

INTERPOLACIÓN DE DATOS FALTANTES MEDIANTE SPLINE CÚBICO Y MÍNIMOS CUADRADOS

S. Haro-Rivera^{1*}, A. Escudero-Villa¹

Grupo de Energías Alternativas y Ambiente,
Facultad de Ciencias, Escuela Superior Politécnica de Chimborazo¹, Riobamba-Ecuador

*e-mail: s_haro@esPOCH.edu.ec; Telephone: 0982752949

R esumen

En la actualidad Ecuador promueve el estudio de variables meteorológicas como fuente principal para generación de energías renovables; por ello, se han instalado diversas estaciones meteorológicas en la provincia de Chimborazo, con varios sensores que miden variables como velocidad de viento. Los datos se registran en un datalogger y de forma automática migran a un servidor, el mayor problema que surge es la pérdidas de datos por falta de mantenimiento permanente en las estaciones debido a su ubicación, motivo por el cual surge la necesidad de buscar métodos de interpolación para la velocidad de viento que permitan realizar análisis confiables. La investigación se basó en la interpolación por spline cúbico y mínimos cuadrados, con la finalidad de determinar el método de mayor ajuste a los datos de velocidad de viento. Para el estudio se tomó la información en intervalos de 10 y 60 minutos; desde 1 hasta 12 datos faltantes. Se halló diferencias significativas en la interpolación mediante spline cúbico cuando el número de faltantes se incrementa al inicio y final de la base de datos, mientras que con mínimos cuadrados cuando los datos faltantes se encuentran dispersos.

Palabras claves: Interpolación, velocidad de viento, spline cúbico, mínimos cuadrados.

A bstract

Ecuador currently encourages the study of meteorological variables as the main source for renewable energy generation; therefore, we have installed various weather stations in the Chimborazo province, with several sensors that measure variables such as wind speed. The data are recorded in datalogger and automatically migrate to server, the biggest problem that present is the data loss by to lack of maintenance on the stations due to its location and arises the need to seek methods interpolation wind speed that allow for reliable analysis. The research was based on the cubic spline interpolation and least squares, in order to determine adjustment method for wind speed data. The information for the study was taken the intervals of 10 to 60 minutes; from 1 to 12 missing data. Significant differences in the interpolation by cubic spline was found when the number of missing increases at the beginning and finally of the database and with least squares when missing data are scattered.

Key words: interpolation, wind speed, spline cubic, least squares

Fecha de recepción: 14-10-2016

Fecha de aceptación: 20-04-2017

INTRODUCCIÓN

El viento es una fuente renovable, que permite generar energía eólica (1) y so-

bre todo proporciona información a los sistemas de vigilancia y alerta temprana; por ello es importante trabajar con bases de datos confiables, pues es común encontrar registros con poca duración, numerosos errores y au-

sencia de datos (2), lo cual dificulta la aplicación de las distintas técnicas estadísticas (3). Para abordar esta problemática se plantearon dos métodos de interpolación: spline cúbico y mínimos cuadrados. La interpolación por medio de splines cúbicos permite crear funciones de orden superior, que aproximan los datos tomados a una curva continua que conserva la suavidad inherente a toda variable física (4), la ecuación 1 determina el polinomio de interpolación (5)

$$X_k(t) = \frac{\sigma_k}{6} \left[\frac{(t_{k+1}-t)^3}{h_k} - h_k(t_{k+1}-t) \right] + \frac{\sigma_{k+1}}{6} \left[\frac{(t-t_k)^3}{h_k} - h_k(t-t_k) \right] + y_k \left[\frac{(t_{k+1}-t)}{h_k} \right] + y_{k+1} \left[\frac{(t-t_k)}{h_k} \right] \quad (1)$$

con; $k=0,1, 2, \dots, n$ donde n es el número de intervalos de tiempo.

El segundo método establece la mejor curva calculando la suma de las distancias al cuadrado entre el valor de la función que aproxima $\varphi(x)$ y el valor de la función dada $y=f(x)$ cuando ésta sea mínima (6). La metodología que determina esta curva radica en el cálculo de los coeficientes mediante derivadas parciales, es decir; dados los datos $\{(x_i, y_i), \text{ con } i=0,1,2, \dots, n\}$ donde n son los intervalos de tiempo. Se debe determinar las funciones $\varphi(x)$ tales que satisfagan la distancia mínima dada por la ecuación 2:

$$\sum_{i=1}^n (y_i - \varphi(x_i))^2 \quad (2)$$

El objetivo principal de la investigación es determinar las condiciones de ajuste de los dos métodos para el relleno de datos faltantes de velocidad de viento.

METODOLOGÍA

Para la investigación se seleccionaron dos bases de datos de velocidad de viento discretizados (7), mismos que fueron registrados en el 2014 por la estación meteorológica de la Escuela Superior Politécnica de Chimborazo (ESPOCH) en el 2014, ubicada a 2754 m.s.n.m., 9816965Y y 758398X. Se trabajó con los promedios cada 10 y 60 min, se tomaron cuatro días (8, 10, 13 y 20 del mes de abril) y se generaron series con 1 hasta 12 datos faltantes de forma aleatoria. La interpolación se realizó mediante MATLAB versión 7.10.0.499 (R2010a).

Para determinar diferencias entre los valores interpolados y los reales se calculó los errores mediante la ecuación 3:

$$E = \left| \frac{\text{Valor}_{real} - \text{Valor}_{calculado}}{\text{Valor}_{real}} \right| \quad (3)$$

El análisis estadístico se realizó mediante SPSS. Debido a que los datos no están distribuidos normalmente se empleó el test de Wilcoxon para dos muestras relacionadas (8) al 95% de confiabilidad, ecuación 4:

$$Z = \frac{W - \left[\frac{n(n+1)}{4} \right]}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \quad (4)$$

Donde W es la mínima suma de rangos entre la suma de rangos de las diferencias positivas y la suma de rangos de las diferencias negativas y n el número de datos.

RESULTADOS

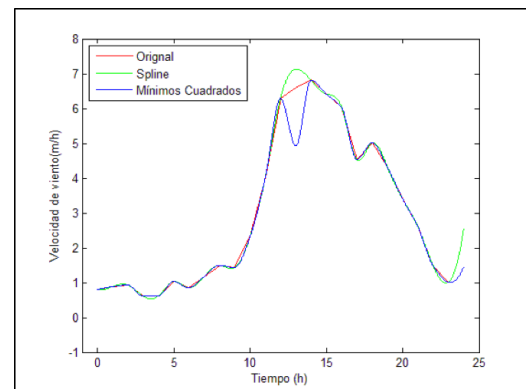


Figura 1: Interpolación por hora con 1 dato faltante.

La Figura 1 muestra los datos reales e interpolados de velocidad de viento con un dato faltante, se observa en el intervalo de 10:00 a 13:00 horas diferencia entre las curvas. El spline cúbico se aproxima de mejor forma a la curva original, mientras que por mínimos cuadrados presenta concavidad. En el intervalo de 21:00 a 23:00 horas el spline cúbico se aleja de la curva original, se puede notar la suavidad del spline cúbico en toda la curva.

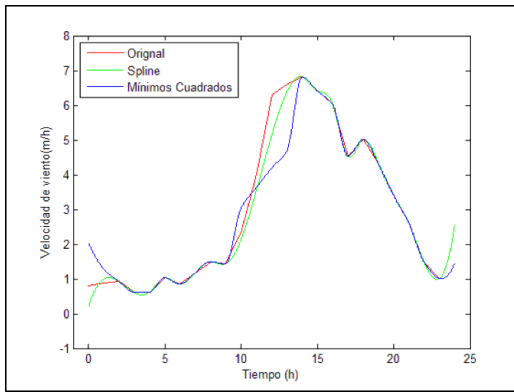


Figura 2: Interpolación por hora con 1 hasta 6 datos faltantes.

La Figura 2 muestra las curvas de los datos originales y las interpolaciones de velocidad de viento con 1 hasta 6 datos faltantes. Se puede observar la variación del spline cúbico al inicio y final de la curva, mientras que en el intervalo de 10:00 a 13:00 horas se aproxima a la original. Por mínimos cuadrados presenta variación al inicio y levemente al final de la función, mientras que en dicho intervalo se aleja de la curva original.

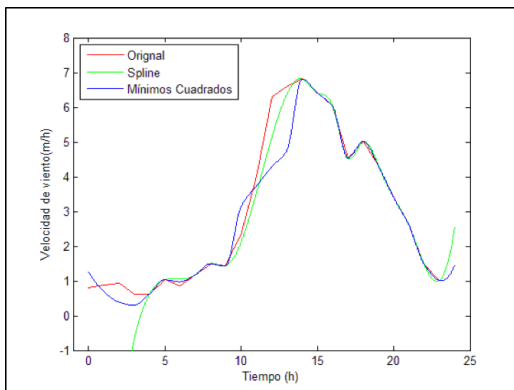


Figura 3: Interpolación por hora con 1 hasta 10 datos faltantes.

La Figura 3 muestra las curvas de los datos originales de velocidad de viento y las interpolaciones por spline cúbico y mínimos cuadrados con 1 a 10 datos faltantes. La curva spline se aleja de la función original al inicio y levemente al final; en el intervalo de 0:00 a 3:00

horas cae notoriamente con una diferencia máxima de 13,41m/h en relación a la curva por mínimos cuadrados.

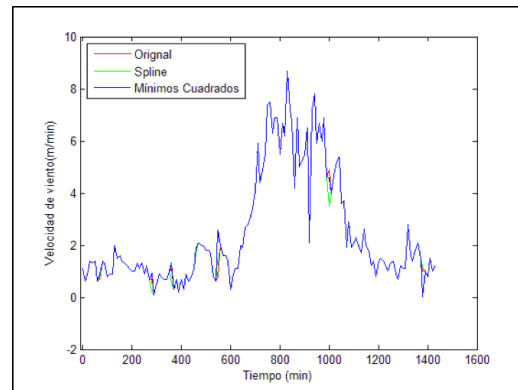


Figura 4: Interpolación por minuto con 1 dato faltante.

La Figura 4 muestra los datos originales e interpolados por minuto con 1 dato faltante. Se puede observar intervalos donde las curvas son similares y cortos intervalos de tiempo donde las curvas se alejan de la original.

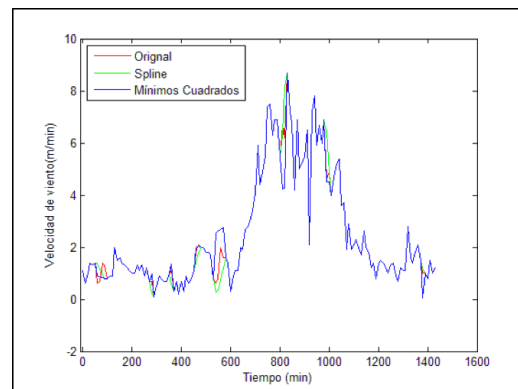


Figura 5: Interpolación por minuto con 1 hasta 6 datos faltantes

La Figura 5 muestra las curvas de las interpolaciones de 1 hasta 6 datos faltantes, se observa notables variaciones del spline cúbico y mínimos cuadrados en los intervalos de 60 a 90 min (01:00 a 01:18) y de 540 a 570 min (09:00 a 09:30).

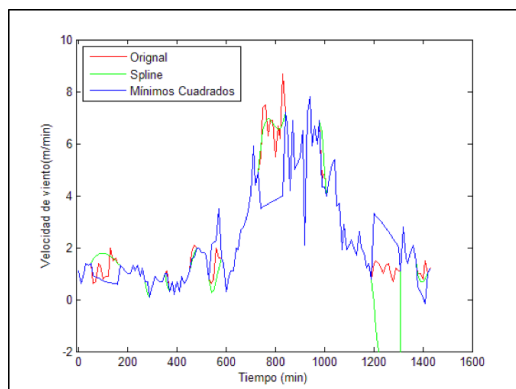


Figura 6: Interpolación por minuto con 1 hasta 11 datos faltantes

La Figura 6 muestra los datos originales e interpolados hasta 11 datos faltantes. La velocidad de viento interpolada mediante spline cúbico cae notoriamente a partir de los 1200 a 1300 min (20:00 a 21:40), mientras que la curva por mínimos cuadrados muestra variación aproximadamente en el intervalo de 700 a 900 min (11:40 a 15:00).

CONCLUSIONES

El análisis mostró un 0,66% de error cuando los datos faltantes son pocos y su distribución es dispersa en la serie, caso contrario alcanzan errores de hasta el 100%.

Con el test de Wilcoxon, en la interpolación por hora, con spline cúbico se hallaron valores de p iguales a 0,001 y 0,015; y con mínimos cuadrados de 0,010 y 0,023, lo cual indica una diferencia significativa entre los datos interpolados y los reales en las series de 11 y 12 datos faltantes. La interpolación con spline cúbico mostró diferencias significativas cuando el número de datos faltantes se incrementan al inicio y final de la serie, mientras que el método por mínimos cuadrados cuando la distribución de los datos faltantes es dispersa.

En la interpolación por minuto, se hallaron valores de p iguales a 0,016, 0,010 y 0,002 con spline cúbico en las series de 9, 10 y 11 datos faltantes; mientras que con mínimos cuadrados la prueba no mostró diferencia significativa, por lo que, se concluye que éste último método aproxima de mejor forma los datos faltantes en intervalos de tiempo por minuto.

AGRADECIMIENTOS

Al Centro de Investigaciones de Energía Alternativa y Ambiente de la ESPOCH, Facultad de Ciencias.

Referencias

1. Guevara J. Cuantificación del perfil del viento hasta 100m de altura desde la superficie y su incidencia en la climatología eólica. Terra Nueva Etapa. 2013 julio-diciembre; XXIX(46):81.
2. Infante S, Ortega J, Cedeño F. Estimación de datos faltantes en estaciones meteorológicas de Venezuela vía un modelo de redes neuronales. Revista de Climatología. 2008 Agosto; 8:51-52.
3. Álvarez O, Montaña T, Quentin EMJ, Solano J. Homogeneización de series de velocidad del viento mensual en las estaciones meteorológicas del INAMHI en Loja, Ecuador. Revista de Climatología. 2013; 13:36.
4. Quintero E, Ardila W, Gallego H. Interfaz gráfica para la interpolación de datos a través de splines cúbicos. Scientia Et Technica. 2010; XVI(44):195.
5. Escudero A, Reclade C, Haro S, Meneses A. Spline cúbico para el tratamiento funcional de la radiación solar global. Información tecnológica. 2016; 27:156.
6. Chapra S, Canale, Raymond. In F. H, editor. Métodos numéricos para ingenieros. Mexico: McGraw-Hill; 2008.p. 498-532.
7. Araújo A, Melo G, de Medeiros A, dos Santos M. Simulación de la producción de energía eléctrica con aerogeneradores de pequeño tamaño. Información Tecnológica. 2009; 20:37.
8. Álvarez R. Estadística multivariante y no paramétrica con SPSS. 3rd ed. Díaz, editor. España; 1995.p. 305-380.