Seminar 4 of the PhD in Mathematical Engineering

Adjoint free variational data assimilation and model order reduce techniques

Student

Andrés Yarce Botero

Advisors

Nicolás Pinel, Olga Lucía Quintero, Arnold W. Heemink

12/10/18

Outline

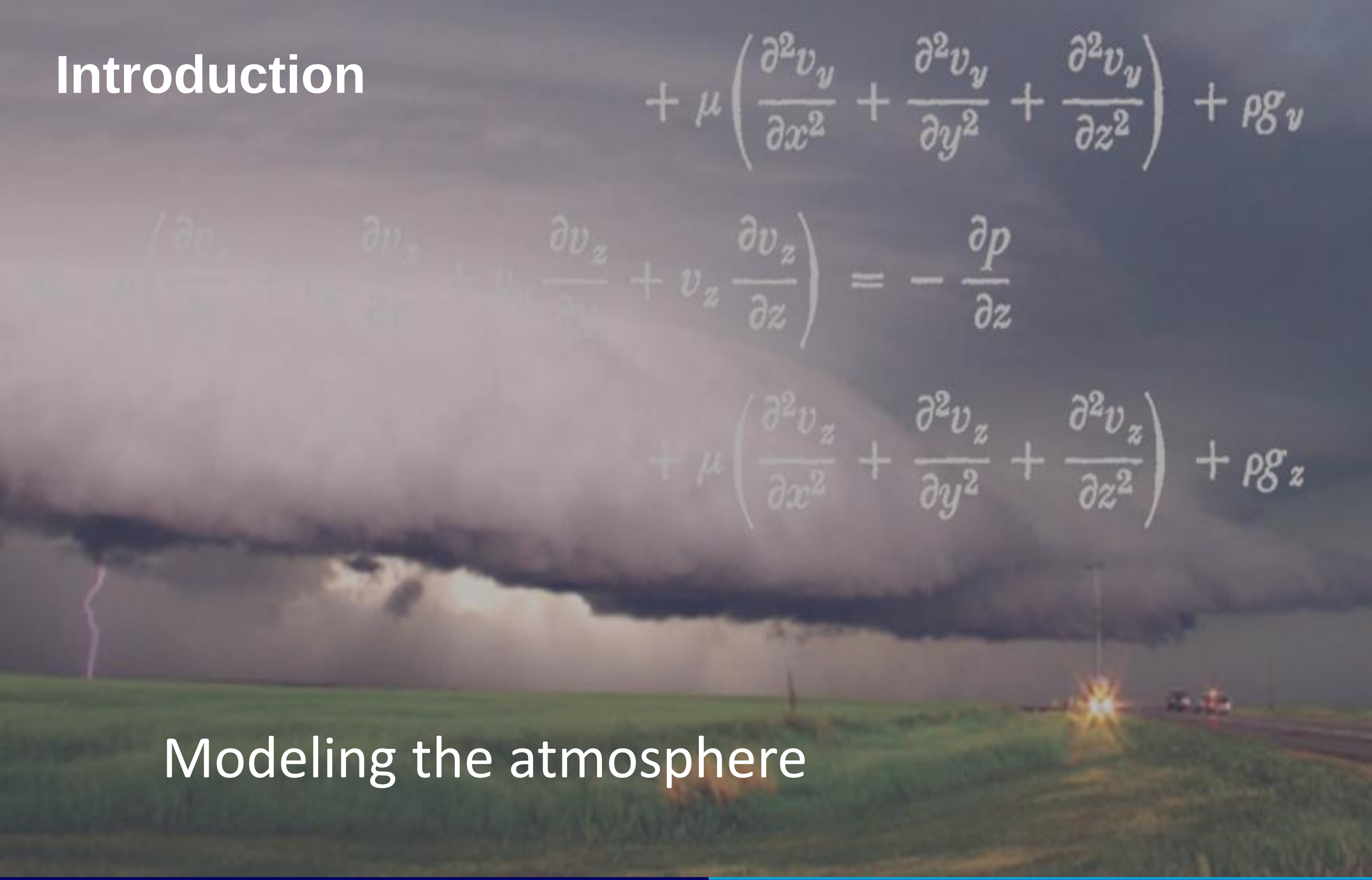
- Motivation
- Introduction
- Model order reduction
- Adjoint free data assimilation techniques
- References

Motivation

¿Can we reduce the cost of data assimilation in the context of atmospheric chemical transport simulation without degrading the results?

¿How can one avoid or minimize the problem of not having an adjoint model for a highly non linear, large scale model?

Introduction



$$+\mu\left(\frac{\partial^2 v_y}{\partial x^2}+\frac{\partial^2 v_y}{\partial y^2}+\frac{\partial^2 v_y}{\partial z^2}\right)+\rho g_y$$

$$\cdots \frac{\partial v_z}{\partial y}+v_z\frac{\partial v_z}{\partial z}\Big)=-\frac{\partial p}{\partial z}$$

$$+\mu\left(\frac{\partial^2 v_z}{\partial x^2}+\frac{\partial^2 v_z}{\partial y^2}+\frac{\partial^2 v_z}{\partial z^2}\right)+\rho g_z$$

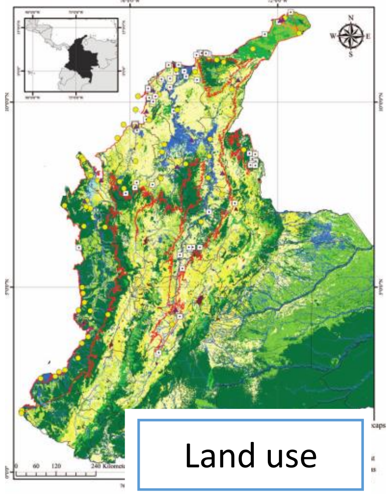
Modeling the atmosphere

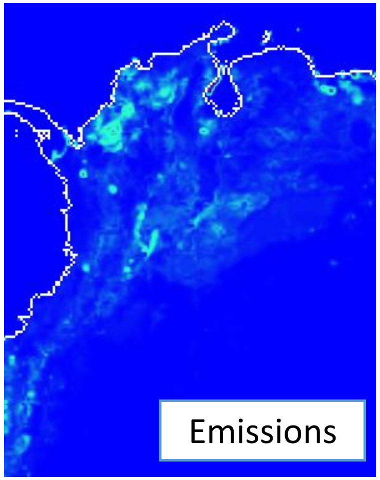
Introduction

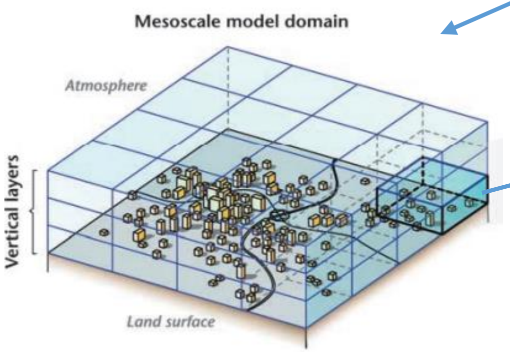
Many modern mathematical models of **real-life** processes pose challenges when used in numerical simulations, due to their complexity and large size (**dimension**)

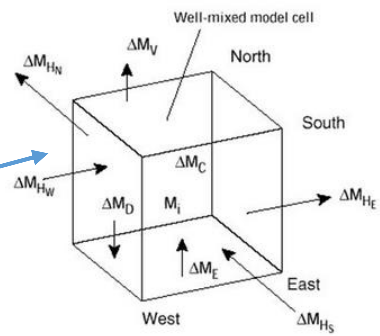
LOTOS-EUROS

(LOng Term Ozone Simulation- EURopean Operational Smog model)

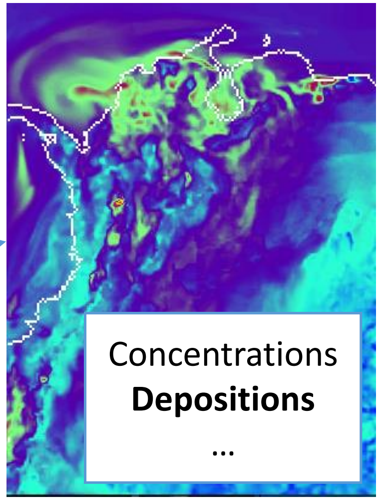








H is horizontal transport
V is vertical transport
E is emissions
D is surface deposition
C is chemical transformations
i is initial
f is final
M is species mixing ratios



Numerical model

TNO innovation for life

LOTOS-EUROS

Introduction

Chemical **T**ransport **M**odel (**CTM**)

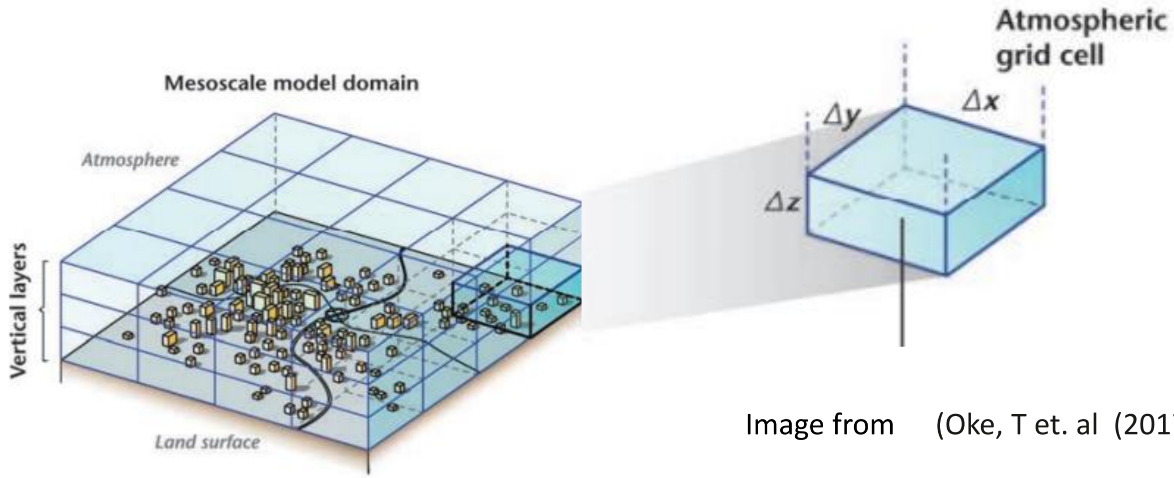


Image from (Oke, T et. al (2017))

$$\frac{\partial C}{\partial t} = -\nabla \cdot (\boldsymbol{u}.\boldsymbol{C}) + \frac{\partial}{\partial \boldsymbol{v}}\left(K_v \frac{\partial C}{\partial \boldsymbol{v}}\right) + E + R + Q - D - W$$

- Change in concentration with time
- Grid resolved transport (Advection)
- Diffusion process
- Entrainment and detrainment
- Generation/Consumption chemical reactions
- Emissions
- Dry and wet deposition proccess

Introduction

$$\frac{\partial C_1}{\partial t} = -\nabla \cdot (\boldsymbol{u}.\boldsymbol{C_1}) + \frac{\partial}{\partial \boldsymbol{v}}\left(K_v \frac{\partial C_1}{\partial \boldsymbol{v}}\right) + E + R + Q - D - W$$

$$\vdots$$

$$\frac{\partial C_n}{\partial t} = -\nabla \cdot (\boldsymbol{u}.\boldsymbol{C_n}) + \frac{\partial}{\partial \boldsymbol{v}}\left(K_v \frac{\partial C_n}{\partial \boldsymbol{v}}\right) + E + R + Q - D - W$$

State space formulation

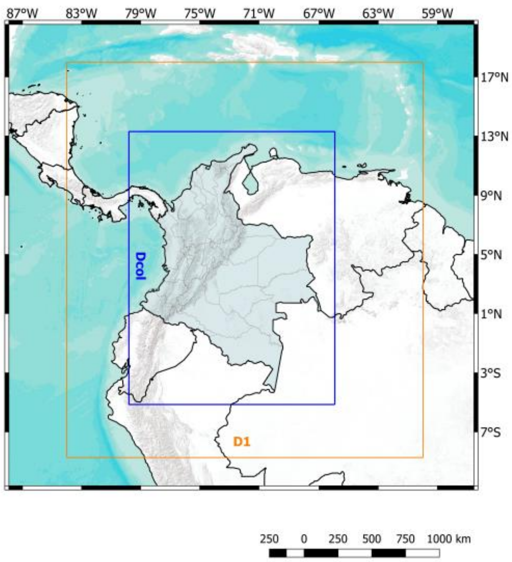
$$\mathbf{X}^{\mathrm{f}}(\mathrm{t}_{\mathrm{i+1}}) = \mathcal{M}[\boldsymbol{X}^f(t_i, \alpha)]$$

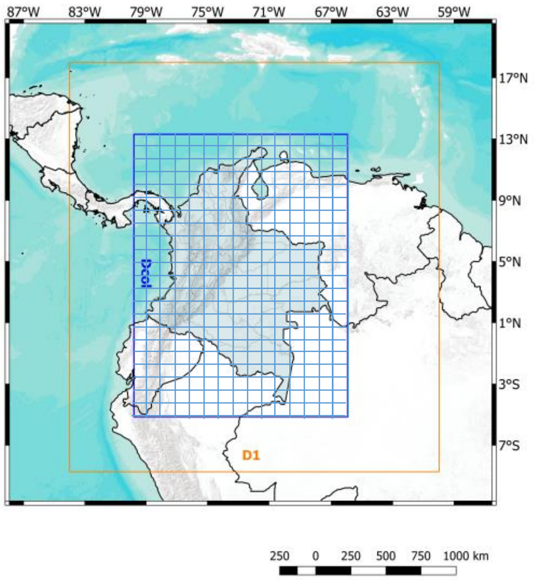
LOTOS-EUROS model

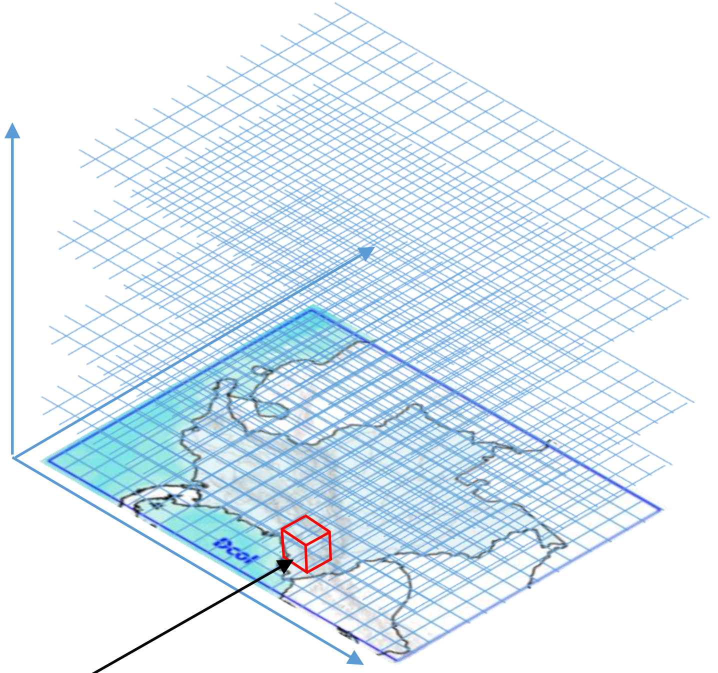
$$\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{n}}$$

$$\mathcal{M}[\boldsymbol{X}^f(t_i, \alpha)] : \mathbb{R}^{\boldsymbol{n}} \Rightarrow \mathbb{R}^{\boldsymbol{n}}$$

Introduction

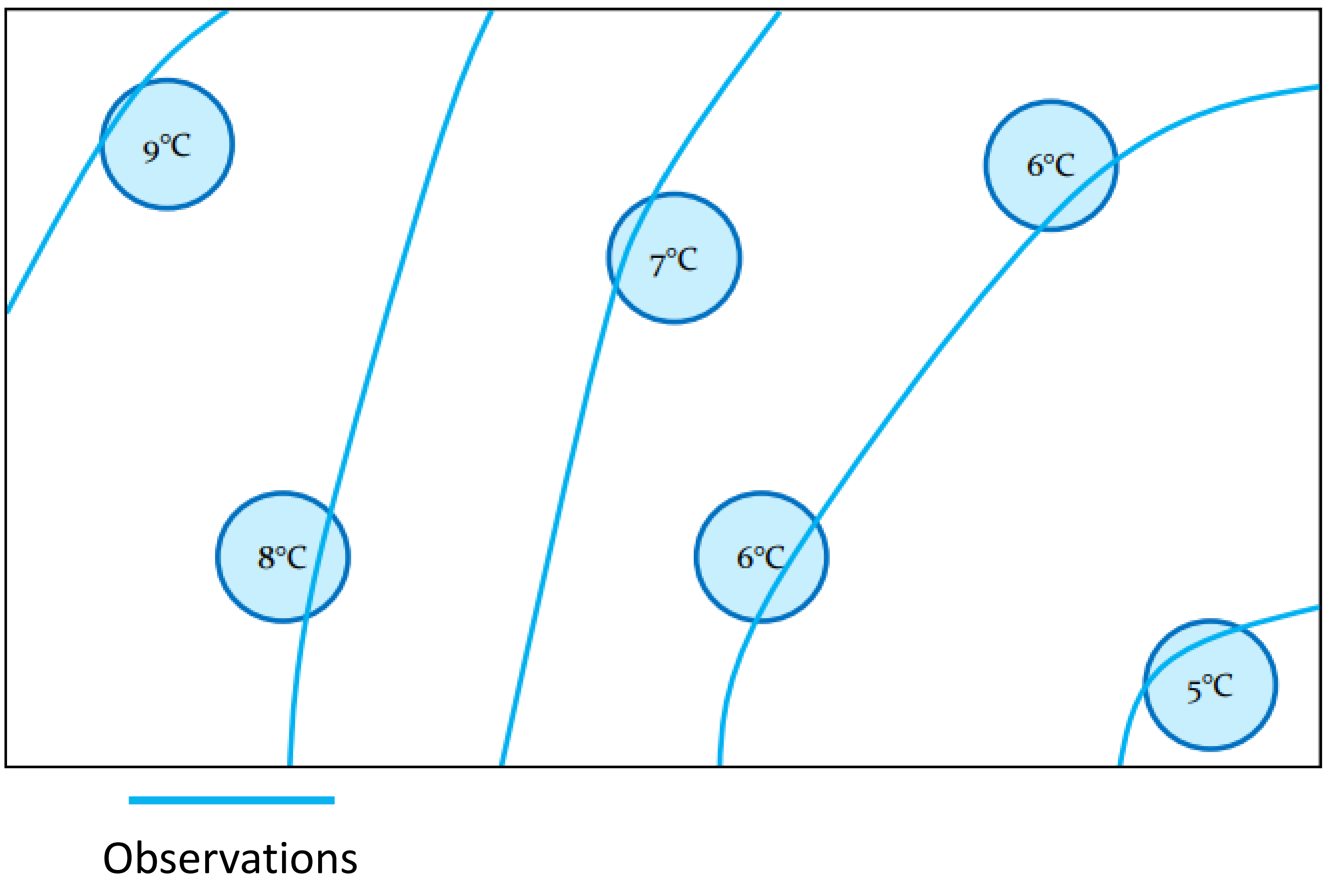




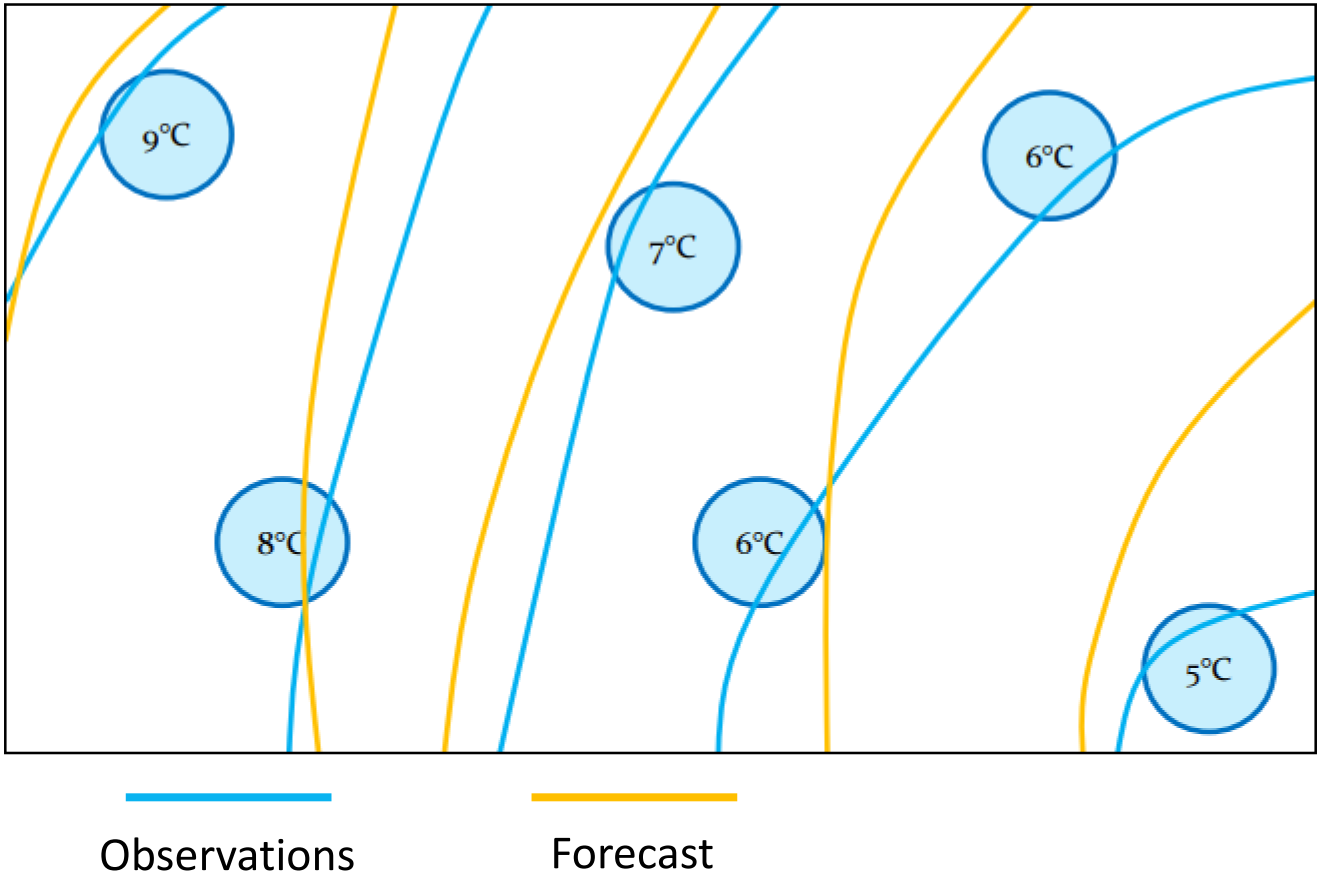


Atmospheric Chemical Transport Model: High-dimensional numerical model $\sim\mathbf{10^6 - 10^8}$states

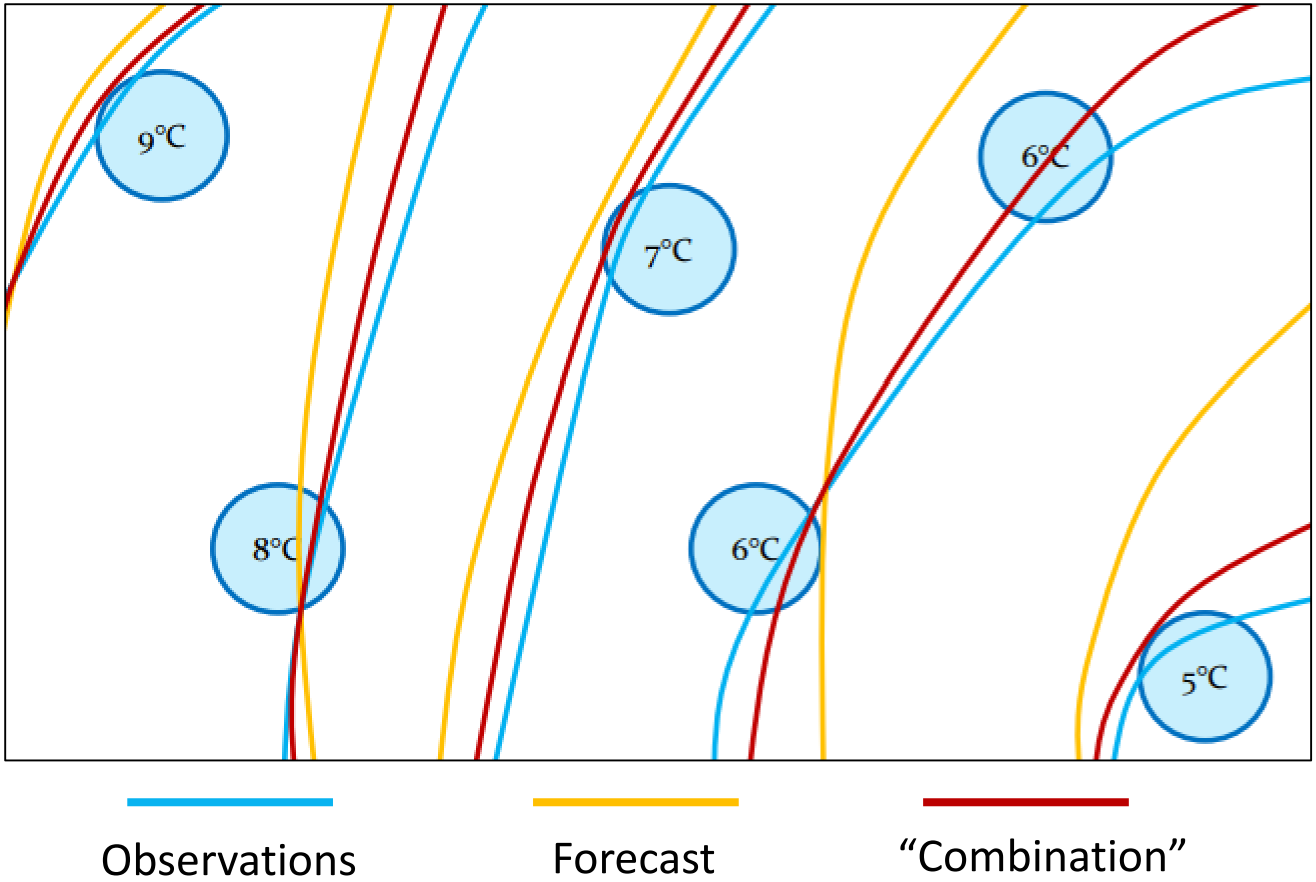
Remember: ¿What is data assimilation?



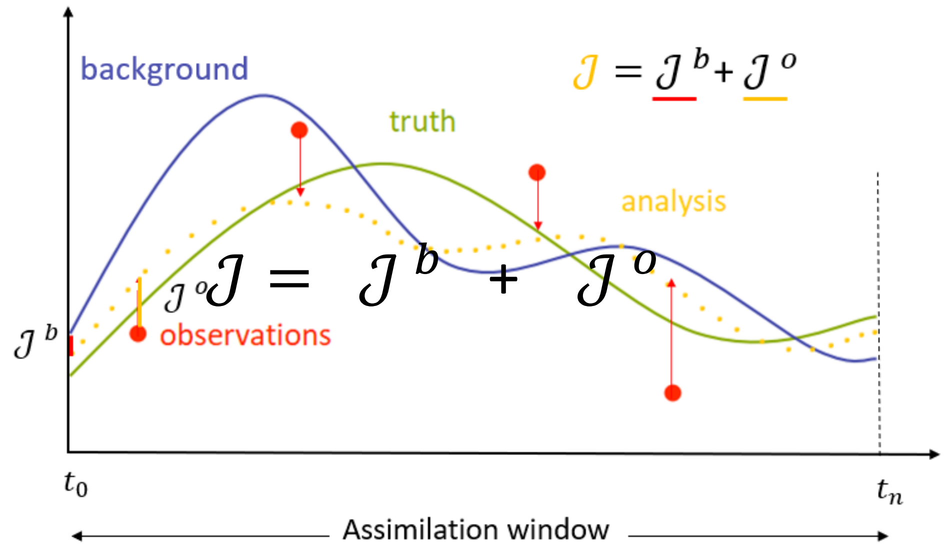
Remember: ¿What is data assimilation?



Remember: ¿What is data assimilation?



Variational Data Assimilation



$$\mathbf{Y}(t_i) = \mathcal{H}\mathbf{X}^f(t_i)$$

$$\mathcal{H}\mathbf{X}^f(t_i) : \mathbb{R}^n \Rightarrow \mathbb{R}^p$$

$$\mathcal{J}(\boldsymbol{x}_0) = \frac{1}{2}\|\boldsymbol{x}_0 - \boldsymbol{x}_b\|_{\boldsymbol{B}}^2 + \frac{1}{2}\|H(\boldsymbol{x}) - y\|_{\boldsymbol{R}}^2$$

3D-Var

Distance to forecast

Distance to observations

$$\mathcal{J}(\boldsymbol{x}_0) = \frac{1}{2}(\boldsymbol{x}_0 - \boldsymbol{x}_b)^T\boldsymbol{B}^{-1}(\boldsymbol{x}_0 - \boldsymbol{x}_b) + \frac{1}{2}(H(\boldsymbol{x}_0) - \boldsymbol{y}_0)^T\boldsymbol{R}^{-1}(\mathcal{H}(\boldsymbol{x}_0) - \boldsymbol{y}_0)$$

Variational Data Assimilation

4D-Var

Strongly constrained

$$\mathcal{J}(\boldsymbol{x}_0) = \frac{1}{2}\|\boldsymbol{x}_0 - \boldsymbol{x}_b\|_{\boldsymbol{B}}^2 + \frac{1}{2}\sum_{i=0}^{s}\|\mathcal{H}M(\boldsymbol{x}) - \boldsymbol{y}_i\|_{\boldsymbol{R}}^2$$

Distance to background — Distance to observations

$$\mathcal{J}(x_o) = \frac{1}{2}\left[(\boldsymbol{x}_0 - \boldsymbol{x}_b)^T\boldsymbol{B}^{-1}(\boldsymbol{x}_0 - \boldsymbol{x}_b) + \sum_{i=0}^{s}(\mathcal{H}(\boldsymbol{x}_i) - \boldsymbol{y}_i)^T\boldsymbol{R}^{-1}(\mathcal{H}(\boldsymbol{x}_i) - \boldsymbol{y}_i)\right]$$

Weakly constrained

$$\mathcal{J}(\boldsymbol{x}_0) = \frac{1}{2}\|\boldsymbol{x}_0 - \boldsymbol{x}_b\|_{\boldsymbol{B}}^2 + \frac{1}{2}\sum_{i=0}^{s}\|\mathcal{H}M_i(\boldsymbol{x}_i) - \boldsymbol{y}_i\|_{\boldsymbol{R}}^2 + \frac{1}{2}\sum_{i=0}^{s}\|\boldsymbol{x} - \boldsymbol{x}_k\|_{\boldsymbol{P}}^2$$

$$\mathcal{J}(x_o) = \frac{1}{2}(\boldsymbol{x}_0 - \boldsymbol{x}_b)^T\boldsymbol{B}^{-1}(\boldsymbol{x}_0 - \boldsymbol{x}_b) + \sum_{i=0}^{s}(\mathcal{H}(\boldsymbol{x}_i) - \boldsymbol{y}_i)^T\boldsymbol{R}^{-1}(\mathcal{H}(\boldsymbol{x}_i) - \boldsymbol{y}_i) + \sum_{i=0}(\boldsymbol{x} - \boldsymbol{x}_k)^T\boldsymbol{P}^{-1}(\boldsymbol{x} - \boldsymbol{x}_k)$$

Variational Data Assimilation

$$x_k = \mathcal{M}\, x_{k-1} \qquad y_k = \mathcal{H} x_s + v_s \qquad v_s \sim N(0, \boldsymbol{R})$$

$$\mathcal{J}(\boldsymbol{x}_0) = \frac{1}{2}(\mathcal{H} x_s - y_s)^T \boldsymbol{R}^{-1} (\mathcal{H} x_s - y_s)$$

$$x_1 = \mathcal{M}\, x_0 \longrightarrow x_2 = \mathcal{M}\, x_1 = \mathcal{M}\mathcal{M} x_1 \;\ldots \longrightarrow x_s = \mathcal{M} x_{s-1} = \mathcal{M}^s\, x_0$$

$$\mathcal{J}(\boldsymbol{x}_0) = \frac{1}{2}(\mathcal{H}\mathcal{M}^s x_0 - y_s)^T \boldsymbol{R}^{-1} (\mathcal{H}\mathcal{M}^s x_0 - y_s)$$

$$\delta\mathcal{J} = -(\mathcal{H}\mathcal{M}^s x_0 - y_0)^T \boldsymbol{R}^{-1} \mathcal{H} \frac{\partial \mathcal{M}^s}{\partial x} \delta\, x_0$$

Variational Data Assimilation

$$\delta \mathcal{J} = \left\langle \mathcal{H}^T \boldsymbol{R}^{-1} (\mathcal{H} \mathcal{M}^s x_0 - y_0)^T, \frac{\partial \mathcal{M}^s}{\partial x} \delta\, x_0 \right\rangle$$

$$\langle x, Ay \rangle = \langle A^T x, t \rangle$$

The adjoint trick

$$\delta \mathcal{J}(\boldsymbol{x}_0) = \left\langle \left[\frac{\partial \mathcal{M}^s}{\partial x} \right]^T \mathcal{H}^T \boldsymbol{R}^{-1} (\mathcal{H} \mathcal{M}^s x_0 - y_0)^T, \delta x_0 \right\rangle$$

$$\delta \mathcal{J}(\boldsymbol{x}_0) = \left\langle \nabla_{\delta(\boldsymbol{x}_0)} \mathcal{J}, \delta \boldsymbol{x}_0 \right\rangle$$

$$\nabla_{\delta(\boldsymbol{x}_0)} \mathcal{J} = \left[\frac{\partial \mathcal{M}^s}{\partial x} \right]^T \mathcal{H}^T \boldsymbol{R}^{-1} (\mathcal{H} \mathcal{M}^s x_0 - y_0)^T$$

Variational Data Assimilation

$$\nabla_{\delta(x_0)}\mathcal{J} = \left[\frac{\partial \mathcal{M}^s}{\partial p}\right]^T \mathcal{H}^T \boldsymbol{R}^{-1}(\mathcal{H}\mathcal{M}^s x_0 - y_0)^T$$

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \frac{\partial f_3}{\partial x_3} & \cdots & \frac{\partial f_3}{\partial x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \frac{\partial f_n}{\partial x_3} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

Low scalability adjoint (Changes in the resolution)

Difficult to mantain if there is **change** in the model

The minimization requires repeated sequential runs of a low resolution **linear model** and its **adjoint**

Atmospheric Chemical Transport Model: High-dimensional numerical model $\sim\mathbf{10^6 - 10^8}$states

¿Can we reduce the cost of data assimilation in the context of atmospheric chemical transport simulation without degrading the results?

Model Order Reduction

By a **reduction** of the model state space dimension, an **approximation** to the original model is computed which is commonly referred to as **reduced order model**

Model Order Reduction

Vector and matrix sizes

$$\mathbf{X}^{\mathrm{f}}(\mathrm{t}_{\mathrm{i}+1}) = \mathcal{M}[\mathbf{X}^{f}(t_i, \alpha)]$$

$$\mathbf{Y}(t_i) = \mathcal{H}\mathbf{X}^{f}(t_i)$$

$[\mathrm{X}]=n$ $[B]=n \times n$ $[M]=\mathrm{n} \times n$

$[\mathrm{Y}]=m$ $[\mathcal{H}]=\mathrm{m}\times n$

For some applications, n and m are large $10^6 - 10^8 \Longrightarrow$ impossible to **store/compute/multiply/inverse** data assimilation matrices B and H and M

Possible solution: rank reduction method

Model Order Reduction

Rank reduction (Square root decomposition**)**

A symmetric definite matrix B can be decomposed into SS^T where S is a $n \times n$

choosing only a small number r of significative columns → $\mathbf{S}_r$ with size $n \times r$

Set $B_r = S_r S_r^T$ with $B_r \approx B$

Model Order Reduction

Singular **V**alue **D**ecomposition (**SVD**)

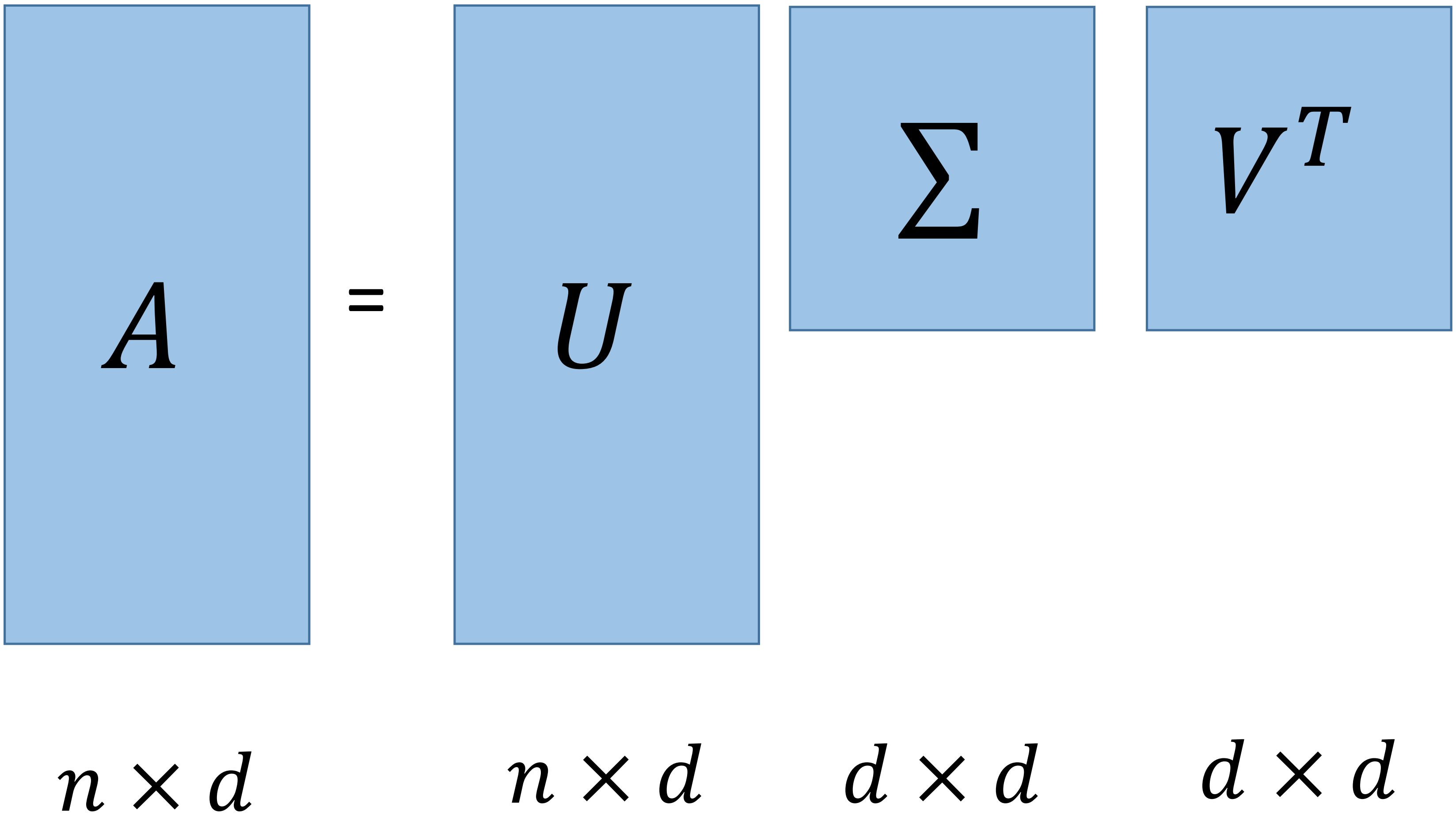
For any $m \times \mathrm{n}$ matrix A we can factor it into:

$$A = U \sum V^T$$

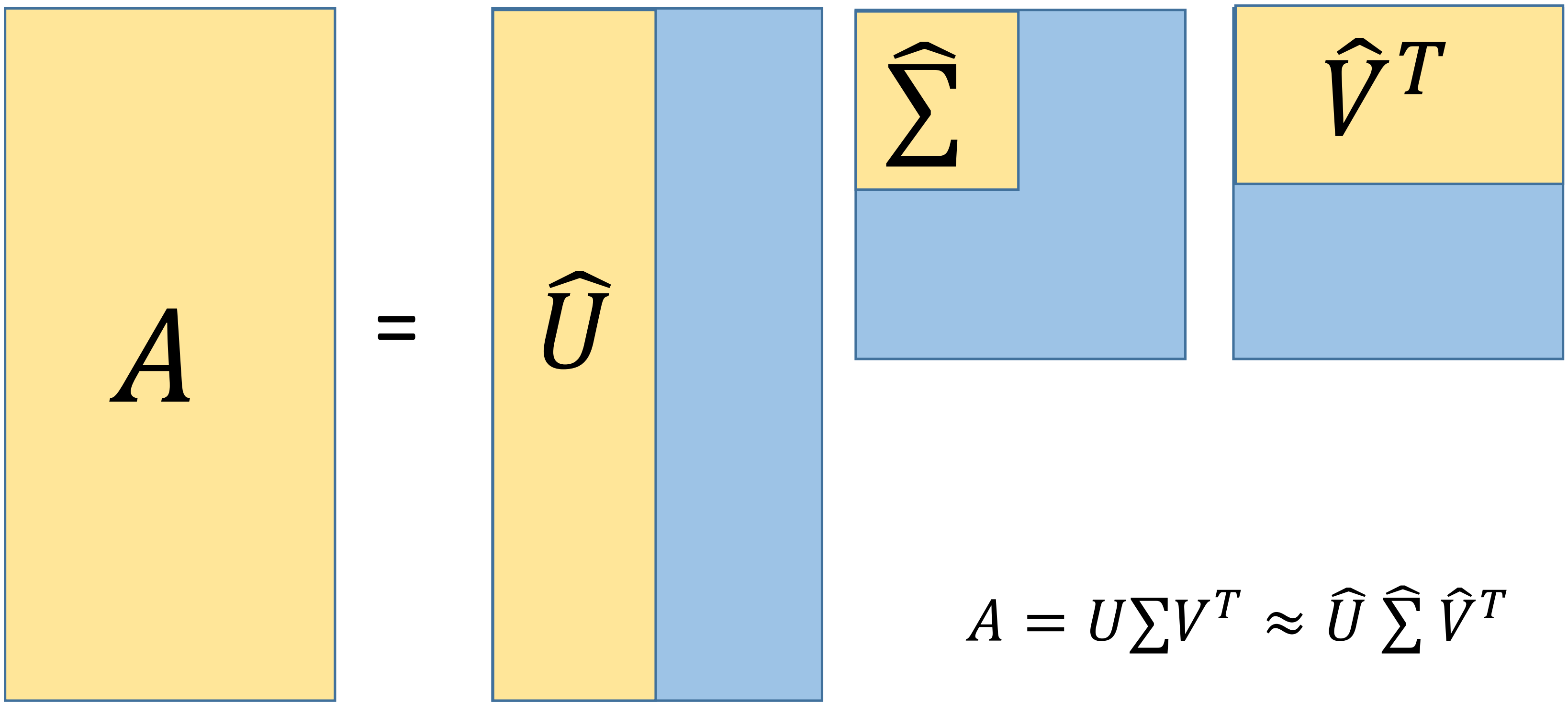
$U = m \times n$ ortogonal matrix
$V = n \times n$ ortogonal matrix
$\sum = m \times n$ diagonal matrix with $\sum_{\{i,j\}} = \sigma_i \geq 0$

$\sigma_i' s$ are ordered $\sigma_i \geq \sigma_{\{i+1\}}$ for $i = 1 \dots n$

Model Order Reduction



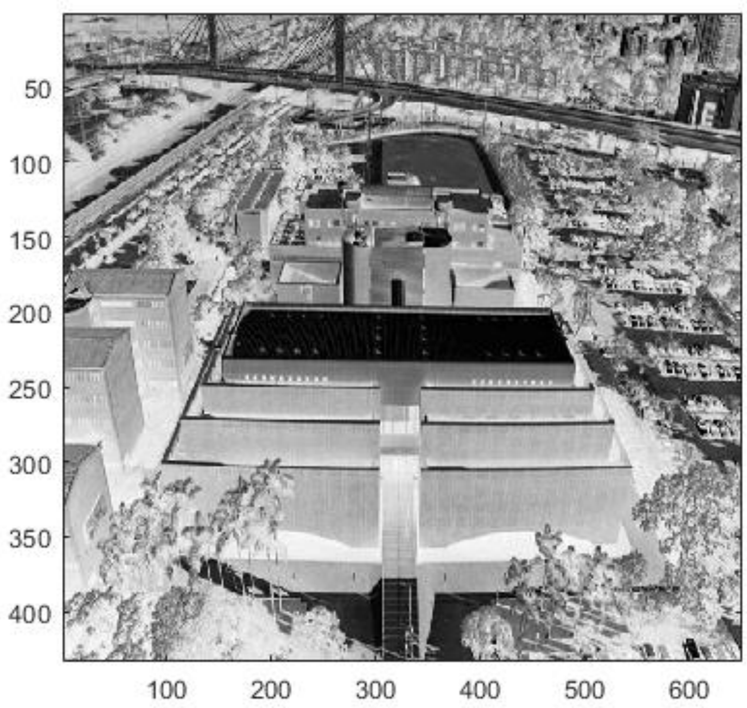
Model Order Reduction

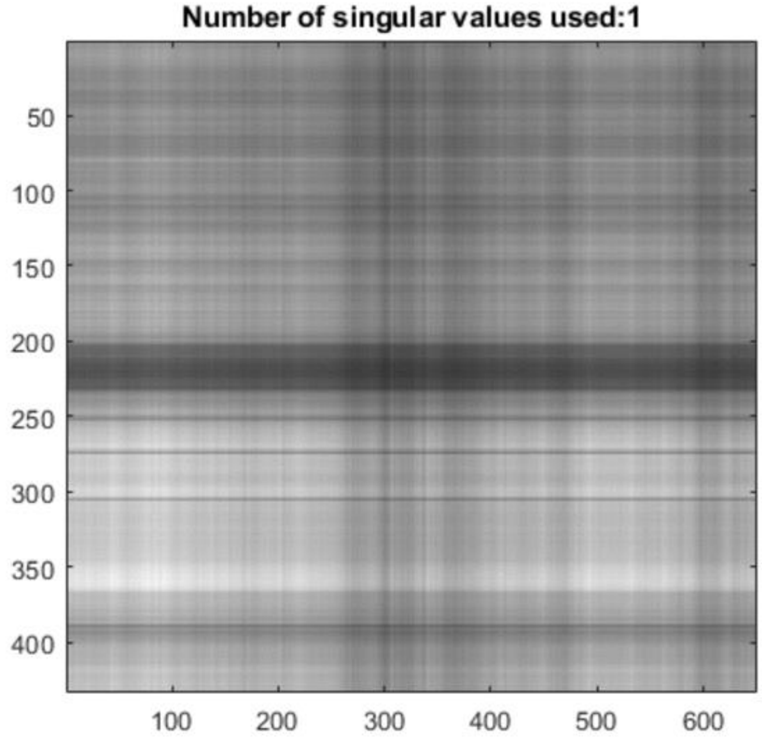


$$A = U\sum V^T \approx \hat{U}\,\hat{\sum}\,\hat{V}^T$$

$n \times d$ $n \times r$ $r \times r$ $r \times d$

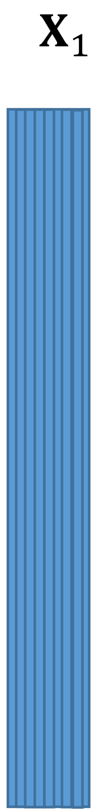
Model Order Reduction



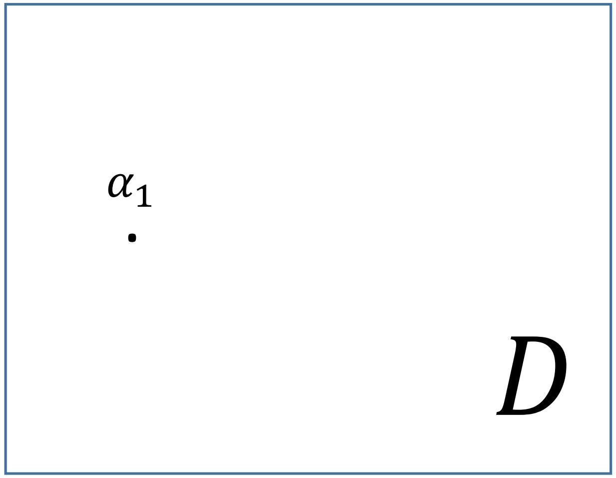


Model Order Reduction

$$\alpha \in D$$



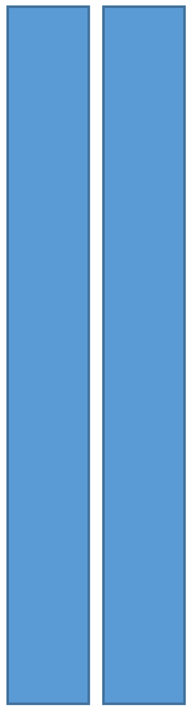
Snapshots of the forward model

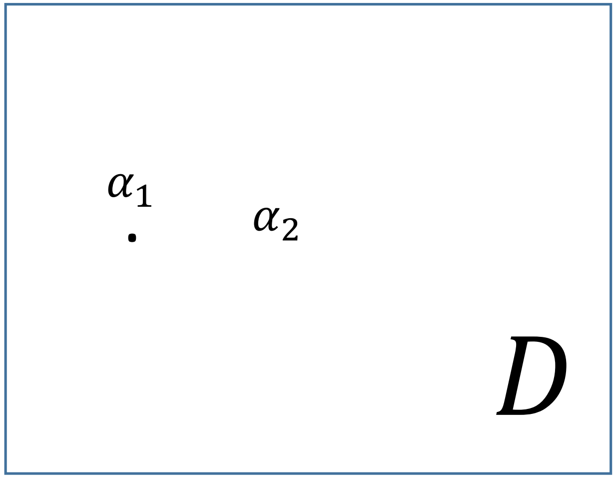


Model Order Reduction

$\alpha \in D$

$\mathbf{X}_1 \quad \mathbf{X}_2$



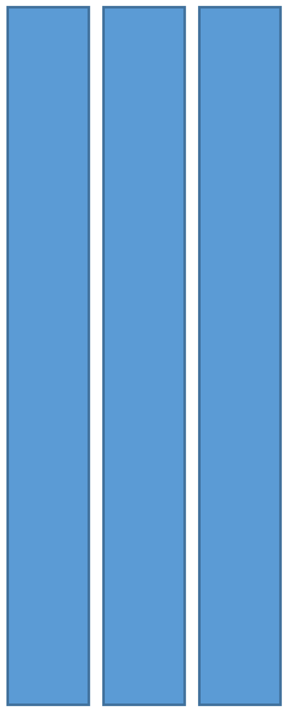


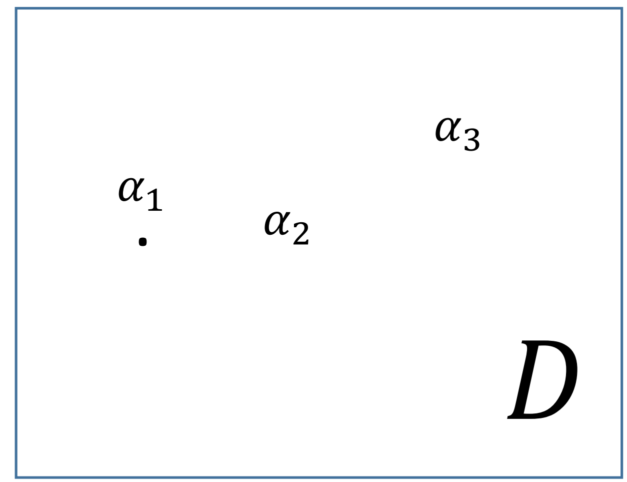
Snapshots of the forward model

Model Order Reduction

$$\alpha \in D$$

$\mathbf{X}_1$ $\mathbf{X}_2$ $\mathbf{X}_3$





$$S = [\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$$

Snapshots of the forward model

Collection of snapshots of the state

Model Order Reduction

Usually **SVD** is calculated for $\boldsymbol{S} = [\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$ or also the following eigenvector problem is solved

$$\boldsymbol{S}^T \boldsymbol{S} \boldsymbol{V}_i = \lambda_\mathrm{i} \mathbf{V}_\mathrm{i}$$

Where $\boldsymbol{V}_i$ and λ_i are the ith eigenvector and eigenvalue respectively and the corresponding basis $\boldsymbol{P}$ can be obtained applying:

$$P_i = SV\lambda_i^{1/2}$$

$$\widehat{\overline{X}}(t_i) = X^b(t_i) + PR(t_i)$$

(Vermeulen P, et al. (2005).)

¿How can avoid or minimize the problem of not having an adjoint model for a highly non linear, large scale model?

Incremental 4D var

- **Incremental DA** (Courtier, 1994) : deals with **perturbation** made to known reference states

Is based in the preconditioning technique of the matrix $\mathbf{B}$

$$\mathbf{B} = \mathbf{U}\mathbf{U}^{\mathbf{T}}$$

Where $\mathbf{U}$ is knowing as the preconditioning matrix

$$\boldsymbol{x_a} = \boldsymbol{x_b} + \boldsymbol{Uw}$$

$$\boldsymbol{w} = \delta\boldsymbol{x}$$

Incremental 4D var

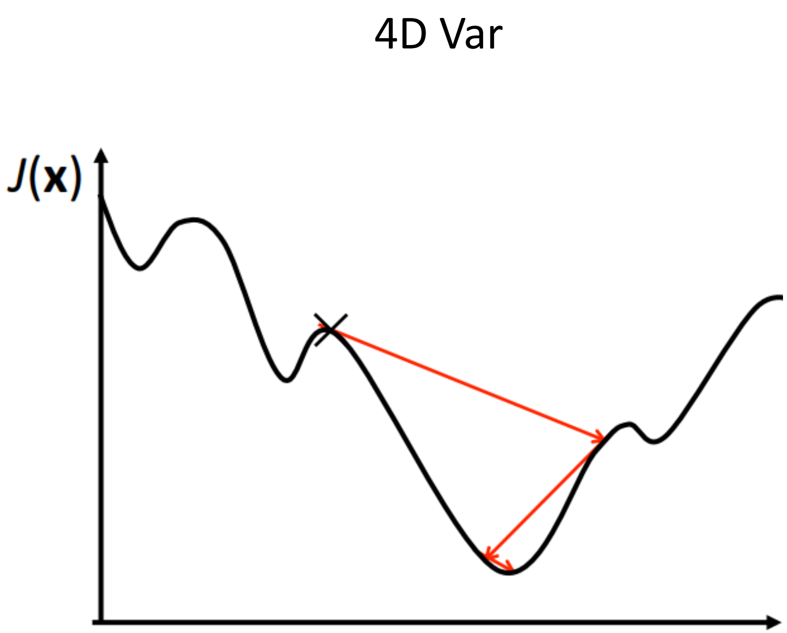
The cost function in control variable space becomes

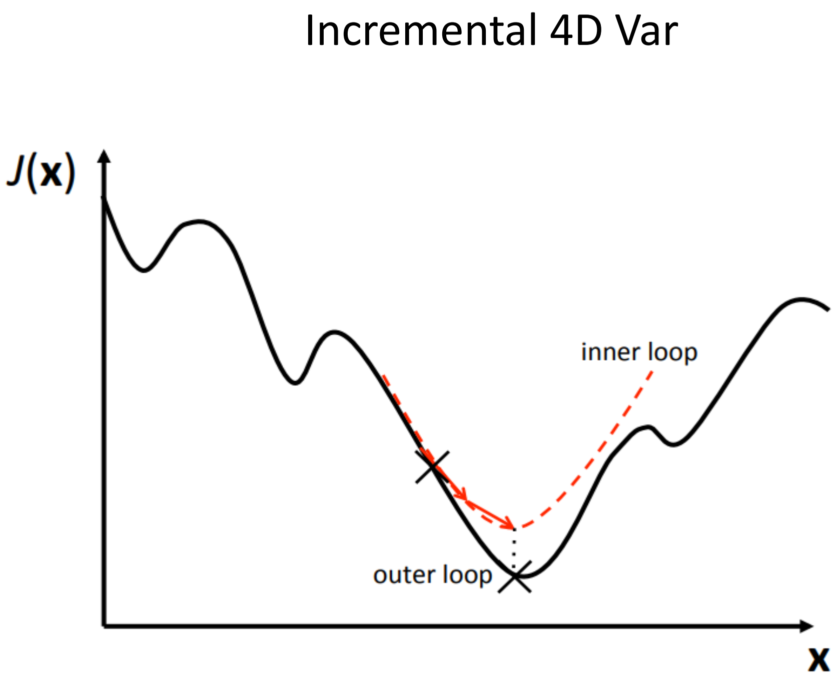
$$J(w) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{1}{2}\sum_{i=0}^{S}(\boldsymbol{HMUw} + \boldsymbol{d}_i)^T\,\boldsymbol{R}^{-1}(\boldsymbol{HMUw} + \boldsymbol{d}_i)$$

$$\nabla_{\boldsymbol{w}}\mathcal{J} = \boldsymbol{w} + \sum_{i=0}^{S}\mathbf{U}^T\,\mathbf{M}^T\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{HMUw} + \mathbf{d}_i)$$

$$\boldsymbol{d_i} = \boldsymbol{HM}(\boldsymbol{x_b}) - \boldsymbol{y_i}$$

Incremental 4D var





Localized minimization procedure

Adjoint-free data assimilation techniques

4D EnVar

Simple idea → linear combination of ensemble members

One ensemble member vector

$$\mathbf{X_b}' = \frac{1}{\sqrt{N-1}}\left(\underbrace{\mathbf{x}_{b1} - \overline{\mathbf{x}_b}}, \mathbf{x}_{b2} - \overline{\mathbf{x}_b}, \ldots, \mathbf{x}_{bN} - \overline{\mathbf{x}_b}\right)$$

N ensemble number

$\mathbf{x}$ state vector

b denotes background

Adjoint-free data assimilation techniques

4D EnVar

The background error covariance calculated approximately as

$$\mathbf{B} \approx \mathbf{X}_b' \mathbf{X}_b'^T$$

From the analisys step in the EnKF we had

$$\boldsymbol{x_a} = \boldsymbol{x_b} + \boldsymbol{BH^T}\left(\boldsymbol{HBH^T} + \boldsymbol{R}\right)^{-1}(\boldsymbol{y} - H\boldsymbol{x_b})$$

$$\mathbf{BH}^T \approx \mathbf{X}_b' \mathbf{X}_b'^T \mathbf{H}^t = \mathbf{X}_b' (\mathbf{HX}_b')^T$$

$$\mathbf{HBH}^T \approx \mathbf{HX}_b' (\mathbf{HX}_b')^T$$

$$\mathbf{HX}_b' = \frac{1}{\sqrt{N-1}}(H\boldsymbol{x}_{b1} - H\overline{\boldsymbol{x}_b}, H\boldsymbol{x}_{b2} - H\overline{\boldsymbol{x}_b}, \dots, H\boldsymbol{x}_{bN} - H\overline{\boldsymbol{x}_b})$$

Adjoint-free data assimilation techniques

$$\mathcal{J}(w) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{1}{2}\sum_{i=0}^{s}(\mathbf{HMX_b}\boldsymbol{w} + \boldsymbol{d}_i)^T \boldsymbol{R}^{-1}(\mathbf{HMX_b}\boldsymbol{w} + \boldsymbol{d}_i)$$

$$\nabla_{\boldsymbol{w}}\mathcal{J} = \boldsymbol{w} + \sum_{i=0}^{s}(\mathbf{HMX'_b})^T\mathbf{R}^{-1}(\mathbf{HMX'_b}\mathbf{w} + \mathbf{d}_i)$$

$$\mathbf{HMX}'_b \approx \frac{1}{\sqrt{N-1}}(HM\boldsymbol{x}_{b1} - HM\overline{\boldsymbol{x}_b}, HM\boldsymbol{x}_{b2} - HM\overline{\boldsymbol{x}_b}, \dots, HM\boldsymbol{x}_{bN} - HM\overline{\boldsymbol{x}_b})$$

NO ADJOINT MODEL NEEDED

(Liu C, et al. (2007))

Thanks

References

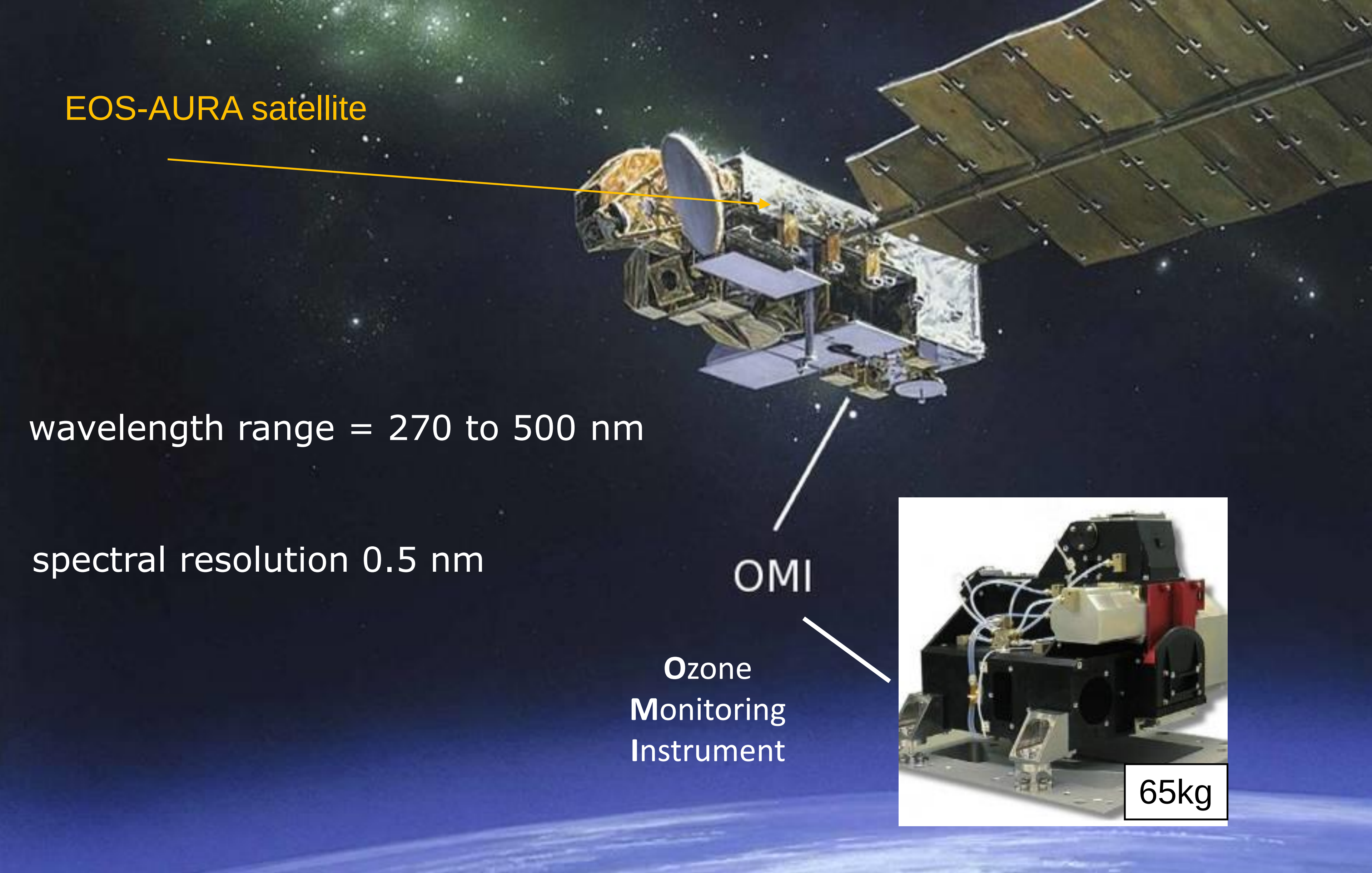
Bannister. (2017). Review article. A review operational methods of variational and ensemble-variational data assimilation. Journal Royal Meteorological Society.

Courtier P, et al. (1994). A strategy for operational implementation of 4DVar using an incremental approach. Journal Royal Meteorological Society.

Liu C, et al. (2007). An Ensemble based four-dimensional Variational Data Assimilation Scheme. Part I: Technical formulation and preliminary test. American Meteorological Society.

Oke, T., Mills, G., Christen, A., & Voogt, J. (2017). Air Pollution. In *Urban Climates* (pp. 294-331). Cambridge: Cambridge University Press. doi:10.1017/9781139016476.012

Vermeulen P, et al. (2005). Model-reduced variational data assimilation. Monthly weather review.

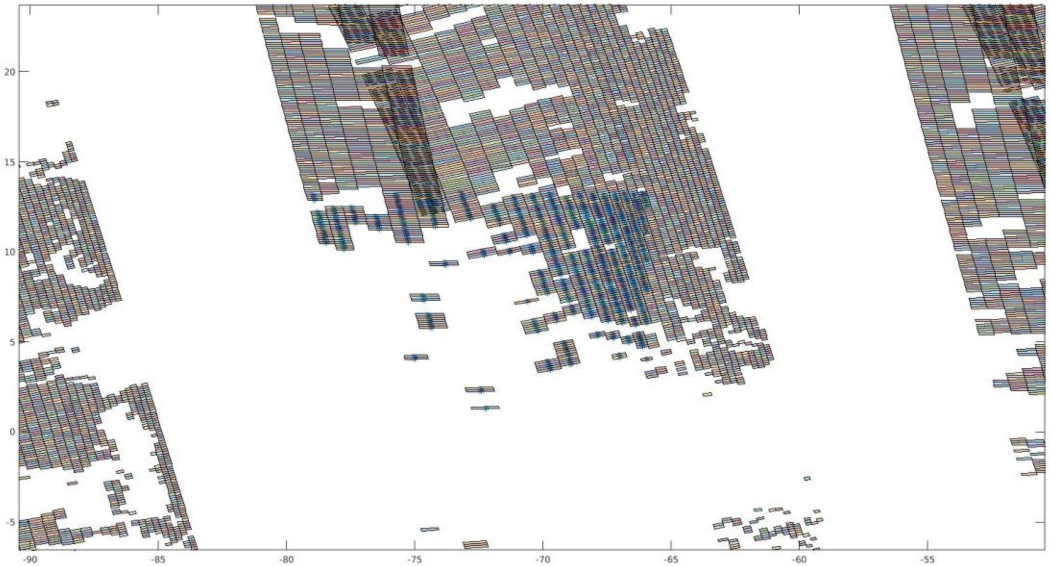


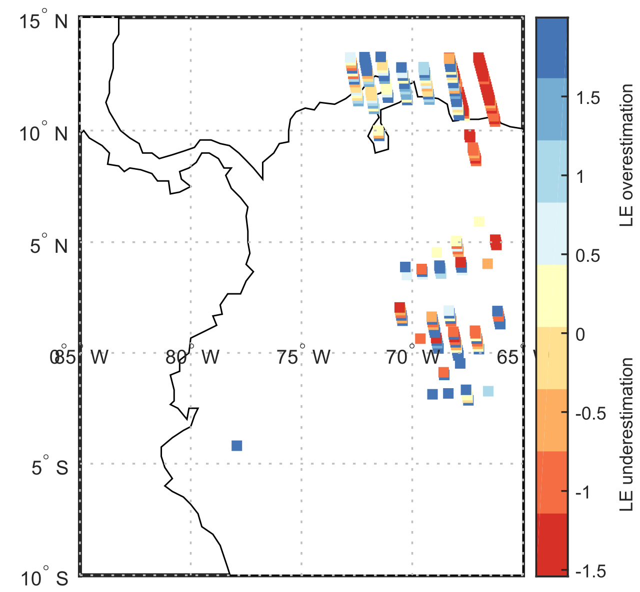
wavelength range = 270 to 500 nm

spectral resolution 0.5 nm

Satellite data acquisition and preprocessing

Crop and Apply Kernel transformation





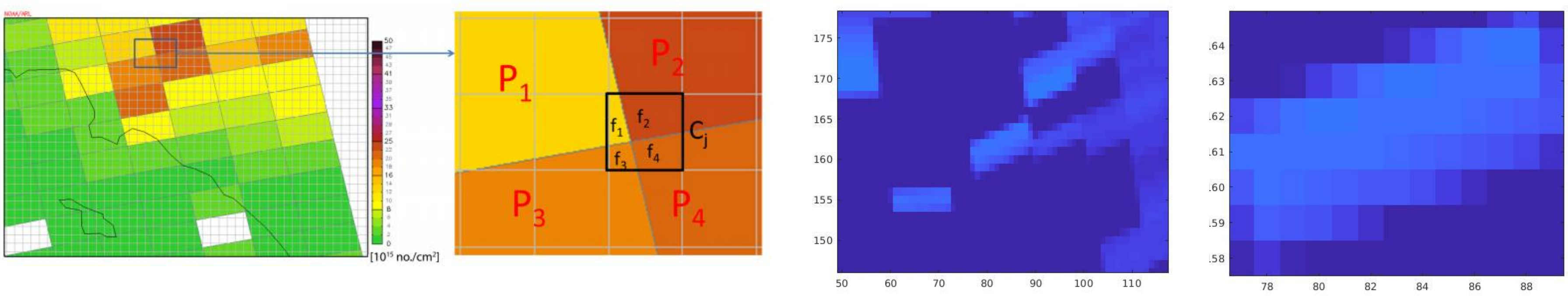
Initial comparison with the model

Procedure to apply the sensitivity kernel transformation to OMI NO_2 concentration

Satellite data acquisition and preprocessing

Concentration mapped
from OMI to LE grid

$$f_{\{i,j\}} = \frac{\mathrm{Area}(\mathrm{P_i} \cap C_j)}{Area(C_j)} \longrightarrow \sum_{i=1}^{4} f_{\{i,j\}} * ConcP_{\{i\}}$$

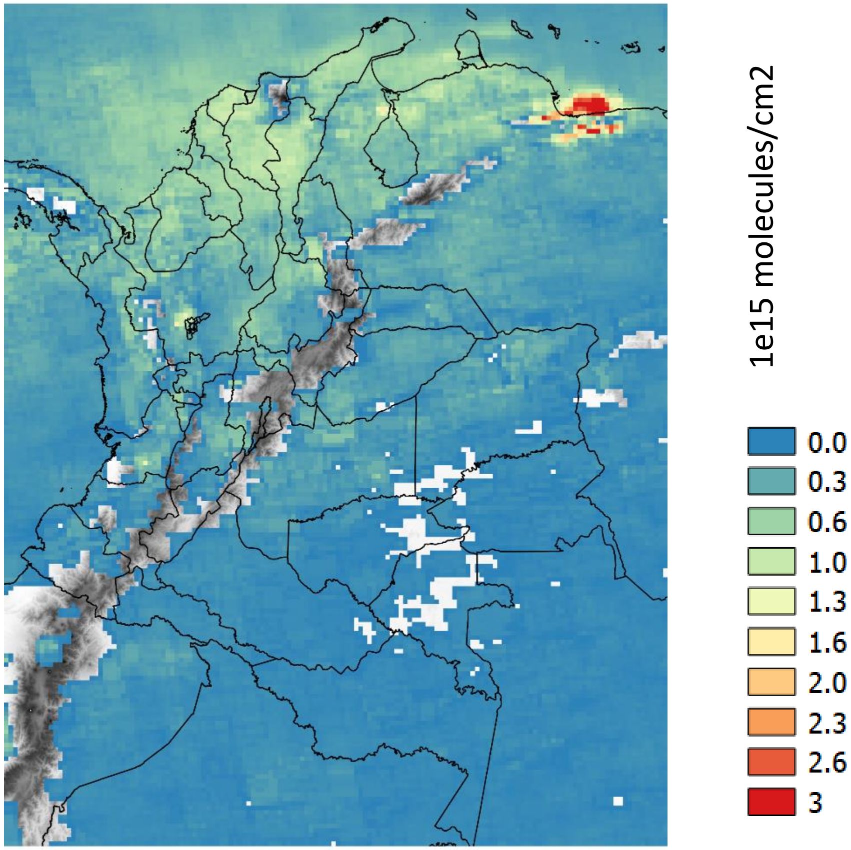


Methodology to map the traces of the satellite information to the LE model GRID

(Kim et al. 2016)

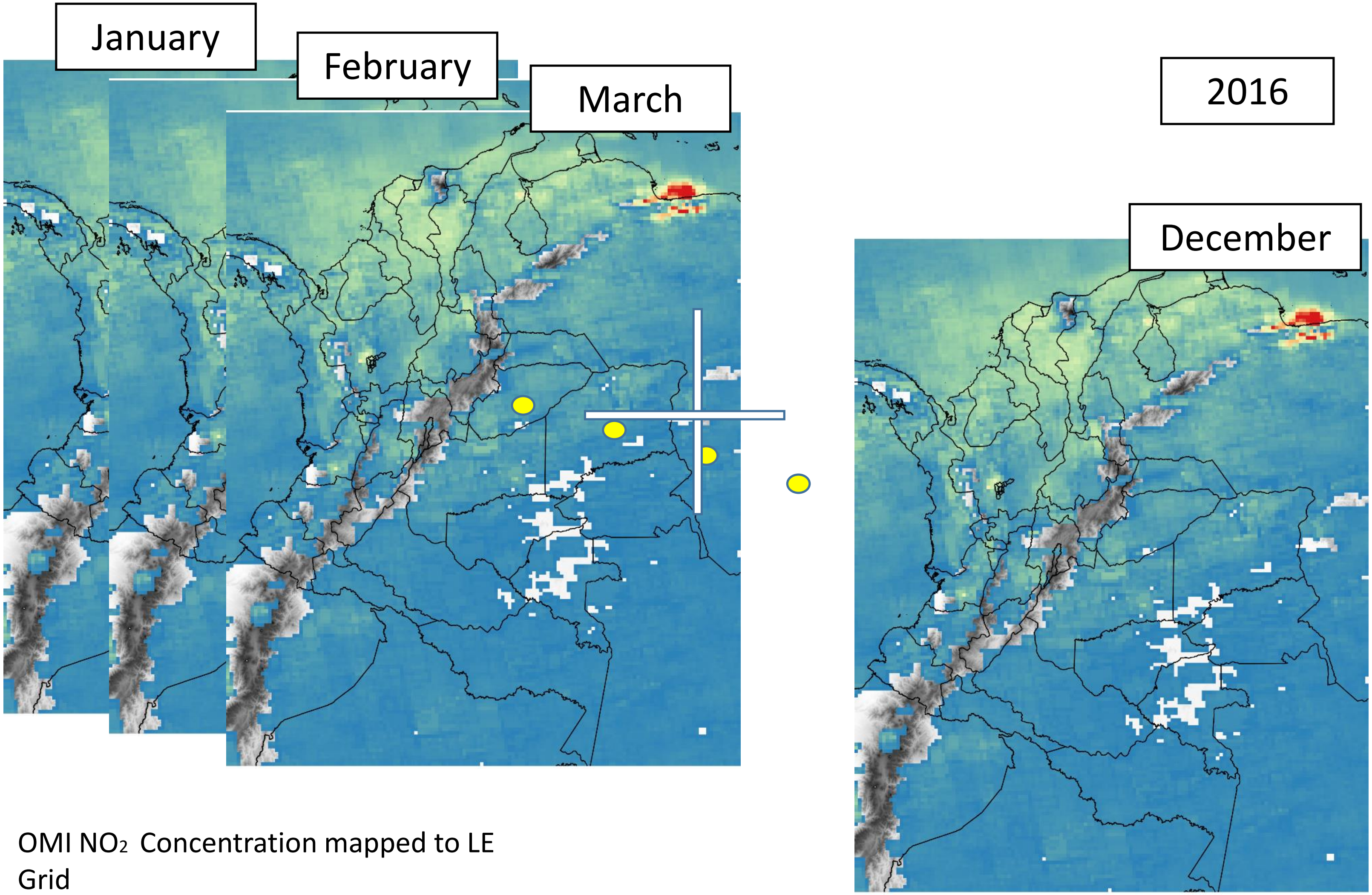
Satellite data acquisition and preprocessing

January



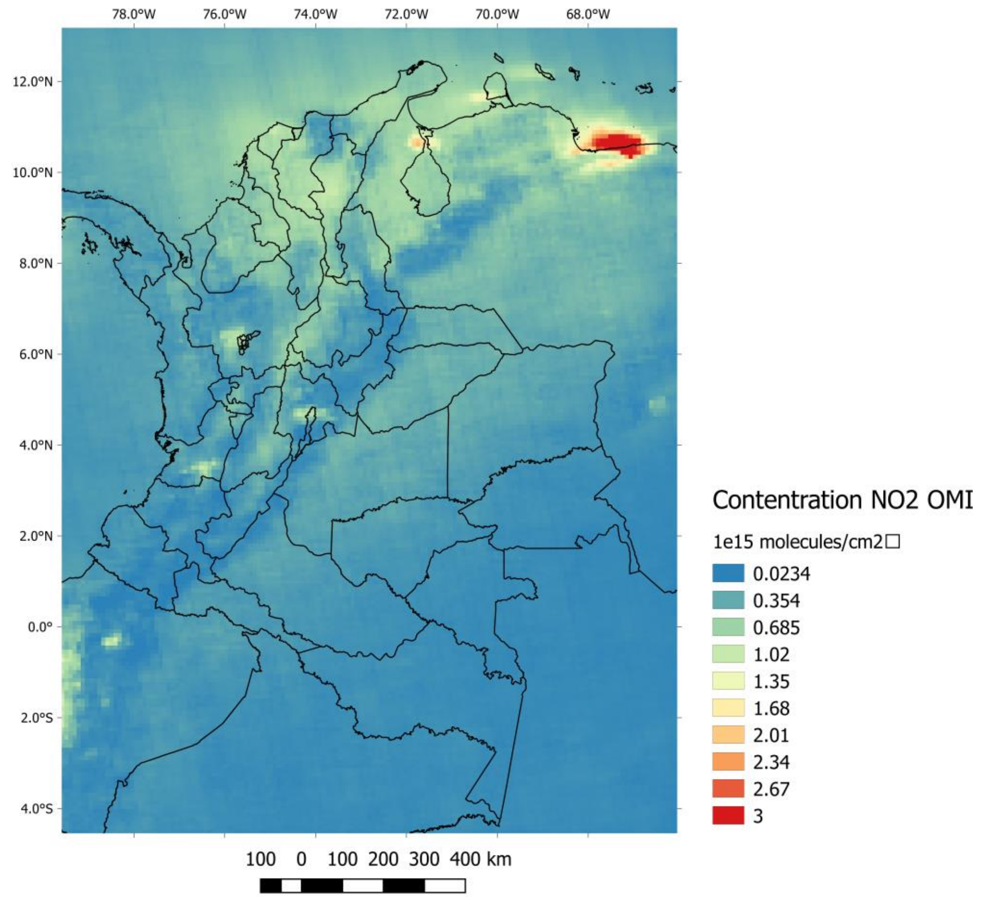
OMI NO_2 Concentration mapped to LE Grid

Satellite data acquisition and preprocessing



OMI NO_2 Concentration mapped to LE Grid

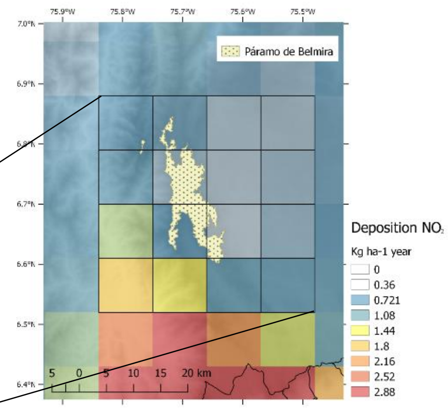
Satellite data acquisition and preprocessing

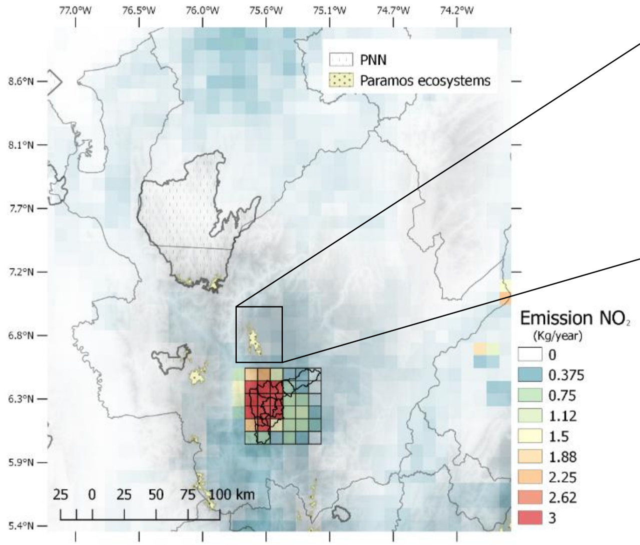


Final product OMI to LE grid (complete 2016)

Future

EFFECTS





CAUSES