

ANÁLISIS DE DOS MÉTODOS DE INTERPOLACIÓN Y SUS PARÁMETROS, PARA TEMPERATURA Y PRECIPITACIÓN MENSUALES, EN EXTREMADURA

Marcelino Núñez Corchero ⁽¹⁾, Manuel del Puerto Centeno ⁽¹⁾,
Carlos Leal Cidoncha ⁽²⁾, Álvaro Camello Lázaro ⁽²⁾

⁽¹⁾ Delegación AEMET en Extremadura, mnunezc@aemet.es

⁽²⁾ Universidad de Extremadura

Resumen

Con este trabajo se pretende optimizar del método de interpolación para temperaturas medias mensuales y precipitaciones mensuales en la comunidad de Extremadura, y así estimar, lo mejor posible, valores de temperatura y precipitación en puntos sin observación.

Para ello, se han analizado que métodos de interpolación, y que parámetros de estos métodos, influyen significativamente en el resultado final del proceso de interpolación. El objetivo principal de este trabajo es determinar cuales son los mejores parámetros para realizar una interpolación de temperaturas y precipitaciones. Es decir, optimizar la herramienta de interpolación para temperaturas y precipitaciones en Extremadura, teniendo en cuenta su escala y el relieve regional.

Para ello, el estudio se ha centrado en dos de los métodos de interpolación más comunes: “Ordinary Kriging” (OK) y “Universal Kriging” (UK), implementados en el software SAGA-GIS, que se ha utilizado conjuntamente con el software R para la automatización del proceso y el análisis de los resultados.

A partir de los resultados obtenidos se concluye que la mejor interpolación para temperaturas en la región de Extremadura es la que se realiza con el método de interpolación “Universal Kriging”, con un modelo de variograma esférico y con cualquier valor de pepita que se le asigne. Asimismo se obtuvo que, para las precipitaciones, el mejor método de interpolación para nuestra Región es también el “Universal Kriging”, con semivariograma esférico y con una pepita de valor 80 mm.

Introducción

La Geoestadística, desarrollada por geólogos e ingenieros de minas en la segunda mitad del siglo XX, ofrece un análisis y descripción de la variabilidad espacial de fenómenos naturales a partir de los datos disponibles, y permite realizar predicciones de estos fenómenos. La Geoestadística

utiliza funciones para modelar la variación espacial, funciones que son utilizadas posteriormente para interpolar en el espacio el valor de la variable en las localizaciones no muestreadas.

La Geoestadística permite por tanto responder a las siguientes preguntas: ¿Cuál es el patrón espacial de la variable de interés? ¿A qué escala se repite este patrón espacial? ¿Existe covariación espacial entre las distintas variables de interés? ¿Cuál es la mejor representación gráfica de mi variable? ¿Cuál es el grado de incertidumbre de estas estimas? Aunque las respuestas a estas preguntas dependen siempre de la escala espacial elegida.

El Krigeado (o Kriging) es un conjunto de técnicas geoestadísticas de interpolación que predicen valores de la variable de interés en lugares no muestreados. El Kriging es un predictor lineal, insesgado y de varianza mínima. Proporciona un mapa de predicciones en todas las ubicaciones de interés y una varianza estimada del error de predicción. Además de la predicción de la variable en cuestión nos ofrece una predicción de los residuos de cada predicción, y la probabilidad de que la variable supere, o no supere, un umbral definido. Exige hipótesis de isotropía y de estacionariedad. Básicamente consiste en asignar unos pesos a cada valor observado de la variable, de tal modo que estos pesos minimicen la varianza de la estimación global, Matheron (1970) y Zhang (1996).

Los dos métodos de Kriging que se han utilizado en este trabajo son el Kriging Ordinario (KO) y el Kriging Universal (KU). El KO es el método kriging más general y utilizado. Presupone que el valor medio de la función aleatoria estacionaria es desconocido. Esa es una hipótesis razonable a menos que haya una razón para rechazarla. El KU hay que utilizarlo cuando exista una tendencia espacial en los datos, es decir cuando los valores medios aumenten o disminuyan en alguna dirección dentro del área de estudio. Para aplicarlo se necesita una variable de apoyo. En este trabajo hemos utilizado el relieve.

A la hora de realizar un kriging hay que, además de suministrar los parámetros del semivariograma isotrópico, o los anisotrópicos, decidir la estrategia

adecuada para la selección de puntos para la interpolación. Si el número de valores en el espacio es suficiente, disminuir el radio de búsqueda de puntos puede ser la mejor estrategia, ya que evita problemas de estacionariedad, aunque hay que procurar que el radio de búsqueda no sea inferior al rango, ya que no se utilizaría toda la información que proporciona el semivariograma. Otra decisión a tomar es el tipo de interpolación: puntual o por bloques.

El semivariograma describe la variabilidad espacial del fenómeno de interés y responde a la siguiente pregunta ¿Cómo de parecidos son los puntos en el espacio a medida que estos se encuentran más alejados? Y proporciona información del comportamiento espacial de la variable.

Sin embargo, es necesario ajustar una función para cuantificar el grado y escala de variación espacial. Existen numerosos modelos de semivariogramas que se utilizan en geoestadística, siendo los más comunes el modelo esférico, el exponencial, el gaussiano y el lineal. El ajuste a una función permite extraer una serie de parámetros que son los que van a ser usados para la interpolación geoestadística y que definen el grado y escala de variación espacial. Estos parámetros son el rango, la pepita, la meseta, y la proporción de la varianza explicada por el espacio.

Resumiendo, las fortalezas de la geoestadística son: 1. Que proporciona una estimación muy robusta y 2. Que, en contraste con otros métodos de interpolación, permite conocer el grado de incertidumbre de la estimación. Por otra parte, teniendo en cuenta que para realizar un Kriging, es necesario analizar las principales hipótesis, elegir del método de Kriging, decidir el radio de búsqueda y el tipo de interpolación, ajustar el semivariograma y obtener sus parámetros, sus principales **desventajas** son: 1. Que no permite automatizar completamente el proceso de estimación y 2. Que su aplicación requiere una inversión considerable de tiempo y esfuerzo.

El objetivo que se han propuesto los autores, con este trabajo es intentar evitar estas desventajas utilizando el paquete estadístico R y el GIS SAGA. Ambas aplicaciones han permitido a los autores, a través de la realización de un gran número de interpolaciones, realizar un análisis de los errores cometidos en cada una de ellas, para establecer el mejor método de interpolación posible, y sus parámetros, en cada caso.

Se han analizado 30 años de datos mensuales de temperatura y precipitación, y se han llegado a las siguientes conclusiones:

R-SAGA GIS, es un tándem que nos permite analizar, de modo automático y sin excesiva capacidad de cálculo, distintos métodos y técnicas de interpolación.

Se ha encontrado objetivamente, para la región de estudio, para cada una de las variables estudiadas, para cada mes, o trimestre (estación), del año, cual es el mejor método, y el mejor conjunto de sus parámetros internos, para conseguir una interpolación lo más exacta posible.

Los autores están convencidos de que el tipo de análisis expuesto, y las herramientas utilizadas, son un camino fácil y cómodo para obtener a unas estimaciones geoestadísticas automatizadas lo más exactas posibles.

Objetivos

Los principales objetivos de este trabajo pueden resumirse como sigue:

Elaboración de un “script” de R que permita automatizar los métodos de interpolación en estudio, OK y UK, teniendo en cuenta todos los parámetros posibles de estos métodos.

Mediante el análisis de los resultados obtenidos al aplicar estos métodos de interpolación a temperaturas y precipitaciones, en el periodo 1981 a 1990, y en Extremadura, obtener los valores óptimos de los métodos y parámetros del proceso de interpolación.

Metodología

El Kriging es un procedimiento de geoestadística avanzado. Se basa en la suposición de que entre dos puntos cercanos debe existir algún tipo de correlación que no existe entre puntos lejanos, de forma que asigna pesos, a la distribución de la variable en estudio, en función de este hecho.

El Kriging está basado en modelos estadísticos que incluyen la autocorrelación, es decir, las relaciones estadísticas entre los puntos medidos. Gracias a esto, las técnicas de geoestadística no solo tienen la capacidad de estimar una superficie sino que también proporcionan alguna medida de certeza o precisión de las predicciones.

Este método es más adecuado cuando se sabe que hay una influencia direccional o de la distancia correlacionada espacialmente en los datos.

La fórmula general para ésta interpolación se forma como una suma ponderada de los datos:

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i)$$

donde $Z(s_i)$ es el valor medio de la variable en cuestión en la ubicación i , λ_i es una ponderación desconocida para el valor medio en la ubicación i , s_0 es la ubicación de la estimación y N la cantidad de valores observados.

Este método de interpolación puede resumirse en dos pasos: 1. Ajustar el modelo: Se crea el variograma y las funciones de covarianza para calcular los valores de dependencia estadística (autocorrelación espacial) que depende del modelo de autocorrelación.

2. Estimar los valores desconocidos en las localizaciones en las que no existe observación.

Semivariograma

De especial relevancia para esta familia de métodos de interpolación es el semivariograma, se trata de un gráfico en él que se representa la semivarianza en función de la distancia a la que se encuentran los datos. Es decir, el gráfico describe la variabilidad espacial de la variable. En general es función de la distancia y de la dirección pero normalmente se supone, así se ha hecho aquí, que solo varía con la distancia (hipótesis isotrópica).

El semivariograma muestra la degradación de la correlación espacial entre dos puntos cuanto aumenta la distancia que hay entre ellos.

Del ajuste del semivariograma se pueden extraer una serie de parámetros que se utilizarán en la interpolación, y que definen el grado y la escala de la variación espacial: la pepita (“nugget”), la meseta (“sill”) y el rango (“range”).

El rango es la distancia a la que los datos no aportan nada, delimita la zona de influencia del fenómeno en estudio, y es la distancia que hay entre el primer punto en el que se alcanza la meseta y el origen de coordenadas. Ver figura 1.

La meseta, o silo, es el valor de varianza constante que se alcanza cuando la distancia aumenta y se alcanza el rango. Es la máxima varianza.

La pepita es el valor que corta al eje Y del semivariograma, y representa la variación de los

datos cuando están a distancia nula. En teoría debería tomar un valor cero, pero en realidad la pepita recoge los errores experimentales y el hecho de que puede haber cambios de la variable en estudio a distancias menores que la distancia de muestreo.

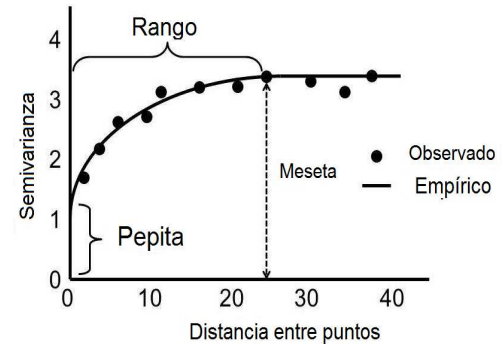


Fig. 1.- Semivariograma típico y parámetros asociados

A continuación, en la figura 2, se presentan las formas generales y las ecuaciones de los modelos matemáticos más utilizados para describir la semivarianza.

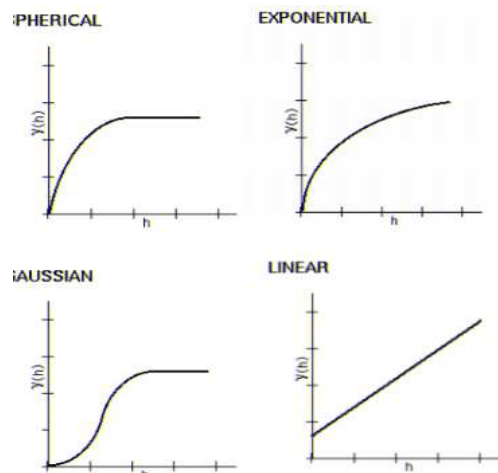


Fig 2. Modelos de semivariograma.

El modelo de semivariograma seleccionado influye en la predicción de los valores desconocidos, en particular cuando la forma de la curva cercana al origen difiere significativamente de un modelo a otro. Cuando más pronunciada sea la forma de la curva cercana al origen, más influirán los vecinos cercanos en la predicción. Como resultado la superficie de salida será menos suave. Cada modelo está diseñado para ajustarse a diferentes tipos de fenómenos de forma más precisa.

Las dos variantes del método Kriging estudiado OK y UK se han descrito arriba en la introducción.

Para el análisis de los métodos de interpolación se ha realizado un programa (“script”) que permite procesar y analizar datos de forma automática sin la necesidad de estar constantemente seleccionando los diferentes módulos, variables de estudio, etc.

Para la implementación de este programa se ha utilizado el lenguaje de programación R que mantiene una perfecta relación con SAGA-GIS mediante la carga de una serie de paquetes que facilitan su conexión.

Fijados los métodos de interpolación de estudio, y resuelta la programación, es necesario hallar los parámetros que más influyen en la interpolación.

En este trabajo se ha utilizado un tamaño de celda de 1 km que a posteriori ofrece buenos resultados, pero que se ha considerado como una hipótesis y se desconoce si este es el tamaño óptimo de la rejilla (esta hipótesis será objeto de un trabajo posterior).

Se ha considerado también un rango de 500 km y una meseta (varianza - pepita) que se calcula para cada mes en función de la varianza inicial de los datos observados. De este modo los parámetros a estudiar teniendo en cuenta los dos tipos de interpolación que se están realizando son el modelo de variograma y la pepita, de la cual se consideran como posibles valores óptimos las cantidades 1, 2, 5 y 10 décimas de grado para las temperaturas y 10, 20, 50 y 100 décimas de mm para las precipitaciones. El periodo de los datos del trabajo es de 10 años, y va desde 1981 hasta 1990.

Para realizar el análisis, no se ha fijado un número concreto de series de estaciones, sino que se parte del total de datos disponibles para cada mes, de temperatura y precipitación que ofrecen las estaciones de AEMET en Extremadura. Como el número total de estaciones del que se dispone para cada mes en cuestión es más que aceptable se divide ese conjunto de estaciones en dos, un 75% del total de las estaciones disponibles y un segundo conjunto con el 25 % de estaciones restante. Con el 75% de las estaciones se realiza la interpolación en estudio, y se extraen, de la rejilla que proporciona dicha interpolación, los resultados situados exactamente en el lugar donde se encuentran el 25% de las estaciones han permanecido al margen de la interpolación. De este modo, se evalúan muy rápidamente los resultados procedentes de la interpolación en el 25% de las estaciones que se consideran de referencia. Para poder validar los métodos de interpolación para un mes, variable, y conjunto de parámetros en estudio, se calcula la diferencia entre el valor estimado y el valor observado, así como la desviación cuadrática media para el conjunto del 25% de las estaciones con datos en ese mes.

Resultados y discusión

Se han analizado para el periodo en estudio un total de 3840 interpolaciones de temperatura, y otras tantas de precipitación, entre las cuales se han analizado para cada variable: 4 valores diferentes de pepita, cada una con 960 interpolaciones; 4 tipos de variogramas, con 960 interpolaciones cada uno; dos métodos de interpolación con 1920 interpolaciones cada uno; y los 12 meses del año, con 320 interpolaciones cada uno. A continuación, se analizan los diferentes parámetros y métodos con detalle.

Agrupando resultados según el método de interpolación se obtiene que, para la temperatura y para la precipitación, el mejor método de interpolación (con un valor más bajo del error cuadrático medio para todos los casos analizados) es el denominado en SAGA Universal Kriging. Ver figura 3.

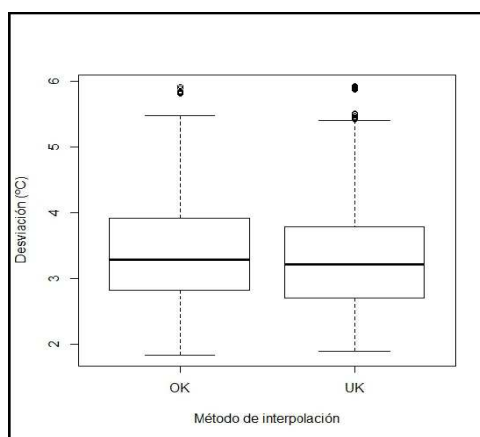


Fig 3. Representación de las desviación media obtenida con cada método de interpolación analizado

En la figura 4, en la que se muestran los resultados para los dos métodos de interpolación de temperaturas utilizados para cada mes del año, se observa también que el mejor método es el UK, pues así se cumple para cada uno de los meses del año. Se aprecia también que el mes para el que se consigue un mejor ajuste en la interpolación de las temperaturas es Diciembre mientras que el peor mes es Septiembre. Si agrupamos los meses por estaciones, la de mejor estimación sería el otoño y la peor resultado el verano.

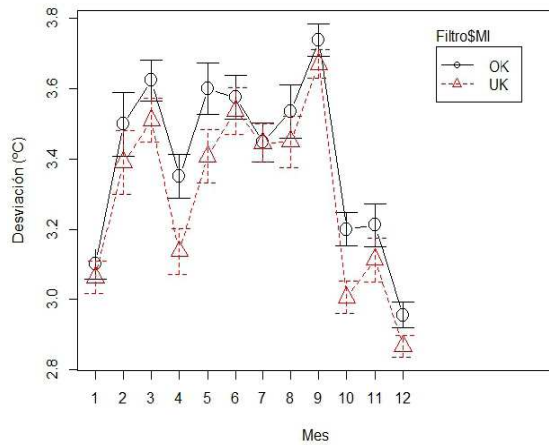


Fig. 4. Resultados para los dos métodos de interpolación de temperaturas utilizados para cada mes del año.

En la figura 5 se presentan los resultados de la interpolación de temperaturas para los cuatro tipos de variograma analizados. Los resultados de las interpolaciones muestran que el mejor modelo de variograma para temperaturas es el vGA (modelo gaussiano). Si se tiene en cuenta, además, el método de interpolación, como se muestra en la figura 6, se obtiene que el mejor modelo de variograma para el método de interpolación UK efectivamente es el vGA, pero si realizamos una interpolación mediante el método de interpolación OK el mejor modelo es el vSP (modelo esférico) muy seguidamente del vEX (modelo exponencial).

Analizando las interpolaciones por tipo de variograma, y para cada uno de los meses, se obtiene que el variograma que, para temperaturas, mejor se ajusta para la mayoría de los meses es el vGA excepto para los meses de Mayo, Junio, Julio y Octubre en los que el mejor modelo de variograma es el vSP.

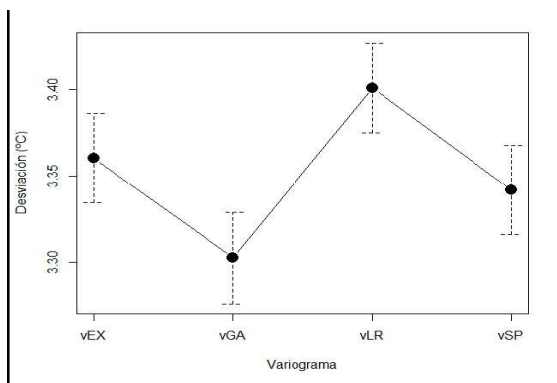


Fig. 5. Resultados de la interpolación de temperaturas para los cuatro tipos de variograma analizados

En la figura 5 se presentan los resultados de la interpolación de temperaturas para los cuatro tipos de variograma analizados. Los resultados de las interpolaciones muestran que el mejor modelo de variograma para temperaturas es el vGA (modelo

gaussiano). Si se tiene en cuenta, además, el método de interpolación, como se muestra en la figura 6, se obtiene que el mejor modelo de variograma para el método de interpolación UK efectivamente es el vGA, pero si realizamos una interpolación mediante el método de interpolación OK el mejor modelo es el vSP (modelo esférico) muy seguidamente del vEX (modelo exponencial).

Analizando las interpolaciones por tipo de variograma, y para cada uno de los meses, se obtiene que el variograma que, para temperaturas, mejor se ajusta para la mayoría de los meses es el vGA excepto para los meses de Mayo, Junio, Julio y Octubre en los que el mejor modelo de variograma es el vSP.

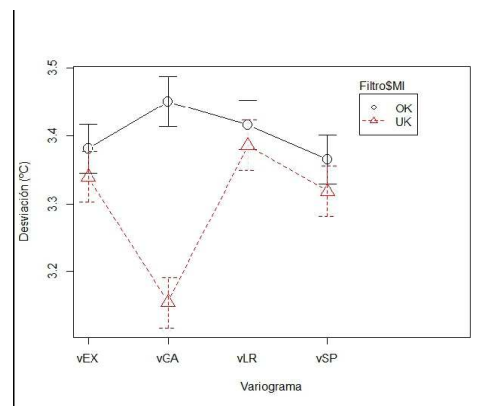


Fig. 6. Resultados de métodos UK y OK para los distintos modelos de semivariograma e interpolación de temperaturas.

Por el contrario, para las precipitaciones se ha obtenido que el modelo de semivariograma gaussiano es el que produce los peores resultados, siendo el que menores desviaciones produce el modelo de semivariograma esférico. Este resultado se obtiene para ambos métodos de krigado UK y OK.

Respecto a los distintos valores de pepita y para la interpolación de temperaturas, se ha obtenido para los distintos los valores ensayados, 1, 2, 5 y 10 décimas de °C, que tal y como se muestra en la figura 7, el valor óptimo de pepita es el 10, aunque esta conclusión no es concluyente pues cabe pensar que habría que tomar valores de pepita mayores con el fin de ver en qué momento el gráfico pasa de tener una pendiente negativa a una positiva y de así encontrar cuál es el valor de pepita que mejor se ajusta a nuestros datos.

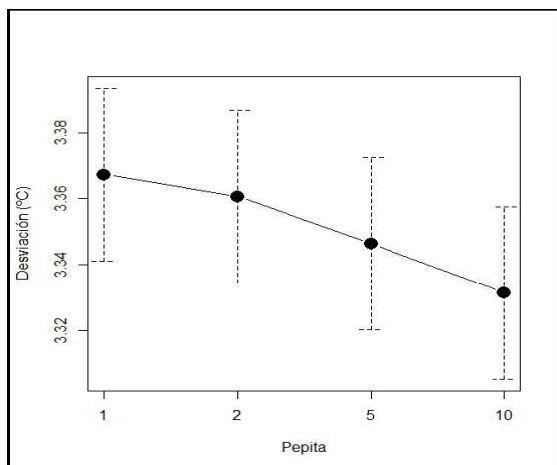


Fig. 7. Desviación media obtenida para los distintos valores de pepita

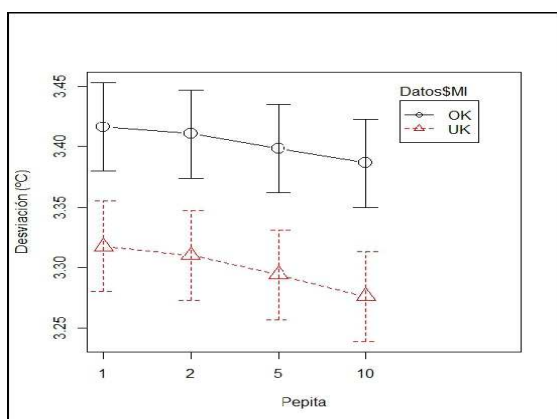


Fig 8. Desviación media obtenida para distintos valores de pepita y métodos de interpolación

La comprobación de mayores pepitas, sugerida en el párrafo anterior, no se llevó a cabo pues tal y como se muestra en la figura 9, para el modelo de semivariograma gaussiano seleccionado como el mejor para la interpolación de temperaturas mediante el método UK, el valor de la pepita no influye en el resultado, así pues esa discusión resulta innecesaria.

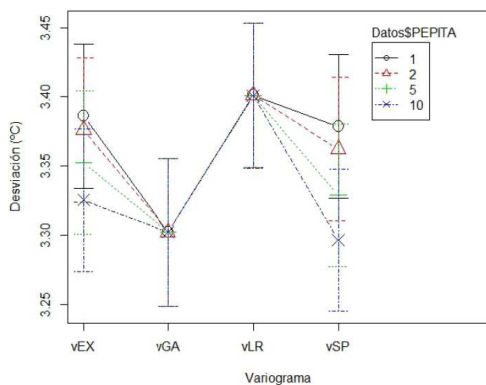


Fig 9. Resultados para distintas pepitas y modelos de semivariograma, para temperaturas.

Para las precipitaciones se ha encontrado que las menores desviaciones se producen para el valor de

pepita 80 mm, no obstante la sensibilidad de los métodos de interpolación de precipitaciones no se muestran muy sensibles a los distintos valores de pepita.

Conclusiones

Realizado el análisis de los diferentes parámetros de los métodos de interpolación "Ordinary Kriging" y "Universal Kriging" para temperaturas medias mensuales y precipitaciones totales mensuales, registradas en la comunidad de Extremadura, se han alcanzado las siguientes conclusiones:

- No es aconsejable generalizar métodos, ni siquiera parámetros de estos métodos para el problema de la interpolación a esta escala.
- Las precipitaciones de los meses lluviosos hay que tratarlas como variables distintas a las variables de los meses secos.
- La programación y el manejo del SIG SAGA mediante el paquete estadístico R resulta sencilla y permite una funcionalidad total.
- El mejor método para la interpolación óptimo de temperaturas y precipitaciones en esta Región es el denominado en SAGA "Universal Kriging", utilizando como variable de apoyo el relieve.

- Para interpolar temperaturas se recomienda el modelo de semivariograma esférico y pepita igual a 1.0 °C para los meses de mayo, junio, julio y octubre, y para el resto de los meses se recomienda el modelo de semivariograma esférico y cualquier valor de pepita.

- Para interpolar precipitaciones se recomienda un modelo de semivariograma esférico y cualquier valor de pepita.

Referencias

- A.C. Cressie , Statistics for Spatial Data Edición: Wiley and sons, Inc. 1993
- Olaya, V., Una introducción amable a SAGA GIS. Edición 1.1: Diciembre de 2004
- SAGA www.saga-gis.org
- R para Principiantes 2002, Emmanuel Paradis
- The R Project for Statistical Computing www.r-project.org
- Matheron G. (1979). La theorie des variables regionalisées et ses applications. Centre Geostatistique et de Morphologie

Mathematique, Ecole National Superior des Mines
de Paris

- Zhang, J., An iterative stochastic inverse approach
for steady-state flow in heterogeneous, variably
saturated porous media, Ph.D. dissertation, Dep. of
Hydrol. and Water Resour., Univ. of
Ariz., Tucson, 1996.