Inspire Create Transform

Vigilada Mineducación

Delft
University of
Technology

Ensemble-based Data Assimilation For High-uncertainty systems: Case of study, PM10 and PM2.5 in the Aburrá Valley

Santiago Lopez-Restrepo

Advisors:

Olga Lucía Quintero Montoya
Nicolás Pinel Pelaez

Doctoral Seminar 4
Medellín, Colombia
October, 2018

Inspire Create Transform

TU Delft Delft University of Technology

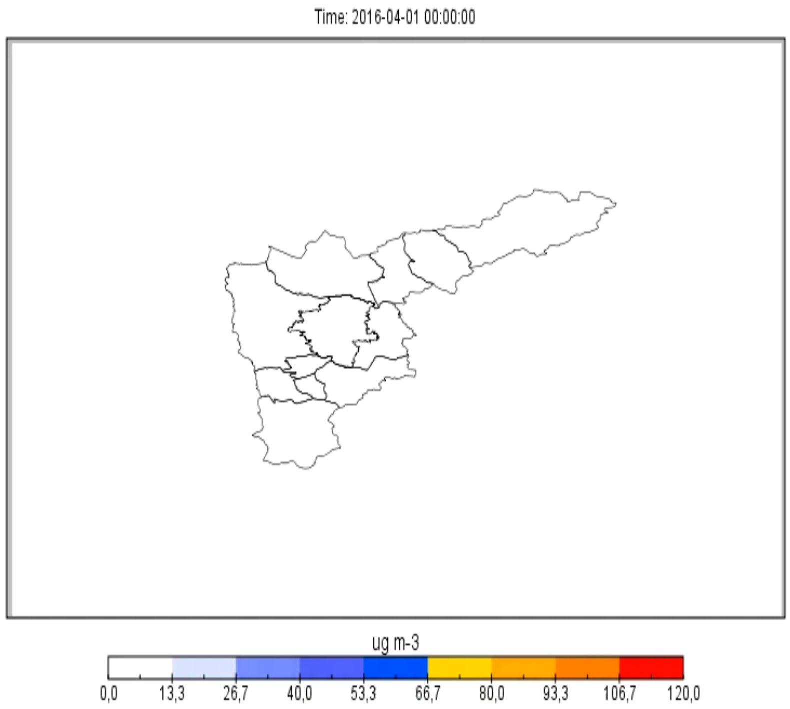
UNIVERSIDAD EAFIT

Outline

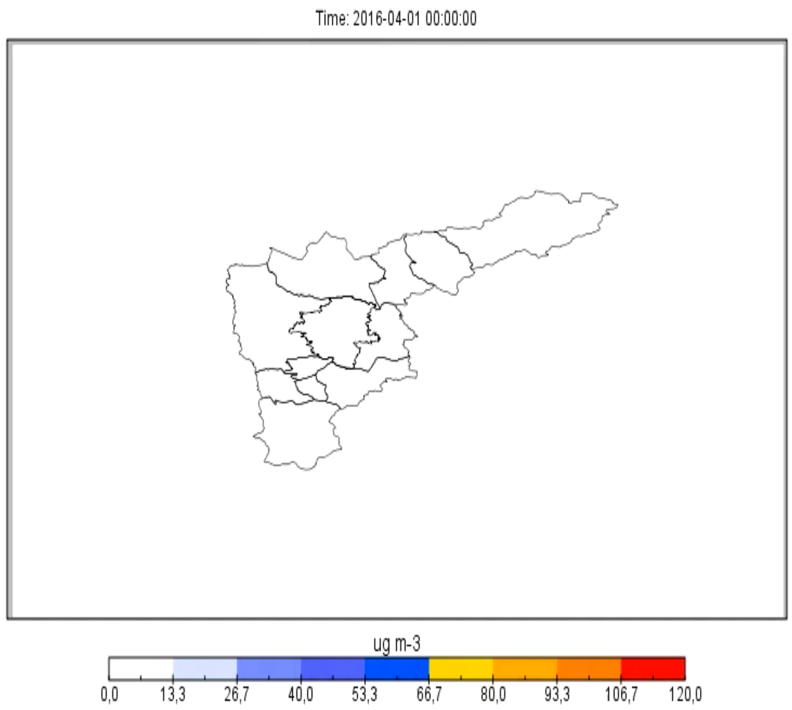
- Introduction
- LOTOS-EUROS Model
- Data Assimilation
- Preliminary Results

Research Status

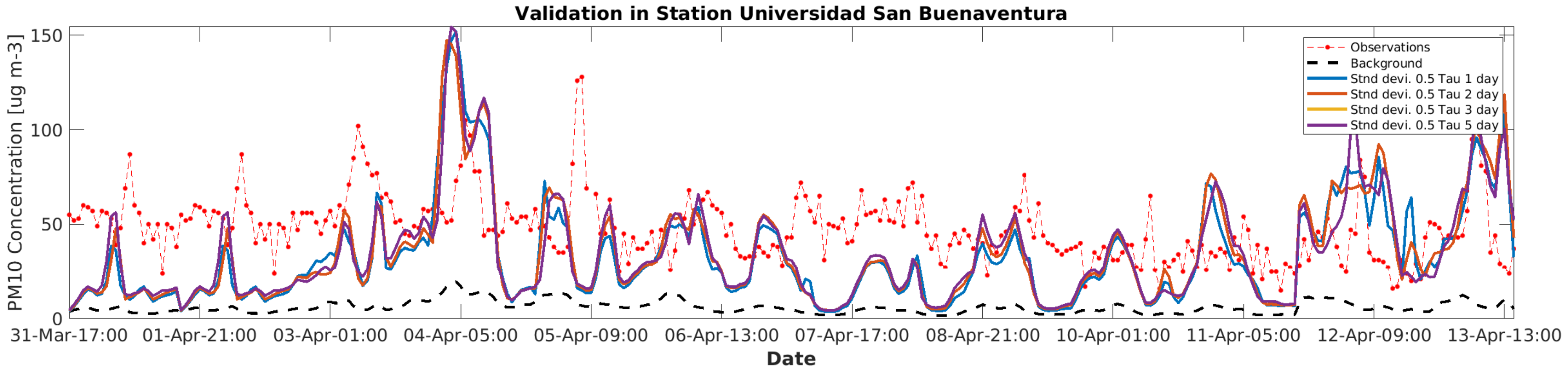
Background

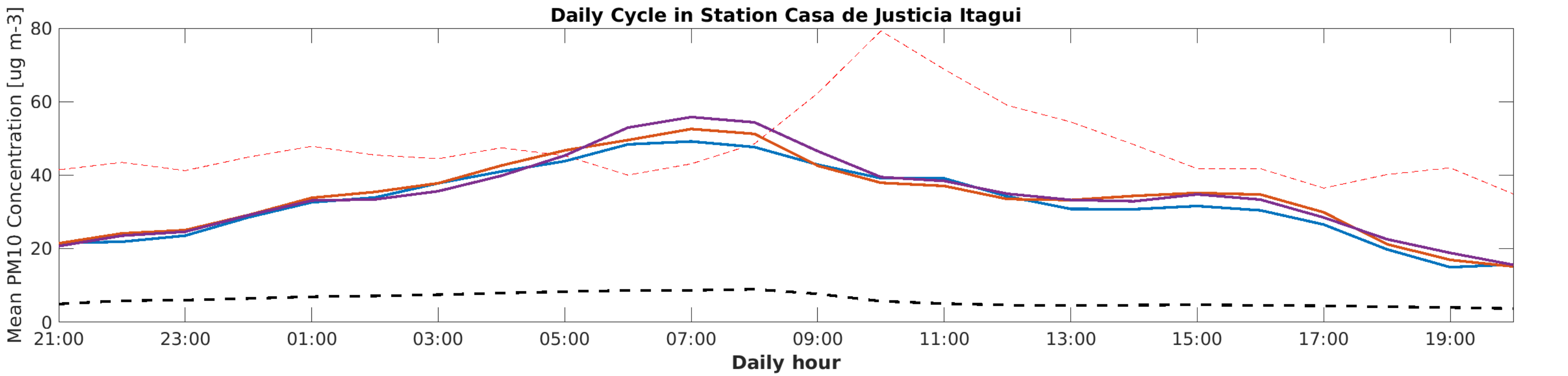


EnKF

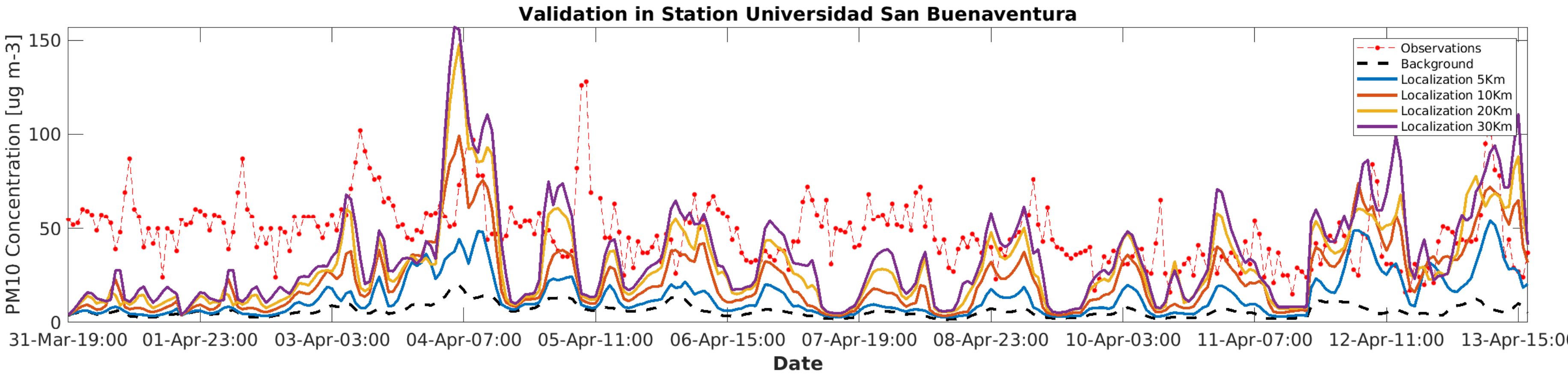


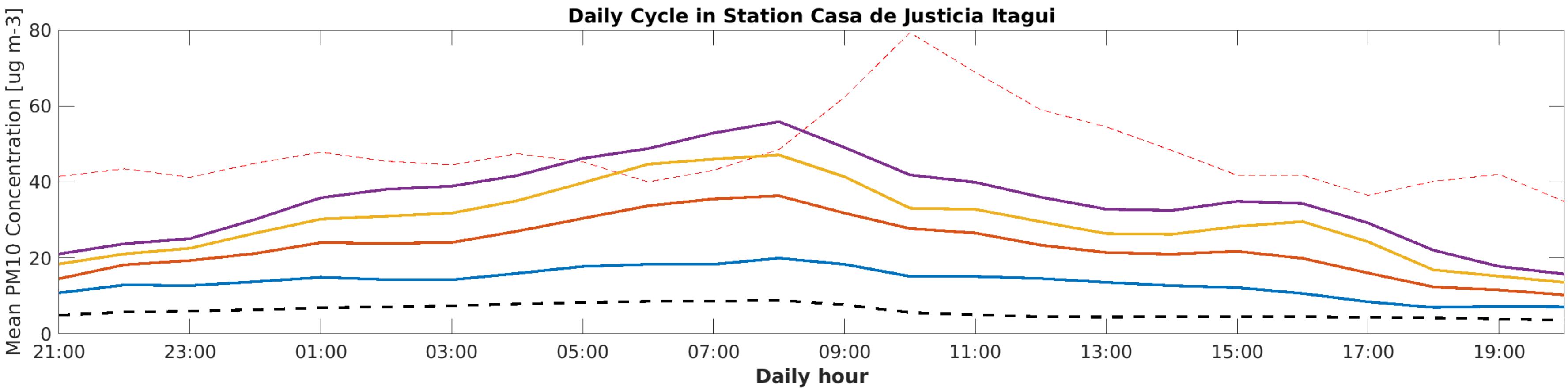
Research Status



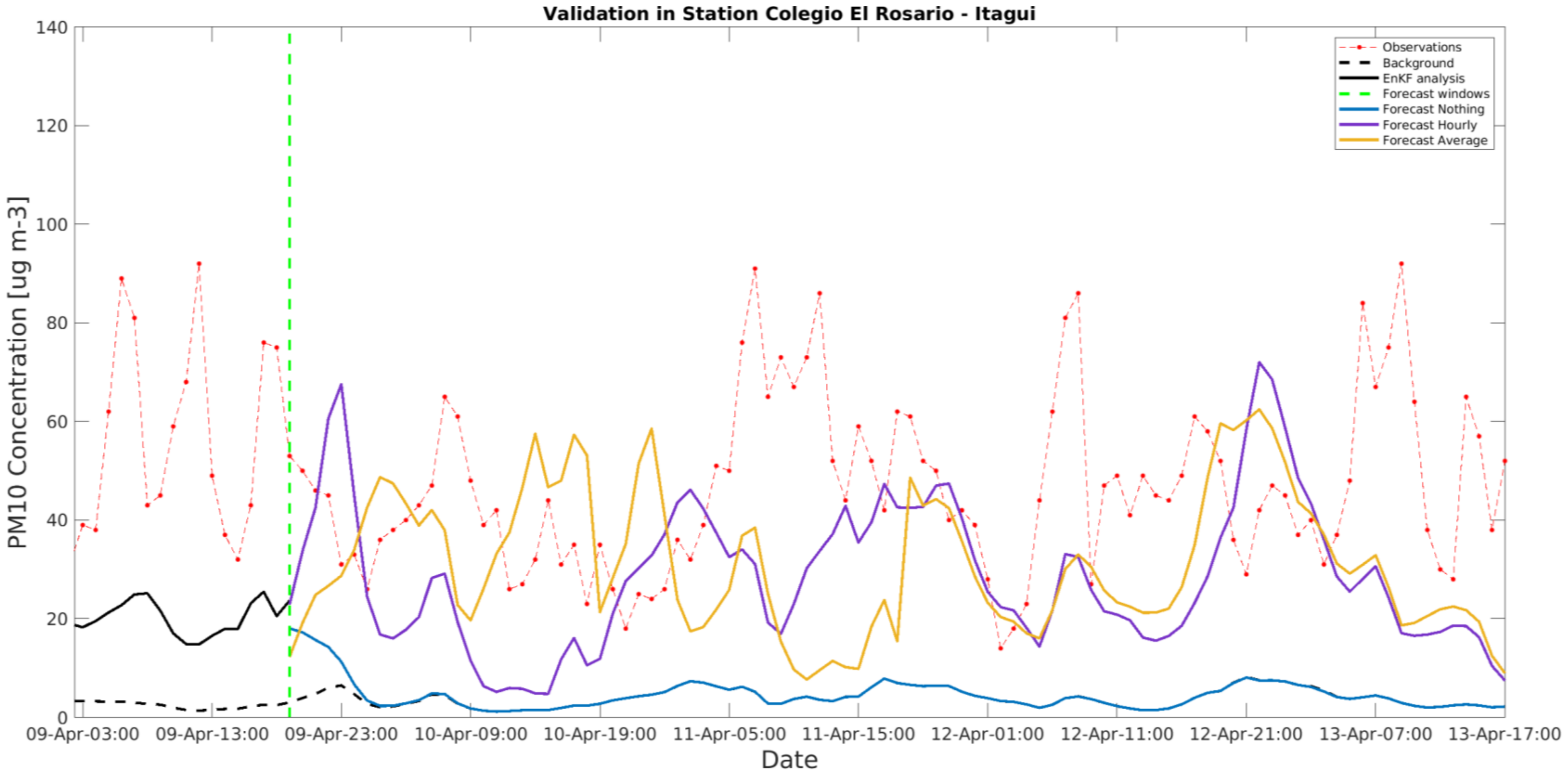


Research Status



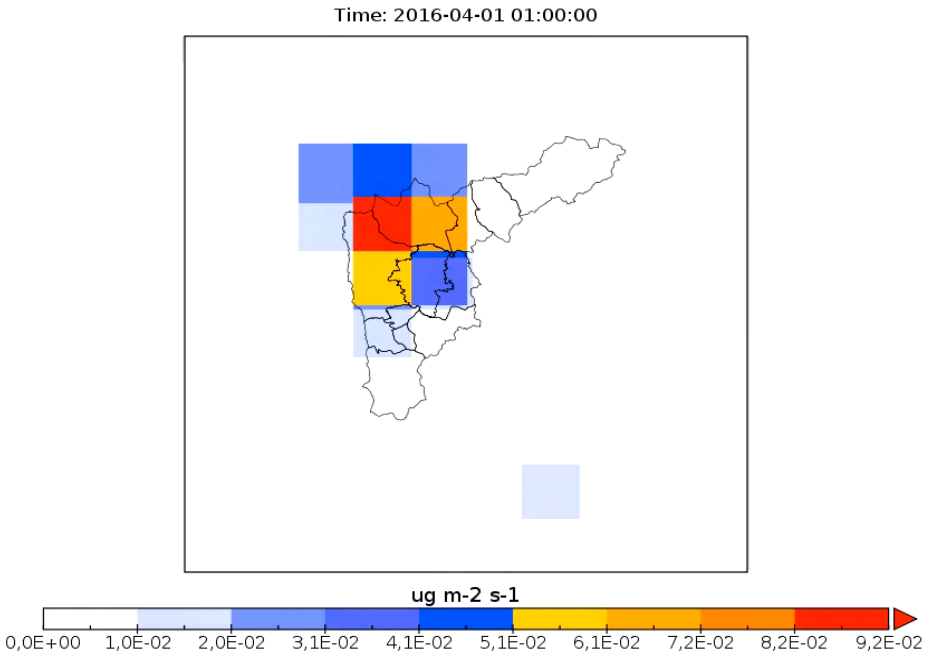


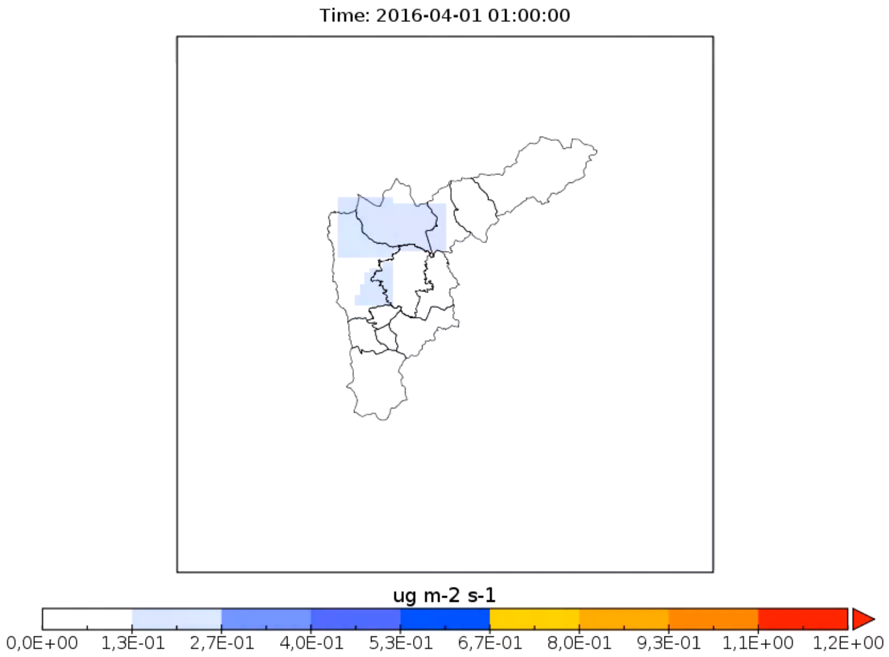
Research Status



Research Status

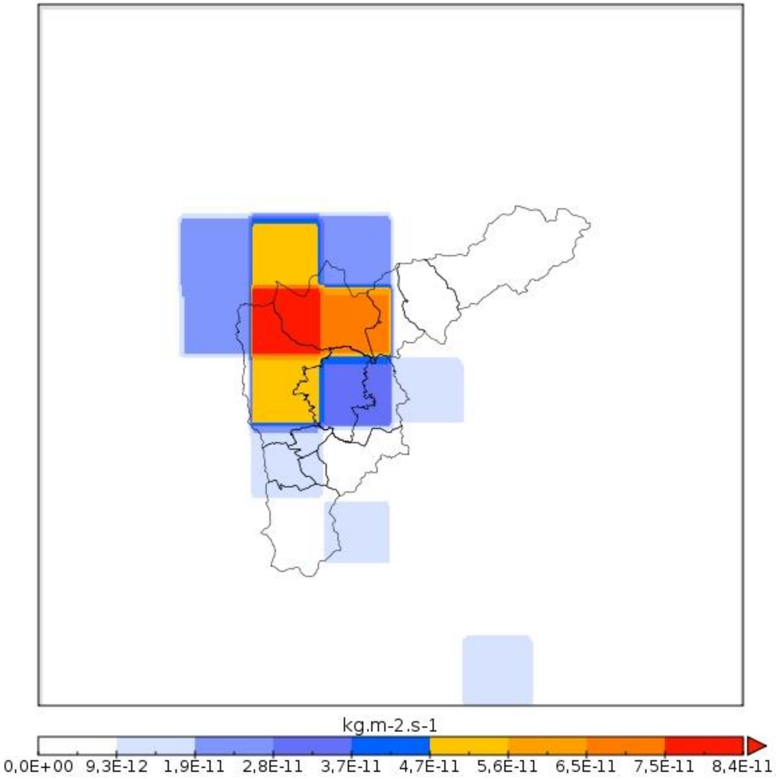
Just Model Emissions PM2.5 | **Data Assimiliation Emissions PM2.5**



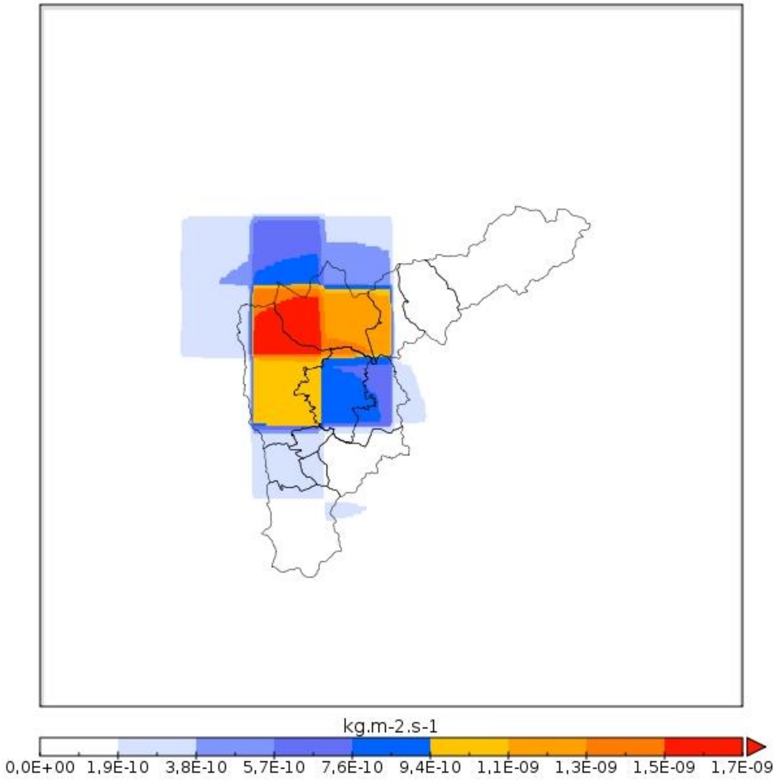


Research Status

Just Model Emissions Means



Data Assimilation Emissions Means



Research Description

The KF is an optimal method when different assumptions about the statistical behavior of the uncertainties are met.

But in many real applications there are not enough information to characterize the system uncertainties or there are too many uncertainties sources.

It is proposed a three-year plan to develop a Ensemble-based data assimilation scheme to lead with high-uncertainty and high-dimensional systems.

Research Description

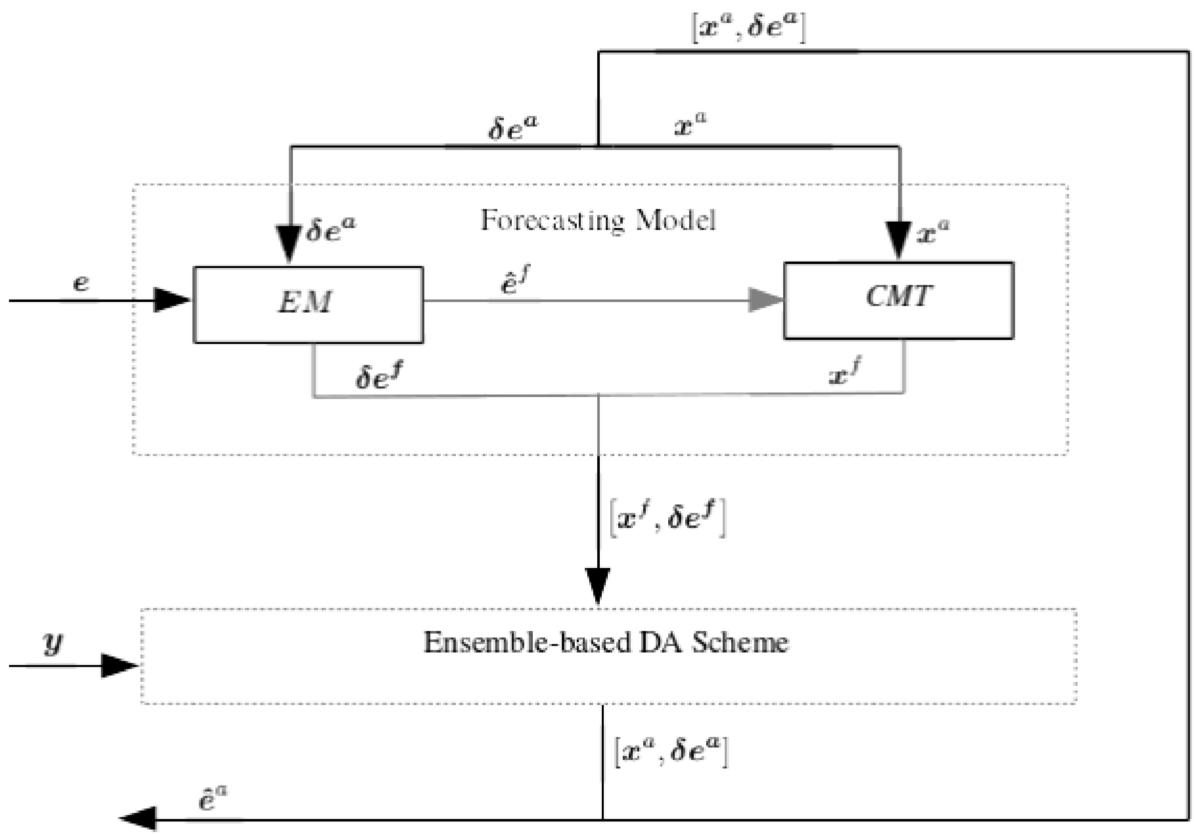
The scheme will be focus in three important aspects of the data assimilation process.

To implement a ensemble robust filter to evaluate its performance in a high-uncertainty system.

To develop a covariance localization technique that incorporates knowledge about the system.

To develop an uncertainty propagation model using phenomenological knowledge of the parameters to be estimated.

Research Description



Schematic representation of parameter estimation using a model to propagate the emission uncertainty and a Ensemble-based DA scheme. Based in (Peng et al., 2017).

Robust Ensemble-Based DA

- In the EnKF it is necessary to make some statistical assumptions related to the uncertainty in the model and the observations, that in many of real applications are no truth. For instance, the Gaussian distribution of the state error.

- A different approach when the systems condition does not satisfy the requirement of the KF-based methods are the robust filters or robust estimators.

- The robust filters emphasize the robustness of the estimation, so that they may have better tolerances to possible uncertainties in assimilation. Since its purpose is not the optimality in the estimation, the robust estimator does not require an exactly statistical representation of the error, showing a better performance that the KF-based methods in scenarios with poor statistical representation of the uncertainty.

Robust Ensemble-Based DA

- Unlike the KF that minimize the variance of the estimation error, the HF is based on the criterion of minimizing the supremum of the L_2 norm of the uncertainty sources (initial conditions, parameters, boundary conditions, etc.) (Han et al., 2009). The HF requires that the total energy of the estimation errors, be no longer than the uncertainty source energy times a factor $1/\gamma$:

$$\sum_{t=0}^{K} \left|\left|x_t^t - x_t^a\right|\right|_{S_t}^2 \leq \frac{1}{\gamma}\left(\left|\left|x_0^t - x_0^a\right|\right|_{\Delta_0^{-1}}^2 + \sum_{t=0}^{K} \left|\left|u_t\right|\right|_{Q_t^{-1}}^2 + \sum_{t=0}^{K} \left|\left|v_t\right|\right|_{R_t^{-1}}^2\right) \quad (1)$$

Where x_t^t is the truth state, x_t^a is the analysis state, S_t is a user-chosen matrix of weights, u_t and v_t are the model and observation uncertainty, Δ_t , Q_t and R_t are the uncertainty weights matrices with respect to the initial conditions, model error and observations error.

Robust Ensemble-Based DA

To solve equation (1), it is defined first the following cost function J^{HF}:

$$J^{HF} = \frac{\sum_{t=0}^{K} \left|\left|x_t^t - x_t^a\right|\right|_{S_t}^2}{\left|\left|x_0^t - x_0\right|\right|_{\Delta_0^{-1}}^2 + \sum_{t=0}^{K} \left|\left|u_t\right|\right|_{Q_t^{-1}}^2 + \sum_{t=0}^{K} \left|\left|v_t\right|\right|_{R_t^{-1}}^2}$$

Then the inequality (1) is equivalent to $J^{HF} \leq \frac{1}{\gamma}$ Let γ^* be the value such that:

$$\frac{1}{\gamma^*} = \underset{\boldsymbol{x_t^a}}{inf} \; \underset{x_0, \{u_t\}, \{v_t\}}{sup} J^{HF}, t \leq K$$

The optimal HF is achieved when $\gamma = \gamma^*$. In this sense, the evaluation of $\gamma*$ is an application of the minimax rule, a strategy that aims to provide robust estimates and is different from its Bayesian counterpart.

Robust Ensemble-Based DA

The γ is then know as performance level of the HF. The inequality (1) can be solve iteratively, similar that in the KF:

$$x_t^f = M_{t-1}(x_{t-1}^a)$$

$$\Delta_t^f = M_{t-1}\Delta_{t-1}^a M_{t-1}^T + Q_t$$

$$(\Delta_t^a)^{-1} = \left(\Delta_t^f\right)^{-1} + H^T(R_t)^{-1H} - \gamma S_t$$

$$G_t = \Delta_t^{aH^T}(R_t)^{-1}$$

$$x_t^a = x_t^f + G_t\left(y_t - Hx_t^f\right)$$

Subscript to the constraint:

$$(\Delta_t^a)^{-1} = \left(\Delta_t^f\right)^{-1} + H^T(R_t)^{-1}H - \gamma S_t \geq 0.$$

Robust Ensemble-Based DA

where Δ denotes the uncertainty matrix, analogues to the covariance matrix P in the KF, and G_t is the HF Gain matrix analogues to the Kalman Gain K_t . To compare directly the HF and the KF, let rewrite:

$$P_t^a = \left(I - K_{tH_t}\right) P_t^f .$$

as:

$$(P_t^a)^{-1} = \left(P_t^f\right)^{-1} + H^T (R_t)^{-1} H$$

In addition, let

$$(\Sigma_t^a)^{-1} = \left(\Delta_t^f\right)^{-1} + H^T (R_t)^{-1} H$$

Then, it is clear that Σ_t^a is a uncertainty matrix created updating Δ_t^f trough a KF, where:

$$(\Delta_t^a)^{-1} = (\Sigma_t^a)^{-1} - \gamma S_t < (\Sigma_t^a)^{-1}$$

Robust Ensemble-Based DA

The EnLTHF proposed in (Luo and Hoteit, 2011) is a time-local version of HF which utilizes only the current state and observations of the system rather than the entire available history (Nan and Wu, 2017). Unlike the HF where the cost function J^{HF} is defined in all the assimilation windows, in the EnLTHF a local cost function is proposed:

$$J_t^{HF} = \frac{\left|\left|x_t^t - x_t^a\right|\right|_{S_t}^2}{\left|\left|x_0^t - x_0\right|\right|_{\Delta_0^{-1}}^2 + \left|\left|u_t\right|\right|_{Q_t^{-1}}^2 + \left|\left|v_t\right|\right|_{R_t^{-1}}^2}$$

Similarly to equation (1), it is required that:

$$\left|\left|x_t^t - x_t^a\right|\right|_{S_t}^2 \leq \frac{1}{\gamma_t}\left(\left|\left|x_0^t - x_0^a\right|\right|_{\Delta_0^{-1}}^2 + \left|\left|u_t\right|\right|_{Q_t^{-1}}^2 + \left|\left|v_t\right|\right|_{R_t^{-1}}^2\right)$$

Robust Ensemble-Based DA

where γ_t is a suitable local performance level, which satisfies:

$$\frac{1}{\gamma_t} \geq \frac{1}{\gamma_t^*} = \inf_{\boldsymbol{x_t^a}} \sup_{x_0, \{u_t\}, \{v_t\}} J_t^{HF}, t \leq K$$

with $\frac{1}{\gamma_t^*}$ being the minimax point of the local cost function J_t^{HF}.

Robust Ensemble-Based DA

The EnLTHF can be expressed in terms of the EnKF algorithm using the notation of (Luo and Hoteit, 2011)

$$[\Sigma_t^a, K_t] = EnKF(x_t^a, Q_t, H)$$

$$G_t = [I_m - \gamma_t \Sigma_t^a S_t]^{-1} K_t$$

$$\xi_t^{a(i)} = \xi_t^{f(i)} + G_t \left[y_t - H_t \xi_t^{f(i)} + v_t^i \right]$$

$$x_t^a = \left(\sum_{i=1}^{N} \xi_t^{a(i)} \right) / N$$

$$\Delta_t^a = [I_m - \gamma_t \Sigma_t^a S_t]^{-1} K_t$$

subject to the constraint

$$(\Delta_t^a)^{-1} = (\Sigma_t^a)^{-1} - \gamma_t S_t \geq 0$$

where the operator $EnKF()$ means that Σ_t^a and K_t are obtained through the EnKF.

Robust Ensemble-Based DA

3938 MONTHLY WEATHER REVIEW VOLUME 139

Robust Ensemble Filtering and Its Relation to Covariance Inflation in the Ensemble Kalman Filter

XIAODONG LUO AND IBRAHIM HOTEIT

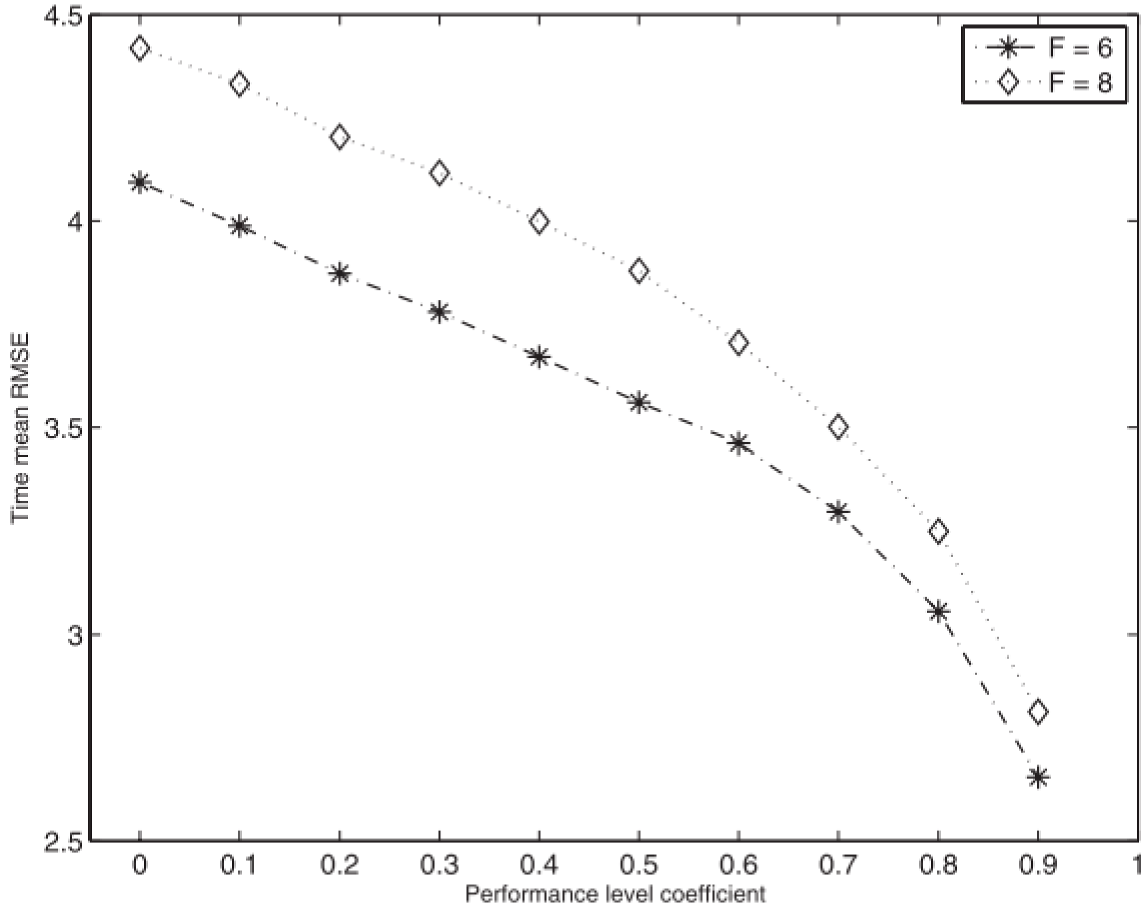
King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

(Manuscript received 1 December 2010, in final form 11 April 2011)

Numerical experiment with the model Lorenz 96

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F(t),\ i = 1, \dots, 40$$

Robust Ensemble-Based DA



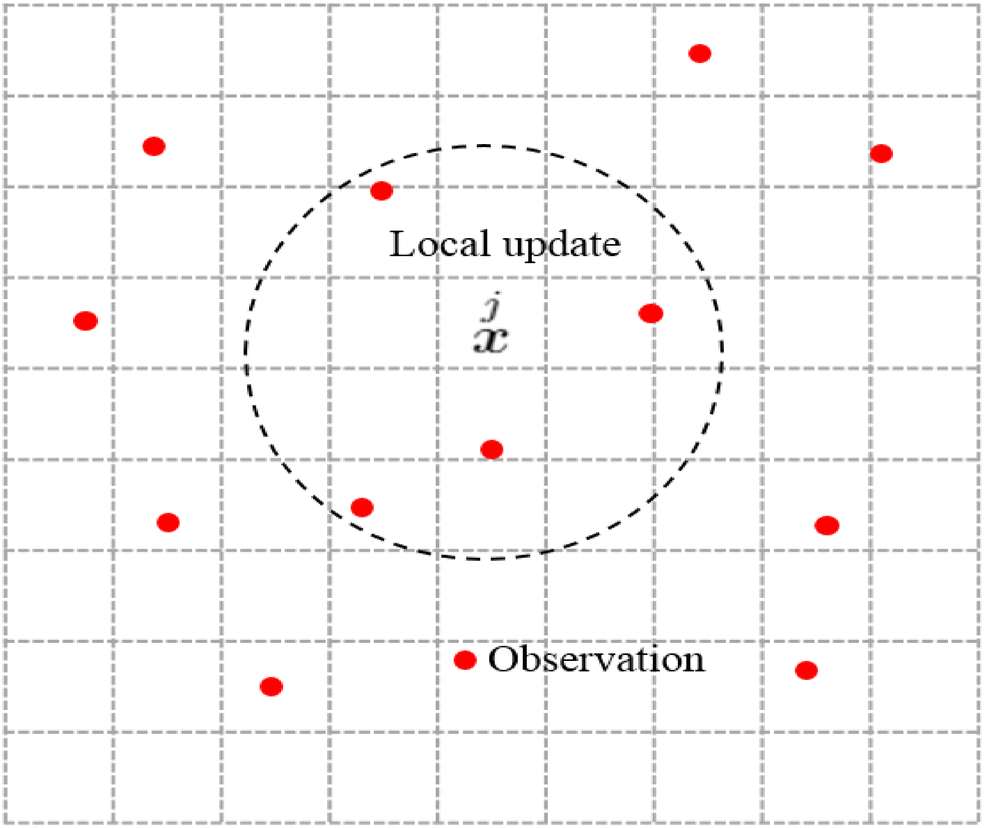
Dynamic-spatial localization

- The idea of distance-dependent localization technique where the localization radius can vary according with knowledge about the system has been little studied in atmospheric DA applications.

- In History Matching problems, there are more applications that try to do something related with theses ideas. Specifically in Soares et al., 2018) two different localization methods in a reservoir parameter estimation case are compared. In the first one, are used influences areas of the observations and delimit the localization windows. The second method is based in streamlines and selection of the historical time when they presented the biggest area and trace the influence area

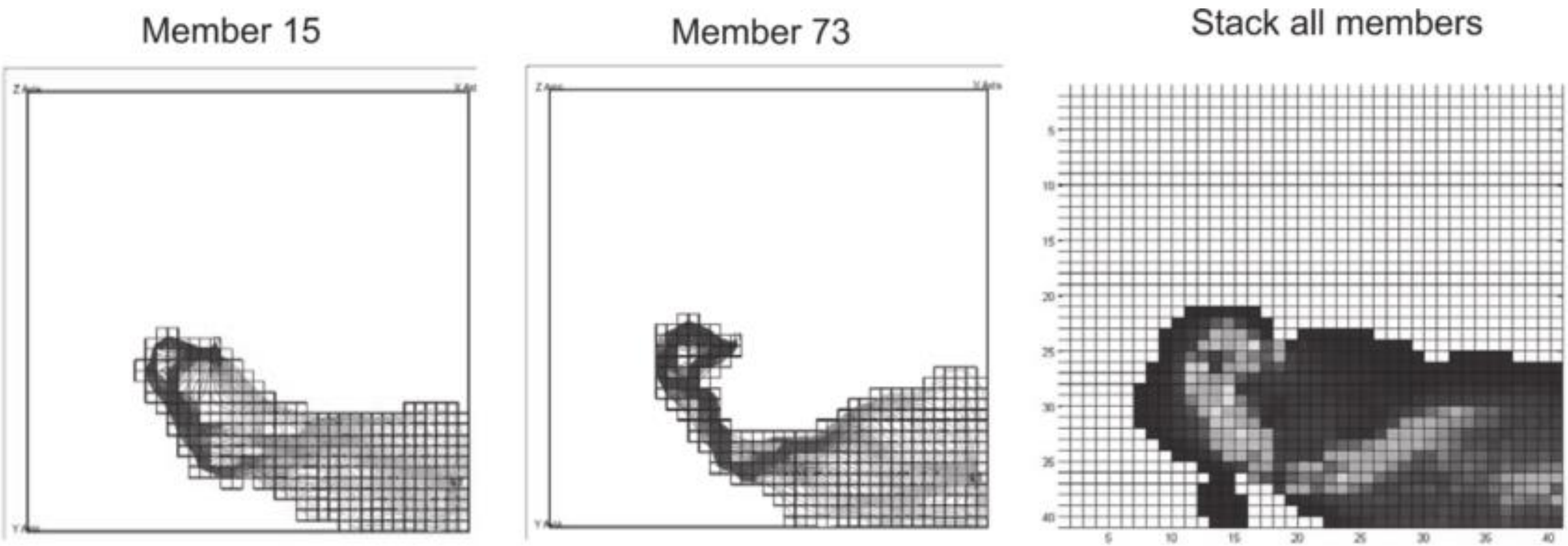
- Localization technique based in streamlines simulation are proposed as an alternative to the distance-dependent methods and using more efficiently the model information.

Dynamic-spatial localization



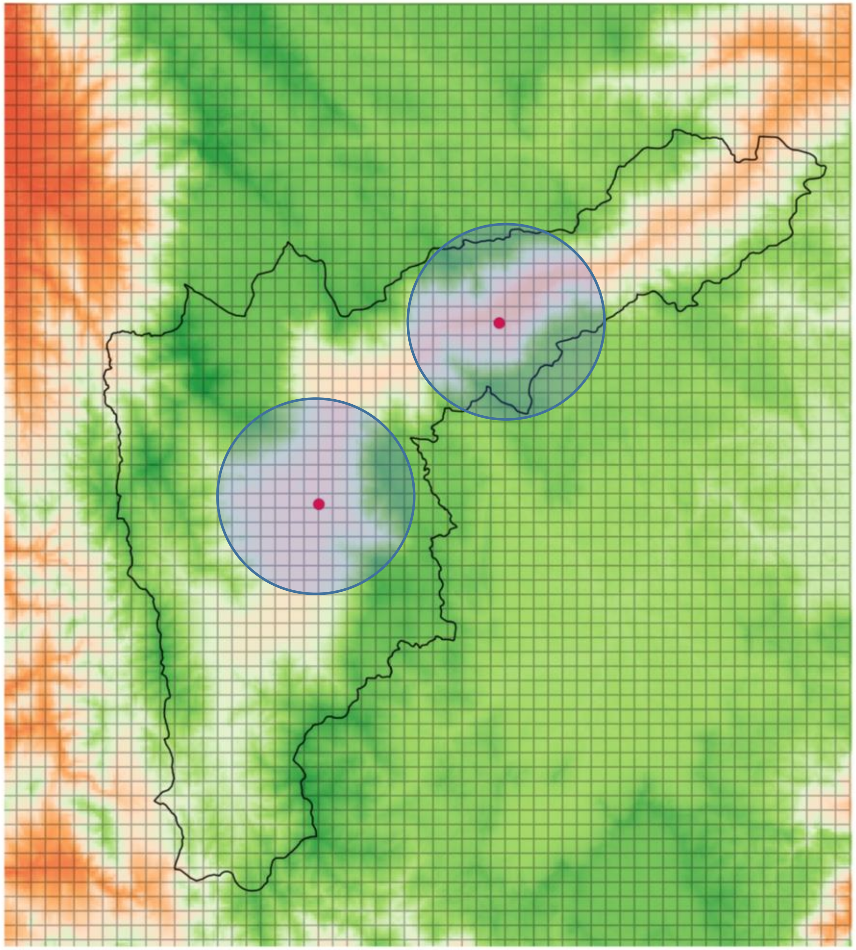
Distance-dependent localization

Dynamic-spatial localization

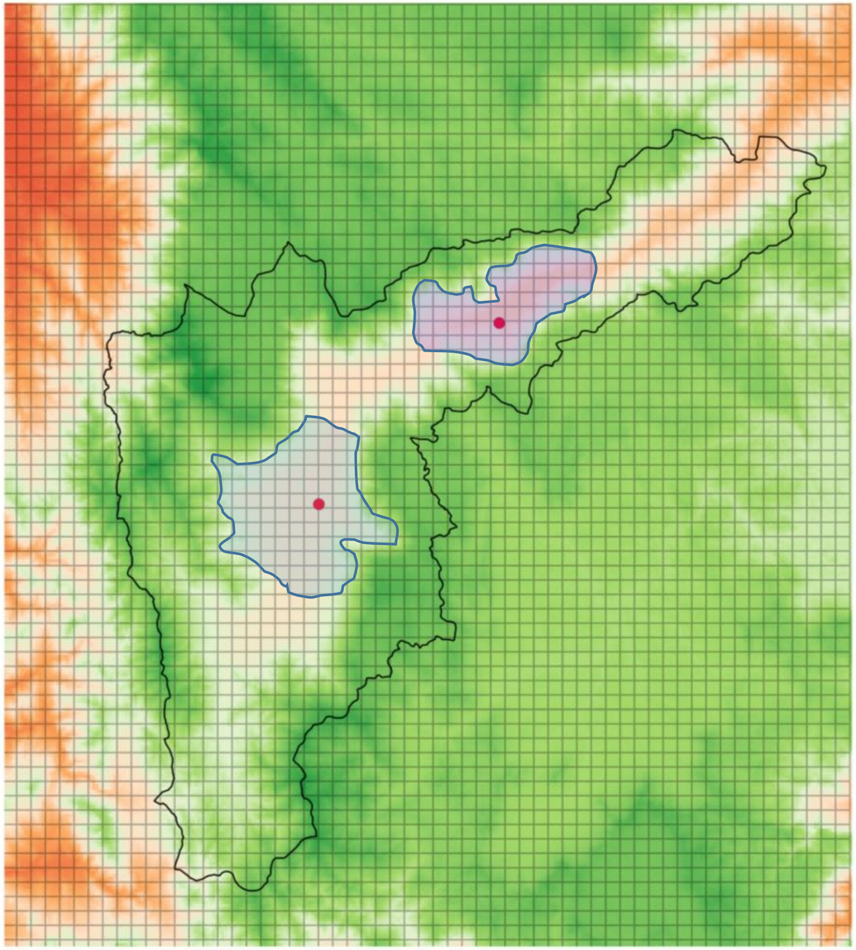


Streamlines simulation localization

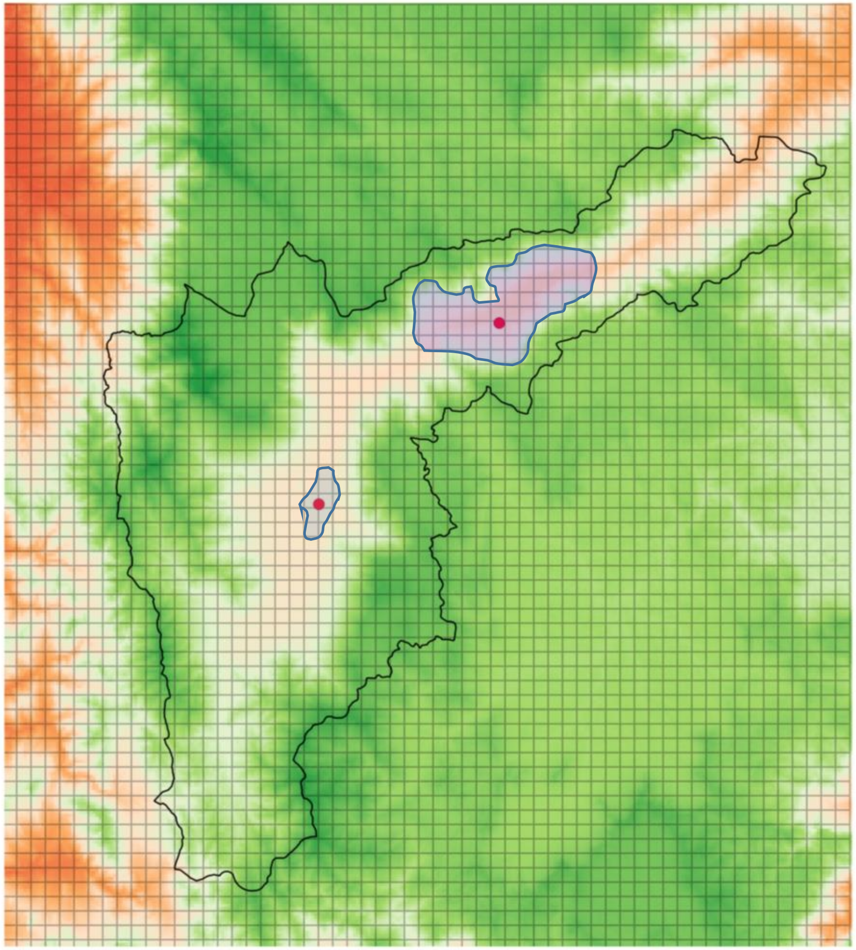
Dynamic-spatial localization



Dynamic-spatial localization



Dynamic-spatial localization



Uncertainty Model Propagation

- In most application of parameter estimation and uncertainty modelling using Ensemble-based DA are followed two approaches: to model the uncertainty as a stochastic process (colored noise) like is described in (Heemink and Segers, 2002) or as a combination of old value like in (Peng et al., 2017).

- The two methods are explain next taking as example the emission estimation in CTMs applications.

Uncertainty Model Propagation

In (Heemink and Segers, 2002) the deterministic model state is represented in discrete time as:

$$x_t = M(x_{t-1})$$

Since the emissions are an important source or error, the uncertainty in the emission are modeled as a stochastic process, for this case, as a colored noise (Jazwinski, 1970):

$$\delta e_t = \alpha \delta e_{t-1} + \sqrt{1-\alpha^2}\, w_t$$

where w_t is a white noise and δe_t is the emission correction factor. Thus, he stochastic model state is formed by augmenting the state vector with the correction factor δe_t:

$$\begin{bmatrix} x_t \\ \delta e_t \end{bmatrix} = \begin{bmatrix} M(x_{t-1}) \\ \alpha \delta e_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \sqrt{1-\alpha^2} \end{bmatrix} w_t$$

Uncertainty Model Propagation

On the other hand, the method proposed in (Peng et al., 2017) uses a persistence forecasting operator that serve as the forecast model for the emission correction factors. This forecast model is built by a smooth operator using the state ensemble and the previous analysis value of emission correction factors λ_t^a :

$$\lambda_t^{p(i)} = \frac{\xi_t^{f(i)}}{x_t^f}$$

$$\lambda_t^{f(i)} = \frac{1}{T}\left(\sum_{j=t-T+1}^{t-1} \lambda_j^{a(i)} + \lambda_t^{p(i)}\right)$$

Here, T is the time windows of the smooth operator.

Uncertainty Model Propagation

The state vector is augmented with the correction factors λ and can be estimated through Ensemble-based DA. With this, is created a forecast emission ensemble and an analysis forecast ensemble following:

$$\hat{e}_t^f(i) = \lambda_t^{f(i)} e_t$$
$$\hat{e}_t^a(i) = \lambda_t^{a(i)} e_t$$


THANK
YOU
GRACIAS
ARIGATO
SHUKURIA
GOZAIMASHITA
EFCHARISTO
DANKSCHEEN
TASHAKKUR ATU
YAQHANYELAY
SUKSAMA
EKHMET
GRAZIE
MEHRBANI
PALDIES
KOMAPSUMNIDA
MAAKE
JUSPAXAR
BOLZÏN
MERCI
BİYAN
SHUKRIA
TINGKI


COLOMBIA

References

Barbu AL, Segers AJ, Schaap M, Heemink AW, Builtjes PJH. A multi-component data assimilation experiment directed to sulphur dioxide and sulphate over Europe. *Atmos Environ*. 2009; 43(9):1622–1631.

Fu, G., Prata, F., Lin, H. X., Heemink, A., Segers, A., and Lu, S. *Data assimilation for volcanic ash plumes using a satellite observational operator: a case study on the 2010 Eyjafjallajökull volcanic eruption* Atmos. Chem. Phys., 2017, 17, 1187-1205, doi:10.5194/acp-17-1187-2017,

Fu, G., Heemink, A., Lu, S., Segers, A., Weber, K., and Lin, H.-X. *Model-based aviation advice on distal volcanic ash clouds by assimilating aircraft in situ measurements* Atmos. Chem. Phys., 2016,16, 9189-9200, doi:10.5194/acp-16-9189-2016.

Green J, Sánchez S. Air Quality in Latin America: An Overview. Clean air Institute. Washington D.C., USA. 2012.

Hendriks C, Kranenburg R, Kuenen J, van Gijlswijk R, Kruit RW, Segers A, van der Gon HD, Schaap M. The origin of ambient particulate matter concentrations in the Netherlands. Atmos Environ. 2013; 69, 289–303.

References

Kumar A, Jiménez R, Belalcázar L, Rojas N. Application of WRF-Chem Model to Simulate PM10 Concentration over Bogota. *Aerosol Air Qual Res*. 2016; 16:1206–1221.

Lu, S., H. X. Lin, A. Heemink, A. Segers, and G. Fu. *Estimation of volcanic ash emissions through assimilating satellite data and ground-based observations* J. Geophys. Res. Atmos., 2016, 121, 10,971-10,994, doi:10.1002/2016JD025131

Lu, S., Lin, H. X., Heemink, A. W., Fu, G., and Segers, A. J. Estimation of volcanic ash emissions using trajectory-based 4D-Var data assimilation Monthly Weather Review, 2015, e-View, doi:10.1175/MWR-D-15-0194.1

Marécal V, Peuch VH, Andersson C, Andersson S, Arteta J, Beekmann M, *et al*. A regional air quality forecasting system over Europe: The MACC-II daily ensemble production. *Geosci Model Dev*. 2015; 8(9):2777–2813.

Rendón AM, Salazar JF, Palacio CA, Wirth V, Brötz B. Effects of urbanization on the temperature inversion breakup in a mountain valley with implications on air quality. *J App Meteorol Climatol*. 2014; 53:840–858.