

Proyecto Análisis Numérico

Métodos de interpolación

Estación Meteorológica

Región Ceará

Nicolás Camacho Plazas Daniela Cortes Antonio
Mateo Florido Sanchez

Octubre 10, 2019

1. Descripción del problema planteado

A partir de la información sumistrada, indicadores meteorológicos de Brasil tales como la temperatura, humedad relativa, velocidad de viento y presión atmosférica , para realizar el proyecto del curso.Sin embargo, para evidenciar que método de interpolación se ajusta mejor para cada tipo de datos faltantes, sacando ciertas filas y así los compararlas con los datos reales y así evidenciar la precisión de cada uno de los métodos.// Adicionalmente, se podrá evidenciar la ubicación de las estaciones por medio de un mapa satelital así como las curvas de nivel para las estaciones seleccionadas.

2. Informe del estado de la base de datos

2.1. Notación de los datos

En un principio los datos se encontraban en archivo que proporcionó la profesora (Abril 2013 (1 em 1 hora).xsl), debido al formato de la fecha que se encontraba en día juliano se realizó la conversión a formato de fecha gregoriano, esto por medio de fórmulas y concatenaciones en excel, lo que nos generó 8 archivos en formatos .CSV UTF-8, es con el fin de facilitar el importe de los datos y la creación de los dataframes.

Cada uno de los archivos contiene la información de cuatro estaciones (Araripe, Fortaleza, Itatira, Quixada), esta fueron seleccionadas ya que contenían datos consistentes en la mayoría de las columnas lo que nos permitía comprobar el error de forma más precisa.
Para cada una de las estaciones tomamos las siguientes columnas, las cuales cada medida se encontraba en una métrica específica:

1. **La temperatura interna:** Se encontraba en grados centígrados con cuatro cifras significativas (dos decimales).
2. **Humedad relativa del aire:** Se encontraba en porcentaje, tomada a 2 metros del suelo, con tres cifras significativas (tres decimales).
3. **Velocidad del viento:** Fue tomada a 10 metros del suelo en $\frac{m}{s}$, con cuatro cifras significativas (tres decimales).
4. **Presión atmosférica:** Se encuentra en hectopascales (hPa), con cinco cifras significativas (dos decimales).

Decidimos tomar estas columnas de valores ya que los datos eran completos y sólidos, su porcentaje en fallos (-6999) y de ceros en los datos obtenidos y ecogidos es del 0 por ciento, ideal para interpolar.

2.2. Herramientas utilizadas

- **Excel:** Utilizamos la versión 19.0 para la revisión y conversión de los datos y poderlos exportar a los dataframes.
- **Jupyter-Python:** Utilizamos un servidor con linux en el que montamos un entorno Jupyter-Python versión 4.7.4 (Interactive python), esto permite guardar todos los datos en el server para no tener que cargar los datos para cada prueba, adicionalmente utilizamos las siguientes librerías:
 - Pandas 0.25.1: Permite realizar el análisis de datos de gran volumen, agrupando esta en mapas para no generar un overhead al realizar la interpolación de los mismos.
 - Numpy 1.17.2: Librería numérica de python con acceso a los algoritmos de interpolación y transformadas, con implementaciones eficientes y ajuste de precisión.
 - Matplotlib 3.1.1 : Es una librería que permite graficar en python, que provee una API orientada a objetos.
 - SciPy 1.3.1: Es una librería utilizada para realizar algunas pruebas con distintos métodos de interpolación dados sus métodos fáciles de usar y eficientes.
- **QGIS:** En su versión 3.8.3 Zanzibar (Software libre), es un sistema de información geográfica con soporte a inverse distance weighting, el cuál fue el método utilizamos para realizar la interpolación de la temperatura con los datos de las cuatro estaciones. Así como se observa en la Figura 1, se encuentra la imagen de ubicación del satélite consultado para realizar la interpolación.

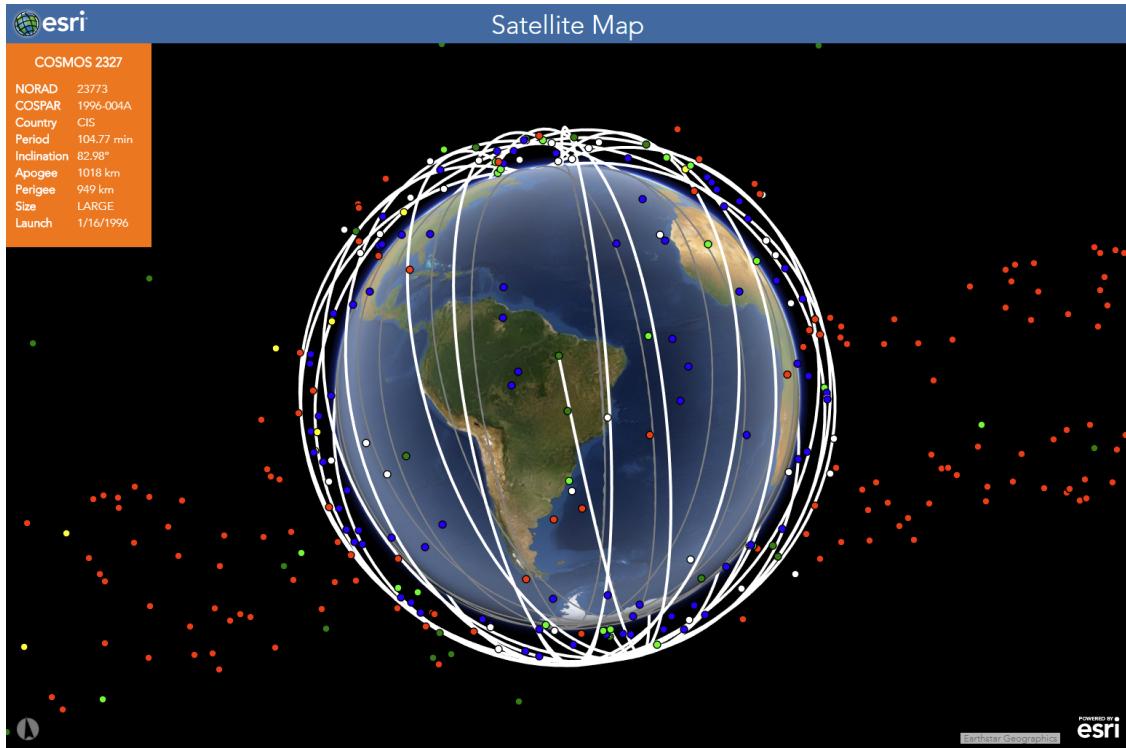


Figura 1: Información satélite

3. Componente Teórico: Métodos Interpolación

3.1. Lagrange Polynomial Interpolation

En primer lugar implementamos el método de interpolación polinómica de Lagrange, tenemos en un principio un conjunto de puntos el cual le falta información, por tanto por medio de estos puntos y del método antes mencionado construimos un polinomio el cual nos ayude a encontrar la información faltante del conjunto de datos, para esto utilizamos la fórmula para obtener el polinomio interpolador de Lagrange y así mismo sus bases polinómicas. Para un conjunto de n puntos (x_i, f_i) , tenemos sus bases Polinómicas :

$$l_i(x) = \frac{\prod_{k=0; k \neq i}^n (x - x_k)}{\prod_{k=0; k \neq i}^n (x_i - x_k)}$$

Polinomio resultante :

$$p(x) = \sum_{j=0}^n f_j l_i(x)$$

Sin embargo partiendo de esta fórmula y un número pequeño de n, sin embargo para un gran cantidad de datos obtenemos lo siguiente:

1. La evaluación del polinomio $p(x)$ requiere $O(n^2)$ multiplicaciones y adiciones.(Lo que requiere una gran capacidad computacional)

2. Al añadir una nueva pareja de puntos requiere un nuevo calculo desde cero.
3. El cálculo del polinomio es numericamente inestable.

3.2. Barycentric Lagrange Interpolation

Es una variante del método estándar de Lagrange por polinomios, y así mismo es un método más estable y rápido. Se realiza con el fin de evaluar y recalcular el polinomio resultante en $O(n)$ operaciones. Partiendo del numeradores de la base polinómica $l_i(x)$ siendo dividida por $(x - x_j)$, definimos los pesos baricentricos de la siguiente forma:

$$w_j = \frac{1}{\prod_{k=1}^n (x_j - x_k)}$$

Esto se puede reescribir como $w_j = \frac{1}{l'(x_j)}$ podemos reescribir l_j como:

$$l_j(x) = l(x) \frac{w_j}{(x - x_j)}$$

Ahora como vimos en la ecuación de Lagrange por polinomios todos los términos de la sumatoria contiene el factor $l(x)$ el cual no depneded de j , lo que nos permite reescribir esta fórmula de interpolación de Lagrange baricentrica como:

$$1 = \sum_{j=0}^n l_j(x) = l(x) \sum_{j=0}^n \frac{w_j}{(x - x_j)}$$

Como vimos antes la interpolación de Lagrange por polinomios tenia una complejidad de $O(n^2)$, sin embargo la modifiaciación de esta fórmula seguida de una complejidad de $O(n)$ hace que este método genere una menor carga computacional y debido a la cantidad de datos a interpolar es mucho más eficiente y estable.

3.3. Cubic Spline

Es un método de interpolación por intervalos, lo que le permite ser más presiso que otros métodos, donde realizan el calculo sobre todos los puntos ingresados. Un spline construido a partir de polinomios de tercer orden que pasan por un conjunto de puntos de control. La segunda derivada de este polinomio es comunmente cercano a cero en los puntos finales, mientras que provee una condición en los límites del polinomio que completa el sistema de $m - 2$ ecuaciones. A partir de n puntos (x, y) :

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

Satisface:

1. $S(x) \in C^2[a, b]$

2. Para cada subintervalo $[x_{i-1}, x_i]$, $S(x)$, existe un polinomio de grado 3.
3. $S(x_i) = y_i$ para todo i

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

Donde cada $C_i = a_i x + b_i x^2 + c_i x^3 + d_i x^4$, con $d_i \neq 0$

3.4. Akima

Este es un método de interpolación que funciona a partir de la interpolación sub-spline continua y diferenciable. Utiliza polinomios de tercer grado a trozos que utilizan la posición de los vecinos de un punto que pertenece al grupo que se desea interpolar para determinar los coeficientes de interpolación de dicho polinomio.

Dado que no es necesario resolver ningún sistema de ecuaciones complejo, y además por su sistema de vecinos locales, este es un método eficiente y con resultados suaves sin comportamientos erráticos.

Dado un grupo de puntos $s_i = s(x_i)$, donde $1 \leq i \leq h$, la función interpolante resultante está definida por:

$$s(x) = a_0 + a_1(x - X_i) + a_2(x - x_i)^2 + a_3(x - x_i)^3, \text{ donde } x_1 \leq x \leq x_{i+1}$$

Los coeficientes a están determinados por el intervalo local $[x_1, x_{i+1}]$

3.5. Piecewise Cubic Hermite Interpolation (PChip)

Siendo x_0, \dots, 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 polinomio 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.

3.6. IDW

El método de interpolación IDW (Inverse Distance Weighting) permite estimar valores desconocidos de acuerdo a su distancia específica de un punto con cuyo valor es conocido. Es ampliamente utilizado en todo el mundo para realizar estimaciones de datos relacionados a un espacio geográfico como la densidad poblacional, la lluvia, la elevación frente al nivel del mar, etc.

Una vez se tienen unos valores conocidos asociados a una ubicación se estiman los valores de

todas las locaciones cercanas bajo el principio de que los puntos cercanos comparten un valor mucho más similar a los puntos lejanos y por esto es su aplicación y uso en las interpolaciones de GIS. La fórmula utilizada es la siguiente:

$$Z_P = \frac{\sum_{i=1}^n \left(\frac{Z_i}{d_i^p} \right)}{\sum_{i=1}^n \frac{1}{d_i^p}}$$

4. Resultados Métodos Interpolación

4.1. Metodología:

En primer lugar se realizo una revisión de los datos para hallar las estaciones y columnas ideales para depues realizar la interpolación correspondiente. Luego, se realizó el cambio de formato días juliano a días gregoriano. Posteriormente importamos los datos en dataframes para realizar el manejo de los mismo por medio de la herramienta Jupyter, desde la cual utilizabamos cada uno de los métodos de interpolación y generabamos para cada uno de estos la gráfica donde podemos comparar sus resultados con los datos reales.

4.2. Resultados:

Se encuentra la gráfica con los resultados de dos estaciones, para poder realizar la comparación de cada de uno de los métodos graficamos el valor de la temperatura interna:

- Itatira: Para realizar la interpolación de los datos de esta estación se eliminaron 101 filas aleatorias.
- Araripe: Para realizar la interpolación de los datos de esta estación se eliminaron 476 filas dejando solo los picos y los bajos, que representan las más bajas y altas temperaturas.

Para cada una de las gráficas podemos observar los datos reales representado con una cruz verde y los datos que obtenemos por medio de la interpolación representado con un círculo azul, esto con el fin de poder observar los fallos en los datos obtenidos. Por comodidad del lector, las imágenes de las pruebas mencionadas anteriormente se encuentran en la sección de Anexos.

5. Análisis de Errores

5.1. Error Araripe

Errores Estación Araripe		
Método	Error Absoluto	Error Relativo
Cubic Spline	2.4612223631458914	0.09850402698849006
PChip	1.9827295765408255	0.07738853925044341
LaGrange	2.0197330560132105	0.07787129246377333
Barycentric	2.156653776121743	0.08698143494456347
Akima	2.0749339627705785	0.08163754061083135
Krogh	0.0	0.0

5.2. Error Itatira

Errores Estación Itatira		
Método	Error Absoluto	Error Relativo
Cubic Spline	2.8596817069415357	0.11312817244906336
PChip	0.09467757659317301	0.0031186412196265046
LaGrange	0.10095514186632883	0.003378312766415002
Barycentric	0.10440509259259258	0.0034683754428674637
Akima	0.09507752737890036	0.0031681970387511466
Krogh	0.0	0.0

Los datos obtenidos por medio del método de Krogh fueron inconsistentes por lo tanto no se tendrán en cuenta en la medición del error absoluto, criterio por el cual se eligió el método con mejor desempeño para este tipo de datos.

Al revisar las medias de los errores en las pruebas anteriores, se decidió optar como método preferido la Interpolación PChip dado su bajo error. Con el método seleccionado se hicieron pruebas interpolando la temperatura interna, humedad relativa, velocidad del viento y presión atmosférica en las cuatro estaciones. En la sección 8.2, en los anexos, se encuentran de estas imágenes y las demás en un archivo comprimido.

Para cada una de las estaciones se decidió generar una tendencia distinta en los datos faltantes (los que se pretende interpolar), para así poder ver el nivel de impacto que cada uno de estos puede tener.

- Para la estación de Quixadá se retiraron 111 filas en total. Las filas que se eliminaron estaban ubicadas en los sumideros de la gráfica de la humedad relativa (esta fue tomada como referencia pero se realizó la interpolación en las cuatro columnas anteriormente mencionadas), es decir que son filas que pertenecen a los registros que están entre dos picos en los datos.

- En los datos suministrados por la estación de Fortaleza se eliminaron 53 filas en total. Se tomaron como referencia los picos de la gráfica de temperatura y se eliminaron (la temperatura fue tomada como referencia pero se realizó la interpolación en las cuatro columnas anteriormente mencionadas).
- Los registros eliminados en la estación de Araripe sumaban un total de 476 filas. Se decidió dejar únicamente los picos altos y bajos en la función de la temperatura.
- Para la estación de Itatira se decidieron eliminar 101 filas sin ningún criterio específico, todas se eligieron de forma aleatoria.

6. Análisis GIS

6.1. Metodología:

1. En primer lugar creamos un archivo .CSV con las coordenadas de cada estación, junto con su nombre y temperatura en la fila que corresponden a la fecha 4/4/2013 12:00 ya que nuestro objetivo era interpolar los datos en un mismo instante de tiempo. Por medio de la herramienta QGis leimos los datos de cada una de las estaciones junto su ubicación en el espacio, para así crear la capa donde se puede evidenciar los puntos donde se ubican las mismas en el mapa satelital por medio de ESRI (herramienta que podemos evidenciar en la Figura 1.)
2. Clasificamos cada punto por cuartiles iguales de acuerdo a su temperatura para posteriormente realizar la interpolación IDW y así generar la interpolación del mapa; más adelante utilizamos el Raster de esta misma herramienta para generar las curvas de nivel con intervalos de 1.5°
3. Una vez hecho se realizo el mismo procedimiento pero esta vez sin la estación Quixadá para poder comparar los dos mapas de resultados y verificar si es posible realizar la interpolación de una estación.

Realizamos un video de la pantalla al manejar la herramienta QGis y generar los mapas con las curvas de nivel de la temperatura, es video lo puede encontrar en el siguiente link: <https://www.youtube.com/watch?v=tRVB3UpTbE0> En las siguientes figuras se observan los mapas los datos de las curvas de nivel con todas las estaciones, mientras que en la última se observan los resultados de la interpolación del mapa sin la estación Quixadá.

6.2. Resultados:



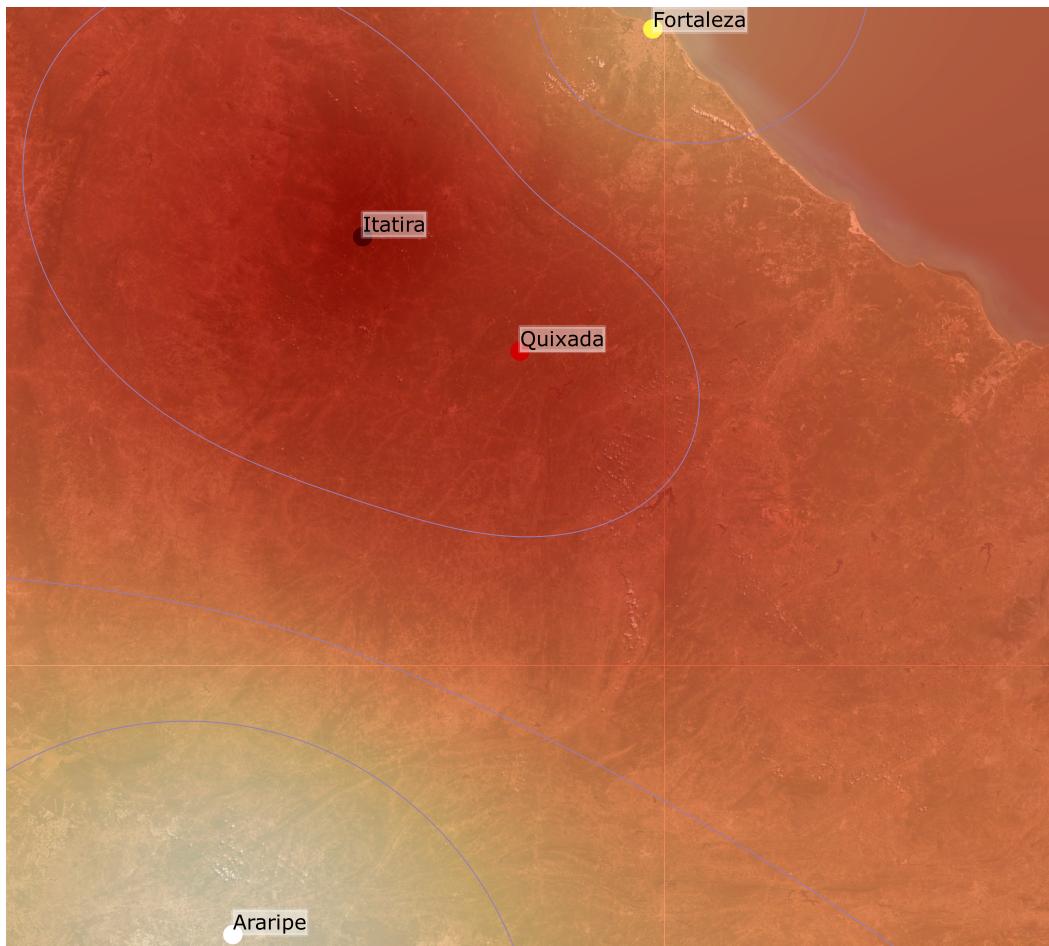
**Pontificia Universidad
Javeriana
Análisis Numérico
Mapa Satelital ESRI**

Estado de Ceará
República Federativa de Brasil



0 25 50 km

Copyright 2019 - Nicolás Camacho Plazas, Daniela Cortes Antonio, Mateo Florido Sanchez



Pontificia Universidad Javeriana Análisis Numérico Mapa Curvas de Nivel

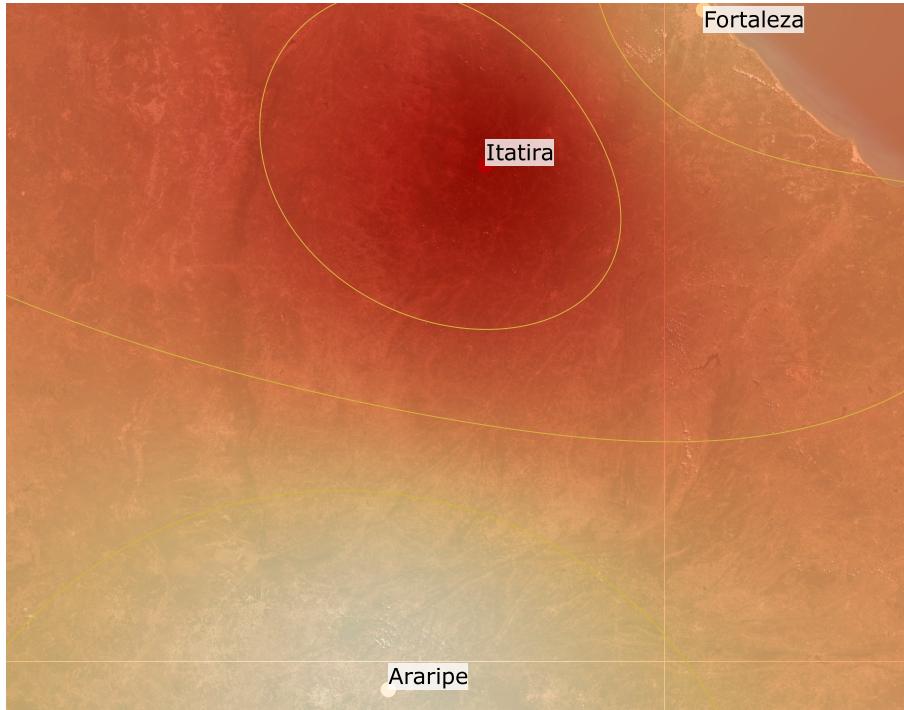
Estado de Ceará
República Federativa de Brasil



- Stations
- 27.95 - 29.33
 - 29.33 - 30.80
 - 30.80 - 32.44
 - 32.44 - 33.28
 - 33.28 - 33.71
- Curvas de nivel
- Interpolado
- 27.950008
 - 29.3899985
 - 30.829989
 - 32.2699795
 - 33.70997
- ESRI Satellite

0 25 50 km

Copyright 2019 - Nicolás Camacho Plazas, Daniela Cortes Antonio, Mateo Florido Sanchez



Copyright 2019 - Nicolás Camacho Plazas Daniela Cortes Antonio, Mateo Florido Sanchez

Pontificia Universidad Javeriana
Análisis Numérico
Mapa Interpolación Estimación
Quixadá

Región de Ceará
República Federativa de
Brasil



StationsDrop
● 27.9500 - 29.1020
● 29.1020 - 30.2540
● 30.2540 - 31.4060
● 31.4060 - 32.5580
● 32.5580 - 33.7100
— Curvas de nivel

Interpolado
■ 27.950038
■ 29.39001725
■ 30.8299965
■ 32.26997575
■ 33.709955
ESRI Satellite

0 50 100 km

7. Conclusiones

1. Una de las observaciones más interesantes encontradas al revisar los resultados de la interpolación PChiper, con todos los datos seleccionados, fue ver como se comportaba la gráfica de la temperatura en comparación con la gráfica de la humedad relativa del aire. Pese a que no era proporcional, se encontró una clara realción inversa pues cuando exsistian tendencias o registros altos en la temperatura, se lograba observar el comportamiento opuesto en la humedad, dejando a la vista su relación ya mensionada como inversa con una uncreible exactitud.
2. Cuando se realizó en análisis del error, se utilizaron cuatro tipos de tendencias en las filas de los datos faltantes para observar su impacto en la interpolación. La tendencia más nociva para la cualquier tipo de interpolación es la de los picos, puesto que se suavizan los puntos bajos inmediatamente anteriores y posteriores al pico, generando una curva suave nada inclinada que deja por fuera todos los datos que conformaban a los picos faltantes, generando un fallo y una distancia enorme entre los datos interpolados y los datos reales. Esto se puede ver en la sección 8.2, en las gráficas de interpolación de la temperatura o humedad en Fortaleza.
3. Siguiendo el concejo de la profezora, se eliminaron todos los registros que no hacían parte de los extremos superiores e inferiores en un escenario de prueba, por los exce- lentes resultados de error en esta interpolación, y considerando la cantidad gigantesca de filas eliminadas, se concluye que los picos altos y los bajos son los datos más impor- tantes y que mayor peso tienen con respecto a la veracidad de la información generada a partir de la Interpolación.
4. Después de ver el promedio de error absoluto y error relativo que presentaron los seis métodos seleccionados, se puede afirmar que el método de interpolación PChip es el más adecuado para el escenario de prueba de los datos generado por las estaciones seleccionadas. La curvatura de las curvas se ajusataba de forma ideal y representaba de manera cercana a la curvatura real que formaban los datos de las cuatro categorías elegidas, aunque el algoritmo de akima, por sus coeficientes locales, también genera resultados bastante confiables en el caso en que se interpole para estimar registros sin una tendencia marcada (aleatorios).

8. Anexos

8.1. Resultados: (4.2)

8.2. Resultados análisis del error:

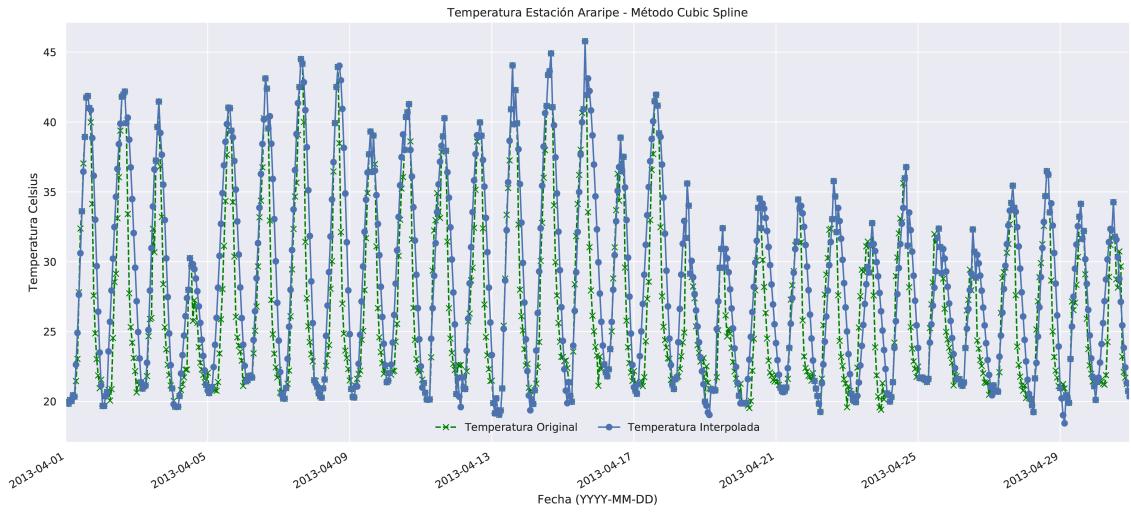


Figura 2: Polinómica de Lagrange-Itatira (Filas aleatorias)

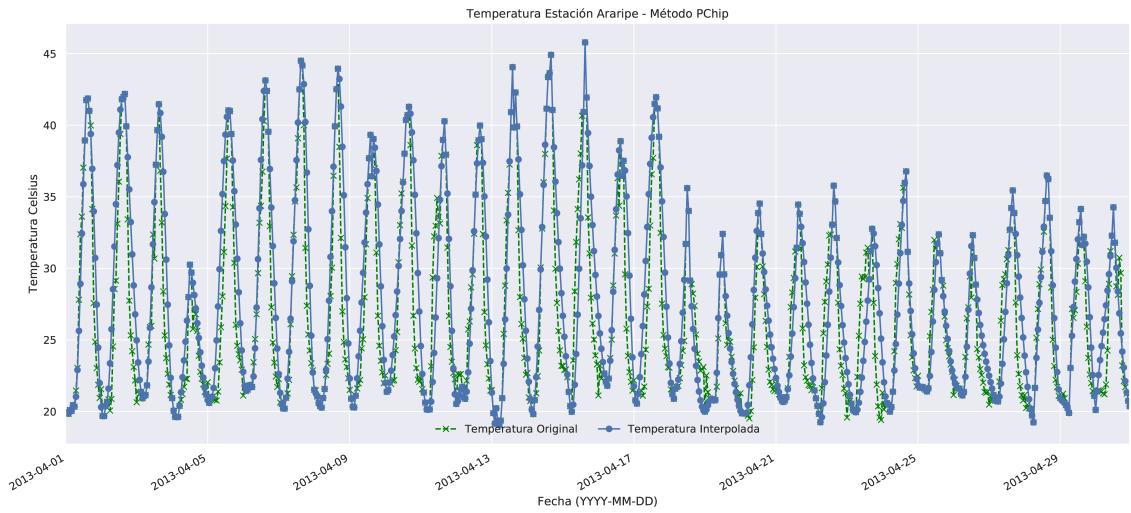


Figura 3: Polinómica de Lagrange-Araripe (Picos y Bajos)

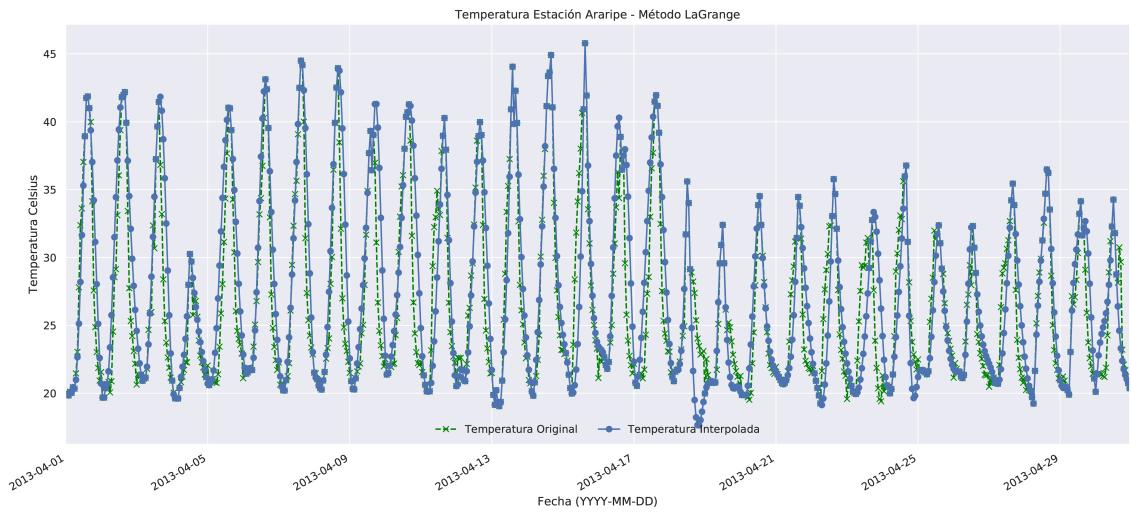


Figura 4: Barycentric Lagrange-Itatira (Filas aleatorias)

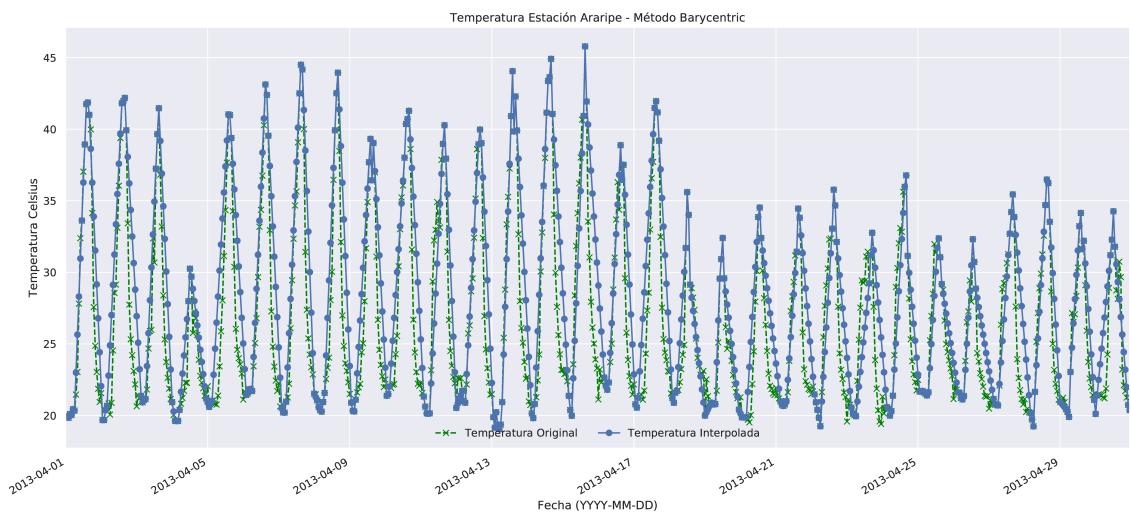


Figura 5: Barycentric Lagrange-Araripe (Picos y Bajos)

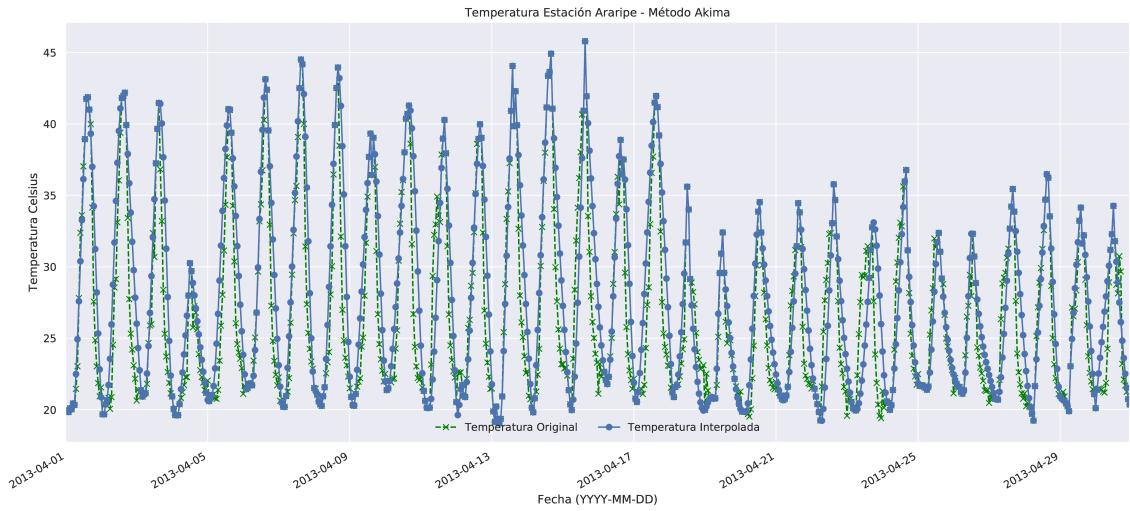


Figura 6: Cubic Spline-Itatira (Filas aleatorias)

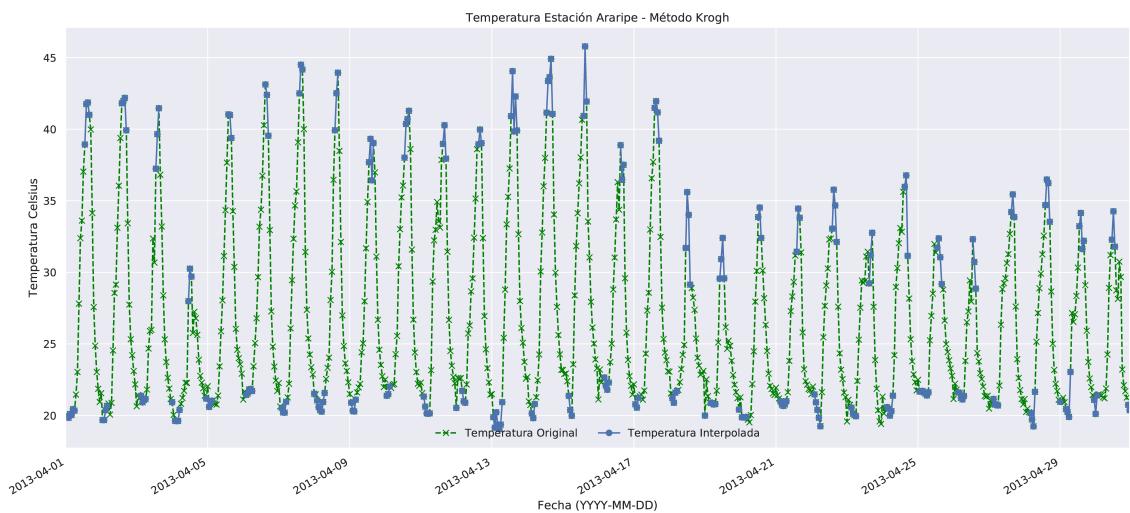


Figura 7: Cubic Spline-Araripe (Picos y Bajos)

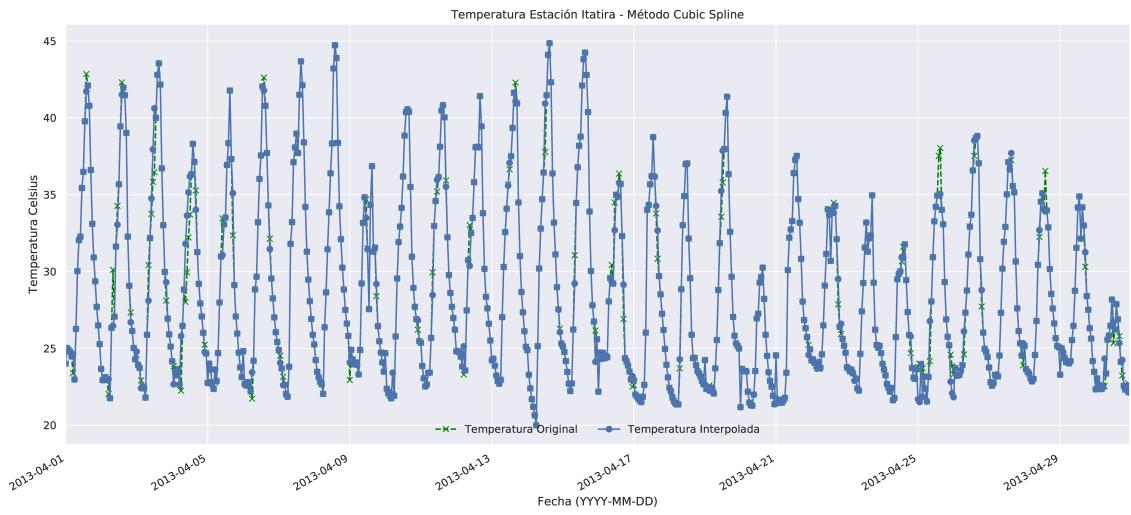


Figura 8: PChip-Itatira (Filas aleatorias)

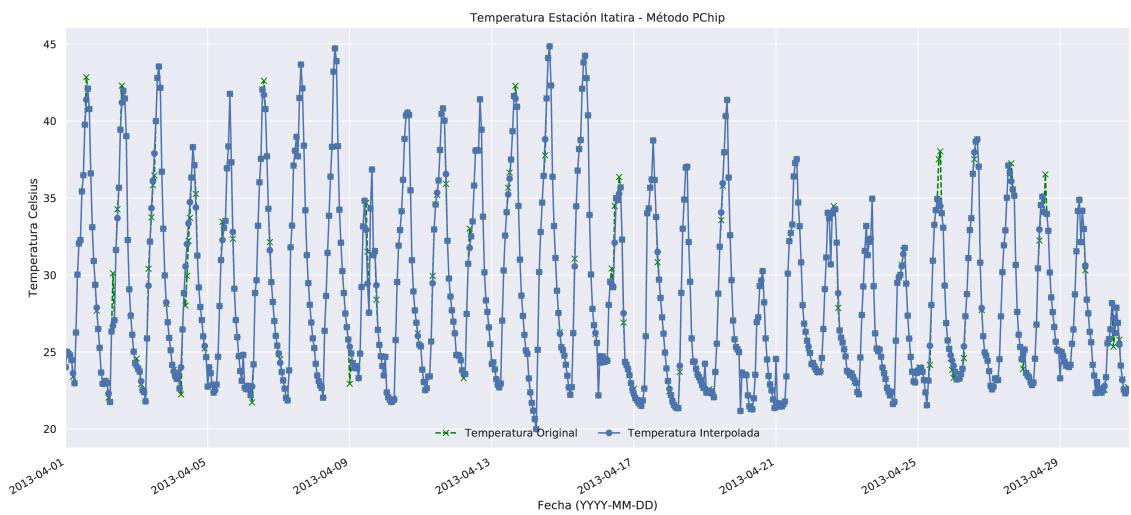


Figura 9: PChip-Araripe (Picos y Bajos)

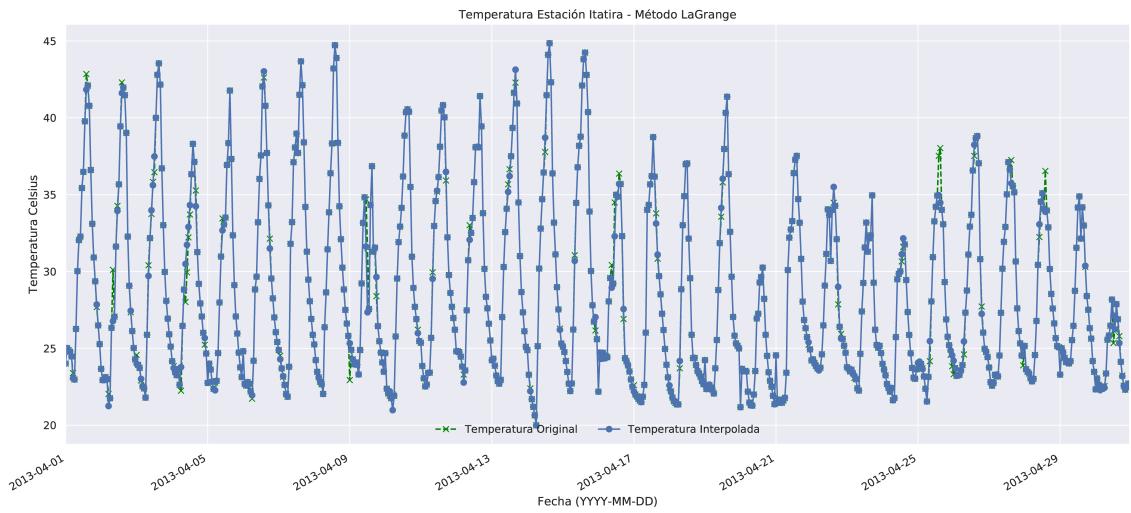


Figura 10: Akima-Itatira (Filas aleatorias)

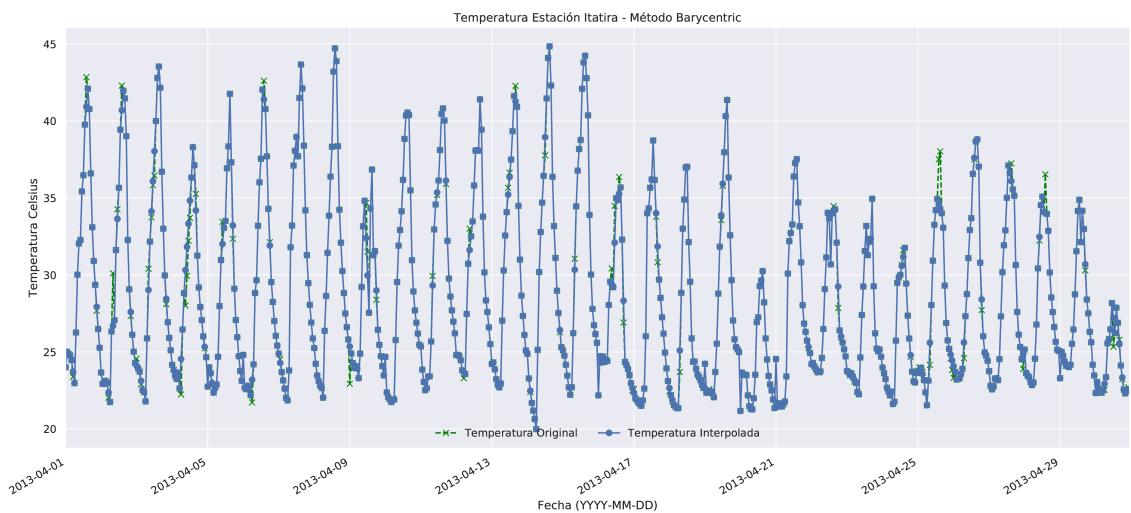


Figura 11: Akima-Araripe (Picos y Bajos)

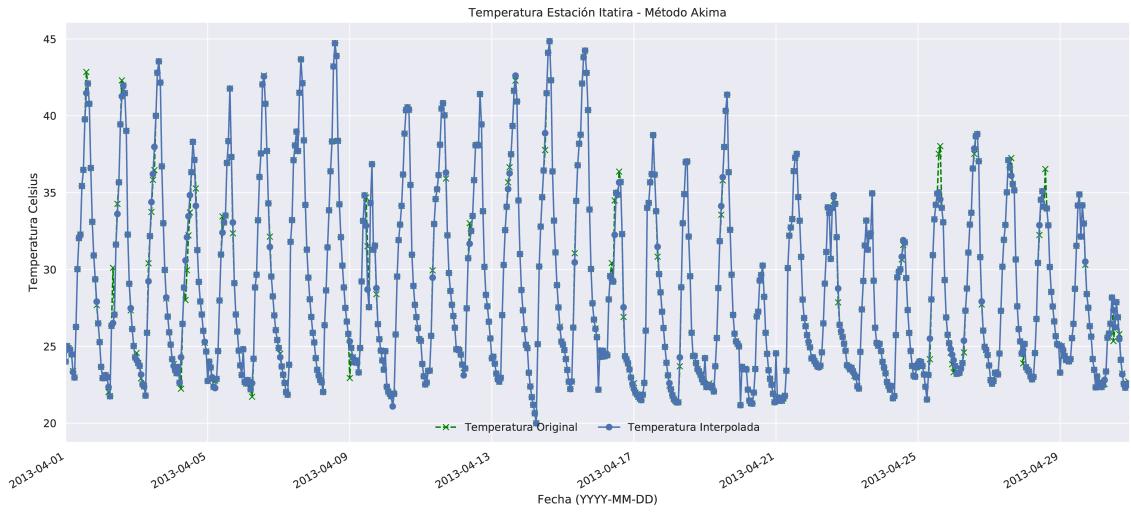


Figura 12: Krogh-Itatira (Filas aleatorias)

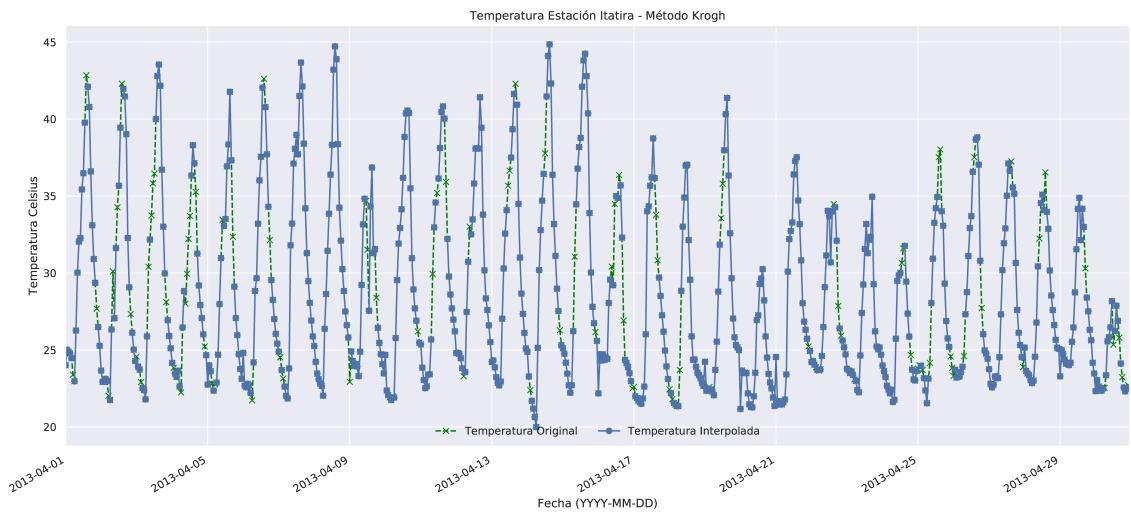


Figura 13: Krogh-Araripe (Picos y Bajos)

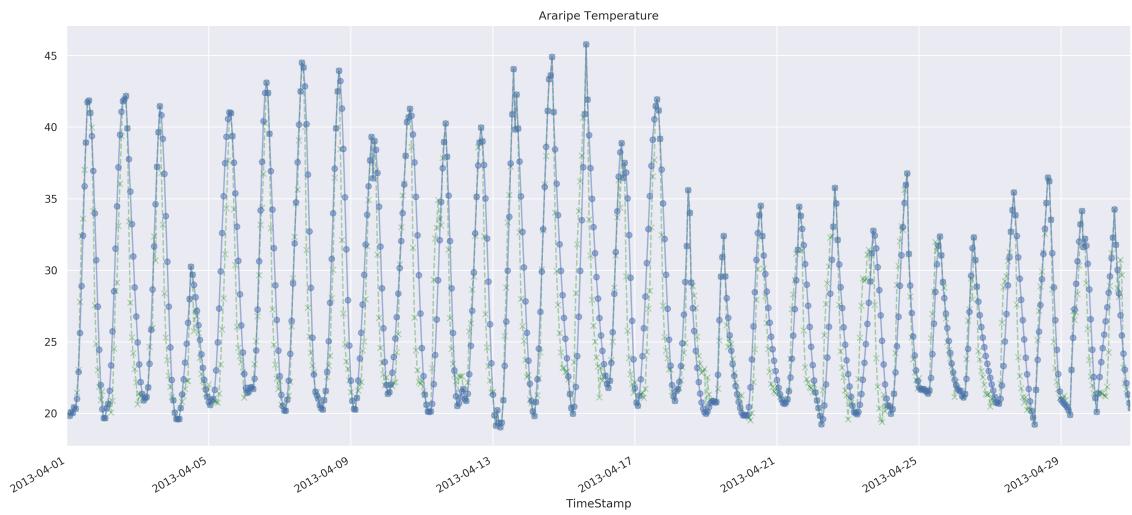


Figura 14: Araripe-Temperatura (Picos y bajos)

Humidity2.png Humidity2.png

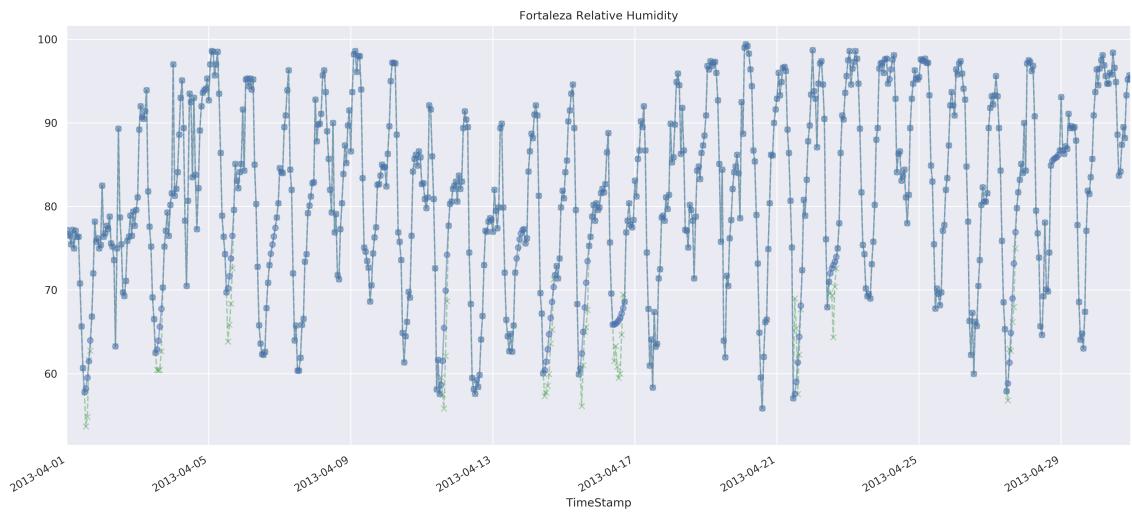


Figura 15: Fortaleza-Humedad Relativa (Picos en Temperatura)

Speed3.png Speed3.png

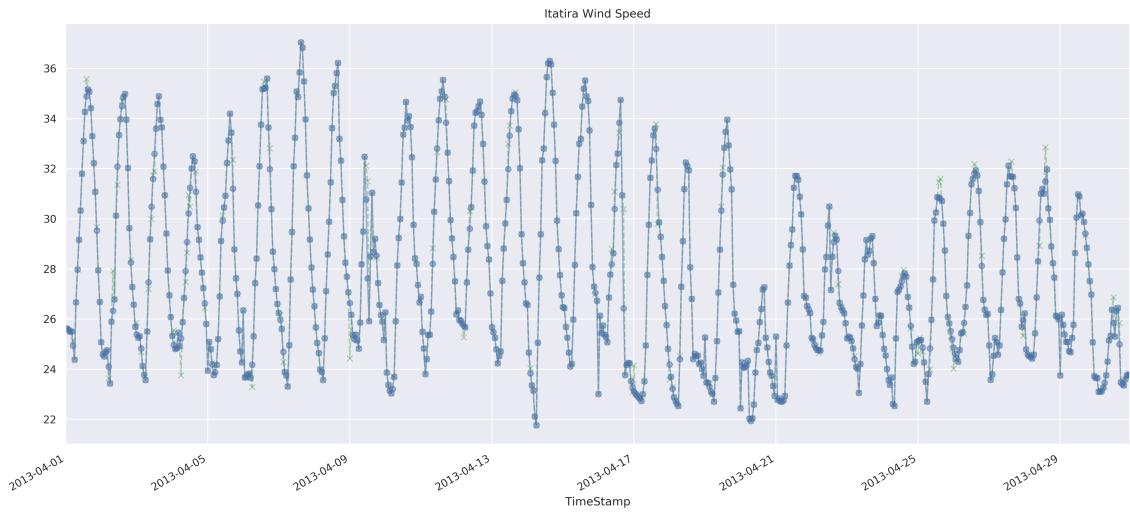


Figura 16: Itatira-Velocidad del Viento (Aleatorio)

pressure0.png pressure0.png

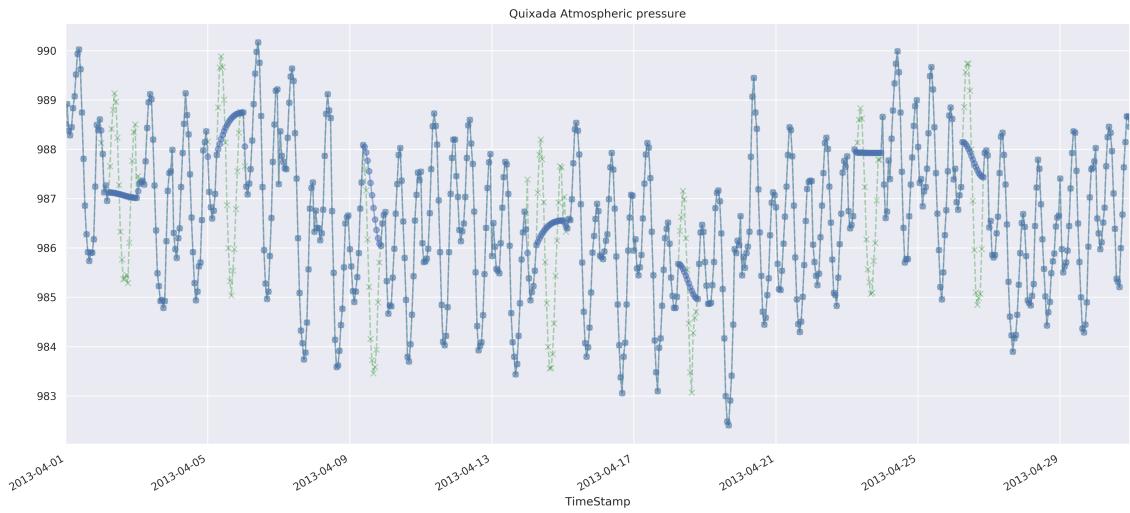


Figura 17: Quixadá-Presión Atmosférica (Sumideros en Humedad Relativa)