Reto : Interpolación

Diana Sofía Carrillo   Nelson Alejandro Mosquera
Camilo José Narvaez   Sergio Esteban Triana

Pontificia Universidad Javeriana
{ds-carrillog, nelson.mosquera, camilonarvaez, triana_se}@javeriana.edu.co

October 29, 2021

Contents

List of Figures

1 Introducción

Dado un conjunto de valores asociados a variables climáticas, que están indexados en el tiempo y en el espacio se busco determinar numéricamente los valores de la variable Y cada media hora en una estación de monitoreo seleccionada, utilizando interpolación ó ajuste de curvas. También se determinaron los valores de la variable Y cada hora en una estación de monitoreo, utilizando los datos de una estación cercana.

Se leyeron los 720 filas de datos de la estación meteorológica para después usar los métodos de spline interpolación cubica, approx interpolación lineal y pchip interpolación cubica.

2 Metodología

2.1 Métodos Numéricos

2.1.1 Cubic Spline

Una spline cúbica es una spline construida por polinomios de tercer orden por partes que pasan por un conjunto de m puntos de control, gracias a esto se vuelve mas precisa y rápida que otros métodos. La segunda derivada de cada polinomio se establece comúnmente en cero en los puntos finales, ya que esto proporciona una condición de frontera que completa el sistema de ecuaciones $m-2$. Esto produce una denominada spline cúbica "natural" y conduce a un sistema tridiagonal simple que puede resolverse fácilmente para dar los coeficientes de los polinomios. Sin embargo, esta elección no es la única posible, y en su lugar se pueden utilizar otras condiciones de contorno.[1]

A partir de n puntos (x, y):

$$a = x_0 < x_1 < ...x_n = b; = S(x)$$

satisface:

- $S(x) \in C^2[a, b]$
- Para cada subintervalo$[x_{i-1}, x_i], S(x), existeunpolinomiodegrado3.$

- $S(x_i) = y_i$ para todo i

$$M[i][j] = \begin{cases} C_1(x);\ x_0 < x < x_1 \\ C_i(x);\ x_{i-1} < x < x_i \\ C_n(x);\ x_{n-1} < x < x_n \end{cases}$$

Donde cada $C_i = a_i x + b_i x + c_i x^2 + d_i x^3$, con $d_i \neq 0$[2]

2.1.2 Piecewise Cubic Hermite Interpolation

Siendo $x_0, ..., x_n$ puntos distintos.Conociendo los valores de la función f y su derivada f' en estos mismos puntos, se busca un polinomio de grado el menor posible que coincida con f y con su derivada en los puntos señalados.

Existe un polinimio que interpola los puntos de la función que pasa por dichos, de grado $2n + 1$, dicho polinomio se le llama polinomio de interpolación de Hermite de f en los puntos x_i.Sin embargo, este método se puede extender considerando valores de derivadas de la función de orden mayor que uno.[3]

2.1.3 Índice de Jaccard

El índice de Jaccard, también conocido como coeficiente de similitud de Jaccard, es una estadística que se utiliza para comprender las similitudes entre conjuntos de muestras. La medición enfatiza la similitud entre conjuntos de muestras finitas y se define formalmente como el tamaño de la intersección dividido por el tamaño de la unión de los conjuntos de muestras [4]. La representación matemática del índice se escribe como:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

2.1.4 Error relativo

con el fin de garantizar la calidad de una medida, el error relativo corresponde a el cociente entre el error absoluto y el valor exacto.

$$E_r = \frac{|valor\ verdadero -\ \ valor\ apoximado|}{Valor\ verdader}$$

2.1.5 Interpolación lineal

La interpolación lineal consiste en estimar la ubicación de un punto dentro de un intervalo numérico, suponiendo que los valores extremos de dicho intervalo están unidos por una recta. Conocida la ecuación de esta recta, es posible ubicar el punto desconocido. También se puede aproximar la curva que une a los puntos dados mediante una función cuadrática u otro polinomio. Sin embargo la recta tiene la ventaja de su sencillez matemática, por lo cual resulta fácil de manejar, aunque siendo la interpolación más simple de todas, es posible que el resultado no sea tan preciso como el que se obtiene al emplear otras funciones.[5]

2.2 Diagramas de Flujo

Los siguientes diagramas de flujo corresponden a la parte 1 y parte 2 del ejercicio. Su desarrollo se llevó acabo con el objetivo de planificar la implementación, por lo que es posible la existencia de pequeñas adiciones en este proceso no mencionadas en las figuras.

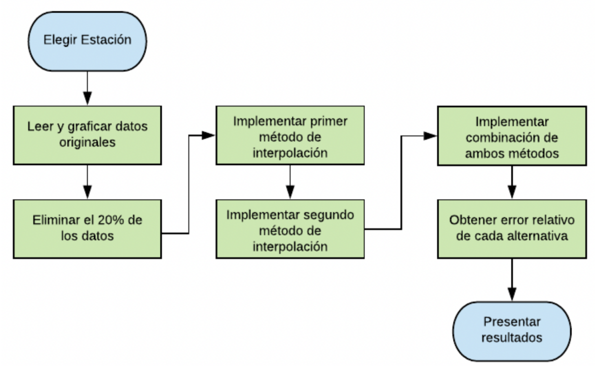

Figure 1: Diagrama de flujo Parte 1.

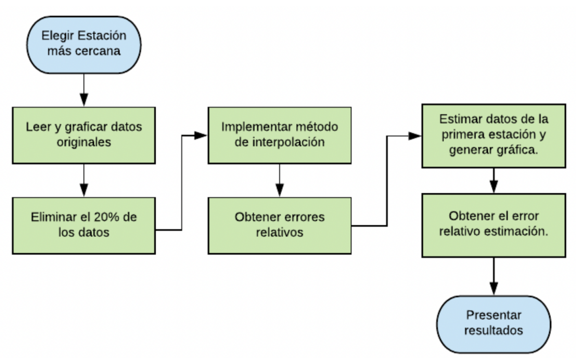

Figure 2: Diagrama de flujo Parte 2.

3 Implementación y Análisis de Resultados

Para la implementación iniciamos con una investigación de métodos de interpolación en Python y en R. A partir de la anterior se decidió hacer los programas necesarios para ambas partes en R usando las funciones "spline", "approx" y "pchip"; siendo la primera una función de interpolación cúbica, la segunda de interpolación lineal y la tercera, de nuevo, de interpolación cúbica.

3.1 Primera Parte

A fin del desarrollo de esta primera parte elegimos la estación Itatira, la cual cuenta con 720 conjuntos de datos (filas).

3.1.1 Implementación

Iniciamos leyendo los datos de esta estación presentes en el documento de excel suministrado y guardando la tabla completa en una variable llamada "temp", para esto hicimos uso de la librería "readxl". En esta misma variable, modificamos el tipo de dato de las columnas "Día Juliano" y "Hora" a numérico, además, dividimos los valores de la última entre 10000. Lo anterior fue con el objetivo de tener una nueva columna ("DH", por día y hora) que contuviera un número que correspondiera al día de forma entera y a la hora de forma decimal.

Continuamos ilustrando los valores iniciales en una gráfica (Figura 3), guardando en las variables 'x' y 'y' los valores de las columnas "DH" y "Temp. Interna (ºC)" respectivamente.

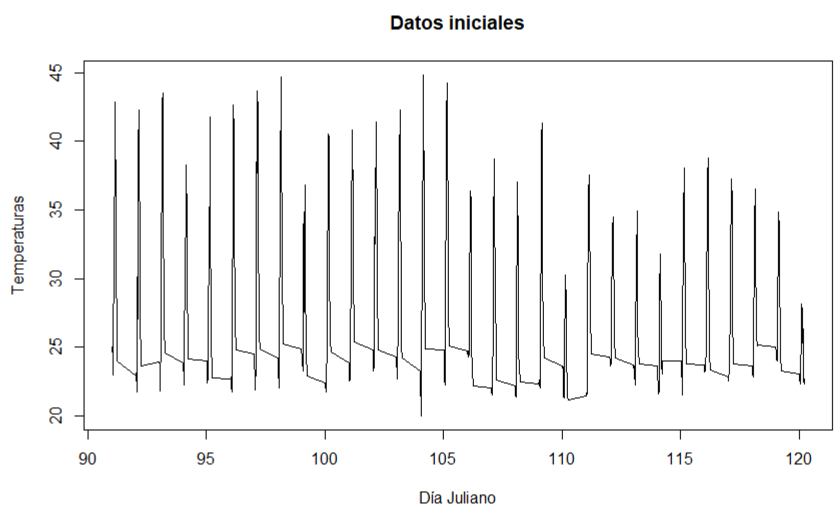

Figure 3: Datos iniciales de temperatura vs día de la estación Itatira.

Para probar la efectividad de los métodos de interpolación, eliminamos el 20% de datos del conjunto inicial y proseguimos con la implementación de cada uno de los algoritmos. A continuación se ilustran los datos obtenidos por interpolación en contraste con los datos iniciales(Figuras 4, 5, 6).

A objetivo de una mejor comparación, en la Figura 7 se puede observar los datos iniciales en contraste con los datos de todos los métodos de interpolación utilizados.

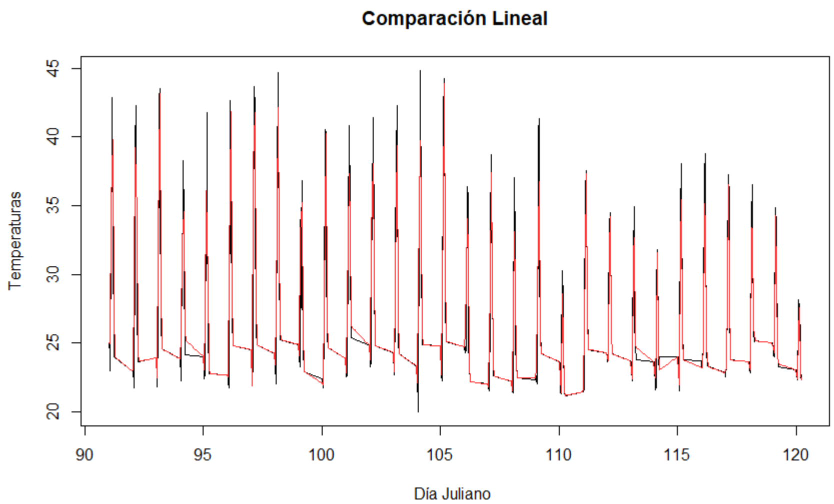

Figure 4: Datos iniciales vs Datos por Interpolación Lineal.

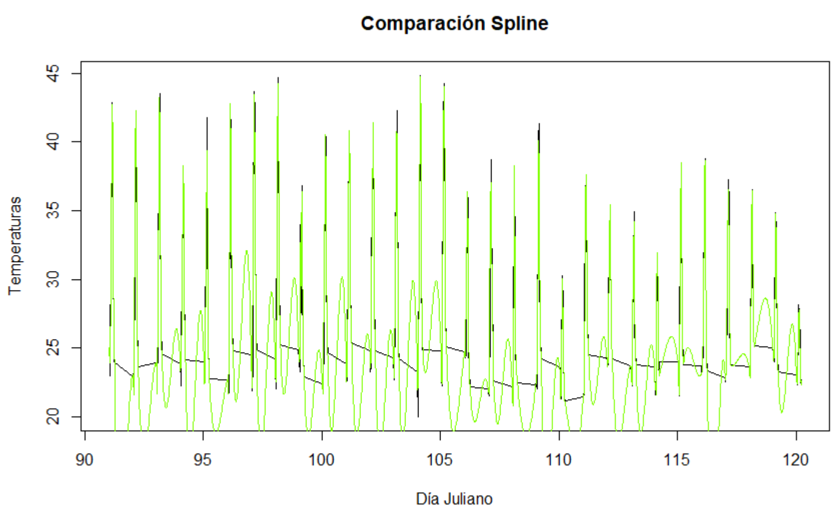

Figure 5: Datos iniciales vs Datos por Interpolación Spline.

3.1.2 Resultados

Siguiendo con los resultados, utilizamos los métodos numéricos error relativo e índice de Jaccard para evaluar el desempeño de cada uno de los métodos de interpolación. Los resultados se evidencian en las siguientes tablas:

	Spline	Approx	PCHIP
Numero de Errores	102	120	102
Error Maximo	0,25	0,34	0,28
Error Minimo	0,01	0,01	0,01
Error Medio	0,02	0,02	0,03
Indice Jaccard	0,8583	0,83333	0,8583

Table 1: Resultados con el 80 por ciento de los datos

	Spline	Approx	PCHIP
Numero de Errores	240	234	212
Error Maximo	0,37	0,38	0,33
Error Minimo	0,01	0,01	0,01
Error Medio	0,02	0,03	0,03
Indice Jaccard	0,6667	0,675	0,7056

Table 2: Resultados con el 60 por ciento de los datos

	Spline	Approx	PCHIP
Numero de Errores	383	372	355
Error Maximo	0,46	0,38	0,33
Error Minimo	0,01	0,01	0,01
Error Medio	0,03	0,03	0,03
Indice Jaccard	0,4681	0,4833	0,5069

Table 3: Resultados con el 40 por ciento de los datos

	Spline	Approx	PCHIP
Numero de Errores	540	NA	523
Error Maximo	30,81	NA	17,44
Error Minimo	0,01	NA	0,01
Error Medio	0,06	NA	NA
Indice Jaccard	0,25	NA	0,2736

Table 4: Resultados con el 20 por ciento de los datos

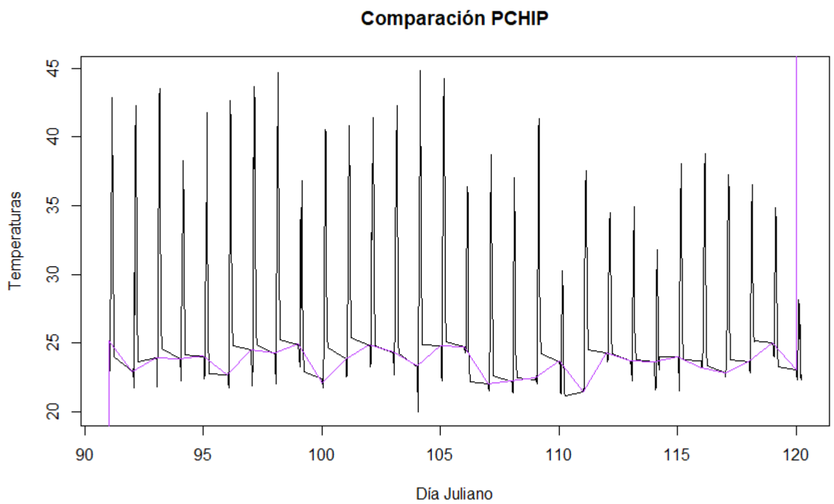

Figure 6: Datos iniciales vs Datos por Interpolación PCHIP.

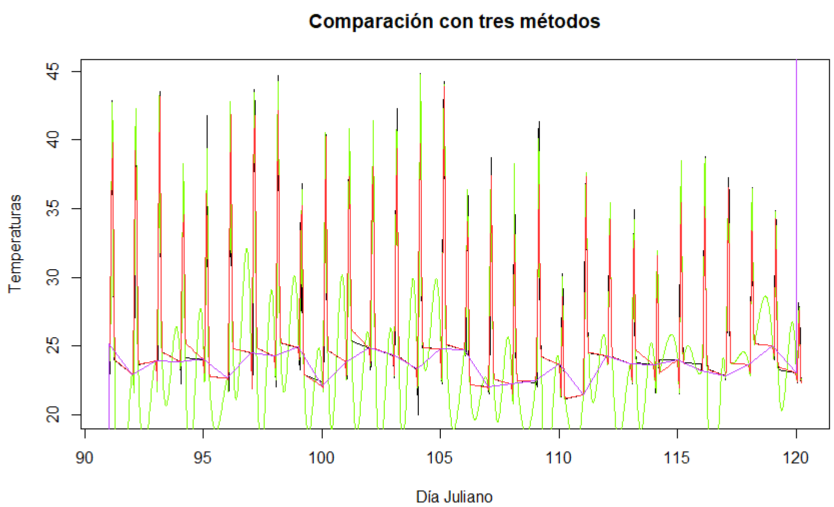

Figure 7: Datos iniciales vs Datos por las tres interpolaciones.

3.2 Segunda Parte

3.2.1 Implementación

Para la Segunda parte del reto, La implementación inicia con una lectura de datos desde el archivo excel (Base de Datos) por medio de la función readexcel la cual esta implementada en la librería readxl, de la estación de Itatira. a continuación de esto se realiza un procedimiento para hallar y guardar las horas y los días julianos. Seguido de esto realizamos la lectura de las estaciones a evaluar las cuales son Santa Quitaria y Pentecoste y se le realiza el mismo procedimiento que a la estación de itatira. Seguido del anterior paso, se realiza un algoritmo para eliminar el 20 por ciento de los datos con el fin de aplicarle seguido de esto la función spline para hacer la interpolación. También utilizamos la función Splinefun y round para redondear los valores obtenidos, para finalmente hacer el plot y tener como resultado la gráfica de la interpolación.

3.2.2 Resultados

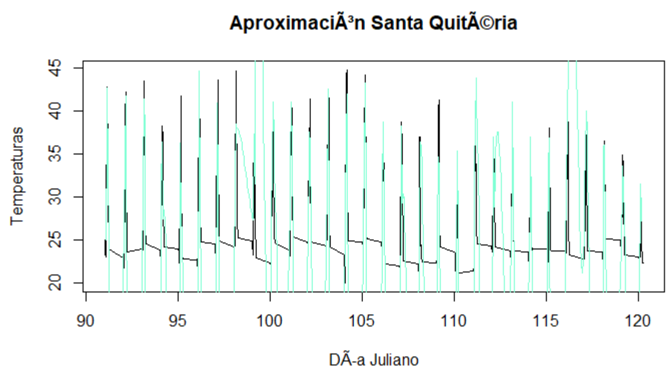

Figure 8: Datos iniciales vs Datos por Interpolación Lineal.

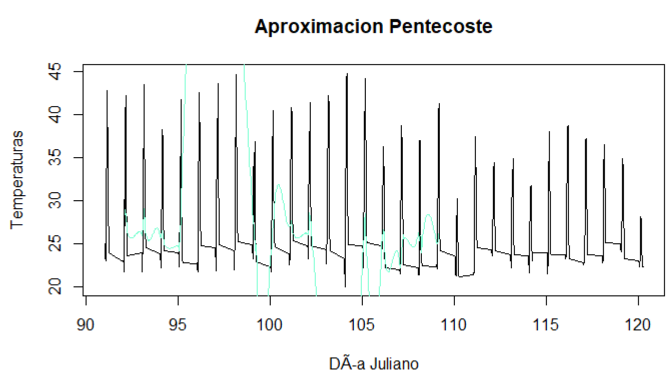

Figure 9: Datos iniciales vs Datos por Interpolación Lineal.

Siguiendo con los resultados, utilizamos los métodos numéricos error relativo e índice de Jaccard para evaluar el desempeño de cada uno de los métodos de interpolación. Los resultados se evidencian en las siguiente tabla:

	Sante Quiteria	Pentecoste
Numero de Errores	717	718
Error Maximo	29532332	235992,8
Error Minimo	0,01	0,01
Error Medio	10660,15	4404,81
Indice Jaccard	0,0042	0,0028

Table 5: Errores e Indice Jaccard

4 Conclusiones

Podemos concluir de este reto que los métodos de interpolación tienen una aplicación importante en cosas cotidianas, además que su efectividad sera definida por factores como la eficiencia del algoritmos, los datos, los intervalos. se evidencia que hay una función de interpolación mas eficiente que el resto la cual es PCHIP ya que tiene unos mejores resultados con diferentes cantidades de datos. Finalmente podemos ver por el índice de jaccard que en la primera parte se va reduciendo la similitud a medida que se reducen los datos y a pesar que las gráficas sean muy parecidas no tienen un grado significativo de similitud en la segunda parte,

References

[1] Weisstein, Eric W. "Cubic Spline." From MathWorld–A Wolfram Web Resource. https://mathworld.wolfram.com/CubicSpline.html " [Online].

[2] Cubic Spline Interpolation. (2021, March 25). Wikiversity. Retrieved 14:04, March 25, 2021 from https://en.wikiversity.org/w/index.php?title=Cubic_Spline_Interpolation&oldid=2270598.

[3] Wikipedia contributors. (2021, September 13). Cubic Hermite spline. In Wikipedia, The Free Encyclopedia. Retrieved 03:10, October 29, 2021, from https://en.wikipedia.org/w/index.php?title=Cubic_Hermite_spline&oldid=1044149380

[4] "Jaccard Index", DeepAI. [En línea]. Disponible en: https://deepai.org/machine-learning-glossary-and-terms/jaccard-index#: :text=Breakingtfind0. [Accedido: 09- Nov- 2020].

[5] "Interpolación lineal: fórmulas, cómo hacerla, ejemplos, ejercicios", Lifeder, 2021. [Online]. Available: https://www.lifeder.com/interpolacion-lineal/. [Accessed: 21- Oct- 2021].