

REJILLA DIARIA PARA LA TEMPERATURA MÁXIMA Y MÍNIMA A PARTIR DE DATOS DEL BANCO NACIONAL DE DATOS CLIMATOLÓGICOS EN AEMET

Elías Criado Pinto.

AEMET, Leonardo Prieto Castro 8, 28071 Madrid, ecriadop@aemet.es

Resumen

El establecimiento de rejillas de temperatura diaria máxima, mínima y media es una necesidad demandada en distintos sectores. En AEMET se han realizado estudios precedentes que consideran la temperatura máxima, mínima y media diaria” Mapas de Riesgo: Heladas y horas frío en la España peninsular. (Periodo 2002-2012)”.

El presente estudio pretende determinar un método de interpolación recomendable para las rejillas diarias de temperatura máxima y mínima, en las regiones espaciales de: Península, Baleares, Ceuta y Melilla, por un lado; y de Canarias por otro.

El método de interpolación recomendado en este estudio tras comparar con otros, es el Krigeado Universal, ya que presenta buenas características respecto a la validación, así como de representación de fenómenos locales.

Suponiendo además que la rejilla sería generada en un intervalo de tiempo lo más cercano al día de medición de la variable (entre 1 y 3 días más tarde), cabe considerar posibles fallas en la disponibilidad de datos de la red de observación. Propone por ello optimizar el parámetro asociado “radio de búsqueda” (search-radius) dentro del Krigeado Universal, con la intención de amortiguar este efecto y así obtener una rejilla lo mejor posible dentro de estas limitaciones.

En este sentido se realizan simulaciones de no disponibilidad de datos moderada o severa en las regiones espaciales ya mencionadas, recomendando finalmente un valor del radio de búsqueda de 400 km.

Adicionalmente se expone una herramienta para la depuración manual de las observaciones asociadas a la rejilla generada, para el posterior establecimiento de la versión final de la misma.

Para el desarrollo del presente estudio se utilizaron las herramientas SAGA GIS y R.

Introducción

La obtención de una rejilla diaria, frente a una mensual o anual, tiene la dificultad añadida de ser más sensible a datos singulares, ya bien sea este considerado erróneo, dudoso, o bueno asociado a un fenómeno local dentro del proceso de depuración. Los fenómenos locales juegan por tanto un papel destacable, y por ello necesitamos disponer de métodos de interpolación que dentro de lo posible genere una rejilla con una representación adecuada de los mismos.

Pretendemos así establecer un método de interpolación adecuado en términos de estadísticos de validación, y de representación de los fenómenos locales. Se realizará

también una comparación visual de las rejillas generadas con ellos, a fin de identificar diferencias, y determinar cuál de ellos es el que más nos interesa.

La depuración manual de las observaciones asociadas a una rejilla generada por interpolación, necesita de una herramienta que con intervención humana, pueda contrastar valores observados entre estaciones cercanas aparte de otras características que mencionaremos más adelante. Ha de ser por tanto una herramienta flexible.

Datos utilizados y zona de estudio

Dentro del contexto de la determinación del mejor método de interpolación, hemos tomado datos de temperatura máxima y mínima del BNDC (Banco Nacional de Datos Climatológicos) que se utilizaron en el anterior estudio “ Mapas de Riesgo: Heladas y horas frío en la España peninsular. (Periodo 2002-2012).”, en concreto los del año 2003, a fin de comparar los distintos métodos de interpolación de la Península y Baleares, y adicionalmente se han tomado otros similares para Canarias. Por otra parte también está la conveniencia de tratar con datos recabados en un intervalo de tiempo cercano a la fecha de estudio. Es decir, que hayan pasado entre uno y tres días de la fecha en que se midieron, con objeto de estudiar la generación de una rejilla en tiempo más cercano al de medición. Se utilizaron datos de estas características en fechas concretas del año 2017, como veremos más adelante.

Las zonas de estudio fueron por un lado, Península, Baleares, Ceuta y Melilla, (véase como ejemplo la fig. 1, del 12 de enero de 2003 con 1269 estaciones), y por otro Canarias. (Véase como ejemplo la fig. 2, del 29 de diciembre de 2003 con 60 estaciones).

En cuanto a la representatividad de la red de estaciones, se puede ver que se distribuyen irregularmente, tendiendo a tener menor densidad en zonas poco pobladas y de mayor altitud. En cuanto a las zonas de montaña se observa una tendencia a tener más estaciones en los valles, y menos en zonas de elevada altitud. Una consecuencia negativa posible de la escasez de datos en regiones montañosas, es que los métodos de interpolación pueden generar extrapolaciones en lugar de las interpolaciones esperadas. (O. Tveito. 2010).

Métodos de interpolación.

Para la interpolación de la temperatura diaria, vamos a tratar de estudiar distintos métodos de interpolación.

En referencia a los métodos de krigado para la interpolación de la temperatura diaria y horaria, cabe destacar el artículo de (Jabot. E 2012). Este muestra cómo los métodos de Regresión con Interpolación de los Residuos por Krigado Ordinario (RKO) y Krigado Universal (KU) mejoran considerablemente al Krigado Ordinario (KO) al incluir variables externas (en concreto la altitud), y cómo a su vez, el KU (también denominado KED), mejora levemente al RKO. El KU tiene además la capacidad de representar fenómenos locales, tales como inversiones térmicas (A.Chazarra, 2014).

A parte de los métodos ya mencionamos incluimos otros métodos adicionales a fin de extender el estudio: Análisis de la Regresión (REG), Regresión con Interpolación de los Residuos por Inverso de la Distancia (RIDW), y Regresión con Interpolación de los Residuos por Multilevel Beta Spline (RMBS).

Los métodos con Regresión y el KU, utilizan variables externas para obtener la predicción, mientras que el KO no. Las variables externas que se tomaron en consideración para la zona de Península, Baleares, Ceuta y Melilla, fueron la altitud, la distancia al mar y la latitud para los métodos con regresión previa, y la altitud y la distancia al mar para el KU. Para Canarias, únicamente se considera la altitud. Todas estas variables están en formato de rejilla con resolución de 1 x 1 km, y por tanto la interpolación se obtendrá con la misma resolución. Esta resolución puede ser útil para algunos estudios que requieran mayor resolución, pero no aporta en principio mayor información dada la densidad de datos de la que se parte (hay estudios que apuntan a una resolución de 5 x 5 km).

Todos estos métodos, se contrastaron por un lado, a partir de la comparación de estadísticos, y por otro, comparando visualmente los mapas diarios obtenidos en todo un año.

Los estadísticos estudiados son el coeficiente de correlación (R), el BIAS, el error absoluto medio (MAE), y el error cuadrático medio (RMSE). Se calculan separando aleatoriamente los datos diarios en un 75% y 25%, para después interpolar con cada método sobre el 75%, y verificar contra el 25% restante.

A partir de las observaciones de los 365 días del año 2003, se calculan las rejillas con los métodos ya mencionados, así como las tablas estadísticas promediadas anualmente para Península y Baleares, y Canarias.

Resultados

Tablas de validación: El proceso de validación se realizó con las observaciones de los 365 días del año 2003.

A continuación, se muestran los resultados estadísticos promediados en formato de tabla para Península y Baleares (Tabla 1 y 2), y para Canarias (Tabla 3 y 4):

	KO	REG	RIDW	RMBS	RKO	KU
R	0,79	0,57	0,84	0,82	0,83	0,83
BIAS	-0,02	-0,04	-0,01	0,00	-0,01	-0,05
MAE	1,68	2,46	1,47	1,54	1,51	1,52
RMSE	2,32	3,11	2,03	2,15	2,10	2,11

Tabla. 1.- Tabla de estadísticos para la variable temperatura máxima promediada en 365 días del año 2003 en la Península y Baleares.

	KO	REG	RIDW	RMBS	RKO	KU
R	0,79	0,63	0,81	0,79	0,79	0,80
BIAS	-0,02	-0,04	-0,02	-0,02	-0,01	-0,02
MAE	1,67	2,23	1,56	1,69	1,66	1,64
RMSE	2,22	2,79	2,06	2,25	2,21	2,19

Tabla. 2.- Tabla de estadísticos para la variable temperatura mínima promediada en 365 días del año 2003 en la Península y Baleares.

	KO	REG	RIDW	RMBS	RKO	KU
R	0,34	0,55	0,54	0,49	0,50	0,53
BIAS	-0,15	0,17	0,10	0,11	0,11	0,17
MAE	2,52	1,76	1,91	2,16	2,12	2,12
RMSE	3,28	2,29	2,50	2,83	2,76	2,76

Tabla. 3.- Tabla de estadísticos para la variable temperatura máxima promediada en 365 días del año 2003 en Canarias.

	KO	REG	RIDW	RMBS	RKO	KU
R	0,36	0,72	0,64	0,58	0,60	0,63
BIAS	-0,24	0,13	0,07	0,06	0,09	0,12
MAE	2,75	1,74	1,93	2,14	2,09	2,07
RMSE	3,52	2,25	2,53	2,82	2,76	2,74

Tabla. 4.- Tabla de estadísticos para la variable temperatura mínima promediada en 365 días del año 2003 en Canarias.

Centrándonos en los parámetros R, MAE y RMSE, vemos que en el caso de Península y Baleares, para la temperatura máxima presentan mejores resultados los métodos con interpolación de residuos y el KU, frente a los otros, siendo mejor el RIDW, seguido del RKO y KU (con escasa diferencia entre ellos). Para la temperatura mínima destacan en orden decreciente, RIDW, KU, RKO, KO y RMBS, y se puede diferenciar en este caso el KU frente al RKO de forma más significativa que en el caso de la temperatura máxima.

En Canarias, se aprecia para ambas variables peores valores del KO y una mejora sustancial para la REG, siendo la que mejores resultados obtiene, y de forma más marcada en la temperatura mínima. Los siguientes métodos que destacan son el RIDW y el KU.

En Canarias, cabe comentar que el método de validación mediante “muestra separada” (“split-sample”) que utilizamos para la obtención de estadísticos, nos permitió comparar los distintos métodos de interpolación, no obstante, en ciertos días, la muestra de datos para aplicar esta validación, no era suficientemente extensa, y daba lugar a valores de los estadísticos dudosos. Además, se puede observar que el coeficiente de correlación para la temperatura mínima, es destacablemente mejor en REG, que en los otros métodos de interpolación. Esto queda corregido al utilizar el método de validación cruzada dejando uno fuera” o “leave one out cross validation” (loocv) ya que trabaja con una muestra más amplia (ver la tabla 5 para la temperatura mínima, y la tabla 6 para la temperatura máxima). En ambas tablas recogemos los resultados del R y RMSE para KO, REG, y KU (destaca como el mejor). Al comparar loocv con el método de muestra separada se observa que: R mejora a salvo para el REG en la temperatura máxima que prácticamente se mantiene igual, así como cierto aumento del RMSE.

	KO	REG	KU
R	0,59	0,70	0,75
RMSE	3,94	2,64	3,09

Tabla. 5.- Tabla de estadísticos para la variable temperatura mínima promediada en 365 días del año 2003 en Canarias con el método de validación “loocv”.

	KO	REG	KU
R	0,60	0,54	0,72
RMSE	3,90	3,04	3,25

Tabla. 6.- Tabla de estadísticos para la variable temperatura máxima promediada en 365 días del año 2003 en Canarias con el método de validación "loocv".

Mapas para la comparación visual: Mostramos mapas de temperatura máxima en fechas en que tras una inspección (de los 365 días del año 2003), se ha observado una mayor diferencia entre los métodos. En base a los resultados estadísticos previos nos centraremos en el RIDW, RMBS, RKO, y KU tanto en Península, Baleares, Ceuta y Melilla, como en Canarias.

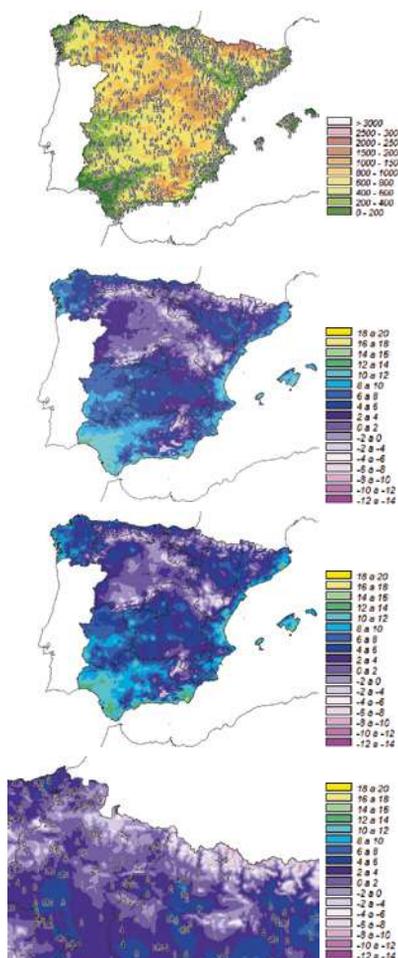


Fig. 1.- Estaciones utilizadas el 12 de enero de 2003 con el mapa de altitudes en la primera imagen. Le siguen en orden de lectura, respectivamente, los mapas correspondientes de temperatura máxima para los métodos RIDW, RMBS, RKO, KU, con detalle de la mitad oeste de Pirineos para el RKO y KU en las últimas dos imágenes.

- Península, Baleares, Ceuta y Melilla: día 12/01/2003, para la temperatura máxima: Caracterizado por altas presiones al norte de la Península. Baja en altura sobre la mitad este peninsular. Nuboso en el litoral cantábrico y norte de Castilla León, y en el extremo sureste de la Península donde dejaron algo precipitación. Heladas generalizadas en la Península.

Los métodos con interpolación de residuos, y el KU, no parecen mostrar diferencias destacables a primera vista (ver fig. 1), pero si nos centramos en la zona de pirineos podemos observar una diferencia clara entre el KU y los demás métodos (RIDW, RMBS y RKO presentan un patrón más parecido entre ellos). En concreto, la zona de temperatura máxima inferior a -2 °C representada en el RKO, corresponde a la zona de alta montaña, y presenta valores mínimos de la temperatura máxima. En esta misma región, el KU presenta en la mayor parte de la misma temperaturas máximas de mayor magnitud. Esto se debe a que en el KU la relación de la variable con las variables externas, no se estima en un proceso de calibración o de regresión previo, sino en el propio krigado, por lo que esta relación puede variar de unos puntos a otros del espacio (Goovaerts, 1997).

En las inversiones de temperatura, permite un gradiente vertical de temperatura distinto en cada punto del espacio, (A. Chazarra, 2014).

En nuestro caso tendremos por tanto un gradiente vertical de la variable estudiada (en este caso la temperatura máxima) distinto en cada punto del espacio para el KU. Por el contrario, los otros métodos estudiados (RKO, RIDW, RMBS) tienen una relación entre la variable y las variables externas (análisis de regresión) que es la misma para todos los puntos.

- Canarias: Día 29/12/2003, para la temperatura máxima:

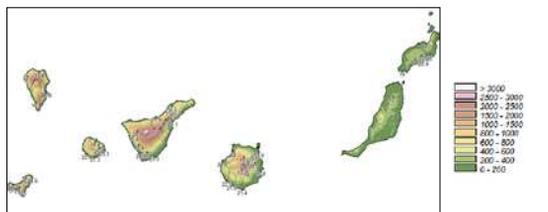
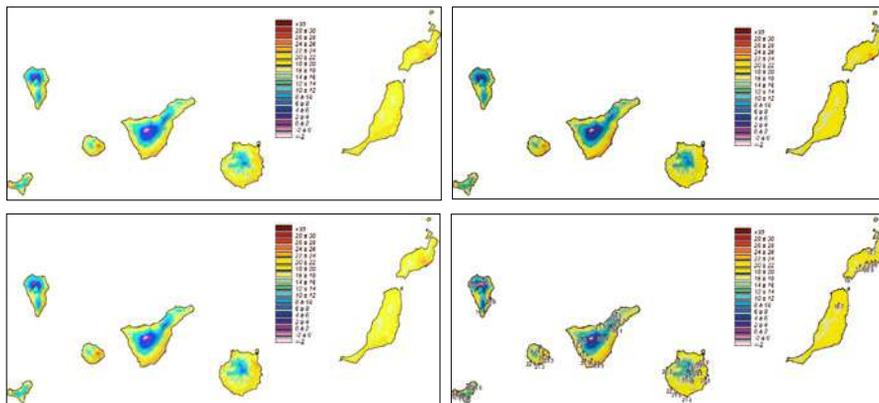


Fig. 2.- Estaciones utilizadas el 29 de diciembre de 2003 con el mapa de altitudes en la primera imagen. Le siguen en orden de lectura, respectivamente, los mapas correspondientes de temperatura máxima para los métodos RIDW, RMBS, RKO, KU.



Día caracterizado por altas presiones. Intervalos nubosos, con nuboso en el norte de las islas. Descenso de temperaturas ligero a moderado.

En términos generales en la fig. 2, se observa a primera vista escasa diferencia entre los métodos con interpolación por residuos, y el KU. En cambio si nos fijamos en el detalle de La Palma (isla del noroeste), se observa que en el KU alcanza valores más bajos de las máximas en las zonas de mayor altitud de la mitad norte por argumentos similares a los ya expuestos en el caso de Península y Baleares.

Notar: no mostramos mapas de temperaturas mínimas, ya que las conclusiones son las mismas.

Conclusiones del estudio de los distintos métodos de interpolación:

1. De los resultados estadísticos, los métodos que incluyen regresión con ajuste por residuos, y el KU presentan resultados buenos y hasta cierto grado comparable dependiendo de la variable y la región de estudio.
2. Se recomienda, no obstante el KU al ser el que presenta mayor capacidad de reproducir fenómenos locales.
3. Cabe mencionar que el KU en zonas con una densidad de estaciones considerablemente baja, o el caso de datos anómalos, puede dar lugar a algunas extrapolaciones exageradas. (Hudson y Wackernagel, 1994), (A. Chazarra, 2014).

Ajuste del parámetro radio de búsqueda en el KU:

Una vez seleccionado el KU como método recomendable, es necesario hacer un estudio del valor más adecuado del parámetro radio de búsqueda (“search radius”), dentro del módulo “Universal Kriging” de SAGA. Este radio delimita las observaciones que pueden entrar a formar parte del cálculo entorno al punto en cuestión.

Se trata de encontrar un valor del parámetro que revierta en la mejor calidad del krigado posible ante eventuales fallas en la recopilación de datos de las estaciones. Por ello se estudian primeramente dos casos para Península, Baleares, Ceuta y Melilla (PBCM), y luego otro para Canarias (CAN). Todos estos con una muestra de observaciones dispuestas inferior a lo habitual.

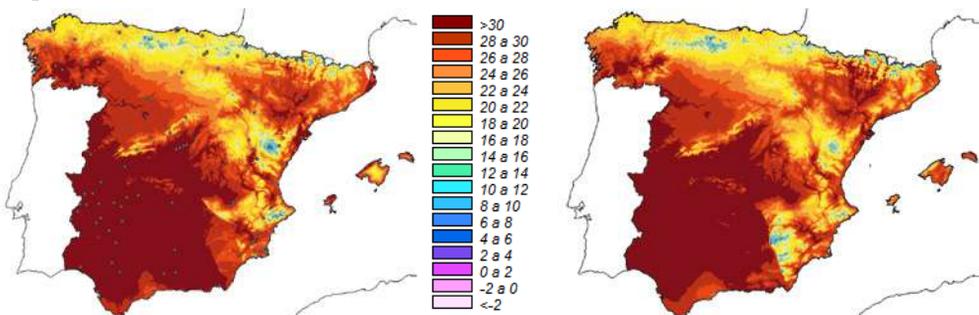


Fig. 3.- Temperatura máxima del 2 de julio de 2017, obtenida por KU con radio de búsqueda de 200 y 400 km en el mapa de la izquierda y derecha respectivamente. Las estaciones se representan como círculos grises en el primero de ellos.

Caso primero de PBCM: día 02/07/2017, con 130 estaciones disponibles. Partimos de un radio de 200 km, y aumentamos a 300, 400 y 450 km. El límite de mejora se obtiene cuando llega a 400 km. En la fig.3 mostramos esta mejora, que se evidencia en la región del sureste peninsular, y en las islas Baleares.

Caso segundo de PBCM: día 01/07/2017. Estudio para la temperatura máxima eliminando estaciones y cambiando el radio. La imagen de la izquierda de la figura 4 es la de referencia (obtenida con todas las estaciones disponibles, 673, y con radio de búsqueda 200 km). La imagen de la derecha se ha obtenido eliminando las estaciones de Baleares, Ceuta y Melilla, con un radio de 400 km. Vemos que esta consigue representar razonablemente bien la región de las islas Baleares

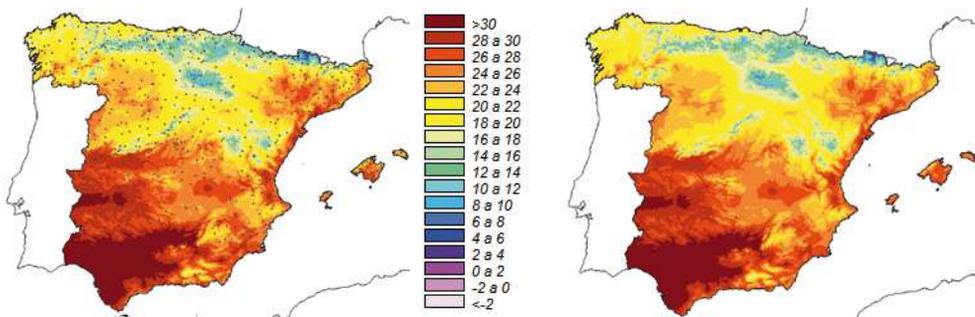


Fig. 4.- Temperatura máxima del 1 de julio de 2017, obtenida por KU. En la imagen de la izquierda con radio de búsqueda de 200 km y todas las estaciones disponibles (se muestran como círculos grises). La imagen de la derecha es la obtenida con radio de 400 km, y sin las estaciones de Baleares, Ceuta y Melilla.

Nota: Se realizaron cálculos para radios intermedios entre los 200 y 400 km que muestran una mejora progresiva de los resultados. Aquí solo mostramos los de 200 y 400 km por simplicidad.

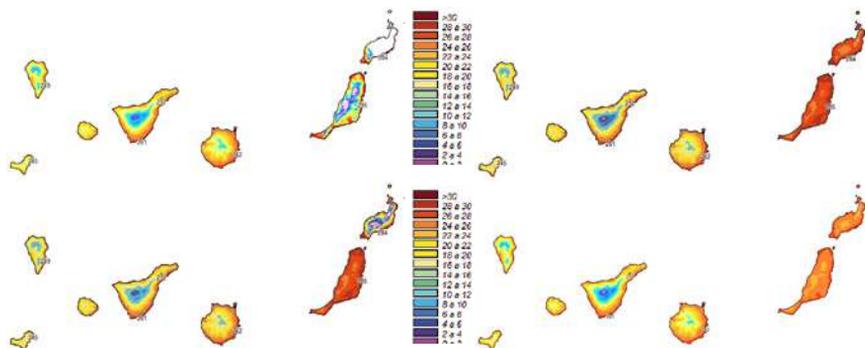


Fig. 5.- Mapas obtenidos por KU para la temperatura máxima del 2 de julio de 2017 en Canarias. En orden de lectura respectivamente tenemos: las tres primeras imágenes con 7 estaciones disponibles para un radio de 200, 250 y 300 km, y la última imagen, con sólo 3 estaciones disponibles; dos en Tenerife y una en Gran Canaria y radio de 300 km.

Caso de CAN: día 02/07/2017, con 7 estaciones disponibles, en las que se ve el efecto de aumentar el radio, de 200 km, a 250, y 300 km. Y también exploramos la falta de datos extrema (sólo 3 estaciones disponibles); suponemos que tenemos datos únicamente en Tenerife (dos estaciones) y Gran Canaria (una estación). Ver fig. 5.

En el caso de 7 estaciones disponibles, se observa que aunque hay datos en Fuerteventura y Lanzarote, no obtenemos para un radio de 200 km, una interpolación adecuada en las mismas. Esta situación mejora al aumentar el radio, pero no es hasta aumentarlo en 300 km, en que consigue una representación adecuada.

En el caso extremo de sólo 3 estaciones disponibles obtenemos una rejilla razonable con 300 km en la zona sin datos. Si la comparamos con la figura anterior, se ve que el patrón es similar, pero muestra una gama de colores de aproximadamente 2 grados inferior. En cualquier caso es un resultado significativamente bueno.

Conclusiones respecto del ajuste del parámetro radio de búsqueda en el KU:

1. Para salvaguardar situaciones extremas en que queden zonas con pocos datos de estaciones, es mejor tener la posibilidad de captar datos adicionales del entorno (aumentando el radio de búsqueda) para mejorar la interpolación.
2. En Península y Baleares, Ceuta y Melilla, como el incremento en el coste computacional al aumentar el radio de 200 km a 400 km es despreciable, y dado que el estudio contempla casos extremos de falta de datos, consideramos que el usar un radio de búsqueda de 400 km puede salvaguardar posibles casos más extremos en ciertas regiones (ejemplo: falta de datos en las islas Baleares y en la costa peninsular cercana). Recomendamos también un radio de 400 km en Canarias por argumentos similares.

Herramienta de visualización y depuración manual de la rejilla

Dada una fecha, dependiendo de los días transcurridos tras la misma, tendremos más o menos datos de estaciones recabadas correspondientes a esa fecha, ya que no todas las observaciones se transmiten al mismo tiempo, y parte de ellas se eliminan, o incorporan de un día para otro tras pasar diversos controles de calidad.

Se plantea por tanto disponer de una herramienta prototipo que permita comparar visualmente las tres versiones de datos disponibles para una fecha concreta, con sus KU asociados, según hayan pasado uno, dos o tres días respectivamente de la misma, para estudiar si hay una mejora o perjuicio significativo entre estas versiones, y seleccionar la que consideremos más adecuada.

Podremos entonces depurar manualmente las observaciones implicadas generando un nuevo KU que contrastaremos visualmente con el previo, para optar por considerarlo como definitivo y generar los productos derivados que establezcamos.

En este sentido apuntar brevemente que dicha herramienta es viable; se ha desarrollado en R aprovechando las virtudes de SAGA.

Productos derivados

La consecución de la rejilla diaria de temperatura máxima y mínima, abre un amplio espectro de productos derivados, demandados en distintos sectores, y más amplio aún

al combinarlo con otras rejillas tales como la de precipitación. Destacamos dentro de los productos derivados que se podrían generar los siguientes: Grados día, número de días de helada, número de días con temperaturas mínimas mayores o iguales que 20 °C, etc.

En la figura 6 se muestran respectivamente en orden de lectura los mapas interpolados con KU de la temperatura mínima, máxima y como producto derivado, la temperatura media del 7 de enero de 2018 en la Península, Baleares, Ceuta y Melilla.

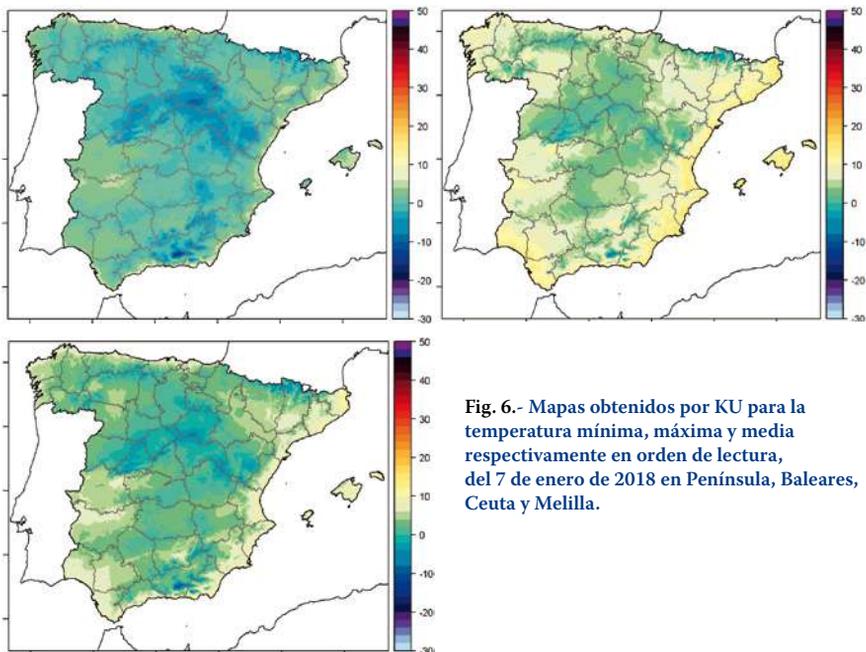


Fig. 6.- Mapas obtenidos por KU para la temperatura mínima, máxima y media respectivamente en orden de lectura, del 7 de enero de 2018 en Península, Baleares, Ceuta y Melilla.

Conclusiones

El KU es un método de interpolación recomendable para generar rejillas diarias de temperatura máxima y mínima en Península, Baleares, Ceuta y Melilla por un lado, y en Canarias por otro, ya que tiene por un lado buen comportamiento respecto a los estadísticos, y es el más indicado de entre los estudiados para representación de fenómenos locales, como son las inversiones de temperatura para la temperatura mínima. Por el contrario puede dar lugar a extrapolaciones exageradas en zonas con escasez de datos.

En el supuesto de querer generar una rejilla diaria en tiempo cercano al del día de medición, es recomendable establecer un valor de 400km del parámetro “radio de búsqueda”, dentro del módulo de KU de SAGA GIS, para solventar de la mejor manera posibles fallas en la disponibilidad de una cantidad significativa de las observaciones diarias implicadas. Además, se recomienda establecer una herramienta de depuración manual

de las observaciones implicadas en la rejilla base diaria con características similares a las ya indicadas en el presente artículo con el objeto de conseguir una rejilla definitiva de mayor calidad, y generar así los productos asociados que se consideren convenientes.

Referencias

- Martínez, L., Moreno, J., Chazarra, A., Gallego, T., Avello, M., Botey, R. (2015): *Mapas de Riesgo: Heladas y Horas Frío en la España Peninsular*, MAPAMA AEMET, 50 pág.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources*. Oxford University Press
- Hudson, G. and Wackernagel, H. (1994). *Mapping temperature using kriging with external drift: Theory and an example from Scotland*. Int. J. Climatol., 14: 77–91. doi: 10.1002/joc.3370140107
- Dodson, R. and Marks, D. (1997). *Daily air temperature interpolated at high spatial resolution over a large mountainous región*. Journal: Climate Research, vol. 8, 1-20.
- Jabot, E., Zin, I., Lebel, T., Gautheron, A. and Obled, C. (2012). Spatial interpolation of sub-daily air temperatures for snow and hydrologic applications in mesoscale Alpine catchments. Hydrol. Process., 26: 2618–2630. doi: 10.1002/hyp.9423
- Tveito, O. (2010). “The developments in Spatialization of Meteorological and Climatological Elements”. *Spatial Interpolation for Climate Data. The Use of GIS in Climatology and Meteorology*. Hartwig Dobesch, Pierre Dumolard Izabela Dyras, 73-86.
- Chazarra, A. (2014). AEMET. *Interpolación espacial de la temperatura mínima diaria mediante Krigeado Universal*, 5 pág.
- Edward H. Isaaks, R. Mohan Srivastava. (1989). *An Introduction to Applied Geostatistics*, Oxford University Press 1989, 561 pág.
- R Core Team, 2013: R. A language and environment for statistical computing. Technical report, R Foundation for Statistical Computing, Vienna, Austria, 2547 pág.