Data Assimilation Schemes in Colombian Geodynamics - Cooperative Research Plan for 2017 - 2020 Between Universidad EAFIT and TUDelft, With the Help of Universidad de Antioquia and Universidad Nacional de Colombia Sede Medellin

Inicio: 1 de Enero de 2017
Final: 30 de Diciembre de 2020

Technical Report – Methodology for the analysis of satellite based data

Medellín Air qUality Initiative MAUI

MAUI-RT-01

Entidad Ejecutora	Universidad EAFIT Cra 49 No 7sur - 50 Medellín, Colombia
	Grupo de investigación en modelado matemático – GRIMMAT Grupo reconocido por COLCIENCIAS Categoría A
Responsables	Prof. Olga Lucia Quintero Montoya Prof. Nicolás Pinel Peláez. Investigadores
Entidades Cooperadoras	Department of Applied Mathematics - Tu Delft, Delft The Netherlands TNO
Responsables	Arnold Heemink Arjo Segers

CONTROL DE EDICIÓN Y DISTRIBUCIÓN

Edición	Control*	Nombre y Cargo	Firma	Entidad	Fecha (Día/Mes/Año)
	Creation			Universidad EAFIT	18/04/2017

* Especificar tipo de control: Creación - Revisión - Modificación – Distrib

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

This report presents the proposed methodology for the analysis of satellite-based model results for atmospheric composition. The objective of the method is the characterization of the behavior and development of the variables of interest, and the possible comparison between different data sets. The proposed methods include spectral analysis, temporal scattering, and temporal distribution.

INTRODUCTION

Due to the lack of ground and satellite data on atmospheric composition, it is not possible to determinate the state of meteorological and gas composition at every moment all throughout the globe. Therefore, the use of mathematical models based of the physical evolution of this variables is an interesting alternative to estimate the current state of variables of interest, specially where direct observations cannot be made.

However, this models must me constantly corrected and analyzed, as they try to forecast the evolution of complex dynamical processes using limited information. Methods that allow the evaluation of the temporal and spatial evolution of the variables are need, such that an assessment of the correct behavior of the models. Also, when different datasets are being compared, robust and just comparisons are preferred. This report provides an insight on the first approach to develop such analyses, for the characterization, comparison and analyses of atmosphere composition model data. The proposed methods focus on the identification of tendencies, periodicities, and distributions of the present data.

1. DATABASE DESCRIPTION

In this section, we describe the origin of the current data under analysis. The present database was provided by TNO, and is the result of an atmospheric composition model developed by the Monitoring Atmospheric Composition and Climate (MACC) project (1). This project is the undergoing operational version of the Copernicus Atmosphere Service (CAMS) (2). Copernicus is an European programme for the development of information services based on satellite Earth observations and in-situ data.

The MACC project obtains information directly from satellite and ground based observations, which is then processed by re-analysis and data assimilation to generate state-of-the-art atmospheric models, which are used to generate near-real-time forecast and asses the atmospheric conditions (3). Additionally, the project provides this information freely to the open public, via the Meteorological Archival and Retrieval System (MARS) (4), which is a dedicated repository which can be accessed to download re-analysis results and forecasts.

Currently, the group is working with a dataset that go from March 2015 to April 2016. The MACC model used was the C-IFS-CB05 (5). This model results from an extension of the Integrated Forecasting System (IFS) meteorological model developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), for forecast and atmospheric composition data assimilation. This extension is named Composition-IFS (C-IFS). The calculations are based on a spectral model that apply semi-Lagrangian and semi-implicit methods to solve the dynamical equations (5). The chemical mechanism is based on the Carbon Bond mechanism 5 (CB05), which describes the chemical reactions. As boundary conditions, the MACCity emission inventory was used (1).

The available variable list is:

- Ozone
- Nitrogen Dioxide
- Sulphur Dioxide
- Carbon Monoxide
- Formaldehyde
- Sea Salt Aerosol
- Dust Aerosol
- Hydrophobic Organic Matter Aerosol
- Hydrophilic Organic Matter Aerosol
- Hydrophobic Black Carbon Aerosol
- Hydrophilic Black Carbon Aerosol
- Sulphate Aerosol

The spatial domain under analysis is defined from 15 degrees south to 30 degrees north (latitude), and from 60 to 85 degrees west (longitude). The spatial grid used by the model is a reduced Gaussian N128 grid, which results in 128 latitude values between the pole and the equator, providing and approximate resolution of 70 km. In the vertical direction, the IFS model provides 60 layers using a hybrid-sigma pressure coordinate system, with a range from 0.1 hPa to 1000 hPa approximately (6). However, a selection of 36 layers ranging from 900 hPa to 100 hPa was made for the current database. Samples are provided with a 3-hour sampling period. Figure 1 shows the corresponding grid map over the domain of interest.

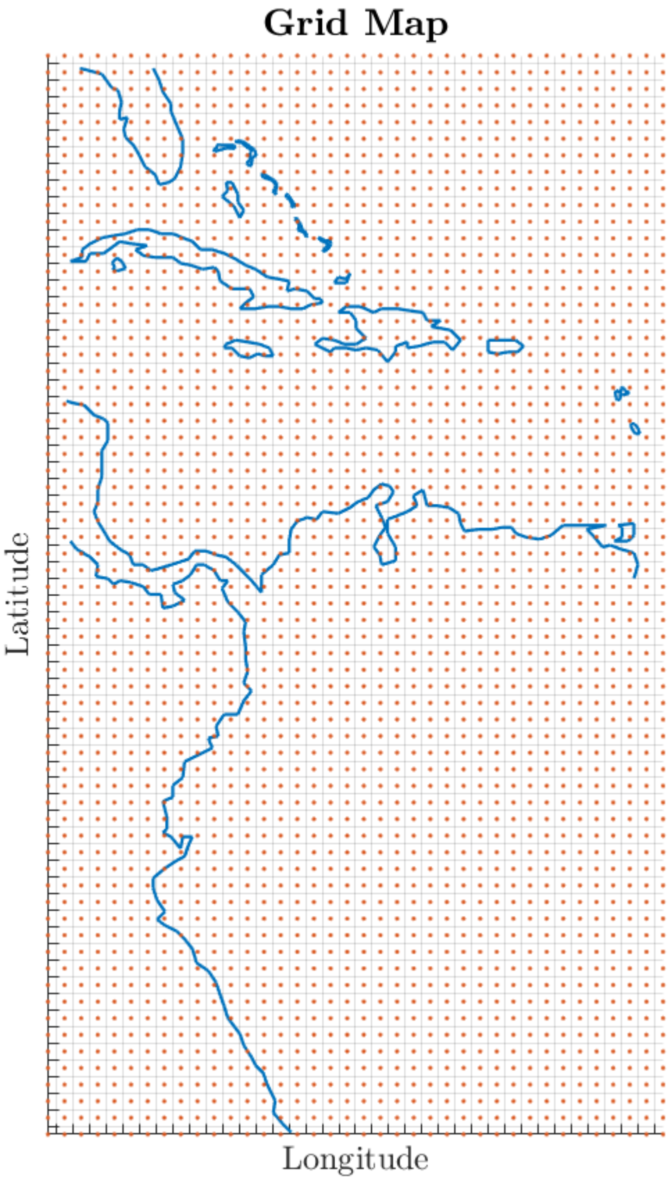

Figure 1: Grid map over the whole domain

The group has special interest in the behavior and development of the variables over the Aburrá Valley. Figure 2 shows the corresponding grid map. It is seen that the model resolution poorly represents the valley's topography, as its area is divided in four grid cells.

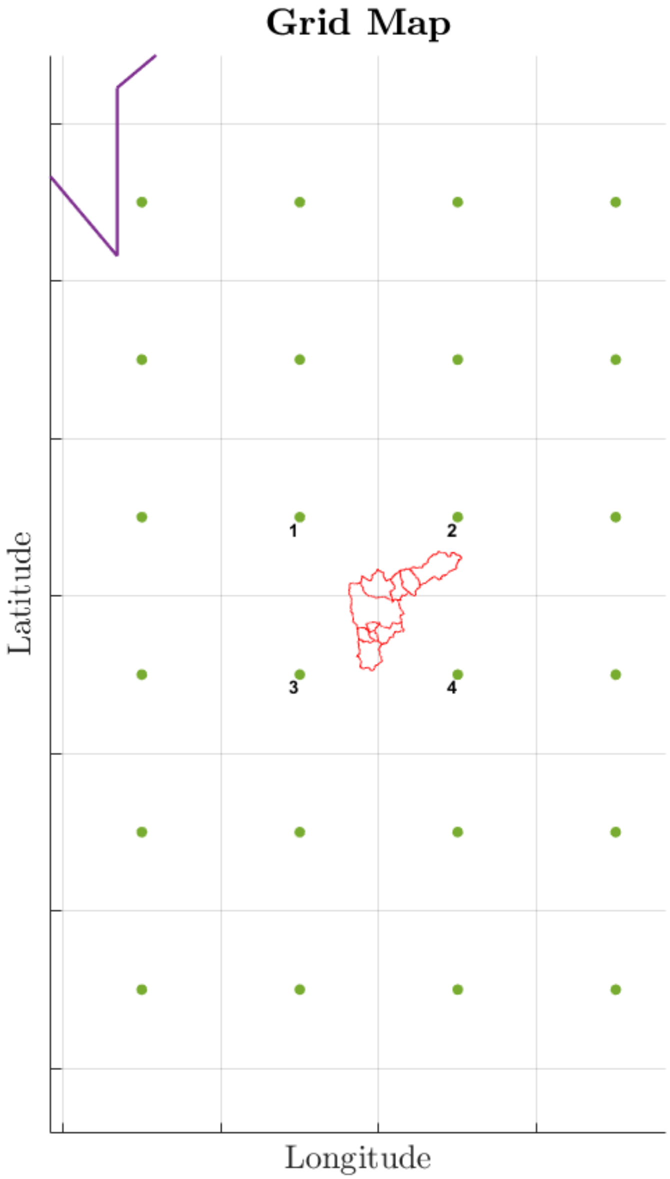

Figure 2: Grid map over the Aburrá Valley

As a reference for the vertical dimension coordinate system, Figure 3 shows the pressure values for the vertical layers over point 1 (see Figure 2).

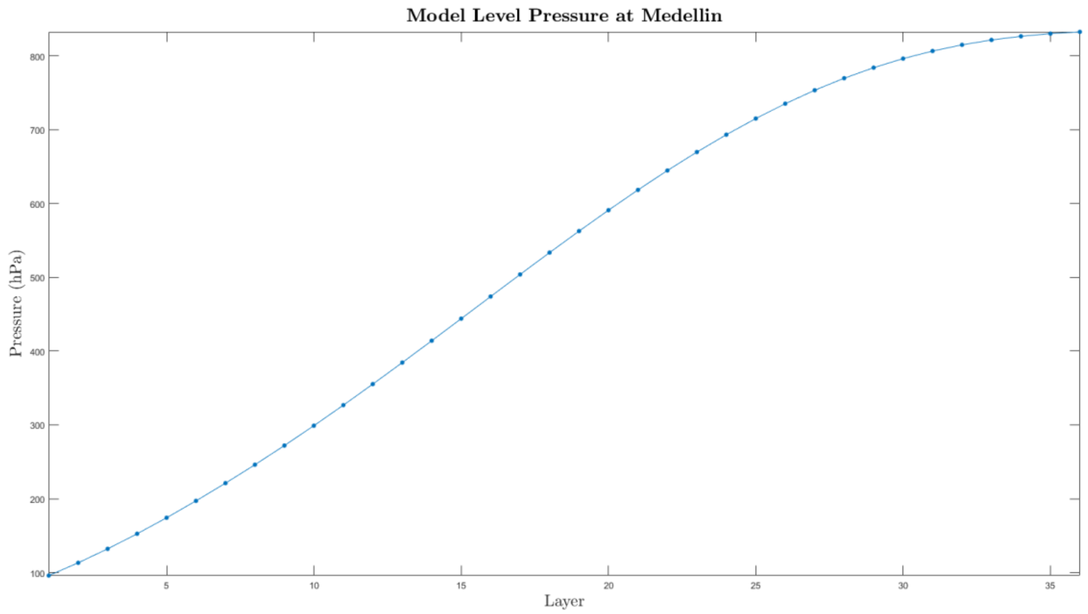

Figure 3: Pressure for the vertical layers over the Aburrá Valley

2. SPECTRAL ANALYSIS

Spectrum calculation is an essential analysis tool used to transform a time signal into the frequency domain, where periodicities and cycles can be evidenced. The N-point discrete Fourier transform is defined as

$$X[k] = \frac{1}{N}\sum_{n=0}^{N-1} x[n]e^{-\frac{j2\pi kn}{N}},$$

where $x[n]$ is the variable of interest time signal, and $X[k]$ is defined as its spectrum. If the number of points N is larger than the number of samples of $x[n]$, then the signal is zero padded. The index k goes from 0 to N-1 and can be related to the corresponding frequency f_k by

$$f_k = \frac{k}{N} f_s$$

where f_s is the sampling frequency. The case $k = 0$ correspond to the mean of the time signal. As the variables under consideration are positive with small variations respect a mean value, this first component is extremely high relative to the rest of the spectrum, and is dominant in a graphical representation. Therefore, the signals are mean-subtracted to avoid this situation. Figures 4 and 5 show the plots for the spectra of ozone at two pressure levels as an example.

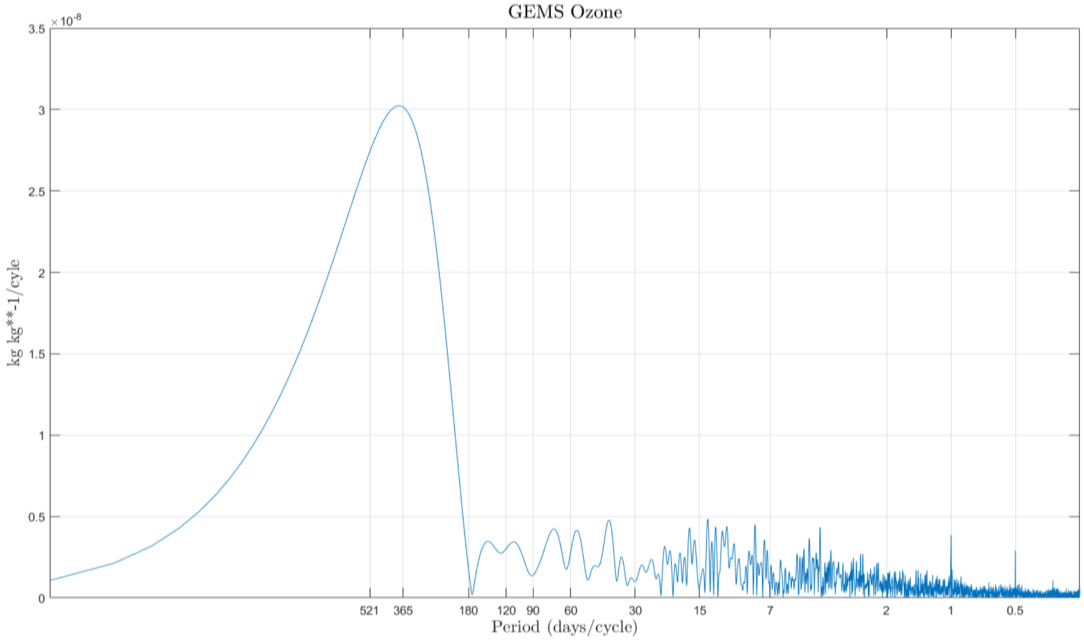

Figure 4: Ozone spectrum at a pressure of 96 hPa

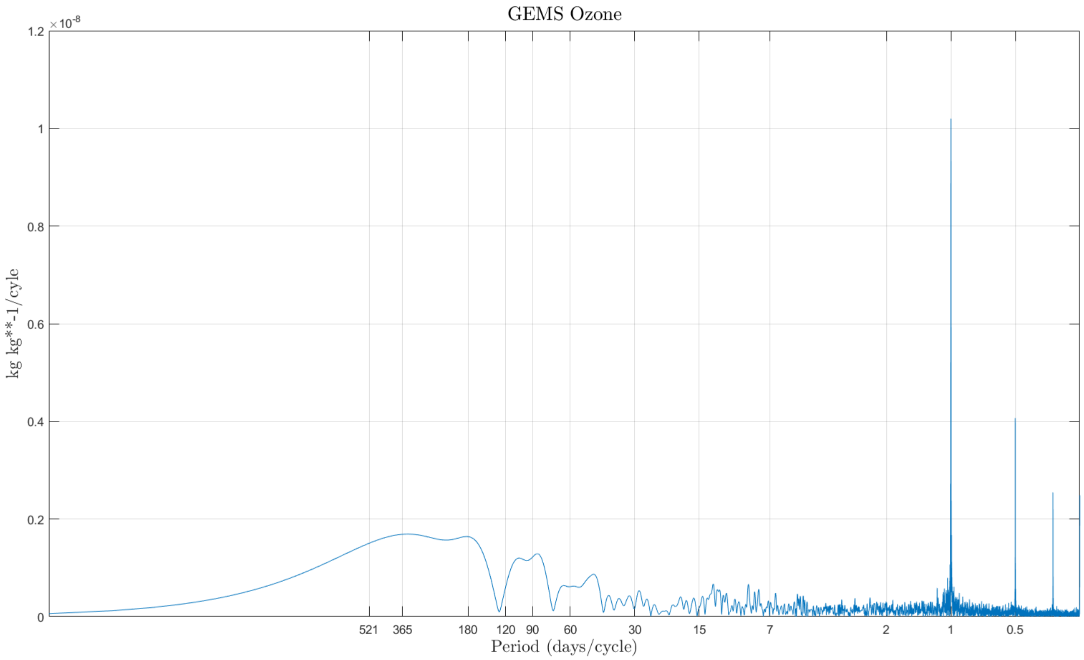

Figure 5: Ozone spectrum at a pressure of 914 hPa

It can be seen that the vertical position is a critical factor in order to characterize the behavior of this substance. At higher levels, lower frequencies are predominant (annual cycle), whilst at lower levels the daily cycle has the highest amplitude. These periodicities are related to physical and chemical processes, and their study is of great importance for the description and modelling of the substances under study.

To evidence more clearly the dependence of the spectrum on the vertical dimension, the following method is proposed: the spectrum for each pressure level is calculated independently and normalized with respect to its maximum value. Afterwards, the resulting spectra are concatenated in a single matrix, which show the change of the most relevant frequencies components as the height varies. Figure 6 shows a visual example of the resulting concatenated spectra for ozone.

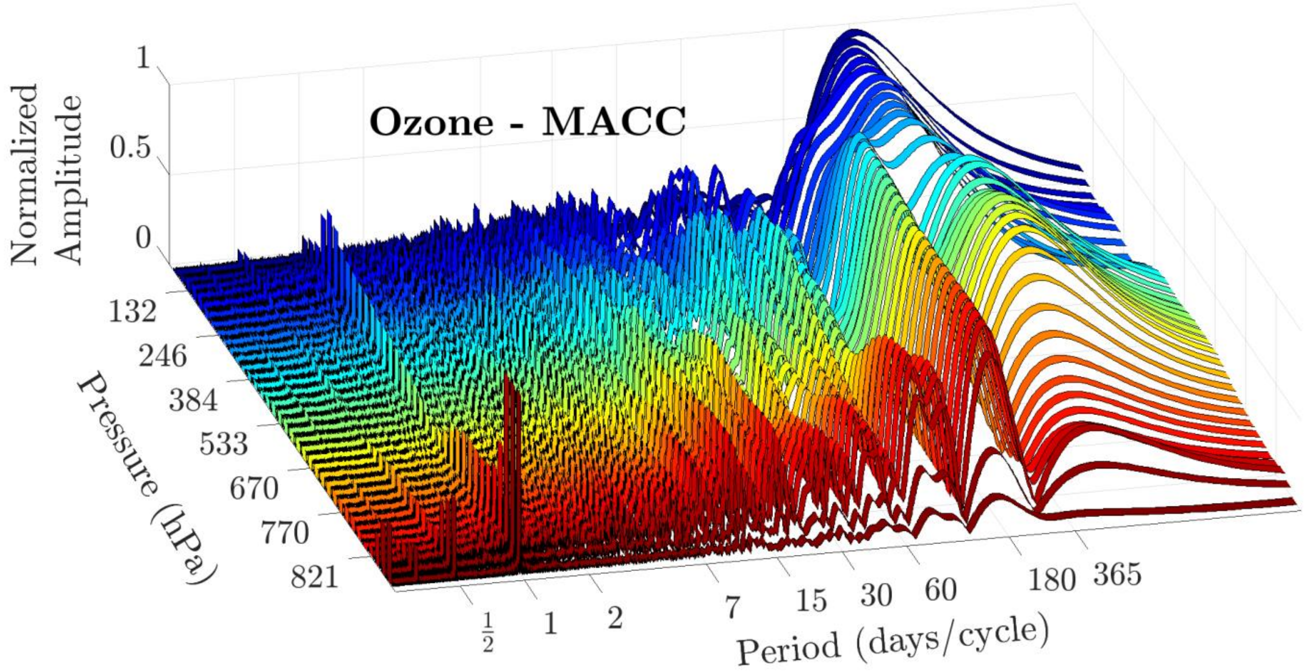

Figure 6: Concatenated spectra for ozone at different pressure levels.

3. DAILY CYCLE ANALYSIS

The daily cycle, variation of the substance concentration throughout the day, is of vital importance as it allows to detect critical moments, and associate this moments to possible causes, man-made or not. To provide a qualitative analysis, a scatter plot is proposed, where each point represents the concentration for a certain substance at a specific hour of day. Figure 7 shows an example of this analysis for the concentration of ozone given by the C-IFS model.

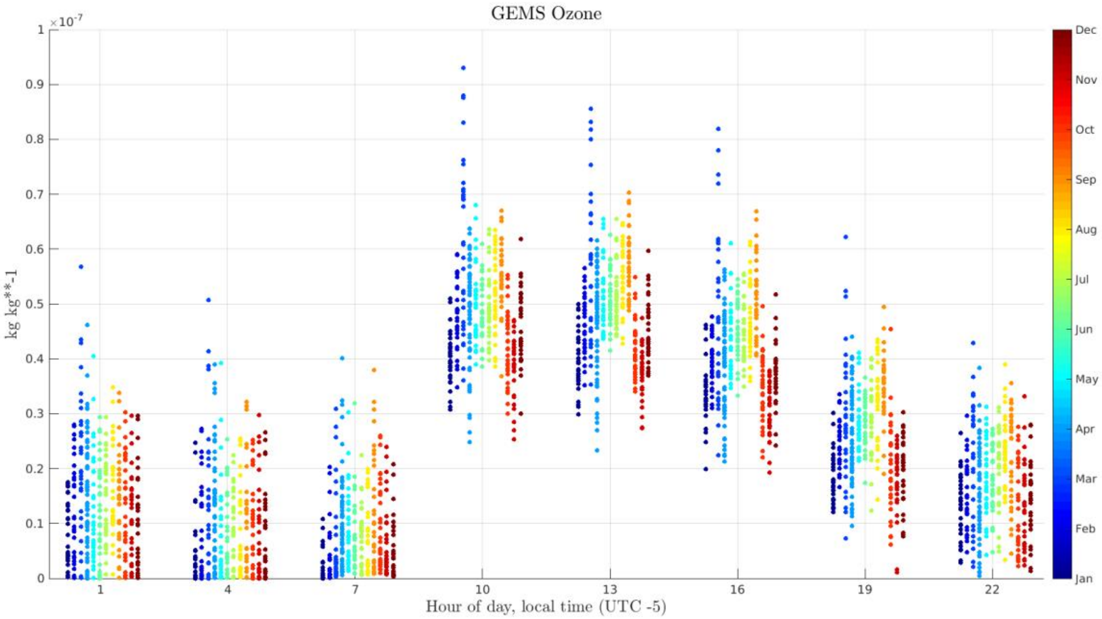

Figure 7: Ozone scatter plot according to the hour of day

In addition to the information of the time of the day, the color is associated to the corresponding month. This allows the identification of variations of the daily cycle associated to the yearly cycle. For example, it can be seen that for ozone, the highest concentrations are found between 10 am and 4 pm, with extreme peaks during march. Figure 8 shows another example, where the effect of the time of the year is drastic.

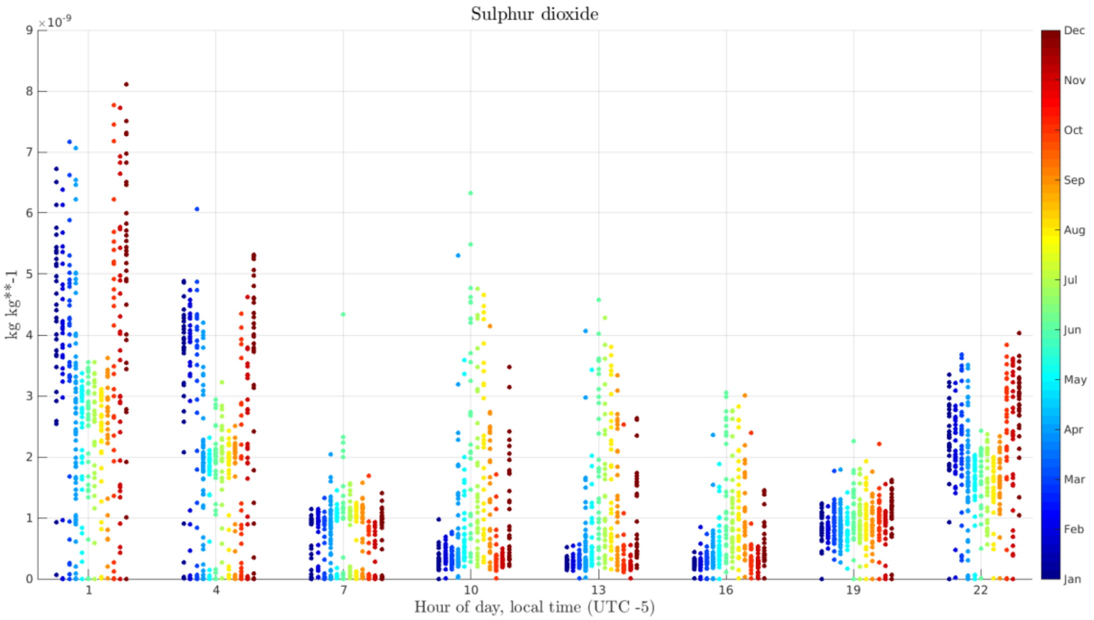

Figure 8: SO2 scatter plot according to the hour of day

In this figure, the highest concentrations of Sulphur dioxide during the rainy season (jan-feb-mar and oct-nov-dec) are found during nighttime, whereas during the dry season (jun-jul-aug), the highest values are found during daytime (10 am to 4 pm).

4. TEMPORAL DISTRIBUTION ANALYSIS

As discussed in the previous section, the concentration levels of the different substances change depending on the time of day and the time of year. Also, as the development of the variables under study is generally unknown, it can be understood as a stochastic process. Thus, the analysis of the distribution is required. Additionally, it is inferred that these are not stationary processes, so the variation of the distributions versus time must be taken in consideration. To evaluate these aspects, the following analysis is proposed:

- Define a time window of arbitrary duration, which is used to segment the variable of interest.
- Calculate the histogram of the windowed data.
- Move the window through the time domain, obtaining the empiric distributions for each position.
- Concatenate the histograms to obtain the temporal distribution matrix (moving histogram).

Figure 9 shows an example of the resulting moving histogram matrix for the case of ozone. In the figure, the x-axis shows the starting point of the time window (35 day duration), y-axis shows the concentration bins nominal value, and color is associated with the relative frequency for each bin.

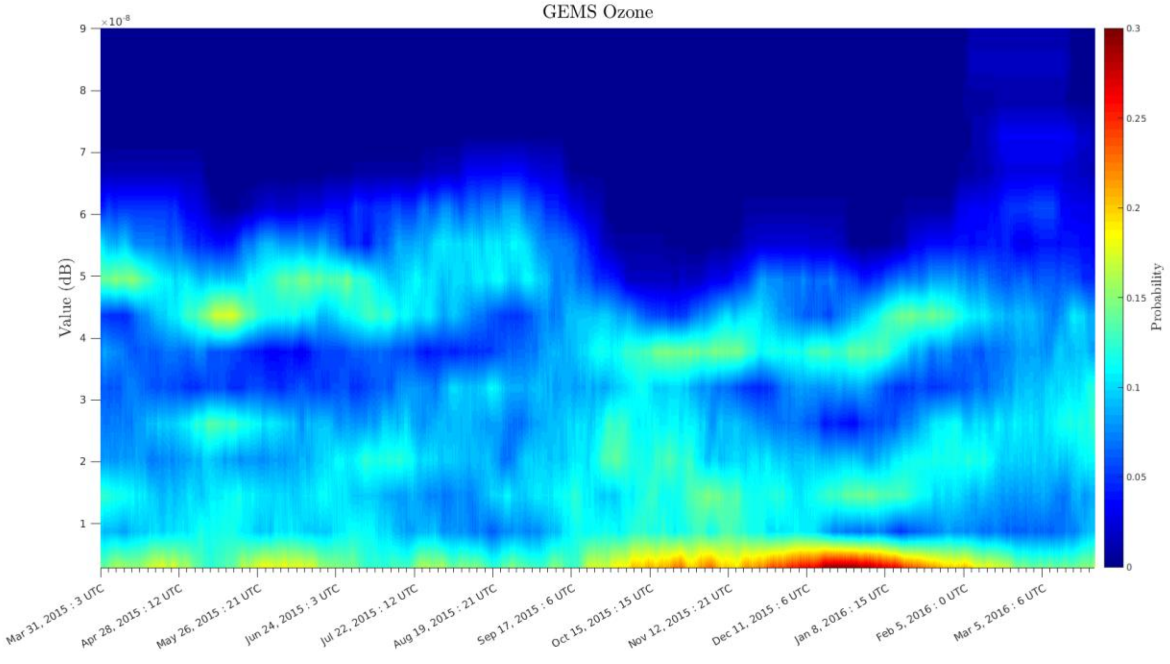

Figure 9: Moving histogram plot for ozone

This analysis can be extended to more specific cases, for instance, an arbitrary set of hours can be selected to calculate the moving histogram. Figure 10 shows the moving histogram for ozone between 10 am and 4 pm, whereas Figure 11 contains the interval between 7 pm and 7 am.

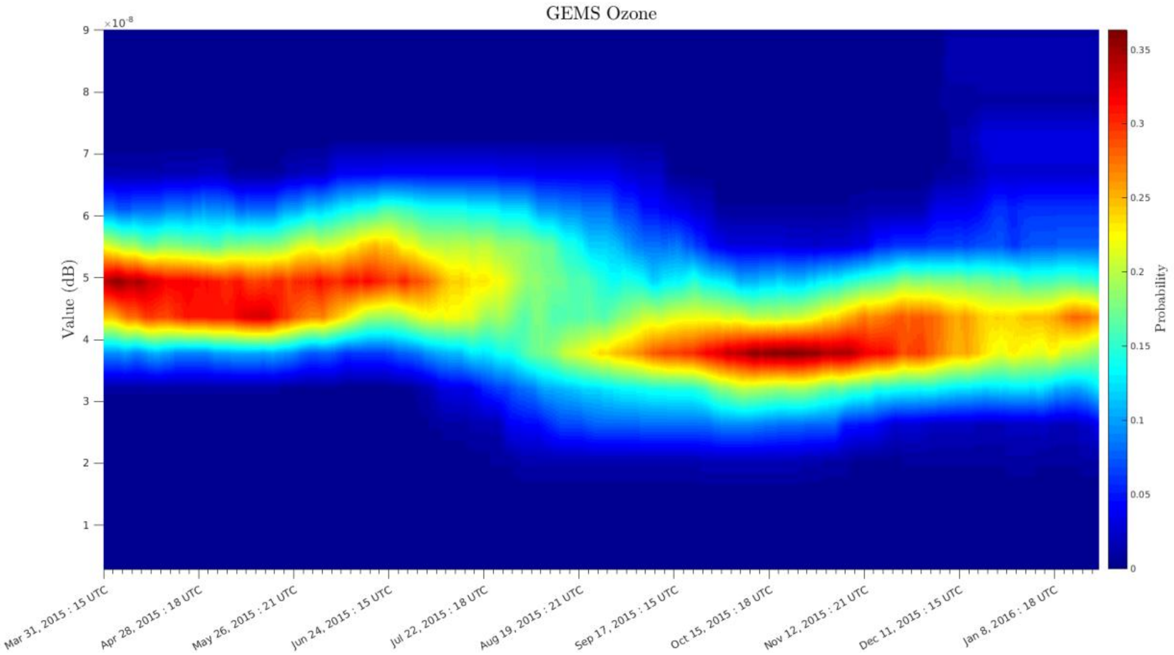

Figure 10: Moving histogram for ozone (day)

It is evidenced that the bimodal behavior present in Figure 9 can be associated to the distributions induced by the daily cycle. This analysis also allows to identify non-stationary situations and identify critical moments associated to a specific time of year.

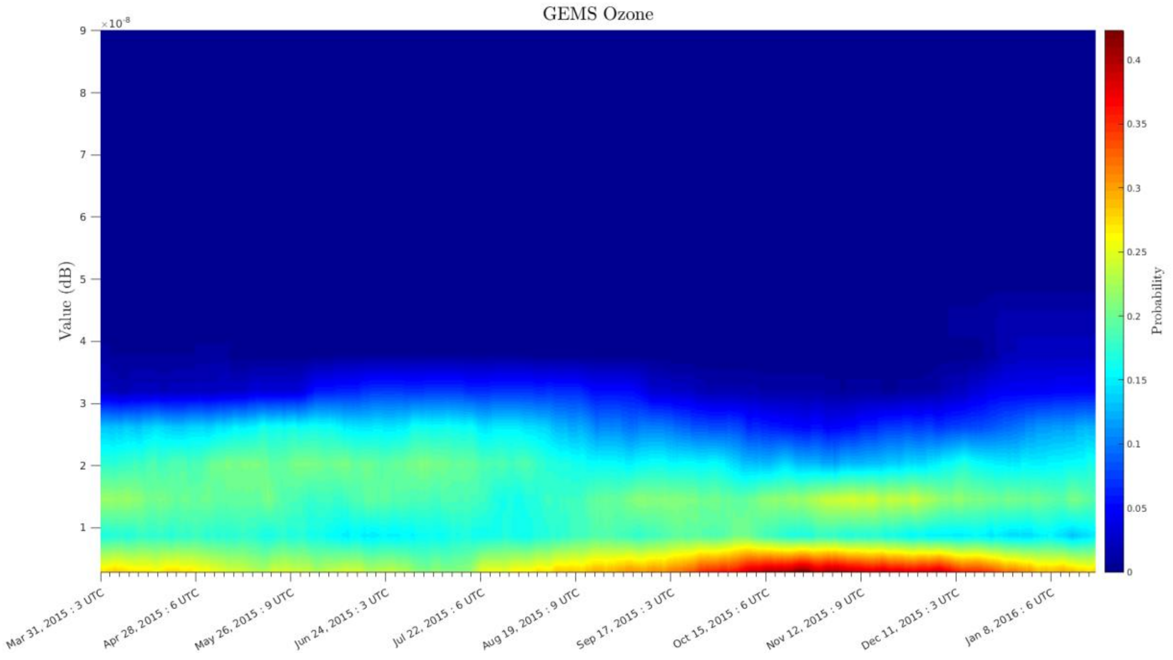

Figure 11: Moving histogram for ozone (night)

5. CONCLUSIONS

This text described the first approach to a methodology for the characterization of atmospheric composition data, obtained from the model developed by the MACC project. The proposed analyses focused on the identification of trends, cycles and tendencies present in the data, as these elements are key to evaluate the underlying physical processes.

Frequency analysis was used to identify the principal periodicities of the variables, and the extended frequency-altitude matrix shows the variation of the spectrum in the vertical direction. Day-cycle scatter plots can be used evaluate the changes of the concentration levels throughout the day, and variations of this cycle throughout the year. The moving histogram analysis is a robust tool to evaluate the temporal evolution of the variables, and allows the selection of specific moments.

REFERENCES

1. *The CAMS interim Reanalysis of Carbon Monoxide, Ozone and Aerosol for 2003-2015.* Flemming, Johannes, et al. 2017, Atmoshperic Chemsitry and Physics, pp. 1945-1983.

2. Copernicus. MACC Project Description. *Moniroring Atmospheric Composition and Climate.* [Online] 2015. http://www.gmes-atmosphere.eu/about/project/.

3. Arola, A., et al. *Validation report of the MACC near-real time global atmospheric composition service system - System evolution and performance statistics status up to 1 march 2015.* s.l. : Copernicus, 2015.

4. ECMWF. MARS user documentation. [Online] September 20, 2016. https://software.ecmwf.int/wiki/display/UDOC/MARS+user+documentation.

5. Flemming, Johannes, et al. *Tropospheric chemistry in the Integrate Forecasting System of ECMWF.* s.l. : Technical Memorandum, 2014.

6. European Centre for Medium-Range Weather Forecasts. L60 model level definitions. *ECMWF.* [Online] 2017. http://www.ecmwf.int/en/forecasts/documentation-and-support/60-model-levels.