

PAQUETE RegRAIN

Interpolador de variables climáticas que
combina las técnicas MLR, Spline e IDW

Manual de Usuario Final

2017

Diego Fernando Alzate Velásquez
Investigador Msc.
CI Tibaitatá
CORPOICA

CONTENIDO

1. ¿Qué es RegRAIN?	3
2. Acerca del modelo RegRAIN.....	3
3. Lenguaje de programación.....	4
4. Requerimientos del sistema.....	4
5. Paquetes necesarios.....	5
6. Entrada de datos al paquete RegRAIN	5
a. RegRAIN.....	6
b. RegTEMP	9
7. Rutinas de ejemplo del paquete RegRAIN	11
a. Rutina ejemplo función RegRAIN	12
b. Rutina ejemplo función RegTEMP.....	14
8. Referencias bibliográficas	16

Este documento es una guía detallada del manejo del **paquete RegRAIN v0.1.0**.

1. ¿Qué es RegRAIN?

RegRAIN es un interpolador regionalizado de variables climáticas que incorpora una Regresión Lineal Múltiple (MLR) en el proceso a partir de variables de terreno como la elevación, la pendiente y el aspecto. Para su cálculo se utiliza un modelo digital de elevación y datos a escala diaria de estaciones climáticas de superficie. El método utiliza también los interpoladores Spline e IDW. La versión 0.1.0 incluye las variables: precipitación acumulada diaria y temperatura (max, min, med) diaria.

Incorpora una función de validación cruzada para verificar el ajuste del modelo en cada interpolación realizada. Los resultados de RegRAIN son: ráster en formato *.tif con el resultado de la interpolación y un plot en formato *.png con una gráfica y estadísticos de error y bondad de ajuste. En la Figura 1 se muestra un ejemplo de las salidas o resultados del paquete RegRAIN.

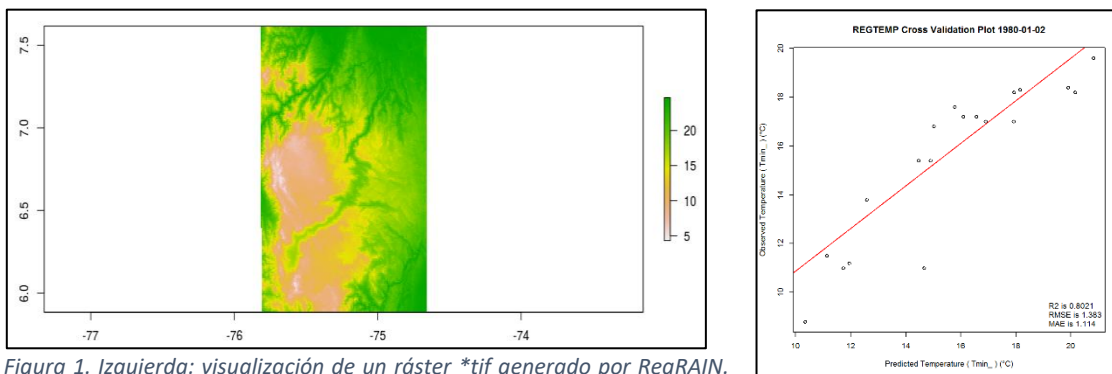


Figura 1. Izquierda: visualización de un ráster *.tif generado por RegRAIN. Derecha: gráfico de validación cruzada con estadístico de error y ajuste.

2. Acerca del modelo RegRAIN

RegRAIN y RegTEMP son modelos regionalizados de lluvia y temperatura del aire, respectivamente, que integran el uso de modelos digitales de elevación (DEM), sus transformaciones (pendiente y aspecto) e información climática homogenizada. El cálculo se basa en una regresión lineal múltiple (MLR).

$$y_i = a_1 \cdot x_{i1} + a_2 \cdot x_{i2} + a_3 \cdot x_{i3} + a_4 \cdot x_{i4} + a_5 \cdot x_{i5} + E_i \quad (\text{Ec.1})$$

Donde,

- y_i = variable meteorológica a interpolar en la estación "i".
- a_{ik} = Coeficiente de regresión en la estación climática "i" para la variable climática "k".
- x_{ik} = k variables en la estación climática "i" con $x_{i1} = 1$.
- E_i = Residuo.

Las 5 variables k que corresponden con factores geográficos y derivaciones de terreno son:

- Latitud y longitud (grados decimales o metros).
- Elevación (a partir del DEM en metros).
- Pendiente y aspecto para cada estación climática (obtenida del DEM en grados).

3. Lenguaje de programación

El paquete fue desarrollado en lenguaje R y para su utilización es necesaria la última versión del software R. Es altamente recomendado que se utilice también la interfaz gráfica R Studio.



R es un entorno y lenguaje de programación con un enfoque al análisis estadístico y es una derivación e implementación libre del lenguaje S. Es uno de los lenguajes más utilizados en investigación por la comunidad estadística, siendo además muy popular en el campo de la minería de datos, la investigación biomédica, la bioinformática y las matemáticas financieras. A esto contribuye la posibilidad de cargar diferentes bibliotecas o paquetes con funcionalidades de cálculo y gráficas.

R es parte del sistema GNU y se distribuye bajo la licencia GNU GPL. Está disponible para los sistemas operativos Windows, Macintosh, Unix y GNU/Linux.

4. Requerimientos del sistema

El requerimiento del paquete depende del volumen de información a interpolar, por lo cual las características mínimas sugeridas para su adecuado funcionamiento y del software R son las siguientes:

- 1 gigahertz (GHz) or faster 32-bit (x86) or 64-bit (x64) processor
- 1 gigabyte (GB) RAM (32-bit) or 2 GB RAM (64-bit)
- 16 GB available hard disk space (32-bit) or 20 GB (64-bit)

R es requerido y se recomienda también R Studio.

<https://cran.r-project.org/bin/windows/base/>

<https://www.rstudio.com/products/rstudio/download/>



5. Paquetes necesarios

Para el funcionamiento de RegRAIN se debe previamente instalar y activar los siguientes paquetes en el entorno de R:

- **dplyr**: provee una serie de herramientas para la manipulación de grandes volúmenes de datos.
- **rgdal**: herramientas para la abstracción de datos geoespaciales.
- **rgeos**: interface al motor geométrico open source GEOS.
- **fields**: herramientas para el manejo de datos espaciales. La función utilizada fue *Tps* (*Thin Plate Spline*) que interpola a través del método Spline las coordenadas geográficas de entrada del modelo.
- **raster**: herramientas para el análisis y modelamiento de datos geográficos. Las funciones utilizadas son: *crs*, *extent*, *extract*, *intrepolate*, *projection*, *raster*, *res*, *resample*, *terrain* y *writeRaster*.
- **sp**: métodos y clases para datos espaciales. Las funciones utilizadas son: *bbox*, *coordinates*, *gridded* y *proj4string*.
- **hydroGOF**: funciones de “bondad de ajuste” para comparar datos observados y simulados de series hidrológicas. Las funciones utilizadas son: *mae* y *rmse*
- **gstat**: herramientas para el modelamiento geoestadístico espacial y espacio-temporal, predicción y simulación. Las funciones utilizadas son: *idw*.

6. Entrada de datos al paquete RegRAIN

El paquete RegRAIN se compone de dos funciones principales que realizan las interpolaciones regionalizadas para las variables de precipitación diaria y temperatura diaria: RegRAIN y RegTEMP.

```
RegRAIN(datos, dem, ini, fin, crossv)
```

```
RegTEMP(datos, dem, ini, fin, crossv, var)
```

Cada función tiene unos argumentos de entrada con requerimientos específicos, como se describe a continuación:

a. RegRAIN

Para ejecutar la función RegRAIN se debe escribir en la consola de R lo siguiente:

```
RegRAIN(datos, dem, ini, fin, crossv)
```

Los argumentos de la función se explican a continuación:

datos es un objeto de la clase “data.frame” que contiene los datos de lluvia de entrada y que debe tener la siguiente estructura:

FECHA	CODIGO	LONGITUD	LATITUD	PPT
2010-01-02	11025010	-76.08433	5.878333	0.0
2010-01-03	11025010	-76.08433	5.878333	0.0
2010-01-04	11025010	-76.08433	5.878333	0.0
2010-01-05	11025010	-76.08433	5.878333	0.0
2010-01-06	11025010	-76.08433	5.878333	0.0
2010-01-07	11025010	-76.08433	5.878333	0.0
2010-01-08	11025010	-76.08433	5.878333	0.3
2010-01-09	11025010	-76.08433	5.878333	0.0
2010-01-10	11025010	-76.08433	5.878333	2.6

Figura 2. Modelo de datos de precipitación de entrada

Las columnas presentes en el data.frame datos son las siguientes:

- **FECHA:** en formato AA-MM-DD. Está columna debe ser de clase “Date” y el formato debe coincidir con AA-MM-DD, de lo contrario se pueden presentar errores en la interpolación.
- **CODIGO:** código o identificador de la estación climática.
- **LONGITUD:** con las longitudes o coordenadas X de cada estación climática.
- **LATITUD:** con las latitudes o coordenadas Y de cada estación climática.
- **PPT:** con el acumulado diario de lluvia en milímetros (mm). El decimal recomendado es ".". Los valores NA pueden ser "99999" o "NA" para que sean tomados por el programa.

Con el paquete activo ☒ RegRAIN ejecute **View(datos)** en la consola para ver los datos de precipitación de ejemplo.

dem es un objeto de la clase “RasterLayer” que contiene el modelo digital de elevaciones (DEM) para el área a interpolar.

Las características del **dem** de entrada como: sistema de proyección, tamaño de pixel, número de columnas, filas y tipo de coordenadas, serán utilizadas y se asignarán al resultado de RegRAIN. Incluido en el paquete se encuentra un DEM de ejemplo en formato *.tif (Figura 3), que coincide espacialmente con los datos de precipitación detallados anteriormente, y que cuenta con las siguientes características:

- Sistema de proyección: WGS84.
- Tipo de coordenadas: geográficas en grados decimales.
- Tamaño de pixel: 0.0091° x 0.0091°
- Filas, columnas, pixeles: 191, 128, 24448

Con el paquete activo ☒ RegRAIN ejecute **plot(dem)** en la consola para ver el modelo digital de elevación de ejemplo.

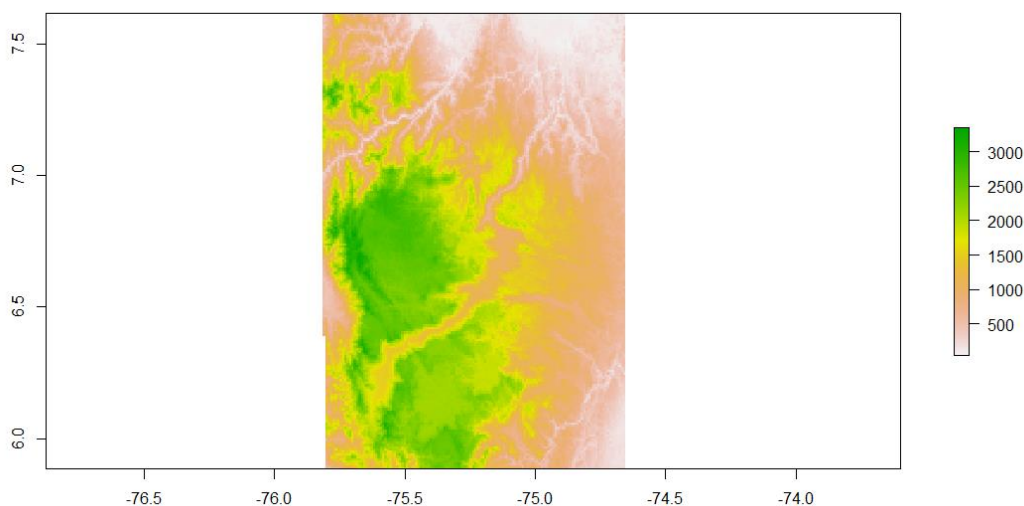


Figura 3. Visualización del DEM de ejemplo incluido en el paquete RegRAIN. Clase “RasterLayer”

ini es un número que indica la fecha inicial de la interpolación y que no es de la clase “Date”, si los datos de lluvia se encuentran entre el 01.01.2010 y el 31.01.2010, se tienen en total

31 días a interpolar. Si se desea interpolar entre el día 10.01.2010 y el 20.01.2010, **ini** debe ser 10.

fin es un número que indica la fecha final de la interpolación y que no es de la clase “Date”, si los datos de lluvia se encuentran entre el 01.01.2010 y el 31.01.2010, se tienen en total 31 días a interpolar. Si se desea interpolar entre el día 10.01.2010 y el 20.01.2010, **fin** debe ser 20.

crossv es de la clase “logical” con valores posibles “FALSE” o “TRUE”. Cuando es “TRUE”, un proceso de validación cruzada será realizado y adicional al ráster de precipitación, se generará y exportará al directorio de trabajo un plot mostrando la dispersión entre lo observado y simulado e indicadores de error y de bondad de ajuste. Cuando es “FALSE” solo se genera el o los rásteres con los resultados de la interpolación sin el gráfico de validación cruzada.

El proceso de validación cruzada genera una superficie de interpolación descontando una a una las estaciones y extractando los datos simulados por el modelo RegRAIN, por lo que cuando se usan muchas estaciones puede ser un proceso dispendioso. Al final se obtiene una matriz de igual número de estaciones con las columnas de datos de PPT observados y simulados, a partir de los cuales se realiza el gráfico de dispersión y se calculan los estadísticos de error (mean absolute error mae y root mean square error RMSE) y de ajuste (R^2). En el caso de la precipitación diaria, el ajuste de la curva se realiza a partir de un modelo exponencial y el coeficiente de determinación (R^2) se calcula como 1 menos el cociente entre la suma residual de cuadrados y la suma total de cuadrados.

En la Figura 4 se muestra un ejemplo de gráfico de validación cruzada para precipitación generado con **crossv** = TRUE.

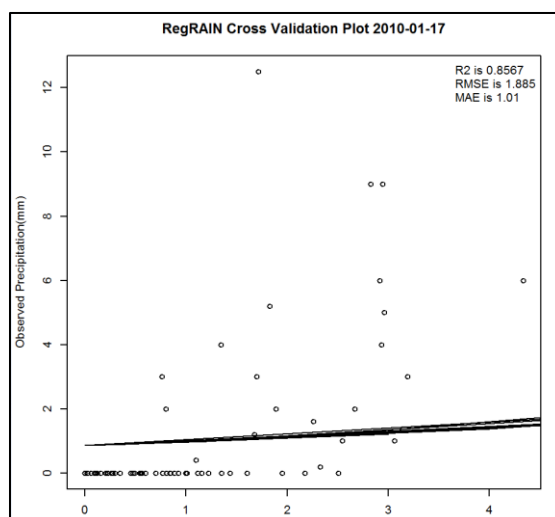


Figura 4. Gráfico de validación cruzada a partir del dataset “datos”, ejemplo para el día 17.01.2010.

b. RegTEMP

Para ejecutar la función RegTEMP se debe escribir en la consola de R lo siguiente:

```
RegTEMP(datos, dem, ini, fin, crossv, var)
```

Los argumentos de la función se explican a continuación:

datos es un objeto de la clase “data.frame” que contiene los datos de temperatura de entrada y que debe tener la siguiente estructura:

	CODIGO	FECHA	LONGITUD	LATITUD	TEMP
1	11025010	1980-01-01	-76.08433	5.878333	12.2
2	11025010	1980-01-02	-76.08433	5.878333	12.0
3	11025010	1980-01-03	-76.08433	5.878333	13.0
4	11025010	1980-01-04	-76.08433	5.878333	12.4

Figura 5. Modelo de datos de temperatura de entrada

Las columnas presentes en el data.frame datos son las siguientes:

- **CODIGO:** código o identificador de la estación climática.
- **FECHA:** en formato AA-MM-DD. Esta columna debe ser de clase “Date” y el formato debe coincidir con AA-MM-DD, de lo contrario se pueden presentar errores en la interpolación.
- **LONGITUD:** con las longitudes o coordenadas X de cada estación climática.
- **LATITUD:** con las latitudes o coordenadas Y de cada estación climática.
- **TEMP:** con la temperatura (min,max,med) del día en grados celsius (°C). El decimal recomendado es ".". Los valores NA pueden ser "99999" o "NA" para que sean tomados por el programa.

Con el paquete activo ☒ RegRAIN ejecute **View(datos_tmin)** en la consola para ver los datos de temperatura mínima de ejemplo.

dem es un objeto de la clase “RasterLayer” que contiene el modelo digital de elevaciones (DEM) para el área a interpolar.

Con el paquete activo ☒ RegRAIN ejecute **plot(dem)** en la consola para ver el modelo digital de elevación de ejemplo.

ini es un número que indica la fecha inicial de la interpolación y que no es de la clase “Date”, si los datos de

temperatura se encuentran por ejemplo entre el 01.01.1980 y el 31.01.1980, se tienen en total 31 días a interpolar. Si se desea interpolar entre el día 15.01.1980 y el 22.01.1980, **ini** debe ser 15.

fin es un número que indica la fecha final de la interpolación y que no es de la clase “Date”, si los datos de lluvia se encuentran por ejemplo entre el 01.01.1980 y el 31.01.1980, se tienen en total 31 días a interpolar. Si se desea interpolar entre el día 15.01.1980 y el 22.01.1980, **fin** debe ser 22.

crossv es de la clase “logical” con valores posibles “FALSE” o “TRUE”. Cuando es “TRUE”, un proceso de validación cruzada será realizado y adicional al ráster de temperatura, se generará y exportará al directorio de trabajo un plot mostrando la dispersión entre lo observado y simulado e indicadores de error y de bondad de ajuste. Cuando es “FALSE” solo se genera el o los rásteres con los resultados de la interpolación sin el gráfico de validación cruzada.

El proceso de validación cruzada genera una superficie de interpolación descontando una a una las estaciones y extractando los datos simulados por el modelo RegTEMP, por lo que cuando se usan muchas estaciones puede ser un proceso dispendioso. Al final se obtiene una matriz de igual número de estaciones con las columnas de datos de PPT observados y simulados, a partir de los cuales se realiza el gráfico de dispersión y se calculan los estadísticos de error (mean absolute error mae y root mean square error RMSE) y de ajuste (R^2). En el caso de la temperatura (min,max,med) diaria, el ajuste de la curva se realiza a partir de un modelo lineal y el coeficiente de determinación (R^2) se obtiene de la regresión lineal.

En la Figura 6 se muestra un ejemplo de gráfico de validación cruzada para temperatura generado con **crossv** = TRUE.

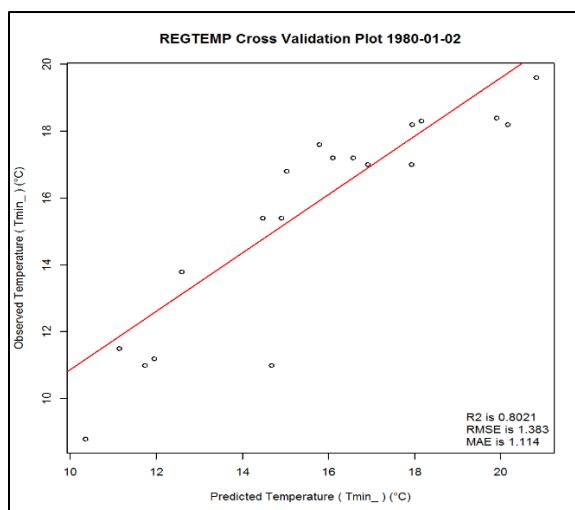


Figura 6. Gráfico de validación cruzada a partir del dataset “datos_tmin”, de ejemplo para el día 02.01.1980.

var es un objeto de clase “character” que indica el tipo de temperatura diaria a interpolar, entre las siguientes opciones: tmin, tmax y tmed.

7. Rutinas de ejemplo del paquete RegRAIN

Para familiarizarse y conocer el funcionamiento de RegRAIN se sugiere que el usuario utilice y ejecute los dos ejemplos disponibles. Para esto debe estar instalado (ver manual de instalación) y activo el paquete RegRAIN, lo cual se puede verificar en la esquina inferior derecha de R Studio, como se muestra a continuación:

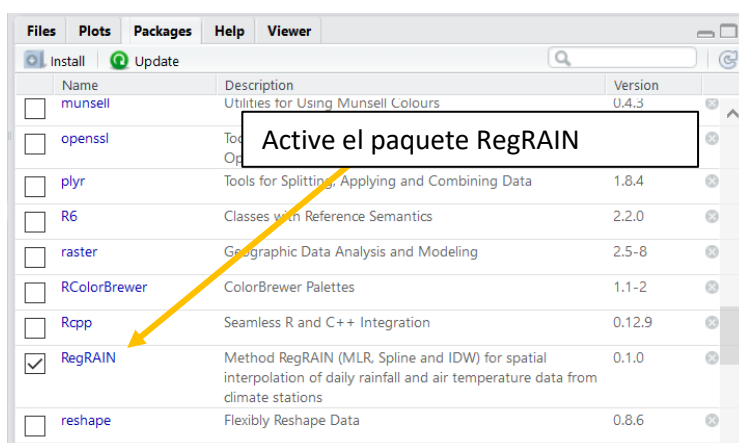


Figura 7. Vista inferior derecha de R Studio indicando la correcta instalación y activación del paquete RegRAIN.

En esta misma vista, haciendo clic en el nombre del paquete RegRAIN se puede acceder a la vista de información general, todo dentro de la misma interface de R Studio:

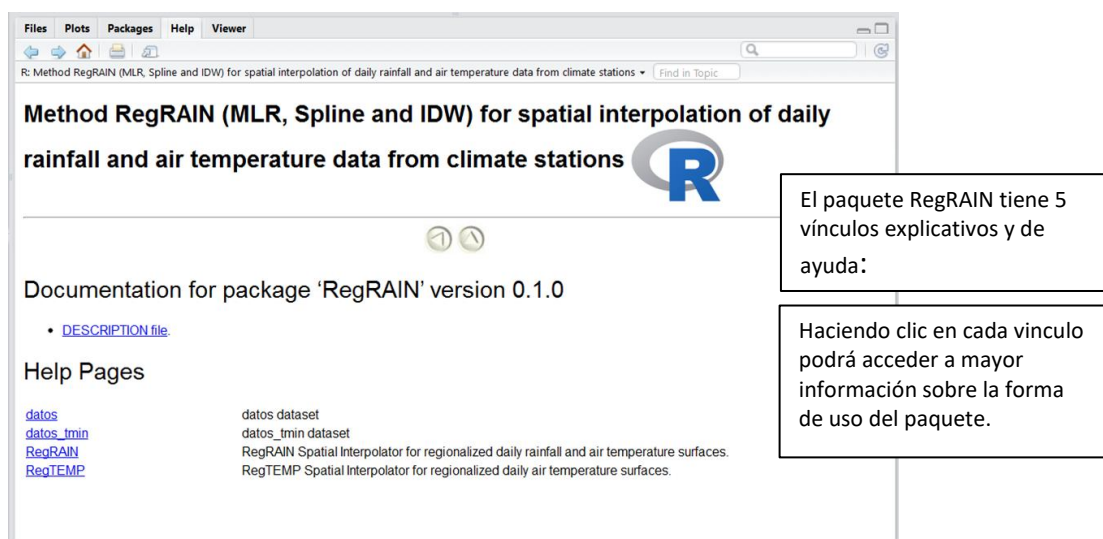


Figura 8. Página de información general del paquete RegRAIN, visible desde la interfaz de R Studio.

A partir de la página de información del paquete, se puede acceder a detalles de las funciones de precipitación y temperatura y correr las rutinas de ejemplo, como se explica a continuación:

a. Rutina ejemplo función RegRAIN

En la página de información general se hace clic en el link RegRAIN:

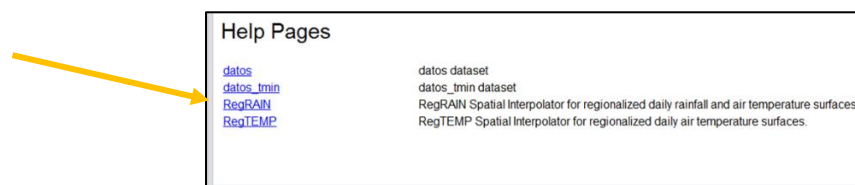


Figura 9. Vista en R Studio de la página general del paquete RegRAIN

Una vez se hace clic se accede a la página de la función RegRAIN, se tiene la siguiente vista:

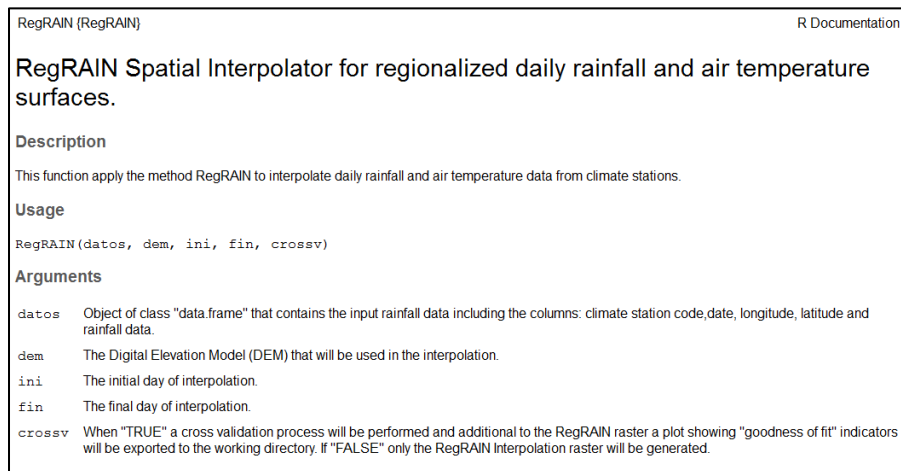


Figura 10. Documentación en R Studio del paquete RegRAIN, se destaca la descripción, uso y argumentos.

Al final de esta página se encuentran las líneas de ejemplo a ejecutar:

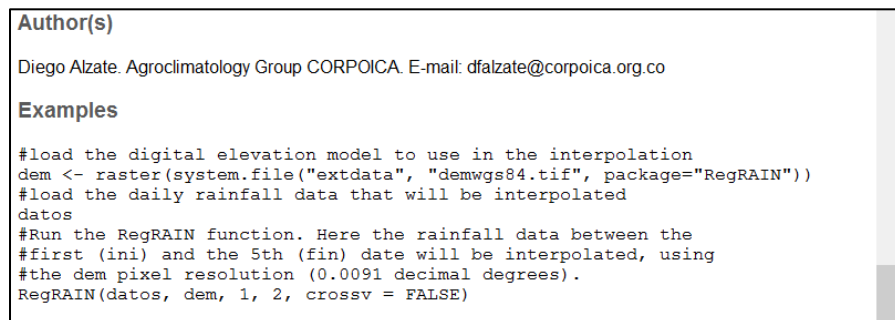


Figura 11. Parte final de la documentación de RegRAIN con la rutina de ejemplo.

Esas líneas de código se deben copiar y pegar en una página en blanco en la parte superior derecha de R Studio o directamente en la consola, luego de esto se hace clic en **Run**:

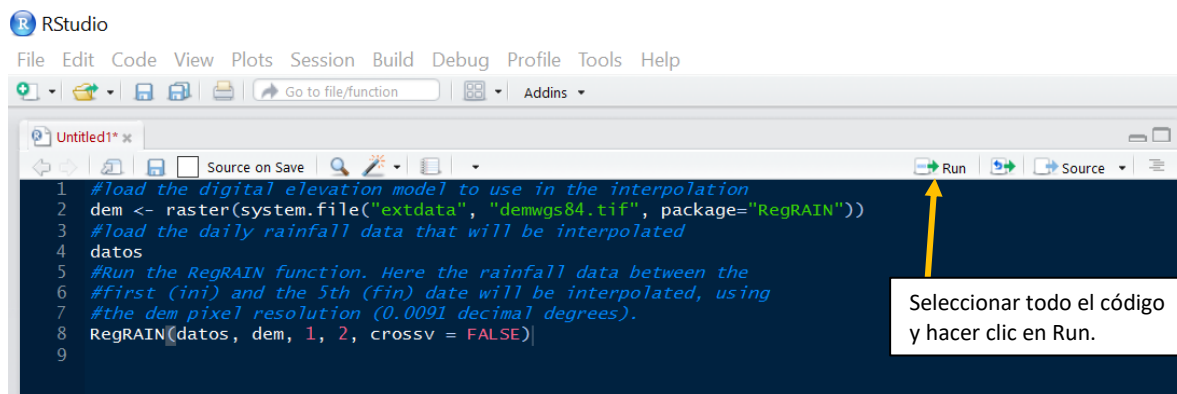


Figura 12. Código listo para correr en una hoja en blanco en la parte superior izquierda de R Studio.

La aplicación debe estar activa para que funcione correctamente. En la consola (parte inferior izquierda de R Studio) se observan los procesamiento del paquete.

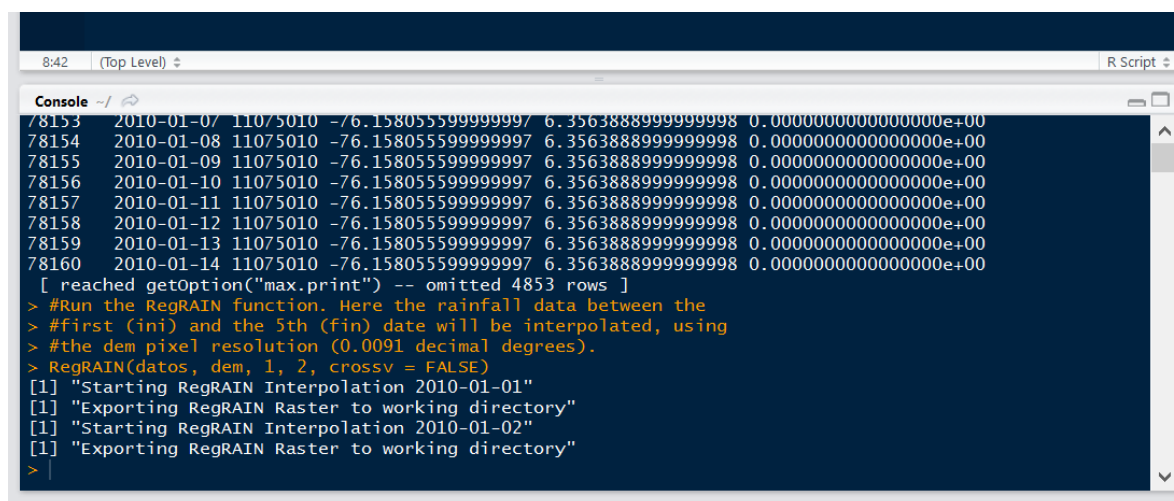


Figura 13. Vista de la consola de R Studio con los resultados del procesamiento de las líneas de ejemplo.

- "Starting RegRAIN Interpolation 2010-01-01" indica que el proceso de interpolación ha iniciado.
- "Exporting RegRAIN Raster to working directory" indica que ha incluido la interpolación y se puede consultar el raster resultante en el directorio de trabajo que el usuario defina.

Para conocer el directorio de trabajo en el cual se exportó el raster se puede escribir en la consola **getwd()**:

```
> getwd()
[1] "C:/Users/S5-371-53RL/Documents"
>
```

En este caso el raster resultante se guardó en la carpeta de Documentos:

Este equipo > Documentos



Nombre	Fecha de modificación	Tipo	Tamaño
 RegRAIN_PPT_2010-01-02	23/03/2017 4:36 p. m.	Archivo TIF	6 KB
 RegRAIN_PPT_2010-01-01	23/03/2017 4:36 p. m.	Archivo TIF	127 KB

Figura 14. Carpeta Mis Documentos en Windows mostrando los rásteres resultantes de la interpolación.

El archivo generado se encuentra en formato *.tif y es compatible con cualquier software SIG, se puede visualizar y procesar en R y contiene la información de precipitaciones de forma espacial para el día y área interpolada.

b. Rutina ejemplo función RegTEMP

En la página de información general se hace clic en el link RegTEMP:

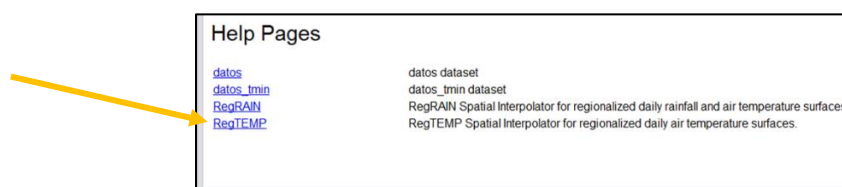


Figura 15. Vista en R Studio de la página general del paquete RegRAIN

Una vez se hace clic se accede a la página de la función RegTEMP, se tiene la siguiente vista:

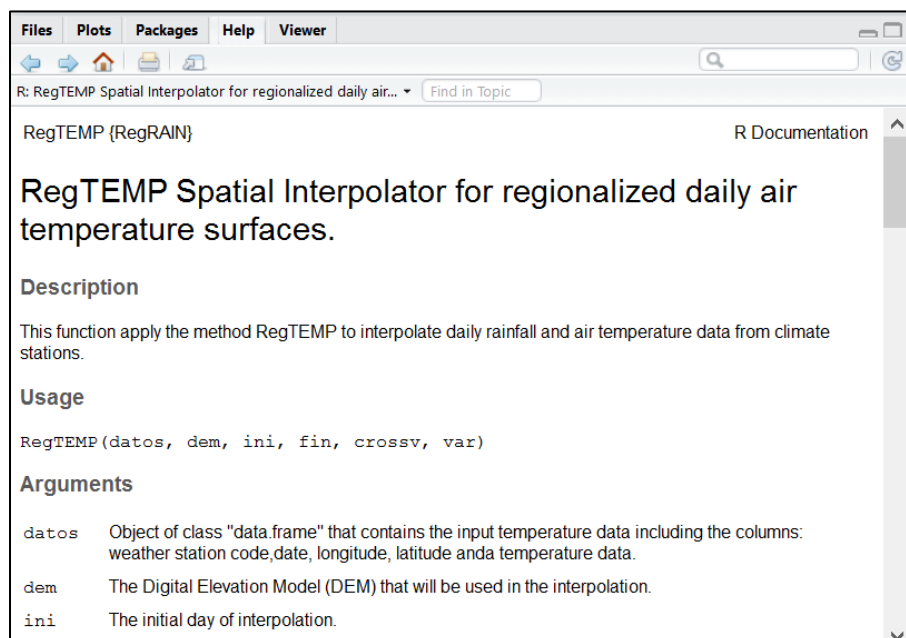


Figura 16. Documentación en R Studio del paquete RegTEMP, se destaca la descripción, uso y argumentos.

Al final de esta página se encuentran las líneas de ejemplo a ejecutar:

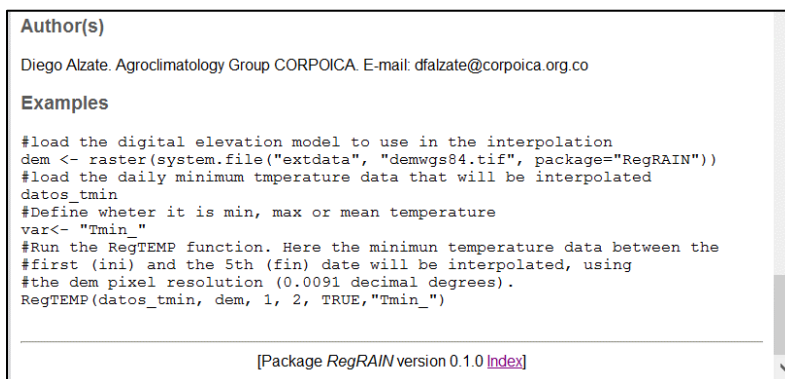


Figura 17. Parte final de la documentación de RegTEMP con la rutina de ejemplo.

Esas líneas de código se deben copiar y pegar en una página en blanco en la parte superior derecha de R Studio o directamente en la consola, luego de esto se hace clic en **Run** o se presiona enter:

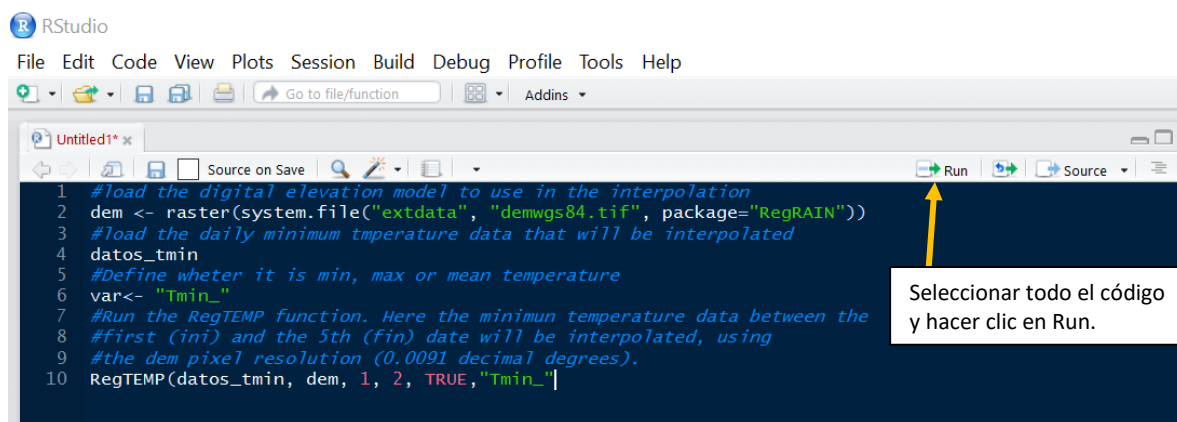


Figura 18. Código listo para correr en una hoja en blanco en la parte superior izquierda de R Studio.

La aplicación debe estar activa para que funcione correctamente. En la consola (parte inferior izquierda de R Studio) se observan los procesamiento del paquete.

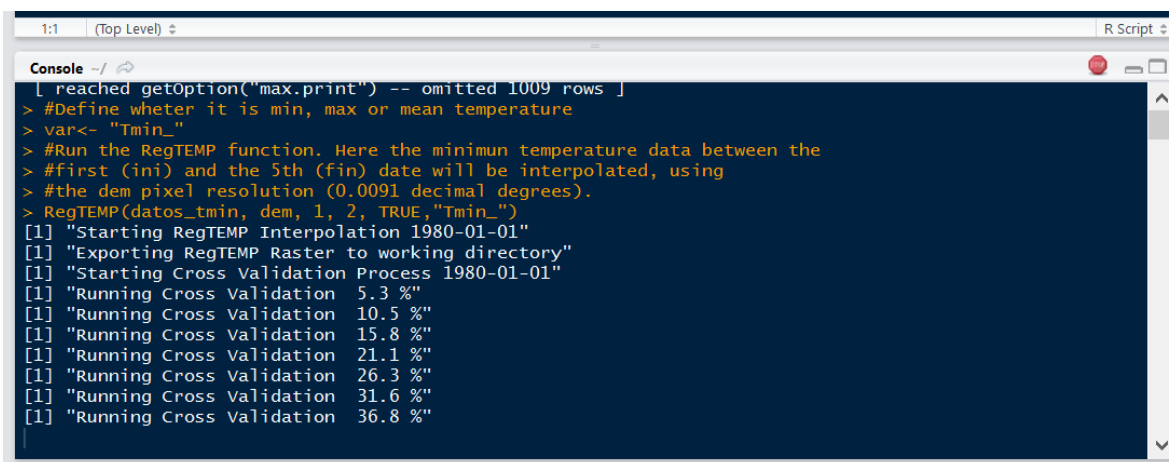


Figura 19. Vista de la consola de R Studio con los resultados del procesamiento de las líneas de ejemplo.

- “Starting RegTEMP Interpolation 1980-01-01” indica que el proceso de interpolación ha iniciado.
- “Exporting RegTEMP Raster to working directory” indica que ha incluido la interpolación y se puede consultar el raster resultante en el directorio de trabajo que el usuario defina.
- “Starting Cross Validation Process 1980-01-01” indica que el proceso de validación cruzada inició para ese día.
- “Running Cross Validation 5.3%” indica el avance del proceso de validación cruzada.

Para conocer el directorio de trabajo en el cual se exportó el ráster se puede escribir en la consola **getwd()**:

```
> getwd()
[1] "C:/Users/S5-371-53RL/Documents"
>
```

En este caso, los rásteres resultantes y las gráficas de validación cruzada se guardaron en la carpeta **Documentos**:

Este equipo > Documentos

Nombre	Fecha de modificación	Tipo	Tamaño
RegTEMP_Tmin_1980-01-02	23/03/2017 4:55 p. m.	Archivo TIF	222 KB
RegTEMP_Tmin_1980-01-01	23/03/2017 4:54 p. m.	Archivo TIF	230 KB
RegTEMP_Cross_Validation_Plot_1980-01-02	23/03/2017 4:56 p. m.	Archivo PNG	14 KB
RegTEMP_Cross_Validation_Plot_1980-01-01	23/03/2017 4:55 p. m.	Archivo PNG	14 KB

Figura 20. Carpeta Mis Documentos en Windows mostrando los rásteres resultantes de la interpolación.

El archivo generado se encuentra en formato *.tif y es compatible con cualquier software SIG, se puede visualizar y procesar en R y contiene la información de temperaturas de forma espacial para el día y área interpolada.

8. Referencias bibliográficas

Abteilung Hydrometeorologie. (2013). REGNIE (Regionalisierte Niederschläge): Verfahrensbeschreibung & Nutzeranleitung. Offenbach: Deutscher Wetterdienst - DWD. 9 p.

Núñez Lopez, D., Treviño Garza, E., Reyes Gomez, V., Muñoz Robles, C., Aguirre Calderón, O., & Jiménez Pérez, J. (2014). Uso de modelos de regresión para interpolar espacialmente la precipitación media mensual en la cuenca del río Conchos. Revista mexicana de ciencias agrícolas, 5 (2), 201-213.

Rauthe, M., Steiner, H., U. Riediger, A., Mazurkiewicz, A., & Gratzki, A. (2013). A Central European precipitation climatology - Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS). Meteorologische Zeitschrift, 22(3), 235-256.