocorrelación.

La multicolinealidad también puede ser calculado en el SPSS. La colinealidad puede analizarse por medio del Nivel de Tolerancia (valores menores a 0,10 indican multicolinealidad); o por medio del VIF o FAV, algunos autores dicen que existen problemas graves de multicolinealidad cuando éste es mayor de 10; sin embargo valores mayores a 4 merecen atención.

Algebra de mapas

El proceso de regresión lineal múltiple pasa por el proceso de obtener matrices ráster de las distintas variables independientes: Modelo de elevación digital, latitud y longitud (ver figura 2). Estos ráster (variables independientes), se relacionan con la variable climática (variable dependiente).

El método consiste en aplicar un conjunto de operadores sobre una o varias capas ráster de entrada para producir uno o varios ráster de salida. Estos operadores, que pueden ser aritméticos, trigonométricos, lógicos o condicionales, se aplican mediante ecuaciones a cada celda del ráster de entrada. Para realizar esta operación se hizo uso de la herramienta Map Calculator, del programa ARCVIEW 3.4; el cual además permite realizar cálculos matemáticos entre mapas temáticos o varios mapas resultantes.

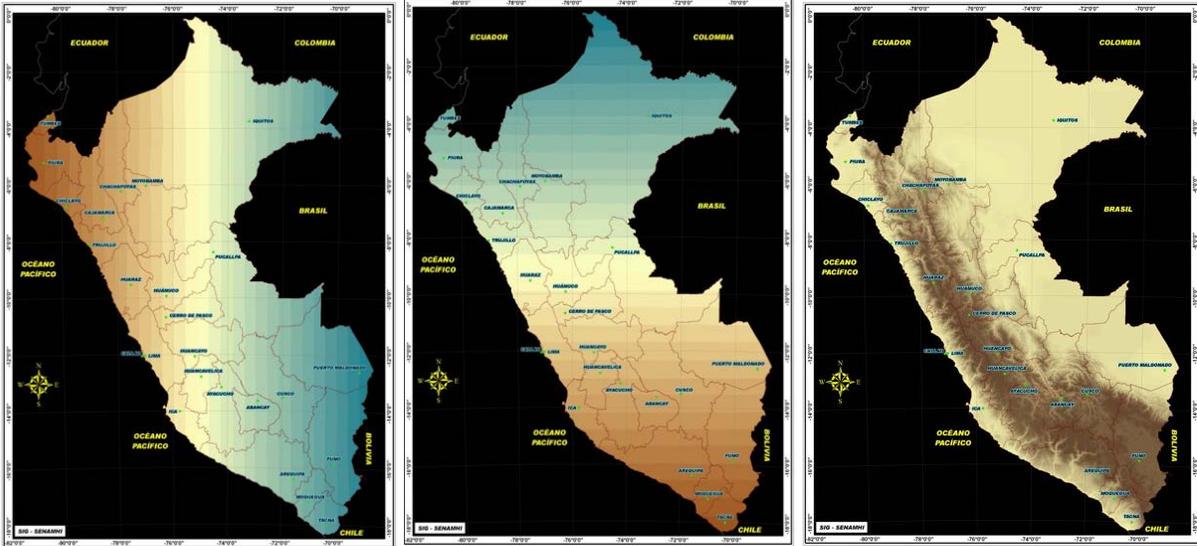


Fig. 2. Izquierda: ráster de Longitud, Centro: ráster de Latitud, Derecha: ráster de Altitud (DEM)

Calibración y Validación

El proceso de calibración consiste en sumar el ráster de residuales, con el ráster del modelo de regresión lineal múltiple obtenido de la data de temperatura del mes que corresponda o del periodo de tiempo que se requiera representar; para así obtener el mapa final.

Los residuales son obtenidos de la diferencia entre los valores medidos en la estación meteorológica y los obtenidos mediante el proceso de regresión lineal múltiple, este último puede ser obtenido mediante el uso de un procesador de datos como el programa EXCEL o de algún programa estadístico como el SPSS. Fue necesario un método de interpolación simple para representar los residuales de la variable temperatura, acción necesaria debido a que los residuales representan aquel porcentaje del comportamiento de la variable que no puede ser explicado mediante el comportamiento físico normal de la variable, es decir para este caso la disminución de la variable con la altura, comportamiento que tiene que ver más con el comportamiento local de la variable, el cual es originado por diversas razones como son la presencia de obstáculos, vegetación predominante, etc.

Comparaciones entre la interpolación por Kriging ordinario (OK), con el SPLINE y el IDW, muestran que el OK y el IDW son superiores frente al SPLINE (QiaoJing *et al.*,

2005). Los resultados mostrados por QiaoJing *et al.*, 2005, para la cuenca del Haihe en China, muestran cierta superioridad del OK frente al IDW, donde el Error Absoluto Medio (MAE) del OK es menor tanto para temperatura máxima como para temperatura mínima, mientras que lo contrario ocurre con los resultados de la Desviación Estándar (SDEV), donde muestra ser menor para el IDW.

En nuestro caso se usó el IDW para la interpolación de los residuales, debido a que da un peso mayor a los pixeles cercanos al pixel conocido, y por tanto tienen a tener un comportamiento similar a éste; mientras que los que se encuentran lejanos tienen un peso menor, y otra es la disponibilidad de este método en el programa ARCVIEW.

Para la validación de los mapas se utilizó el método de la validación cruzada generalizada (GCV, Generalized Cross-Validation), el cual se define como el error cuadrático promedio entre los valores estimados y los valores observados, cuando el valor observado en cuestión es dejado fuera del proceso de estimación. Esta es una medida de qué tan bien la superficie es capaz de estimar un valor observado cuando éste no es utilizado en el proceso de estimación. Para la comprobación de la exactitud de los resultados obtenidos por del método de Regresión Lineal Múltiple, se procesan los mapas con una aproximación de

un 60% de la data, y el restante 40% para su validación.

Como indicador de la bondad del ajuste se usó el coeficiente de determinación (R^2), y como medida de error el Root Mean Square Error (RMSE), además del Sesgo (Bias); mientras menores sean estos dos últimos índices, mejor será el funcionamiento del modelo. Además se aplicó una fórmula que indicaría la eficiencia del modelo definido como ME (ver fórmulas 5 a 8).

Las estaciones utilizadas en total para el proceso de validación fueron 35, los cuales debían tener como requisito un mínimo de 10 años de registro, los cuales debían además estar distribuidos a nivel nacional, para procurar representar el comportamiento de la variable en cada departamento o región (ver tabla 2).

$$R^2 = \frac{\sum_{j=1}^m (\hat{Y}_j - F)^2}{\sum_{j=1}^m (Y_j - F)^2}, 0 \leq R^2 \leq 1 \quad (5)$$

Donde el numerador corresponde a la suma de cuadrados debido a la regresión (SCE), y el denominador a la suma de cuadrados total (SCT)

$$RMSE = \sqrt{\frac{1}{n} \sum (Q_{est} - Q_{obs})^2} \quad (6)$$

$$Bias = \frac{1}{n} \sum (Q_{est} - Q_{obs}) \quad (7)$$

$$ME = 1 - \frac{\sum (Q_{obs} - Q_{est})^2}{\sum (Q_{obs} - \bar{Q}_{obs})^2} \quad (8)$$

RESULTADOS Y DISCUSIÓN

Previamente a la regresión lineal múltiple se realizó la prueba de normalidad en el SPSS. Los resultados muestran que la variable temperatura tiene un comportamiento normal, tal como se observa en las Figuras 3 y 4.

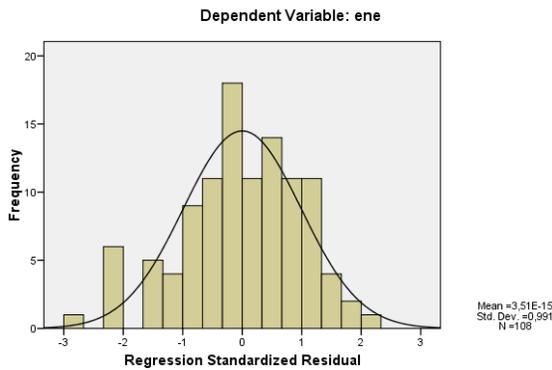


Fig. 3. Histograma de la temperatura máxima del mes de enero.

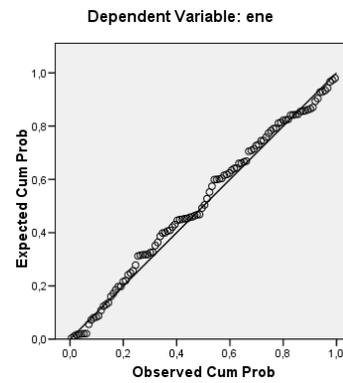


Fig. 4. Gráfico P-P plot de la temperatura máxima del mes de enero.

Tabla 2. Ubicación de estaciones utilizadas en la validación.

Estación meteorológica	Periodo	Años	Código	Latitud	Longitud	Alt. DEM
Aucayacu	1986-2000	12	474	08°36'01''	75°56'01''	1181
Bellavista	1971-2000	15	382	07°03'03''	76°33'33''	265
Canchan	1988-2000	11	457	09°56'56''	76°17'17''	2260
Chancay Baños	1987-2000	13	395	06°34'33''	78°52'52''	2034
Chirinos	1988-2000	10	260	05°18'17''	78°54'53''	1259
Contumaza	1988-2000	13	354	7°21'21''	78°49'49''	2801
Curahuasi	1984-2000	15	677	13°33'09''	72°44'06''	2924
Cutervo	1987-2000	13	352	06°22'21''	78°48'47''	2982
Desaguadero	1990-2000	11	883	16°34'34''	69°02'02''	3959
Donoso	1984-2000	15	546	11°28'01''	77°14'01''	179
El Salto	1971-2000	21	135	03°26'26''	80°19'19''	5
Hacienda Bernales	1972-2000	24	650	13°45'45''	75°57'57''	312
Pacarán	1971-2000	18	638	12°51'43''	76°03'18''	854
Lomas de Lachay	1971-2000	18	534	11°22'01''	77°22'01''	410
Pampa De Villacuri	1971-2000	19	637	13°57'01''	75°48'01''	432
Punta Atico	1971-2000	20	830	16°13'43''	73°41'40''	40
Sama Grande	1971-2000	30	875	17°47'2''	70°29'17''	562
Jauja	1991-2000	10	503	11°46'45''	75°28'27''	3359
Ayaviri	1971-2000	19	776	14°52'52''	70°35'34''	3933
Juli	1971-2000	30	880	16°12'12''	69°27'26''	3830
Lagunillas	1983-2000	18	763	15°46'46''	70°39'38''	4199
Pampahuta	1971-2000	15	762	15°29'29''	70°40'39''	4315
Ferreñafe	1971-2000	30	331	06°43'42''	79°46'45''	41
Huánuco	1986-2000	14	404	09°57'56''	76°14'13''	1961
Fonagro Chinchá	1986-2000	13	791	13°28'28''	76°08'08''	15
La Unión	1974-2000	15	384	07°10'10''	76°29'29''	304
Lamas	1990-2000	11	383	06°25'25''	76°31'31''	639
Marcapomacocha	1988-2000	11	549	11°24'17''	76°19'31''	4551
Oyón	1987-2000	12	541	10°40'01''	76°46'01''	3660
Paucartambo	1990-2000	10	689	13°19'28''	71°35'26''	3434
Rioja	1987-2000	14	377	06°02'01''	77°10'10''	815
San Ignacio	1985-2000	15	242	05°08'07''	78°59'58''	1074
Santa Cruz	1987-2000	11	351	06°37'36''	78°56'55''	2215
Tinajones	1986-2000	14	335	06°38'37''	79°25'25''	189
Urubamba	1985-2000	15	683	13°18'38''	72°07'26''	3177

Tabla 3. Resultados de la prueba de autocorrelación por Durbin Watson.

	Ene	Feb	Mar	Abr	May	Jun	Jul	Agos	Set	Oct	Nov	Dic
TMAX	1.841	1.839	1.792	1.816	1.779	1.752	1.729	1.729	1.731	1.751	1.798	1.822
TMIN	1.771	1.861	1.808	1.639	1.552	1.529	1.539	1.532	1.513	1.526	1.526	1.614

Luego fue necesario realizar la prueba de homocedasticidad, indicando que las varianzas de los residuos eran homogéneas. La Figura 5 muestra el resultado para la temperatura máxima del mes de enero.

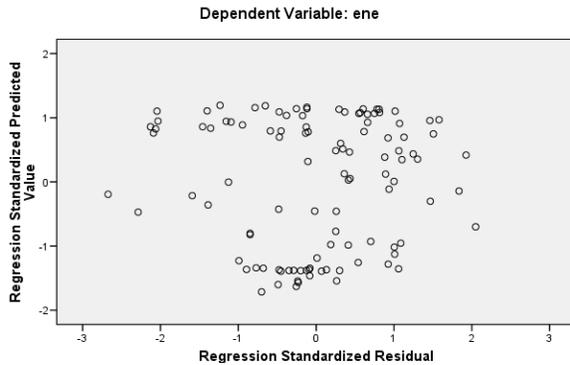


Fig. 5 Varianza de los residuos.

Al igual los resultados de Durbin y Watson, mostraron que no existía autocorrelación entre las variables predictoras, es decir entre la longitud, latitud y el DEM (ver Tabla 3). Mientras que los resultados de colinealidad mostraron que la variable longitud es la que menos representaba a la temperatura, con resultados de tolerancia de 0.23, seguido por el de latitud con 0.268 y el de altura con 0.668.

Adicionalmente se muestra los resultados de la influencia que ejerce cada variable predictor (latitud, longitud y altura), sobre la temperatura (ver Figura 6). Se observa que la latitud tiene mayor influencia en el comportamiento de la variable, con un R^2 de 0.9; mientras que las variables independientes latitud y longitud tienen una influencia mucho menor, que apenas llega a un R^2 de 0.3; sin embargo entre ambas (latitud y longitud) se observa que la dispersión de la latitud es mucho menor que el de la longitud, motivo por el cual la latitud tendría cierta influencia sobre el comportamiento de la variable, logrando aparecer en algunos meses. Esto último se sustenta en el análisis de significancia (ver Tabla 4).

Previamente a la regresión, se realizó la prueba de significancia en el SPSS. Al encontrar que los resultados eran menores a 0.05, se rechazaba la hipótesis nula ($H_0: \beta_i = 0$); por tanto la variable independiente tenía influencia en el comportamiento de la variable (ver tabla 4). La Tabla 5 muestra los coeficientes hallados para la estimación de la temperatura máxima y mínima por el método de regresión lineal múltiple.

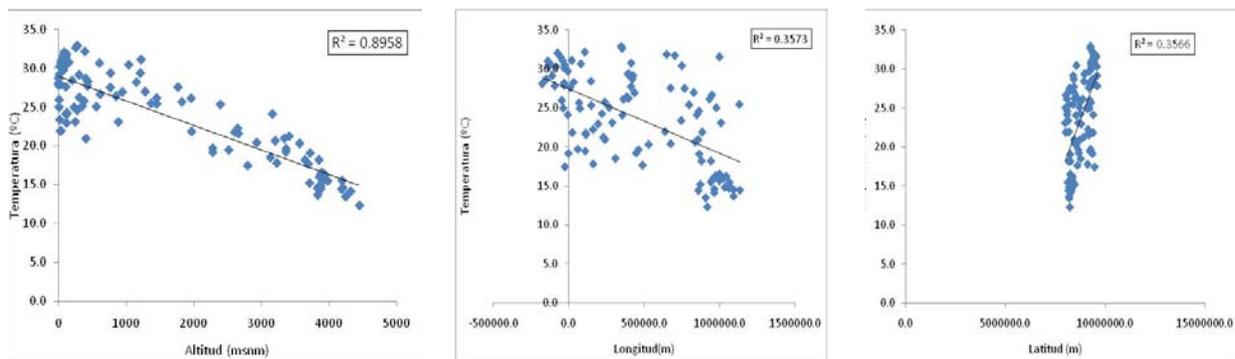


Fig. 6. Relación entre la temperatura máxima media anual vs altitud, longitud y latitud.

Tabla 4. Significancia de las variables geográficas y topográficas utilizadas en las fórmulas de regresión lineal múltiple

Mes	Temperatura Máxima			Temperatura Mínima		
	(Long)	(Latitud)	(altitud)	(Long)	(Latitud)	(altitud)
Enero	3.54E-01	1.10E-02	<1.00E-03	3.00E-03	<1.00E-03	<1.00E-03
Febrero	7.89E-01	1.20E-01	<1.00E-03	1.94E-01	<1.00E-03	<1.00E-03
Marzo	8.51E-01	7.00E-02	<1.00E-03	1.91E-01	<1.00E-03	<1.00E-03
Abril	3.72E-01	2.00E-03	<1.00E-03	5.10E-02	<1.00E-03	<1.00E-03
Mayo	5.70E-02	<1.00E-03	<1.00E-03	5.00E-02	<1.00E-03	<1.00E-03
Junio	8.00E-03	<1.00E-03	<1.00E-03	9.70E-02	<1.00E-03	<1.00E-03
Julio	2.00E-03	<1.00E-03	<1.00E-03	8.40E-02	<1.00E-03	<1.00E-03
Agosto	1.00E-03	<1.00E-03	<1.00E-03	1.70E-02	<1.00E-03	<1.00E-03
Setiembre	1.00E-03	<1.00E-03	<1.00E-03	2.00E-03	<1.00E-03	<1.00E-03
Octubre	1.00E-03	<1.00E-03	<1.00E-03	<1.00E-03	<1.00E-03	<1.00E-03
Noviembre	5.00E-03	<1.00E-03	<1.00E-03	<1.00E-03	<1.00E-03	<1.00E-03
Diciembre	4.80E-02	<1.00E-03	<1.00E-03	<1.00E-03	<1.00E-03	<1.00E-03

Tabla 5. Coeficientes de la regresión lineal múltiple para la temperatura máxima y mínima.

Mes	Temperatura Máxima				Temperatura Mínima			
	β_1 (Long)	β_2 (Latitud)	β_3 (altura)	Conste	β_1 (Long)	β_2 (Latitud)	β_3 (altura)	Conste
Enero	*	1.412E-06	-0.003	18.33	1.72E-06	3.14E-06	-0.0042	-7.77
Febrero	*	*	-0.004	31.59	*	2.29E-06	-0.004	0.98
Marzo	*	*	-0.004	31.45	*	2.66E-06	-0.004	-2.66
Abril	*	1.982E-06	-0.003	12.44	*	4.005E-06	-0.004	-15.74
Mayo	*	3.08E-06	-0.002	0.93	2.13E-06	6.63E-06	-0.005	-41.10
Junio	3.47E-06	6.16E-06	-0.0025	-29.18	*	5.71E-06	-0.005	-33.36
Julio	4.45E-06	7.03E-06	-0.0023	-38.03	*	5.53E-06	-0.004	-32.61
Agosto	5.24E-06	7.54E-06	-0.0022	-42.59	3.14E-06	7.25E-06	-0.0043	-48.96
Setiembre	5.19E-06	7.21E-06	-0.0022	-38.85	3.92E-06	7.41E-06	-0.0039	-50.33
Octubre	4.80E-06	6.07E-06	-0.0025	-27.72	4.50E-06	7.11E-06	-0.0038	-47.26
Noviembre	3.82E-06	4.92E-06	-0.0027	-16.30	4.63E-06	6.27E-06	-0.0039	-39.06
Diciembre	2.30E-06	3.28E-06	-0.0032	-0.07	3.38E-06	4.81E-06	-0.0040	-24.41

*: No significativa al 5%

Las Tabla 6 y 7, muestran los coeficientes de determinación (R^2), obtenidos de los modelos de regresión lineal múltiple y kriging, además de los resultados del error medio cuadrático (RMSE), el sesgo (BIAS) y la eficiencia del Modelo (ME) obtenidos del proceso de validación del modelo.

Los resultados de la tabla 6 muestran que el método de regresión lineal múltiple es bastante preciso tanto para temperatura máxima como para mínima, siendo incluso más preciso para el

caso de las temperaturas mínimas (TMIN), con R^2 que van de 0.9 en los meses de agosto a octubre, y de 0.98 en los meses de febrero y marzo. Para el caso de la temperaturas máxima varía de 0.69 en el mes de setiembre a 0.92 para el mes de febrero (ver Figuras 11 y 12).

El RMSE, muestra mejores resultados en los meses de verano, entre diciembre y marzo, para ambos casos, tanto para TMIN como para TMAX; observándose mejores resultados en la TMIN, con valores que van de 1.4 a 2.9;

mientras que en TMAX los resultados van de 2.0 a 2.4; observándose mejores resultados en febrero y marzo.

El BIAS, muestra que los resultados de TMAX son más precisos, con diferencias entre lo pronosticado y lo observado que van de -0.1 a 0.1 entre abril y octubre, y de 0.2 de noviembre

a marzo. Y en la TMIN el BIAS varía de -0.3 a -0.5 en los meses de noviembre a abril, y de -0.6 a -0.8 en los meses restantes.

En cuanto a la ME, se muestra muy buenos resultados y similares entre ambas temperaturas, con valores de 0.8 a 0.9 en la TMAX, y de 0.9 a 1.0 en la TMIN.

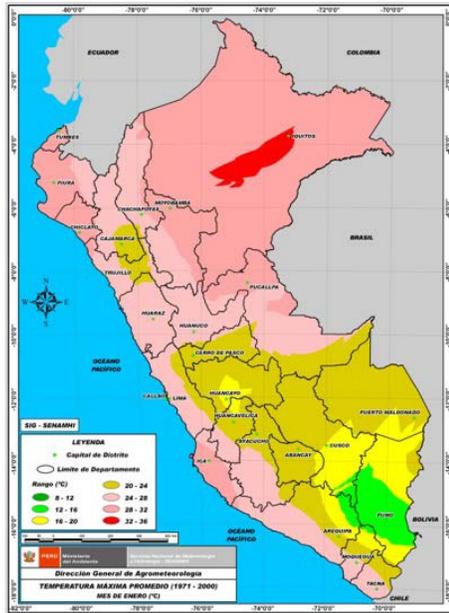


Fig. 7 Mapa climático de la temperatura máxima para el mes de enero por el método Kriging.

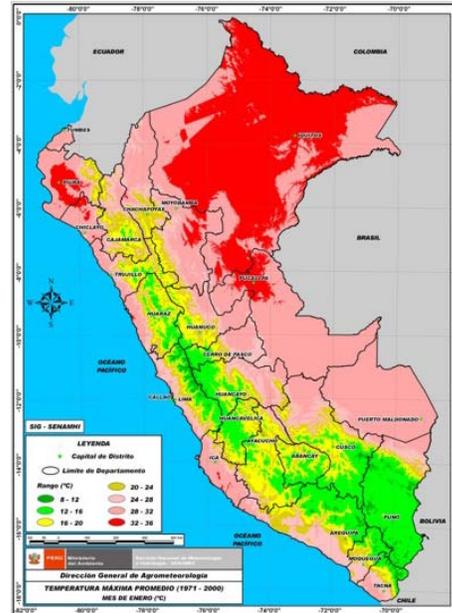


Fig. 8. Mapa climático de la temperatura máxima para el mes de enero por el método de Regresión Lineal Múltiple.

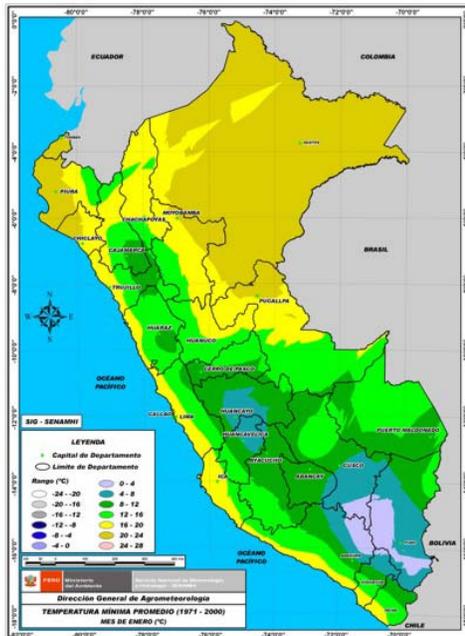


Fig. 9. Mapa climático de la temperatura mínima para el mes de enero por el método Kriging

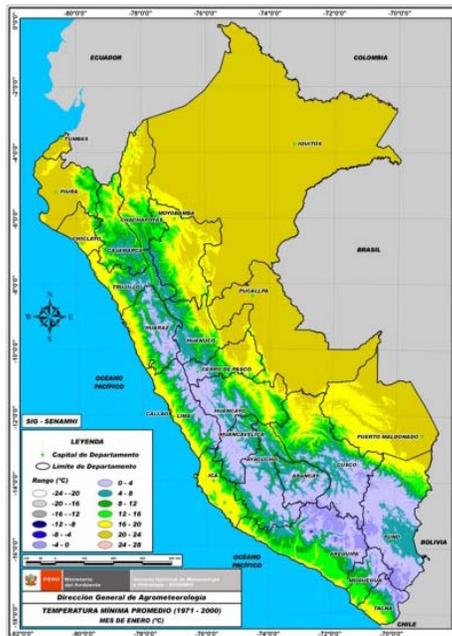


Fig. 10. Mapa climático de la temperatura mínima para el mes de enero por el método de Regresión Lineal Múltiple.

Tabla 6. Resultados por mes de la interpolación por regresión lineal múltiple de la temperatura máxima y mínima. R^2 : Coeficiente de determinación, RMSE: Error Medio Cuadrático, BIAS: Sesgo, ME: Eficiencia del Modelo.

Mes	Temperatura Máxima				Temperatura Mínima			
	R^2	RMSE	BIAS	ME	R^2	RMSE	BIAS	ME
Enero	0.90	2.1	0.2	0.9	0.97	1.4	-0.4	1.0
Febrero	0.92	2.0	0.2	0.9	0.98	1.3	-0.3	1.0
Marzo	0.91	2.0	0.2	0.9	0.98	1.3	-0.3	1.0
Abril	0.88	2.1	0.1	0.9	0.96	1.7	-0.5	0.9
Mayo	0.82	2.2	0.0	0.8	0.94	2.5	-0.8	0.9
Junio	0.78	2.3	-0.1	0.8	0.92	2.7	-0.8	0.9
Julio	0.74	2.4	-0.1	0.8	0.91	2.9	-0.8	0.9
Agosto	0.70	2.4	0.0	0.8	0.90	2.7	-0.8	0.9
Setiembre	0.69	2.4	-0.1	0.8	0.90	2.2	-0.6	0.9
Octubre	0.70	2.4	0.1	0.8	0.90	1.9	-0.7	0.9
Noviembre	0.75	2.3	0.2	0.8	0.92	1.8	-0.5	0.9
Diciembre	0.84	2.1	0.2	0.9	0.95	1.6	-0.5	0.9

La Tabla 7 muestra los resultados del coeficiente de determinación (R^2) de la interpolación por el método Kriging, para temperatura máxima y mínima; con valores que van en la TMAX de 0.82 a 0.84 entre los meses de junio a noviembre, y de 0.85 a 0.9 en los meses restantes; y para la TMIN, entre 0.90 a 0.93 de mayo a octubre, y de 0.94 a 0.96 de noviembre a abril (ver figuras 13 y 14).

Los resultados de RMSE, muestra mejores resultados para la TMIN que para la TMAX; oscilando los valores para la TMIN de 2.5 en octubre, a 3.2 en enero y febrero; y en TMAX se observó el mínimo valor de 3.4 en los meses

de julio y agosto, a un máximo de 4.4 en febrero.

De la misma forma los resultados del BIAS son menores para la TMIN, así se observan resultados de 0.3 de mayo a octubre, y de 0.5 a 0.9 de noviembre a marzo. En la TMAX los menores valores, de 0.6 y 0.7, se observaron entre julio y setiembre; y las mayores diferencias, de 0.8 a 1.4, en los meses restantes.

Los resultados de ME por kriging, muestra mejores resultados para la TMIN, con valores de 0.8 a 0.9; mientras que los valores de la TMAX estuvieron entre 0.5 a 0.7.

Tabla 7. Resultados por mes de la interpolación por Kriging de la temperatura máxima y mínima. RMSE: Error Medio Cuadrático, BIAS: Sesgo, ME: Eficiencia del modelo.

Mes	Temperatura Máxima				Temperatura Mínima			
	R^2	RMSE	BIAS	ME	R^2	RMSE	BIAS	ME
Enero	0.90	4.3	1.3	0.5	0.95	3.2	0.9	0.8
Febrero	0.90	4.4	1.4	0.5	0.96	3.2	0.9	0.8
Marzo	0.89	4.3	1.4	0.5	0.94	3.1	0.9	0.8
Abril	0.87	3.9	1.1	0.6	0.94	2.6	0.5	0.9
Mayo	0.85	3.6	0.8	0.6	0.91	2.7	0.3	0.9
Junio	0.84	3.7	0.8	0.6	0.90	2.8	0.3	0.9
Julio	0.83	3.4	0.6	0.7	0.89	2.9	0.3	0.9
Agosto	0.83	3.4	0.7	0.7	0.90	2.8	0.3	0.9
Setiembre	0.83	3.5	0.6	0.6	0.92	2.6	0.3	0.9
Octubre	0.82	3.7	0.8	0.6	0.93	2.5	0.3	0.9
Noviembre	0.84	3.8	0.9	0.6	0.94	2.6	0.5	0.9
Diciembre	0.86	4.0	1.1	0.5	0.95	2.8	0.5	0.8

Los resultados de los estadísticos utilizados tal como se muestra en las Tablas 6 y 7, muestra la superioridad del método de regresión lineal múltiple frente al Kringin para espacializar la variable temperatura. Esto se puede además apreciar claramente en las Figuras 7, 8, 9 y 10,

donde se muestra dos ejemplos de interpolación por ambos métodos, ambos para el mes de enero, para la temperatura máxima y temperatura mínima; donde se observa claramente la superioridad en la representación espacial de la variable.

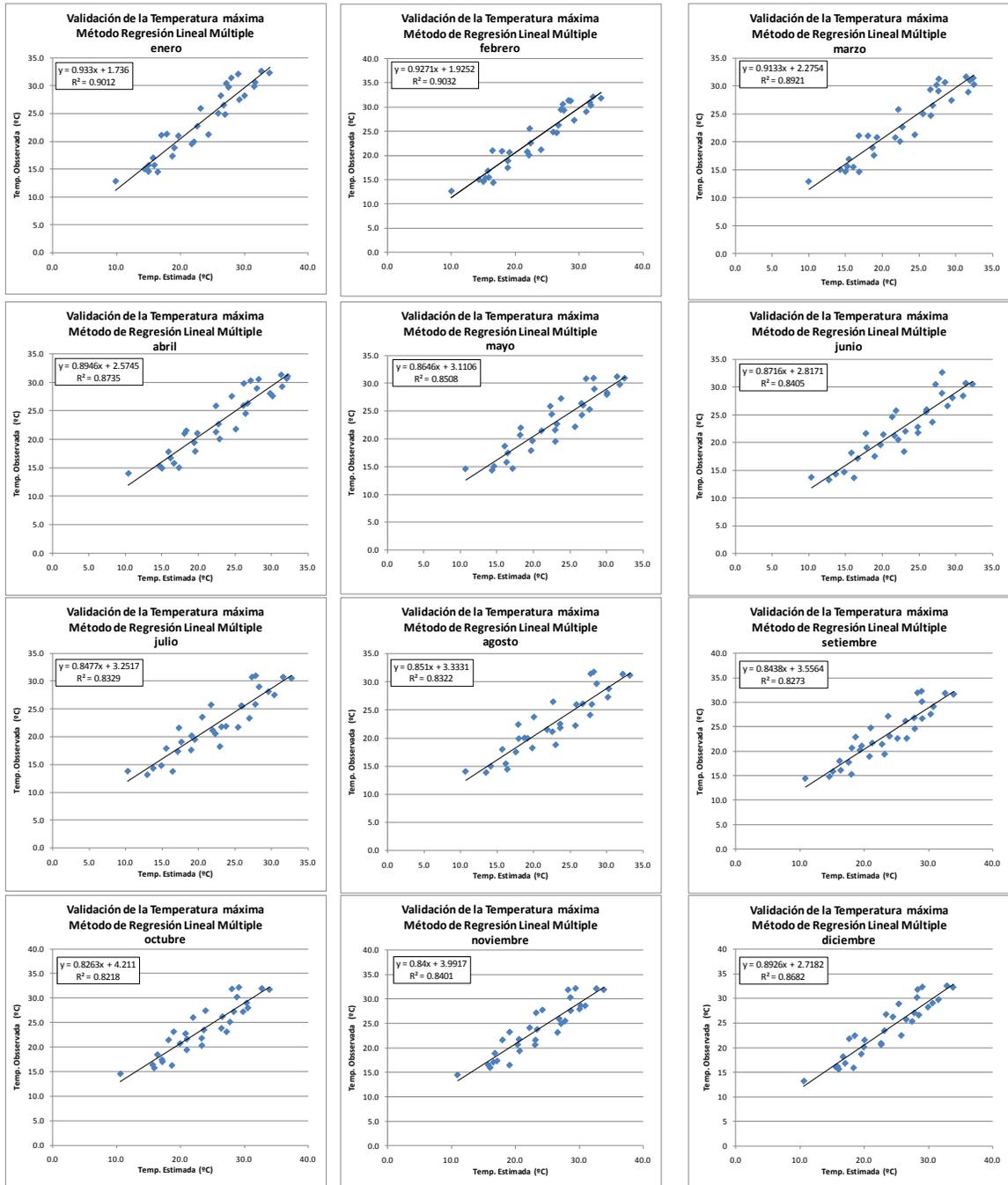


Fig. 11. Resultados de la validación de la temperatura máxima por el método de regresión lineal múltiple

COMPARACIÓN DE DOS MÉTODOS DE INTERPOLACIÓN

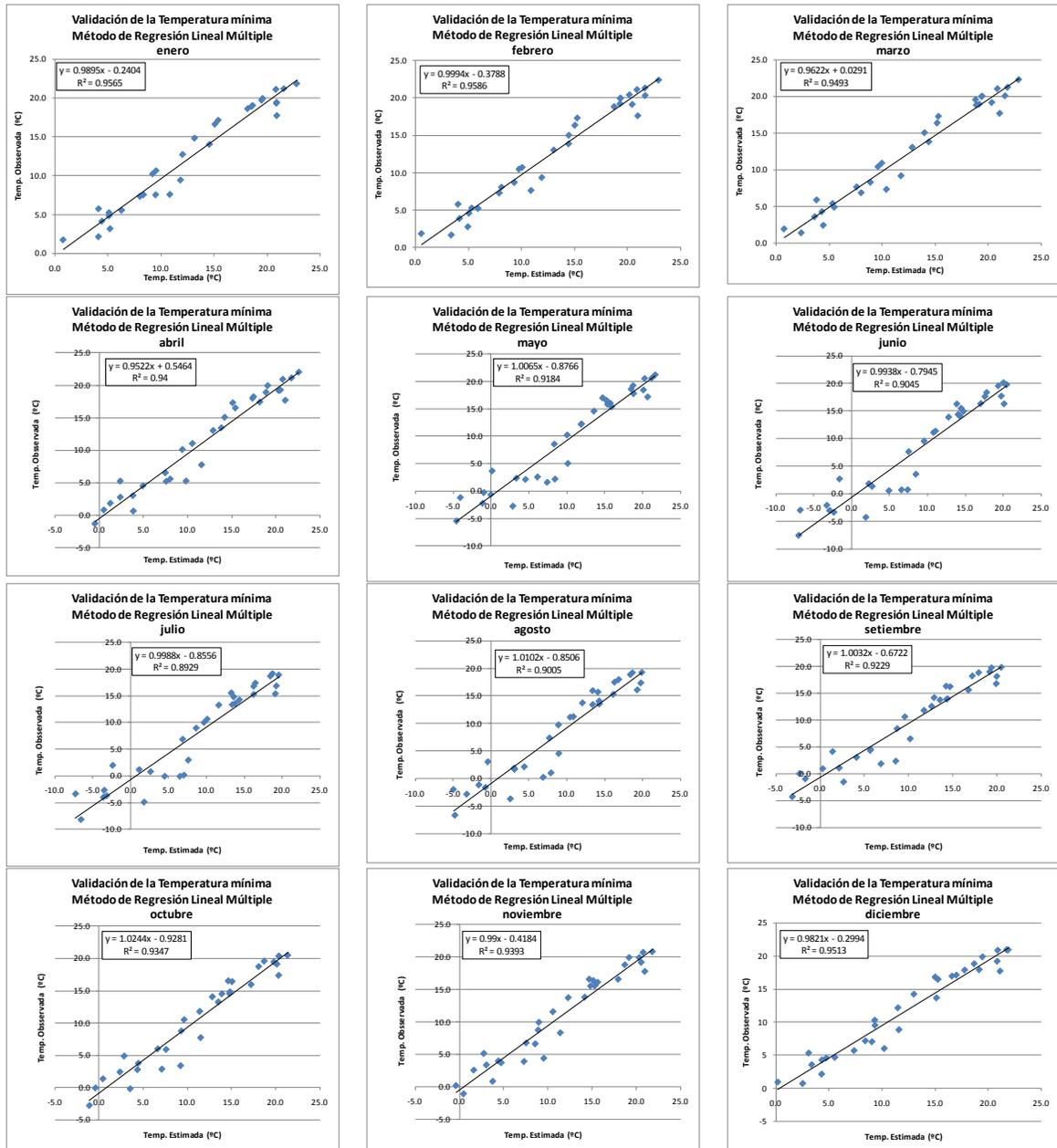


Fig. 12. Resultados de la validación de la temperatura mínima por el método de regresión lineal múltiple.

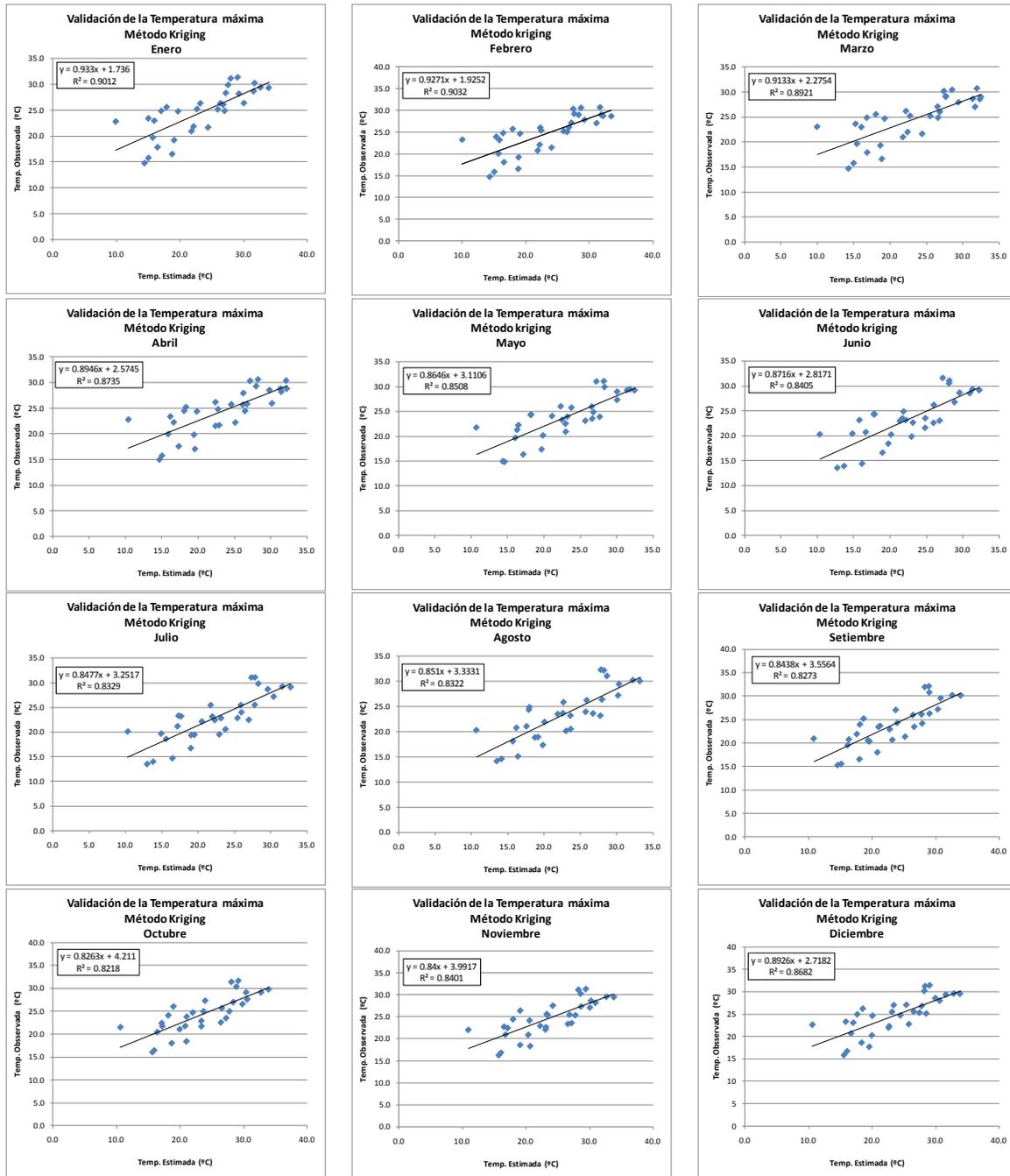


Fig. 13. Resultados de la validación de la temperatura máxima por el método Kriging.

COMPARACIÓN DE DOS MÉTODOS DE INTERPOLACIÓN

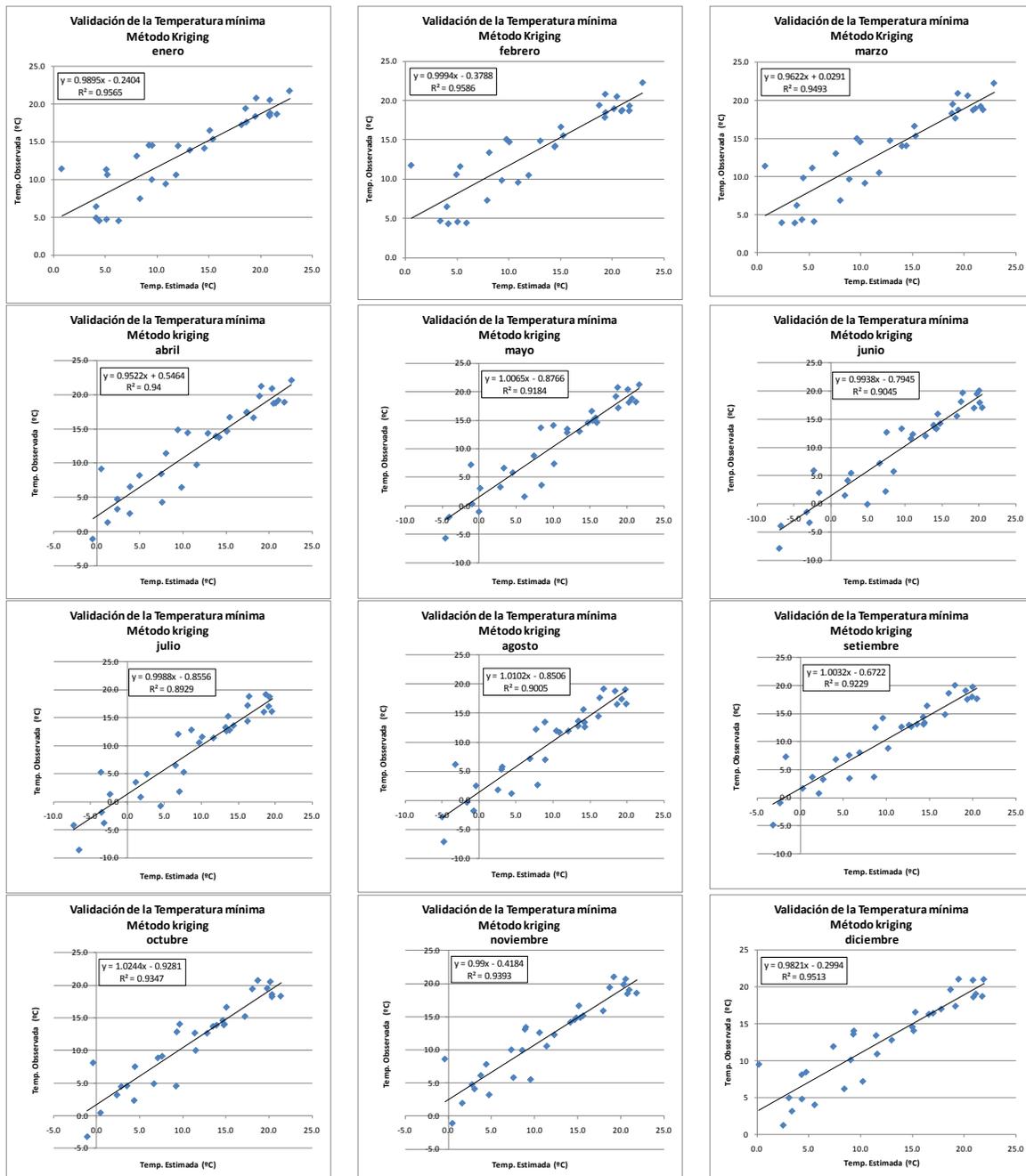


Fig. 14. Resultados de la validación de la temperatura mínima por el método Kriging.

Agradecimientos: Agradezco a Philippe VAUCHEL investigador del Instituto de Investigación para el Desarrollo de Francia (IRD), por su contribución en la revisión del presente artículo y a sus valiosos aportes; y al PhD. Guillermo A. Baigorria, investigador del departamento de Ingeniería Agrícola y Biológica de la Universidad de Florida, por apoyarme en la revisión bibliográfica, permitiéndome tener una

visión más amplia de las metodologías empleadas en la interpolación de variables climáticas.

REFERENCIAS

Cristóbal, J.; Ninyerola, M. y Pons, X. (2005). Aportación de variables obtenidas mediante Teledetección a la modelización de la temperatura

del aire. XI Congreso Nacional de Teledetección. Puerto de la Cruz. Tenerife

Cuadrat J. M., Pita M. F. (1997). Climatología. Madrid. Cátedra. Capítulo II: 19-40.

Díaz, G. Sánchez, I.; Quiroz, R.; Garatuza, J.; Watts, C. y Cruz, I. (2008). Interpolación Espacial de la Precipitación Pluvial en la zona de Barlovento y Sotavento del Golfo de México. Agricultura Técnica de México. 34(3): 279-287.

McVicar, T., Van Niel, T., LingTao L., Hutchinson, M., XingMin M., ZhiHong L. (2007). “Spatially distributing monthly reference evapotranspiration and pan evaporation considering topographic influences” ScienceDirect. Journal of Hydrology 338: 196-220.

Ninyerola, M.; Pons, X. y Roure, JM. (2000). A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques”. International Journal of Climatology. 20: 1823-1841.

OMM.(1996). Guía de Instrumentos y Métodos de Observación Meteorológicos.

Philips, D., Marks, D. (1996). “Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models”. Elsevier. Ecological Modelling 91 : 213-229

QiaoJing Q., Bingfang W., Jun X. (2005). “Interpolation System for Generating Meteorological Surfaces Using to Compute Evapotranspiration in Haihe River Basin”. Institute of Remote Sensing Applications, Chinese Academy Sciences.

Scholoeder, C.A., Zimmerman, N.E., Jacobs, M.J. (2001). Comparison of methods for interpolating soil properties using limited data. Soil Science Society of American Journal 65: 470-479.

Vicente, S., Lanjeri, S., López-Moreno, J. (2007). Comparison of different procedures to map reference evapotranspiration using geographical information systems and regression-based techniques. International Journal of Climatology 27: 1103-1118

Villaroto, M., Henríquez, C., Sancho, F. (2008) Comparación de los interpoladores IDW y Kriging en la variación espacial de PH, CA, CICE, y P del suelo. Agronomía Costarricense 32(1): 95-100.