

PAPER

Analisis Geo Estadistico

J. Saavedra.*

Estudiante pregrado, Universidad Catolica Del Maule, Talca, Chile

*Corresponding author. Email: javier.saavedra@alu.ucm.cl

Abstract

The Atmospheric Infrared Sounder (AIRS) is a cross-track scanning grating spectrometer mounted on NASA's Aqua satellite. It covers the spectral range from 3.74 μm to 15.4 μm with 2378 channels. AIRS Level 2 products include cloud-cleared infrared radiances and retrieved profiles of atmospheric temperature, water vapor, and ozone at a nominal spatial resolution of 45 km. The retrieval algorithm assumes a linear, time-variable CO₂ climatology to ensure the algorithm's linearity. Although CO₂ profiles are part of Version 5 (V5), the algorithm does not retrieve CO₂. This is left to a post-processing stage using the Vanishing Partial Derivative (VPD) algorithm.

Keywords: Geoestadística, Qgis, Python

En el contexto de la exploración atmosférica y el análisis de datos científicos, este estudio se centra en la importancia de los datos proporcionados por el Espectrómetro de Red de Barrido de AIRS (Atmospheric Infrared Sounder) a bordo del satélite NASA Aqua. A través de este preámbulo, se pretende establecer el marco general de investigación antes de adentrarnos en los detalles.

1. Introduction.

La observación y el estudio de la atmósfera de la Tierra desempeñan un papel fundamental en la comprensión de los procesos climáticos y meteorológicos. En este contexto, el Espectrómetro de Red de Barrido de AIRS (Atmospheric Infrared Sounder) se erige como una herramienta esencial en la obtención de datos atmosféricos cruciales. Montado en el satélite NASA Aqua, AIRS realiza exploraciones espectrales en un amplio rango de longitudes de onda, proporcionando información detallada sobre la temperatura, el vapor de agua, el ozono y otros componentes atmosféricos.

Este instrumento se ha convertido en una piedra angular en la investigación climática y meteorológica, permitiendo un monitoreo preciso de la atmósfera terrestre. En este estudio, exploraremos la importancia de los datos generados por AIRS y su contribución a la comprensión de los fenómenos atmosféricos. Además, se analizará el uso de un algoritmo de recuperación específico en la obtención de datos de CO₂, vital para el seguimiento de cambios en el balance de gases de efecto invernadero en

la atmósfera. Estos datos son esenciales para la investigación y la toma de decisiones relacionadas con el cambio climático y la calidad del aire.

A lo largo de esta investigación, se abordará la relevancia de AIRS en el ámbito científico, su contribución a la modelización climática y su influencia en la toma de decisiones para abordar los desafíos ambientales actuales. Para comprender plenamente el impacto de AIRS, es fundamental explorar tanto sus capacidades como las limitaciones de su algoritmo de recuperación. A medida que avanzamos en esta investigación, profundizaremos en estos aspectos, destacando la importancia de esta herramienta en la exploración y comprensión de la atmósfera terrestre Maddy et al. 2008

2. Marco Teorico(on working...)

2.1 Introducción a AIRS y su Misión Espacial

En esta subsección, se proporciona una breve introducción al instrumento AIRS y su colocación en el satélite Aqua. También se explica la importancia de esta misión espacial para la recopilación de datos atmosféricos.

2.2 Técnicas de Observación por Infrarrojos

Se exploran las técnicas y tecnologías de observación por infrarrojos utilizadas por AIRS. Esto incluye la descripción de la gama espectral cubierta y la resolución espectral.

2.3 Preprocesamiento de Datos de AIRS

Se describen las etapas de preprocesamiento de datos que se aplican a las observaciones de AIRS, como la eliminación de efectos de nubes y la corrección de datos.

2.4 Niveles de Datos y Procesamiento

Se explica el proceso de generación de los diferentes niveles de datos de AIRS, desde las observaciones crudas hasta los productos de nivel 3. También se describe el procesamiento y la reducción de datos.

2.5 Estudios Previos sobre CO₂ y AIRS

Se presenta un resumen de estudios previos que han utilizado datos de AIRS para investigaciones relacionadas con el dióxido de carbono (CO₂) en la atmósfera. Esto puede incluir investigaciones sobre la distribución espacial y temporal del CO₂.

2.6 Función Aleatoria (FA)

En esta subsección, se introduce el concepto de función aleatoria y se explora la diferencia entre variables aleatorias y funciones aleatorias. Se aborda la relación entre la posición y las variables aleatorias espacialmente distribuidas.

2.7 Variogramas y Semivariogramas

Se presenta una explicación detallada de los variogramas y semivariogramas, que son herramientas fundamentales en el análisis de funciones aleatorias espaciales. Se discute cómo se utilizan para caracterizar la dependencia espacial de los datos.

2.8 Función de Distribución de una Función Aleatoria

Se describe la función de distribución de probabilidad de una función aleatoria y cómo se utiliza para caracterizar la distribución espacial de variables aleatorias en un dominio. Se discuten los elementos clave de esta función. Giraldo Henao, año

2.9 Momentos de una Función Aleatoria

Esta subsección se centra en los momentos de segundo orden de una función aleatoria, como la semivarianza, la varianza y la covarianza. Se explica su importancia en la caracterización de la dispersión espacial de los datos.

2.10 Estacionaridad de una Función Aleatoria

Se aborda el concepto de estacionaridad en una función aleatoria, explicando lo que significa que una función aleatoria sea estrictamente estacionaria y cómo esto afecta a los primeros momentos.

2.11 Propiedades de la Función Aleatoria

Esta subsección explora otras propiedades importantes de las funciones aleatorias, como la continuidad espacial y la regularidad.

2.12 Variogramas Experimentales

Se introduce el concepto de variogramas experimentales y cómo se calculan a partir de datos observados. También se discute su utilidad en la estimación de parámetros.

2.13 Modelado de Funciones Aleatorias

Se abordan técnicas de modelado de funciones aleatorias, incluyendo la estimación de parámetros de variogramas y la interpolación espacial.

3. Análisis Exploratorio

La exploración de datos es el primer paso en el viaje hacia la comprensión y el conocimiento. Antes de sumergirnos en el análisis detallado de nuestras variables, es esencial que comprendamos la naturaleza de los datos con los que estamos trabajando. Como ya nos hemos introducido en el contexto de nuestra investigación, es hora de adentrarnos en la descripción de las variables de nivel 3 mensuales. Cada variable es un punto de partida para desvelar patrones, tendencias y secretos ocultos en nuestros datos. Comencemos esta travesía con la latitud, una de las características fundamentales en la exploración de la atmósfera terrestre.

3.1 Descripción de Variables nivel 3 mensuales

- **Latitude (Latitud):**
 - **Tipo de Dato:** Numérico (32-bit FLT)
 - **Dimensiones:** [144, 91]
 - **Descripción:** Esta variable representa la latitud del centro de cada celda o bin en el conjunto de datos. Cada valor indica la latitud en grados del punto central de la celda geoespacial.
- **Longitude (Longitud):**
 - **Tipo de Dato:** Numérico (32-bit FLT)
 - **Dimensiones:** [144, 91]
 - **Descripción:** Esta variable representa la longitud del centro de cada celda o bin en el conjunto de datos. Cada valor indica la longitud en grados del punto central de la celda geoespacial, con indicación de si se encuentra al este o al oeste del meridiano de Greenwich.
- **Mole Fraction of Carbon Dioxide in Free Troposphere (Fracción Molar de Dióxido de Carbono en la Troposfera Libre):**
 - **Tipo de Dato:** Numérico (32-bit FLT)
 - **Dimensiones:** [144, 91]
 - **Descripción:** Esta variable representa la fracción molar de dióxido de carbono (CO₂) en la troposfera libre. La unidad es adimensional (unitless), y los valores indican la concentración de CO₂ en cada celda geoespacial.
- **Mole Fraction of Carbon Dioxide in Free Troposphere Standard Deviation (Desviación Estándar de la Fracción Molar de CO₂ en la Troposfera Libre):**
 - **Tipo de Dato:** Numérico (32-bit FLT)
 - **Dimensiones:** [144, 91]
 - **Descripción:** Esta variable representa la desviación estándar de la fracción molar de dióxido de carbono (CO₂) en la troposfera libre. Al igual que la variable anterior, se mide en unidades adimensionales (unitless) y proporciona información sobre la variabilidad de la concentración de CO₂ en cada celda.
- **Mole Fraction of Carbon Dioxide in Free Troposphere Count (Conteo de la Fracción Molar de CO₂ en la Troposfera Libre):**
 - **Tipo de Dato:** Entero (32-bit INT)

- **Dimensiones:** [144, 91]
- **Descripción:** Esta variable representa el conteo de observaciones o mediciones utilizadas para calcular la fracción molar de dióxido de carbono (CO₂) en la troposfera libre en cada celda geoespacial. Proporciona información sobre la cantidad de datos disponibles para cada celda.

3.2 *Recopilación y Lectura de Datos*

En esta sección, se describe en detalle el proceso de recopilación y lectura de datos relacionados con los productos Level 3. Para acceder a estos datos, se requiere autenticación a través de NASA Earthdata. A continuación, se detallan los pasos seguidos para obtener los datos necesarios:

1. **Autenticación en NASA Earthdata:** Para acceder a los datos, se realizó una autenticación en NASA Earthdata utilizando las credenciales proporcionadas. Esto permitió la descarga de datos desde fuentes autorizadas.
2. **Recopilación de enlaces:** Se recopilaron enlaces a los datos de interés y se almacenaron en un archivo de texto. Cada enlace apuntaba a archivos netCDF que contenían información geoespacial relevante.
3. **Descarga de archivos:** Los enlaces recopilados se utilizaron para descargar archivos de datos en formato netCDF desde las fuentes externas correspondientes. Estos archivos eran esenciales para el análisis exploratorio.
4. **Lectura de archivos netCDF:** Una vez descargados, los archivos netCDF se leyeron y procesaron. Esto implicó la extracción de información relacionada con las variables de interés, como concentraciones de dióxido de carbono (CO₂) y datos geoespaciales.
5. **Cambios en la Dimensionalidad:** Los archivos netCDF originalmente tenían una estructura de dimensiones de [91, 144], lo que representaba una cuadrícula geoespacial con 91 valores de latitud y 144 valores de longitud. Estos datos se transformaron en una estructura que facilitaba el análisis y la creación de identidades de geometría.
6. **Creación de Identidades de Geometría:** Para convertir los datos en un formato geoespacial compatible, se crearon identidades de geometría para cada punto de datos. Cada identidad de geometría representaba una ubicación única con sus coordenadas geoespaciales correspondientes.

3.3 Visualizaciones Básicas.

El proceso de recopilación y lectura de datos es fundamental para la investigación, ya que proporciona los conjuntos de datos necesarios para el análisis exploratorio y los análisis subsiguientes.

A continuación, se presentan los resultados de esta recopilación y lectura de datos, lo que permitirá un análisis más profundo de las concentraciones de CO₂ en la troposfera libre y su distribución geoespacial.

Como observamos en 1 se resalta un claro sesgo positivo y al ser tan drástico el cambio

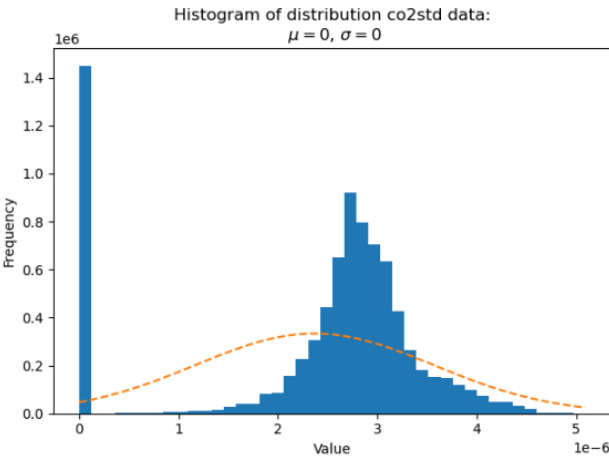


Figure 1. Distribución CO₂_stdv

de variabilidad podemos inferir algún tipo de error en los captura de datos ya que entendemos que son operaciones complicadas y nos adentraremos en entender estos sesgos a lo largo de esta investigación. Otro punto a mencionar es que la cantidad de datos aislados(outlayers), no supera el 10% del total de la data, por lo cual no es descabellado pensar en simplemente omitir esa parte de la data sobre todo cuando dependemos de análisis multidimensionales ya sean visuales o matemáticos con números pequeños como los son estos.

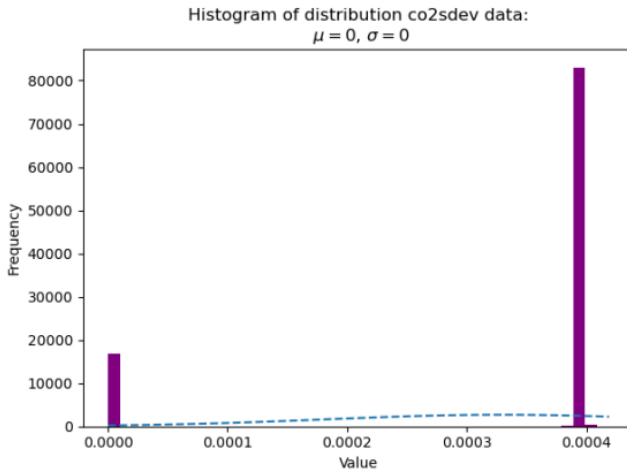


Figure 2. Distribución CO2_sdes

En este Caso, el fenómeno se repite, sin embargo, podríamos encontrarnos con opiniones contrarias frente a esta situación dato que da la impresión de que solo existe 2 valores en tendencia... pero, si pensamos en el gráfico anterior y comparamos la frecuencia podemos notar una correlación por lo que asumimos que se deben al mismo fenómeno el sesgo de los datos con tendencia al 0, por otro lado, si nos centramos con atención en la base de la pila al costado derecho del gráfico, podemos identificar una variabilidad... la cual, si ponderamos por la diferencia entre las frecuencias de ambas pilas se vuelve significativa, es por esto, que tenemos en consideración la eliminación de datos atípicos.

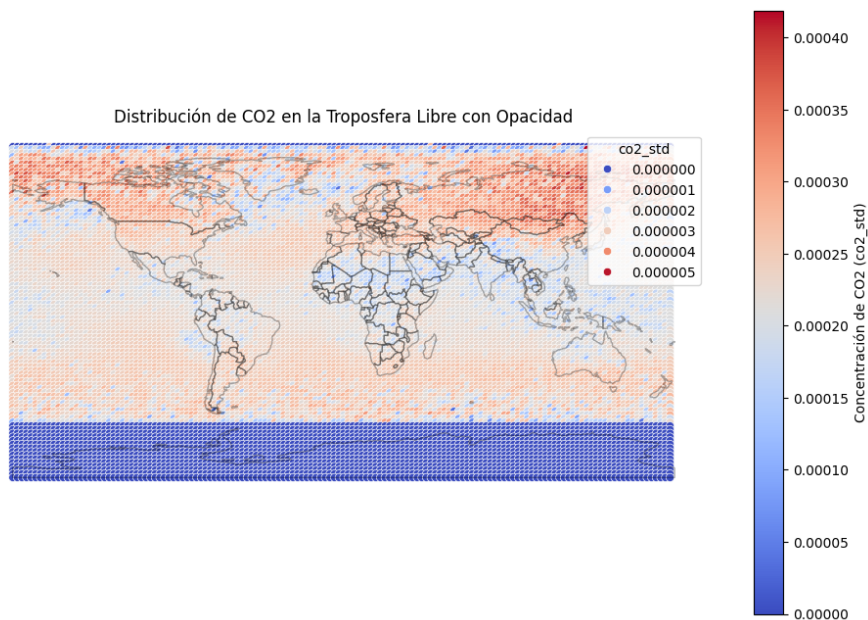


Figure 3. Enter Caption

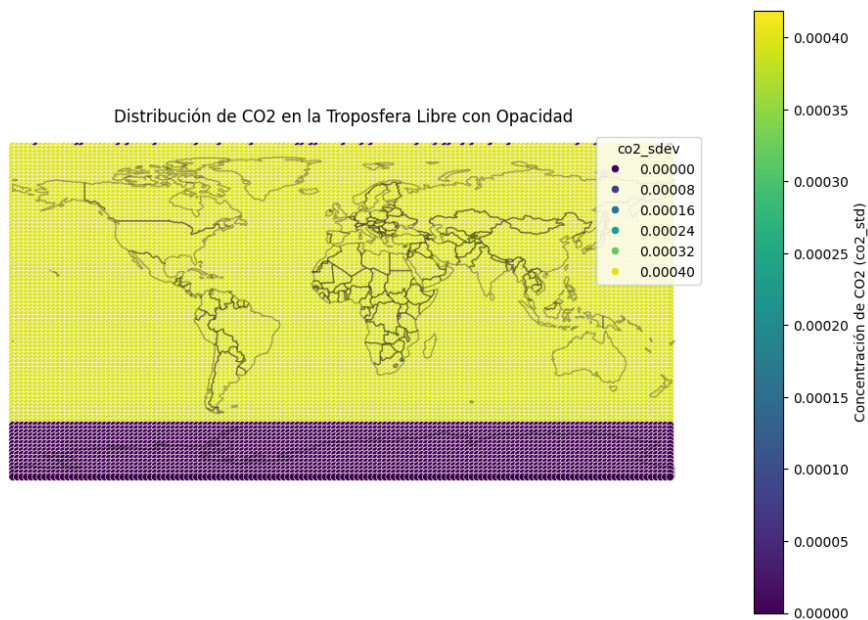


Figure 4. Enter Caption

En estos Mapas se hace notar, y en consecuencia de la naturaleza de la recopilación de datos que mientras mayor sea la latitud en valor absoluto (distancia menor a polos) ocurre una variación en la cobertura dimensional del registro de datos ("teoría de la distancia ortodrómica") lo cual además de ser influenciado en que los datos sean diferentes (democráticamente), impacta en una mayor concentración de datos lo cual genera un peso en esas coordenadas el cual se intento abordar en la reducción de dimensionalidad (código) sin embargo la mejor opción hasta el momento ha sido dejar fuera a los datos, aun estamos expectantes por implementar la matemática aplicada para extraer la mayor información de estos datos... continuara...

3.4 Datos limpiados

En relación a lo mencionado anteriormente, se aplico un filtro a los datos eliminando todas las observaciones con un valor de $co2_std \leq 0.000000001$ de esta forma obtuvimos un total de 10803 observaciones lo cual equivale al 82.44% de las observaciones iniciales.

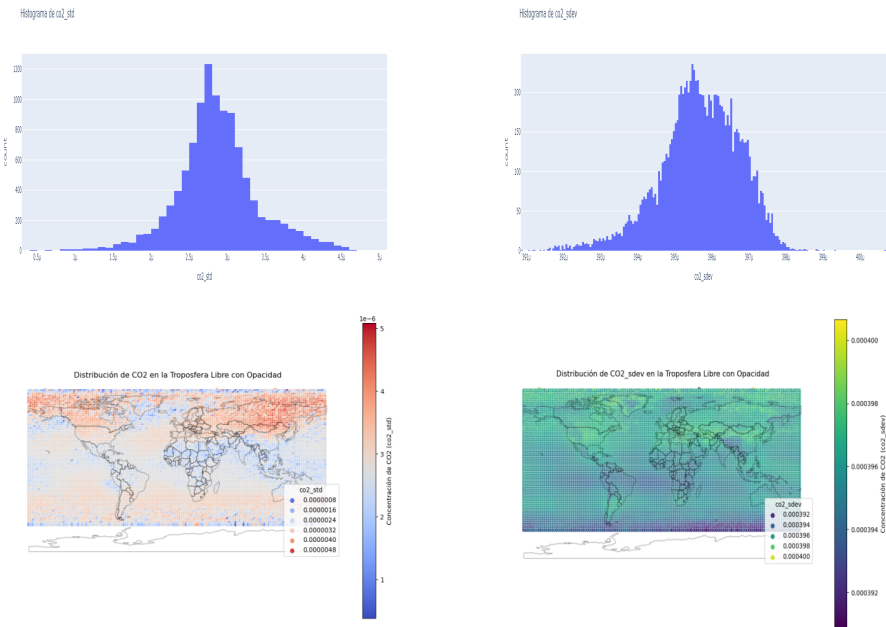


Figure 5. Graficos Exploratorios sin datos atipicos

3.5 Analisis distribucional(normal)

Comprobar si los datos siguen una distribución normal es importante en muchos análisis estadísticos, ya que varios métodos y pruebas estadísticas asumen o funcionan mejor cuando los datos se distribuyen normalmente. Para verificar la normalidad de los datos, se pueden utilizar diferentes métodos y pruebas estadísticas como la prueba de normalidad de Shapiro-Wilk, la prueba de Kolmogorov-Smirnov, histogramas, gráficos de probabilidad normal (QQ-plots) y otras herramientas visuales o estadísticas.

Al asegurarse de que los datos sigan una distribución normal, se puede tener más confianza en la aplicación de ciertos modelos estadísticos y en la interpretación de los resultados derivados de esos modelos.

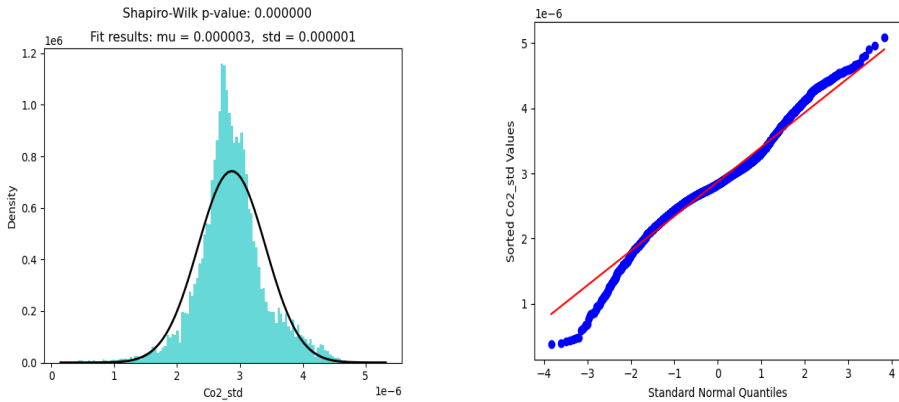


Figure 6. Graficos Analisis normalidad

4. Analisis de estacionariedad

El concepto de estacionariedad es fundamental en el análisis de series temporales y espaciales. Se refiere a la propiedad de una serie de datos donde las características estadísticas, como la media, la varianza y la autocovarianza, permanecen constantes a lo largo del tiempo o el espacio.

En esta sección, exploraremos la noción de estacionariedad en nuestro conjunto de datos. Comenzaremos por examinar la estabilidad de las propiedades estadísticas a lo largo del tiempo o el espacio, lo que nos permitirá comprender mejor la variabilidad y patrones inherentes en nuestros datos. Exploraremos técnicas y herramientas específicas para evaluar la estacionariedad, identificar tendencias y detectar posibles cambios significativos en el comportamiento de nuestros datos.

Al comprender la estacionariedad de nuestros datos, estaremos mejor equipados para realizar pronósticos precisos, modelar adecuadamente los procesos subyacentes y extraer conclusiones significativas de nuestro análisis espacial.

Para esto utilizaremos pruebas de raíz unitaria. Una raíz unitaria es una característica de los procesos que evolucionan a través del tiempo y que puede causar problemas en inferencia estadística en modelos de series de tiempo. Un proceso estocástico lineal tiene una raíz unitaria si el valor de la raíz de la ecuación característica del proceso es igual a 1, por lo tanto tal proceso es no estacionario. Si las demás raíces de la ecuación característica se encuentran dentro del círculo unitario es decir, tienen un valor absoluto menor a uno. entonces la primera diferencia del proceso es estacionaria.

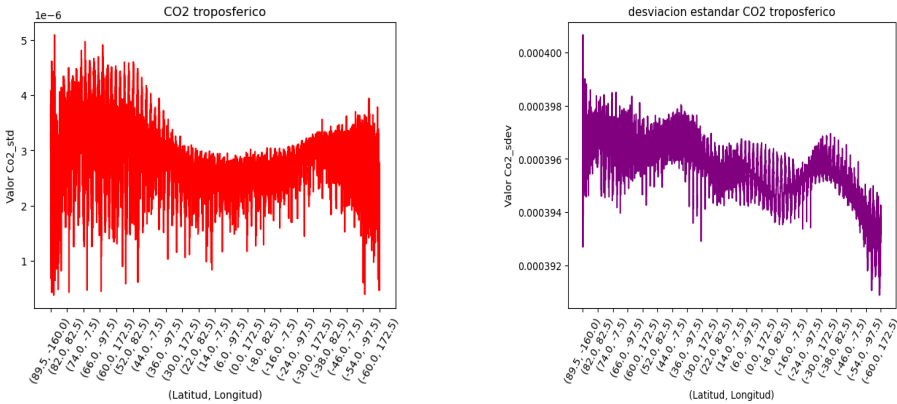


Figure 7. Grafico de series espaciales

4.1 Pruebas de raíz unitaria para *co2_std*

Unitary root test	Test Statistic	P-value	Critical Value	Alternative Hypothesis
Argumented Dickey Fuller	-9.2705	1.3301e-15	-2.8618	The process is weakly stationary.
Phillips Perron	-101.0728	0.0	-2.8618	The process is weakly stationary.
KPSS	3.5542	0.0001	0.4614	The process contains a unit root.
Dickey-Fuller GLS	-8.4868	6.7302e-14	-1.9456	The process is weakly stationary.

En resumen, las pruebas de Dickey-Fuller Aumentada, Phillips-Perron y Dickey-Fuller GLS indican que el proceso es débilmente estacionario, ya que los valores P son muy cercanos a cero, y los estadísticos de prueba están por debajo de los valores críticos.

Por otro lado, la prueba KPSS indica que el proceso contiene una unidad raíz, lo que significa que no es estacionario. Esto crea una discrepancia entre las pruebas, y en este caso, se daría preferencia a las pruebas de Dickey-Fuller Aumentada, Phillips-Perron y Dickey-Fuller GLS, que indican estacionariedad. Sin embargo, es importante considerar la interpretación en el contexto de los datos y los objetivos del análisis. Esta incertidumbre debido a la discrepancia en los resultados, nos motivan a considerar otras técnicas de análisis.

4.2 Pruebas de raíz unitaria para *co2_sdev*

Unitary root test	Test Statistic	P-value	Critical Value	Alternative Hypothesis
Argumented Dickey Fuller	-4.4384	0.0002	-2.8618	The process is weakly stationary.
Phillips Perron	-25.8534	0.0	-2.86181	The process is weakly stationary.
KPSS	12.4155	0.0001	0.4614	The process contains a unit root.
Dickey-Fuller GLS	-0.7686	0.3939	-1.9456	The process is weakly stationary.

En este caso, la Prueba de Dickey-Fuller Aumentada y la Prueba de Phillips-Perron indican que el proceso es débilmente estacionario, ya que los valores P son muy bajos, y los estadísticos de prueba están por debajo de los valores críticos. Estas pruebas respaldan la estacionariedad de los datos.

Por otro lado, la Prueba de Kwiatkowski-Phillips-Schmidt-Shin (KPSS) sugiere que el proceso contiene una unidad raíz, lo que significa que no es estacionario. Esta prueba está en contradicción con las otras dos pruebas.

La Prueba de Dickey-Fuller GLS no es tan concluyente, ya que el valor P es relativamente alto, y el estadístico de prueba no rechaza la hipótesis de no estacionariedad, pero tampoco la confirma con firmeza.

4.3 En contexto

Dado el contexto, es importante considerar que la estacionariedad de estas series temporales puede ser crucial en análisis geoestadísticos relacionados con el CO₂ atmosférico, ya que las tendencias y patrones temporales pueden influir en la interpretación de los datos y en la toma de decisiones.

Las pruebas de raíz unitaria que has realizado indican que la Prueba de Dickey-Fuller Aumentada y la Prueba de Phillips-Perron sugieren que las series de datos "co2_std" y "co2_sdev" son débilmente estacionarias, lo que implica que pueden ser utilizadas en análisis estadísticos con la suposición de estacionariedad.

Por otro lado, la Prueba de Kwiatkowski-Phillips-Schmidt-Shin (KPSS) sugiere que las series de datos contienen una unidad raíz, lo que implicaría no estacionariedad. Esta discrepancia entre las pruebas podría deberse a diferentes supuestos y enfoques de las pruebas.

4.3.1 *En esta situación, consideramos las siguientes opciones.*

Evaluar visualmente los datos: Realiza gráficos de las series temporales para identificar tendencias, estacionalidades u otros patrones visuales que puedan ayudarte a tomar una decisión más informada sobre la estacionariedad.

Realizar más pruebas: Podrías considerar realizar pruebas adicionales o explorar técnicas estadísticas más avanzadas para evaluar la estacionariedad de las series de datos.

Diferenciación: Si no puedes llegar a una conclusión clara sobre la estacionariedad de las series de datos, podrías aplicar técnicas de diferenciación para transformar los datos y hacerlos estacionarios antes de continuar con tu análisis geoestadístico.

Logaritmos: En nuestro caso particular y debido que los valores son pequeños, aplicar logaritmos aún puede ser una estrategia útil para estabilizar la varianza y reducir la magnitud de las fluctuaciones en la serie temporal. Sin embargo, es importante tener en cuenta que, en este caso, el logaritmo natural (\ln) puede ser más apropiado, ya que tiende a funcionar mejor con valores pequeños. Como datos son muy pequeños, es probable que tengan una magnitud similar a la de los valores logarítmicos resultantes. Esto puede facilitar la interpretación de los resultados y hacer que los datos transformados sean más adecuados para análisis geoestadísticos, incluyendo pruebas de estacionariedad.

5. Discusión

La aplicación de la transformación logarítmica se contempla debido a varios factores identificados durante el análisis exploratorio de las variables. Entre las razones clave que respaldan esta decisión se encuentran:

5.1 *Estabilización de la varianza*

Durante la exploración inicial, se observó una alta variabilidad en los datos de la fracción molar de CO_2 en la troposfera libre. La aplicación de logaritmos a estos valores puede ayudar a reducir la magnitud de las fluctuaciones y estabilizar la varianza, lo que facilita el ajuste de modelos estadísticos y reduce la influencia de valores extremos.

5.2 *Interpretación más clara de los resultados*

El logaritmo natural tiende a funcionar bien con valores pequeños, comunes en mediciones de fracciones molares de CO_2 en la atmósfera. La transformación logarítmica podría facilitar la interpretación de los resultados, alineando la escala de los datos con la magnitud de las fluctuaciones.

5.3 Cumplimiento de supuestos estadísticos

En muchos análisis, se asume la normalidad de los datos. La transformación logarítmica podría acercar los datos a una distribución más normal, lo que puede mejorar la aplicabilidad de ciertos modelos estadísticos y pruebas que requieren normalidad en los datos. Genton 1998

5.4 Reducción de sesgo en la inferencia

La transformación logarítmica puede disminuir el sesgo inherente en los datos, lo que permite una mejor inferencia sobre los patrones y cambios en la fracción molar de CO₂ en la troposfera libre.

Es importante resaltar que la decisión de aplicar logaritmos a los datos se toma en función de los objetivos específicos del análisis y la naturaleza de los datos observados. Esta transformación no es una solución universal, pero se considera una estrategia potencialmente beneficiosa para mejorar la idoneidad de los datos para ciertos tipos de análisis estadísticos espaciales, como es nuestro caso donde nuestra variable no posee una unidad de medida.

6. Resultados

Una vez echa la transformación, por fin podemos acercarnos al mundo de la estadística espacial, dado que cumplimos los supuestos necesarios para la aplicación de esta (véase apéndice).

6.1 Lag y distancia óptima

Determinar la distancia óptima de "lag" es importante porque impacta directamente en la estimación de la variabilidad espacial y en la interpolación de valores en ubicaciones no muestreadas. Un "lag" demasiado pequeño puede no capturar la verdadera estructura espacial de los datos, mientras que un "lag" demasiado grande puede perder detalles importantes.

El uso de la función `scattergram()` (véase anexo imágenes scattergram) ayuda a visualizar cómo cambian las diferencias entre los valores de los pares de puntos a medida que aumenta la distancia entre ellos. Esto puede proporcionar pistas sobre la estructura de correlación espacial y ayudar a identificar la distancia óptima de "lag" donde se alcanza un equilibrio entre la autocorrelación y la variabilidad espacial. Matheron 1963

$$\gamma(h) = \frac{1}{2N(h)} * \sum_{i=1}^{N(h)} (Z(x_i) - Z(x_{i+h}))^2 \quad (1)$$

Donde:

- h es el lag
- $N(h)$ es el número de pares ordenados (lat, lon) en h -lag class
- $Z(x_i)$ es el valor de la observación i en la ubicación teórica x

6.2 Semivariogramas

En el contexto del análisis geoestadístico, el semivariograma es una herramienta crucial que muestra la variabilidad espacial de los datos a medida que la distancia entre las observaciones cambia. El "sill" se refiere a un valor constante en el semivariograma que indica la máxima variabilidad espacial entre los datos. Visualmente, en un gráfico de semivariograma, el sill puede representarse como una línea horizontal en el punto donde la variabilidad entre las muestras ya no cambia, lo que indica la máxima distancia a la que las muestras están correlacionadas espacialmente.Mälicke 2021

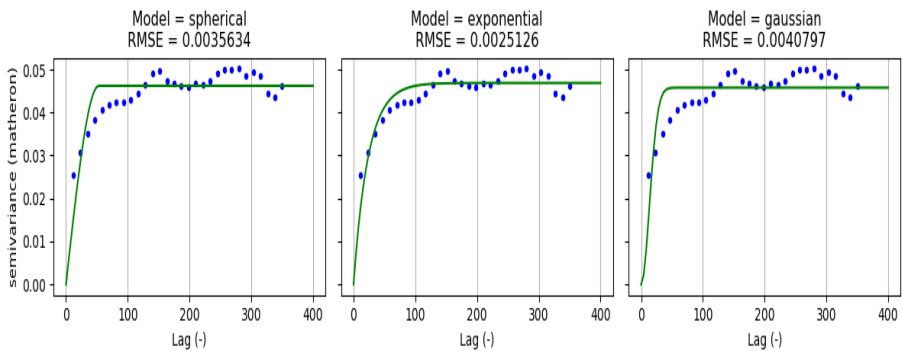


Figure 8. Variogramas

6.2.1 Parámetros

Table 1. Modelos de variograma con sus respectivos rangos, umbrales y valores de pepita.

Modelo	Rango	Umbral	Pepita
Spherical	52.93	0.046	0
Exponential	70.87	0.047	0
Gaussian	37.75	0.046	0

rango: se refiere a la distancia a la cual el variograma alcanza una meseta, lo que indica la distancia máxima a la cual los datos están correlacionados espacialmente en cada modelo.

umbral: es una medida de la variabilidad dentro de los datos. Representa la altura del variograma en su meseta, lo que refleja la variación intrínseca de los datos a esa distancia específica.

pepita: es una medida de la variabilidad espacial que no se puede explicar mediante la distancia; es la varianza intrínseca que se encuentra a distancias muy cortas entre los datos.

6.2.2 Métricas

Los resultados de las métricas de evaluación de los modelos se presentan en la Tabla 2. Estas métricas proporcionan una evaluación cuantitativa del rendimiento de los

Table 2. Métricas de evaluación de modelos con sus respectivos valores de RMSE, MSE, MAE y NRMSE.

Modelo	RMSE	MSE	MAE	NRMSE
Spherical	0.00356	1.27e-05	0.00280	0.07994
Exponential	0.00251	6.31e-06	0.00204	0.05637
Gaussian	0.00408	1.66e-05	0.00322	0.09152

modelos considerados.

El RMSE (Root Mean Square Error): mide la raíz cuadrada del promedio de los errores al cuadrado entre las predicciones y los valores observados. Los modelos Spherical, Exponential y Gaussian tienen RMSE de 0.00356, 0.00251 y 0.00408 respectivamente. Esto indica la discrepancia entre las predicciones del modelo y los valores reales, siendo menor el RMSE para el modelo Exponential, lo que sugiere un mejor ajuste de este modelo a los datos observados.

El MSE (Mean Square Error): es la media de los errores al cuadrado entre las predicciones y los valores observados.

El MAE (Mean Absolute Error): representa el promedio de los valores absolutos de los errores entre las predicciones y los valores observados. Un menor valor de MAE indica un mejor ajuste del modelo a los datos observados, siendo el modelo Exponential el que tiene el MAE más bajo.

El NRMSE (Normalized Root Mean Square Error): normaliza el RMSE por la desviación estándar de los valores observados. Los valores de NRMSE son 0.07994, 0.05637 y 0.09152 para los modelos Spherical, Exponential y Gaussian respectivamente. Un menor NRMSE indica una mejor capacidad predictiva en relación con la variabilidad de los datos, mostrando nuevamente el mejor desempeño del modelo Exponential en este aspecto.

En resumen, basándonos en estas métricas, el modelo Exponential parece presentar el mejor rendimiento en términos de RMSE, MSE, MAE y NRMSE en comparación con los modelos Spherical y Gaussian.

6.3 Kriging

En esta sección se presentan los resultados obtenidos mediante la aplicación del método de Kriging, siendo una técnica versátil en estadística espacial, no solo se emplea para la estimación precisa de valores desconocidos en ubicaciones no muestreadas, sino que también se utiliza en la simulación de campos espaciales y el diseño óptimo de redes de muestreo. Este método se destaca por su capacidad para generar estimaciones interpoladas y mapas continuos que representan la distribución espacial de variables de interés. Los resultados detallados de este proceso de interpolación espacial se presentan a continuación, ofreciendo una visualización exhaustiva de la variación y distribución de los valores estimados en el área de estudio.

Resumen: Basicamente, este metodo consiste en calcular un promedio ponderado como se menciona en 1, esto se puede expresar de forma mas simple

$$\mathbf{Z}(\mathbf{x}_0) = \frac{\lambda}{N(h)} * \sum_{i=1}^{i=N(h)} \mathbf{Z}(\mathbf{x}_i) \quad (2)$$

tal que la suma de todos los lamdas sea igual a uno, cumpliendo asi imparcialidad e insesgades.

Esta es la ecuación de Kriging para Kriging ordinario que se puede encontrar en los libros de texto. Agregamos los unos a la matriz de resultados y a la matriz de semivarianzas. μ es un multiplicador lagrangiano que se utilizará para estimar la varianza de Kriging, que se tratará más adelante. El Kriging ordinario todavía supone que la observación y sus residuos están distribuidos normalmente y son estacionarios de segundo orden. Dowd 1984

$$\begin{pmatrix} \gamma(s_1, s_1) & \gamma(s_1, s_2) & \gamma(s_1, s_3) & \gamma(s_1, s_4) & \gamma(s_1, s_5) & 1 \\ \gamma(s_2, s_1) & \gamma(s_2, s_2) & \gamma(s_2, s_3) & \gamma(s_2, s_4) & \gamma(s_2, s_5) & 1 \\ \gamma(s_3, s_1) & \gamma(s_3, s_2) & \gamma(s_3, s_3) & \gamma(s_3, s_4) & \gamma(s_3, s_5) & 1 \\ \gamma(s_4, s_1) & \gamma(s_4, s_2) & \gamma(s_4, s_3) & \gamma(s_4, s_4) & \gamma(s_4, s_5) & 1 \\ \gamma(s_5, s_1) & \gamma(s_5, s_2) & \gamma(s_5, s_3) & \gamma(s_5, s_4) & \gamma(s_5, s_5) & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix} * \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \mu \end{bmatrix} = \begin{pmatrix} \gamma(s_0, s_1) \\ \gamma(s_0, s_2) \\ \gamma(s_0, s_3) \\ \gamma(s_0, s_4) \\ \gamma(s_0, s_5) \\ 1 \end{pmatrix} \quad (3)$$

6.3.1 Validacion Cruzada "jackknife"

La validación cruzada "jackknife" es una técnica comúnmente utilizada para evaluar la robustez y la precisión de un modelo. En este enfoque, se analiza la estabilidad de un modelo al excluir iterativamente una observación y evaluar su desempeño. Esta técnica proporciona una estimación del error de predicción y permite evaluar cómo el modelo responde a pequeñas modificaciones en los datos de entrada, lo que brinda información valiosa sobre su fiabilidad.

Resultado obtenido(MSE): 0.010141525886376004

6.3.2 Correlacion entre valores y dimensiones

En esta subsección se presentará el gráfico de tendencia de ubicación generado mediante la función `location_trend()`. Este gráfico traza los valores respecto a cada dimensión de las coordenadas en un diagrama de dispersión. Su finalidad es proporcionar una representación visual que permita observar posibles correlaciones entre los valores y las diferentes dimensiones de coordenadas. La detección de una dependencia del valor en relación con la ubicación podría implicar una violación de la hipótesis intrínseca, lo que sugiere una estacionariedad de segundo orden de manera más sutil y precisa.

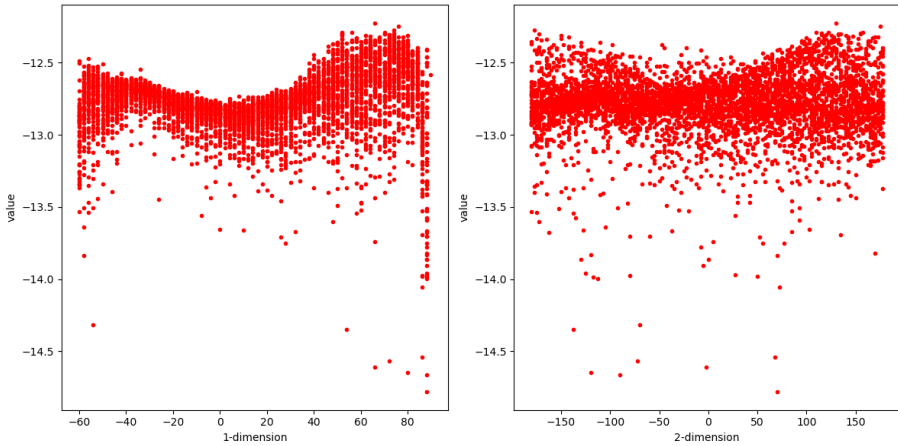


Figure 9. location trends

6.3.3 Gráficos y Comparaciones

En el análisis geoestadístico, la comparación entre el variograma experimental y el modelo teórico ajustado es fundamental para evaluar la idoneidad del modelo en la representación de la estructura espacial de los datos. Los siguientes gráficos muestran esta comparación y permiten entender mejor la discrepancia entre ambos.

6.3.4 Gráfico de Dispersión: Variograma Experimental vs. Modelo Ajustado

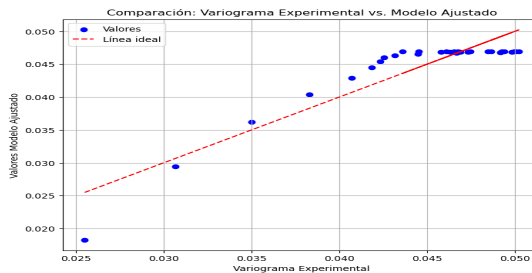


Figure 10. location trends

El primer gráfico muestra una comparación visual entre los valores del variograma experimental y los valores predichos por el modelo ajustado. Cada punto en el gráfico representa un par de valores: el variograma experimental en el eje x y los valores del modelo ajustado en el eje y . La línea roja punteada representa la igualdad perfecta entre ambos valores. Se espera que los puntos azules, que representan los valores, sigan aproximadamente esta línea ideal para un buen ajuste del modelo.

Este gráfico permite una rápida apreciación de la relación entre los valores observados y los predichos por el modelo ajustado. Si los puntos azules se encuentran cercanos a la línea roja punteada, indica una buena correspondencia entre el variograma experimental y el modelo ajustado.

6.3.5 Gráfico de Diferencia: Variograma Experimental - Modelo Ajustado

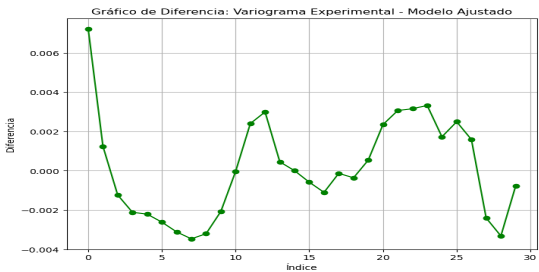


Figure 11. location trends

El segundo gráfico muestra la diferencia entre los valores del variograma experimental y los valores del modelo ajustado. La diferencia se representa en función del índice o posición en los datos. Cada punto en el gráfico indica la discrepancia entre el variograma experimental y el modelo ajustado para un determinado índice. Aquí, una alineación cercana a cero en la línea horizontal implica una concordancia cercana entre los valores.

Este gráfico ayuda a identificar patrones en las discrepancias entre el variograma experimental y el modelo ajustado. Si la mayoría de los puntos están cerca de cero, indica que el modelo ajustado está generalmente en consonancia con el variograma experimental.

Estos gráficos proporcionan una visualización útil para evaluar la calidad del ajuste del modelo respecto al variograma experimental, permitiendo identificar discrepancias significativas o tendencias en los datos analizados.

7. Conclusion

The conclusion text goes here.

Acknowledgement

Insert the Acknowledgment text here.

Funding Statement This research was supported by grants from the <funder-name> <doi> (<award ID>); <funder-name> <doi> (<award ID>).

Competing Interests A statement about any financial, professional, contractual or personal relationships or situations that could be perceived to impact the presentation of the work — or ‘None’ if none exist.

References

- Dowd, P. A. 1984. The variogram and kriging: robust and resistant estimators. In *Geostatistics for natural resources characterization*, edited by G. Verly et al., 91–106. D. Reidel, Dordrecht.
- Genton, M. G. 1998. Highly robust variogram estimation. *Math. Geol.* 30:213–221.
- Giraldo Henao, Ramón. año. *Introducción a la geoestadística: teoría y aplicación*. Sede Bogotá: Universidad Nacional de Colombia.
- Maddy, E. S., C. D. Barnet, M. Goldberg, C. Sweeney, and X. Liu. 2008. Co2 retrievals from the atmospheric infrared sounder: methodology and validation. *Journal of Geophysical Research* 113:D11301. <https://doi.org/10.1029/2007JD009402>.
- Mälicke, M. 2021. Scikit-gstat 1.0: a scipy flavoured geostatistical variogram estimation toolbox written in python. [preprint], *Geosci. Model Dev. Discuss.*, <https://doi.org/10.5194/gmd-2021-174>. <https://doi.org/10.5194/gmd-2021-174>.
- Matheron, G. 1963. Principles of geostatistics. *Economic Geology* 58 (8): 1246–1266. <https://doi.org/10.2113/gsecongeo.58.8.1246>.

Appendix 1. Example Appendix Section

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.