



Bootstrap para calcular intervalos de predicción mediante modelos no paramétricos de regresión funcional de radiación solar

Bootstrap for estimating prediction intervals using nonparametric functional regression models of solar radiation

Bootstrap para calcular intervalos de previsão usando modelos de regressão funcional não paramétricos de radiação solar

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

Manuel Antonio Meneses-Freire ^{II}
ameneses@unach.edu.ec
<https://orcid.org/0000-0003-3192-5084>

Correspondencia: david.zuniga@unach.edu.ec

Ciencias Técnicas y Aplicadas
Artículo de Investigación

* **Recibido:** 12 de junio de 2024 * **Aceptado:** 18 de julio de 2024 * **Publicado:** 30 de agosto de 2024

- I. Universidad Nacional de Chimborazo, Riobamba, Ecuador.
- II. Universidad Nacional de Chimborazo, Riobamba, Ecuador.

Resumen

La presente investigación tuvo como objetivo el desarrollo de intervalos de predicción en series de tiempo de radiación solar, aplicando la metodología bootstrap con modelos no paramétricos de regresión funcional. Para lo cual se registró por cada segundo la radiación solar durante el año 2023, en la estación meteorológica de la Facultad de Ingeniería de la Universidad Nacional de Chimborazo. Se realizó la limpieza de valores atípicos, la interpolación de datos faltantes y se calculó la media horaria mensual. Los modelos de regresión se ajustaron utilizando B-splines, que capturo adecuadamente las características esenciales de cada serie temporal mensual. Se generaron 5000 réplicas de las series temporales mediante bootstrap, la cuales se usaron para predecir la radiación solar de diciembre de 2023. La predicción obtenida fue comparada con los datos reales, que mostraron una alta consistencia. Además, se suavizaron las predicciones con bases de Fourier y se calcularon intervalos de confianza al 95%. Los resultados demuestran que la metodología empleada es fiable, ya que proporcionó intervalos de predicción robustos. Este enfoque no solo mejora la comprensión de los patrones de radiación solar a lo largo del año, sino que también ofrece una herramienta valiosa para futuras aplicaciones en meteorología.

Palabras clave: radiación solar; series de tiempo; bootstrap; regresión funcional; modelo no paramétrico; Fourier.

Abstract

The present research aimed to develop prediction intervals in solar radiation time series, applying the bootstrap methodology with non-parametric functional regression models. For this, solar radiation was recorded for every second during the year 2023, at the meteorological station of the Faculty of Engineering of the National University of Chimborazo. Outliers were cleaned, missing data was interpolated, and the monthly hourly average was calculated. The regression models were adjusted using B-splines, which adequately captured the essential characteristics of each monthly time series. 5,000 replicas of the time series were generated using bootstrap, which were used to predict solar radiation for December 2023. The prediction obtained was compared with the real data, which showed high consistency. In addition, the predictions were smoothed with Fourier bases and 95% confidence intervals were calculated. The results demonstrate that the methodology used is reliable, since it provided robust prediction intervals. This approach not only improves the

understanding of solar radiation patterns throughout the year, but also offers a valuable tool for future applications in meteorology.

Keywords: solar radiation; time series; bootstrap; functional regression; nonparametric model; Fourier.

Resumo

O objetivo desta investigação foi desenvolver intervalos de previsão em séries temporais de radiação solar, aplicando a metodologia bootstrap com modelos de regressão funcional não paramétricos. Para o qual foi registada a radiação solar a cada segundo durante o ano de 2023, na estação meteorológica da Faculdade de Engenharia da Universidade Nacional de Chimborazo. Foi realizada a limpeza de outliers, interpolação de dados em falta e cálculo da média horária mensal. Os modelos de regressão foram ajustados através de B-splines, que capturaram adequadamente as características essenciais de cada série temporal mensal. Foram geradas 5000 réplicas da série temporal através de bootstrap, que foram utilizadas para prever a radiação solar para dezembro de 2023. A previsão obtida foi comparada com os dados reais, que apresentaram uma elevada consistência. Além disso, as previsões foram suavizadas com bases de Fourier e foram calculados intervalos de confiança a 95%. Os resultados demonstram que a metodologia utilizada é fiável, uma vez que forneceu intervalos de previsão robustos. Esta abordagem não só melhora a compreensão dos padrões de radiação solar ao longo do ano, como também oferece uma ferramenta valiosa para futuras aplicações em meteorologia.

Palavras-chave: radiação solar; séries temporais; inicialização; regressão funcional; modelo não paramétrico; Fourier.

Introducción

El estudio del comportamiento de las variables meteorológicas han sido un tema de gran importancia, debido a la aleatoriedad de datos influenciada por varios factores externos. Tradicionalmente, se han ajustado modelos de regresión de series de tiempo con resultados satisfactorios, especialmente para predicciones a corto y mediano plazo. Sin embargo, al intentar ampliar el tiempo de análisis, las predicciones tienden a desviarse significativamente (Chariguamán Maurisaca & Meneses Freire, 2024). En este contexto, la presente investigación se

centra en la aplicación de un modelo no paramétrico funcional para series de tiempo de radiación solar, con el objetivo de mejorar la precisión de las predicciones. A través de la metodología bootstrap y la utilización de modelos B-Spline, se busca generar intervalos de predicción más confiables y representativos. Este enfoque no paramétrico adaptado a datos funcionales proporciona una herramienta robusta para el análisis y predicción de series temporales.

En esta investigación se debe definir algunos conceptos importantes que se detallan a continuación:

Modelos de regresión no paramétrico B-Spline

Los modelos de regresión no paramétricos ofrecen una estimación adaptable de efectos no lineales, sin imponer restricciones relacionadas con una forma funcional específica. Cuando se trata de una única covariable continua "x" (Toalombo et al., 2022), el modelo estándar para la regresión no paramétrica se describe de la siguiente manera:

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

En relación con la variable de error ϵ_i , se asume que la función f posee ciertas propiedades de suavidad, como la continuidad o diferenciabilidad, sin que se especifique una forma paramétrica concreta (Horváth & Kokoszka, 2012). Un spline es una función que se construye en segmentos a partir de funciones polinomiales, y su nombre se origina del concepto de "listón elástico" (Racine, 2023). Los splines "son curvas polinómicas por secciones que son continuamente diferenciables hasta un orden determinado" (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 máxima diferenciabilidad, que actúa como una extensión de la curva de Bézier. Los B-splines se caracterizan por su orden m y la cantidad de nodos interiores N (existen dos puntos extremos que también son nodos, por lo que el total de nodos será $N + 2$) (Peña, 2015). El grado del polinomio B-spline corresponde al orden de la spline menos uno, es decir, $m-1$ (Racine, 2022). Una B-spline de grado n (con un orden de spline $m = n+1$) es una curva paramétrica formada por una combinación lineal de B-splines base $B_{i,n}(x)$ de grado n (Boor, 2001). Se expresa de la siguiente manera:

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

La función $f(x)$ tiene la siguiente estructura:

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

Donde:

K: números 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 se denomina 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 de la siguiente manera:

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

Donde:

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

El spline $f(x)$ se puede representar 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 que se presentan a continuación conforman una base de funciones polinómicas para el spline:

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

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

La Figura 1 muestra las curvas B-spline de base correspondientes a un grado específico del polinomio, el número de vértices del polígono de control y los nodos internos.

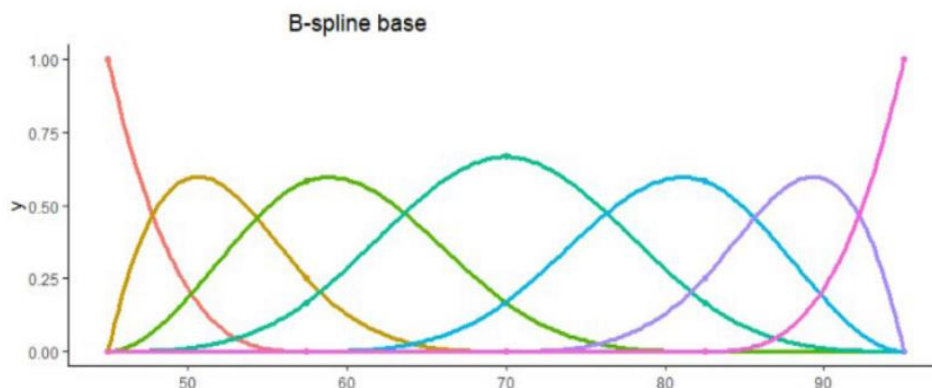


Figura 1. B-spline de base para un modelo de regresión.

Fuente: (Toalombo et al., 2022)

Metodología Bootstrap

El método Bootstrap es una técnica estadística utilizada para aproximar la distribución de muestreo, generalmente de un estadístico. Este método se basa en remuestreo, lo que implica generar nuevas muestras de manera aleatoria a partir de la muestra original ((Meneses et al., 2018). Basándose en los principios generales de este método y manteniendo la hipótesis nula, se calcula el intervalo de confianza para la diferencia de medias de dos muestras pareadas.

Modelo no paramétrico de series de tiempo funcionales

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

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

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

En esta investigación, se utilizó el estimador de Nadaraya-Watson para $m(X)$, basado en las ideas de (Aneiros-Pérez et al., 2011), tal como se muestra 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 : ventana de suavizado

N : tamaño de la muestra

K : función de núcleo

d : semimétrica que mide la distancia entre curvas utilizando el método de análisis de componentes principales funcionales

El trabajo se centró en el espacio L^2 de funciones integrables módulo al cuadrado como lo sugiere (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 a través de validación cruzada (CV), como se muestra 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).

Suavizado mediante bases de Fourier

El suavizado con bases de Fourier es una técnica utilizada para representar y analizar funciones o señales mediante una combinación de senos y cosenos de diferentes frecuencias. Esta técnica es especialmente útil para el análisis de datos periódicos o cíclicos. Se usa para descomponer una serie de tiempo en componentes cíclicos y eliminar fluctuaciones ruidosas (Pérez González, 2006).

Metodología

La investigación es cuantitativa con diseño experimental, se aplica la metodología bootstrap para calcular intervalos de predicción utilizando modelos no paramétricos de regresión funcional en series de tiempo de variables meteorológicas.

Los datos son obtenidos representan el registro por segundo de la radiación solar, que han sido obtenidos en la estación meteorológica de la Facultad de Ingeniería de la UNACH durante el año 2023, a los cuales fueron se realizó una limpieza de valores atípicos en Excel y la interpolación de datos faltantes. Posteriormente, se calculó la media de radiación solar por horas para cada mes del año 2023.

El procedimiento en el software estadístico R fue el siguiente:

- Exploración gráfica de los datos procesados de radiación solar, para identificar patrones, tendencias y posibles anomalías.
- Se emplearon B-splines para el ajuste de modelos de regresión, debido a su capacidad para adaptar curvas suaves y flexibles a los datos observados, asegurando que los modelos capturen adecuadamente las características esenciales de cada serie temporal mensual de radiación solar.
- Aplicación de la metodología bootstrap, generando 5000 réplicas de las series temporales mensuales de radiación solar. Cada réplica fue modelada utilizando los B-splines ajustados previamente, proporcionando una base robusta para la predicción.
- Aplicación del modelo no paramétrico funcional para predecir la serie temporal correspondiente al mes de diciembre del año 2023. Esto se realizó para cada una de las réplicas, obteniendo así 5000 predicciones bootstrap. Se incluyó también una predicción basada en los datos reales, permitiendo comparar y evaluar la precisión y variabilidad de las predicciones obtenidas a través del bootstrap.
- Suavizar las irregularidades presentes en las 5000 predicciones y mejorar la precisión mediante bases de Fourier.
- Finalmente, se calcularon bandas de confianza al 99% para las predicciones, proporcionando un intervalo donde se encuentren los valores de radiación solar para el mes de diciembre del año 2023.

Resultados y discusión

Gráfico exploratorio de los datos:

En la figura 2, se observa las doce series temporales de radiación solar, cada una representa la media de radiación solar por hora para cada mes del año 2023. Cada serie temporal presenta un pico alrededor del mediodía. Este patrón es consistente a lo largo de todos los meses.

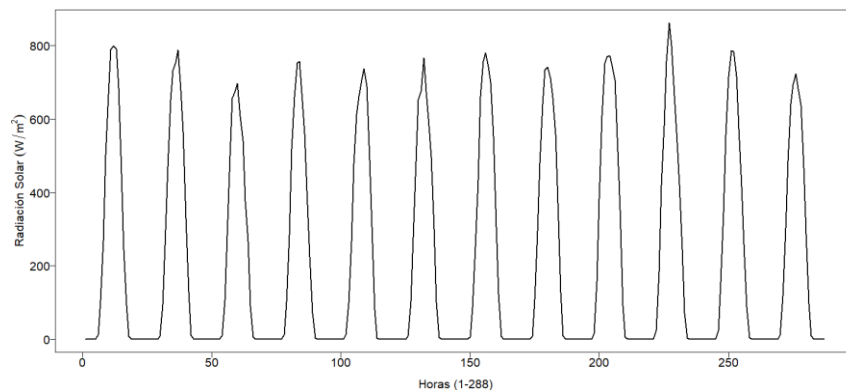


Figura 2: Curvas medias mensuales de radiación solar en el año 2023
Realizado por: Zúñiga Jorge, 2024

Radiación solar media en los meses de enero a noviembre

En la figura 3, se observa las series temporales de enero a noviembre de 2023, son semejantes a una forma de campana. Cada serie temporal muestra un aumento gradual de la radiación solar desde las 6:00 horas alcanzando un pico máximo alrededor del mediodía, y luego disminuyendo gradualmente hasta las 18:00 horas.

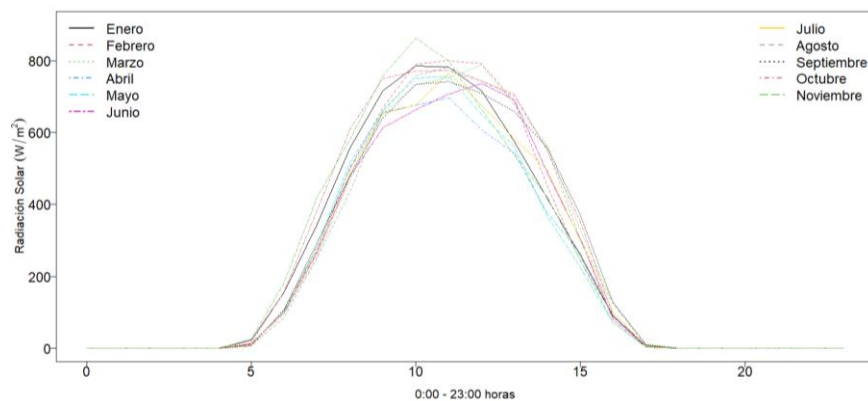


Figura 3: Perfil diario de radiación solar media de enero a noviembre en el año 2023
Realizado por: Zúñiga Jorge, 2024

Predicciones de radiación solar 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.

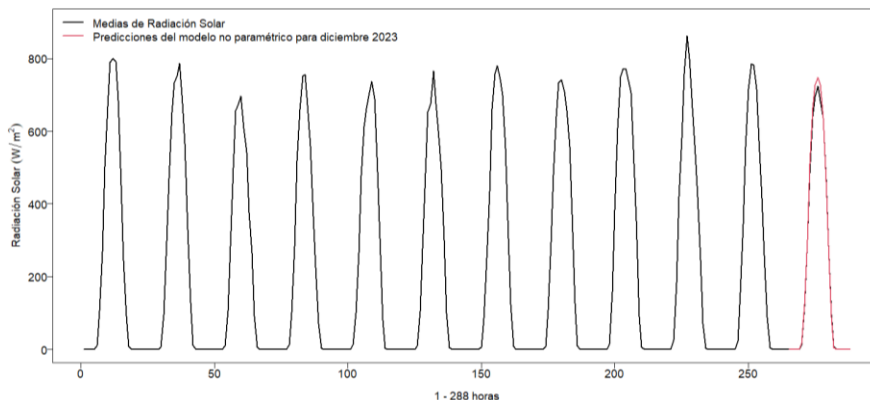


Figura 4: Medias de radiación solar mensual con predicciones para diciembre del año 2023
Realizado por: Zúñiga Jorge, 2024

Predicciones del modelo no paramétrico de regresión funcional de radiación solar 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.

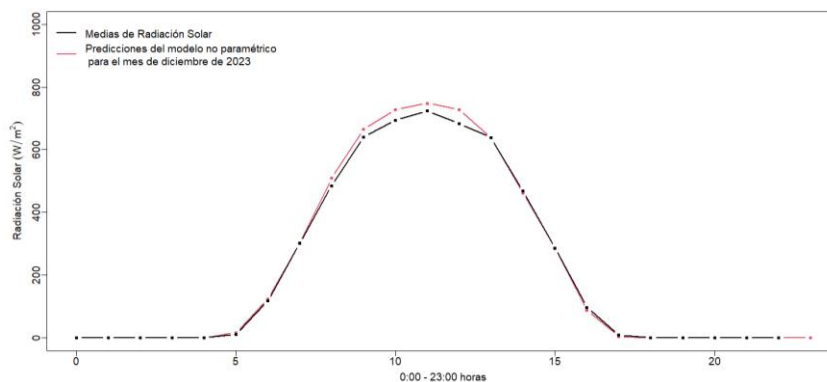


Figura 5: Serie temporal de radiación solar y predicción del modelo no paramétrico de regresión funcional para diciembre de 2023
Realizado por: Zúñiga Jorge, 2024

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

En la figura 6, se observa la variación de las simulaciones de la radiación solar para cada mes del año 2023.

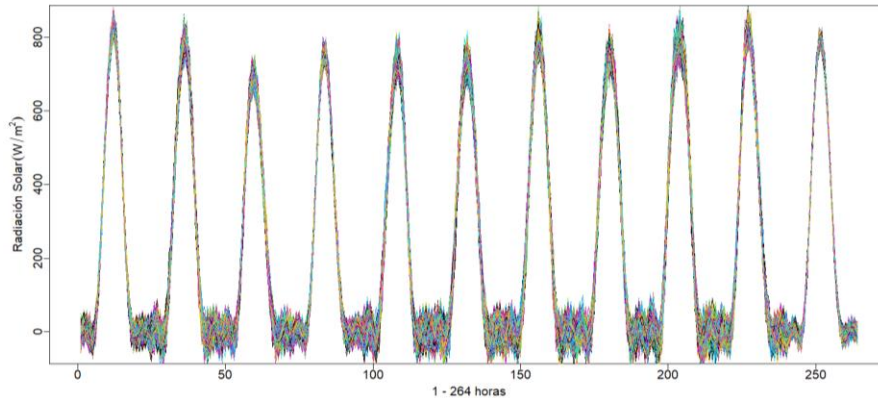


Figura 6: 5000 simulaciones mediante modelos B-Spline de la serie temporal de radiación solar de enero a noviembre del año 2023

Realizado por: Zúñiga Jorge, 2024

Metodología Bootstrap para obtener 5000 réplicas de predicciones de series de tiempo de radiación solar para diciembre del año 2023

En la figura 7, se observa las 5000 réplicas formando una banda de predicciones de series de tiempo de radiación solar para diciembre, ya que son muy semejantes, con una ligera variación de las cuales se obtendrán los intervalos puntuales de predicción para cada hora.

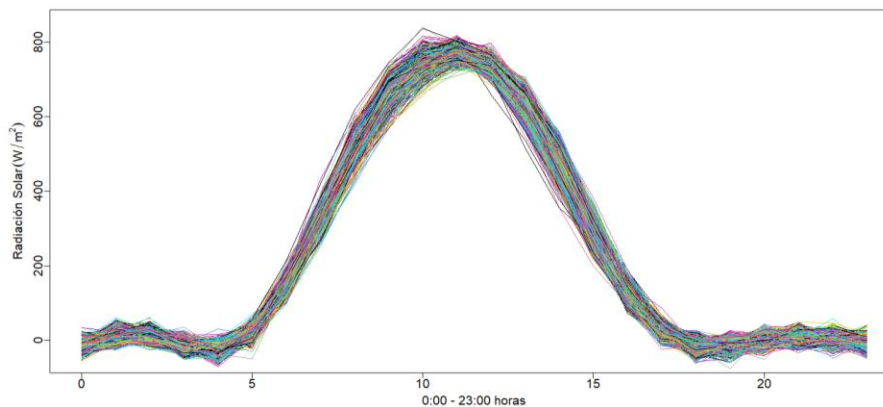


Figura 7: Predicciones de series de tiempo de radiación solar utilizando el modelo no paramétrico funcional

Realizado por: Zúñiga Jorge, 2024

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

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

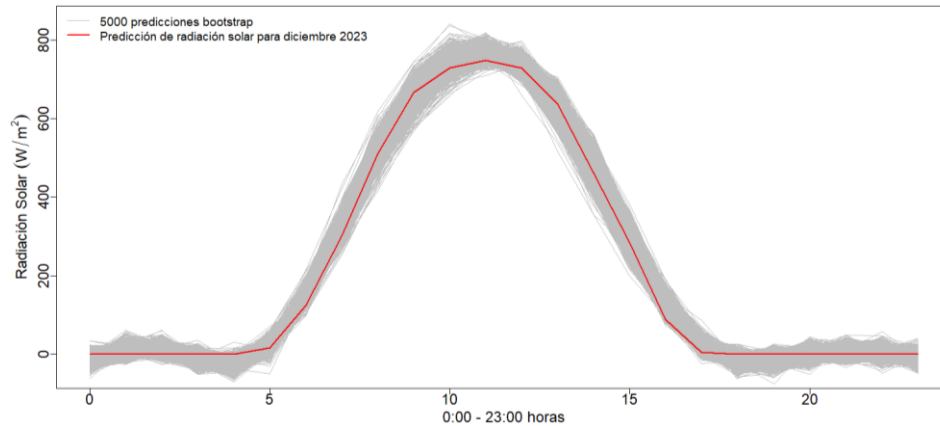


Figura 8: 5000 predicciones bootstrap de radiación solar junto a la predicción de diciembre 2023
Realizado por: Zúñiga Jorge, 2024

Suavizado de las 5000 predicciones bootstrap con bases de Fourier

En la figura 9, para mejor representación de la banda se realiza un suavizado mediante bases de Fourier de las 5000 réplicas Bootstrap, se aprecia que luego de este proceso, la predicción del modelo no paramétrico de radiación solar del mes de diciembre, se mantiene dentro de los límites de la banda de predicción suavizada.

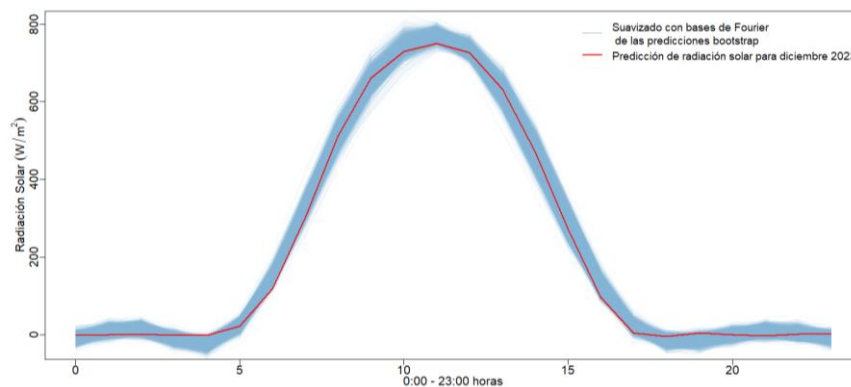


Figura 9: Suavizado de las 5000 predicciones bootstrap de radiación solar con bases de Fourier y predicción para diciembre 2023
Realizado por: Zúñiga Jorge, 2024

Intervalos de predicción puntuales bootstrap al 95% de la serie temporal de radiación solar

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 radiación solar mediante el método no paramétrico, se encuentra dentro de los intervalos puntuales bootstrap al 95% de confianza.

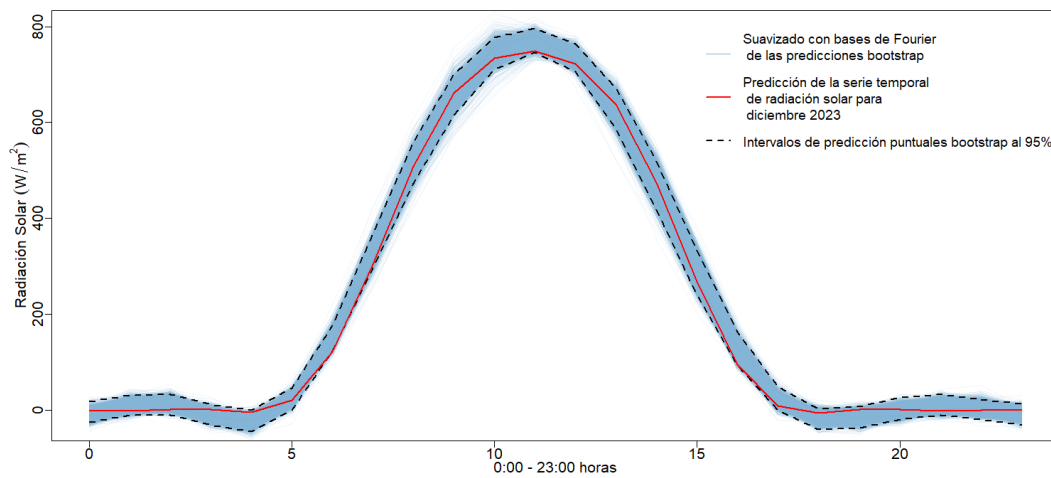


Figura 10: Intervalos de predicción puntuales bootstrap al 95% de la serie temporal de radiación solar.
Realizado por: Zúñiga Jorge, 2024

Conclusiones

En el presente estudio se observó que los datos de radiación solar en series de tiempo son muy semejantes en cada mes del año 2023, presentando un patrón consistente con picos alrededor del mediodía. Sin embargo, se identifican ligeras variaciones en los picos máximos de radiación solar en algunos meses. Estos hallazgos permiten comprender mejor los patrones de radiación solar a lo largo del año y pueden ser útiles para futuras investigaciones en meteorología.

Mediante los modelos de regresión B-Spline se obtuvo réplicas bootstrap muy similares a las series temporales de radiación solar 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 radiación solar del mes de diciembre.

Se obtuvieron intervalos de confianza puntuales bootstrap al 95% para la serie temporal de radiación solar 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 radiación solar 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>
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 de tiempo 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. Pérez González, J. (2006). Ecuaciones diferenciales series de fourier transformadas de fourier y laplace. https://www.ugr.es/~fjperez/textos/eedd_laplace_fourier.pdf
12. Racine, J. S. (2022). A primer on regression splines. https://cran.r-project.org/web/packages/crs/vignettes/spline_primer.pdf
13. Racine, J. S. (2023). The crs Package. <https://doi.org/https://doi.org/10.32614/CRAN.package.crs>
14. 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>

© 2024 por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0) (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).