PhD Agreement

Andres Yarce Botero

January 2018

1 Introduction

Colombia is characterized by its high biodiversity (one of the 17 megadiverse countries of the world (1)) and diverse ecosystems. The sustainable utilization of the biodiversity and agricultural production systems are seen nowadays as major pillars of growth (2; 3). Its territory is a composition of a complex mountainous system in the center, a wide plains over the east, on the south the thick jungles of the Amazon and the dense Chocó tropical forest at west.

Human activities generate pollution in the place where they are being carried out, the pollutants travel to remote areas carried by the wind and rivers currents, generating problems in the places where they are being deposited. The transport of urban atmospheric primary and secondary pollutants towards natural areas represents a threat to ecosystem functions.

Medellín, the second principal city of the country, is a source of pollutants with the potential to reach vulnerable ecosystems. Conceptualization of the valley (Medellín is located in a deep narrow seated valley) as a point-source, dubbed "The Volcano of the Aburraes", will permit the assimilation of local emissions data into regional models, and thus improve the assesment of urban-generated atmospheric pollution on vulnerable ecosystems. These vulnerable ecosystems are mainly located at the North and at the Northwest of the city, the reason why the contaminants can reach this two zones is due to the average annual trend of the trade winds in the global circulation patterns(4). The motivation for this research work is oriented to understand the atmospheric chemistry and transport processes affecting the mobilization of several compounds, from urban areas and agriculture production centers, to natural zones protected areas, so as to initiate the identification of those areas that need more than local conservation efforts for the preservation of their ecological functions.

Mathematical models are used to describe the interactions among diverse variables through certain atmospheric physical processes (5). Models aims to describe reality accurately, but generally, uncertainty exists between the simulated outputs of the model and the measured reality. Higher resolution estimations, via regional transport-chemical models is urgently needed to identify accutely vulnerable areas. To reduce the gap between the model output and the measurements, there is a technique known as Data Assimilation (DA). DA is a structured way to combine data from observations with the models with the objective of reducing the uncertainty of the forecast, calibrate the model, suggest new places for taking strategic measurements, and make uncertainty analysis over the model (6). This methodology is used to produce a regular, physically consistent, four-dimensional representation of the state of the atmosphere from a heterogeneous array of in-situ and remote instruments that sample imperfectly and irregularly in space and time(7).

The ultimate target of a DA is to have the residuals of the state of the system compared with the observational networks of sensors (if they exist) or another source of remote sensing data, as small as possible, under DA name a variety of methods exist which all try to reach this target (8). DA refers to the fact that all methods try to merge model forecast and measurements using the benefit of both sources of information(9).

There exist principal two classes of DA techniques: Variational and filters methods. The Variational DA methods are based on the minimization of a cost function using proper orthogonal decomposition adjoint method (10). This cost or objective function is typically the sum of squared differences between the data and the corresponding model values. From variational perspective exist basically two methods, the 3DVar and the 4DVar. The 3DVar method uses a static, flow-independent, background error covariance that is often spatially homogeneous and anisotropic (11). The 4DVar method allows the fitting of the model forecast trajectory to observations distributed over a period of time so as to provide more accurate model state estimations that are also more consistent with the prediction model (12).

The ensemble Kalman filter is a linear adaptive filter. The state analyzed is a linear combination of the forecast state and the data elements where the analyzed state is adapted proportional to the residue (13). With a Kalman filter is possible to make estimation to predict some response value and be seen as an extension of the Optimum Interpolator method (8), accounting for the evolution of errors from previous time.

An important difference between 4DVar and Kalman filtering is the form of the result. 4DVar provides an assimilated result in the form of piece-wise model evaluation, with discontinuities at the evaluation interval; the Kalman filters provides the result in terms of mean and covariance (8). Kalman and variationally methods are suitable to be used in online forecast applications, for offline applications such a parameter estimation, the variation approach is often favored due to its clear insight into how parameters are optimized, by comparison of model forecast based on certain parameter values measured. Compared to the 4DVar method the Kalman Filter is in general quite simple to implement since only the forward model is in use (14).

It is important to know how the results of the system are affected by these data incorporated to the model, for this reason through techniques of analysis it is possible to measure the impact of the data that is externally introduced at the assimilation process, this is intended to be other focus of my research project. In accordance with the preceding outlook, the following research questions have been proposed:

What are the ecosystem impacts from urban centers to natural areas of Colombia through the emission of pollutants transported via long-distance atmospheric mechanisms?

How can we adapt atmospheric Chemical Transport Models (CTM) to our region, so that they allow us to answer the question above and to formulate additional research hypotheses?

What kind of remote sensing data available could we use on DA experiences in order to reduce the uncertainty levels that the LE model is currently given for the interest natural protected regions in Colombia?

What DA perspective (Kalman or Variational) would be more appropiate for the regional interest on the natural protected areas, and under what criteria?

2 Research plan

In the 1^{st} year I have been studying about Data Assimilation techniques and have been working in the implementation of LE in different scenarios like for example the Aburrá Valley using nested domains. I have also been dealing with the updating of the meteorology input from one with lower resolution (ECMWF) to one with higher resolution (WRF). A similar situation with the updating of the current land use (GLC Global Land Cover) from another one more complete (CCI Climate Change Initiative ESA) that represent better the reality

One result of this first year was the participation on the 3^{rd} South American CMAS Conference in Brazil. For the next 8 months, I will focus my attention in the recent problems we have found with the running of the LE model in Colombia updating the LE inputs and coupling the LE with a better meteorology representation from the WRF model.

By the end of the first year, a literature review about the Data Assimilation techniques in Chemical Transport Models will be completed, and a paper showing the results of the LE implementation with the update sources will be submitted.

In the 2^{nd} year, the first version of the system for analize the impact of the deposition of contaminants in the protected areas will be finished applying different techniques of Data Assimilation with the observations available. A conference publication and a journal paper of the results will be completed.

In the 3^{rd} - 4^{th} the issues and problems of the variational technique used will be studied. The formalism and the method to improve the variational techniques will be developed. Two journal papers will be submitted and the PhD thesis will be completed.

The courses objective in this first year was to develop a mathematical formalism and approach the the Data Assimilation techniques. The following courses were or will be taken:

- Introduction to Real Analysis (Course taken in EAFIT University). **GS credits=5.**
- Measurement theory and Probability (Course taken in EAFIT University). **GS credits=5.**
- Elements of set theory (Course taken in EAFIT University). **GS credits=5.**
- Data Assimilation (Course will be taken in TU Delft, 2017/2018 - 2^{nd} Semester). **GS credits=5.**
- Data Assimilation Summer School (Course will be taken in Summer 2018). **GS credits=5.**

References

[1] E. Arbelaez, M. Torres, D.Lopez, J.Palacio, A. Mendoza, and C. Medina, *Colombian frozen biodiversity: 16 years of the tissue collection of the Humboldt Institute.* Acta Biologica Colombiana, 2015.

[2] C. Moreno, *Landscaping a Biofuture in Latin America.* The Centre for Research and Documentation Chile-Latin America, 2016.

[3] P. Rocha, *The Bioeconomy, A New Paradigm for Economic Growth and Sustainability?* ALCUE-KBBE IICA Anniversary, 2012.

[4] N. Pinel, J. Salazar, J. A. Posada, M. Rodrguez, A. M. Rendon, O. L. Quintero, and A. Yarce, *Potential urban pollution impacts on protected areas in Colombia through atmospheric teleconnection.* CMAS SouthAmerica, 2017.

[5] G. Evensen, *Data Assimilation.* Springer, 2009.

[6] M. Verlaan, N. Velzen, S. Hummel, and H. Gerritsen, *OpenDA, a generic toolbox for data assimilation in numerical modelling.* Deltares, VORtech,TU Delft, 2010.

[7] R. Daley, *Atmospheric data analysis.* Cambridge University Press, 1991.

[8] A. Segers, *Data assimilation in atmospheric chemistry models using Kalman filtering.* Delft University Press, 2002.

[9] M. Asch, M. Bocquet, and M. Nodet, *Data assimilation Methods, Algorithms and applications.* SIAM Society for Industrial and Applied Mathematics, 2016.

[10] M. Altaf, M. E. Gharamti, A.W.Hemmink, and I.Hoteit, *A reduced adjoint approach to variational data assimilation.* Comput. Methods Appl. Mech Engrg., 2012.

[11] S. Lu, *Variational data assimilation of satellite observations to estimate volcanic ash emissions.* Ipskamp Printer, 2017.

[12] C.Liu, M.Xue, and C.Liu, *Relationships among Four-Dimensional Hybrid Ensemble–Variational Data Assimilation Algorithms with Full and Approximate Ensemble Covariance Localization.* Monthly Weather Review, 144(2), 591–606., 2016.

[13] G. Fu, *Improving volcanic ash forecast with ensemble-based data assimilation.* Ipskamp Printer, 2016.

[14] M. Verlaan and J. Sumihar, *Observation Impact methods for storm surge forecasting systems.* Ocean Dynamics 66(2), 221–241., 2016.

[15] A. Barbu, *Ensemble-based data assimilation schemes for atmospheric chemistry models.* Wohrmann Print Service, 2010.