



Rafael Meléndez Surmay



Luis Castillo Suarez



Albert De La Hoz Navas



Universidad de La Guajira, Colombia

OPEN ACCESS

Recibido: 15/01/2024

Aceptado: 10/04/2024

Publicado: 08/05/2024

Correspondencia de
autores:

rmelendez@uniguajira.edu.co



Copyright 2024
by Investigación
e Innovación en
Ingenierías

Resumen

Objetivo: Comparar la variación espacio temporal de los niveles de partículas suspendidas PM_{10} y de dióxido de nitrógeno NO_2 en ciudades chilenas durante el periodo del 2021 a través del enfoque del Análisis de Datos Funcionales ADF. **Metodología:** A partir de la base de datos de PM_{10} y NO_2 se aplicaron pruebas de hipótesis para datos funcionales para contrastar entre ciudades y métodos gráficos funcionales para detectar curvas atípicas. **Resultados:** Se evidencia diferencias significativas entre las curvas medias de NO_2 entre la ciudad de Valparaíso y la Región Metropolitana de Santiago RMS, mientras que la prueba ANOVA funcional muestra significancia de las curvas medias de PM_{10} entre las cuatro ciudades estudiadas. Finalmente, los diagramas de cajas funcionales identificaron dos estaciones atípicas de PM_{10} (El Bosque y Pudahuel) en la RMS. **Conclusiones:** Utilizar el análisis de datos funcionales permite tener una visión global del problema al considerar cada objeto funcional como unidad de estudio y facilita identificar aquellas estaciones donde hay producción extrema de particulado PM_{10} evidenciando un factor de riesgo de morbilidad y mortalidad.

Palabras clave: Análisis de datos funcionales, diagramas de caja para datos funcionales, ANOVA para datos funcionales, calidad del aire.

Abstract

Objective: This work aims to compare the spatio-temporal variation of the levels of particulate matter PM_{10} and nitrogen dioxide NO_2 in Chilean cities during 2021 through the ADF Functional Data Analysis approach. **Methodology:** Hypothesis tests for functional data and functional graphical methods are applied to detect atypical curves from the PM_{10} and NO_2 database. **Results:** significant differences are evident between the mean NO_2 curves between the city of Valparaíso and the Metropolitan Region of Santiago RMS, while the functional ANOVA test shows the significance of the mean PM_{10} curves between the four cities studied. Finally, functional box plots identified two PM_{10} outlier stations (El Bosque and Pudahuel) in the RMS. **Conclusions:** Using the analysis of functional data allows us to have a global vision of the problem by considering each functional object as a unit of study and facilitates identifying those stations where there is extreme production of PM_{10} particles, evidencing a risk factor for morbidity and mortality.

Keywords: functional data analysis; boxplots for functional data; ANOVA for functional data; air quality.

Como citar (IEEE): R. Meléndez Surmay, L. Castillo Suarez, and A. De La Hoz Navas, "Monitoreo de la calidad del aire urbano utilizando análisis de datos funcionales" Investigación e Innovación en Ingenierías, vol. 12, no. 1, pp. 71–82, 2024, doi: <https://doi.org/10.17081/invinno.12.1.7067>

Introducción

La contaminación del aire es un problema ambiental de gran relevancia en las grandes ciudades [1,2]. El aire nunca está perfectamente limpio y por lo tanto la contaminación del aire es una amenaza continua para la salud y el bienestar de los seres humanos [3,4]. Durante las últimas décadas, ha crecido el número de estudios que señalan que la calidad del aire es un problema importante en las ciudades de todo el mundo [5]. Además, existe un consenso general sobre los efectos nocivos sobre la salud de los seres humanos [6]. Incluso en el mejor de los casos, este problema aparentemente endémico que afecta la vida en las grandes ciudades ya se considera una de las principales causas de mortalidad tanto directa como indirecta [7].

La información sobre contaminación de la calidad aire, como la producida por monóxido de carbono (CO), óxidos de nitrógeno (NO y NO₂), dióxido de azufre (SO₂), ozono (O₃) y material particulado (PM₁₀) y (PM_{2.5}), es cada vez más importante debido a los efectos nocivos sobre salud humana [2].

En la actualidad se puede obtener mediciones automatizada de la concentración de estos contaminantes del aire, las cuales se registran en tiempo real. A su vez, permite a los ciudadanos informarse de posibles daños colaterales relacionado con altos niveles de contaminantes presente en el aire en los lugares donde habitan [3,4].

Los sistemas de previsión de la calidad del aire permiten avisar a la población a corto plazo de los próximos episodios de alta contaminación, de manera que se puedan tomar las medidas adecuadas para minimizar en lo posible los daños causados por estos episodios. Por ejemplo, la ciudad de Madrid, para cumplir con la normativa de Unión Europea,[8] elaboró un protocolo de calidad del aire que incluye restricciones al uso de vehículos cuando las concentraciones de NO₂ alcanzan determinados umbrales [9].

En la actualidad se desarrollan nuevas técnicas de análisis de datos funcionales (ADF) que presentan muchas aplicaciones en las ciencias e ingeniería. Un aspecto importante en la ADF es la forma de curvas observadas que resumen las propiedades de los datos de alguna manera específica. Por ejemplo, puede ser interesante saber cómo afecta un determinado tratamiento al nivel medio de proteínas de los pacientes. Uno estaría interesado en construir una curva que describa cómo cambia este nivel a lo largo del tiempo en un grupo de pacientes [10]. La aparición de esta nueva metodología estuvo motivada por la ineficiencia de las técnicas clásicas de minería de datos para abordar el tratamiento de estos cuando involucran valores cronológicos o sucesiones vectoriales [11]. Las aplicaciones del ADF son tan diversas que a menudo se utilizan para la investigación medioambiental y médica, sensores electrónicos y métodos industriales [12].

El abordar el análisis de particulado en el contexto del modelo funcional presenta dos características importantes: en primer lugar, tiene en cuenta la estructura de correlación temporal de los datos y, en segundo lugar, las comparaciones son más generales, conduciendo a una visión global del problema [11].

El propósito de esta investigación es examinar los niveles de partículas suspendidas PM₁₀ y dióxido de nitrógeno NO₂ a lo largo del tiempo, tratándolos como objetos funcionales, y comparar las curvas resultantes entre diferentes ciudades. Utilizamos pruebas de hipótesis no paramétricas diseñadas para datos funcionales con el fin de contrastar las curvas medias obtenidas de las muestras. Adicionalmente, aplicamos técnicas gráficas para identificar curvas atípicas que se desvían de la norma establecida, basándonos en los métodos propuestos por [13].

Metodología

En la actualidad los sistemas de monitoreo ambiental proporcionan información en tiempo real permitiendo crear grandes bases de datos. Sin embargo, es muy habitual encontrar datos faltantes en este tipo de estudios. Esto no permite obtener información completa para su análisis, aunque en el ADF la ausencia de observaciones discretas no es un problema grave, dado que se puede recuperar las curvas a través de la expansión de Karhunen-Loève.

Para la realización de este trabajo se tomaron los datos a través del Sistema de Información de Calidad del Aire SINCA de Chile para el periodo del 2021 que incluye los centros urbanos más poblados. Estos archivos de EXCEL extensión *csv* fueron monitoreados diariamente, por comunas y estaciones con las mediciones de PM_{10} y NO_2 .

Inicialmente, se obtiene una serie discretizadas de datos perteneciente al periodo del 2021, los cuales se transforman en objetos funcionales utilizando la técnica expansión de bases B-spline, que consiste en estimar cada curva como combinación lineal de funciones bases,

$$x_i(t) \approx \sum_{m=1}^k C_{im} \phi_m(t) \quad 1 \leq i \leq N, \quad (1)$$

Donde,

N = número total de curvas.

C_{im} = coeficientes estimados para k es suficientemente grande.

$\phi_m(t)$ = son las funciones bases seleccionada.

En el ADF los métodos más utilizados son las bases de Fourier, B-splines, wavelet y de polinomios. Por otra parte, los coeficientes de base suelen estimarse por el método de mínimos cuadrados (con o sin penalización) a partir de las observaciones con ruido en el tiempo discretizado.

Prueba para dos muestras funcionales

El problema general de dos muestras para datos funcionales con una función de covarianza común se puede formular de la siguiente manera. Supongamos que tenemos dos muestras funcionales que provienen de un proceso estocástico como sigue a continuación:

$$\{x_{1i}(t)\}_{i=1,\dots,N} \quad y \quad \{x_{2i}(t)\}_{i=1,\dots,M} \quad (2)$$

Curva media estimada

$$\hat{\mu}(t) = \frac{1}{n} \sum_{i=1}^n x_i(t)$$

Función de covarianza

$$\hat{\gamma}(s, t) = \frac{1}{n-1} \sum_{i=1}^2 \sum_{j=1}^{n_i} [x_{ij}(s) - \bar{x}_i(s)] [x_{ij}(t) - \bar{x}_i(t)]$$

donde $n_1(t)$ y $n_2(t)$ son las funciones medias desconocidas de las dos muestras con función de covarianza común generalmente desconocida. Entonces la hipótesis de interés es la siguiente;

$$H_0: \mu_1(t) = \mu_2(t) \text{ vs } H_1: \mu_1(t) \neq \mu_2(t) \quad (3)$$

Prueba t punto a punto

El problema de comparar dos muestras de datos funcionales con una función de covarianza común puede abordarse a través de dos enfoques distintos: el enfoque paramétrico y el no paramétrico. En la práctica, el supuesto de una distribución gaussiana no siempre es válido. Por lo tanto, los trabajos existentes se centran en la detección de la igualdad de las curvas medias mediante pruebas univariadas punto a punto, bajo el supuesto de que las muestras provienen de un proceso estocástico gaussiano. Además, el enfoque paramétrico, según se discute en el trabajo [14], considera la prueba de dos muestras basada en la norma L^2 .

Por otro lado, en el enfoque no paramétrico para datos funcionales, se emplean pruebas puntuales t y z , utilizando métodos de permutación o remuestreo "bootstrap" en el caso de dos muestras (3) bajo diversas condiciones. La idea fundamental detrás de una prueba puntual es evaluar la hipótesis nula en cada punto del tiempo, lo que descompone el problema en subproblemas individuales en cada punto t , como se describe a continuación.

$$H_{0t}: \mu_1(t) = \mu_2(t) \text{ vs } H_{1t}: \mu_1(t) \neq \mu_2(t). \quad (4)$$

Con base a las funciones media muestrales y la función de covarianza común muestral, el estadístico de la prueba pivotal está dado por:

$$Z(t) = \frac{\sqrt{n} \{ \bar{x}_1(t) - \bar{x}_2(t) \}}{\sqrt{\{1/n_1 + 1/n_2\} \hat{\gamma}(t, t)}} \quad (5)$$

Donde,

n_1, n_2 = tamaño respectivo de las muestras.

$\hat{\gamma}(t, t)$ = función de covarianza común.

En muchas situaciones, las muestras son pequeñas y el supuesto gaussiano no siempre es válido. Es decir, el supuesto se cumple aproximadamente. Entonces para cierta, se tiene $Z(t) \sim t_{(n-2)}, t \in T$. Entonces [14] presentó un procedimiento de remuestreo para probar la igualdad de las curvas medias. En este caso se calculan un número de B veces muy grande las medias muestrales funcionales $x_1^{-B}(t)$ y $x_2^{-B}(t)$ con $B \rightarrow \infty$ y se estima la función de covarianza común $\gamma^B(s, t)$. Entonces, se procede a estimar la estadística para cada punto t . Finalmente se obtienen los valores críticos de la distribución empírica estimada.

Luego, la prueba de remuestreo funcional puntual esta dada por

$$Z^B(t) = \frac{\{\bar{x}_1^B(t) - x_2^B(t)\} - \{\bar{x}_1(t) - \bar{x}_2(t)\}}{\sqrt{\{1/n_1 + 1/n_2\}} \hat{\gamma}^B(t,t)}. \quad (6)$$

Se realiza el cálculo de $Z^B(t)$. Para asegurar que la distribución de $Z^B(t)$ se ajuste adecuadamente a la distribución original $Z(t)$, es necesario repetir el proceso de remuestreo múltiples veces. Luego, se calcula el percentil 100 $(1 - \alpha/2)$ para determinar las bandas de confianza basadas en el remuestreo, lo que permite realizar inferencias pertinentes.

ANOVA para datos funcionales

En los modelos lineales clásicos, el análisis de varianza de una vía (ANOVA) es una técnica utilizada para comparar las medias de tres o más muestras. Las muestras se agrupan de acuerdo con una variable categórica, conocida como factor. A continuación, se muestra esta técnica extendida al caso funcional.

El Análisis de Varianza Funcional (FANOVA) es una herramienta que se utiliza para efectuar comparaciones entre k grupos de curvas. Aunque existen métodos convencionales de prueba funcional, como la F funcional presentada en [15] y la estadística de prueba funcional propuesta por [16], estos se concentran en evaluar respuestas a lo largo de todo el dominio temporal. Por otro lado, [17] introdujo una técnica para realizar inferencias en contextos de FANOVA sobre subdominios específicos. No obstante, es importante destacar que estos métodos no incorporan la técnica de prueba punto a punto discutida en [18,19], que facilita la realización de comparaciones directas entre medias funcionales.

ANOVA funcional

La prueba ANOVA de una vía para datos funcionales es una extensión natural de la prueba F clásica ANOVA para variables reales. Por lo tanto, se realiza de manera similar a la prueba punto a punto la prueba F propuesta por [20].

Para definir el problema ANOVA de una vía para datos funcionales asumimos que se tiene k muestras de curvas independientes.

Donde $n_i(t)$ es la función media del grupo i en el tiempo t , $i = 1, \dots, k$, j indexa una respuesta funcional dentro de un grupo $j = 1, \dots, n_i$, $y_{ij}(t)$ la función residual.

Entonces, para algún tiempo t de interés, especificado como $[a, b]$. Adicionalmente, el FANOVA también se conoce como el problema de k muestras para datos funcionales. La hipótesis de interés para este caso está dada por,

$$H_0 = \mu_1(t) = \dots = \mu_k(t), \text{ con } t \in T \quad (7)$$

El problema de k muestras de curvas tiene como objetivo verificar si el efecto de un factor o un tratamiento es estadísticamente significativo.

Modelo de efecto principal

En el caso del ANOVA clásico para valores continuo satisfacen

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij}, i = 1, \dots, k, j = 1, \dots, n_i \quad (8)$$

donde μ se conoce como la función media global de las k muestras y α_i es la i -ésima función de efecto principal para $i = 1, \dots, k$. De manera análoga, el modelo (8) se puede escribir como el siguiente modelo ANOVA de una vía estándar para datos funcionales:

$$x_{ij}(t) = \mu(t) + \alpha_i(t) + \epsilon_{ij}(t), i = 1, \dots, k, j = 1, \dots, n_i \quad (9)$$

En esta ecuación, la hipótesis nula (8) se puede expresar de manera equivalente como

$$\alpha_1(t) \equiv \alpha_2(t) \equiv \dots \equiv \alpha_k(t) \equiv 0, \quad t \in T \quad (10)$$

es decir, probar si las funciones de efecto principal son iguales a 0.

Prueba par a par

Cuando se rechaza (7), es necesario comparar par a par las curvas medias de los tratamientos. Por ejemplo, uno puede querer saber si dos funciones de efecto principal $\alpha_i(t)$ y $\alpha_j(t)$ son iguales, donde i y j son dos enteros tales que $1 \leq i < j \leq k$. Esta prueba se puede escribir como

$$H_0: \alpha_i(t) \equiv \alpha_j(t), t \in T$$

Para el problema k muestras funcionales generalmente se asumen que la función de covarianza es común $c(s, t), t \in T$ para $T = [a, b]$ como el caso homocedástico.

Estadística de prueba

El estadístico de prueba de F puntual para cada punto $t \in T$ está dado por

$$F_n(t) = \frac{SSR_n(t)}{k-1} \bigg/ \frac{SSE_n(t)}{n-k} \quad (11)$$

donde

$$SSR_n(t) = \sum_{i=1}^k [\bar{x}_{i\cdot}(t) - \bar{x}_{\cdot\cdot}(t)]^2 \quad (12)$$

$$SSE_n(t) = \sum_{i=1}^k \sum_{j=1}^{n_i} [x_{ij}(t) - \bar{x}_{i\cdot}(t)]^2 \quad (13)$$

La prueba F puntual se rechaza (7) en cada t siempre que a un nivel de significancia α , donde F_{α} denota el percentil α de la distribución F. Por lo tanto, podemos probar (7) en todos los puntos de t usando el mismo valor crítico para cualquier nivel significativo predeterminado.

Para el caso donde se tenga muestras pequeñas se puede optar por la estadística de prueba de remuestreo "bootstrap".

Valores atípicos funcionales

El análisis estadístico de valores atípicos a menudo puede conducir a resultados inexactos. El ADF no es la excepción. Por lo tanto, se describirá algunas técnicas utilizadas para identificar curvas atípicas y su aplicación en el análisis de estudios de parámetros de calidad del aire.

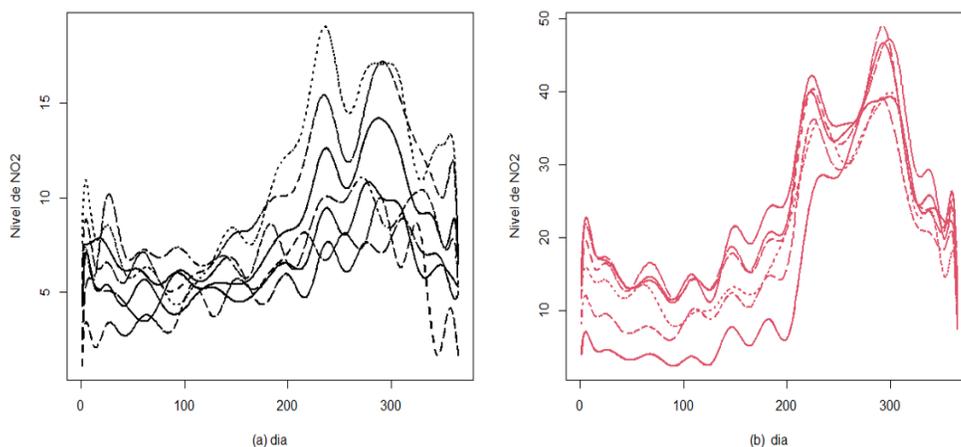
Los métodos gráficos son esenciales para identificar características que no siempre son evidentes a través de modelos matemáticos o resúmenes estadísticos. A pesar de su importancia, la literatura predominante tiende a enfocarse en la modelización, agrupación o predicción de datos funcionales, relegando a los métodos de visualización a un papel secundario, tal como señala [21, 22]. En contextos donde se maneja una amplia gama de curvas, identificar las atípicas se convierte en un desafío, especialmente cuando estas curvas se superponen. En este estudio, se emplean técnicas gráficas como el diagrama arco iris, el diagrama de bolsa funcional y el diagrama de caja de la región de mayor densidad (HDR) funcional, con el objetivo principal de identificar curvas atípicas de particulados PM_{10} .

En este contexto, se han utilizado dos métodos de ordenación basados en un algoritmo que se aplica a las coordenadas bivariadas de las dos primeras componentes principales derivadas de las curvas. Estas coordenadas bivariadas se pueden organizar mediante técnicas como la profundidad de Tukey (1974). A partir de este enfoque, se genera el diagrama de bolsa funcional al aplicar el gráfico bivariado de Rousseeuw et al., 1999 a las dos primeras puntuaciones.

Resultados

Inicialmente, se consideró dos muestras de curvas del dióxido de nitrógeno (NO_2) que es un contaminante producido en el aire principalmente a causa del tráfico automotor y la creación de fuente de energía. Además de dar origen a la lluvia ácida y a otros contaminantes del aire perjudiciales para la salud humana. Para este análisis se consideró las muestras de curvas correspondientes de las ciudades de Valparaíso y la Región Metropolitana de Santiago con el fin de determinar si hay significancia en las curvas medias de NO_2 . Ver la Figura 1.

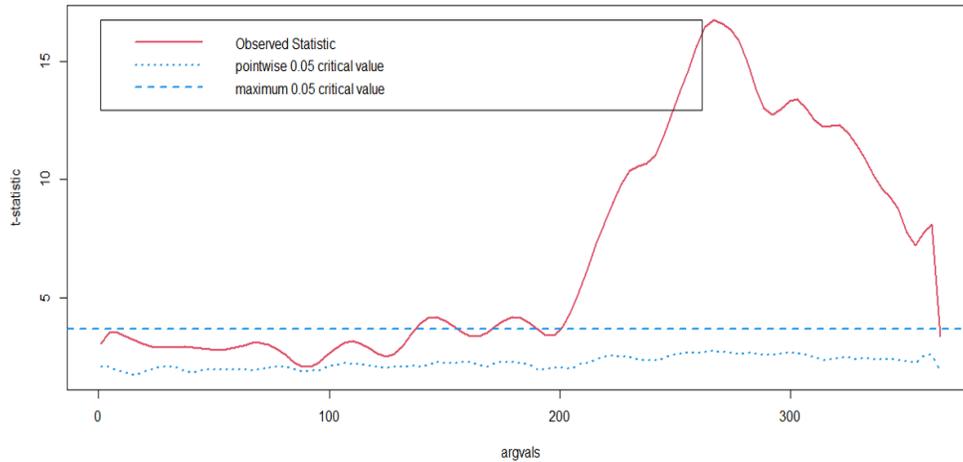
Figura 1. Las curvas de NO_2 de la ciudad de (a) Valparaíso (negro) y de la (b) RMS (rojo)



Fuente: elaboración propia

Se aplico una prueba t puntual Bootstrap para datos funcionales. De la Figura 2, se evidencia que hay diferencias significativas en las curvas medias de NO₂ a un nivel de significancia de 0.05. Para el desarrollo de las pruebas de dos muestras y k muestras funcionales se utilizó el paquete fda del software estadístico R.

Figura 2. La prueba t puntual Bootstrap entre las muestras de Valparaíso y RMS con valor crítico 0.05

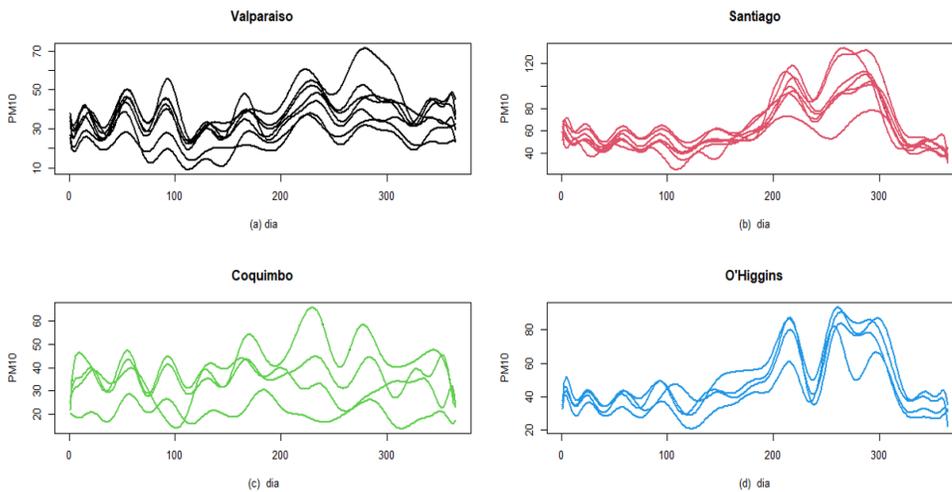


Fuente: elaboración propia

Para los datos PM₁₀ se obtuvo un conjunto de curvas agrupadas por ciudades; Valparaíso, Región Metropolitana de Santiago, Coquimbo y Región Central de O'Higgins ver Figura 3. La idea es verificar la existencia de diferencias entre las curvas medias de las cuatro zonas. Con este fin implementamos una prueba ANOVA de una vía funcional para contrastar las curvas medias.

Para identificar en que estaciones se presentan valores extremos de particulados aplicamos métodos gráficos funcionales expuesto anteriormente.

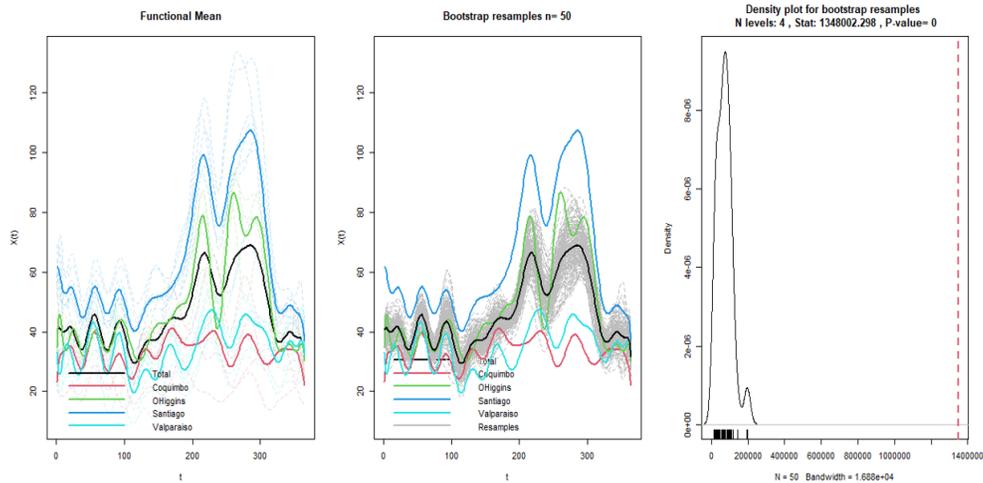
Figura 3. Curvas de PM₁₀ en las ciudades de Valparaíso (línea negra), RMS (línea roja), Coquimbo (línea verde), Región Central de O'Higgins (línea azul)



Fuente: elaboración propia

Se aplico una prueba F puntual de una vía para datos funcionales, que consiste en aplicar en cada punto t del tiempo una prueba F propuesta por [20]. A continuación, se muestra los resultados de la prueba. Ver la Figura 4.

Figura 4. Prueba ANOVA funcional basado en permutaciones con cuatro niveles. La línea media (azul claro) Valparaíso, RMS (línea azul oscuro), Coquimbo (línea roja) Región Central de O'Higgins (línea verde)

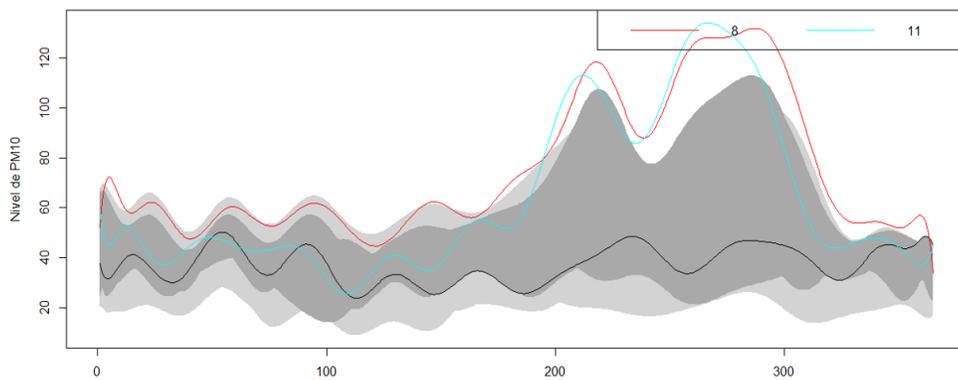


Fuente: elaboración propia

De la Figura 4 se observa que se rechaza la hipótesis de igualdad de curvas medias, lo que indica que hay diferencias significativas (<0.001) en al menos dos curvas medias de PM10. Para detectar que curva se producen mayor particulado PM10 se utiliza métodos gráficos para detectar curvas extremas.

A partir de la Figura 5 se identifican dos curvas atípicas de escala correspondientes a las estaciones 8 y 11. Es así como la curva 8 es representada por la estación El Bosque y la curva 11 representada por la estación Padahuel, las dos perteneciente a la Región Metropolitana de Santiago. Esto evidencia altos niveles de partículas suspendidas de PM10 específicamente en estas dos comunidades.

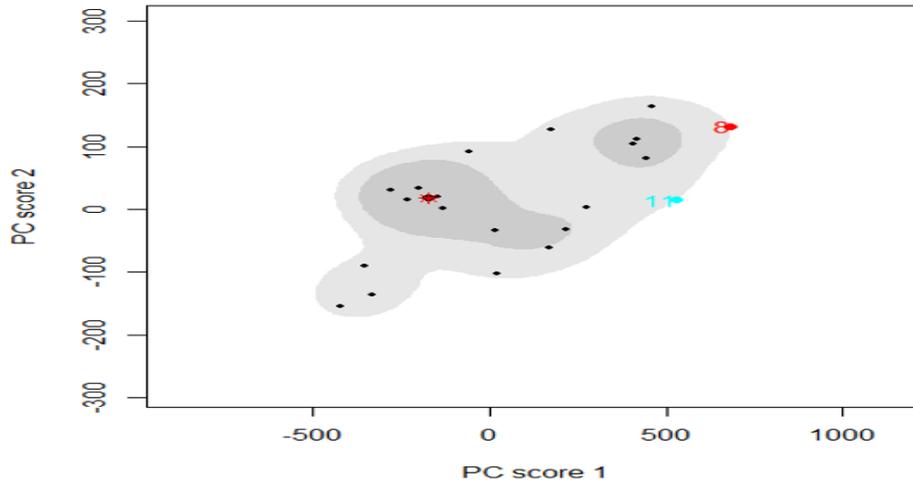
Figura 5. Gráfico HDR identifica las curvas atípicas Estación el Bosque (8) y Estación Padahuel (11) .de particulado PM10 en el Metropolitano de Santiago



Fuente: elaboración propia

En la Figura 5 las regiones de color gris claro y oscuro muestran la región que cubre el 95% de las curvas. Similarmente en la Figura 6 el asterisco rojo es la mediana obtenida usando la profundidad de Tukey. Las curvas de la región exterior se muestran como valores atípicos de diferentes colores, que representan las dos curvas de las estaciones El Bosque (punto rojo) y Padahuel (punto azul).

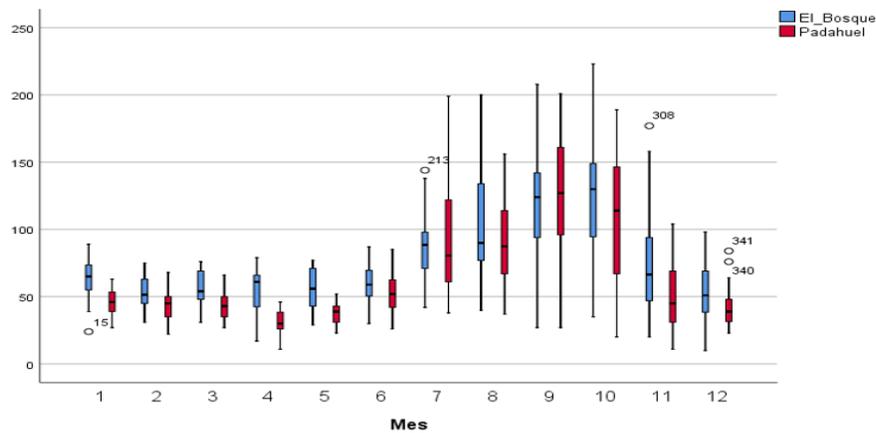
Figura 6. Gráfico de bolsa bivariado con la estación (8) El Bosque (roja) y (11) Padahuel (azul claro) de partículas PM10 de la RMS



Fuente: elaboración propia

La estación El Bosque se evidencia valores extremos para el día número 285 ($223 \text{ ng}/\text{m}^3$), el día 262 ($208 \text{ ng}/\text{m}^3$), el día 218 ($200 \text{ ng}/\text{m}^3$) y el día 215 ($192 \text{ ng}/\text{m}^3$). Similarmente, para la estación Padahuel se presentaron en el día 254 ($201 \text{ ng}/\text{m}^3$), el día 210 ($199 \text{ ng}/\text{m}^3$), el día 287 ($189 \text{ ng}/\text{m}^3$) y el día 300 ($189 \text{ ng}/\text{m}^3$). Esto permite confirmar que los meses de septiembre y octubre y algunos días del mes de noviembre se presenta niveles extremos de PM₁₀.

Figura 7. Diagrama de cajas de PM10 de la estación El Bosque (azul) y Padahuel (rojo) en la RMS.



Fuente: elaboración propia

Conclusiones

Al tratar los valores de las partículas suspendidas PM10 como un objeto funcional, obtenemos una perspectiva integral del problema, considerando también el comportamiento cíclico climático a lo largo del tiempo. Adoptando este enfoque ADF en lugar de recurrir a la estadística descriptiva tradicional, no solo abordamos eficazmente el problema, sino que también facilitamos un análisis espaciotemporal de la dinámica del particulado a lo largo del tiempo.

“En esta investigación, abordamos el desafío de analizar k muestras de curvas utilizando el FANOVA no paramétrico, especialmente cuando el número de muestras de curvas es muy pequeño. Además, hemos implementado una técnica orientada a la detección de curvas atípicas desde una perspectiva funcional, basándonos en los métodos gráficos propuestos por Hyndman [12].

Además, se han registrado niveles extremadamente altos de PM10 durante los meses de octubre y noviembre de 2021 en la Región Metropolitana de Santiago, específicamente en las estaciones El Bosque y Papahuel. Estas áreas deberían ser el foco de atención para las autoridades ambientales, con el objetivo de estudiar y proponer soluciones a esta problemática.

Para investigaciones futuras, resulta imperativo extender el análisis de la varianza funcional a un diseño multidireccional. Además, se sugiere explorar la posibilidad de aplicar la regresión espacial en el contexto de respuestas funcionales, lo cual facilitaría la explicación de la variable dependiente a través de su componente espacial.

Referencias bibliográficas

1. P.J. García-Nieto, Parametric study of selective removal of atmospheric aerosol by coagulation, condensation and gravitational settling, *Int. J. Environ. Health Res*, vol 11, pp. 151–162, 2001.
2. P.J. García-Nieto, Study of the evolution of aerosol emissions from coal-fired power plants due to coagulation, condensation, and gravitational settling and health impact, *J. Environ. Manage*, vol 79, no 4, pp.372–382, 2006.
3. T. Godish, *Air Quality*, Lewis Publishers, Boca Raton, Florida, 2004.
4. L.K. Wang, N.C. Pereira, Y.T. Hung, *Air Pollution Control Engineering*, Humana Press, New York, 2004.
5. M.E Héroux, H.R. Anderson, et al, Quantifying the health impacts of ambient air pollutants: recommendations of a WHO/Europe project. *Int. J. Publ. Health*, vol 60, pp. 619–627, 2015. <https://doi.org/10.1007/s00038-015-0690-y>.
6. K.H Kim, E. Kabir, S. Kabir, A review on the human health impact of airborne particulate matter., *Environ. Int*, vol 74, pp. 136–143, 2015. <https://doi.org/10.1016/j.envint.2014.10.005>.
7. A.J. Badyda, J. Grellier, P. Dąbrowiecki, Ambient PM2.5 exposure and mortality due to lung cancer and cardiopulmonary diseases in polish cities. *Adv. Exp. Med. Biol*, vol 944, pp. 9-17, 2017. https://doi.org/10.1007/5584_2016_55
8. UNION, PEAN, et al. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. *Official Journal of the European Union*, 2008.
9. R. de Medrano, V. de Buen Remiro, J. L. Aznarte, SOCAIRE: Forecasting and monitoring urban air quality in Madrid, *Environmental Modelling & Software*, Volume 143, 2021.
10. J. Martínez, Á. Saavedra, P.J. García-Nieto, J.I. Piñeiro, C. Iglesias, J. Taboada, J. Sancho, J. Pastor, Air quality parameters outliers detection using functional data analysis in the Langreo urban area (Northern Spain), *Applied Mathematics and Computation*, Volume 241, 2014.
11. A. C. Comrie, J. E. Diem, Climatology and forecast modeling of ambient carbon monoxide in Phoenix, Arizona, *Atmospheric Environment*, Volume 33, Issue 30, 1999.

12. J. Sancho, J.J. Pastor, J. Martínez, M.A. García, Evaluation of Harmonic Variability in Electrical Power Systems through Statistical Control of Quality and Functional Data Analysis, *Procedia Engineering*, Volume 63, 2013.
13. R. J. Hyndman y H. L. Shang, Rainbow Plots, Bagplots, and Boxplots for Functional Data, 2008 *Journal of Computational and Graphical Statistics*, vol 19, no 1, 2008. DOI:10.2307/25651298.
14. J.T. Zhang, *Analysis of variance for functional data*, Chapman & Hall, London, 2014.
15. Q. Shen, y J. Faraway, An F Test for linear models with functional responses, *Statistica Sinica*, vol 14, no 4, pp. 1239-1257, 2004.
16. A. Cuevas, M. Febrero, y R. Fraiman, On the use of the bootstrap for estimating functions with functional data, *Computational Statistics and Data Analysis*, In press, 2006.
17. O. A. Vsevolozhskaya, M. C. Greenwood, S. Powell, D. Zaykin, Resampling-based multiple comparison procedure with application to point-wise testing with functional data. *Environ Ecol Stat.* 2015.
18. D.D Cox, J.S. Lee, Pointwise testing with functional data using the Westfall–Young randomization method. *Biometrika*, vol 95, no 3, pp. 621–634, 2008.
19. J.O., Ramsay, B. W. Silverman, *Functional Data Analysis*, Second ed. Springer 2005.
20. T. Gorecki, y L. Smaga, Comparison of tests for the one-way anova problem for functional data, *Comput. Stat*, vol 30, no 4, pp. 987–1010, 2015.
21. A. Maluf Blanco y E. S. Michelena Fernández, “Estudio de factores que afectan la integración de los sistemas de gestión en el Centro Nacional de Biopreparados, Cuba”, *Investigación e Innovación en Ingenierías*, vol. 8, n.º 1, pp. 37–53, ene. 2020. DOI: <https://doi.org/10.17081/invinno.8.1.3592>
22. R. J. Hyndman y M. S. Ullah, “Robust forecasting of mortality and fertility rates: A functional data approach”, *Computational Statistics & Data Analysis*, vol 51, no 10, pp. 4942-4956. 2007. <http://ideas.repec.org/a/eee/csdana/v51y2007i10p4942-4956.html>