



DOI: <https://doi.org/10.23857/dc.v11i1.4243>

Ciencias Técnicas y Aplicadas
Artículo de Investigación

Predicción de series temporales de humedad relativa mediante bootstrap y modelos de regresión funcional no paramétrica

Relative humidity time series prediction using bootstrap and nonparametric functional regression models

Previsão de séries temporais de humidade relativa usando modelos bootstrap e de regressão funcional não paramétrica

Jorge David Zúñiga-Lema ^I
jorged.zuniga@unach.edu.ec
<https://orcid.org/0009-0006-0621-0967>

Lourdes del Carmen Zúñiga-Lema ^{II}
lzuniga@epoch.edu.ec
<https://orcid.org/0000-0001-9060-6273>

Stephani Belén Ruiz-Zambrano ^{III}
stephani.ruiz@epoch.edu.ec
<https://orcid.org/0009-0005-6097-1620>

Estalin Fabián Mejía-Hidalgo ^{IV}
estalin.mejia@unach.edu.ec
<https://orcid.org/0009-0006-0215-2237>

Correspondencia: jorged.zuniga@unach.edu.ec

***Recibido:** 11 de noviembre de 2024 ***Aceptado:** 17 de diciembre de 2024 * **Publicado:** 28 de enero de 2025

- I. Universidad Nacional de Chimborazo, Riobamba, Ecuador.
- II. Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador.
- III. Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador.
- IV. Universidad Nacional de Chimborazo, Riobamba, Ecuador.

Resumen

La investigación tuvo como objetivo, desarrollar intervalos de predicción para series temporales de humedad relativa utilizando la metodología bootstrap combinada con modelos de regresión funcional no paramétrica. Para ello, se registraron los valores de humedad relativa en porcentaje, cada segundo, durante el año 2023 en la estación meteorológica de la Facultad de Ingeniería de la Universidad Nacional de Chimborazo. La investigación incluyó la limpieza de valores atípicos, la interpolación de datos faltantes y el cálculo de la media horaria mensual. Los modelos de regresión funcional se ajustaron mediante B-splines, los cuales capturaron de manera efectiva las características esenciales de cada serie mensual. Se generaron 5000 réplicas de las series temporales usando la metodología bootstrap para predecir la humedad relativa de diciembre de 2023. La predicción fue contrastada con los datos reales obtenidos, mostrando una alta consistencia. Para mejorar la precisión, las predicciones fueron suavizadas con bases de Fourier y se calcularon intervalos de confianza al 95%. Los resultados demuestran que la metodología empleada es sólida, proporcionando intervalos de predicción robustos. Este enfoque no solo facilita una comprensión más profunda de los patrones de humedad relativa a lo largo del año, sino que también ofrece una herramienta valiosa para futuras investigaciones. La metodología implementada puede ser aplicada en la predicción y análisis de series temporales en variables meteorológicas.

Palabras clave: humedad relativa; series temporales; bootstrap; regresión funcional; modelo no paramétrico; Fourier.

Abstract

The objective of the research was to develop prediction intervals for relative humidity time series using the bootstrap methodology combined with non-parametric functional regression models. To do this, relative humidity values were recorded in percentage, every second, during the year 2023 at the meteorological station of the Faculty of Engineering of the National University of Chimborazo. The research included the cleaning of outliers, the interpolation of missing data and the calculation of the monthly hourly average. The functional regression models were adjusted using B-splines, which effectively captured the essential characteristics of each monthly series. 5000 replicas of the time series were generated using the bootstrap methodology to predict the relative humidity for December 2023. The prediction was contrasted with the real data obtained, showing high consistency. To improve accuracy, the predictions were smoothed with Fourier bases and 95% confidence intervals

Predicción de series temporales de humedad relativa mediante bootstrap y modelos de regresión funcional no paramétrica

were calculated. The results demonstrate that the methodology used is solid, providing robust prediction intervals. This approach not only facilitates a deeper understanding of relative humidity patterns throughout the year, but also offers a valuable tool for future research. The implemented methodology can be applied in the prediction and analysis of time series of meteorological variables.

Keywords: relative humidity; time series; bootstrap; functional regression; nonparametric model; Fourier.

Resumo

A investigação teve como objetivo desenvolver intervalos de previsão para séries temporais de humidade relativa utilizando a metodologia bootstrap combinada com modelos de regressão funcional não paramétricos. Para tal, foram registados valores de humidade relativa em percentagem, a cada segundo, durante o ano de 2023 na estação meteorológica da Faculdade de Engenharia da Universidade Nacional de Chimborazo. A investigação incluiu a limpeza de valores discrepantes, a interpolação de dados em falta e o cálculo de médias horárias mensais. Os modelos de regressão funcional foram ajustados utilizando B-splines, que capturaram eficazmente as características essenciais de cada série mensal. Foram geradas 5000 réplicas da série temporal utilizando a metodologia bootstrap para prever a humidade relativa para dezembro de 2023. A previsão foi contrastada com os dados reais obtidos, mostrando uma elevada consistência. Para melhorar a precisão, as previsões foram suavizadas com bases de Fourier e foram calculados intervalos de confiança a 95%. Os resultados demonstram que a metodologia utilizada é robusta, fornecendo intervalos de previsão robustos. Esta abordagem não só facilita uma compreensão mais profunda dos padrões de humidade relativa ao longo do ano, como também oferece uma ferramenta valiosa para futuras pesquisas. A metodologia implementada pode ser aplicada na previsão e análise de séries temporais em variáveis meteorológicas.

Palavras-chave: humidade relativa; séries temporais; inicialização; regressão funcional; modelo não paramétrico; Quatro.

Introducción

El análisis del comportamiento de las variables meteorológicas es crucial debido a la variabilidad de los datos, la cual está influenciada por diversos factores externos. Aunque los modelos de regresión de series temporales han sido utilizados tradicionalmente con éxito para predicciones a corto y

Predicción de series temporales de humedad relativa mediante bootstrap y modelos de regresión funcional no paramétrica

mediano plazo, la extensión del período de análisis a menudo resulta en predicciones menos precisas (Chariguamán Maurisaca & Meneses Freire, 2024). En este contexto, el presente estudio se enfocó en aplicar un modelo funcional no paramétrico para series temporales de humedad relativa, a fin de mejorar la precisión en las predicciones.

Se utilizó la metodología bootstrap junto con modelos B-Spline, a fin de generar intervalos de predicción más fiables y representativos (Zúñiga & Meneses, 2024). Este enfoque no paramétrico adaptado a datos funcionales ofrece una herramienta sólida para el análisis y la predicción de series temporales de humedad relativa. A continuación, se consideran algunos conceptos relevantes para la presente investigación.

Modelos de regresión no paramétrico B-Spline

Los modelos de regresión no paramétricos permiten estimar efectos no lineales de manera flexible sin la necesidad de imponer una forma funcional predeterminada. Esto es especialmente útil cuando se analiza una sola covariable continua, como “x” (Toalombo et al., 2022), el modelo estándar para la regresión no paramétrica se define de la siguiente manera:

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, \dots, n \quad (1)$$

En cuanto a la variable de error ϵ_i , se considera que la función f tiene ciertas propiedades de suavidad, como la continuidad o la diferenciabilidad, sin especificar una forma paramétrica concreta (Horváth & Kokoszka, 2012). Un spline es una función construida en segmentos utilizando funciones polinomiales, y su nombre proviene del concepto de “listón elástico” (Racine, 2023). Los splines “son curvas polinómicas por secciones que son continuamente diferenciables hasta un cierto orden” (Paluszny et al., 2005). Ejemplos comunes incluyen C_0 , un spline lineal por tramos, y C_1 , un spline cúbico.

Función base B-spline

Una función B-spline es una función base interpolativa con alta diferenciabilidad, que extiende la curva de Bézier. Los B-splines se definen por su orden m y el número de nodos interiores N (incluyendo dos nodos extremos, sumando un total de $(N + 2)$ nodos) (Peña, 2015). El grado del polinomio B-spline es $m-1$ (Racine, 2022). Una B-spline de grado n (con un orden de spline $m = n+1$) es una curva paramétrica compuesta por una combinación lineal de B-splines base $B_{i,n}(x)$ de grado n (Boor, 2001). Se define de la siguiente manera:

$$B(x) = \sum_{i=0}^{N+n} \beta_i * B_{i,n}(x), \quad x \in [t_0, t_{N+1}] \quad (2)$$

$f(x)$ tiene la siguiente estructura:

$$f(x) = a_1 * B_1(x) + \dots + a_k * B_k(x) \quad (3)$$

Donde:

K: cantidad de bases

a_1, \dots, a_k : parámetros desconocidos

B_1, \dots, B_k : funciones polinomiales

β_i : puntos de control o puntos de Boor

La secuencia t es la secuencia de nodos, donde cada término individual de la secuencia se conoce como un nodo (Racine, 2022). La regresión que utiliza B-splines de orden p se estructura como:

$$f(x) = a_0 + a_1x + \dots + a_px^p + B_1(x - C_1)^p + \dots + B_k(x - C_k)^p \quad (4)$$

Donde:

a_0, \dots, a_p y B_1, \dots, B_k : coeficientes a encontrar.

El spline $f(x)$ se puede expresar como una combinación lineal en la siguiente forma:

$$f(x) = a_0 * B_0(x) + \dots + a_p * B_p(x) + a_{p+1} * B_{p+1}(x) + \dots + a_{p+M} * B_{p+M}(x) \quad (5)$$

Las funciones siguientes conforman una base de funciones polinómicas para el spline:

$$B_0(x) = 1, B_1(x) = x, B_p(x) = x^p \quad (6)$$

$$B_{p+1}(x) = (x - C_1), \dots, B_{p+M} = (x - C_M) \quad (7)$$

Metodología Bootstrap

La metodología Bootstrap es una técnica empleada para aproximar la distribución de un muestreo estadístico. Este método se fundamenta en el remuestreo de datos, lo que significa generar nuevas muestras aleatorias a partir del original (Meneses et al., 2018). Utilizando los principios generales de este método y manteniendo la hipótesis nula, se puede calcular el intervalo de confianza para la diferencia de medias de dos muestras pareadas.

Modelo no paramétrico de series temporales funcionales

Uno de los métodos más comunes en el análisis de series temporales es el modelo de regresión no paramétrico. Este modelo puede aplicarse a series de tiempo funcionales de la siguiente manera:

$$X_{i+1}(t) = m(X_i) + \varepsilon_{i+1}(t) \quad (8)$$

$X_i(t)$ es el i -ésimo dato funcional, con t variando dentro de un intervalo de tiempo. La ecuación (8) modela una serie temporal para cada instante t , ajustada mediante un modelo no paramétrico y un término de error, $\varepsilon_{i+1}(t)$ (Melo Martínez et al., 2020).

En este estudio, se empleó el estimador de Nadaraya-Watson para $m(X)$, siguiendo las ideas de (Aneiros-Pérez et al., 2011), como se ilustra en la ecuación (9):

$$\hat{m}_h(X) = \sum_{i=1}^{N-1} w_h(X, X_i) X_{i+1}(t) \quad (9)$$

$$w_h(X, X_i) = \frac{K\left(\frac{d(X, X_i)}{h}\right)}{\sum_j K\left(\frac{d(X, X_j)}{h}\right)} \quad (10)$$

Donde:

h : apertura de suavizado

N : tamaño de muestra

K : función de núcleo

d : distancia entre curvas

El estudio se enfocó en el espacio L^2 de funciones cuadráticamente integrables, siguiendo las recomendaciones de (Beyaztas & Shang, 2020). Para ajustar el modelo, se minimizó el error cuadrático medio descrito en la ecuación (11), utilizando una ventana de suavizado estimada \hat{h} obtenida mediante validación cruzada (CV), como se indica en la ecuación (12).

$$MSE = \frac{1}{N} \sum_{i=1}^{N-1} \int (m(X_i(t)) - X_{i+1}(t))^2 dt \quad (11)$$

$$\hat{h} = \min_h CV(h) \quad (12)$$

donde,

$$CV(h) = \frac{1}{N} \sum_{i=1}^{N-1} \int (m^i(X_i(t)) - X_{i+1}(t))^2 dt$$

Siendo $m^i(X_i(t))$ el estimador de Nadaraya-Watson calculado excluyendo la observación i -ésima de los datos (Meneses Freire et al., 2022).

Metodología

La investigación es cuantitativa con un diseño experimental, utilizando la metodología bootstrap para calcular intervalos de predicción mediante modelos no paramétricos de regresión funcional aplicados a series temporales de humedad relativa.

Los datos utilizados corresponden al registro de la humedad relativa por segundo, obtenidos en la estación meteorológica de la Facultad de Ingeniería de la UNACH durante el año 2023. Se realizó una limpieza de valores atípicos en Excel y una interpolación de datos faltantes. Luego, se calculó la media horaria de la humedad relativa para cada mes del año 2023.

El procedimiento en el software estadístico R fue el siguiente:

- **Exploración gráfica:** Se llevaron a cabo gráficos exploratorios usando la librería *ggplot2* de los datos de humedad relativa para identificar patrones, tendencias y posibles anomalías.
- **Ajuste de modelos:** Se utilizaron B-splines para ajustar los modelos de regresión, que permiten la adaptación flexible de curvas en los datos observados, asegurando que los modelos capturen adecuadamente las características esenciales de cada serie temporal mensual de humedad relativa.
- **Generación de réplicas:** Se aplicó la metodología bootstrap, generando 5000 réplicas de las series temporales mensuales de humedad relativa. Cada réplica fue modelada usando los B-splines ajustados previamente, estableciendo una base sólida para la predicción.
- **Predicción:** Se utilizó el modelo no paramétrico funcional para predecir la serie temporal del mes de diciembre de 2023. Se generaron 5000 predicciones bootstrap, y se incluyó una predicción basada en los datos reales para comparar y evaluar la precisión y variabilidad de las predicciones obtenidas.
- **Suavización y precisión:** Se suavizaron las 5000 predicciones mediante bases de Fourier para mejorar la precisión.

Predicción de series temporales de humedad relativa mediante bootstrap y modelos de regresión funcional no paramétrica

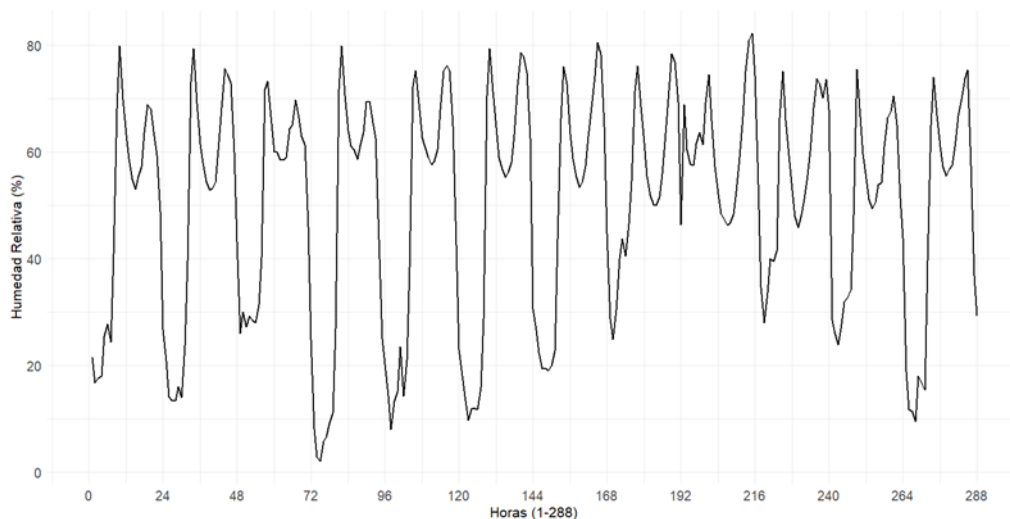
- **Cálculo de intervalos:** Finalmente, se calcularon intervalos de confianza del 95% para las predicciones, proporcionando un rango en el que se espera que se encuentren los valores de humedad relativa para diciembre de 2023.

Resultados y discusión

Gráfico exploratorio de los datos:

En la figura 2, se observa las doce series temporales de humedad relativa en porcentaje, representa la media horaria en cada mes del año 2023. Cada serie temporal presenta dos picos máximos en el transcurso del día. Este patrón es consistente a lo largo de todos los meses.

Figura 1: Curvas medias mensuales de humedad relativa en el año 2023



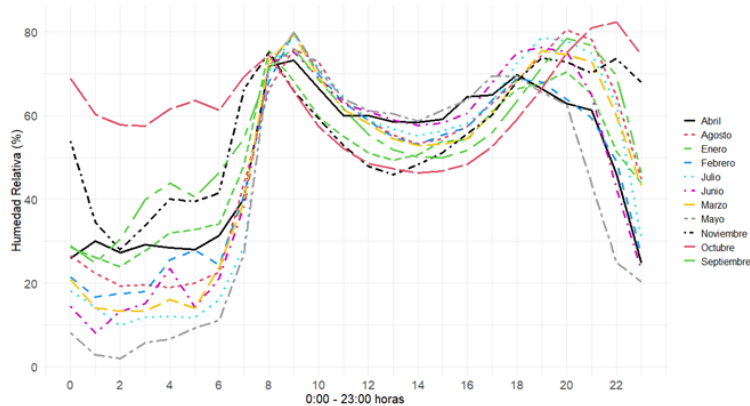
Realizado por: Zúñiga Jorge, 2024

Humedad relativa media en los meses de enero a noviembre

En la figura 3, se observa las series temporales de enero a noviembre de 2023. Cada serie temporal muestra un aumento gradual de porcentaje de humedad relativa desde las 00:00 horas alcanzando un pico máximo alrededor de las 8:00 horas y otro pico máximo alrededor de las 20:00 horas, luego disminuye gradualmente para el resto de la noche.

Predicción de series temporales de humedad relativa mediante bootstrap y modelos de regresión funcional no paramétrica

Figura 2: Perfil diario de humedad relativa media de enero a noviembre en el año 2023

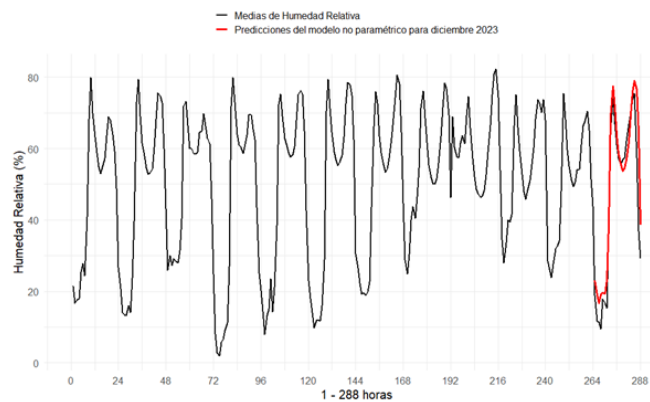


Realizado por: Zúñiga Jorge, 2024

Predicciones de humedad relativa con el modelo no paramétrico funcional para el mes de diciembre del año 2023

En la figura 4, se observa que la predicción de la serie temporal del mes de diciembre mediante el modelo no paramétrico funcional, es muy semejante a la serie temporal de los datos reales para este mes.

Figura: Medias de humedad relativa mensual con predicciones para diciembre del año 2023



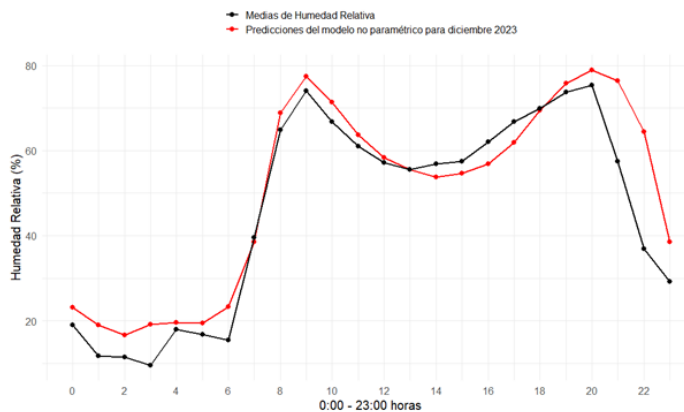
Realizado por: Zúñiga Jorge, 2024

Predicciones del modelo no paramétrico de regresión funcional de humedad relativa para el mes de diciembre del año 2023

En la figura 5, se observa la comparación gráfica entre la predicción del modelo no paramétrico de la serie temporal del mes de diciembre con los datos reales de este mes.

Predicción de series temporales de humedad relativa mediante bootstrap y modelos de regresión funcional no paramétrica

Figura 3: Serie temporal de humedad relativa y predicción del modelo no paramétrico de regresión funcional para diciembre de 2023

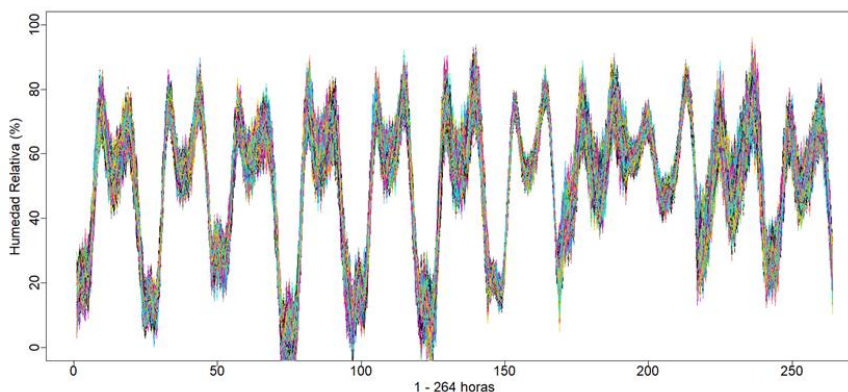


Realizado por: Zúñiga Jorge, 2024

Cálculo de 5000 simulaciones mediante modelos B-Spline de la serie temporal de humedad relativa de enero a noviembre del año 2023

En la figura 6, se observa la variación de las simulaciones del porcentaje de humedad relativa para cada mes del año 2023.

Figura 4: 5000 simulaciones mediante modelos B-Spline de la serie temporal de humedad relativa de enero a noviembre del año 2023



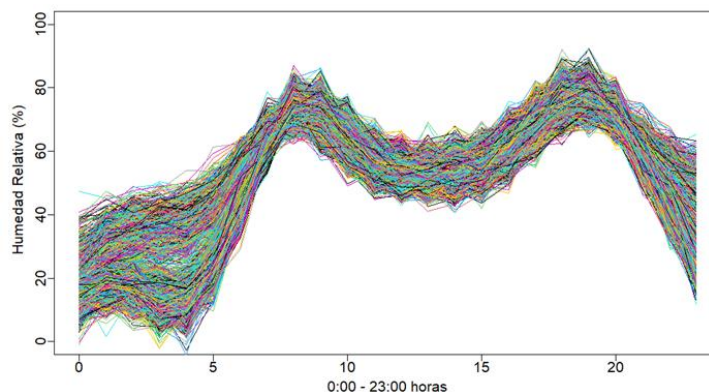
Realizado por: Zúñiga Jorge, 2024

Predicción de series temporales de humedad relativa mediante bootstrap y modelos de regresión funcional no paramétrica

Metodología Bootstrap para obtener 5000 réplicas de predicciones de series de tiempo de humedad relativa para diciembre del año 2023

En la figura 7, se observa las 5000 réplicas formando una banda de predicciones de series de tiempo de humedad relativa para diciembre 2023, tienen una ligera variación de las cuales se obtendrán los intervalos puntuales de predicción para cada hora.

Figura 5: Predicciones de series de tiempo utilizando el modelo no paramétrico funcional

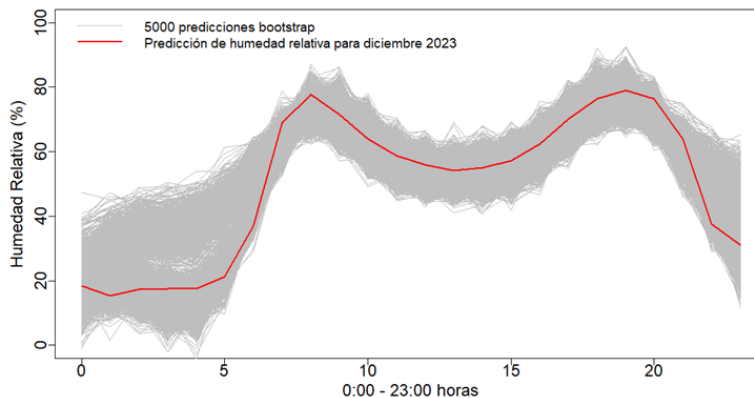


Realizado por: Zúñiga Jorge, 2024

Aplicación del modelo no paramétrico funcional para obtener 5000 réplicas Bootstrap de series de tiempo de humedad relativa para diciembre de 2023

En la figura 8, se observa que la predicción del modelo no paramétrico de humedad relativa del mes de diciembre, se encuentra dentro de los límites de la banda de las 5000 predicciones bootstrap.

Figura 6: 5000 predicciones bootstrap de humedad relativa junto a la predicción de diciembre 2023

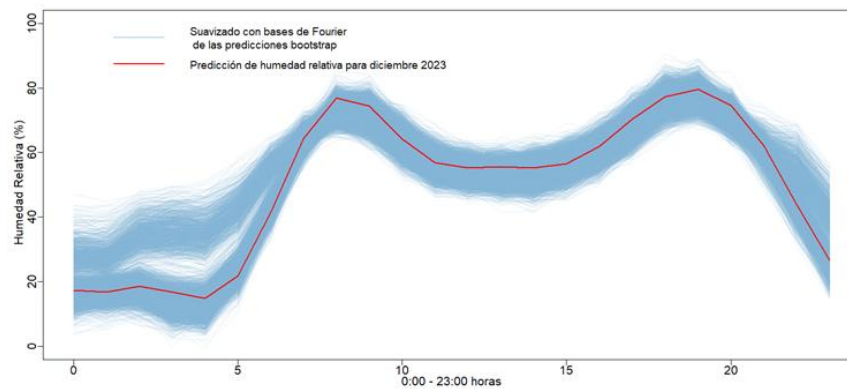


Realizado por: Zúñiga Jorge, 2024

Suavizado de las 5000 predicciones bootstrap con bases de Fourier

En la figura 9, se muestra un suavizado usando bases de Fourier de las 5000 réplicas Bootstrap, se aprecia la eliminación de picos e irregularidades presentes en la figura anterior, la predicción del modelo no paramétrico de humedad relativa del mes de diciembre, se mantiene dentro de los límites de la banda de predicción suavizada.

Figura 7: Suavizado de las 5000 predicciones bootstrap de humedad relativa con bases de Fourier

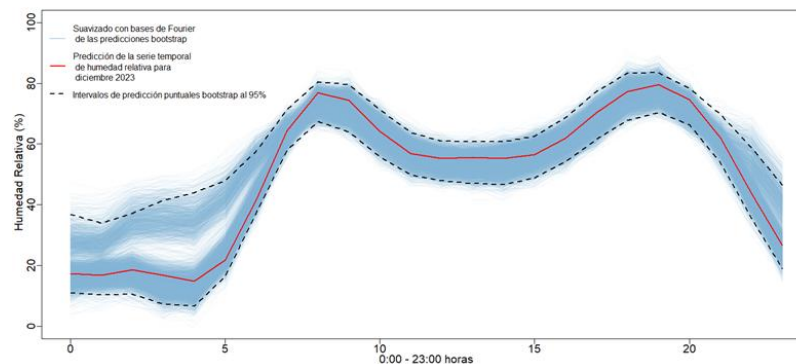


Realizado por: Zúñiga Jorge, 2024

Intervalos de predicción puntuales bootstrap al 95% de la serie temporal de humedad relativa.

En la figura 10, se observan los intervalos de predicción puntuales de 00:00 horas a 23:00 horas del día. La curva roja corresponde a la serie temporal de predicción de humedad relativa mediante el método no paramétrico y se encuentra dentro de los intervalos puntuales bootstrap al 95% de confianza.

Figura 8: Intervalos de predicción puntuales bootstrap al 95% de la serie temporal de humedad relativa.



Realizado por: Zúñiga Jorge, 2024

Conclusiones

En el presente estudio se observó que los datos de humedad relativa en series de tiempo son muy semejantes en cada mes del año 2023, presentando un patrón consistente con dos picos máximos, uno alrededor de las 8:00 horas y otro alrededor de las 20:00 horas. Sin embargo, se identifican ligeras variaciones en estos picos máximos en algunos meses. Estos hallazgos permiten comprender mejor los patrones de humedad relativa a lo largo del año y pueden ser útiles para futuras investigaciones en meteorología.

Mediante los modelos de regresión B-Spline se obtuvieron réplicas bootstrap muy similares a las series temporales de humedad relativa de cada mes. Estas réplicas se aplicaron mediante un método no paramétrico para generar una banda de predicción bootstrap, la cual contiene a la predicción de la serie temporal de humedad relativa del mes de diciembre.

Se obtuvieron intervalos de confianza puntuales bootstrap al 95% para la serie temporal de humedad relativa del mes de diciembre de 2023, demostrando que la predicción se encuentra dentro de estos intervalos. Esto confirma la fiabilidad del modelo en la estimación de los datos. Este intervalo de confianza proporciona una medida robusta de la variabilidad y la incertidumbre en las predicciones, validando así la efectividad del enfoque bootstrap en la modelización de datos de humedad relativa y en futuras predicciones de otras variables meteorológicas.

Referencias

1. Aneiros-Pérez, G., Cao, R., & Vilar-Fernández, J. M. (2011). Functional methods for time series prediction: A nonparametric approach. *Journal of Forecasting*, 30(4), 377-392. <https://doi.org/10.1002/for.1169>
2. Beyaztas, U., & Shang, H. L. (2020). On function-on-function regression: partial least squares approach. *Environmental and Ecological Statistics*, 27(1), 95-114. <https://doi.org/10.1007/S10651-019-00436-1/METRICS>
3. Boor, C. de. (2001). *A Practical Guide to Splines - Revised Edition*. En Springer-Verlag, New York (Vol. 115). Springer. <https://link.springer.com/book/9780387953663>
4. Chariguamán Maurisaca, N. E., & Meneses Freire, M. A. (2024). Modelos de regresión funcional con respuesta funcional aplicados a las variables meteorológicas, temperatura, humedad, presión y radiación. *Tesla Revista Científica*, 4(1), e317. <https://doi.org/10.55204/trc.v4i1.e317>

Predicción de series temporales de humedad relativa mediante bootstrap y modelos de regresión funcional no paramétrica

5. Horváth, L., & Kokoszka, P. (2012). Inference for Functional Data with Applications. 200. <https://doi.org/10.1007/978-1-4614-3655-3>
6. Melo Martínez, C. E., Diego, &, Malagón Márquez, A., David, D., & Forero, R. (2020). Interpoladores determinísticos espacio-temporales, series detiempo y análisis de datos funcionales para el estudio y predicción de la precipitación en Cundinamarca y Bogotá D.C. 2017-2020. <http://revistas.udistrital.edu.co/ojs/index.php/azimut>
7. Meneses, A., Zúñiga, L., Santos, C., Haro, S., Chariguamán, N., & Vera, L. (2018). Método bootstrap para hipótesis concernientes a la diferencia de medias para muestras pareadas: aplicaciones. *Perfiles*, 2(20), 100-106. <https://doi.org/10.47187/PERF.V2I20.39>
8. Meneses Freire, A., Muñoz Cargua, J., Zúñiga Lema, L., Lara Sinaluisa, J., & Acurio Martínez, W. (2022). Modelo no paramétrico funcional predictivo en series de tiempo funcionales. aplicación en variables meteorológicas. *Perfiles*, 1(28), 83-89. <https://doi.org/10.47187/PERF.V1I28.186>
9. Paluszny, M., Prautzsch, H., & Boehm, W. (2005). Métodos de Bézier y B-splines. Universitätsverlag Karlsruhe, 303. <https://doi.org/10.5445/KSP/1000002481>
10. Peña, D. (2015). Análisis de series temporales (Alianza Editorial). Alianza Editorial. <https://www.alianzaeditorial.es/libro/manuales/analisis-de-series-temporales-daniel-pena-9788420669458/>
11. Racine, J. S. (2022). A primer on regression splines. https://cran.r-project.org/web/packages/crs/vignettes/spline_primer.pdf
12. Racine, J. S. (2023). The crs Package. <https://doi.org/https://doi.org/10.32614/CRAN.package.crs>
13. Toalombo, B., Meneses, A., Zúñiga, L., & Espín, R. (2022). Modelos de regresión paramétricos polinomiales y no paramétricos b-splines. una aplicación en ingeniería. *Perfiles*, 1(28), 72-82. <https://doi.org/10.47187/PERF.V1I28.185>
14. Zúñiga, J., & Meneses, A. (2024). Bootstrap para calcular intervalos de predicción mediante modelos no paramétricos de regresión funcional de radiación solar. *CASADELPO*, 9, 3726-3740. <https://doi.org/10.23857/pc.v9i8.7909>