

# Interpolación en variables meteorológicas

Julián Camilo Builes, Daniel Santiago Bermudez Rivera, Daniel Camilo Fierro Garcia y Daniel Andrés Reyes Muñoz

**Abstract** — *In this document you will find a study of the behavior of the climatological variables of different stations in Brazil. Showing an analysis through interpolation methods, which will allow an estimation of the behavior of the chosen variables, analyzing the error that occurs in their calculation.*

**Key words**— Hermite, Spline, Interpolation

**Resumen**— *En este documento se va a encontrar un estudio del comportamiento de las variables climatológicas de diferentes estaciones en Brasil. Mostrando un análisis por medio de métodos de interpolación, lo cual va a permitir realizar una estimación del comportamiento de las variables elegidas, analizando el error que se presenta en su calculación.*

**Palabras clave**— Hermite, Spline, Interpolación

## I. INTRODUCCIÓN

Históricamente, las predicciones meteorológicas son fundamentales para planificar las actividades humanas. Actualmente en el mundo se existen muchos estudios del clima, ya que los datos que este puede arrojar representan gran importancia para evaluar diferentes situaciones, desde el momento más urgente hasta el momento más habitual como puede ser, plantar, cosechar, elegir el tiempo adecuado para construir un edificio o estudiar la crisis del cambio climático. En algunos países y regiones, estos servicios meteorológicos muestran una precisión asombrosa dando aproximaciones que han sido de gran utilidad.

Dentro de este artículo se va a enfocar principalmente en un estudio realizado en Brasil a la fecha de abril de 2012, mostrando las diferentes fluctuaciones en el ambiente, las cuales dependen de variables climatológicas como, la temperatura, humedad, presión atmosférica, entre otros. Para el estudio de estas variables, Brasil cuenta con diferentes estaciones climatológicas, ubicadas de forma estratégica para poder efectuar una evaluación correcta.

## II. DEFINICIÓN DEL PROBLEMA

Como bien se ha mencionado antes, tener información precisa de los datos de la variación del clima, es crucial para muchas actividades, actualmente existen muchas herramientas que permiten estimar con mayor precisión los valores de los cambios de las diferentes variables. Dentro este artículo se van a explicar dos posibles soluciones que llevan aun resultado aproximado. Estos son:

- El primero, consiste en seleccionar puntualmente una base meteorológica, y tomar una de las variables que

este registra. Con los datos ya escogidos sacar una muestra (el 80% o menos del total de los valores) y aplicar mínimo dos métodos de interpolación.

- Como segundo punto tratado en el artículo, se va a escoger un base y a partir de las distancias con las otras bases analizar cual de ellas es la más cercana geográficamente y así aplicar un método de interpolación para estudiar su comportamiento.

## III. MÉTODOS USADOS

Para realización de esta investigación se usó la herramienta Rstudio, la cual permite una fácil manipulación de tablas y archivos .csv o .xlsx, ya que los datos de estudio se encontraban almacenados en tablas de Excel, lo cuales contenían registros de las variables que obtienen las 10 estaciones meteorológicas. Algunas de estas variables son, el día juliano, hora, temperatura, humedad, presión, entre muchas otras.

Para realizar el estudio de estos datos, se usó el método matemático de interpolación, el cual tiene como objetivo construir nuevos puntos de datos, dentro de un rango de un conjunto discreto de puntos de datos conocidos. Permitiendo obtener una nueva sucesión de puntos o una función que pueda evaluarse en todos los puntos conocidos.

Dentro de las librerías que provee Rstudio, existen funciones que implementan diferentes algoritmos de interpolación, para este artículo se usaron tres métodos distintos con el fin de poder analizar cual es el que mejor se aproxima. Estas funciones son, *spline()* el cual evalúa en la interpolación por medio de spline cubico, el segundo *splinefun()* evalúa la interpolación de Hermite y el ultimo método es *aprox()* el cual permite evaluar por medio de una interpolación lineal.

Para evaluar los resultados que se obtienen, se va analizaron los errores, medio, máximo, mínimo y error medio cuadrático. Y para poder definir si los valores son aceptados se determinó un rango de aceptación, donde al menos el 95% de las interpolaciones tiene que corresponder al valor original para ser concluida como aceptada.

## IV. PROBLEMA 1

Para el análisis de este problema se escogió la base que denominada Fortaleza-UECE la cual consta de 703 datos registrados por cada variable. Dentro de las estas se tomó como estudio la variable **humedad relativa** mostrando el

resultado que se ilustra en la imagen 1, esta muestra la variabilidad de la humedad con la toma del 100% de los datos originales.

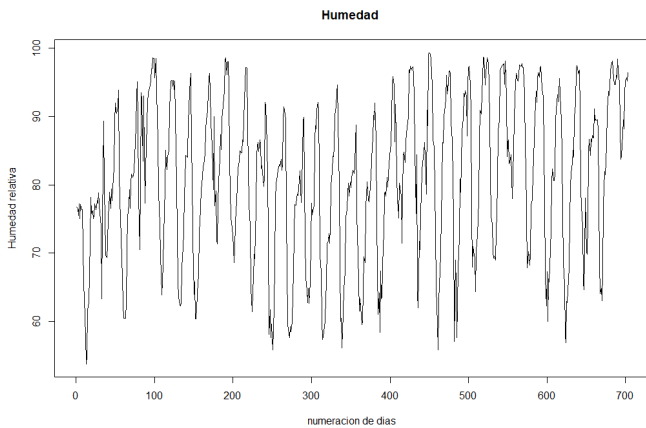


Imagen 1: totalidad de los datos de humedad.

A partir de la información anterior, se tomo una muestra del 80% de los datos, lo cuales son 562 registros para aplicar las interpolaciones, esto se hizo de forma aleatoria para no sesgar la información, como puede ser el caso de que se eliminaran el último o primer 20% de datos lo cual dañaría por completo el análisis. Ya con la muestra seleccionada se aplicaron los métodos mencionados, obteniendo un resultado de 703 cifras interpoladas por cada método, permitiendo así sobreponer y comparar los resultados, los cuales se explican a continuación.

- Para el caso de **spline** se obtuvo los datos de la imagen dos, que como se puede observar, la proyección de la interpolación sobre la gráfica de los datos originales no es de todo consistente debido a el método que busca aproximar de una forma cuadrática valores que son estocásticos, es decir que varían de forma aleatoria y dependen de la probabilidad.

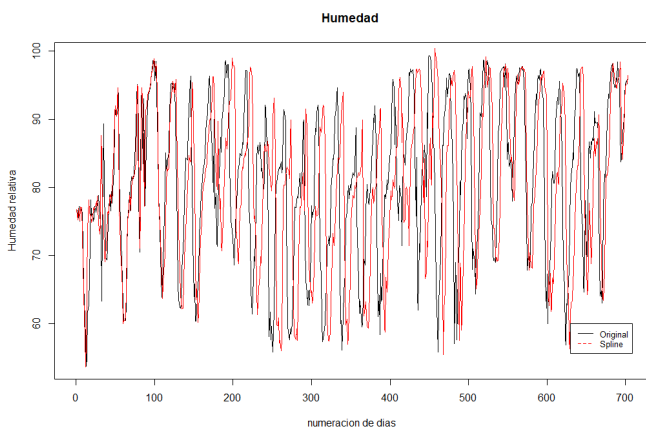


Imagen 2: Spline vs Original

Con los datos obtenidos se realizo el calculo de los direntes errores, los cuales se encuentran en la **Imagen 3**, estos muestran un error relativamente

grande, lo cual explica efecto en el desfase que se evidencia en la **Imagen 2**.

```
> cat("error EMC del spline ",emcS,'\n')
error EMC del spline 13.26743
> cat("error medio del spline ",(sum(eSpline))/tam,'\n')
error medio del spline 10.15884
> cat("error minimo del spline ",min(eSpline),'\n')
error minimo del spline 0
> cat("error maximo del spline ",max(eSpline),'\n')
error maximo del spline 40.08896
```

Imagen 3: Errores del Spline

- El método lineal, a diferencia del método anterior muestra un resultado as simple vista que es un poco más aproximado, sin embargo, podemos evidenciar en el error (**Imagen 5**) que no difieren mucho con respecto al método del **spline()** (**Imagen 2**).

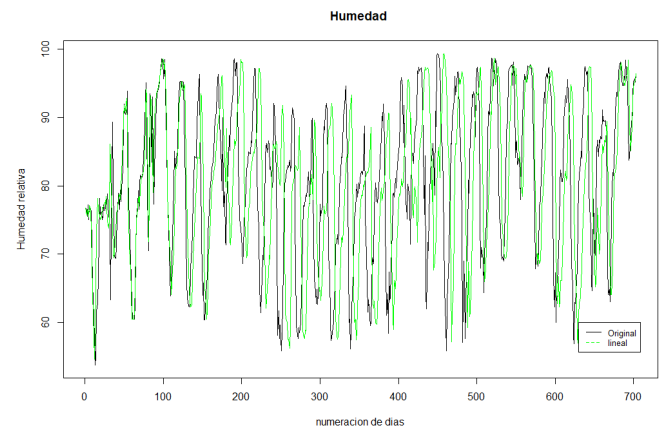


Imagen 4: Lineal vs Original

```
error EMC del lineal 13.11293
> cat("error medio del lineal ",(sum(el))/tam,'\n')
error medio del lineal 10.05017
> cat("error minimo del lineal ",min(el),'\n')
error minimo del lineal 0
> cat("error maximo del lineal ",max(el),'\n')
error maximo del lineal 39.44556
```

Imagen 5: Error método Lineal

- Para el caso del splinefun o el de Hermite, realmente los datos se sobreponen casi en su totalidad a los datos originales (**imagen 6**), lo cual nos dice que esta muestra una mejor aproximación a diferencia de los métodos anteriores, lo cual también se evidencia mostrando un error significativamente menor en comparación con los otros dos (**imagen 7**).

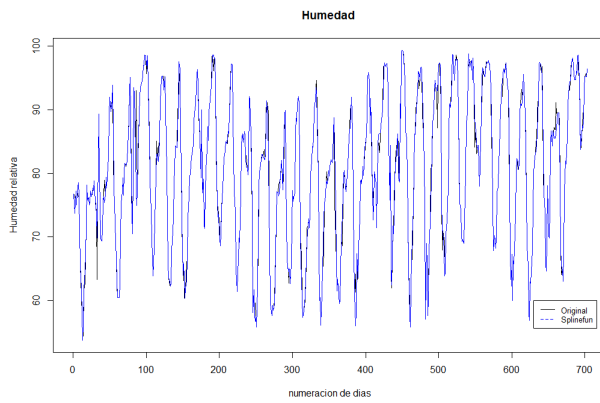


Imagen 6: Splinefun vs Humedad original

```
> cat("error EMC del hermite ",emcC,'\n')
error EMC del hermite 1.355292
> cat("error medio del hermite ",(sum(ec))/tam,'\n')
error medio del hermite 0.4385104
> cat("error minimo del hermite ",min(ec),'\n')
error minimo del hermite 0
> cat("error maximo del hermite ",max(ec),'\n')
error maximo del hermite 16.03784
```

Imagen 7: Error Hermite

Ya concluyendo la solución a este problema, se puede llegar a concluir que el método que mejor aproxima los datos es el splinefun, mostrando un error medio mucho mas pequeño que los otros dos métodos utilizados, también evaluando el error medio cuadrático (EMC) se puede llegar a concluir que la fluctuación del error durante el periodo de la señal de la humedad registrada en el de tiempo dado es pequeño en comparación con los otros dos, lo cual evidencia que hay una aproximación por parte del spline y el método lineal de un 71.1207% y 71.526% de coincidencia respectivamente. Ya por parte del splinefun muestra una consistencia de los datos de un 98.121%, lo cual hace que sea el único método que esta dentro del rango de aceptación del criterio establecido.

## V. PROBLEMA 2

Este problema es muy similar al anterior, debido a que se aplicaron métodos de interpolación para encontrar la estimación del comportamiento de la variable de temperatura interna de la estación de Quixeramobim, pero este tiene la pequeña diferencia, de que los datos que se toman son de la estación que se encuentra más cercana a la escogida, los datos de las 10 estaciones se encuentran en la **tabla 1**.

Estaciones	X	Y
Vicosa do	41,094	3,5683
Ceara	55	033
Sobral	40,344	3,6606
	765	054

Sao Goncalo do	38,979	3,6747
Amarante	811	505
Santa Quiteria	40,149	4,3329
	933	534
Santana do	39,733	7,1830
Cariri	388	59
Quixeramobim	39,283	5,2007
	696	7
Quixada	-	-
	39,03337	4,967000
	4	1
Pentecoste	39,272	3,7956
	307	568
Jati	38,997	7,6823
	926	8
Itatira	39,635	4,5289
	985	383

Tabla 1: Coordenadas de las estaciones

Partiendo de las localizaciones geográficas como se visualiza en la **Imagen 8**, se utilizó el método de distancia cuadrática para encontrar aquella estación más cercana a la de Quixeramobim. Los resultados de este cálculo se visualizan en la **Tabla 2**, mostrando que la estación más próxima es la de Quixada, la razón por la cual se toma la distancia más cercana es porque se busca obtener el resultado más próximo a los verdaderos valores obtenidos en la estación de Quixeramobim.

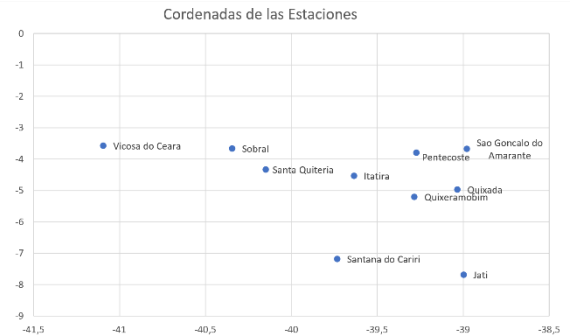


Imagen 8: localización de las estaciones climatológicas

Estaciones	Distancias
Vicosa do Ceara	2,438060442
Sobral	1,870286963
Sao Goncalo do Amarante	1,555982605
Santa Quiteria	1,226161335
Santana do Cariri	2,03265661
Quixeramobim	0
Quixada	0,342504804
Pentecoste	1,405159344
Jati	2,498009729

Tabla 2: Distancias de Quixeramobim

Para la solución se hicieron uso de las mismas tres funciones de interpolación, usadas en el problema anterior, mostrando los siguientes resultados:

- Aplicando el método de **Spline** se obtuvo la gráfica que se visualiza en la **imagen 9**. Se observa semejanza considerable entre los valores originales y los interpolados, prácticamente tienen el mismo valor lo cual indica un error considerablemente bajo (**imagen 10**). Dentro de los errores se puede observar que, aunque no son tan próximos a 0 realmente muestran valores que se podrían clasificar dentro de un rango de aceptación visible al ojo humano.

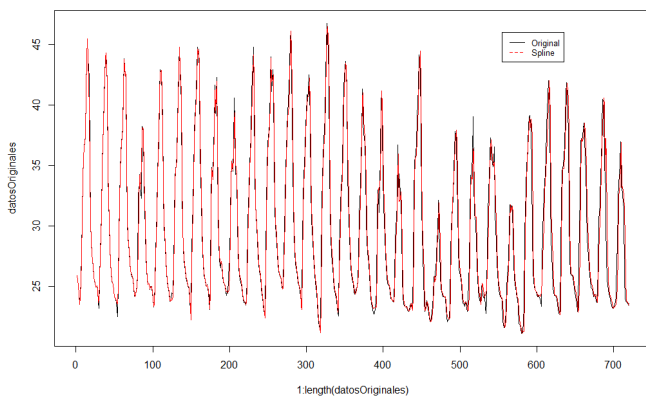


Imagen 9: Spline vs Temperatura original

```
> cat("error EMC del spline ",emcS,'\n')
error EMC del spline 1.343141
> cat("error medio del spline ",(sum(eSpline))/tam,'\n')
error medio del spline 0.8091102
> cat("error minimo del spline ",min(eSpline),'\n')
error minimo del spline 0
> cat("error maximo del spline ",max(eSpline),'\n')
error maximo del spline 8.039371
```

Imagen 10: Error del Spline

- Al utilizar el método lineal, se obtuvo la gráfica que se logra ver en la **imagen 11**, esta grafica es muy parecida a la obtenida por medio el método de spline, con una semejanza considerable entre los datos interpolados y los originales. Esto se hace evidente al observar el poco error producido (**imagen 12**).

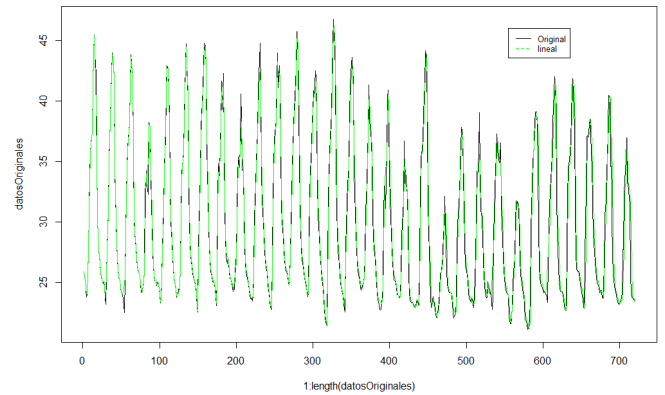


Imagen 11: Lineal vs Temperatura original

```
> cat("error EMC del lineal ",emcL,'\n')
error EMC del lineal 1.323655
> cat("error medio del lineal ",(sum(eL))/tam,'\n')
error medio del lineal 0.8181022
> cat("error minimo del lineal ",min(eL),'\n')
error minimo del lineal 0
> cat("error maximo del lineal ",max(eL),'\n')
error maximo del lineal 7.48879
```

Imagen 12: Error Spline

- Realizando la interpolación de Spliefun o Hermite se obtuvo una gráfica en donde los valores interpolados se sobreponen casi en su totalidad sobre los valores originales (**Imagen 13**), los resultados se evidencian en la imagen, debido a la precisión del método, el error evaluado (**Imagen 14**) es mínimo en comparación con los con los métodos evaluados anteriormente.

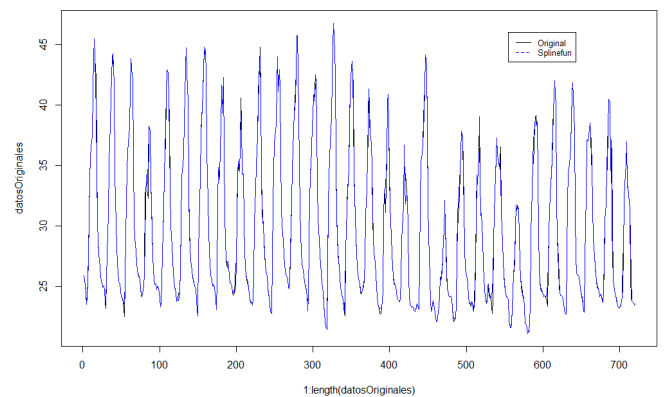


Imagen 13: Splinefun vs Temperatura original

```
> cat("error EMC del hermite ",emcC,'\n')
error EMC del hermite 0.537771
> cat("error medio del hermite ",(sum(ec))/tam,'\n')
error medio del hermite 0.1503838
> cat("error minimo del hermite ",min(ec),'\n')
error minimo del hermite 0
> cat("error maximo del hermite ",max(ec),'\n')
error maximo del hermite 5.785645
```

Imagen 14: Error Splinefun

Lo mencionado en este problema lleva a concluir, que los tres métodos muestran una aproximación precisa y aceptable dentro del rango de aceptación establecido dentro de este documento, el Spline y el lineal se muestran muy iguales, indicando que tienen una parecido a los valores originales que son del 97.289% y del 97.258% de exactitud respectivamente, para el caso del Splinefun se obtuvo un valor prácticamente perfecto ya que su coincidencia es del 99.495% lo cual lo hace el mejor método para aplicar a los datos, pero aun así los otros dos también hacen parte del rango de aceptación ya que superan el 95% de precisión .

## VI. CONCLUSIONES

- Se puede llegar a concluir que el efecto de los errores en los dos puntos a pesar de que fueran evaluados con los mismos métodos, se debe a el tipo de dato evaluado, en el caso del primer problema se evidencia un mayor ruido en los datos de humedad originales, mostrando una gran cantidad de fluctuaciones en cortos intervalos de tiempo, a diferencia del segundo problema que toma datos de la temperatura, los cuales se muestran con una mayor suavidad, permitiendo que las funciones puedan arrojar un valor más aproximado.
- El espacio geográfico en donde se realizan las predicciones climatológicas para el segundo problema nos permite realizar supuestos de cómo es su comportamiento en zonas cercanas, a lo largo de esta investigación se evidenció, por medio de la interpolación el comportamiento de las variables pertinentes al estudio, definiendo así una tendencia prácticamente idéntica. De este modo concluimos que las ubicaciones de las estaciones en donde se analizaron las temperaturas poseen una geografía prácticamente similar, así mismo el comportamiento del clima no difiere en gran medida.
- Se puede decir, en general, que el método splinefun es el más adecuado para resolver los problemas relacionados con el clima, ya que este método se adapta muy bien a las fluctuaciones que le ocurre en los registros de este tipo, y se evidencia en ambos problemas analizados, que este fue el que mejor resultados mostro, proporcionando la mejor aproximación a los datos que ya se habían recolectado. Los demás métodos, sin embargo, muestran que también son una forma bastante confiable para realizar este tratamiento a los datos, aunque en general no arrojan la mejor semejanza con los datos originales, teniendo un poco más de error, en comparación con el splinefun.

## VII. BIBLIOGRAFÍA

- [1] J. L. De la Fuente O'conor, Ingeniería de los algoritmos y metodos numericos, Editoria círculo Rojo, 2017.
- [2] R. L. Burden y J. Douglas Faires, Numeriacal Analysis, Canada: Brooks/cole.