



Estudiantes

Miguel Angel Avila Santos
Juan Andres Martinez Amado
Jorge Luis Esposito Albornoz
Juan Sebastian Herrera Guaitero

Asignatura

Análisis numérico

Profesora

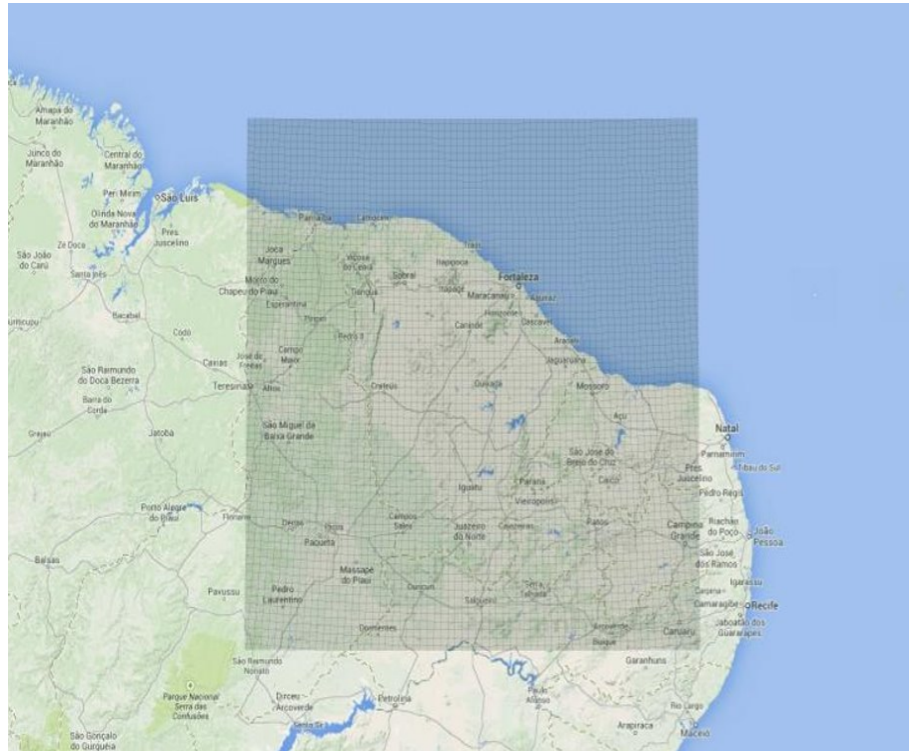
Eddy Herrera Daza




Reto 1

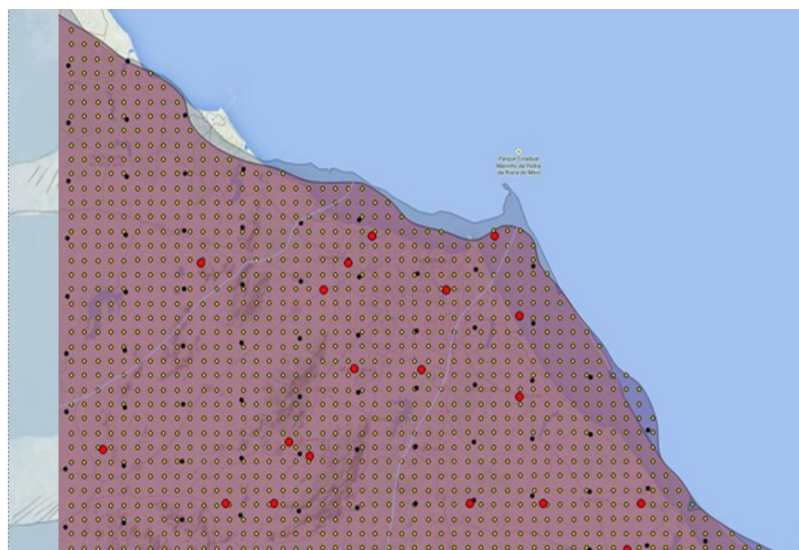
29 de Octubre de 2021

Descripción del problema:

Se proporcionó una base de datos en excel de 17 estaciones climáticas ubicadas cerca a la zona de Fortaleza, Brasil, estas hojas de cálculo poseen variables climáticas como lo son la temperatura, la presión atmosférica, la humedad, entre otras. Dado estos conjuntos de datos indexados en tiempo y espacio se realizaron dos puntos donde se empleó el método de interpolación, el cual tiene como objetivo la obtención de nuevos puntos de una sucesión, o conjunto de ellos, que obedece a algún patrón determinado.



-  Estación Monitoreo
-  Radar Meteorológico
-  Satélite



Primer punto

Se realizó interpolación utilizando el 70% de los datos aleatoriamente escogidos de la estación Aiuaba, esta última escogida tras analizar la integridad de los datos de diferentes estaciones en la variable de Umidade Relativa do Ar 2m (%), la cual fue la variable elegida para este punto, y validar que la estación presentaba valores dentro de un rango de 20 y 100, los cuales coinciden con los valores esperados de la variable, esta integridad es más evidente cuando se contrasta los valores de la estación de Aiuaba con los valores presentados en la estación de Acaraú, pues estos últimos incluyen en repetidas mediciones valores de -6999, los cuales no corresponde a un valor esperado de la variable y los cuales podrían representar complicaciones en el desarrollo de las interpolaciones deseadas.

Para el desarrollo del punto se hizo uso del lenguaje R y la herramienta RStudio, la cual posee librerías y funciones las cuales facilitan: la manipulación de la base de datos proporcionada, la aplicación de la interpolación, la visualización de los resultados obtenidos, el cálculo de errores, entre otros.

Métodos de interpolación utilizados:

Interpolación de Splines Cúbicos con la función `spline()`:

El spline cúbico ($k=3$) es el spline más empleado, debido a que proporciona un excelente ajuste a los puntos tabulados y su cálculo no es excesivamente complejo.

Sobre cada intervalo $[t_0, t_1], [t_1, t_2], \dots, [t_{n-1}, t_n]$, S está definido por un polinomio cúbico diferente.

Sea S_i el polinomio cúbico que representa a S en el intervalo $[t_i, t_{i+1}]$, por tanto:

$$S(x) = \begin{cases} S_0(x) & x \in [t_0, t_1) \\ S_1(x) & x \in [t_1, t_2) \\ \vdots & \vdots \\ S_{n-1}(x) & x \in [t_{n-1}, t_n) \end{cases}$$

Interpolación de tipo Hermite con la función `splinefun()`:

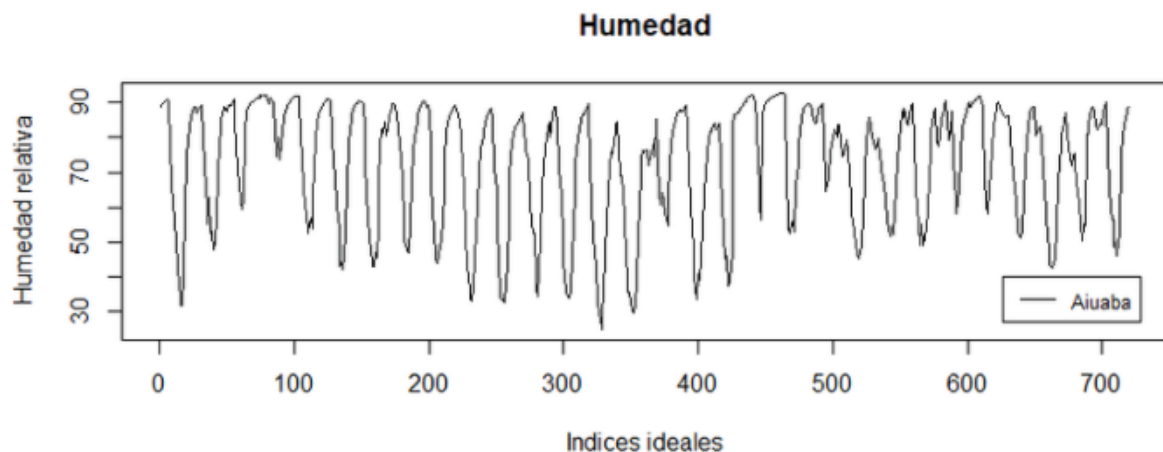
En el análisis numérico, la interpolación de Hermite, nombrada así en honor a Charles Hermite, es un método de interpolación de puntos de datos como una función polinómica. El polinomio de Hermite generado está estrechamente relacionado con el polinomio de Newton, en tanto que ambos se derivan del cálculo de diferencias divididas.

Interpolación Lineal con la función approx() :

La interpolación lineal es un caso particular de la interpolación general de Newton. Con el polinomio de interpolación de Newton se logra aproximar un valor de la función $f(x)$ en un valor desconocido de x . El caso particular, para que una interpolación sea lineal es en el que se utiliza un polinomio de interpolación de grado 1, que se ajusta a los valores en los puntos x_1 y x_2 . Se denota de la siguiente manera:

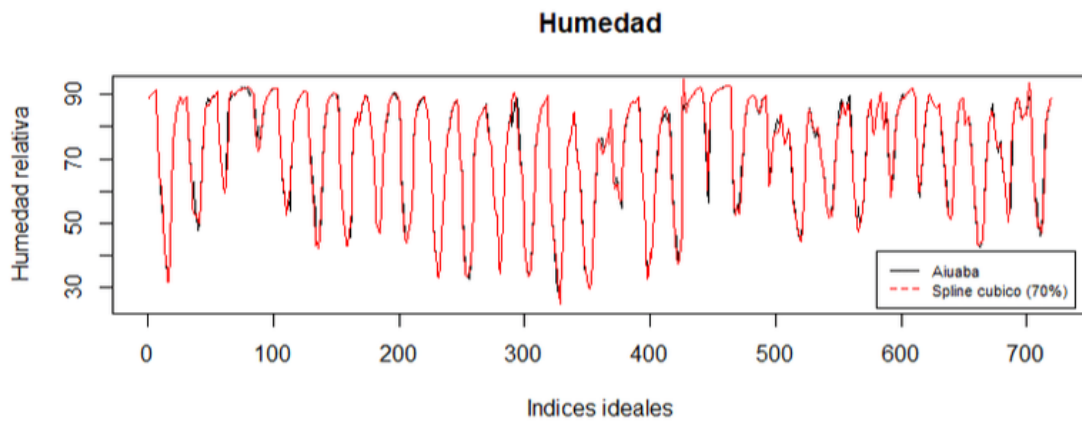
$$f(x|x_1; x_2) = f(x_1) + \frac{f(x_2) - f(x_1)}{(x_2 - x_1)}(x - x_1)$$

Resultados:

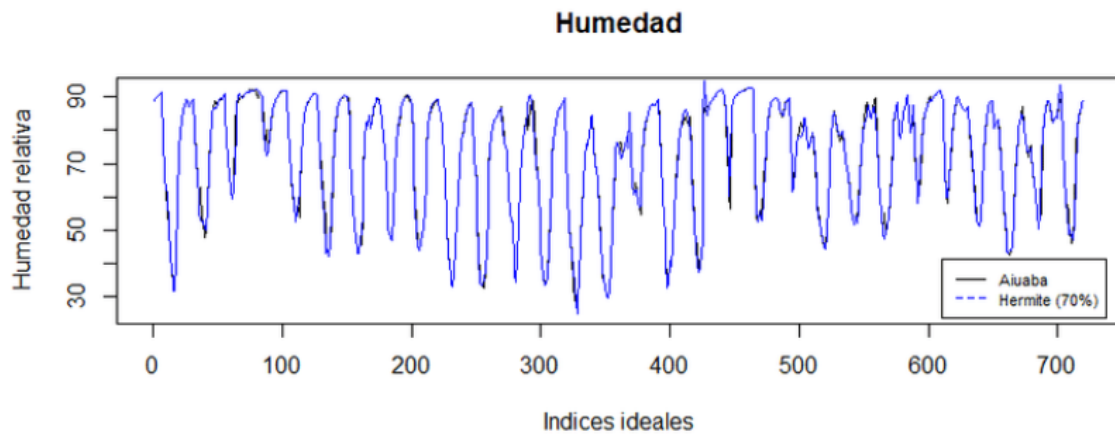


Estos son los datos originales graficados de Aiuaba tomando la columna de humedad relativa.

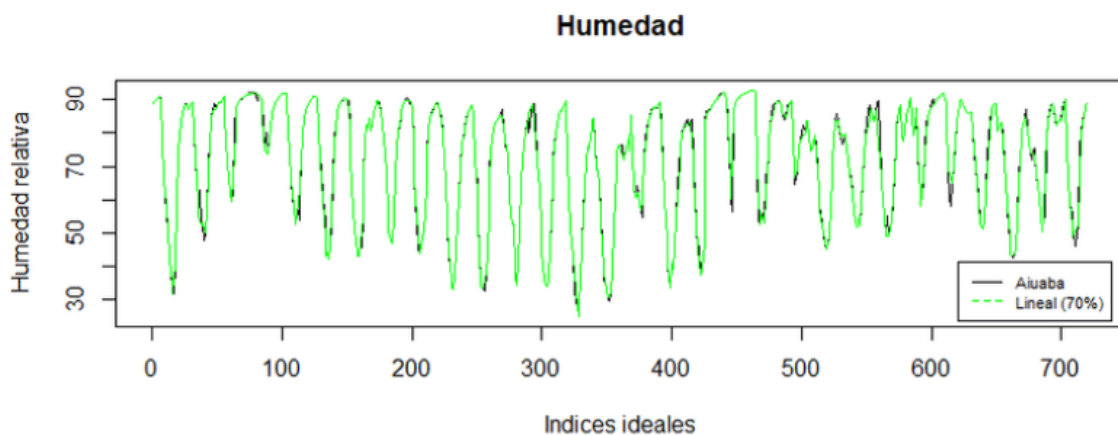
Se escogieron de manera aleatoria el 70% de los datos, como se explicó inicialmente, asegurando que en la muestra aleatoria se encuentre el primero y último dato original, y tras aplicar los métodos de interpolación seleccionados se obtuvieron los siguientes resultados:



Este es el resultado de implementar Spline Cúbico, al haber escogido el 70% de los datos originales, pero se ve que en varios puntos los que pertenecen al 30% no seleccionado la interpolación no fue acertada.



El resultado de implementar Hermite, casualmente y pese a ser métodos y funciones distintas de R, fueron iguales a los obtenidos con la interpolación por Spline Cúbicos.



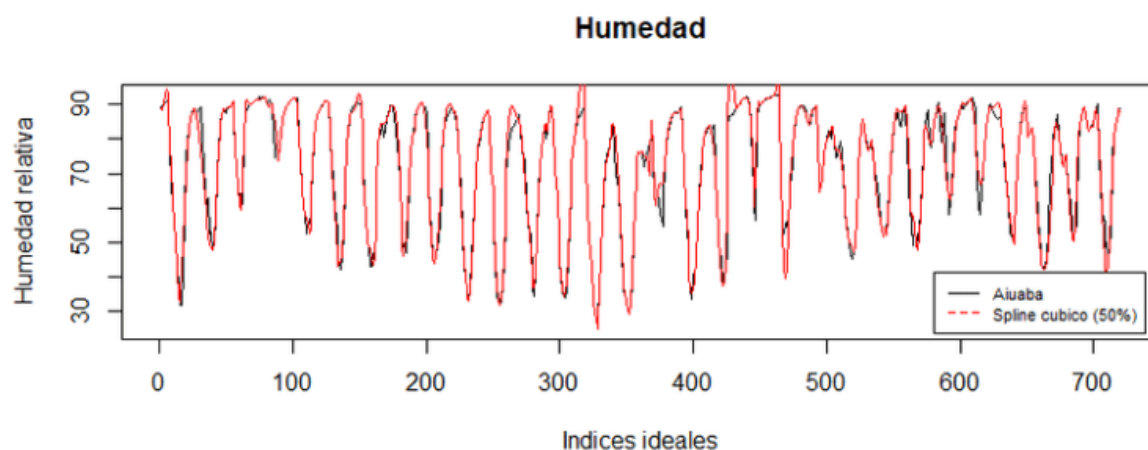
Este es el resultado al implementar interpolación lineal, se ve a simple vista que esta falla mucho más en los puntos del 30% por ciento a comparación de los otros dos métodos anteriormente utilizados.

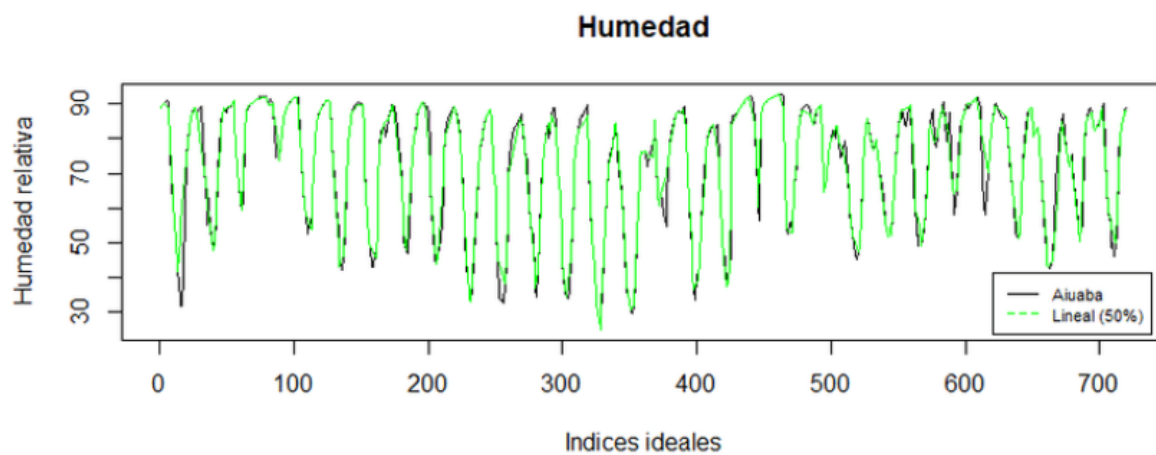
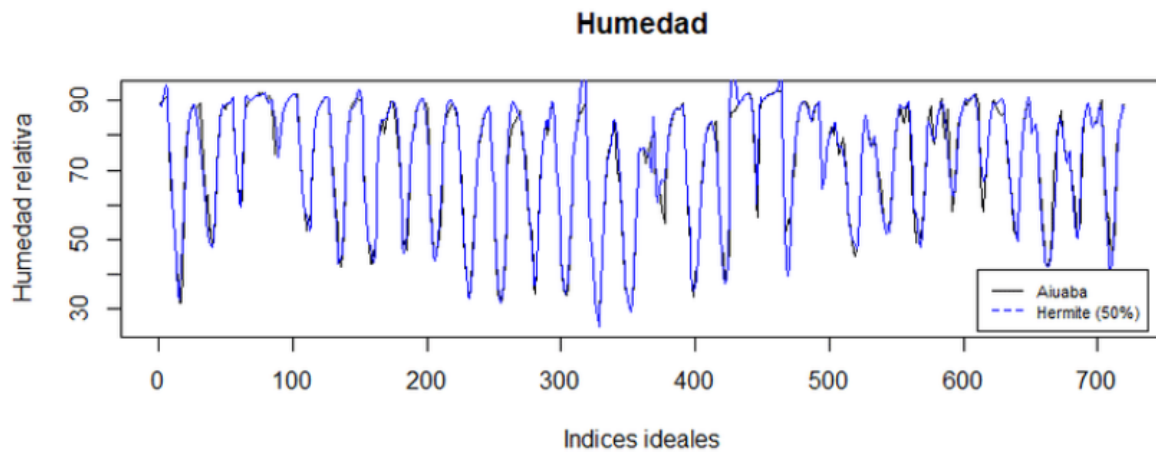
Métricas 70%	Error mínimo	Error máximo	Error absoluto medio	Error cuadrático medio	Índice de Jaccard
Spline cúbico	0	15.55	0.64	1.79	23.96%
Interpolación Hermite	0	15.55	0.64	1.79	23.96%
Interpolación Lineal	0	17.24	0.68	184	22.58%

Al evaluar los errores como el mínimo, el máximo, el error medio absoluto y el error cuadrático. Se deduce, que todos los métodos en algún dato del 30% interpolado es igual al valor de los datos originales. En el error máximo se ve un valor mayor en la interpolación lineal en comparación con los otros dos métodos. Igualmente, esto se evidencia tanto en el error absoluto medio y en el error cuadrático donde el spline cúbico y Hermite presentan errores medios menores.

Para el índice de Jaccard, el cual mide el grado de similitud entre dos conjuntos, sigue evidenciado la superioridad de los dos primeros métodos, con índices del 23,96%. Esto se calculó manejando un rango de error del 0,5, esto quiere decir que al momento de calcular un error si este es menor a ese dato, se considera como igual y por ende perteneciente a ambos conjuntos.

Con el fin de comparar estos resultados y probar en diferentes condiciones los métodos, se repitió el procedimiento con el 50% de los datos aleatoriamente escogidos, asegurando de nuevo que en la muestra aleatoria se encuentre el primero y último dato original. Estos fueron los resultados obtenidos:



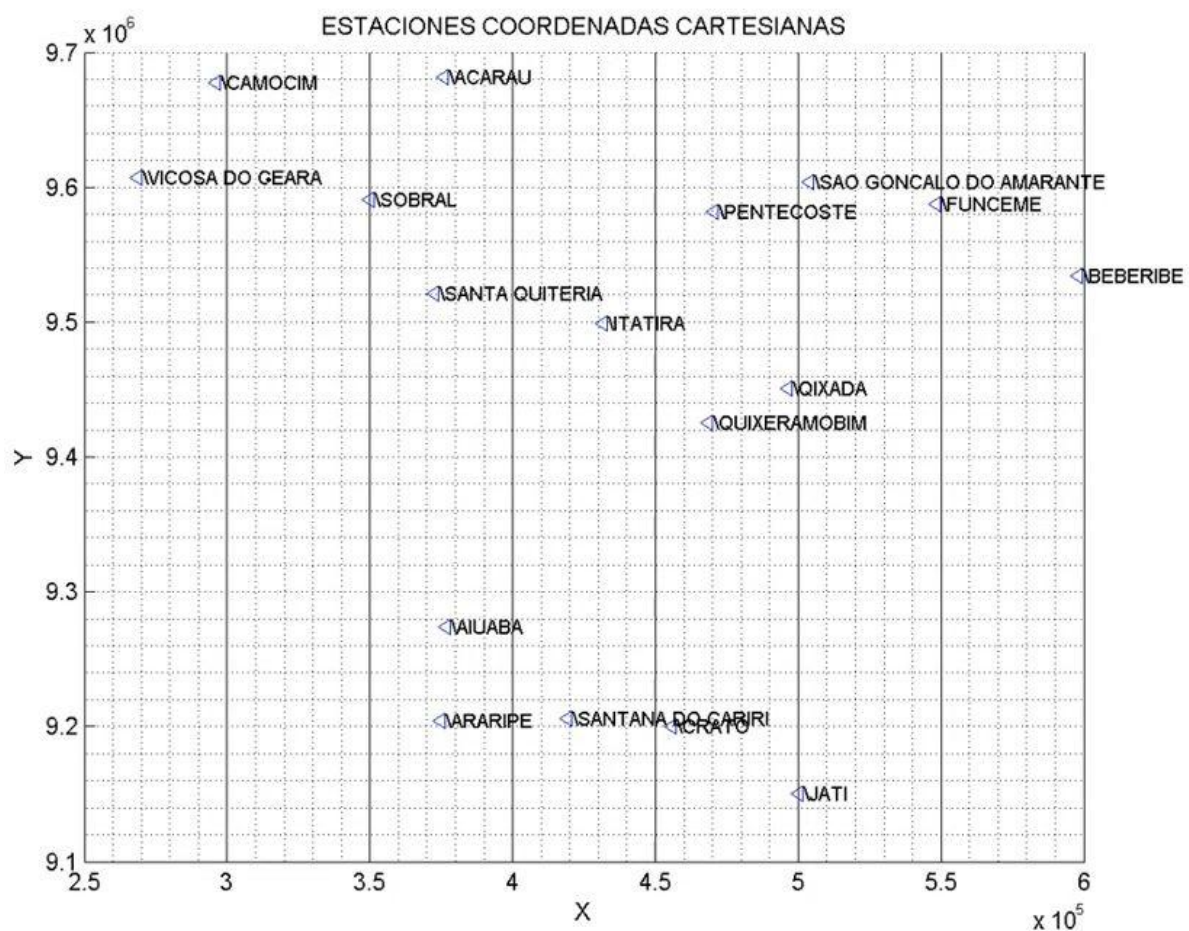


Métricas 50%	Error mínimo	Error máximo	Error absoluto medio	Error cuadrático medio	Índice de Jaccard
Spline cúbico	0	16.37	1.57	3.18	18.61%
Interpolación Hermite	0	16.37	1.57	3.18	18.61%
Interpolación Lineal	0	28.3	1.6	3.47	16.94%

En los resultados a comparación de la muestra del 70%, se evidenció gráficas más imprecisas, el error aumentó en todos los casos, bajo el porcentaje de similitud de Jaccard, por lo que se puede concluir que a menor cantidad de datos frente a los mismos datos reales (originales), la precisión de los métodos de interpolación de spline cúbico, hermite y lineal disminuyen.

Segundo Punto:

Se realizó interpolación utilizando los datos de las estaciones cercanas a la estación de Crato, esta última escogida tras analizar la integridad de los datos de las diferentes estaciones en la variable de Temp, do Ar 2m(°C), la cual fue la variable elegida para este punto, se validó que la estación y sus estaciones cercanas presentaban valores dentro de un rango de 10 y 40, los cuales coinciden con los valores esperados de la variable, está integridad es más evidente cuando se contrasta los valores de la estación de Crato con los valores presentados en la estación de Acaraú, pues estos últimos incluyen en repetidas mediciones valores menores a -39, los cuales no corresponde a un valor esperado de la variable y los cuales podrían representar complicaciones en el desarrollo de las interpolaciones deseadas. Para poder definir las dos estaciones más cercanas a Crato, se hizo una aproximación visual a las coordenadas en base al siguiente plano cartesiano y tomando como referencia el centro del triángulo a la izquierda del nombre de cada estación:



Las coordenadas aproximadas se muestran en la siguiente tabla:

Estación	Coordenada en X (10 ⁵)	Coordenada en Y (10 ⁶)
Camocim	2.96	9,675
Acarau	3.76	9,681
Vicosa do Gera	2.68	9,607
Sao Goncalo do Amarante	5.04	9,603
Sobral	3.50	9,591
Pentecoste	4.71	9,582
Funceme	5.48	9,588
Santa Quiteria	3.73	9,521
Beberibe	4.98	9,535
Itatira	4.32	9,498
Qixada	4.97	9,450
Quixeramobim	4.68	9,427
Aiuaba	4.77	9,272
Araripe	3.74	9,204
Santana do Cariri	4.19	9,206
Crato	4.56	9,198
Jati	5.00	9,150

En base a estas coordenadas calculamos la distancia de Crato al resto de estaciones utilizando Excel y con la siguiente fórmula:

Sean dos puntos sobre el plano cartesiano, $P_1(x_1, y_1)$ y $P_2(x_2, y_2)$. La distancia que hay entre ellos viene dada por la siguiente expresión:

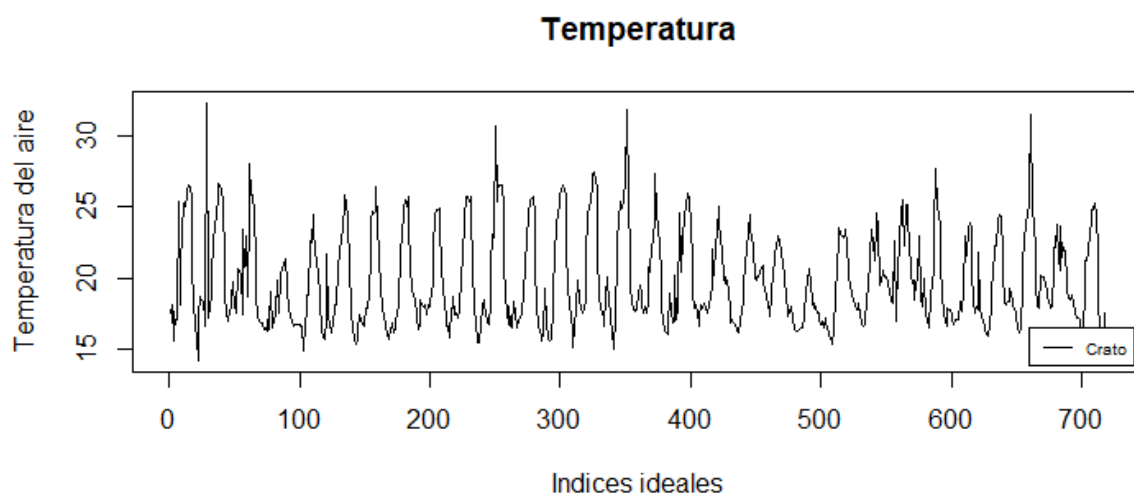
$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Los resultados de los cálculos se pueden observar en la siguiente tabla.

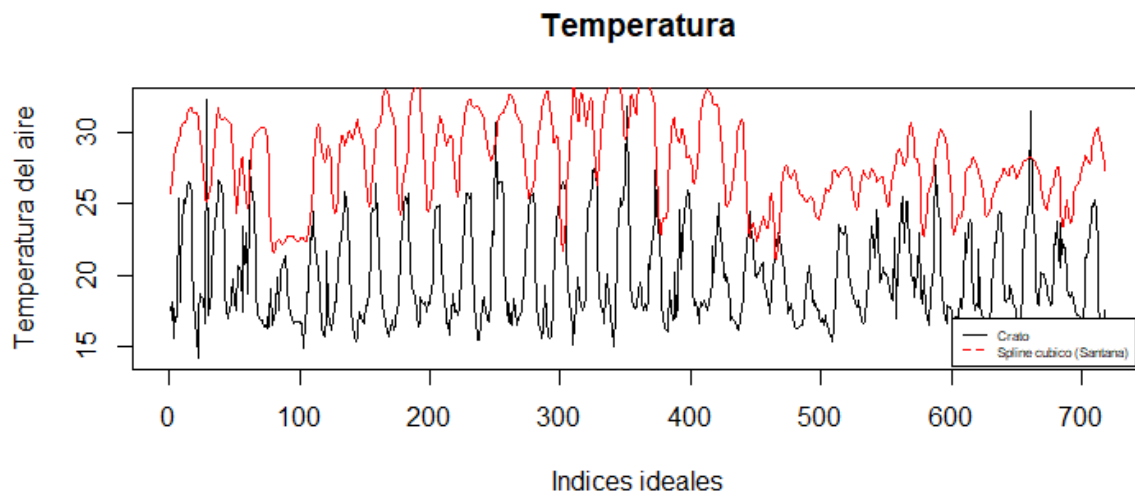
Estación	Distancia
Camocim	5.03119270E+05
Acarau	4.89580433E+05
Vicosa do Gera	4.50138867E+05
Sao Goncalo do Amarante	4.07834525E+05
Sobral	4.07044224E+05
Pentecoste	3.84292857E+05
Funceme	4.00704380E+05
Santa Quiteria	3.33493628E+05
Beberibe	3.39607126E+05
Itatira	3.00958469E+05
Qixada	2.55313533E+05
Quixeramobim	2.29314195E+05
Aiuaba	7.69220385E+04
Araripe	8.22192192E+04
Santa do Cariri	3.78549865E+04
Crato	0.00000000E+00
Jati	6.51152824E+04

En base a la tabla, las dos estaciones más cercanas a la estación de Crato son; primero Santana do Cariri y segundo Jati.

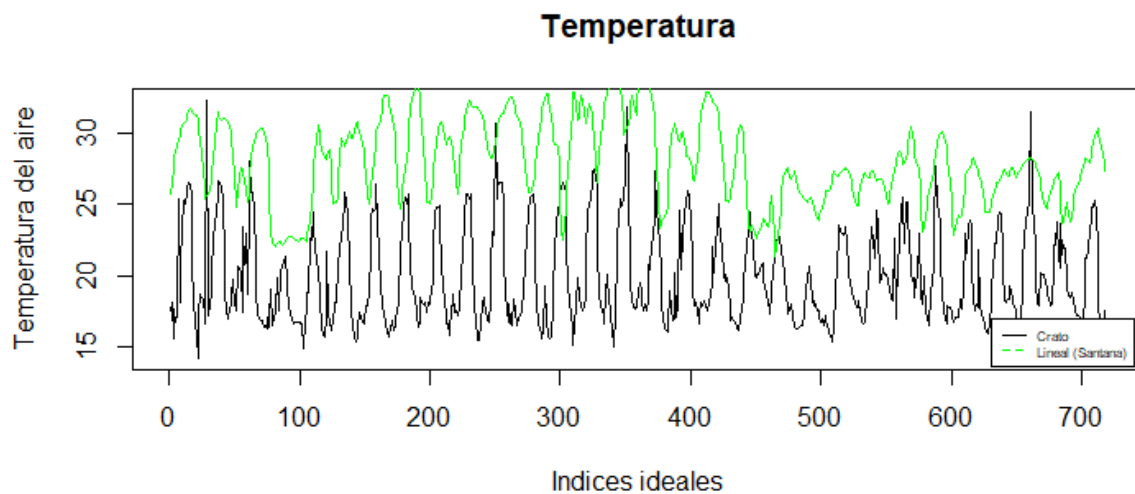
Al igual que en el primer punto para el desarrollo del punto se hizo uso del lenguaje R y la herramienta RStudio.



Esta gráfica inicial corresponde a los datos de temperatura de la estación Crato, la cual será usada como base para el desarrollo del punto.



Teniendo ya los datos de Crato se procede a realizar interpolación, en este caso del tipo spline Cúbico, para comparar los datos obtenidos en la estación Santana do Cariri y ver sus coincidencias con Crato.



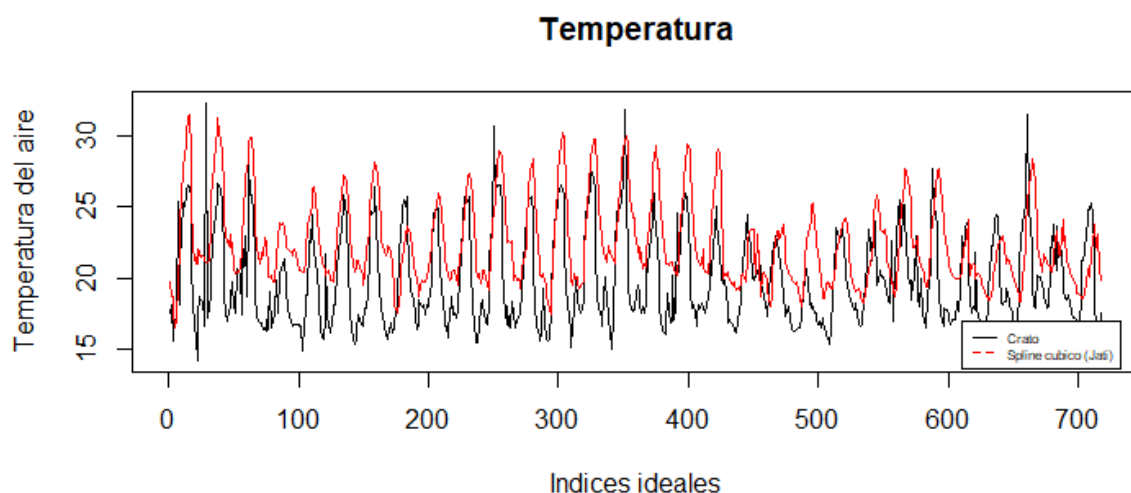
De igual manera se realiza el mismo procedimiento, con la excepción de que en esta ocasión la interpolación es del tipo lineal.

Métricas Estación Santana	Error mínimo	Error máximo	Error absoluto medio	Error cuadrático medio	Índice de Jaccard
Spline cúbico	0.06	18.9	8.2	9.18	0.97%
Interpolación Lineal	0.01	18.87	8.18	9.14	1.39%

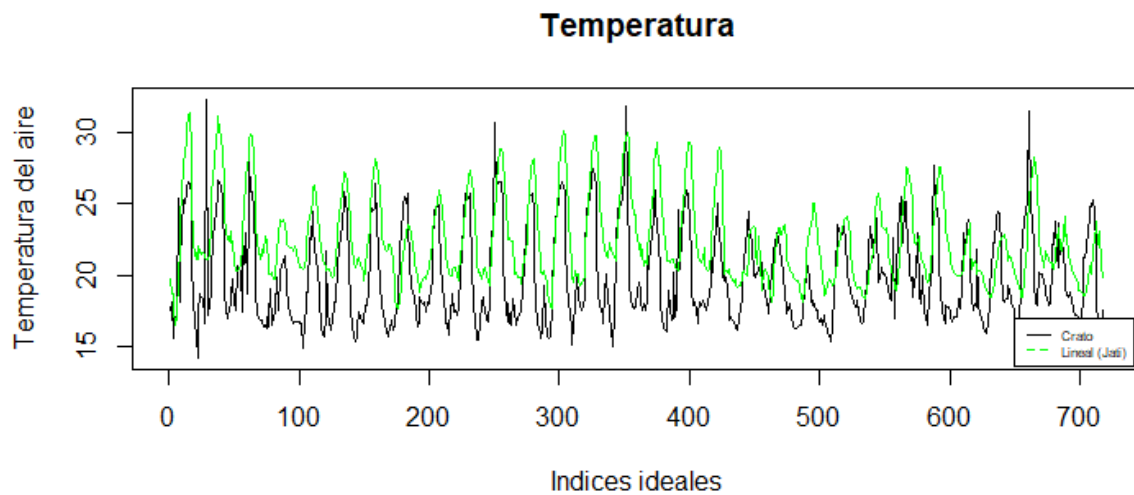
A diferencia de lo que se podría pensar, al observar las métricas calculadas, comprobamos que en este caso fue más precisa la interpolación lineal con respecto a la cúbica. De igual manera, se destaca que la diferencia entre ambas es realmente pequeña.

Con respecto al desarrollo del punto 2 con la estación de Santana do Cariri es importante mencionar que, a pesar de que los valores de la interpolación no dan exactamente los mismos valores que los valores de Crato, si se realizara un desplazamiento, se podría tener una mejor aproximación, esto puede ser evidenciado en la primera parte de la interpolación de Spline Cúbico, donde de 0 a 400, aproximadamente, los valores son los mismos pero con un desplazamiento hacia arriba, por lo que el método de interpolación está sobreestimando los datos, esta sobrestimación se puede ver en los errores por 8.2 grados, el error medio, aunque cabe aclarar que en otros datos de la gráfica, con el desplazamiento, los valores no serían exactos, hasta el punto de tener un comportamiento inverso como en los datos alrededor de 450 aproximadamente.

A continuación procedemos a realizar el mismo procedimiento pero utilizando otra estación para comparar con Crato, en este caso la estación utilizada es Jati.



En la gráfica se realizó un spline cúbico y se pueden observar resultados más cercanos a los obtenidos en Crato.



Por último, se repitió el procedimiento en esta ocasión con una interpolación lineal y de igual manera se logra observar la similitud tanto en resultados como en la gráfica.

Métricas Estación Jati	Error mínimo	Error máximo	Error absoluto medio	Error cuadrático medio	Índice de Jaccard
Spline cúbico	0	11.8	3.15	3.7	7.24%
Interpolación Lineal	0	11.17	3.14	3.69	7.24%

A diferencia del caso Crato-Santana, las métricas calculadas nos muestran una mayor coincidencia en los datos, viéndose reducidos todos los errores y observándose un aumento del índice de Jaccard.

De igual manera cabe mencionar, que en ambos casos se mantiene la sobreestimación de los métodos respecto a la temperatura, pero con una menor cantidad de grados en Jati frente a Santana. Pese a lo anterior, se puede concluir que para la estimación de la temperatura de la estación de Crato, utilizar la información de sus estaciones vecinas, Santana do Cariri y Jati, no es suficiente.