COMPARATIVA DE MÉTODOS DE INTERPOLACIÓN

PARA EL PROYECTO

"ELABORACIÓN DE ESCENARIOS LOCALES Y REGIONALES DE CAMBIO CLIMÁTICO ADAPTADOS AL SEXTO INFORME DEL GRUPO INTERGUBERNAMENTAL DE EXPERTOS SOBRE EL CAMBIO CLIMÁTICO (IPCC) Y SU DIFUSIÓN A TRAVÉS DE LAS TECNOLOGÍAS DE LA INFORMACIÓN Y LA COMUNICACIÓN"

INTRODUCCIÓN 2

1. DATOS UTILIZADOS 3

1.1. SELECCIÓN DE OBSERVATORIOS 5

2. METODOLOGÍA 7

2.1. ZONIFICACIÓN POR ESTRATOS 7

2.2. USO DE COVARIABLES 7

2.3. METODOLOGÍAS DE INTERPOLACIÓN 7

2.3.1. REGRESIÓN LINEAL MÚLTIPLE CON INTERPOLACIÓN DE RESIDUOS 8

2.3.3. INVERSE DISTANCE WEIGHTED (IDW) 10

2.3.4. THIN PLATE SPLINES (TPS) 10

2.3.5. KRIGING 13

3. RESULTADOS DE LA REGRESIÓN LINEAL MÚLTIPLE CON INTERPOLACIÓN DE RESIDUOS POR ESTRATOS 17

3.1. REGRESIÓN LINEAL CON COVARIABLES 17

3.1. INTERPOLACIÓN DE LOS RESIDUOS 19

4. INTERPOLACIÓN A TODA ANDALUCÍA 27

5. CONCLUSIONES 33

INTRODUCCIÓN

El presente documento detalla el proceso de determinación de la metodología de interpolación empleada en el proyecto "Elaboración de escenarios locales y regionales de cambio climático adaptados al sexto informe del grupo intergubernamental de expertos sobre el cambio climático (IPCC) y su difusión a través de las tecnologías de la información y la comunicación" de la Junta de Andalucía.

El objetivo de este estudio ha sido seleccionar el mejor método de interpolación de las principales variables climatológicas del proyecto, la precipitación y la temperatura, principales porque a partir de éstas se pueden determinar todas las demás variables del proyecto. Recordemos que el objetivo final es la creación de una rejilla del territorio andaluz (de un mapa) de una resolución de 200 metros. Estudiando el error cometido por los diferentes métodos de interpolación llegaremos a determinar cuál es el mejor método.

Recordemos que los trabajos de interpolación definidos en el Plan de Trabajo del proyecto se especificaron como

1. Estudio de las regresiones lineales por estratos y
2. Estudio de la mejor forma de interpolar los residuos entre los siguientes métodos: Kriging, Inverse distance weighting (IDW) y Thin Plate Spline (TPS).

Yendo más allá, y de manera adicional, se han realizado estudios complementarios para fortalecer la toma de decisiones para la elección de la mejor forma de interpolación. En este sentido, se han realizado estudios de las interpolaciones de las variables directamente con la altitud como covariable para cada estrato y también para toda Andalucía.

1. DATOS UTILIZADOS

Una de las asunciones de partida del estudio a realizar, especificada en el propio pliego técnico, es la división del territorio andaluz en 5 zonas bioclimáticas; la denominación que vamos a utilizar de tales zonas es Málaga, Cádiz, Guadalquivir, Este y Norte (Figura 1.1). Inicialmente, se presupone que la interpolación debe consistir en interpolar independientemente en cada una de esas 5 zonas y luego unir todos los resultados obtenidos. Esto es así porque, al menos de entrada, y tal y como se especifica en el Pliego Técnico, se considera que en Andalucía las variables climáticas de primer orden están relacionadas con determinados parámetros geográficos tales como la altitud, distancia al océano atlántico y mediterráneo. Sin embargo, el comportamiento de estas variables también está muy determinado por los sectores geográficos donde se encuentran ya que las barreras que conforman el relieve dan lugar a comportamientos diferenciados en poca distancia ya que en función de la dirección de entrada de los vientos dominantes y frentes húmedos, variables como la precipitación se distribuyen de manera muy diferenciada en sotavento respecto a barlovento, indistintamente de la altitud. Precisamente esos estratos geográficos se consideran definidos por el relieve y orientación a los principales tipos de frentes.

Figura 1.1. Zonas bioclimáticas (estratos) en las que se divide Andalucía para este proyecto.

Para la interpolación se han utilizado los datos de los observatorios suministrados, que han sido tratados y corregidos tras el control de calidad. Los datos de los observatorios presentan lagunas de datos y por lo tanto, sólo se han seleccionado observatorios con datos suficientes para caracterizar las variables en cuestión. Recordemos que, en cuanto a las variables climáticas a interpolar, éstas hacen referencia a la media mensual para temperatura y a la media del acumulado mensual para precipitación del periodo 1991-2020.

Para las variables medias de temperatura se han seleccionado observatorios que, de los 30 años empleados (1991 - 2019) tengan al menos el 40% de datos diarios de los meses a representar. Para las variables de medias acumuladas de precipitación se han seleccionado observatorios que en al menos en 15 años del periodo de referencia existen meses con el 60% de dato diario. Los meses incompletos se han completado añadiendo la media diaria de dicho mes del dato de precipitación a los días sin dato.

Así, en total se han utilizado 788 observatorios de temperatura y 1655 observatorios de precipitación.

El Modelo Digital de Elevación (MDE) a 200 metros de resolución utilizado se ha obtenido del Centro de Descargas del Centro Nacional de Información Geográfica (https://centrodedescargas.cnig.es/). Esa elevación será la que se emplee, si así se considera, a la hora de interpolar en cada punto de grid (figura 1.2). Además, a partir de este MDE se ha obtenido la distancia euclidiana tanto a la costa atlántica como a la mediterránea, dos posibles covariables a tener en cuenta a la hora de realizar la interpolación.

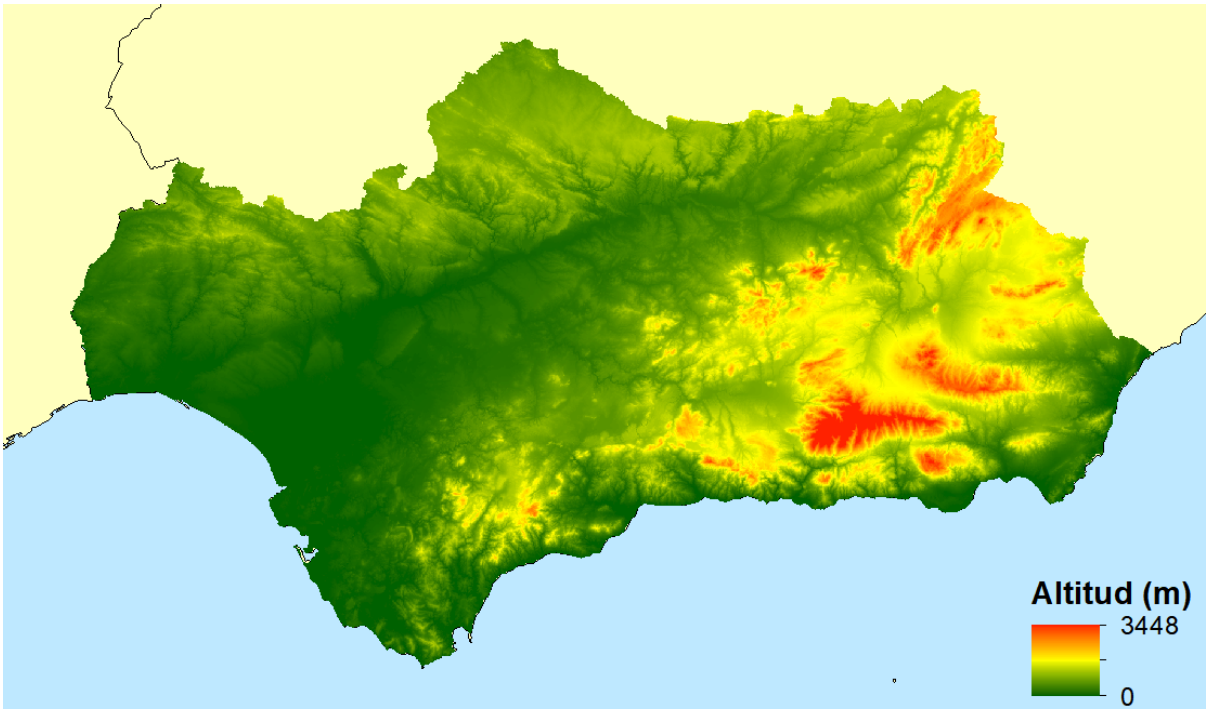

Figura 1.2. Mapa de elevaciones a partir del MDE utilizado en el estudio.

1.1. SELECCIÓN DE OBSERVATORIOS

En las tablas 1.1 (temperatura) y 1.2 (precipitación) se detallan la cantidad de observatorios disponibles en los estratos a estudiar, agrupados por cada variable cuyo posible error vamos a estudiar a la hora de interpolar (como luego veremos, las pruebas realizadas se han ceñido a las precipitaciones medias acumuladas de enero y julio y a las temperaturas medias máxima y mínima de enero y julio). Tales observatorios son

- Para las variables de temperatura máxima y mínima media de enero se han seleccionado los siguientes observatorios para cada estrato: 37 en Cádiz, 61 en Norte, 161 en Guadalquivir, 57 en Málaga y 115 en este.
- Para las variables de temperatura máxima y mínima media de julio se han seleccionado: 32 en Cádiz, 46 en Norte, 148 en Guadalquivir, 51 en Málaga y 11 en Este.
- Para la variables de precipitación acumulada media de enero se han seleccionado los siguientes observatorios para cada estrato: 77 en Cádiz, 154 en Norte, 322 en Guadalquivir, 113 en Málaga y 280 en Este.
- Para la variable de precipitación acumulada media de julio se han seleccionado: 66 en Cádiz, 148 en Norte, 300 en Guadalquivir, 97 en Málaga y 259 en Este.

Tabla 1.1 Observatorios de temperatura seleccionados

Estratos	Observatorios seleccionados		Observatorios disponibles
	Temperatura máxima y mínima media de enero	Temperatura máxima y mínima media de julio	
Cádiz	37	32	60
Norte	61	46	129
Guadalquivir	161	148	275
Málaga	57	51	111
Este	115	100	213
Total	431	377	788

Tabla 1.2 Observatorios de precipitación seleccionados			
Estratos	Observatorios seleccionados		Observatorios disponibles
	Precipitación acumulada media de enero	Precipitación acumulada media de julio	
Cádiz	77	66	121
Norte	154	148	267
Guadalquivir	322	300	521
Málaga	113	97	246
Este	280	259	500
Total	946	870	1655

2. METODOLOGÍA

Pasemos a detallar las diferentes metodologías de interpolación a emplear y los distintos criterios de interpolación a probar. Básicamente, la prueba tiene una dimensión triple:

1. La zonificación por estratos bioclimáticos ¿tiene sentido a la hora de interpolar? ¿O sería mejor una interpolación global de todo el territorio?
2. ¿Tiene sentido usar alguna covariable en la interpolación, además de las coordenadas propias del punto en el que vamos a interpolar?
3. ¿Cuál es el mejor método de interpolación a emplear?

2.1. ZONIFICACIÓN POR ESTRATOS

Tal y como se ha especificado en el pliego técnico, debe considerarse la división del territorio andaluz en 5 zonas bioclimáticas (que vimos en la Figura 1.1) ya que, inicialmente, se presupone que la interpolación debe consistir en interpolar independientemente en cada una de esas 5 zonas y luego unir todos los resultados obtenidos. Sin embargo, es posible que una interpolación global de toda el área de la superficie de Andalucía sea mejor que una interpolación por estratos, por lo que llevaremos a cabo pruebas para comprobar qué estrategia es mejor.

2.2. USO DE COVARIABLES

¿Tiene sentido utilizar covariables en la interpolación, además de las coordenadas propias del punto en el que vamos a interpolar? Como veremos, en el uso de la regresión múltiple como técnica de interpolación se utilizan varias covariables, pero intentaremos ir más allá y comprobar si el uso de las covariables es útil para el método de interpolación que empleemos y según la variable a interpolar.

2.3. METODOLOGÍAS DE INTERPOLACIÓN

Debemos recordar que existe una enorme variedad de métodos de interpolación, basados en diversas aproximaciones matemáticas. No existe un método de interpolación del que se pueda declarar unánimemente que es el mejor puesto que las diferentes condiciones de una interpolación (número de observatorios, densidad geográfica de éstos, naturaleza de la variable a estudiar,…) hacen que los resultados obtenidos puedan variar mucho en función de la región en la que se realiza la interpolación y según la variable a interpolar.

Es importante destacar que, sea cual sea el método a emplear, la facilidad de uso de la aplicación informática a utilizar en su ejecución, y, muy especialmente, el tiempo de computación necesario para su cálculo, son factores de extraordinaria importancia tanto a la hora de acometer las interpolaciones como a la hora de decidir si su utilización es verdaderamente practicable. Para garantizar que ambos requisitos se cumplan de manera

adecuada, se ha empleado el entorno del software estadístico R (R Development Core Team, 2009) para realizar las interpolaciones y los estudios derivados.

Veamos ahora una descripción de la metodología general a emplear (la regresión lineal múltiple con interpolación de residuos) y de las metodologías de interpolación de los propios residuos.

2.3.1. REGRESIÓN LINEAL MÚLTIPLE CON INTERPOLACIÓN DE RESIDUOS

Este método es el inicialmente sugerido para realizar la interpolación en el Pliego Técnico. En este método se obtiene la espacialización de la variable climática a través de una regresión lineal múltiple y la interpolación de los residuos obtenidos. En este caso, se ha realizado un estudio para seleccionar el método de interpolación de residuos que comete menos errores. El método se puede resumir en 7 pasos (figura 2.1) que, en nuestra prueba, han sido aplicados para cada estrato tal y como se especifica en el propio Pliego Técnico. Recordemos que en el propio Pliego Técnico se especifica que el método para la interpolación de los residuos es el método del inverso de la distancia (IDW), pero que aquí hemos decidido ampliar la prueba probando también con otros dos métodos, Thin Plate Splines (TPS) y Kriging.

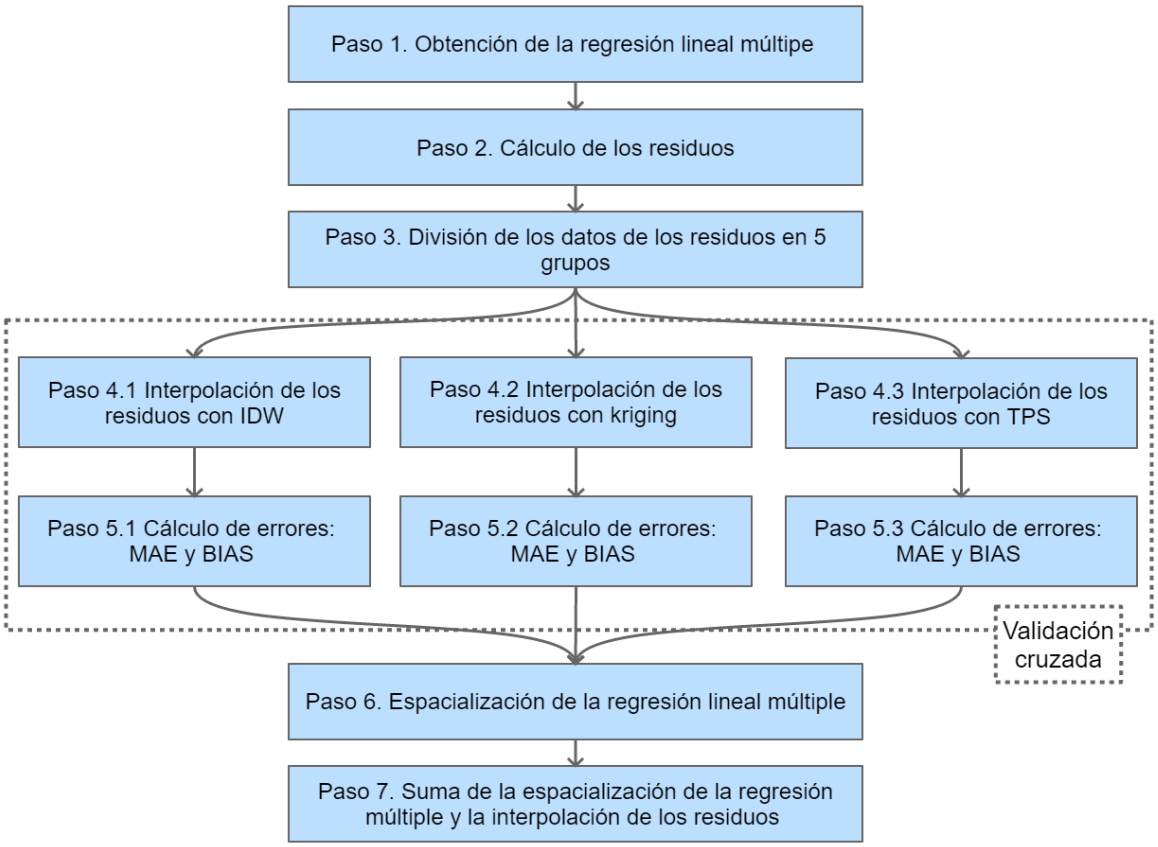

Figura 2.1. Metodología para la espacialización de variables mediante regresión lineal múltiple e interpolación de residuos.

Las pruebas realizadas, dada su extensión y complejidad, se han ceñido a las precipitaciones medias acumuladas de enero y julio y a las temperaturas medias máxima y mínima de enero y julio. En todos los casos, enero y julio se han elegido como meses representativos de dos situaciones antagónicas (un verano tradicionalmente caluroso y seco en julio y un invierno tradicionalmente más húmedo y más frío en enero) que permiten muestrear el comportamiento de las variables en dos situaciones extremas.

Veamos ahora en detalle el proceso de prueba especificado en la figura 2.1.

Paso 1. Obtención de la regresión lineal múltiple

Para cada variable se ha realizado una regresión lineal múltiple con los datos propios de los observatorios. La variable dependiente ha sido la propia variable climática y entre las variables independientes: en todos los casos se ha seleccionado la altitud, y la distancia al mar Mediterraneo y/o al mar Atlántico siempre y cuando fuesen variables significativas.

Paso 2. Cálculo de los residuos para cada observatorio

Una vez obtenida la regresión múltiple se ha obtenido la variable predicha en cada observatorio. Los residuos han sido calculados a través de la diferencia entre la variable observada y la variable predicha.

Paso 3. División de los datos de los residuos en 5 grupos

Los residuos de los observatorios se han dividido en 5 grupos aleatorios para realizar la validación cruzada.

Paso 4. Interpolación de los residuos

Los residuos de cada variable para cada estrato se han interpolado mediante tres métodos: kriging, TPS e IDW. Con el objetivo de conocer el mejor método de interpolación se ha realizado una validación cruzada con los grupos definidos del paso anterior. Cada método de interpolación se realizó con 4 de los 5 grupos definidos, de esta manera se han obtenido 5 interpolaciones excluyendo en cada una de ellas un grupo para cada método.

Paso 5. Cálculo de errores: MAE y BIAS

En cada interpolación realizada, con el grupo de observatorios que no se ha utilizado se ha obtenido el valor predicho en la interpolación en cada observatorio. Por lo tanto para cada observatorio del grupo se obtendrá un valor de residuo predicho y otro calculado. Con la diferencia de ambos valores se calcula el Error Absoluto Medio (MAE) y el error BIAS.

Paso 6. Espacialización de la regresión lineal múltiple

En cada estrato se ha obtenido la espacialización de la regresión lineal múltiple obtenida para cada variable. Los datos de entrada han sido el MDE a 200 metros de resolución y la distancia al mar Atlántico y/o Mediterráneo en el caso de que hubiesen sido seleccionadas como significativas.

Paso 7. Obtención de la variable climática espacializada

El último paso ha consistido en la suma de la regresión múltiple espacializada y las interpolaciones de los residuos. De esta manera, se ha obtenido la variable climática espacializada.

2.3.3. INVERSE DISTANCE WEIGHTED (IDW)

El método IDW (Inverse Distance Weighted o Pesado por Inversa de la Distancia), en alguna de sus variantes, es uno de los más utilizados en la simulación de variables por interpolación. Básicamente consiste en el cálculo del valor de la variable según el valor de ésta en los vecinos más cercanos al punto donde vamos a interpolar; el valor en los vecinos más cercanos será pesado con respecto a nuestro punto según una cierta función peso definida en base a la distancia que separa a tal vecino del punto en el que queremos simular. Podremos elegir tanto el radio máximo de distancia de búsqueda de los vecinos más cercanos como el número de vecinos más cercanos a considerar; tomar uno de los dos criterios o ambos a la vez nos permitirá obtener distintos valores en la interpolación según cómo definamos la vecindad del punto a simular. El método más sencillo y extendido consiste en asociar el peso a la inversa de la distancia elevada a una cierta potencia, generalmente a un valor de 1 o, más usualmente, de 2 (Shepard, 1968). En nuestro caso utilizaremos la potencia de 2 como el ajuste básico de las pruebas.

2.3.4. THIN PLATE SPLINES (TPS)

Thin Plate Splines es un método que ha sido ampliamente probado y apoyado en multitud de estudios (Haylock et al, 2008), y que se ha utilizado en multitud de importantes bases de datos climatológicas dada su elevada capacidad de simulación con un bajo índice de error.

El método de Thin Plate Splines (una variante concreta del más general método de splines para ajuste de curvas y superficies a un grupo de puntos dados) debe su nombre a la analogía física de una fina hoja de metal cuya curvatura es ajustada de tal manera que, aun imponiendo que debe pasar exactamente por los valores iniciales proporcionados (nuestras observaciones), esa curvatura minimice la “tensión” impuesta sobre la superficie. Esa “hoja de metal”, esa superficie con tensión minimizada, sería nuestra interpolación (figura 2.2).

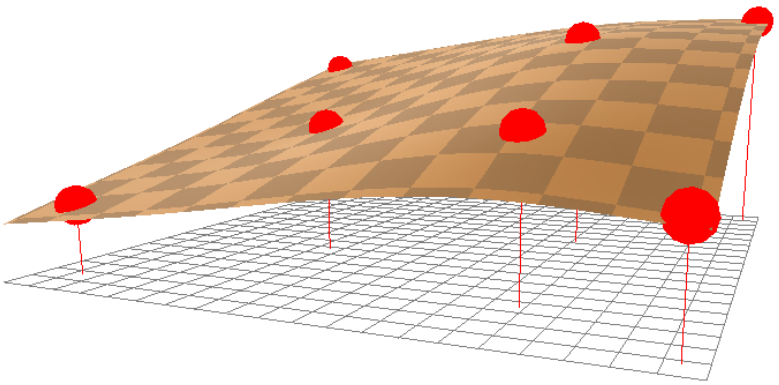

Figura 2.2 Visualización de la forma en que se comporta Thin Plate Splines: una superficie de tensión mínima que se ajusta exactamente a los puntos iniciales (Fuente: Jarno Clonen. URL: http://elonen.iki.fi/code/tpsdemo).

El método de Thin Plate Splines permite, por lo tanto, tratar a la vez la exactitud de la superficie de ajuste y su curvatura. En nuestro caso, vamos a utilizar la metodología de Hutchinson (1.991), una metodología, además, especialmente recomendada para la interpolación de variables meteorológicas. En esta metodología se utiliza una forma generalizada del cross-validation para optimizar la función de suavizado de tal forma que no aparezca una sobre-suavización que produzca una variación significativa de la superficie con respecto a los valores suministrados, es decir, que la "tensión" de la superficie (que es la manera de controlar su forma) se mantenga optimizada. Esa "tensión" se controla mediante la ecuación

$$\sum_{j=1}^{N}\left[z_j - F(r_j) - \sum_{k=1}^{p} B_k \psi_{kj}\right]^2 + \lambda I(F)$$

donde λ es el parámetro de suavizado, I(F) es la "seminormal de suavizado", ψ es la función de escalado de la distancia euclídea con coeficientes B, y p es el número de covariantes parciales lineales. La "seminormal de suavizado" es una medida de la rugosidad (*roughness*) de la superficie a crear (en rigor, de la capacidad de derivar la superficie creada). Es importante comprender que, gracias a ese cross-validation generalizado, la optimización de los parámetros de la superficie (es decir, la interpolación) se realiza de forma automática.

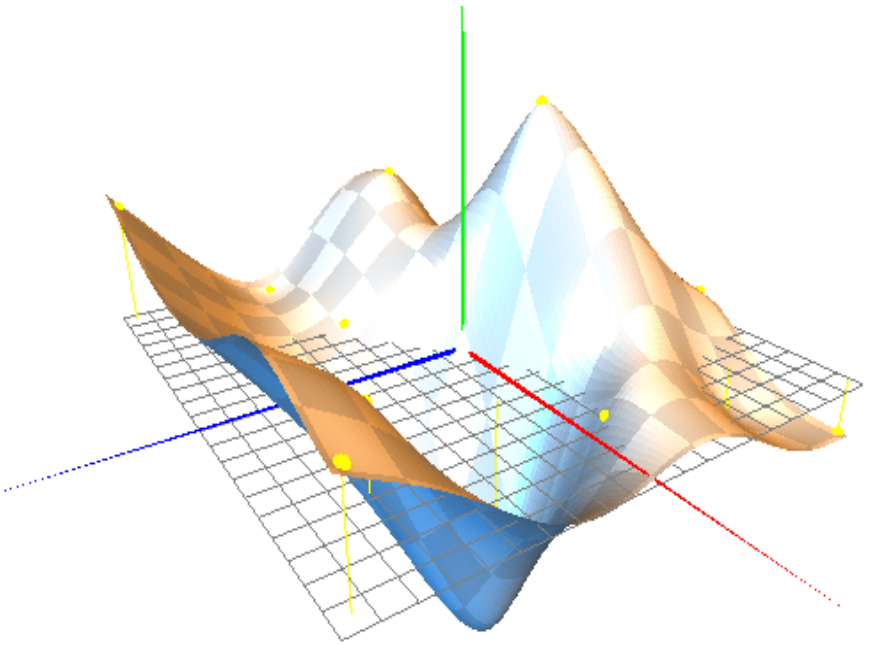

Figura 2.3: visualización de una superficie de interpolación creada por Thin Plate Splines; la superficie pasará por los valores suministrados y tendrá "tensión" mínima (Fuente: Jarno Clonen. URL: http://elonen.iki.fi/code/tpsdemo).

Una de las principales ventajas de Thin Plate Splines es que es un método insesgado y exacto, y además no necesita de estudios previos de la distribución de los datos (como le ocurre al método Kriging de interpolación, que necesita del estudio manual previo de variogramas), con lo que su ejecución es automática. Otra ventaja de este método es que posee diferentes variantes de interpolación, que permiten considerar otras variables; las principales y más utilizadas son la que permite introducir otras variables cuya covarianza tiene relación con la variable a interpolar, y la que considera más de dos dimensiones en la distribución de la variable (como el Thin Plate Splines de 3 dimensiones).

La principal desventaja de este método proviene de que, en ocasiones, y si las condiciones de contorno no están bien definidas, o existen grandes huecos en la densidad de valores suministrados, el ajuste de la superficie puede producir valores que deben ser rechazados por no reales (véase la figura 2.4). Obsérvese que determinar que un cierto valor "debe ser rechazado por no real" implica conocer *a priori* el comportamiento general de la variable a interpolar (por ejemplo, si tiene o no sentido que la temperatura alcance ciertos valores en un cierto mes) y estudiar a posteriori el resultado de una cierta interpolación.

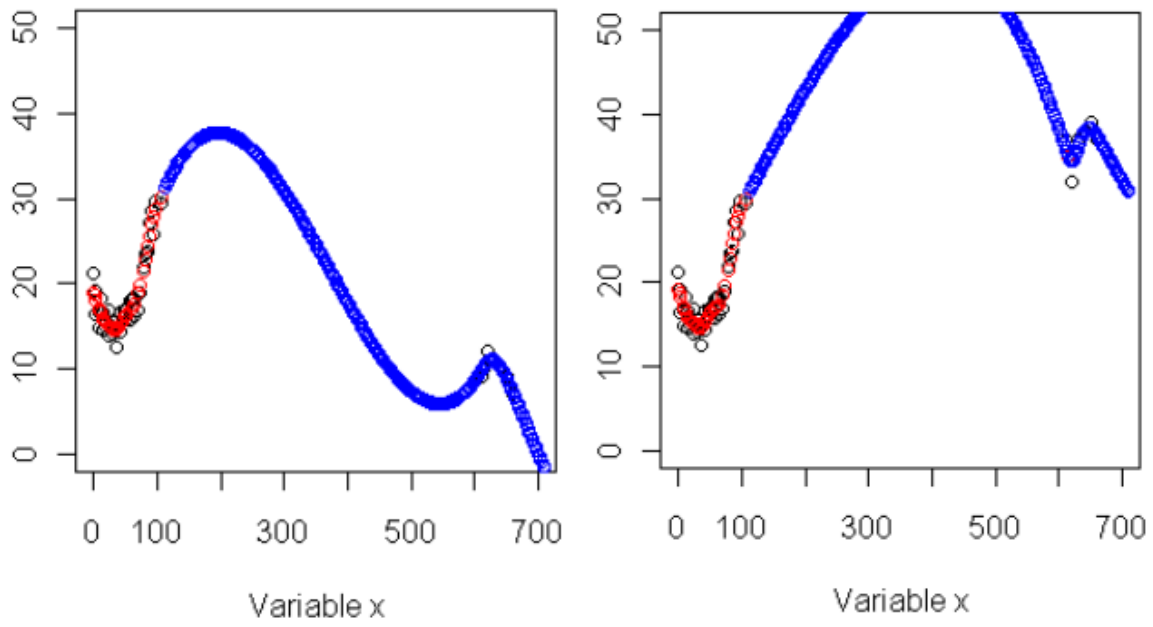

Figura 2.4: ilustración, para un grupo de datos unidimensionales (donde, dado *x*, tengo *f(x)*), del comportamiento de Thin Plate Splines. Obsérvese que el grupo de datos principal, el comprendido entre los valores 0 y 100, y representando mediante los círculos negros, se interpola bien (los círculos rojos) en ambos casos. Los posibles valores del subgrupo de la derecha (los situados aproximadamente en 600), pintados mediante círculos negros, afectan mucho (a la izquierda, valores más bajos; a la derecha, más altos) a la superficie Thin Plate Splines a crear, debido a la fuerte ausencia de valores intermedios.

Debe hacerse notar que este problema puede afectar a cualquier método de interpolación utilizado (una falta de valores tan elevada como la representada en la figura 2.4 puede afectar de maneras muy variadas a cualquier método), y por lo tanto, sea cual sea el método de interpolación a utilizar, siempre debemos realizar el estudio a posteriori de la interpolación.

2.3.5. KRIGING

Kriging es un método de interpolación espacial originado en el campo de la geología minera, y de hecho se denomina así en honor del ingeniero de minas sudafricano Danie Krige, su creador. Su formulación original se encuentra en Matheron (1963), y se recomienda su estudio a través de Bivand et al (2008).

Es uno de los varios métodos que utilizan un conjunto limitado de puntos de datos muestreados para estimar el valor de una variable en un campo espacial continuo. Un ejemplo de un valor que varía a través de un campo espacial aleatorio podría ser el promedio mensual de las concentraciones de ozono en una ciudad, o la disponibilidad de alimentos saludables en los vecindarios. Difiere de métodos más simples, como la Interpolación por Inversa de la Distancia (IDW) o la Regresión Lineal en que utiliza la correlación espacial entre los puntos muestreados para interpolar los valores en el campo espacial: la interpolación se basa en la disposición espacial de las observaciones empíricas, más que en un presunto modelo de

distribución espacial. Una de las mayores fortalezas de Kriging es que también genera estimaciones de la incertidumbre que rodea a cada valor interpolado.

En un sentido general, los pesos de Kriging se calculan de tal manera que se da más peso a los puntos cercanos al lugar de interés que a los más alejados. También se tiene en cuenta la agrupación de puntos, de modo que los grupos de puntos se ponderan menos (en efecto, contienen menos información que los puntos individuales). Esto ayuda a reducir el sesgo de las predicciones.

El predictor Kriging es un "predictor lineal óptimo" y un interpolador exacto, lo que significa que cada valor interpolado se calcula para minimizar el error de predicción para ese punto. El valor que se genera a partir del proceso de kriging para cualquier lugar realmente muestreado será igual al valor observado en este punto, y todos los valores interpolados serán los mejores predictores lineales sin sesgo (BLUP). En general, Kriging no será más eficaz que los métodos más sencillos de interpolación si hay poca autocorrelación espacial entre los puntos de datos muestreados (es decir, si los valores no se coevolucionan en el espacio). Sin embargo, si hay al menos una autocorrelación espacial moderada, el kriging puede ser un método útil para preservar la variabilidad espacial que se perdería si se utilizara un método más sencillo.

Kriging puede entenderse como un proceso de dos etapas: en primer lugar, la estructura de covarianza espacial de los puntos muestreados se determina ajustando un variograma; y en segundo lugar, los pesos derivados de esta estructura de covarianza se utilizan para interpolar los valores de los puntos o bloques no muestreados en todo el campo espacial.

Un variograma (a veces denominado "semivariograma") es una representación visual de la covarianza exhibida entre cada par de puntos en los datos muestreados. Para cada par de puntos en los datos muestreados, el valor gamma o "semivarianza" (una medida de la diferencia media cuadrada entre sus valores) se representa contra la distancia, o "lag", entre ellos. El variograma "experimental" es la representación gráfica de los valores observados, mientras que el variograma "teórico" o "modelo" es el modelo distributivo que mejor se ajusta a los datos. Los modelos de variograma se obtienen a partir de un número limitado de funciones "autorizadas", incluidos los modelos lineal, esférico, exponencial y de potencia.

La elección de un modelo de variograma está fundamentalmente definida por el usuario, aunque los programas informáticos estadísticos a menudo pueden ayudar a definir los modelos que mejor se adaptan utilizando diversos enfoques que incluyen los mínimos cuadrados, la máxima probabilidad y los métodos bayesianos. En el ejemplo que figura a continuación (figura 2.5) de variogramas para las tasas de cólera y disentería en una zona de Bangladesh, se ha elegido un modelo exponencial como el modelo "más adecuado" para un variograma de las tasas de cólera, mientras que un modelo esférico se ajusta mejor a las tasas de disentería. En ambos casos, la curva ascendente en distancias cortas implica que los lugares que están más cerca entre sí son más similares entre sí que los lugares que están separados. En un cierto punto (~300 metros para el cólera y ~1000 metros para la disentería), los valores gamma alcanzan una meseta, indicando que la diferencia entre las tasas en los

lugares separados por esta distancia han alcanzado el valor de la varianza global de la muestra.

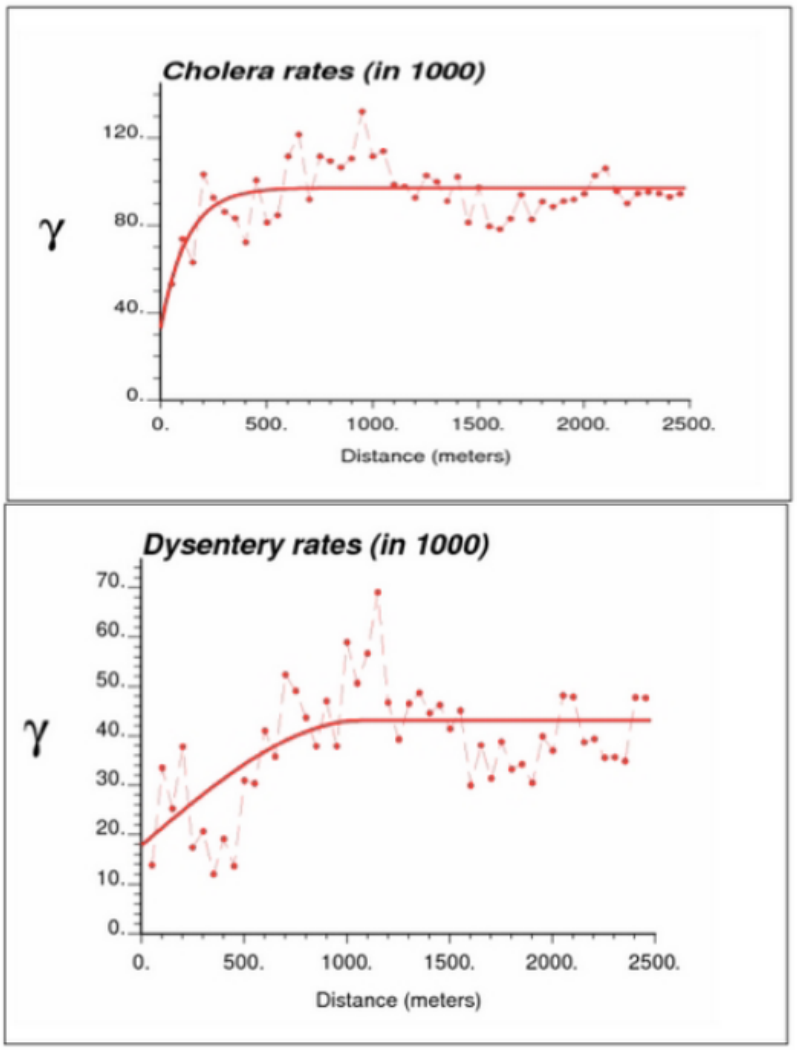

Figura 2.5 Ejemplo de diferentes variogramas experimentales (puntos) y teóricos (líneas de ajuste a los puntos). Tomado de Ali et al (2006).

Existen varios tipos de Kriging; los más utilizados en interpolaciones de variables climáticas son:

• Kriging ordinario, para el cual debe asumirse la suposición de estacionariedad (que la media y la varianza de los valores es constante en todo el campo espacial). Esta es una de las formas más sencillas de kriging, pero el supuesto de estacionariedad no se cumple a menudo en muchas variables atmosféricas.

• Kriging universal, que relaja el supuesto de estacionariedad permitiendo que la media de los valores difiera de forma determinante en diferentes lugares (por ejemplo, a través de algún tipo de tendencia espacial), mientras que sólo la varianza se

mantiene constante en todo el campo. Esta estacionariedad de segundo orden es a menudo una suposición pertinente con las variables atmosféricas.

• Kriging de bloques (Block kriging), que estima valores promediados sobre "bloques" cuadriculados en lugar de puntos individuales. Estos bloques a menudo tienen menores errores de predicción que los que se observan en los puntos individuales.

• Cokriging, en el que se utilizan variables adicionales observadas (que a menudo están correlacionadas entre sí y con la variable de interés) para aumentar la precisión de la interpolación de la variable de interés en cada lugar. Es una técnica habitual a la hora de interpolar variables meteorológicas (por ejemplo, utilizando la altura como covariable para interpolar temperatura).

Como principal problema, Kriging necesita de un ajuste teórico del variograma empírico que relaciona cómo varía la variable a interpolar en función de la distancia (para los observatorios a emplear), y tal ajuste no siempre es todo lo bueno que se desearía. Además, el mismo problema que vimos en la figura 2.4 para TPS puede presentarse aquí para Kriging.

3. RESULTADOS DE LA REGRESIÓN LINEAL MÚLTIPLE CON INTERPOLACIÓN DE RESIDUOS POR ESTRATOS

3.1. REGRESIÓN LINEAL CON COVARIABLES

En el primer paso del método se han obtenido las regresiones múltiples para cada variable y cada estrato (área bioclimática). La tabla 3.1 nos muestra, para cada variable (temperatura y precipitación en enero y julio, recordemos), el resultado de realizar una regresión lineal múltiple con la altitud, la distancia al Mediterráneo y la distancia al Atlántico como covariables. Si alguna de esas covariables es significativa en la variable y el estrato considerados, aparece marcada, y como medida de la bondad del ajuste de la regresión lineal múltiple se ha utilizado el estadístico R cuadrado ajustado.

Tabla 3.1. Variables independientes seleccionadas y R cuadrado ajustado de la regresión lineal múltiple de cada variable climática estudiada en cada estrato.

Variable	Estrato	altitud	Distancia al Mediterráneo	Distancia al Atlántico	R cuadrado ajustado
Precipitación acumulada de enero	Cadiz	x	x	x	0.72
	Norte	x	x	x	0.41
	Guadalquivir	x	x	x	0.13
	Málaga	x		x	0.5
	Este	x	x	x	0.27
Precipitación acumulada de julio	Cadiz	x			0.17
	Norte	x	x		0.14
	Guadalquivir	x			0.03
	Málaga	x			0.0009
	Este	x	x	x	0.36
Temperatura mínima media	Cadiz	x		x	0.54

de enero	Norte	x		x	0.31
	Guadalquivir	x		x	0.5
	Málaga	x	x		0.62
	Este	x	x		0.75
Temperatura máxima media de enero	Cadiz	x			0.9
	Norte	x			0.71
	Guadalquivir	x	x		0.72
	Málaga	x			0.5
	Este	x	x		0.72
Temperatura mínima media de julio	Cadiz	x			0.15
	Norte	x			0
	Guadalquivir	x			0.09
	Málaga	x			0.27
	Este	x	x		0.57
Temperatura máxima media de julio	Cadiz	x	x	x	0.58
	Norte	x	x		0.56
	Guadalquivir	x	x	x	0.56
	Málaga	x	x		0.34
	Este	x	x	x	0.38

De los resultados mostrados en la tabla 3.1 se pueden extraer una serie de importantes conclusiones:

1. Con la única excepción de la precipitación de enero en Cádiz, y como muestran los resultados de R cuadrado ajustado, utilizar la regresión lineal múltiple en la variable de precipitación no aporta ninguna ventaja a la interpolación: R cuadrado es demasiado bajo (y más en julio, es decir, cuando apenas hay precipitación), es decir, la regresión lineal múltiple no es capaz de explicar significativamente la variabilidad de la precipitación, con lo que realizar la regresión no aporta nada a nuestro estudio - aunque continuaremos realizándolo buscando futuras conclusiones.
2. En general, y sin tomar cada estrato por separado, las temperaturas de enero sí muestran que la regresión lineal múltiple puede ser útil para intentar explicar su

variabilidad (los R cuadrado ajustado son moderadamente buenos) pero no así en julio (un mes especialmente caluroso en Andalucía), cuando se muestra que la regresión lineal múltiple no aporta prácticamente ninguna explicación de la variación de la variable (y su uso estaría, por lo tanto, desaconsejado).

3. En cuanto a las covariables empleadas, sólo la altura es significativa en todas las regresiones (por muy malas que éstas sean) y la aparición de las distancias al mediterráneo y al atlántico puede calificarse de esporádica con la excepción de la precipitación de enero (cuando los frentes provenientes del atlántico provocan la aparición de las mayores precipitaciones acumuladas) y la temperatura máxima en julio.

3.1. INTERPOLACIÓN DE LOS RESIDUOS

Veamos ahora los mapas generados al interpolar los residuos de la regresión lineal múltiple. Las figuras 3.1, 3.2, 3.3 y 3.4 muestran algunos ejemplos de la interpolación de las variables en algunos de los estratos con el objetivo de poder apreciar a simple vista la posible coherencia espacial de la distribución de lo obtenido. El resultado más apreciable es que la interpolación de residuos mediante IDW presenta el fenómeno conocido como *dianas*, en el que, alrededor de los observatorios empleados (y de forma más llamativa alrededor de los que tienen mayores valores), se forman dianas (valores concéntricos alrededor de los observatorios) como consecuencia de la dependencia de la interpolación de IDW con respecto al cuadrado de la distancia al punto. Este efecto provoca efectos irreales alrededor de los observatorios (las variables no se comportan así) que impiden una interpolación suave y homogénea en todo el territorio.

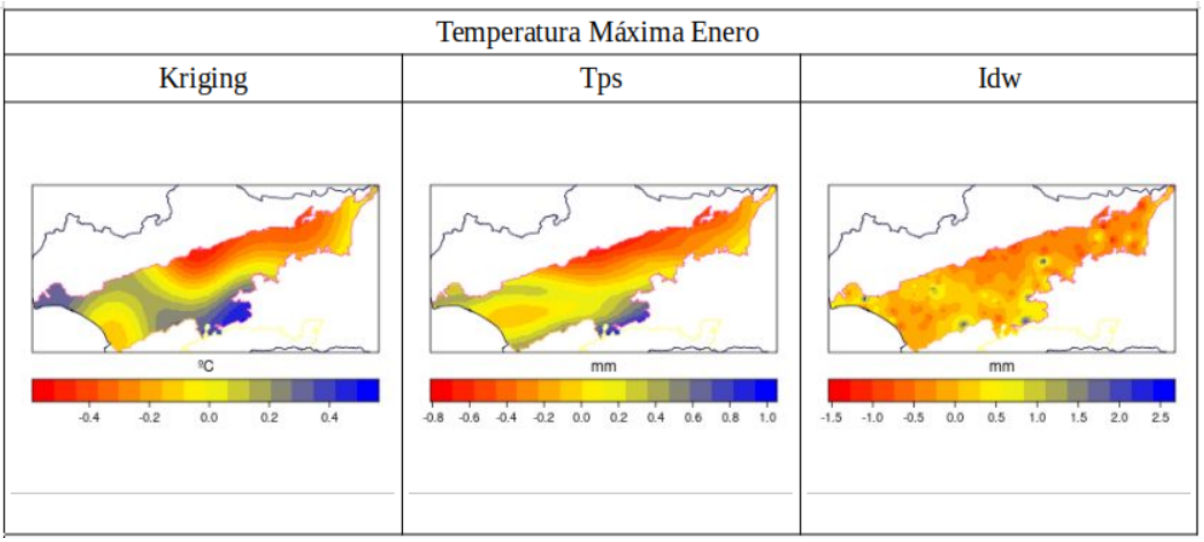

Figura 3.1. Resultados de la interpolación de los residuos de la regresión lineal múltiple de la temperatura máxima de enero en la región del Guadalquivir para las tres metodologías estudiadas.

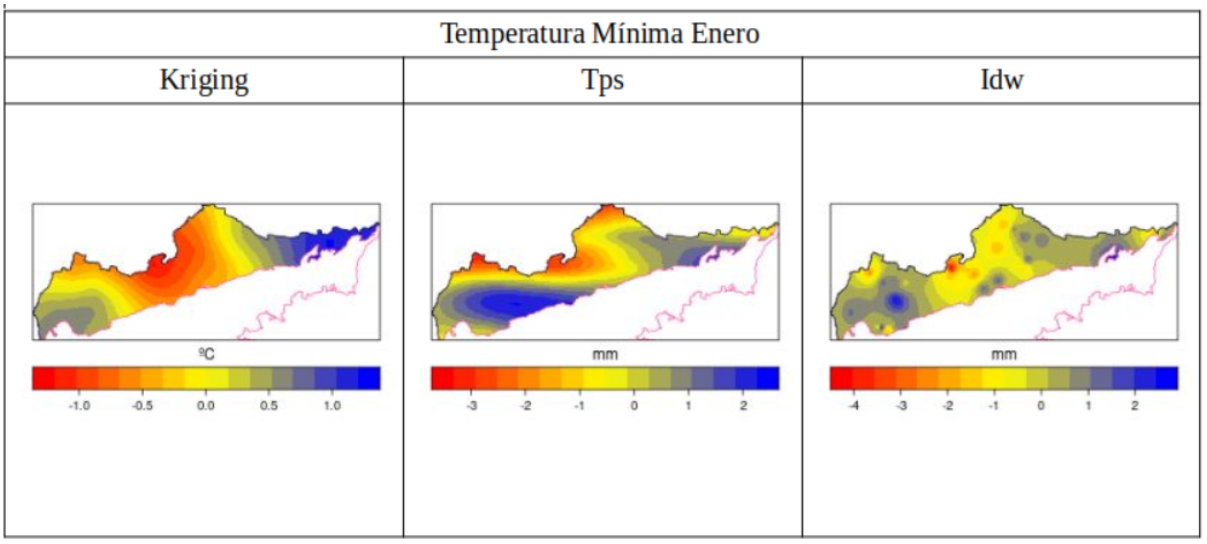

Figura 3.2. Resultados de la interpolación de los residuos de la regresión lineal múltiple de la temperatura mínima de enero en la región Norte para las tres metodologías estudiadas.

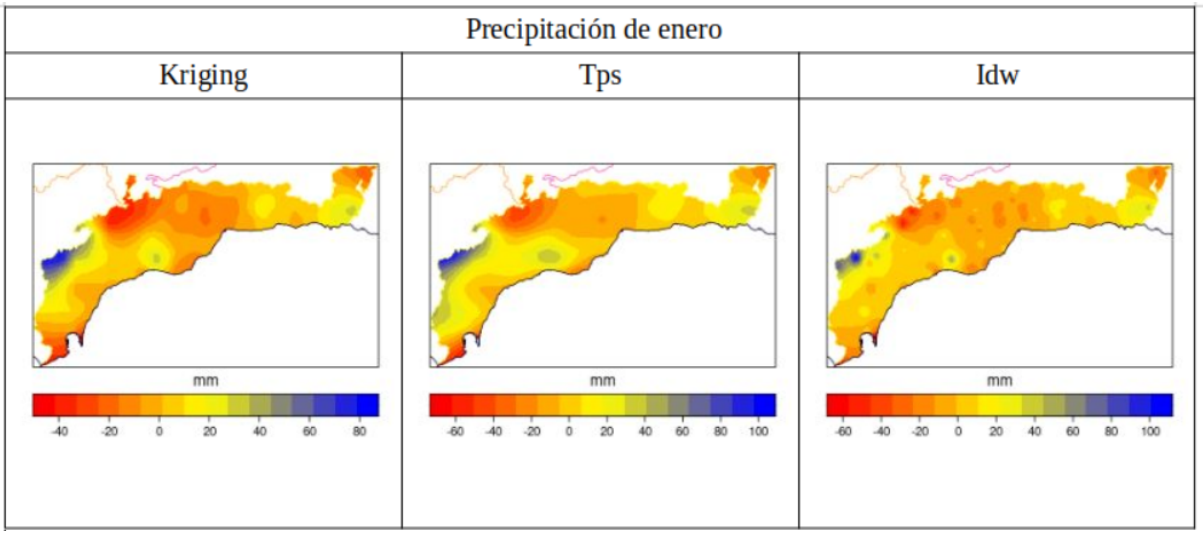

Figura 3.3. Resultados de la interpolación de los residuos de la regresión lineal múltiple de la precipitación de enero en la región de Málaga para las tres metodologías estudiadas.

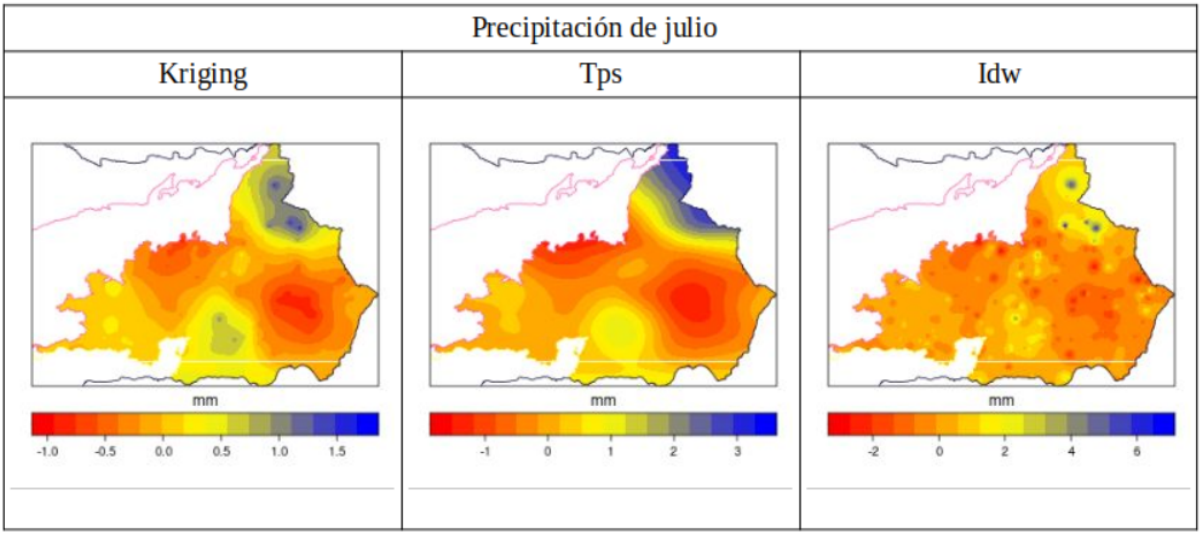

Figura 3.4. Resultados de la interpolación de los residuos de la regresión lineal múltiple de la precipitación de julio en la región Este para las tres metodologías estudiadas.

La figura 3.5 muestra un error grave producido al interpolar con la metodología de Kriging, el tipo de error sobre el que hablábamos al mostrar la figura 2.4: valores relativamente próximos pero muy diferentes producen efectos de interpolación de valores excesivos al intentar que la curva de ajuste pase de forma suave por ambos puntos. Este ejemplo (precipitación de enero en la región del Guadalquivir) aparece en más ocasiones, y aunque se intente controlar evitando tales observatorios implica un control manual de todas las interpolaciones realizadas por Kriging, lo que desaconseja el uso de este método.

Veamos en las siguientes figuras (3.6, 3.7 y 3.8) resultados de la medida del error mediante k-fold de 5 grupos de la interpolación de residuos.

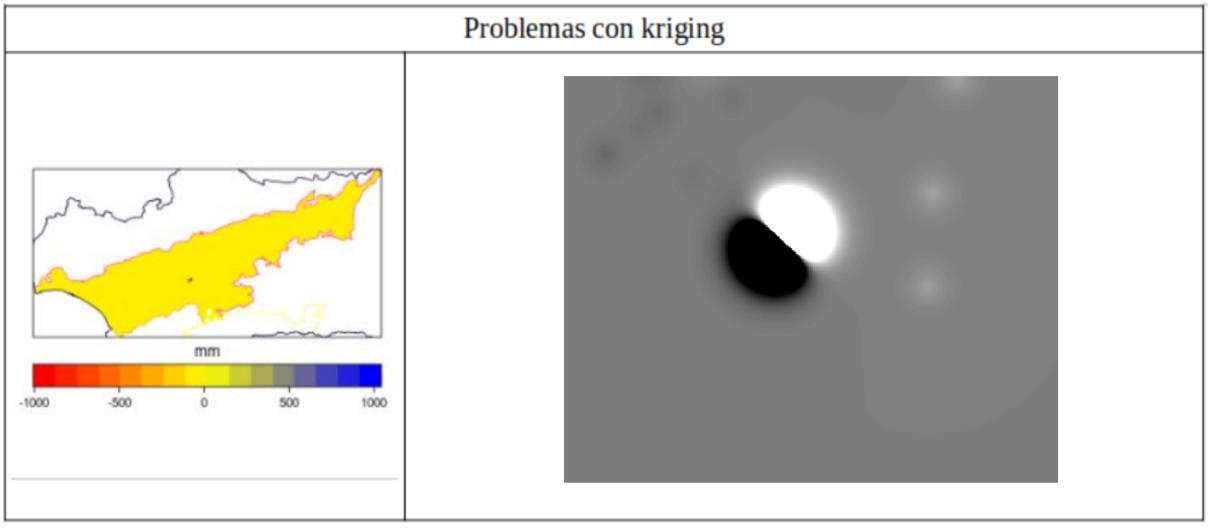

Figura 3.5. Error grave obtenido al interpolar con Kriging en la región del Guadalquivir los residuos de la regresión lineal múltiple de la precipitación de enero.

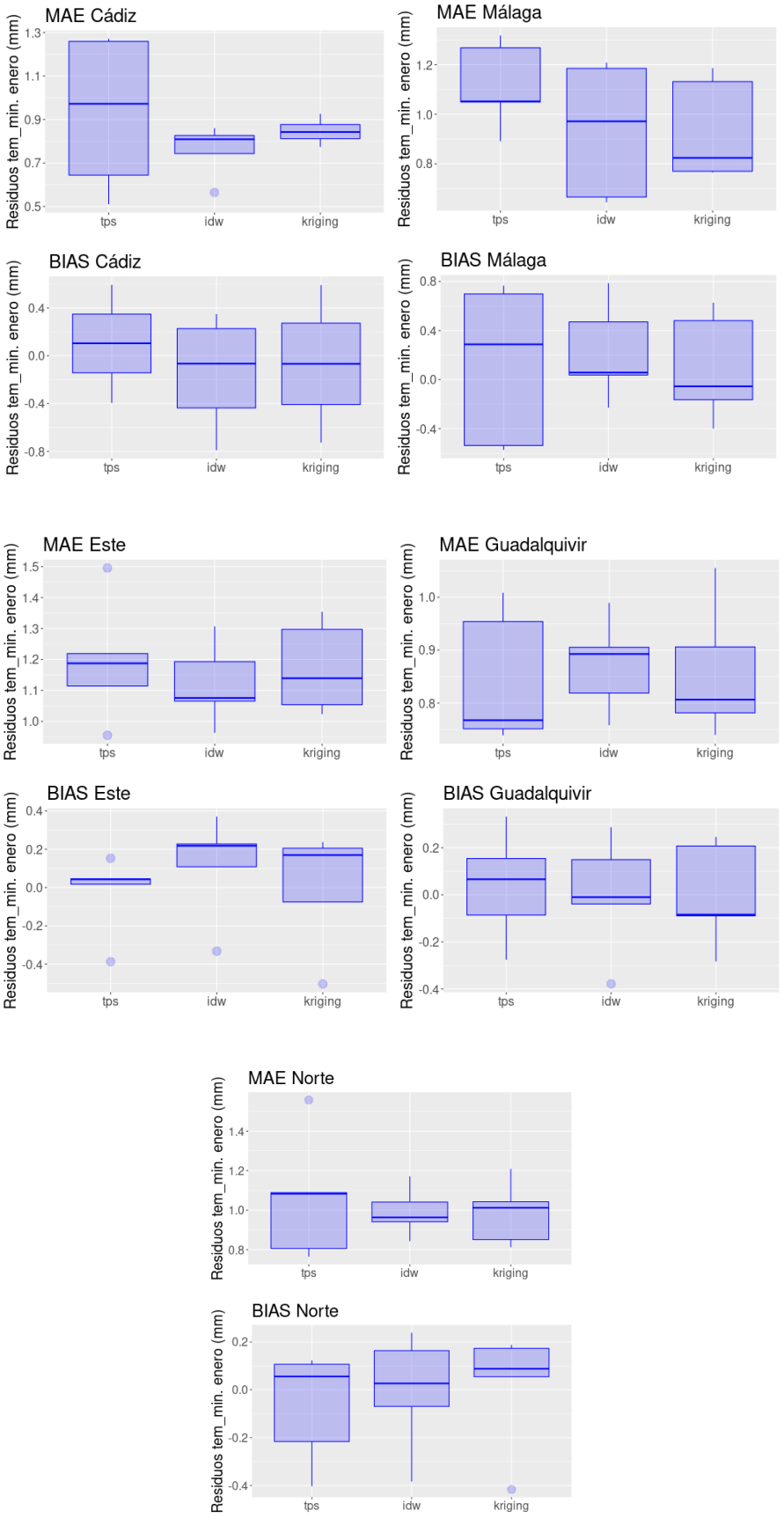

Figura 3.6. Error cometido al interpolar los residuos en los distintos estratos con las técnicas estudiadas para la temperatura media de las mínimas de enero.

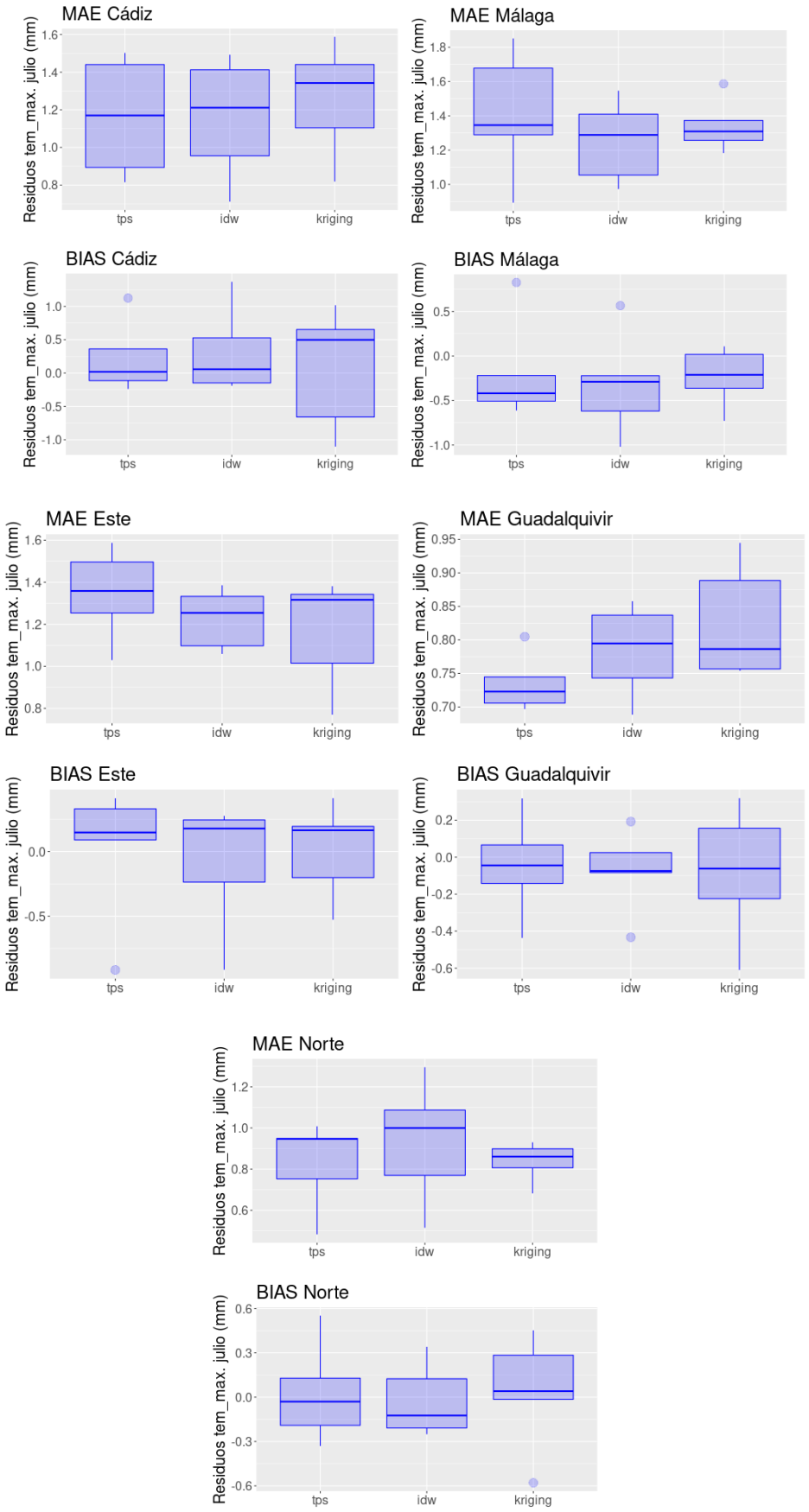

Figura 3.7. Error cometido al interpolar los residuos en los distintos estratos con las técnicas estudiadas para la temperatura media de las máximas de julio.

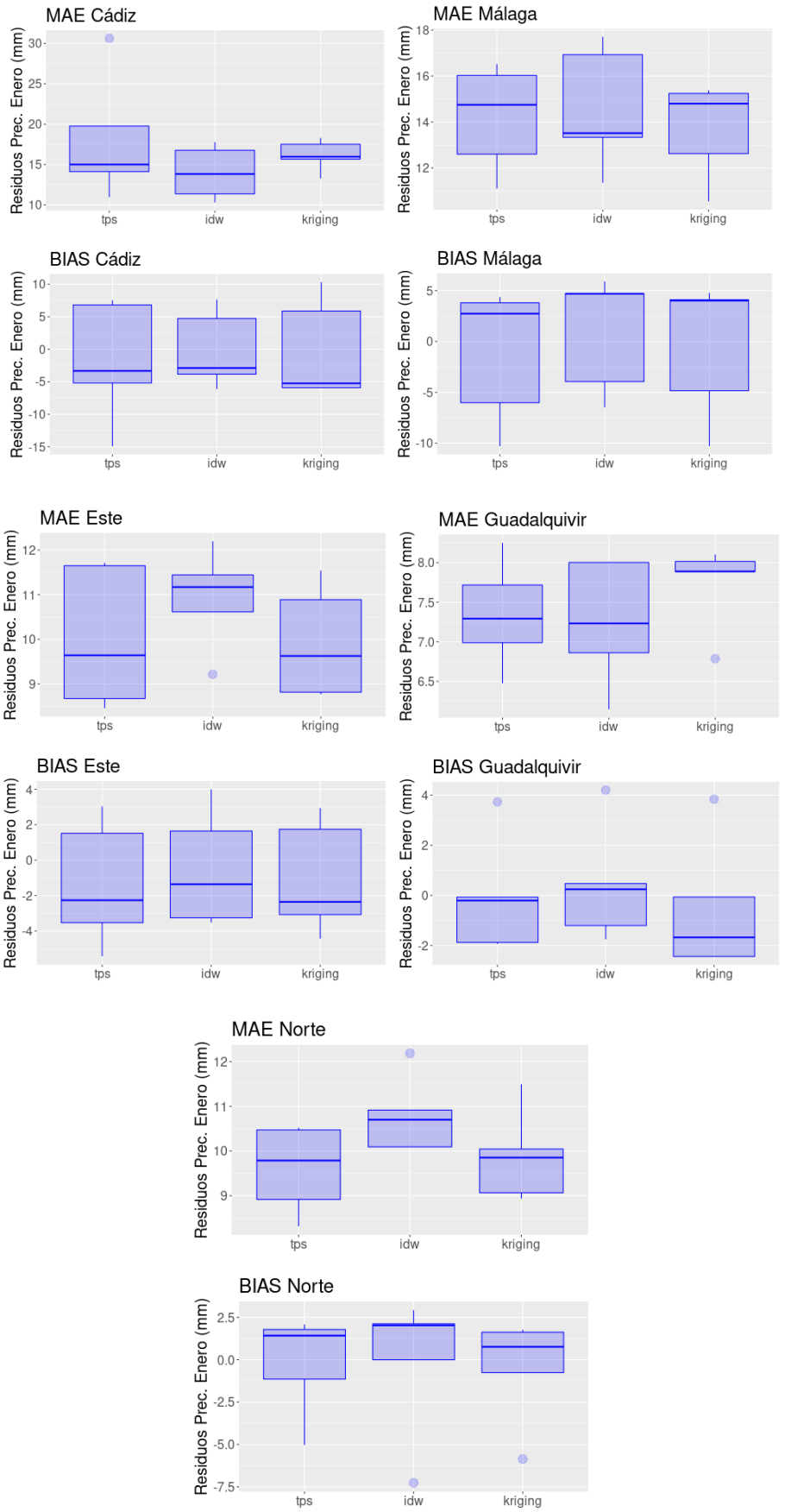

Figura 3.8. Error cometido al interpolar los residuos en los distintos estratos con las técnicas estudiadas para la precipitación acumulada de enero.

Como se puede apreciar en las anteriores figuras, ningún método de interpolación de residuos puede darse por ganador por MAE y/o BIAS simultáneamente en todos los estratos y para todas las variables, sino que dependiendo de unos y otros se puede dar como ganador a un método o a otro. Es por eso por lo que la recomendación de su uso pasa por sus características como método, es decir, intentar evitar el uso tanto de IDW como de Kriging debido a la aparición de artefactos estadísticos (*dianas* y valores extremos falsos) y recomendar TPS dada su capacidad de proporcionar superficies de variación suave, homogénea y continua.

Como apoyo de lo anterior, veamos algunos ejemplos del resultado final de la interpolación de los residuos más los valores de la regresión lineal múltiple, es decir, el verdadero resultado final de la interpolación de las variables.

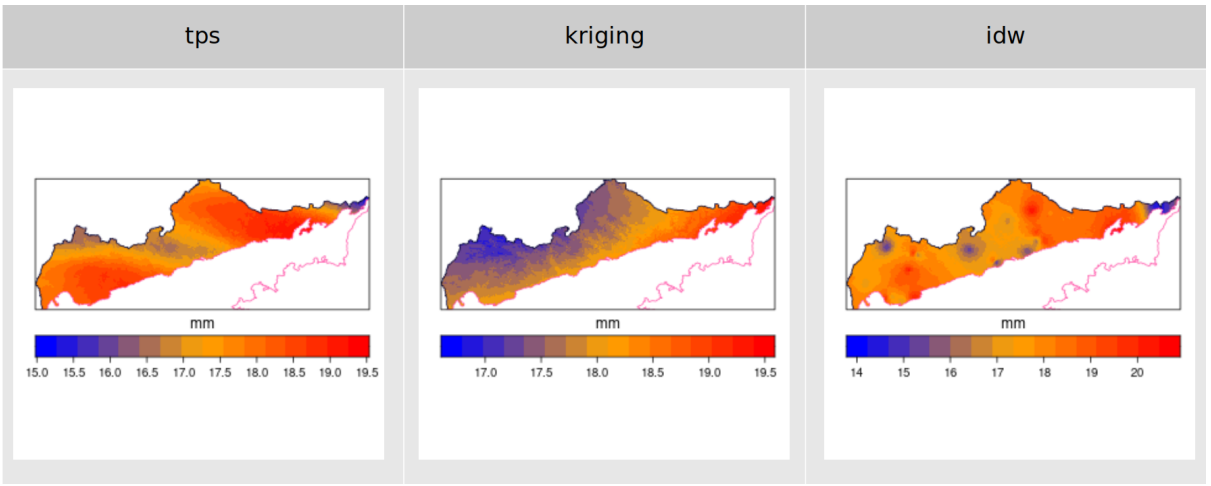

Figura 3.9. Interpolación de la temperatura promedio de las mínimas del mes de julio en el estrato Norte según las diferentes metodologías de interpolación de residuos.

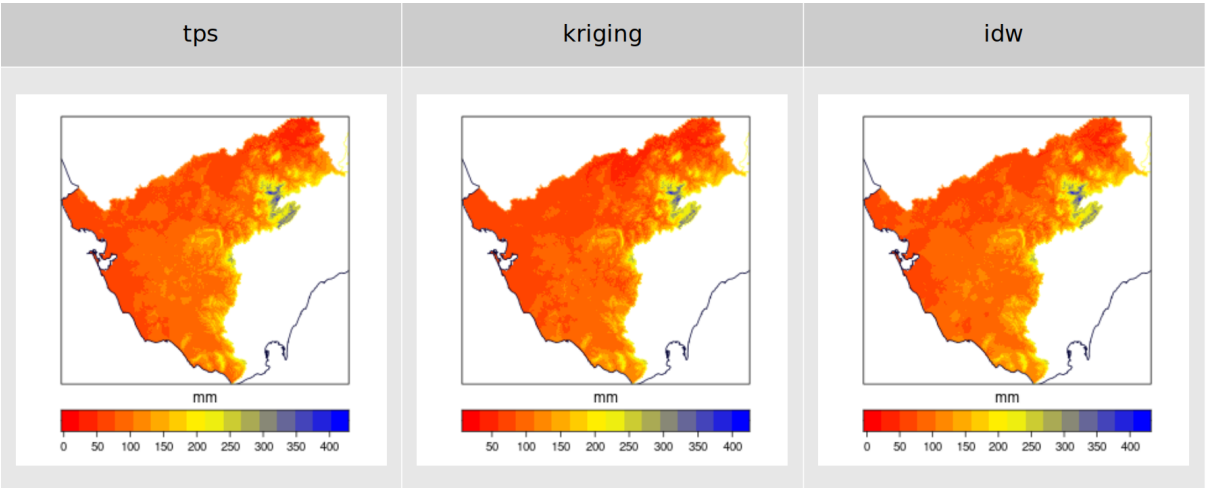

Figura 3.10. Interpolación de la precipitación acumulada de enero en el estrato de Cádiz según las diferentes metodologías de interpolación de residuos.

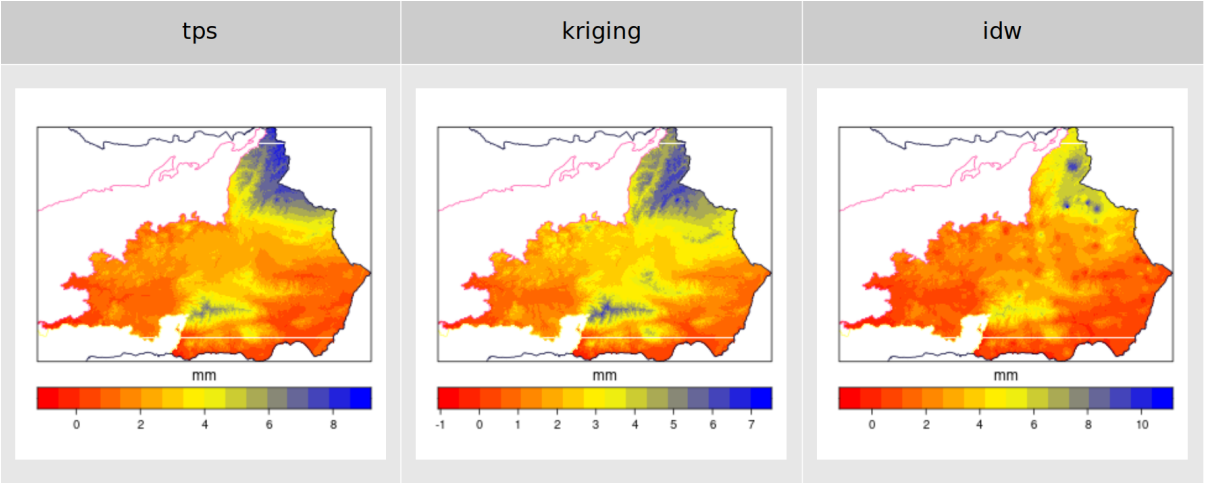

Figura 3.11. Interpolación de la precipitación acumulada de julio en el estrato Este según las diferentes metodologías de interpolación de residuos.

Las figuras 3.9, 3.10 y 3.11 muestran la verdadera interpolación final de las variables (sumando el efecto de la interpolación de los residuos con el valor de la regresión lineal múltiple) en algunos ejemplos. Nótese cómo en la precipitación de enero de la región de Cádiz (figura 3.10) apenas se observan diferencias significativas a simple vista, lo que se debe a los elevados valores obtenidos en la región de Grazalema (la región de Andalucía donde más llueve en enero, a mucha distancia del resto de Andalucía, y por lo tanto la más difícil de estimar), que desdibujan el resto de los valores obtenidos.

Sin embargo, la temperatura promedio de las mínimas del mes de julio en el estrato Norte (figura 3.9) y la precipitación de julio en el estrato Este (figura 3.11) muestran claramente el efecto final en las interpolaciones de las *dianas* producidas por IDW y su menor reconocimiento de la altura como covariable. Nótese que en el ejemplo de la figura 3.9 y a pesar de que debemos recordar que Kriging puede llegar a presentar problemas graves es Kriging como interpolador de residuos el que produce un resultado más realista (suave, homogéneo y gradual) y no TPS, lo que vuelve a incidir en que según el estrato y la variable podríamos llegar a tener comportamientos diferentes.

Por lo anterior, y como conclusión, si vamos a utilizar la interpolación de residuos producida por la regresión lineal múltiple, debemos usar o bien TPS o bien Kriging, pero estamos obligados a un trabajo manual de supervisión para detectar posibles fallos. Nótese que usar diferentes metodologías según el estrato implica que, para una misma variable, no somos capaces de encontrar una forma homogénea de interpolarla, es decir, no somos capaces de encontrar un único método que permita reconocer la naturaleza de su distribución, lo cual es, en sí mismo, malo.

4. INTERPOLACIÓN A TODA ANDALUCÍA

Hemos visto cuál es la mejor interpolación que podemos realizar utilizando la regresión lineal por estratos e interpolando después los residuos resultantes por diferentes métodos de interpolación. Pero ¿y si esas técnicas que hemos usado exclusivamente para interpolar residuos las utilizáramos *directamente* para interpolar los valores de la variable sobre todo el territorio de Andalucía, sin tener en cuenta los estratos? La idea es cuanto menos interesante ya que podríamos

1. utilizar técnicas avanzadas de interpolación (TPS y Kriging, en esta prueba) a las que se puede añadir directamente el uso de una covariable como la altura (algo que no hemos tenido en cuenta al usar esas técnicas en las pruebas de regresión lineal porque es en la propia regresión lineal donde hemos incluido la altura como posible covariable),
2. utilizar observatorios de todo el territorio, lo que puede aportar más información a los modelos (mayor variabilidad) y enriquecer las interpolaciones, y
3. evitar las líneas de unión entre los distintos estratos, algo que, aunque se ha definido de forma precisa para que los cambios se realicen de forma contínua y suave (usando funciones de fusión en integración en los límites de contacto entre superficies, para que de esta manera los valores de las superficies próximos al borde de los estratos sean transformados por una función de fusión del tipo sigmoide), puede provocar inhomogeneidades generales entre las áreas de los distintos estratos.

Así, las siguientes cuatro figuras nos muestran los mapas generados al interpolar la temperatura media de las máximas del mes de enero utilizando TPS (figura 4.1) y Kriging (4.2) como técnicas de interpolación y la precipitación media de enero utilizando TPS (4.3) y Kriging (4.4). Recordemos que estamos usando todos los observatorios disponibles en esos periodos y que hemos introducido la altura como covariable en los propios métodos de interpolación.

A simple vista se puede observar que la interpolación de la temperatura a todo el territorio de Andalucía es similar usando TPS o Kriging, pero la interpolación de la precipitación no. En precipitación TPS muestra mucho más nítidamente las zonas geográficas principales (la Sierra de Grazalema, por supuesto, pero también, por ejemplo, Sierra Nevada, Cazorla o la Sierra de Aracena) que aparecen mucho más desdibujadas en la interpolación realizada por Kriging, que muestra además en todo el territorio efectos de concentración (*dianas* esparcidas por todo el territorio) muy probablemente debidos a la interpolación de observatorios concretos, algo que, como ya vimos, puede llegar a perjudicar especialmente a este método. Por lo tanto, y exclusivamente a simple vista, TPS sería el mejor método de interpolación para todo el territorio ya que su coherencia espacial es mucho mayor.

Pero lo que necesitamos es medir el error cometido por estos dos métodos y compararlo con el cometido con regresión lineal múltiple con interpolación de residuos. Y eso, mediante el uso de boxplots, es lo que nos van a mostrar las figuras 4.5 (para la temperatura) y 4.6 (para la precipitación).

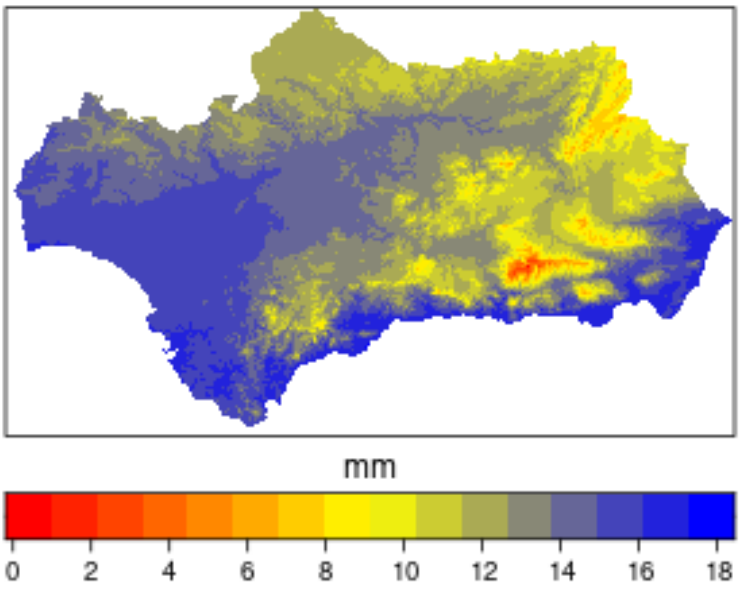

Figura 4.1 Interpolación para todo el territorio de Andalucía de la temperatura media de las máximas de enero, realizada con la técnica de TPS.

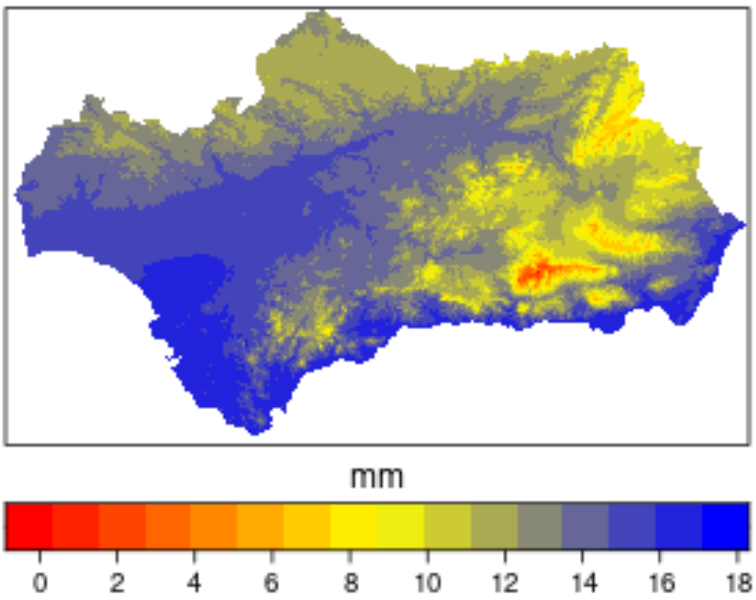

Figura 4.2 Interpolación para todo el territorio de Andalucía de la temperatura media de las máximas de enero, realizada con la técnica de Kriging.

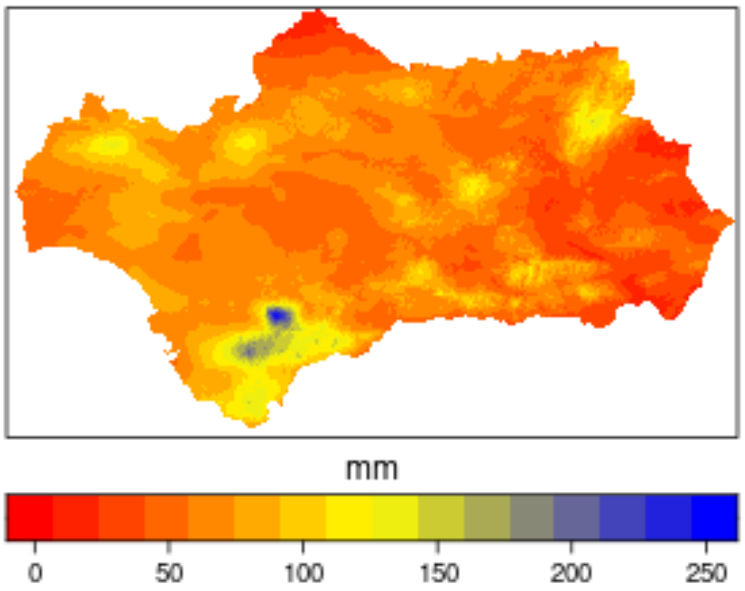

Figura 4.3 Interpolación para todo el territorio de Andalucía de la precipitación media de enero, realizada con la técnica de TPS.

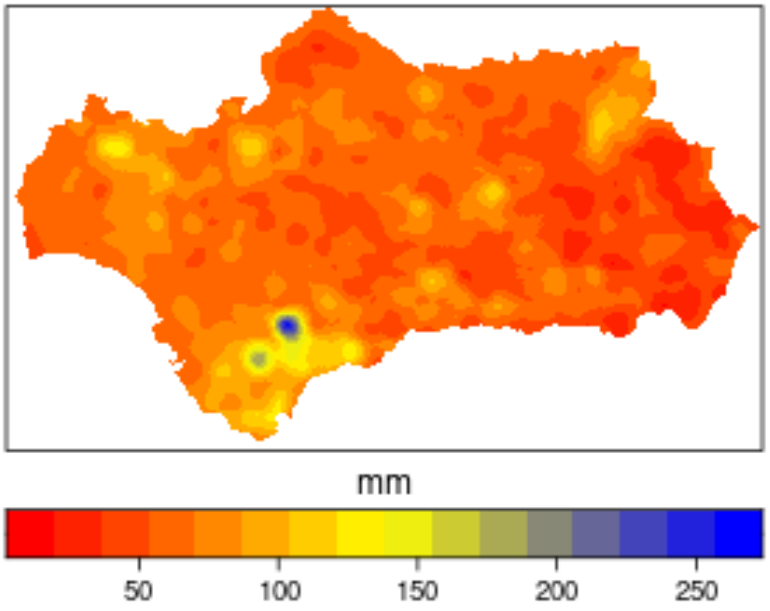

Figura 4.4 Interpolación para todo el territorio de Andalucía de la precipitación media de enero, realizada con la técnica de Kriging.

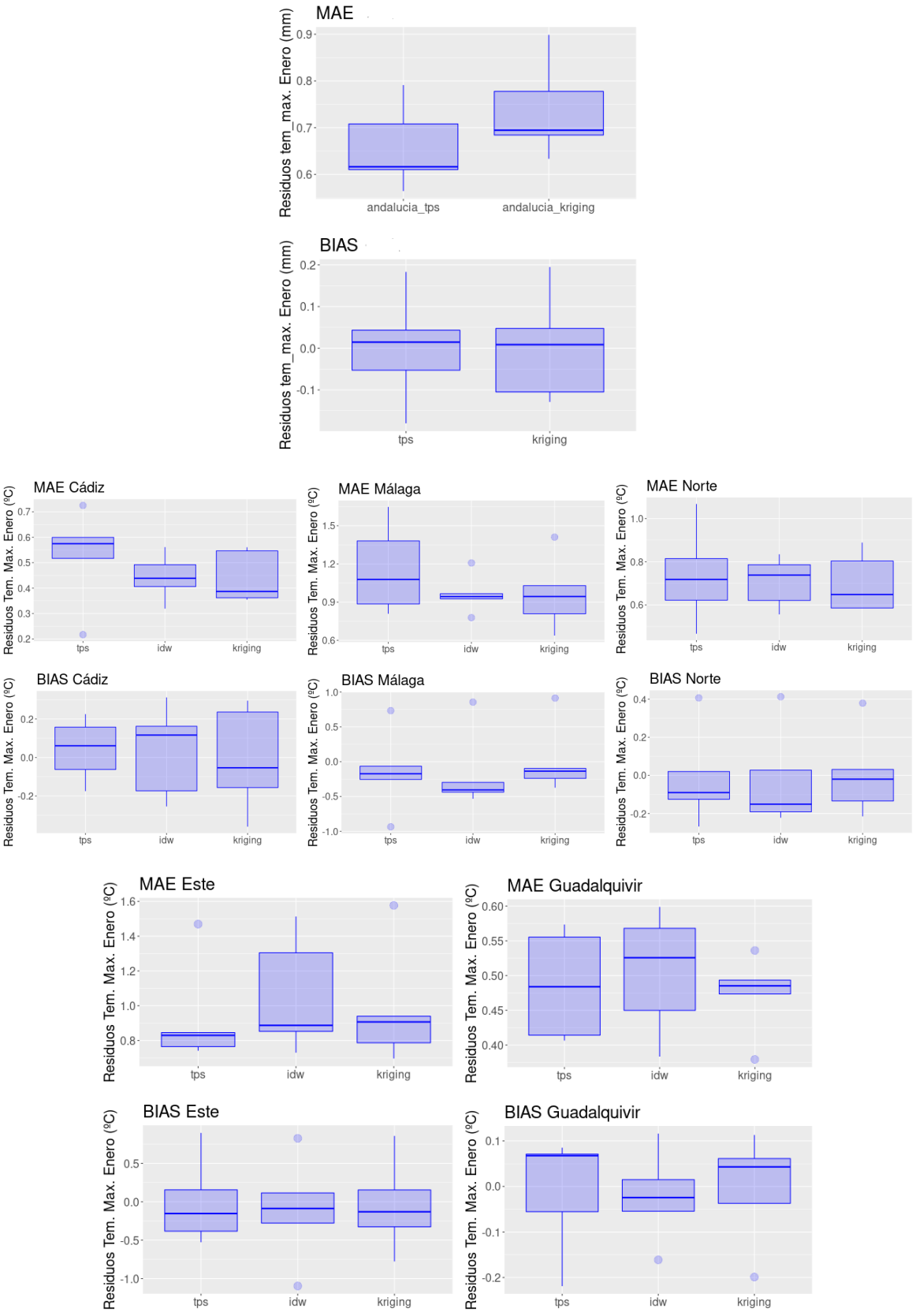

Figura 4.5 Comparativa de los errores de interpolación (MAE y BIAS) de la temperatura media de las máximas de enero para interpolaciones a toda Andalucía (boxplots superiores) y con los métodos de regresión lineal múltiple con interpolación de residuos por estrato.

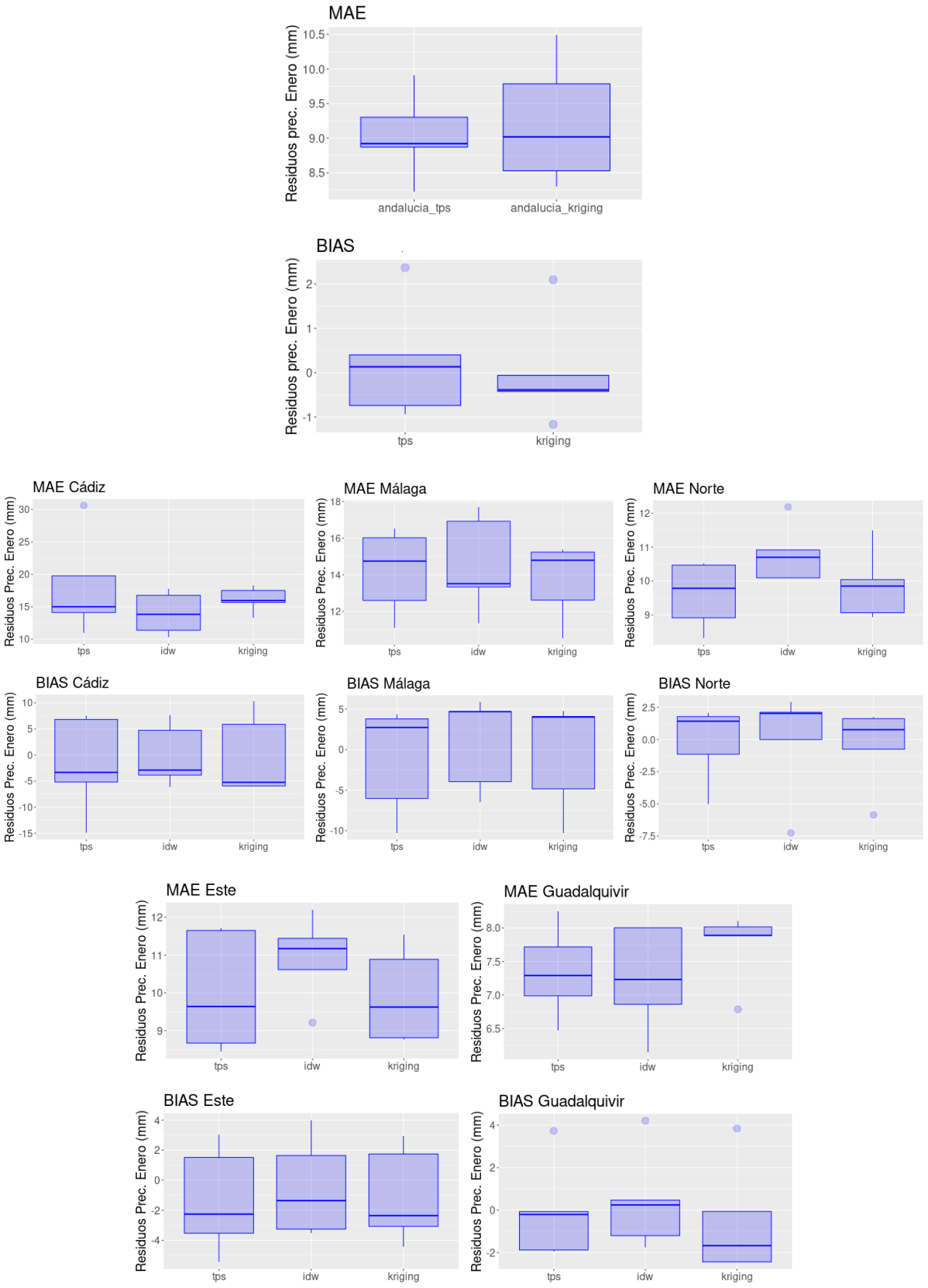

Figura 4.6 Comparativa de los errores de interpolación (MAE y BIAS) de la precipitación media de enero para interpolaciones a toda Andalucía (boxplots superiores) y con los métodos de regresión lineal múltiple con interpolación de residuos por estrato.

Si nos fijamos en la temperatura (figura 4.5), los boxplots superiores muestran que la interpolación a todo el territorio de TPS es sensiblemente mejor que la de Kriging atendiendo al MAE y similar atendiendo al BIAS (aunque la dispersión del BIAS es algo mayor en el Kriging), lo que nos permite decir que la interpolación de temperatura para toda Andalucía es mejor usando TPS - de entre los dos métodos. Pero ¿y en comparación con los métodos de regresión lineal múltiple con interpolación de residuos? Como podemos ver en los boxplots de MAE y BIAS de la regresión lineal múltiple seguida de la interpolación de residuos de los cinco estratos (Cádiz, Málaga, Norte, Este y Guadalquivir) desglosando con el tipo de interpolación de residuo, sólo en las regiones de Cádiz y Guadalquivir encontramos un MAE algo inferior al obtenido para toda Andalucía por TPS, y en ninguna de ellas el BIAS se aproxima tanto a 0 como en las interpolaciones generales de toda Andalucía.

En cuanto a la precipitación (figura 4.6), los boxplots superiores muestran que la interpolación a todo el territorio de TPS posee un error promedio similar a la de Kriging atendiendo al MAE (aunque Kriging muestra una dispersión del error mayor) y también similar atendiendo al BIAS (aunque la dispersión del BIAS es algo mayor en TPS). Y en comparación con los métodos de regresión lineal múltiple con interpolación de residuos en los diferentes estratos, y atendiendo a los boxplots de MAE y BIAS, podemos ver cómo el MAE de la interpolación a toda Andalucía es mejor que el obtenido en todos los estratos excepto en el Guadalquivir, y el BIAS es también mucho mejor (tiende a 0 con mucha mayor nitidez) excepto de nuevo en el Guadalquivir (pero muy levemente).

De todo lo anterior podemos deducir que

1. Para interpolar a toda Andalucía, es mejor TPS: es mejor (tiene menos error) en temperatura y en precipitación produce distribuciones espaciales más coherentes que kriging (con el que tiene un error similar).
2. Es mejor interpolar a toda Andalucía con TPS que con una regresión lineal múltiple con interpolación de residuos para cada estrato: los errores son, en general, menores, tanto de MAE como de BIAS.

5. CONCLUSIONES

De todos los estudios realizados podemos extraer una serie de conclusiones claras:

1. No siempre merece la pena la realización de una regresión lineal múltiple para estudiar una cierta variable climática en el área (o áreas, si hablamos de estratos) de estudio; en el caso de la precipitación, no es en absoluto aconsejable.
2. La única covariable a emplear en todos los casos (todas las variables, todas las áreas geográficas) es la elevación; los resultados sobre las distancias al Atlántico y al Mediterráneo no son concluyentes (o, incluso, no son significativos).
3. La interpolación de residuos mediante IDW muestra la aparición de *dianas* en los entornos de muchos observatorios (un efecto propio de IDW debido a su dependencia de la distancia al observatorio), con lo que su interpolación no es en absoluto suave.
4. La interpolación de residuos mediante Kriging produce a veces *artefactos estadísticos* con errores notables debidos a la proximidad de observatorios con valores muy diferentes.
5. Interpolar una variable con el método de regresión lineal múltiple con interpolación de residuos por estratos obliga a revisiones *a posteriori* de los resultados obtenidos dados los problemas que presentan, según la variable y el estrato, las interpolaciones finales de las variables.
6. Lo anterior implica que el método de regresión lineal múltiple con interpolación de residuos por estratos no es capaz de encontrar una estrategia común de interpolación por variable, es decir, no es capaz de reconocer la naturaleza de la distribución geográfica de las variables climáticas.
7. La interpolación a toda Andalucía (sin estratos) con métodos geoestadísticos directos (TPSy Kriging, sin ningún tipo de regresión lineal múltiple) produce resultados homogéneos y graduales y con errores menores (o, en todo caso, iguales) a los producidos por la regresión lineal.

Todo lo anterior nos conduce a dos conclusiones claras sobre el método de interpolación a usar en este proyecto:

1. No utilizar la regresión lineal múltiple por estratos con interpolación de residuos y
2. Utilizar la interpolación a todo el territorio de Andalucía con la técnica de TPS con la altura como covariable.