

# Uso de *R* y el paquete *Fields* para la interpolación espacial recurrente de variables ambientales \*

**Breiner Dan Bastidas Osejo**    *Universidad de Antioquia*  
**Teresita Betancur Vargas**    *Universidad de Antioquia*

---

**Resumen:** se resalta la utilidad de *R* y el paquete *Fields* como herramientas para el análisis geoespacial, en especial para la interpolación recurrente de variables ambientales, presentando un caso de estudio en la región de Urabá (Colombia). Se ha diseñado un código simple para la interpolación espacial recurrente, usando el método Spline de placa delgada, previamente definido en el paquete *Fields*, permitiendo el ingreso de  $n$  cantidad de puntos de observación y la generación de  $m$  cantidad de campos de interpolación espacial en formato raster. El código diseñado resulta útil para la generación rápida y semi automática de estos campos, los cuales suelen ser variables de entrada para modelos matemáticos distribuidos en el espacio y en el tiempo.

**Keywords:** Paquete *Fields*, distribución espacio temporal, Spline de Placa Delgada

---

## Introducción

En áreas de ingeniería, geociencias y medio ambiente, es común el uso de modelos numéricos para representar procesos físicos, que permitan estimar, valores y/o pronosticar el comportamiento en el espacio y en el tiempo de las variables de interés (Díaz et al., 2017), como es el caso de los estudios de hidrología y clima, donde es de interés conocer las variaciones espacio temporales de variables hidrometeorológicas. Los modelos numéricos distribuidos son de gran utilidad para obtener dichas estimaciones, sin embargo estos modelos también requieren mayor cantidad de información y muchas de sus entradas, como es lógico deben ser espacialmente distribuidas, la obtención de dichas distribuciones no suele representar dificultad cuando la escala temporal de trabajo es corta, sin embargo, cuando se desea realizar simulaciones de largo plazo y en áreas grandes, la obtención de la distribución espacio-temporal de las variables de entrada puede representar una tarea dispendiosa (Díaz et al., 2017). Es aquí donde la programación resulta una herramienta fundamental para optimizar el tiempo y los recursos que se invierten en preparar dichas variables de entrada. En particular, usando el lenguaje de programación *R* Core Team (2019) y el entorno de desarrollo integrado R Studio, es posible escribir códigos sencillos pero eficientes que permitan realizar esta labor, se resalta el uso de *R* por su versatilidad para ser usado como Sistema de Información Geográfica y su creciente uso en áreas de la geociencia como la hidrología (Mas, 2018; Slater et al., 2019), incorporando diversos paquetes que facilitan el análisis geoespacial, tales como *Fields* (Nychka et al., 2015). En este trabajo, se presenta un código simple diseñado para atender a la necesidad de interpolar de manera recurrente una gran cantidad de datos en el espacio y para cada paso de tiempo requerido, el enfoque principal que se aborda es el de la preparación de variables ambientales con variación espacio-temporal para ser usadas como entrada de modelos numéricos distribuidos. A continuación, se describe la metodología general, la configuración de los datos de entrada y la descripción de los datos específicos utilizados para el caso de estudio, finalmente se presenta la síntesis del código en *R* y los resultados específicos del caso de estudio.

---

\* Autor de contacto: [breiner.bastidas@udea.edu.co](mailto:breiner.bastidas@udea.edu.co)

## Materiales y métodos

El código diseñado consiste en un sencillo script de R, que en un conjunto de 54 líneas permite definir un flujo de trabajo para la interpolación de  $m$  campos espaciales de una variable ambiental con base en  $n$  observaciones. Para esto, se han utilizado y son requeridos los paquetes *Raster* (Hijmans, 2017) y *Fields* (Nychka et al., 2015). Los datos a interpolar son requeridos en formato delimitado por comas (.csv), por su facilidad para representar datos en formato tabla. El formato para el cual ha sido definido el código se presenta en la Tabla 1.

Tabla 1: Formato requerido de los datos de entrada para la interpolación (PO: punto observación).

	Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	....	Col. m+3
<b>Fila 1</b>	"Nombre"	"X"	"Y"	Nombre mapa 1	Nombre mapa 2	....	Nombre mapa m
<b>Fila 2</b>	Nombre PO 1	Coord Este PO 1	Coord Norte PO 1	Obs. 1 en PO1	Obs. 2 en PO1	....	Obs. m en PO1
<b>Fila 3</b>	Nombre PO 2	Coord Este PO 2	Coord Norte PO 2	Obs. 1 en PO2	Obs. 2 en PO2	....	Obs. m en PO2
:	:	:	:	:	:	....	:
<b>Fila n+1</b>	Nombre PO n	Coord Este PO n	Coord Norte PO n	Obs. 1 en PO n	Obs. 2 en PO n	....	Obs. m en PO n

El paquete *Raster* permite la manipulación, análisis y generación de archivos matriciales espacialmente georreferenciados (gridded spatial data), mientras que el paquete *Fields* constituye una potente herramienta para el análisis geoespacial, incorporando diversos métodos determinísticos y geoestadísticos de interpolación, entre ellos los métodos Kriging, Spline cubico y Spline de placa delgada, los cuales incorporan funciones para la determinación de los parámetros de suavizado y ajuste, así como los de co-varianza a partir del análisis de validación cruzada (cross-validation) (Nychka et al., 2015).

Para este script se ha elegido el método de base radial de Spline de placa delgada (TPS por sus siglas en inglés) definido y programado en el paquete *Fields*. El uso de la técnica determinística TPS ha sido ampliamente estudiado como una aproximación a la técnica geoestadística Kriging Universal (KU) en diversas aplicaciones (Westenbroek et al., 2010; Donato and Belongie, 2002; Moreles and Mejía, 2010), mostrando que al igual que el Kriging el TPS es útil para representar la variabilidad espacial con suavizado de datos dispersos, esta técnica de interpolación en malla irregular muestra la ventaja de no requerir la parametrización de un variograma teórico como en el caso del Kriging, ni una validación de supuestos estadísticos previos, lo cual facilita su programación para la generación recurrente de gran cantidad de superficies interpoladas, lo cual constituye el objetivo de este trabajo. El ajuste de la función TPS al conjunto de  $n$  observaciones georreferenciadas, así como la generación de los nuevos valores interpolados en una malla regular definida y la posterior escritura del archivo ascii de salida, se anidan en un ciclo *While*, para que dicho proceso se repita hasta generar los  $m$  campos espaciales de interés. En muchas ocasiones, la extensión espacial que cubren los datos es mayor que la requerida y por efectos de baja densidad de muestreo, el tamaño de celda de interpolación debe ser más grueso (menor detalle) de lo requerido, por tanto, también se incluyen algunas líneas que permiten realizar un pos-proceso a los mapas interpolados, ajustándolos a la extensión espacial y al tamaño de celda de un archivo raster de referencia. Adicionalmente, por tratarse de ajustes polinomiales, es probable, dependiendo de

la densidad de datos cercanos a cero, que el ajuste arroje valores interpolados negativos, por tanto se ha configurado una corrección adicional para que la matriz interpolada reemplace dichos valores por cero, también en caso de que todos los valores de las observaciones a interpolar sean cero, no procede realizar un ajuste polinomial TPS, sino que el código configura una matriz de ceros.

A continuación, se presenta la síntesis del código en R para realizar la interpolación de manera recurrente, a partir de los datos consignados en el archivo *formato.csv* y usando el archivo de referencia espacial *referencia.asc*

```
TableFormat <- read.csv("formato.csv",header = TRUE, sep = ",") # lee los datos
x <- TableFormat[,2] # asigna coordenada Este alojada en Col. 2
y <- TableFormat[,3] # asigna coordenada Este alojada en Col. 3
title <- names(TableFormat) # asigna los nombres de los mapas alojados en Fila 1
Referencia <- raster("referencia.asc") # carga el archivo de referencia
m <- 6 # número de mapas a interpolar más 3 (ej. m <- 10 interpolará 7 mapas)
j=4 # columna inicio interpolación (ej. j=4 inicia con datos del mapa 1)
while(j<=m) {
  z <- TableFormat[,j] # asigna los valores a interpolar para el mapa j
  X0 <- seq(min(x),max(x),100) # asigna malla interpolación, dx=100m en este caso
  Y0 <- seq(min(y),max(y),100) # asigna malla interpolación, dy=100m en este caso
  if (sum(z)>0) {
    fit <- Tps(cbind(x,y),z) # ajusta Spline de Placa Delgada a los datos (x,y,z)
    grid.l <- list(X=X0, Y=Y0) # define la malla de interpolación
    outp <- predictSurface(fit,grid.list=grid.l,extrap=T) # genera interpolación
    matriz <- outp$z # asigna los datos interpolados a una matriz
  } else {
    matriz <- matrix(0,length(X0), length(Y0)) # caso datos cero - matriz ceros
    interpolatemap <- writeArcASCII("mapint.asc", matriz, min(x), min(y),
                                   length(X0), length(Y0),100) # Escribe la matriz a un ASCII
    mapint1 <- raster("mapint.asc") # asigna el ASCII para operar
    mapint2 <- resample(mapint1, Referencia) # remuestrea ASCII con referencia file
    mapint3 <- crop(mapint2, Referencia) # corta ASCII a extensión de referencia file
    mapint3[mapint3<0]<-0 # configura en ceros los valores negativos
    plot(mapint3) # grafica el mapa interpolado y ajustado
    titlesp <- title[j] # asigna el nombre del archivo ASCII a exportar
    writeRaster(mapint3,titlesp,format="ascii",overwrite=T) # exporta ASCII
    j = j+1 } # actualiza ciclo While para interpolar el mapa siguiente (j+1)
```

Para el caso de estudio, las variables a interpolar corresponden a valores diarios de precipitación (mm/día), temperatura máxima y temperatura mínima (°C), para 18 estaciones meteorológicas con registro de lluvia, cuatro (4) con registro de temperatura y siete (7) puntos con información auxiliar (observaciones auxiliares) de temperatura. Los registros son suministrados por el Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia - IDEAM y abarcan una extensión de 15.198 km<sup>2</sup> en la región del Eje Bananero del Urabá Antioqueño en Colombia, con una ventana temporal de interés de cinco años (2014-2018). La interpolación se realiza a tamaño de pixel de 100 m, para la extensión que abarcan las estaciones.

## Resultados

En la Figura 1 y en la Figura 2 se muestran algunos ejemplos de superficies diarias de precipitación y temperatura máxima respectivamente, interpoladas usando el código propuesto. Para el caso de la precipitación, es evidente su comportamiento totalmente disperso en la escala diaria, la generalidad de los 1826 mapas muestra el predominio de una estructura espacial altamente caótica y dispersa, aproximando adecuadamente el comportamiento físico de la lluvia a esta escala, en especial en una zona de estudio de carácter regional. En cuanto a la temperatura máxima, como era de esperarse, en la escala diaria no presenta una dispersión tan grande como la precipitación y conserva una estructura espacial con baja dispersión. El código toma un tiempo de ejecución de aproximadamente 56 segundos por cada superficie a interpolar, en un computador con memoria RAM de 8 GB y procesador i5-3470 (3,20 GHz)

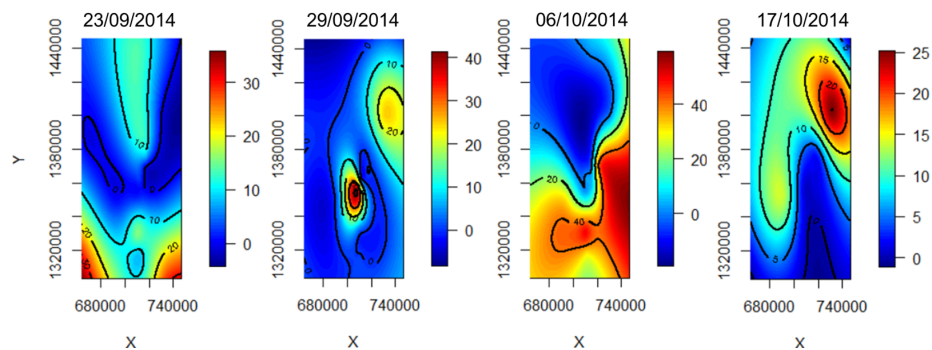


Figura 1: Campos espaciales de precipitación diaria (mm) para algunos días del 2014.

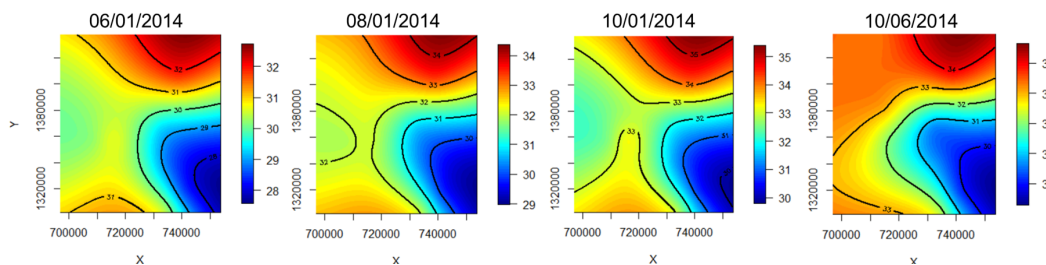


Figura 2: Campos espaciales de temperatura máxima (C) para algunos días del 2014.

## Conclusiones

El código presentado, ilustra de forma sintética el uso de R y paquetes asociados para la generación recurrente de una gran cantidad de campos espaciales de variables ambientales, usando un método determinístico de interpolación que se ajusta de manera adecuada a datos dispersos. El código no permite la validación automática de la interpolación de cada campo espacial, dado que el objetivo es generar campos espaciales de manera rápida y semi-automática, de tal forma, el control de la validación de los resultados de interpolación es tarea del usuario. Para el caso de estudio y otras aplicaciones usando otras variables ambientales, las interpolaciones han resultado representar adecuadamente los datos de entrada.

## Referencias

- Díaz, Diana, Carlos Bojacá, Rodrigo Gil, Francisco de P. Gutiérrez, Juanita Burgos, Javier Riascos, Gabriel Villalobos and Favio Cala Vitery. 2017. *Modelado y simulación de sistemas naturales*. 1 ed. Bogotá D.C.: Universidad de Bogotá Jorge Tadeo Lozano.
- Donato, Gianluca and Serge Belongie. 2002. Approximate Thin Plate Spline Mappings. In *Computer Vision - ECCV 2002*, ed. Springer-Verlag. Vol. 2352 Berlin: pp. 21–31.  
**URL:** <http://link.springer.com/10.1007/3-540-47969-4>
- Hijmans, Robert J. 2017. “raster: Geographic Data Analysis and Modeling.”.  
**URL:** <https://cran.r-project.org/package=raster>
- Mas, Jean-François. 2018. *Análisis espacial con R : Usa R como un Sistema de Información Geográfica*. Kocani: EUROPEAN SCIENTIFIC INSTITUTE.
- Moreles, Miguel Ángel and Francisco Mejía. 2010. “Interpolación con Funciones de Base Radial y el Método del Sistema Diferencial para identificación de parámetros en acuíferos.” *Revista Internacional de Metodos Numericos para Calculo y Diseno en Ingenieria* 26(3):241–247.
- Nychka, Douglas, Reinhard Furrer, John Paige and Stephan Sain. 2015. “fields: Tools for spatial data, v. 8.4-1.”.
- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.  
**URL:** <https://www.R-project.org/>
- Slater, Louise J., Guillaume Thirel, Shaun Harrigan, Olivier Delaigue, Alexander Hurley, Abdou Khouakhi, Ilaria Prodosimi, Claudia Vitolo and Katie Smith. 2019. “Using R in hydrology: a review of recent developments and future directions.” *Hydrology and Earth System Sciences Discussions* (d):1–33.
- Westenbroek, M.S., V.a. Kelson, W.R. Dripps, R.J. Hunt and K.R. Bradbury. 2010. “SWB — A Modified Thornthwaite-Mather Soil-Water- Balance Code for Estimating Groundwater Recharge.” *U.S. Geological Survey Techniques and Methods* 6-A31 p. 60.