Chilean Journal of Statistics
Vol. xx, No. x, Nov 2023, 1–14

Explorando el Aire que Respiramos: Geoestadísticas con el Espectrómetro AIRS

Javier Saavedra[1,*]

[1]Fc Ciencias Basicas, Universidad Catolica Del Maule, Talca, Chile,
(Received: 00 Month 200x · Accepted in final form: 00 Month 200x)

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: Geoestadistica · Python · Kriging · CO2 · AIRS.

Mathematics Subject Classification: Geostatistics 86A32 · Meteorology 86A10.

In the context of atmospheric exploration and scientific data analysis, this study focuses on the significance of data provided by the Atmospheric Infrared Sounder (AIRS) Grating Spectrometer aboard NASA's Aqua satellite. This preamble aims to establish the overall research framework before delving into the specifics.

1. Introduction

Observation and study of Earth's atmosphere play a pivotal role in understanding climatic and meteorological processes. In this context, the Atmospheric Infrared Sounder (AIRS), mounted on NASA's Aqua satellite, stands as a crucial tool in obtaining crucial atmospheric data. Conducting spectral scans across a broad wavelength range, AIRS furnishes detailed information on temperature, water vapor, ozone, and other atmospheric components.

This instrument has evolved into a cornerstone in climate and weather research, allowing precise monitoring of Earth's atmosphere. This study aims to explore the significance of data generated by AIRS and its contribution to understanding atmospheric phenomena. Furthermore, it will analyze the utilization of a specific retrieval algorithm in obtaining CO2 data, vital for monitoring changes in the greenhouse gas balance in the atmosphere. These data are essential for research and decision-making concerning climate change and air quality.

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

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

Throughout this research, we will address the relevance of AIRS in the scientific domain, its contribution to climate modeling, and its influence on decision-making to tackle current environmental challenges. To fully comprehend AIRS's impact, it is crucial to explore both its capabilities and limitations regarding its retrieval algorithm. As we progress in this investigation, we will delve into these aspects, emphasizing the importance of this tool in exploring and understanding Earth's atmosphere [**?**,]]Maddy etal 2008.

2. Section Two

Section Two of this study delves deeply into the investigation of the Atmospheric Infrared Sounder (AIRS), exploring its capabilities and limitations within the realm of environmental research. The section is structured to cover various crucial aspects of AIRS, beginning with an introduction to this pivotal tool and its vital role in climate modeling and decision-making processes concerning current environmental challenges [Maddy et al., 2008]. This comprehensive section further encompasses an analysis of the AIRS retrieval algorithm, emphasizing its impact and constraints. Through a series of subsections, we examine the functioning of AIRS in collecting essential atmospheric data, including its contributions to climate studies, weather forecasting, and atmospheric dynamics. Additionally, it details the role of AIRS within the Aqua satellite, highlighting its collaborative operation with other instruments. The section also includes insightful discussions on the capabilities and limitations of AIRS in retrieving atmospheric information, laying a solid foundation for a comprehensive understanding of its significance within the realm of Earth's atmosphere.

2.1 Introduction to AIRS

The Atmospheric Infrared Sounder (AIRS) is a vital component of the Earth Observing System (EOS) aboard the Aqua satellite. It plays a crucial role in monitoring Earth's atmospheric conditions and collecting data related to temperature, humidity, and trace gases in the atmosphere.

Importance of AIRS: AIRS is instrumental in providing high-quality infrared soundings, enabling the retrieval of valuable atmospheric information. Its precise observations contribute significantly to weather forecasting, climate research, and the study of atmospheric dynamics.

Data Collection by AIRS: AIRS functions by measuring the infrared energy emitted by the Earth and its atmosphere in a range of wavelengths. These measurements aid in understanding the vertical distribution of temperature and humidity, essential for climate studies and weather prediction models.

Role on Aqua Satellite: Mounted on the Aqua satellite, AIRS operates in conjunction with the Advanced Microwave Sounding Unit (AMSU) and the Humidity Sounder for Brazil (HSB), forming a powerful suite of instruments for comprehensive atmospheric observations.

Image of the Aqua Satellite: Below is an image depicting the Aqua satellite housing the AIRS instrument, showcasing its placement and role within the satellite's suite of scientific instruments.

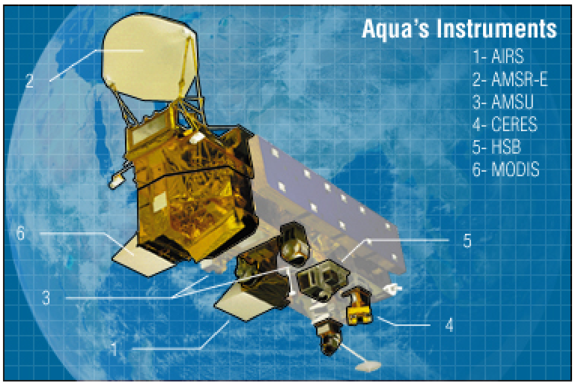

Figure 1. Image of Aqua Satellite with AIRS Instrument

2.2 Preprocessing of AIRS Data

The preprocessing phase of AIRS (Atmospheric Infrared Sounder) data is a crucial step in extracting meaningful insights from the dataset. It involves several preparatory steps to ensure data accuracy, quality, and suitability for further analysis. This section details the various preprocessing procedures applied to the AIRS dataset, encompassing data retrieval, cleaning, transformation, and dimensional reduction techniques.

2.2.1 Description of Monthly Level 3 Variables

In our exploration, we'll begin by understanding the nature of the Level 3 monthly variables:

- **Latitude**:
  - **Data Type**: Numeric (32-bit FLT)
  - **Dimensions**: [144, 91]
  - **Description**: Represents the latitude of the center of each geospatial cell or bin in the dataset. Each value indicates the latitude in degrees of the cell's central point.
- **Longitude**:
  - **Data Type**: Numeric (32-bit FLT)
  - **Dimensions**: [144, 91]
  - **Description**: This variable represents the longitude of the center of each cell or bin in the dataset. Each value indicates the longitude in degrees of the central point of the geospatial cell, indicating whether it is east or west of the Greenwich meridian.
- **Mole Fraction of Carbon Dioxide in Free Troposphere**:
  - **Data Type**: Numeric (32-bit FLT)
  - **Dimensions**: [144, 91]
  - **Description**: This variable represents the mole fraction of carbon dioxide (CO2) in the free troposphere. The unit is dimensionless (unitless), and the values indicate the concentration of CO2 in each geospatial cell.
- **Mole Fraction of Carbon Dioxide in Free Troposphere Standard Deviation**:
  - **Data Type**: Numeric (32-bit FLT)
  - **Dimensions**: [144, 91]
  - **Description**: This variable represents the standard deviation of the mole fraction of carbon dioxide (CO2) in the free troposphere. Similar to the previous variable, it is measured in dimensionless units (unitless) and provides information on the variability of CO2 concentration in each cell.
- **Mole Fraction of Carbon Dioxide in Free Troposphere Count**:
  - **Data Type**: Integer (32-bit INT)
  - **Dimensions**: [144, 91]

- **Description**: This variable represents the count of observations or measurements used to calculate the mole fraction of carbon dioxide (CO_2) in the free troposphere in each geospatial cell. It provides information about the amount of data available for each cell.

2.2.2 Data Collection and Reading

Detailed process of collecting and reading Level 3 data:

(1) **Authentication on NASA Earthdata:** Authentication was performed on NASA Earthdata using provided credentials to access authorized data sources.
(2) **Link Compilation:** Links to the data of interest were gathered and stored in a text file, pointing to netCDF files containing relevant geospatial information.
(3) **File Downloading:** The collected links were used to download netCDF data files from respective external sources, crucial for exploratory analysis.
(4) **Reading netCDF Files:** Once downloaded, the netCDF files were read and processed, involving the extraction of information related to the variables of interest, such as CO_2 concentrations and geospatial data.
(5) **Dimensionality Changes:** The original netCDF files had a [91, 144] dimension structure, representing a geospatial grid with 91 latitude values and 144 longitude values. This data was transformed into a structure facilitating analysis and geometry identities creation.
(6) **Geometry Identities Creation:** To convert the data into a compatible geospatial format, geometry identities were created for each data point. Each geometry identity represented a unique location with its corresponding geospatial coordinates.

2.3 Basic Visualizations

The data collection and reading process play a fundamental role in our research by providing datasets necessary for exploratory analysis and subsequent analyses.

The obtained data allows us to delve deeper into CO2 concentrations in the free troposphere and their geospatial distribution. Here, we present visualizations resulting from this data collection and reading process, enabling a more in-depth analysis.

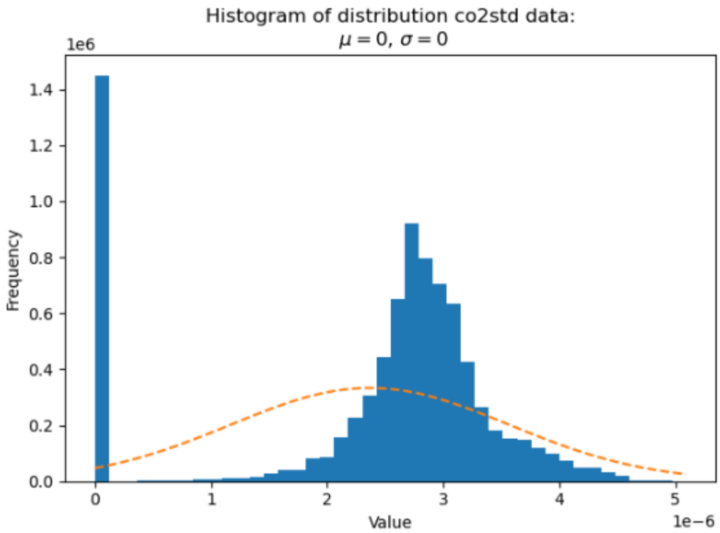

Figure 2. CO2_stdv Distribution

In Figure 2, a noticeable positive bias is highlighted. Such a drastic change in variability leads us to infer potential errors in data capture, considering these are complex operations. Also, note that isolated data points (outliers) do not exceed 10

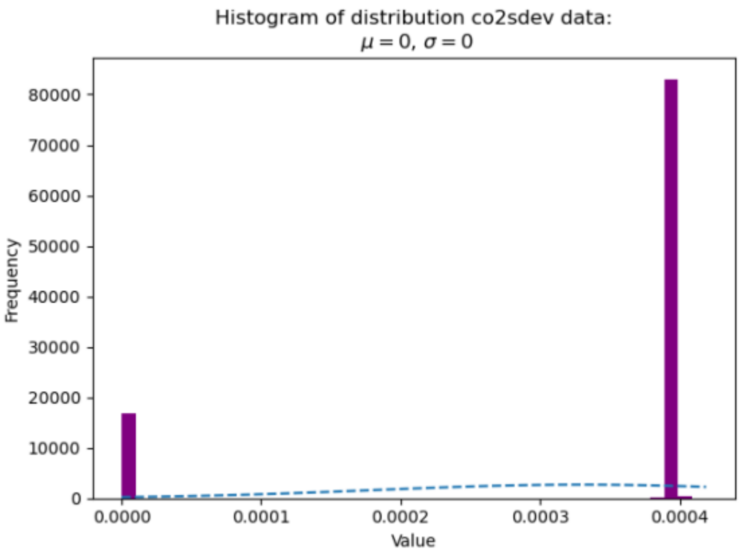

Figure 3. CO2_std Distribution

Similarly, in Figure 3, the phenomenon repeats. While it may appear that only two values are in trend, upon comparison of frequencies with the previous graph, a correlation

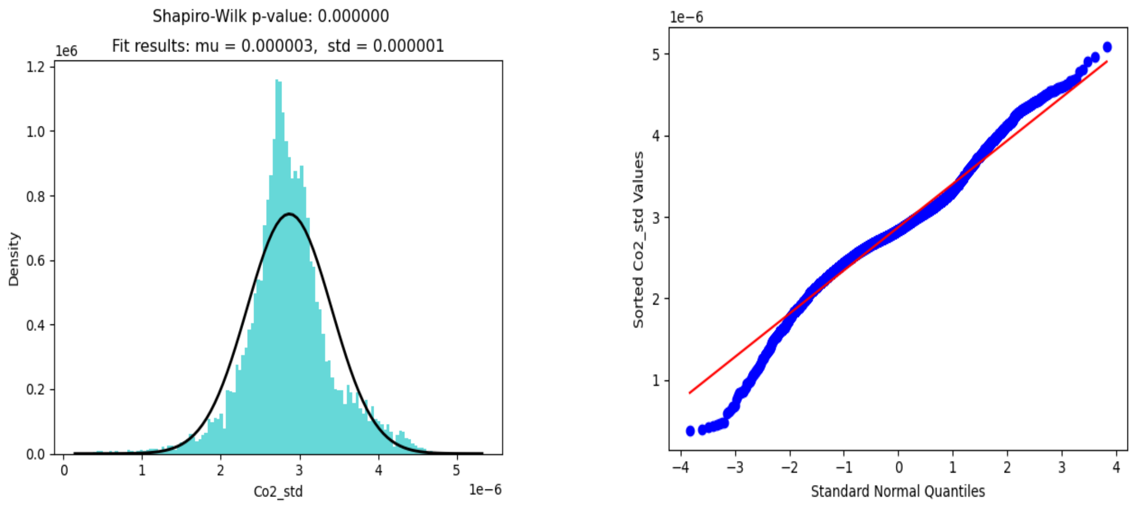

Figure 4. Normality Analysis Plots

is noticeable. Hence, we assume these data biases toward zero are part of the same phenomenon. Furthermore, by focusing on the base of the stack on the right side of the graph, variability is identified. By weighing the difference between the frequencies of both stacks, this becomes significant. Hence, we consider removing outliers.

2.4 Normality Distribution Analysis

Verifying whether the data follows a normal distribution is crucial for many statistical analyses, as several statistical methods and tests assume or perform better when the data are normally distributed. To check the normality of the data, various methods and statistical tests can be employed, such as the Shapiro-Wilk test, Kolmogorov-Smirnov test, histograms, normal probability plots (QQ-plots), and other visual or statistical tools.

Ensuring that the data adheres to a normal distribution allows for increased confidence in applying specific statistical models and interpreting results derived from those models.

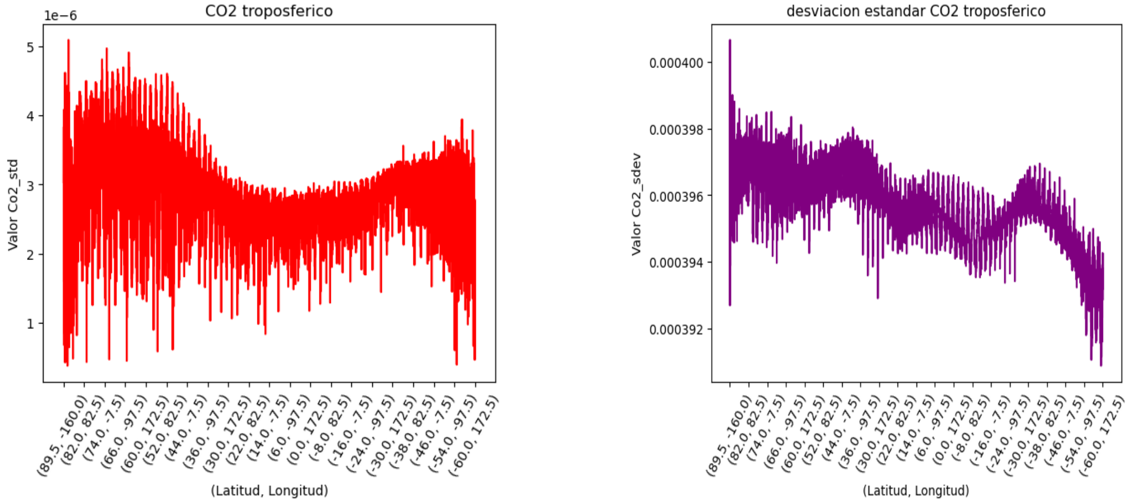

Figure 5. Spatial Series Graphs

2.5 Stationarity Analysis

The concept of stationarity is fundamental in the analysis of time and space series. It refers to the property of a data series where statistical characteristics such as mean, variance, and autocovariance remain constant over time or space.

In this section, we explore the notion of stationarity in our dataset. We start by examining the stability of statistical properties over time or space, which allows us to better understand inherent variability and patterns in our data. We will explore specific techniques and tools to assess stationarity, identify trends, and detect potential significant changes in data behavior.

By understanding the stationarity of our data, we will be better equipped to make accurate forecasts, appropriately model underlying processes, and derive meaningful conclusions from our spatial analysis.

For this purpose, we will use unit root tests. A unit root is a characteristic of processes evolving over time that can cause issues in statistical inference in time series models. A linear stochastic process has a unit root if the root value of the characteristic equation of the process is equal to 1, thus making such a process non-stationary. If the other roots of the characteristic equation lie inside the unit circle, meaning they have an absolute value less than one, then the first difference of the process is stationary.

2.5.1 Unit Root Tests for co2_std

Unitary root test	Test Statistic	P-value	Critical Value	Alternative Hypothesis
A. 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.

In summary, the Augmented Dickey-Fuller, Phillips-Perron, and Dickey-Fuller GLS tests indicate that the process is weakly stationary, as the P-values are very close to zero, and the test statistics are below the critical values.

However, the KPSS test suggests that the process contains a unit root, indicating non-stationarity. This discrepancy between the tests prompts us to consider other analytical techniques.

2.5.2 Unit Root Tests for co2_sdev

Unitary root test	Test Statistic	P-value	Critical Value	Alternative Hypothesis
A. 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.

For co2_sdev, the Augmented Dickey-Fuller and Phillips-Perron tests suggest weak stationarity, while the KPSS test implies non-stationarity. The Dickey-Fuller GLS test results are inconclusive.

2.5.3 In Context

Given this context, it's crucial to consider that the stationarity of these time series might be crucial in geo-statistical analyses related to atmospheric CO_2, as temporal trends and patterns can influence data interpretation and decision-making.

The Augmented Dickey-Fuller and Phillips-Perron tests suggest that the data series "co2_std" and "co2_sdev" are weakly stationary, allowing for statistical analysis under the stationarity assumption.

However, the conflicting results from the KPSS test raise uncertainty regarding stationarity. This discrepancy motivates the exploration of other analysis techniques.

2.5.4 **In this scenario, consider the following options:**

Visual data inspection: Plot time series to identify trends, seasonalities, or other visual patterns to make an informed decision on stationarity.

Conduct further tests: Consider additional tests or explore more advanced statistical techniques to evaluate the stationarity of the data series.

Differentiation: If no clear conclusion is reached on stationarity, consider applying differencing techniques to transform the data into stationary form before proceeding with geo-statistical analysis.

Logarithms: Due to the small values, applying logarithms could still be a useful strategy to stabilize variance and reduce fluctuation magnitude in the time series. Natural logarithm (ln) might be more appropriate here due to its better performance with small values. This transformation may aid in result interpretation and make transformed data more suitable for geo-statistical analyses, including stationarity tests.

2.6 Discussion

The consideration of logarithmic transformation arises due to several factors identified during the exploratory analysis of the variables. Key reasons supporting this decision include:

2.6.1 Variance Stabilization

Initial exploration indicated high variability in atmospheric CO_2 molar fraction data. Applying logarithms may mitigate fluctuations and stabilize variance, aiding in statistical model fitting and reducing the influence of outliers.

2.6.2 Clearer Result Interpretation

Natural logarithms tend to work well with small values, typical in atmospheric CO_2 fraction measurements. This transformation could facilitate result interpretation by aligning data scale with fluctuation magnitude.

2.6.3 Statistical Assumption Fulfillment

Many analyses assume data normality. Logarithmic transformation might approximate data to a more normal distribution, enhancing applicability of specific statistical models and tests relying on data normality.[**?**,]]Genton1998

2.6.4 Reduced Inference Bias

Logarithmic transformation might diminish inherent data bias, allowing for better inference on atmospheric CO_2 molar fraction patterns and changes.

It's important to highlight that the decision to apply logarithms to the data is made based on specific analysis objectives and the nature of the observed data. This transformation isn't a universal solution but is considered a potentially beneficial strategy to improve data suitability for certain types of spatial statistical analyses, particularly when our variable doesn't possess a unit of measurement.

2.7 Geostatistical Analysis

In this section, we delve into the application of geostatistical techniques to derive critical insights from the collected data on CO_2 emissions.

2.7.1 Defining Lag and Optimal Distance

The determination of the optimal "lag" distance is pivotal as it directly impacts spatial variability estimation and value interpolation for unsampled locations. Choosing a "lag" that is too small might overlook the true spatial structure, while an excessively large "lag" could obscure vital details.

Utilizing the scattergram function for visualizing how differences between pairs of points evolve concerning increasing distances. This visualization hints at the spatial correlation structure and aids in identifying the optimal "lag" distance that strikes a balance between autocorrelation and spatial variability [**?**,]]Matheron1963.

THEOREM 2.1 Variogram Equation represent the variogram equation.

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

where:

- h represents the lag,
- $N(h)$ signifies the number of ordered pairs (lat, lon) in the h-lag class, and
- $Z(x_i)$ is the value of observation i at theoretical location x.

2.7.2 Semivariograms

The semivariogram in geostatistical analysis is instrumental in illustrating the spatial variability as observation distances change. The "sill" denotes a constant value in the semivariogram, indicating maximum spatial variability among data points. Visually, the sill, depicted as a horizontal line in the semivariogram plot, represents the point where sample variability ceases to change, signifying the maximum distance at which samples are spatially correlated [**?**,]]Malicke2021.

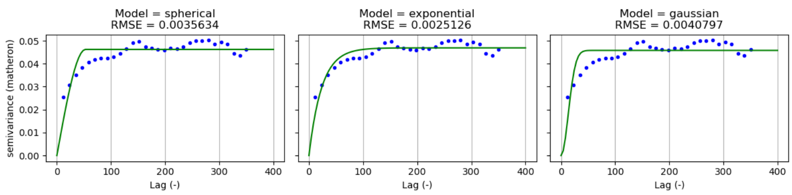

Figure 6. Variograms

2.7.3 Parameters

Model	**Range**	**Threshold**	**Nugget**
Spherical	52.93	0.046	0
Exponential	70.87	0.047	0
Gaussian	37.75	0.046	0

Table 1. Variogram Models with Corresponding Ranges, Thresholds, and Nugget Values.

The **range** signifies the distance where the variogram plateaus, indicating the maximum distance at which data points are spatially correlated in each model.

The **threshold** measures data variability, representing the variogram's height at its plateau, reflecting intrinsic data variation at that specific distance.

The **nugget** is a measure of spatial variability unexplained by distance; it is the intrinsic variance observed at very short distances between data points.

2.7.4 Metrics

Model	**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

Table 2. Model Evaluation Metrics with RMSE, MSE, MAE, and NRMSE Values.

The model evaluation metrics (as shown in Table 2.7.4) offer a quantitative assessment of the considered models' performance.

RMSE (Root Mean Square Error) measures the square root of the average squared errors between predictions and observed values. Lower RMSE, such as the one exhibited by the Exponential model, suggests better fit to observed data.

MSE (Mean Square Error) is the average of squared errors between predictions and observed values.

MAE (Mean Absolute Error) represents the average absolute errors between predictions and observed values. Lower MAE signifies better model fit, with the Exponential model demonstrating the lowest value.

NRMSE (Normalized Root Mean Square Error) normalizes RMSE by the standard deviation of observed values. A lower NRMSE indicates better predictive capacity relative to data variability, favoring the Exponential model's performance in this aspect.

Summing up, based on these metrics, the Exponential model appears to perform the best in terms of RMSE, MSE, MAE, and NRMSE compared to the Spherical and Gaussian models.

2.7.5 Kriging

In this section, we present the results obtained from applying the Kriging method, a versatile technique in spatial statistics. Not only does it estimate unknown values accurately at unsampled locations, but it is also used for spatial field simulation and optimal sampling network design. This method is known for generating interpolated estimations and continuous maps representing the spatial distribution of variables of interest. Detailed results of this spatial interpolation process are presented below, offering a comprehensive visualization of value variation and distribution in the study area.

Summary: This method essentially calculates a weighted average as mentioned in Equation 1, which can be expressed more simply:

Theorem 2.2 Kriging EquationThe Kriging equation is given by:

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

Here, $\mathbf{Z}(\mathbf{x}_0)$ represents the estimation at location $\mathbf{x}_0$, λ is a weight factor, $N(h)$ denotes the number of observations within lag h, and $\mathbf{Z}(\mathbf{x_i})$ denotes the observed values at location $\mathbf{x_i}$.

ensuring the sum of all lambdas equals one, maintaining impartiality and unbiasedness.

This equation represents Kriging for Ordinary Kriging found in textbooks. We added ones to the result matrix and semivariance matrix. μ is a Lagrange multiplier to estimate the Kriging variance, which will be further addressed. Ordinary Kriging still assumes that the observation and its residuals are normally distributed and are second-order stationary [**?**,]]Dowd1984.

2.7.6 Jackknife Cross-validation

Jackknife cross-validation is a commonly used technique to evaluate the robustness and accuracy of a model. In this approach, the model's stability is assessed by iteratively excluding one observation and evaluating its performance. This technique provides an estimation of prediction error and enables assessing how the model responds to small changes in the input data, providing valuable insights into its reliability. **Result obtained (MSE)**: 0.010141525886376004

2.7.7 Correlation between Values and Dimensions

This subsection presents the location trend plot generated through the function 'location_trend()'. This plot showcases the values against each dimension of the coordinates in a scatter plot. Its purpose is to provide a visual representation to observe potential correlations between the values and the different coordinate dimensions. Detecting a dependency of value concerning location might imply a violation of the intrinsic hypothesis, suggesting a second-order stationarity in a more subtle and precise manner.

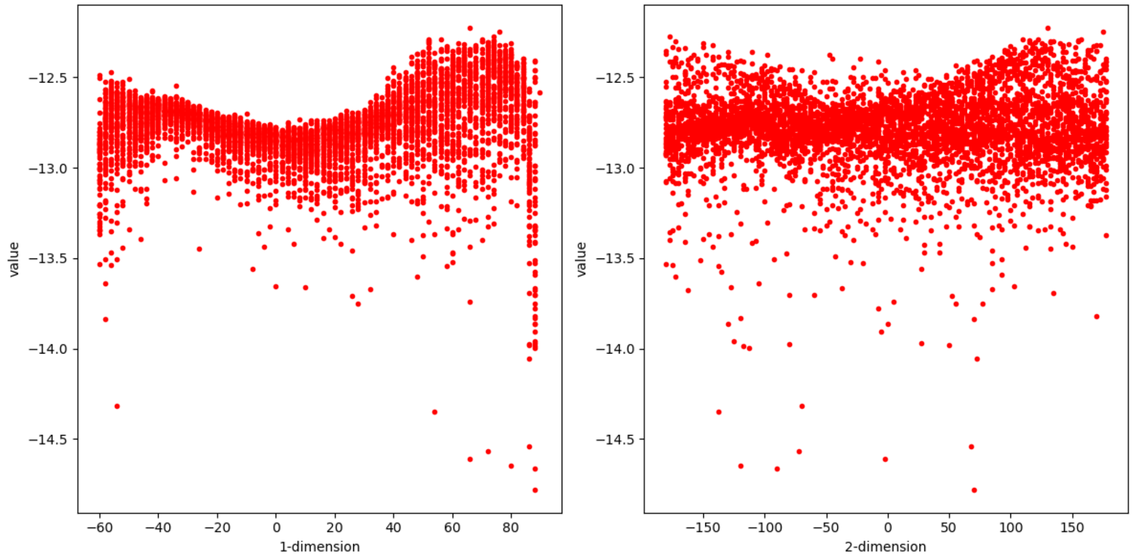

Figure 7. Location Trends

2.7.8 Graphs and Comparisons

In geostatistical analysis, comparing the experimental variogram with the fitted theoretical model is fundamental to assess the adequacy of the model in representing the spatial structure of the data. The following plots showcase this comparison, allowing a better understanding of the discrepancy between them.

2.7.9 Scatter Plot: Experimental Variogram vs. Fitted Model

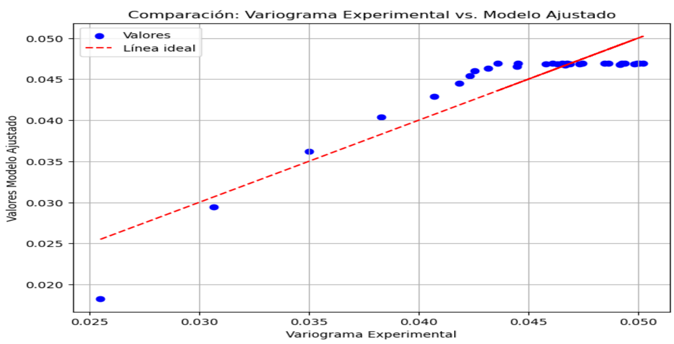

Figure 8. Location Trends

The first graph visually compares the values of the experimental variogram with the values predicted by the fitted model. Each point on the graph represents a pair of values: the experimental variogram on the x-axis and the values from the fitted model on the y-axis. The dashed red line represents perfect equality between both values. Ideally, the blue points representing the values should closely follow this ideal line for a good model fit.

This plot allows for a quick assessment of the relationship between observed values and those predicted by the fitted model. If the blue points are close to the dashed red line, it indicates a good match between the experimental variogram and the fitted model.

2.7.10 Difference Plot: Experimental Variogram - Fitted Model

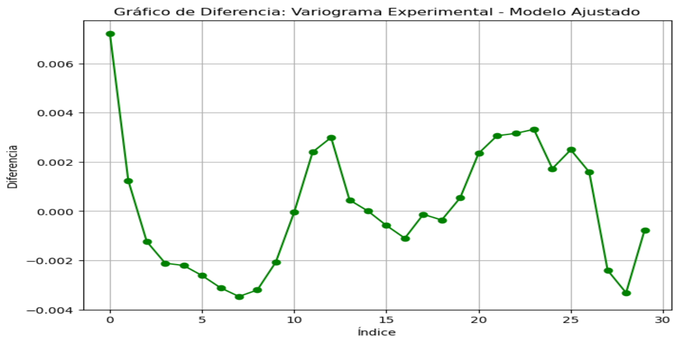

Figure 9. Location Trends

The second graph displays the difference between the values of the experimental variogram and those of the fitted model. The difference is plotted against the index or position in the data. Each point on the graph indicates the discrepancy between the experimental variogram and the fitted model for a specific index. Here, alignment close to zero on the horizontal line implies values' close agreement.

This plot aids in identifying patterns in the discrepancies between the experimental variogram and the fitted model. If most points are close to zero, it suggests that the fitted model generally aligns with the experimental variogram.

These graphs provide a useful visualization to evaluate the quality of the model fit concerning the experimental variogram, enabling the identification of significant discrepancies or trends in the analyzed data.

2.7.11 Conclusions Based on Objectives

The analysis of the spatial distribution of carbon dioxide (CO2) molar fraction in the free troposphere was successfully conducted using data from the Atmospheric Infrared Sounder (AIRS) and advanced Geo-statistical techniques. By fulfilling the general and specific objectives of identifying the characteristics of geo-referenced data and employing statistical techniques for the precise estimation of CO2 concentration, significant strides were made in understanding the spatial distribution of this gas in the atmosphere.

2.7.12 Innovative Ideas and Proposals

This analysis lays a solid foundation for future research and applications of significant relevance. The precise predictions obtained for CO2 concentration in the free troposphere carry important implications. Optimizing parameters, such as the determined range (70 degrees in the best-fitted model), is crucial. These optimized parameters could be employed in creating sampling networks that measure other pollutants with greater accuracy, thus enhancing environmental decision-making and air quality management.

Moreover, there is potential to apply these geo-spatial methods for analyzing other relevant environmental variables. This might encompass the distribution of other greenhouse gases, air quality, or the evolution of climatic phenomena, expanding the impact of these techniques in environmental understanding and management.

2.7.13 Proposed Future Geo-Spatial Studies

Exploring the application of these geo-spatial methods in more detailed studies on CO2 variability across different temporal and spatial scales is suggested. Additionally, developing more complex and multi-dimensional predictive models that incorporate additional atmospheric data could improve estimation accuracy. These future studies may offer significant contributions in addressing global environmental challenges such as climate change and may support adaptation and mitigation strategies at both local and global levels.

References

0

[1] Maddy, E. S., Barnet, C. D., Goldberg, M., Sweeney, C., & Liu, X. (2008). CO2 retrievals from the Atmospheric Infrared Sounder: Methodology and validation. *Journal of Geophysical Research, 113*, D11301. `https://doi.org/10.1029/2007JD009402`

[2] Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

[3] Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

[4] Genton, M. G., 1998, Highly robust variogram estimation, Math. Geol.,30,213-221

[5] Matheron, G. *Principles of geostatistics. Economic Geology*, 58(8), 1246-1266 (1963).

[6] Dowd, P. A. *The variogram and kriging: Robust and resistant estimators.* In *Geostatistics for Natural Resources Characterization*, edited by G. Verly and others, 91-106. Dordrecht: D. Reidel (1984).

[7] M. Mälicke, "SciKit-GStat 1.0: A SciPy flavoured geostatistical variogram estimation toolbox written in Python," *Geosci. Model Dev. Discuss.*, 2021, [preprint], `https://doi.org/10.5194/gmd-2021-174`