Environnement et Changement climatique Canada
Environment and Climate Change Canada

Mathematical Approaches of Atmospheric Constituents Data Assimilation and Inverse Modeling
Session 7 – Parametric Kalman filtering (TCPL 201)

# PvKF Assimilation of GOSAT Methane in the Hemispheric CMAQ: Design and Results using Optimal Error Statistics with an Application for Emissions Inversion


Presenter: Sina Voshtani[1]

Co-authors: Richard Ménard[2], Thomas W. Walker[3], Amir Hakami[3]

[1] Department of Physics, University of Toronto
[2] Air Quality Research Division, Environment and Climate Change Canada
[3] Department of Environmental Engineering, Carleton University


BIRS, Banff, 23 March 2023

# Outline:

## Background

- Data Assimilation and Inverse Modelling
- Importance of Atmospheric Methane

## Motivation

- Gaps and Limitations
- Questions

## Tools

- Model and Observations
- Covariance Modeling and Parameter Estimation

## Project

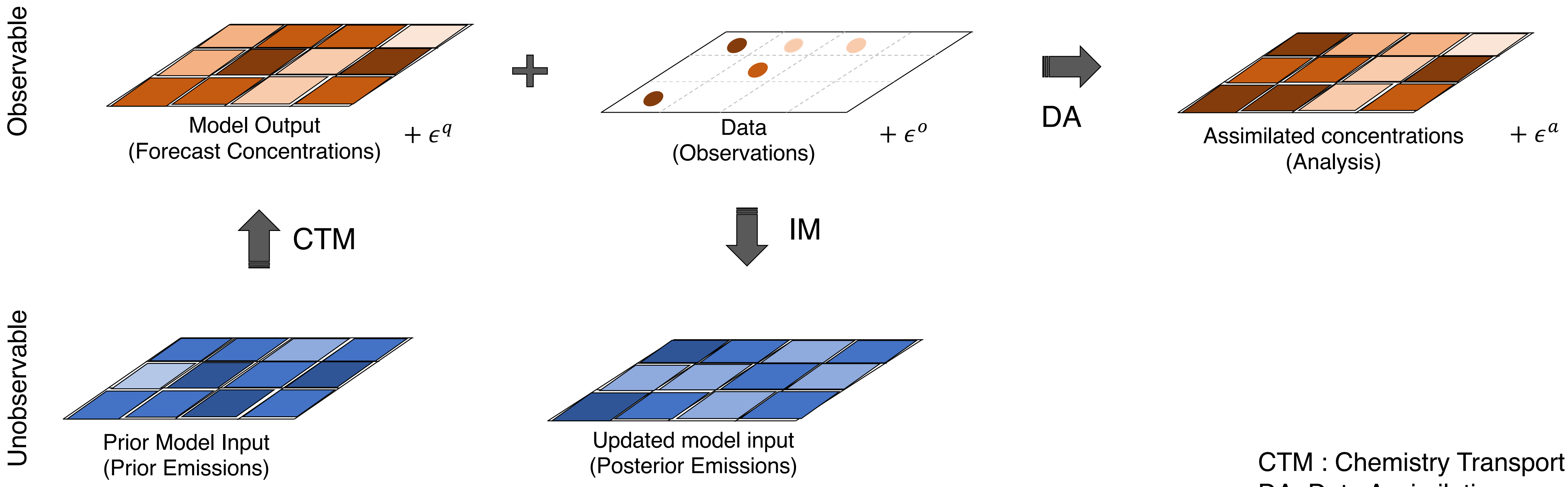
- Part I: **Design of an Assimilation System for Methane**
- Part II: **Assimilation Results with Optimal Error Statistics**
- Part III: **Assimilation Use in Methane Emissions Inversion**

## Future Work

# Data Assimilation and Inverse Modelling

are statistical frameworks to:

- Obtain consistent, precise, and evolving 3-dimensional picture of the atmosphere
- Fill in data gaps and inferring information about unobserved variables

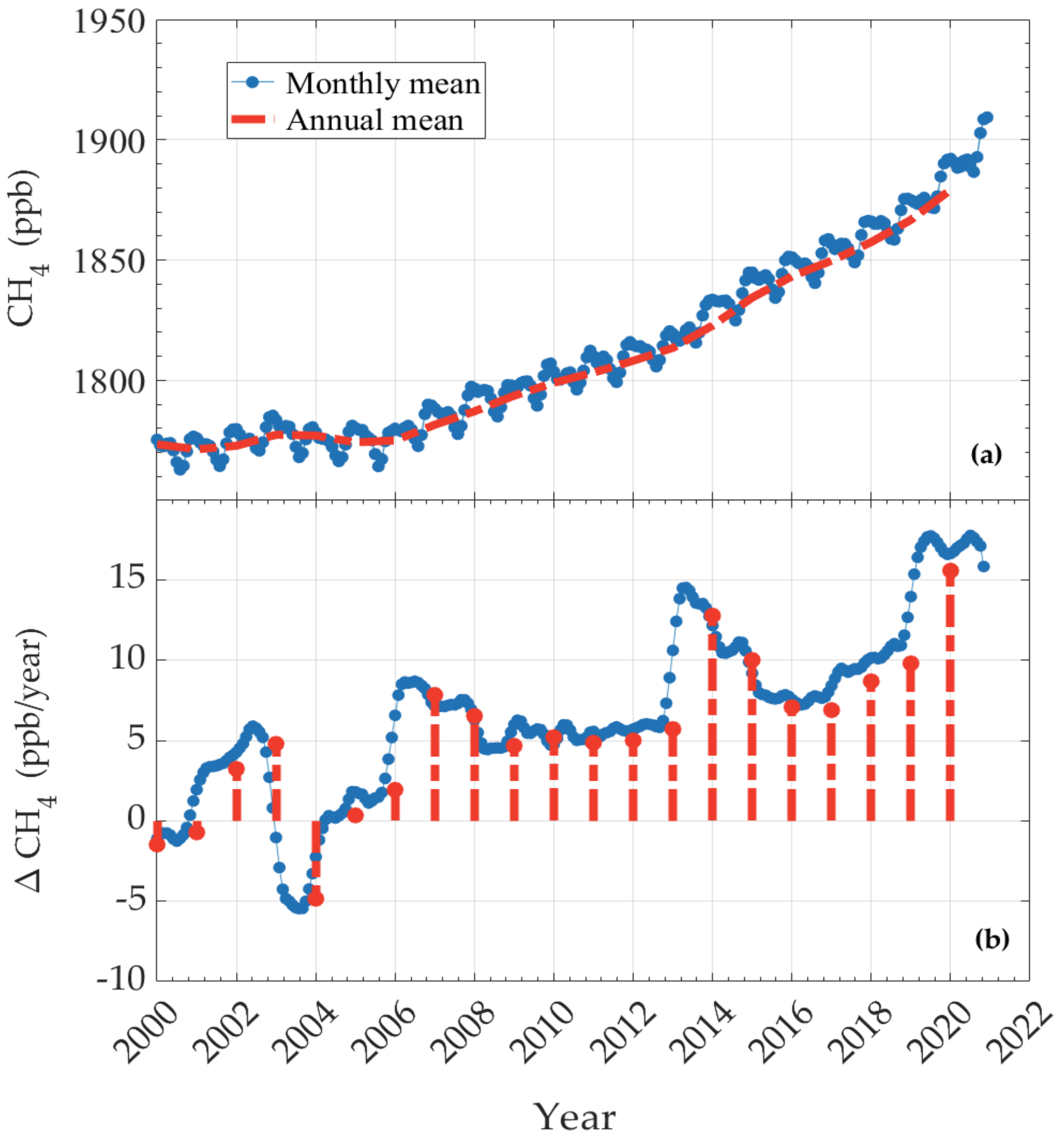



CTM : Chemistry Transport Model
DA: Data Assimilation
IM: Inverse Modelling (or Inversion)

# Why Atmospheric Methane?

Because of the large **climate** and **air quality** impact

- Largest anthropogenic radiative forcing after $CO_2$
- Short lifetime and ~30 times greater GWP than $CO_2$
- Outsized influence on near-term climate change
- Large air quality impact, (e.g., $O_3$ production)
- Global average concentration acceleration after 2007




GWP: Global Warming Potential

Data obtained from Dlugokencky (2022)

# Gaps and Limitations in the Past Methane Studies

## 1. Challenges in Emissions Inversion

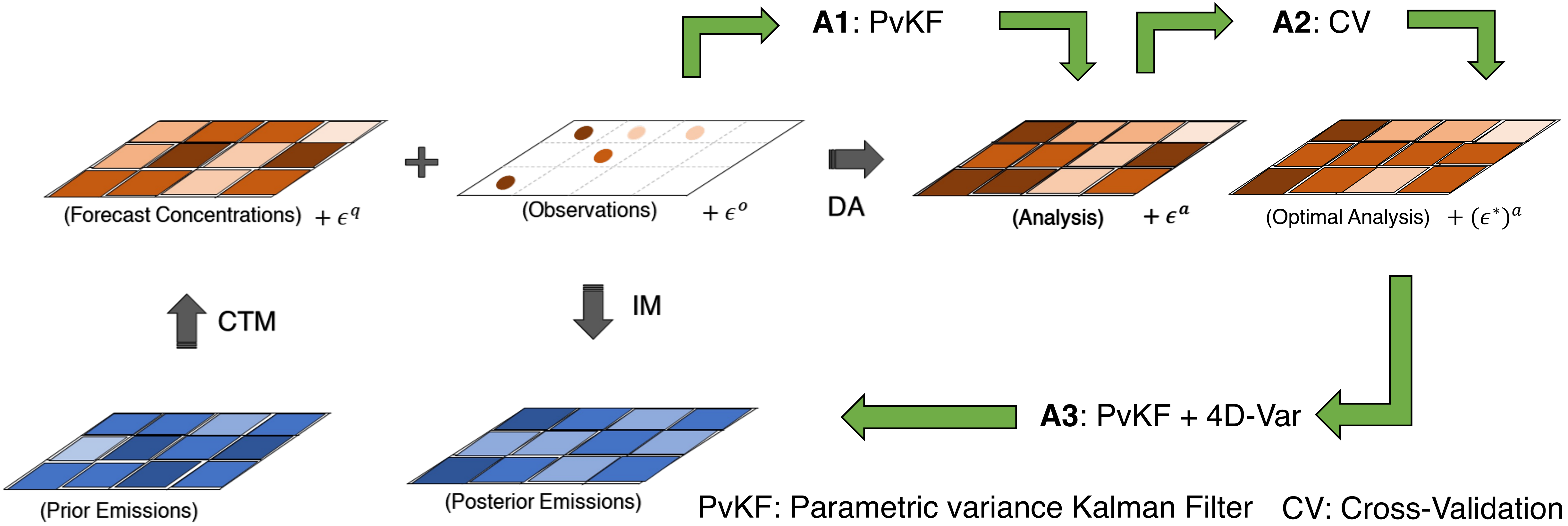
- Scale, temporal, and spatial resolution (Turner et al., 2015; Zavala-Araiza et al., 2017)
- Initial and boundary conditions (Bousserez et al. 2016; Bergamaschi et al. 2018)
- Contradiction in the result of different inversions (Ganesan et al., 2019; Miller et al., 2019)
- High computations to estimate the state errors (Yu et al., 2021; Voshtani et al., 2022a)

## 2. Limitations in Estimation Problem

- Perfect model assumptions (Janardanan et al., 2020; Zhang et al., 2021)
- Error statistics are already optimal (Voshtani et al., 2022b)
- Separate evaluations on the error statistics (Voshtani et al., 2022b)
- Concentration uncertainties and error correlations in the observation space (Voshtani et al., 2023; in review)

# Research Questions?

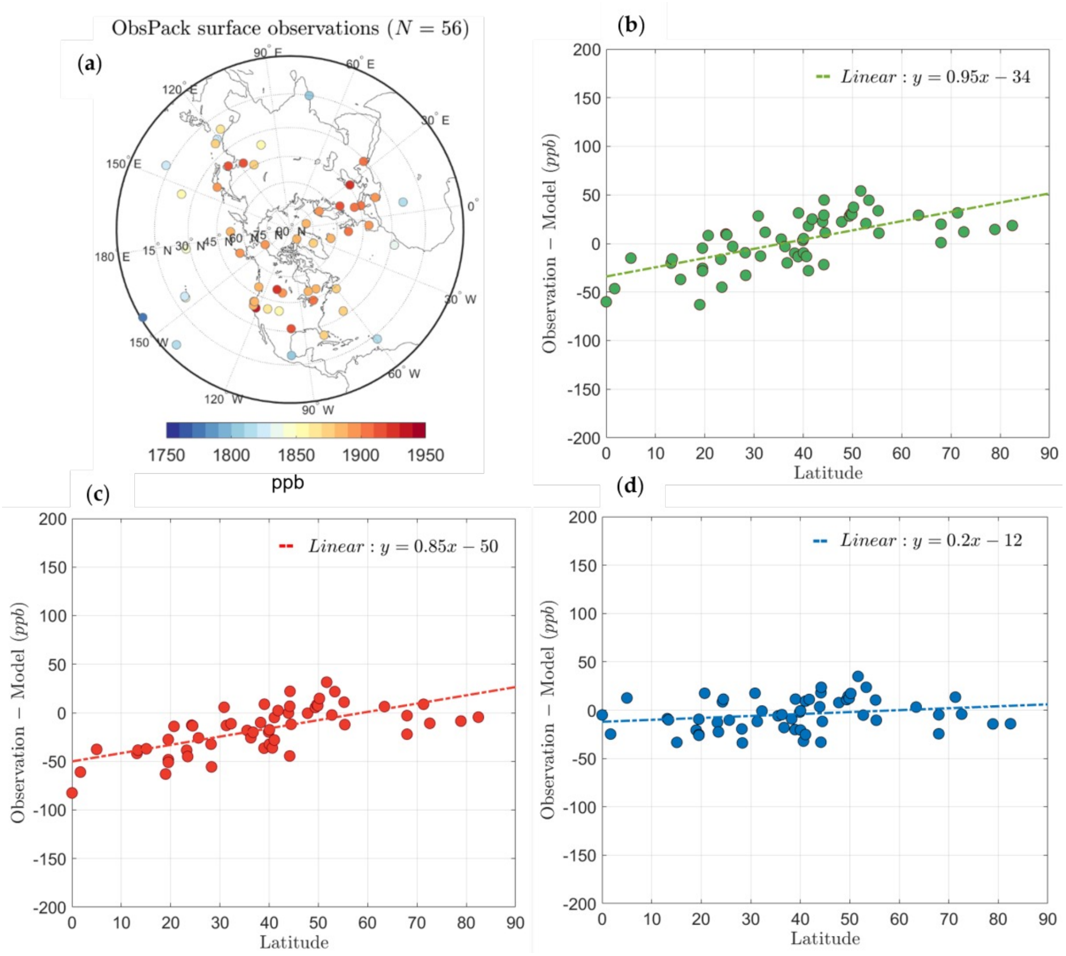
**Q1**: How to obtain a low-cost yet powerful DA system, capable of estimating uncertainties?
**Q2**: What is the impact of optimal error statistics on the analysis?
**Q3**: Can we improve on 4D-Var inversion using optimal analysis and their uncertainties?




PvKF: Parametric variance Kalman Filter CV: Cross-Validation

# Model and Data (+ Adaptation)

Bias correction (I)




**Model**: Hemispheric CMAQ v5 and CMAQ-ADJ

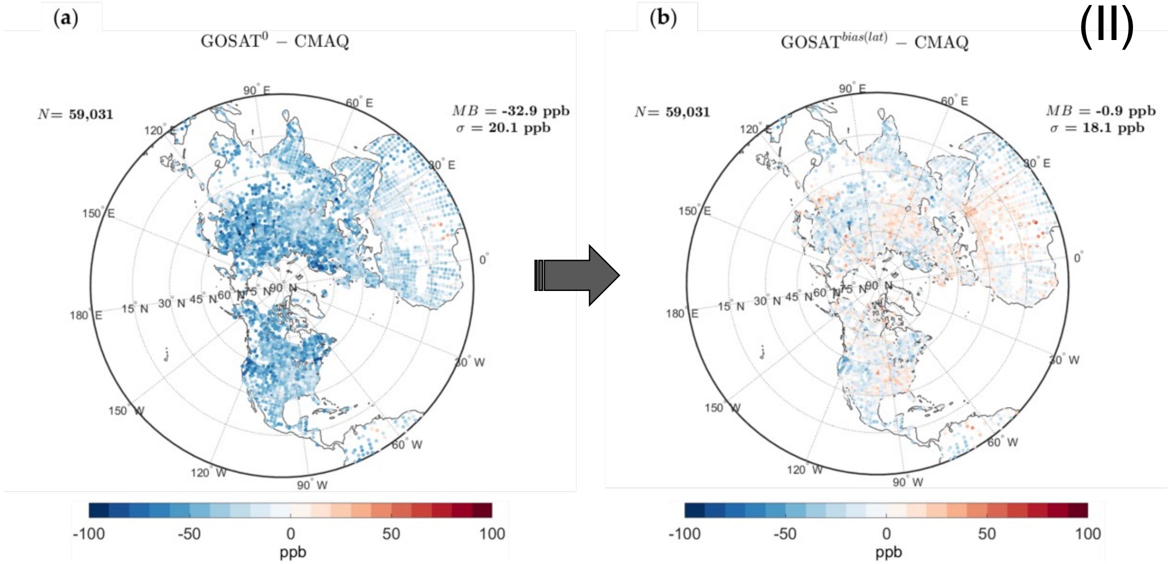
- Processing Emissions:
  Anthropogenic (EDGAR v6) + Natural (WetCHARTs v3.0)
- Modifying chemical mechanism of gas-phase chemistry in CCTM:

$$CH_4 + OH \rightarrow CH_3 + H_2O$$

**Data**: GOSAT observations

- Bias correction relative to ObsPack surface observations
- Quality control (i.e., removing outliers)

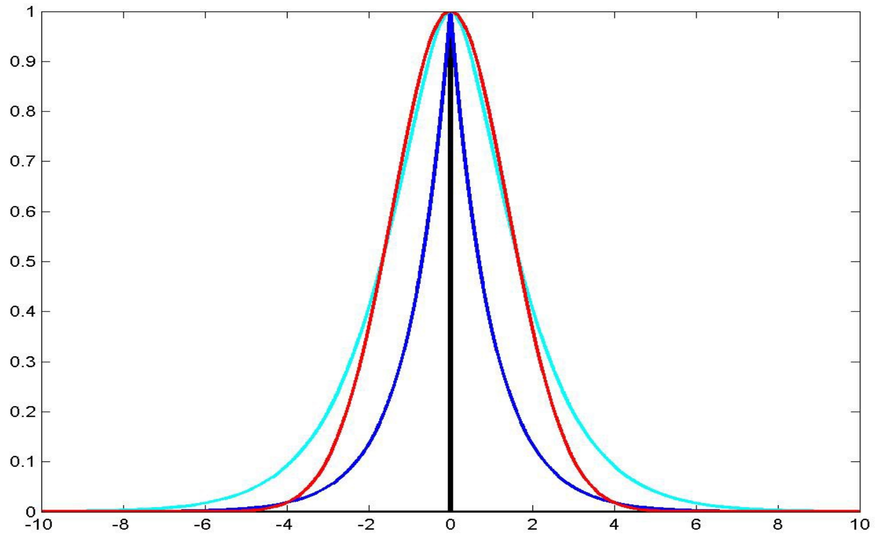
(II)



# Covariance Modelling

Examples of suitable correlation functions:

- FOAR: $C(\|\mathbf{r}-\mathbf{r}'\|) = \exp\left(-\frac{\|\mathbf{r}-\mathbf{r}'\|}{L}\right)$
- SOAR: $C(\|\mathbf{r}-\mathbf{r}'\|) = \left(1+\frac{\|\mathbf{r}-\mathbf{r}'\|}{L}\right)\exp\left(-\frac{\|\mathbf{r}-\mathbf{r}'\|}{L}\right)$
- Gaussian: $C(\|\mathbf{r}-\mathbf{r}'\|) = \exp\left(-\frac{\|\mathbf{r}-\mathbf{r}'\|^2}{2L^2}\right)$




$$\mathbf{P}^f_{n\times n} = \mathbf{\Sigma}^f_t \mathbf{C}\,\mathbf{\Sigma}^f_t$$

Forecast error

$$P(\mathbf{r},\mathbf{r}',t) = \sigma(\mathbf{r},t)C(\mathbf{r},\mathbf{r}',t)\sigma(\mathbf{r}',t)$$

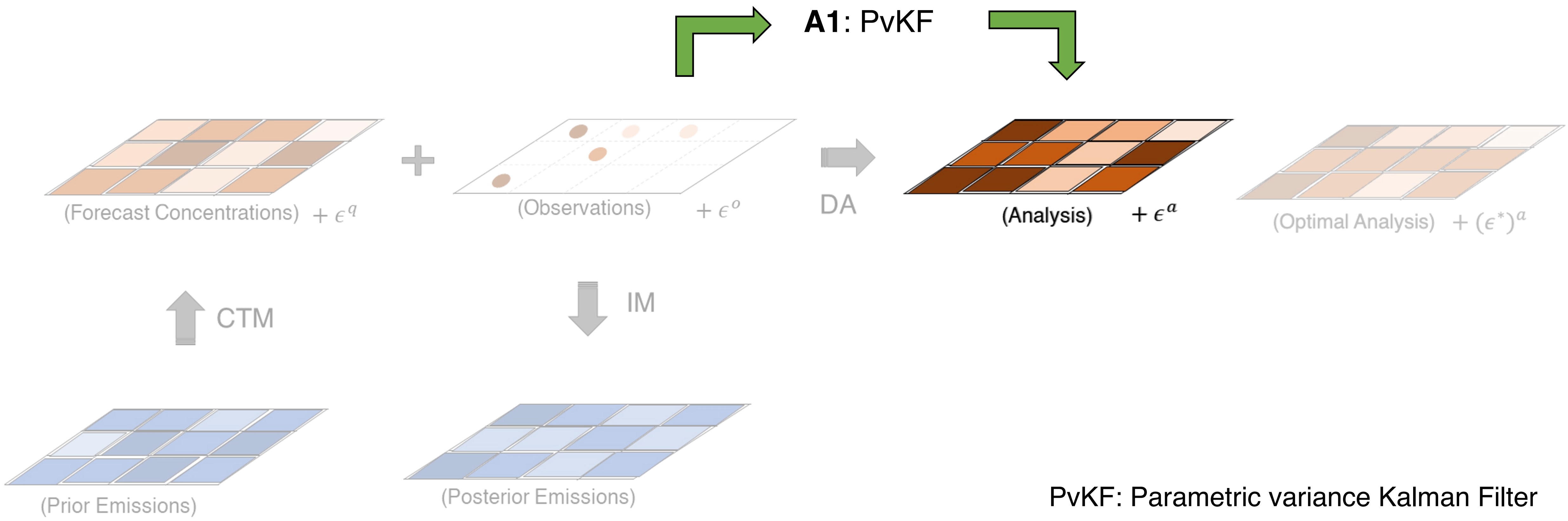
$$C(\mathbf{r},\mathbf{r}') = C(\|\mathbf{r}-\mathbf{r}'\|)$$

$$\mathbf{R}_{m\times m} = \left(f^o\varepsilon^m\right)^2\mathbf{I}$$

$$\mathbf{Q}_{n\times n} = \left(f^q\varepsilon^q\right)^2\mathbf{I}$$

We will estimate $\boldsymbol{\alpha} = \{f^o, f^q, L, \ldots\}$

# **Q1**: How to obtain a low-cost DA system, capable of estimating uncertainties?




PvKF: Parametric variance Kalman Filter

# **Part I**: Development of PvKF Assimilation

Forecast step

1. Transport of the model state

$$\frac{\partial c}{\partial t}+\mathbf{V}\cdot\nabla c-\frac{1}{\rho}\nabla\cdot\left(\rho\mathbf{K}\nabla c\right)+\mathrm{R}=\mathrm{E}$$

2. Advection of error variance

$$\frac{\partial\sigma^2}{\partial t}+\mathbf{V}\cdot\nabla\sigma^2=q$$

Analysis step

1. Covariance function convolution

$$\zeta = 1,\dots,m$$

$$H_{\mathbf{x}}[P^f(\mathbf{x},\mathbf{x}',t)] \Leftrightarrow \mathbf{H}\,\mathbf{P}^f$$

$$\zeta = 1,\dots,m$$

$$H_{\mathbf{x}'}[P^f(\mathbf{x},\mathbf{x}',t)] \Leftrightarrow \mathbf{P}^f\mathbf{H}^T$$

$$H_{\mathbf{x}}\left[H_{\mathbf{x}'}\left(P^f(\mathbf{x},\mathbf{x}',t)\right)\right] \Leftrightarrow \mathbf{H}\mathbf{P}^f\mathbf{H}^T$$

2. Matrix inversion

$$\mathbf{\Gamma}=\mathbf{H}\mathbf{P}^f\mathbf{H}^T+\mathbf{R}$$

$\mathbf{\Gamma}b=d$ $\longrightarrow$ *Cholesky decomposition*

3. Analysis increment/
Analysis error variance reduction

$$\zeta = 1,\dots,m$$

$$\mathbf{X}^a=\mathbf{X}^f+\mathbf{K}(\mathbf{Y}^o-\mathbf{H}\mathbf{X}^f)$$

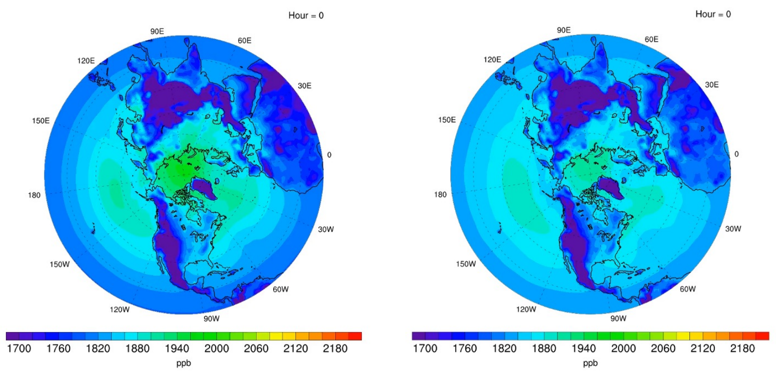
$$\mathbf{P}^a=\mathbf{P}^f-(\mathbf{H}\mathbf{P}^f)^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T+\mathbf{R})^{-1}(\mathbf{H}\mathbf{P}^f)$$

- Large state-space problem (e.g., ~1.5e6 elements)
- Produces forecast and analyses and explicitly evolve its error variance
- Computationally advantageous compared to 4D-Var and EnKF
- Accounts for model imperfection
- High potential for real-time or operational assimilation

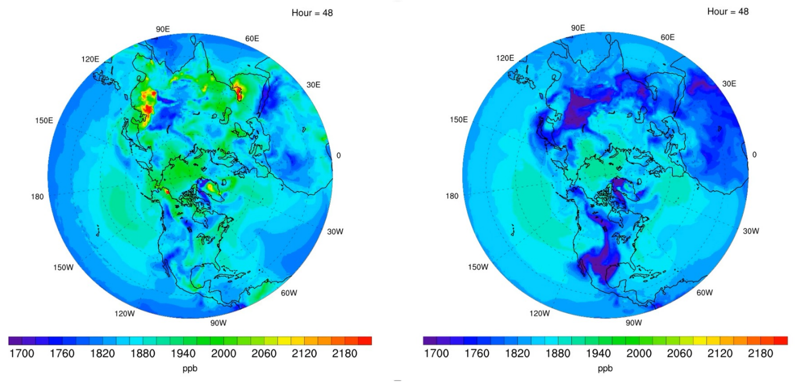
# **Part I**: Evolution of the State

Methane Analysis (with DA) vs. Methane Model (without DA)

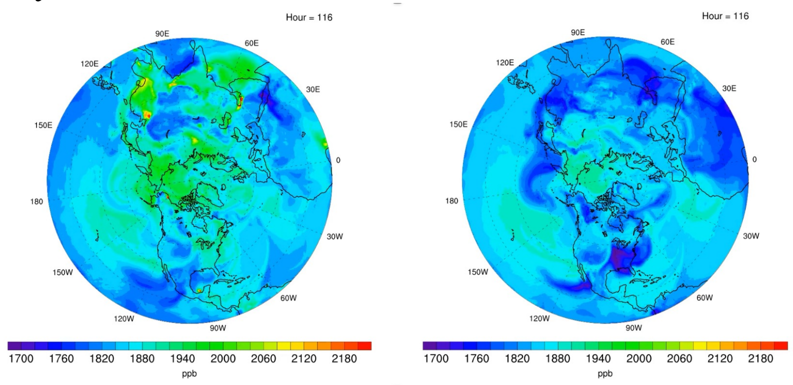
Day 0



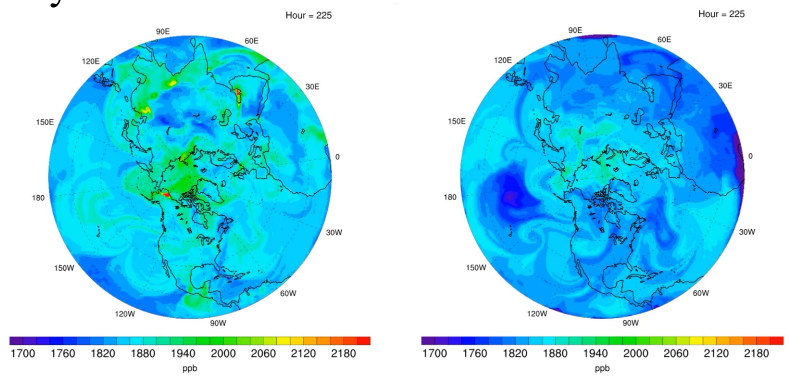

Day 4



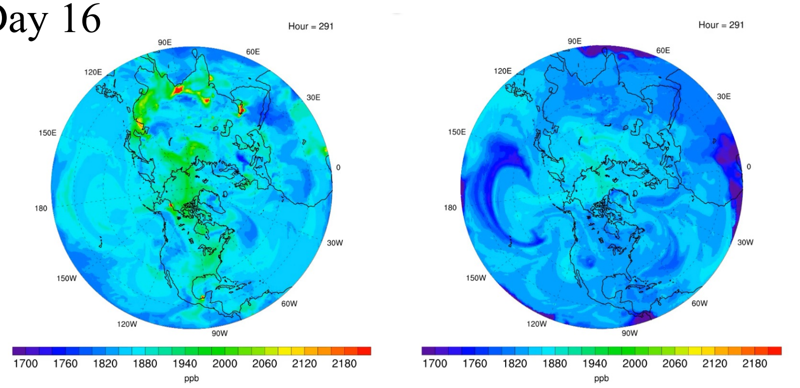

Day 8



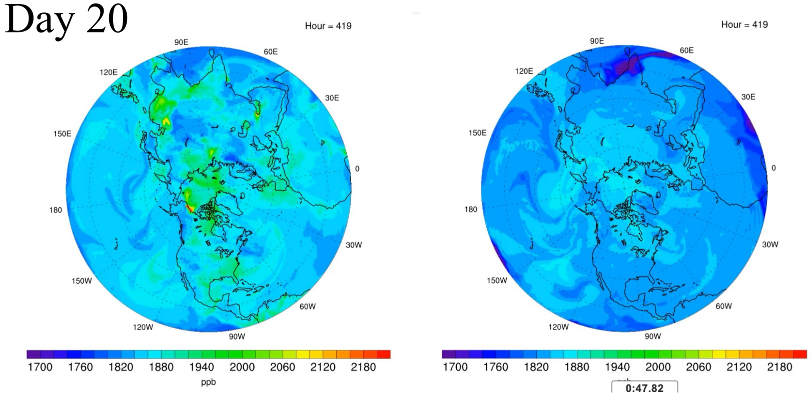

Day 12




Day 16

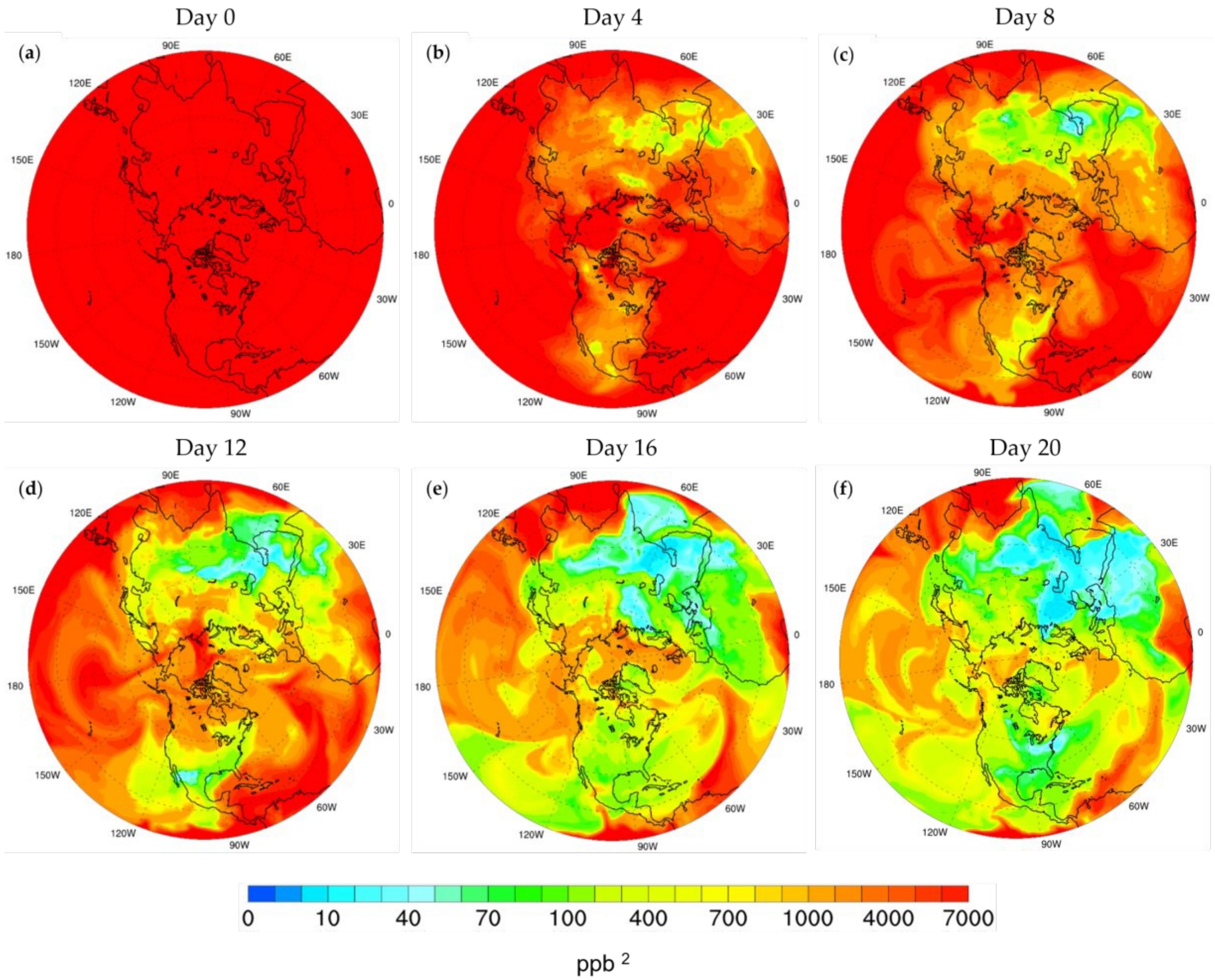



Day 20



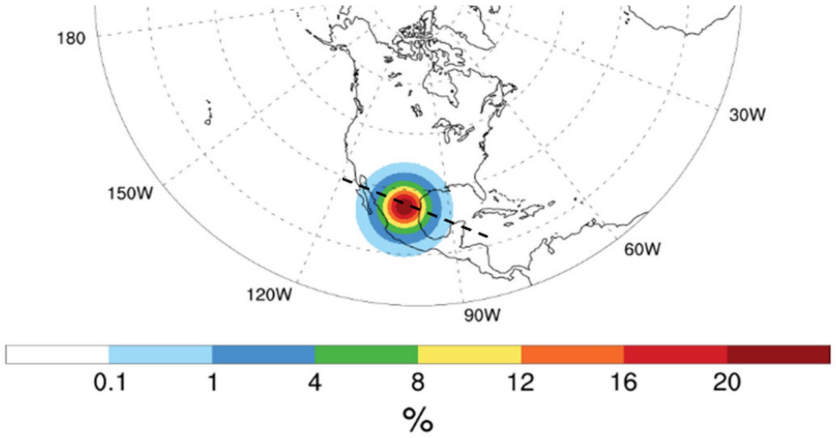
# **Part I**: Evolution of the Error Variance

Methane Analysis Error Variance:

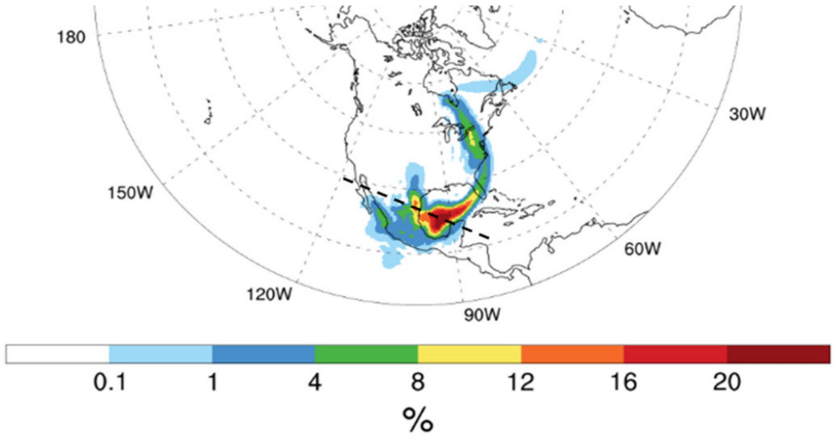


# **Part I**: Verification with Single Observation

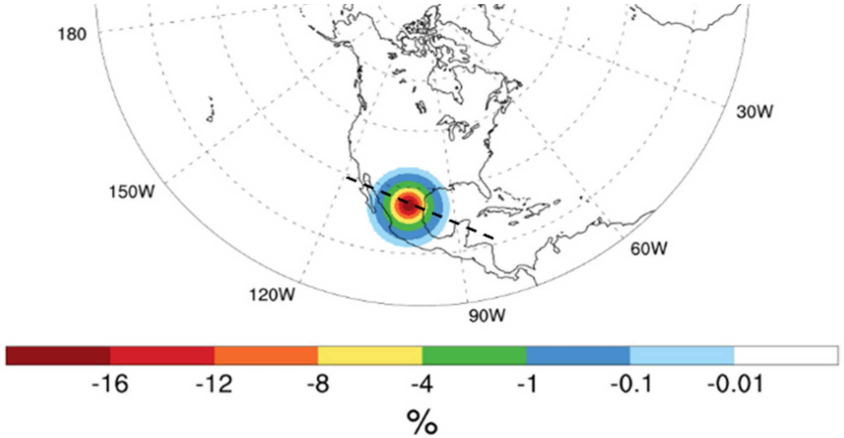
**Analysis** increment (Day 0)



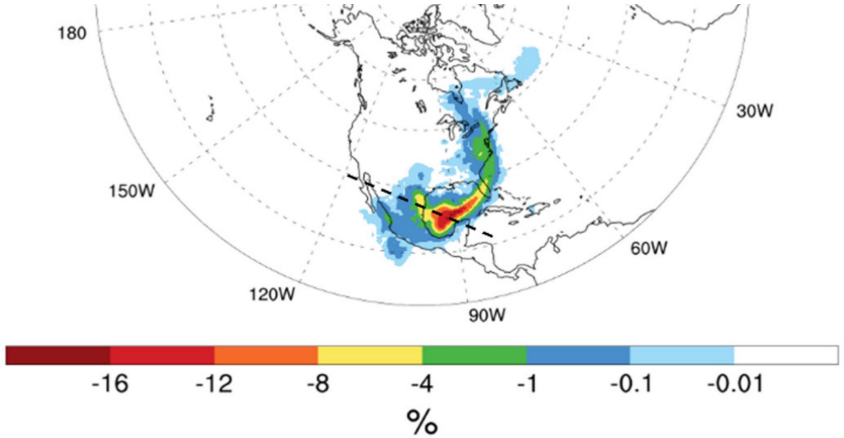

**Analysis** increment (Day 3)




**Error variance** reduction (Day 0)



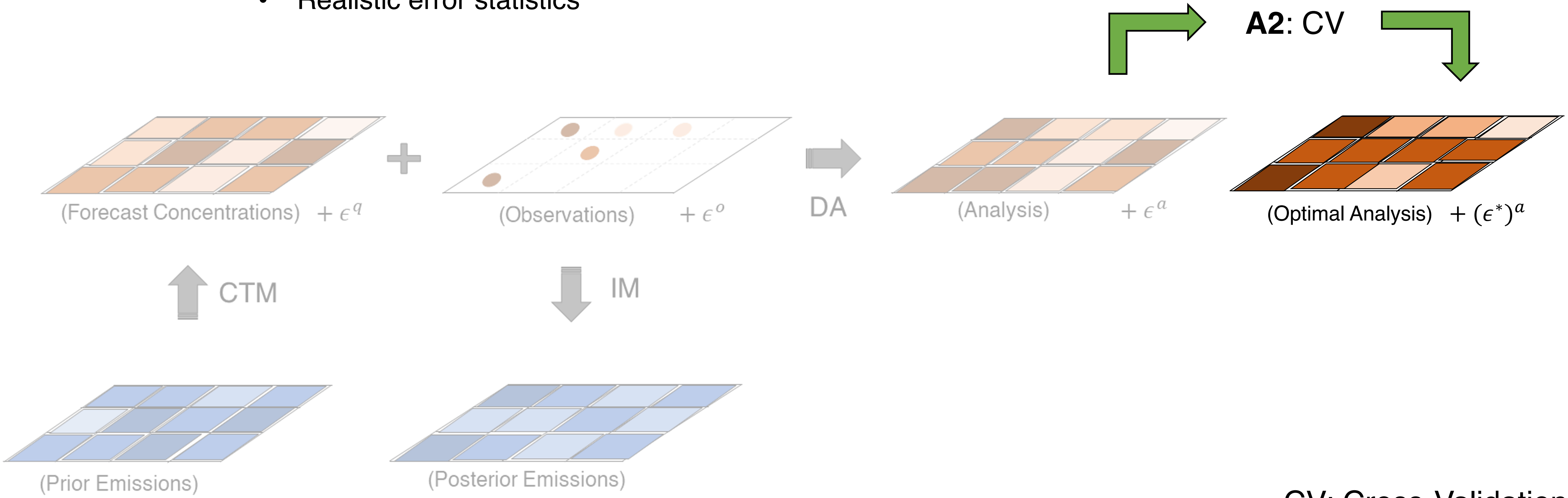

**Error variance** reduction (Day 3)



# **Q2**: What is the impact of optimal error statistics on the Analysis?

We want to obtain

- Optimal (true) analysis
- Realistic error statistics




CV: Cross-Validation

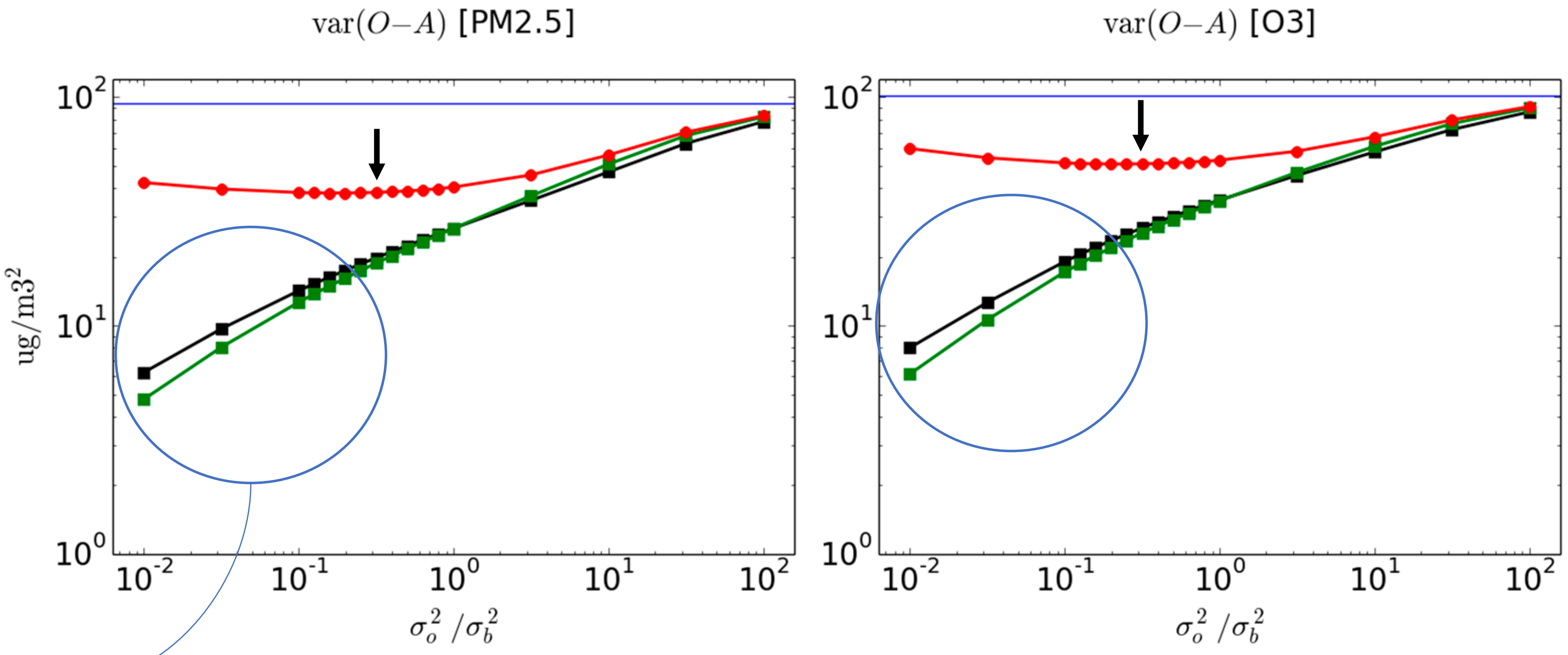
# **Part II**: Why Cross-Validation?

Because it does not assume that the analysis is already optimal

$$tr(\mathbf{A}(\boldsymbol{\alpha})) = J = \left\langle (O-A)^2 \right\rangle \quad \longrightarrow \quad \arg\min_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha}) \Rightarrow \text{optimal } \mathbf{A}(\boldsymbol{\alpha})$$

$$\boldsymbol{\alpha} = \{f^o, f^q, L, \ldots\}$$

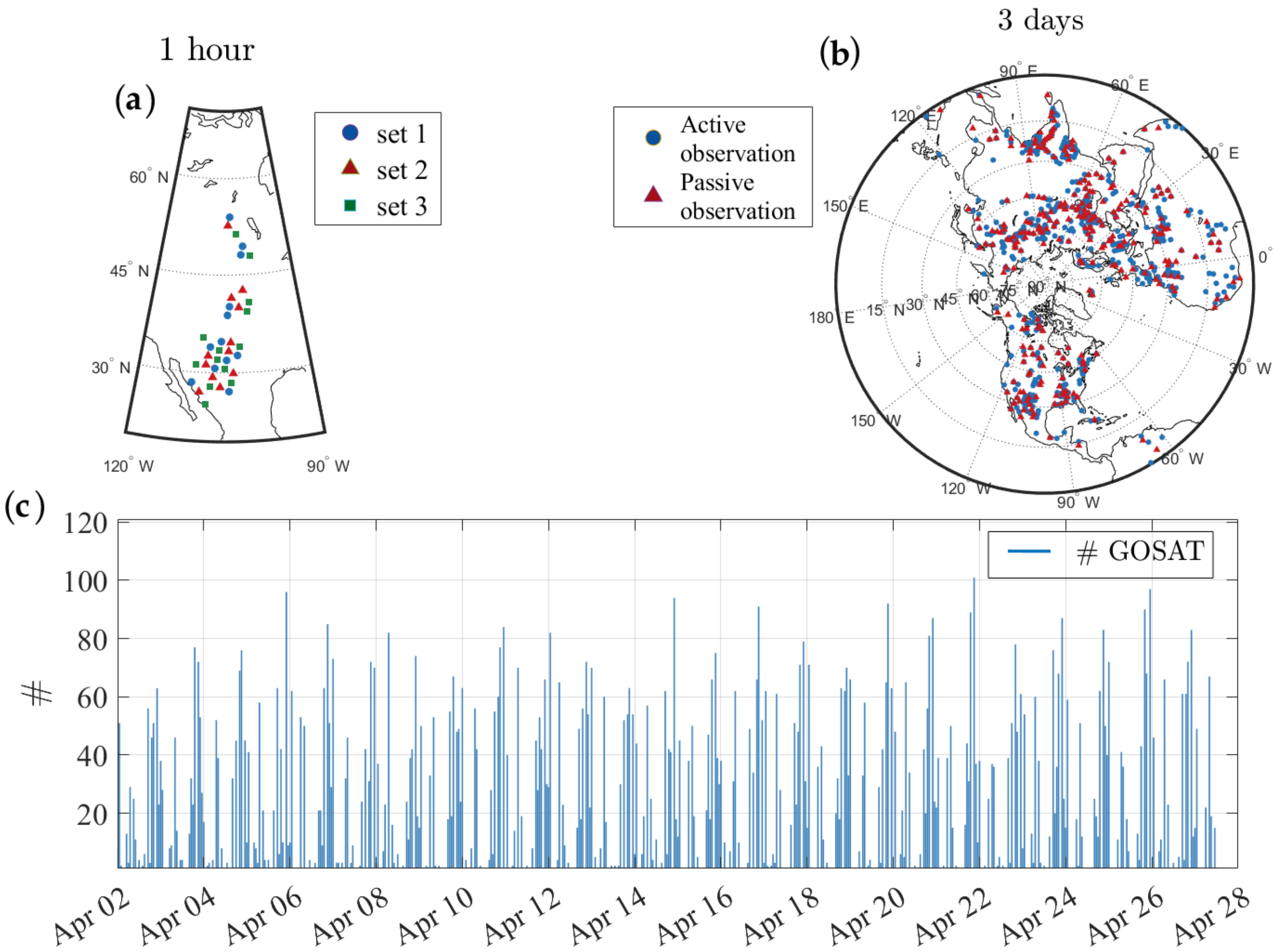
| With CV | Without CV |
| --- | --- |
| | $J = \left\langle (O-A)^2 \right\rangle$ |
| $J_c = \left\langle (O-A)_c^2 \right\rangle$ | $J = \left\langle (O-A)^2 \right\rangle$ |
| Uncorrelated | Correlated |




Overestimation

(Ménard and Deshaies-Jacques, 2018a)

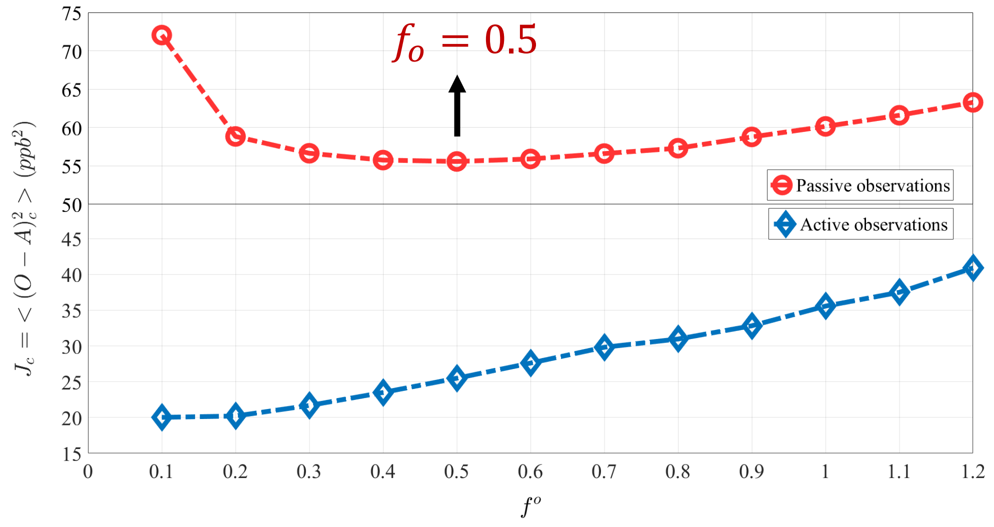
# **Part II**: Cross-Validation with GOSAT

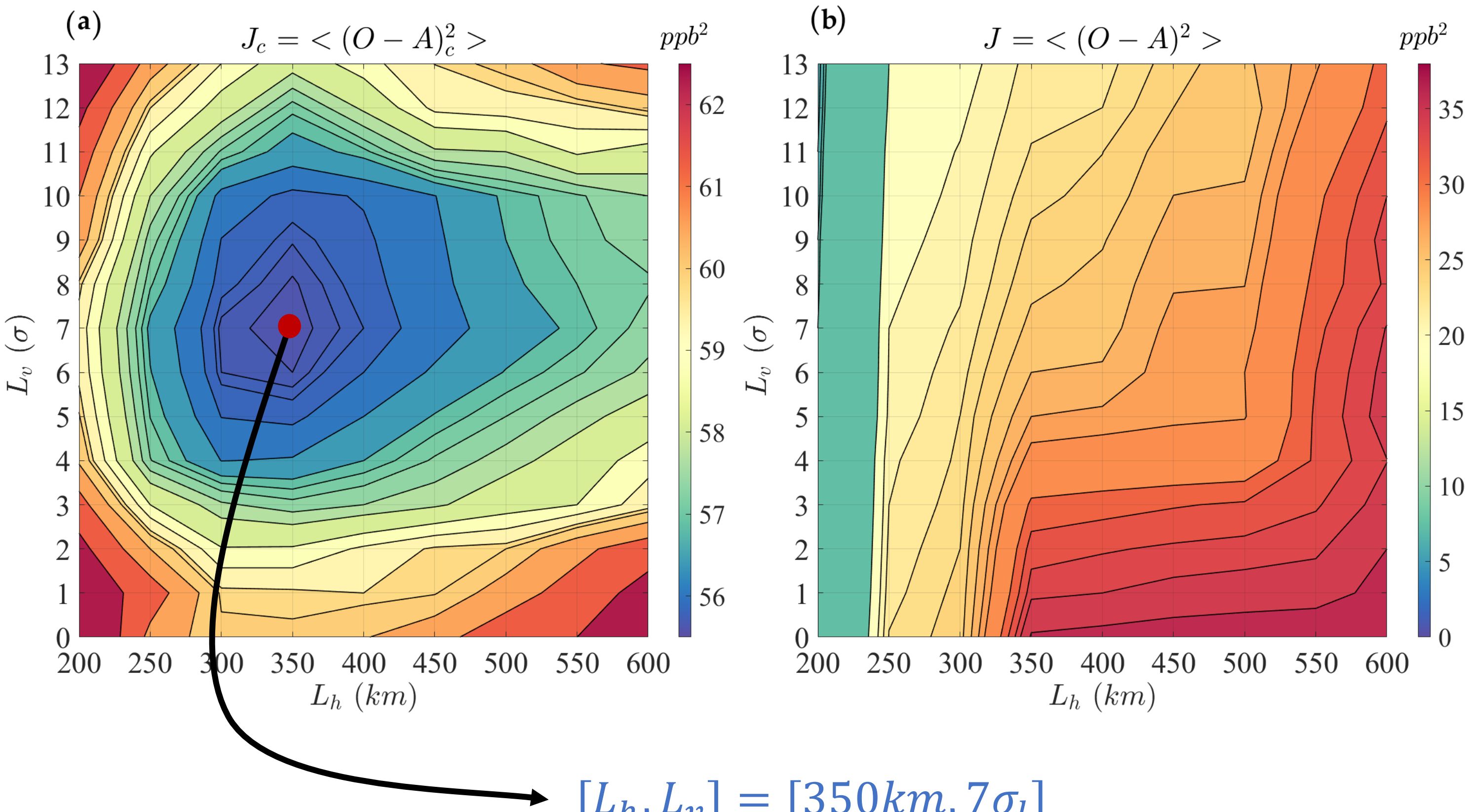


# **Part II**: Estimate Error Covariances Parameters

Optimizing CV cost function to obtain error parameters, corresponding to optimal solution.

| itr/Step \ Parameters | $L_h$ | $L_v$ | $f^o$ | $J_c = \langle (O-A)_c^2 \rangle$ |
|---|---|---|---|---|
| initial | 500 km | $1\sigma$ | 1 | 62.4 ppb$^2$ |
| itr 0/step 1 | 500 km | $1\sigma$ | 0.45 | 60.5 ppb$^2$ |
| itr 0/step 2 | 350 km | $7\sigma$ | 0.45 | 55.9 ppb$^2$ |
| itr 1/step 1 | 350 km | $7\sigma$ | 0.5 | 55.6 ppb$^2$ |
| itr 1/step 2 | 350 km | $7\sigma$ | 0.5 | 55.6 ppb$^2$ |




$[L_h, L_v] = [350km, 7\sigma_l]$

# **Part II**: Impact of Optimal Estimation

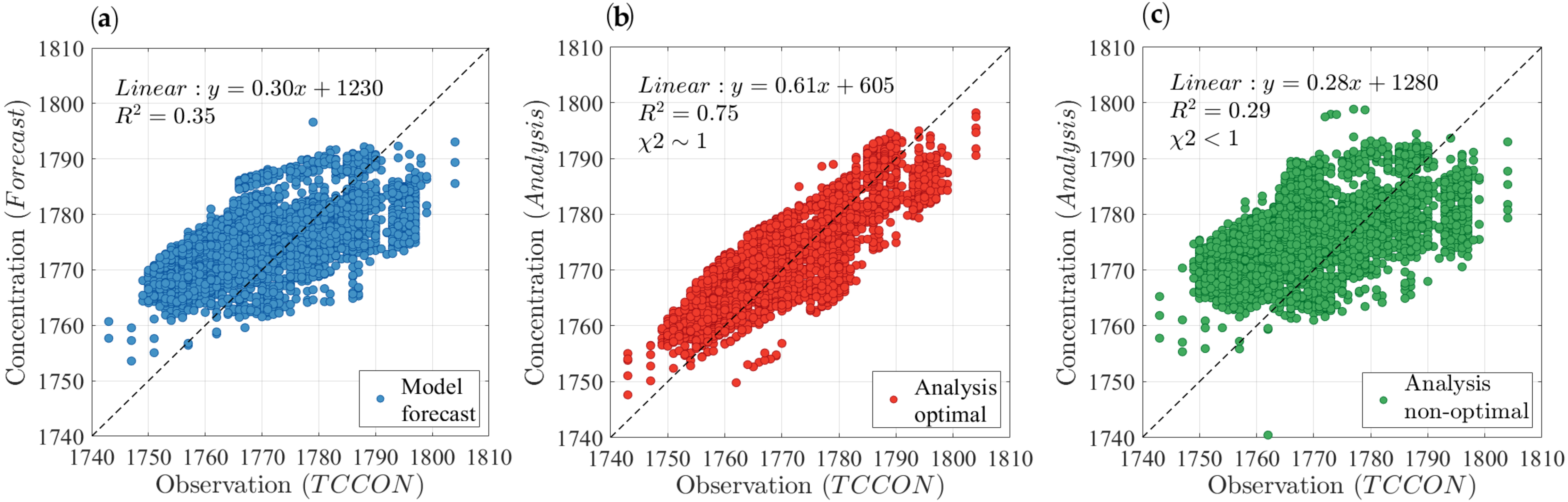
Optimal estimation parameters:

$$f^o = 0.5, f^i = 0.45, f^q = 0.018, L_h = 350\ \mathrm{km}, L_v = 7\sigma$$

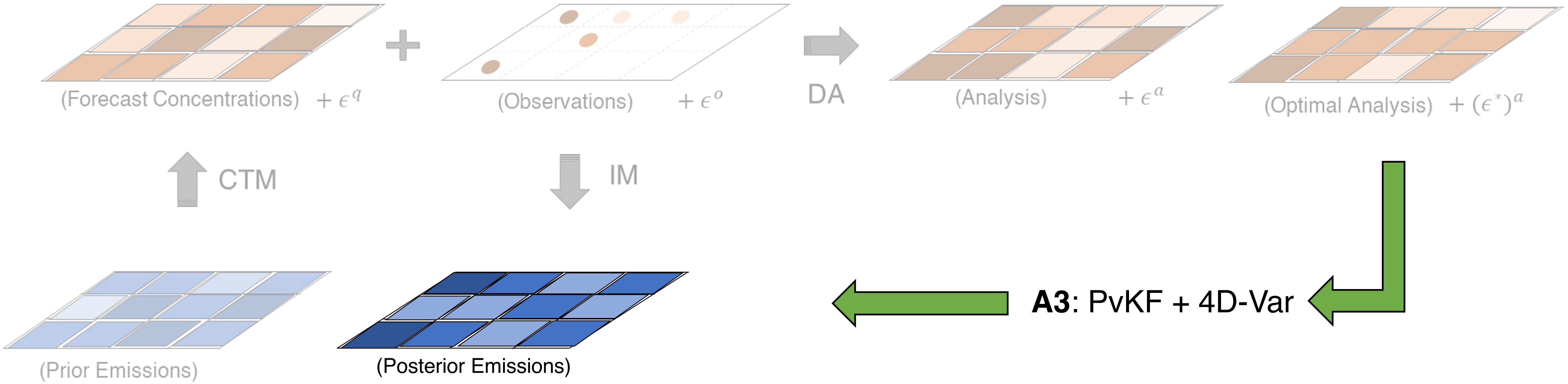
Non-optimal estimation but commonly used parameters:

$$f^o = 1.2, f^i = 0.45, f^q = 0, L_h = 600\ \mathrm{km}, L_v = 1\sigma$$

Optimality of error parameters has a crucial impact on the assimilation result.

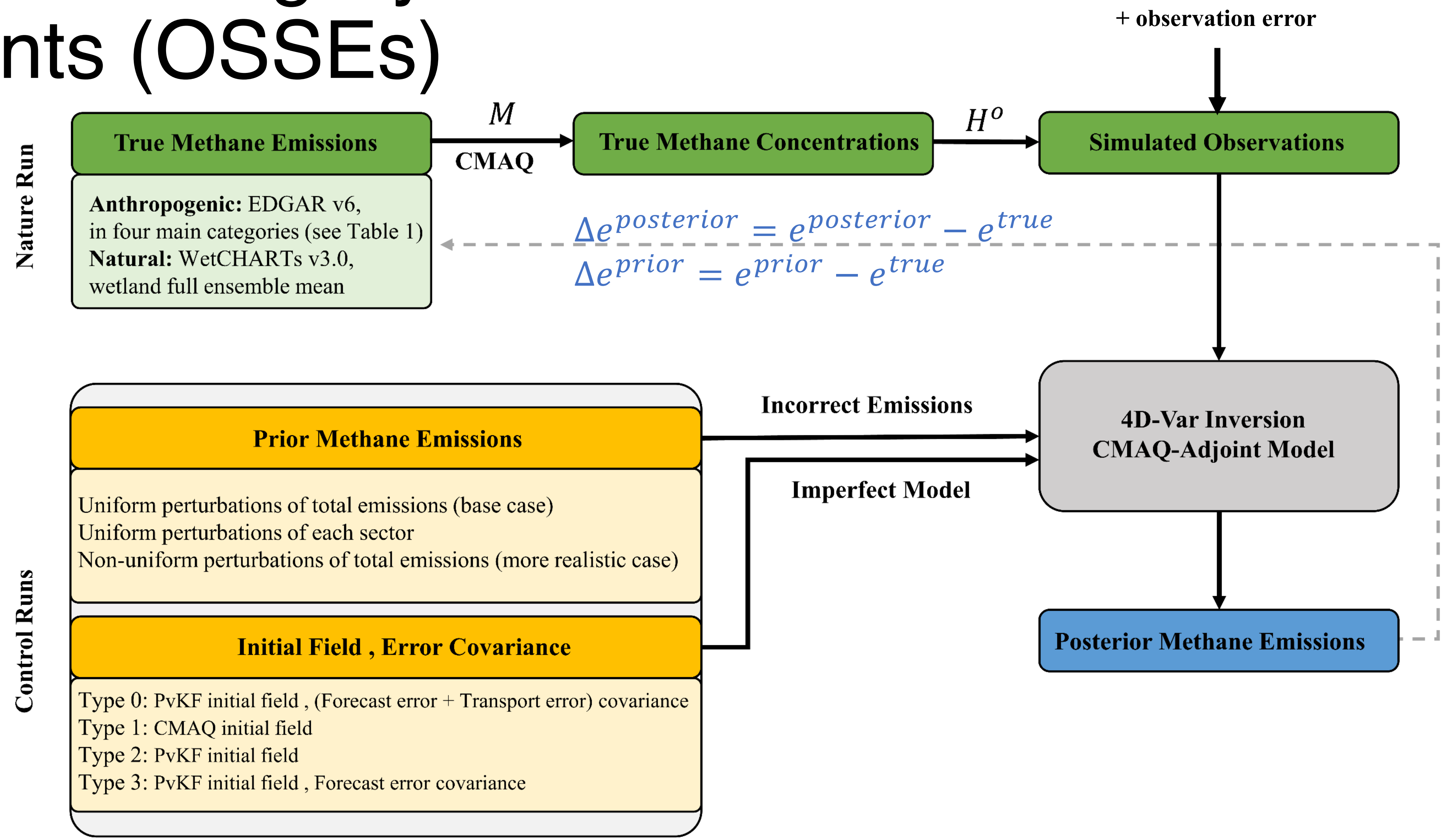


# **Q3**: Can we improve on 4D-Var inversion using optimal analysis and their uncertainties?



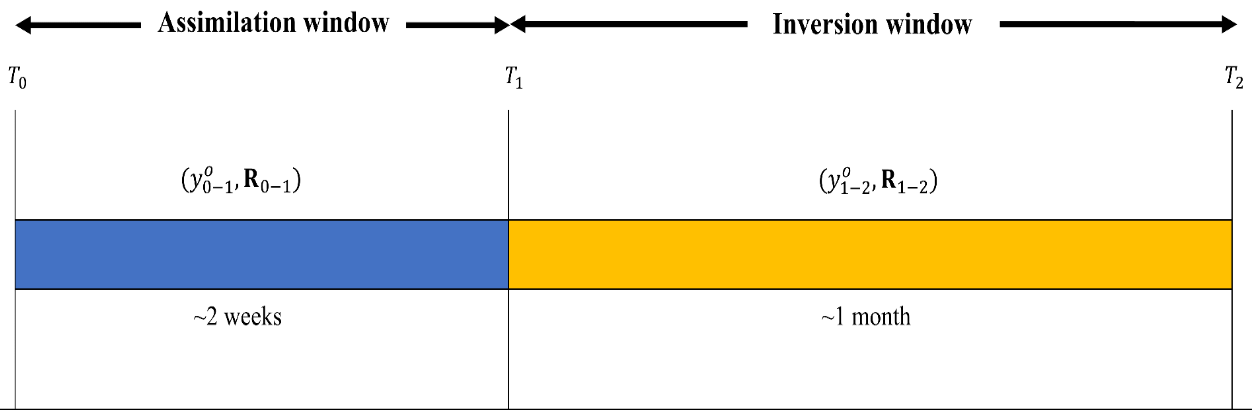
# **Part III**: Observing System Simulation Experiments (OSSEs)

To evaluate the proposed inversion:



# **Part III**: Use of PvKF Assimilation for Emissions Inversion (4D-Var)

$y$: Observations
$x$: Emission scaling factor
$c$: Model/assimilation concentration
$\mathbf{R}$: Observation error covariance
$\mathbf{P}$: Forecast error covariance
$\mathbf{Q}$: Model transport error covariance
$\mathbf{A}$: Analysis error covariance




| | Experiments | | | | Assumptions |
|---|---|---|---|---|---|
| **Classical form (Previous studies)** | 1 | $M$ | $c_1^f$ | $\mathbf{P}^f_{1-2} = \emptyset$ | Perfect forecast field<br>Perfect model |
| | 2 | PvKF | $c_1^a$ | $\mathbf{P}^f_{1-2} = \emptyset$ | Perfect analysis field<br>Perfect model |
| **Other variation** | 3 | PvKF | $(c_1^a, \mathbf{A}_1)$ | $\mathbf{P}^f_{1-2}(\mathbf{A}_1)$ | Imperfect analysis field<br>Perfect model |
| **This study** | 4* | PvKF | $(c_1^a, \mathbf{A}_1)$ | $\mathbf{P}^f_{1-2}(\mathbf{A}_1, \mathbf{Q})$ | Imperfect analysis field<br>Imperfect model |

# **Part III**: Different From of 4D-Var Cost Functions

| Type # | Cost function | | $c_1^a$ | $\mathbf{P}(\mathbf{A}_1)$ | $\mathbf{P}(\mathbf{Q})$ |
|---|---|---|---|---|---|
| Type 0: | $J_0(x)=\frac{1}{2}\gamma(x-x_b)^T\mathbf{B}^{-1}(x-x_b)+\sum_{t=0}^{n}\frac{1}{2}\left(y_t^o-H_t(c_1^a,x)\right)^T\left(H^o\mathbf{P}_t^f(\mathbf{A}_1,\mathbf{Q})H^{o\,T}+\mathbf{R}_t\right)^{-1}\left(y_t^o-H_t(c_1^a,x)\right)$ | This study | ✓ | ✓ | ✓ |
| Type 1: | $J_1(x)=\frac{1}{2}\gamma(x-x_b)^T\mathbf{B}^{-1}(x-x_b)+\sum_{t=0}^{n}\frac{1}{2}\left(y_t^o-H_t(c_1^f,x)\right)^T(\mathbf{R}_t)^{-1}\left(y_t^o-H_t(c_1^f,x)\right)$ | Classical form | ✗ | ✗ | ✗ |
| Type 2: | $J_2(x)=\frac{1}{2}\gamma(x-x_b)^T\mathbf{B}^{-1}(x-x_b)+\sum_{t=0}^{n}\frac{1}{2}\left(y_t^o-H_t(c_1^a,x)\right)^T(\mathbf{R}_t)^{-1}\left(y_t^o-H_t(c_1^a,x)\right)$ | | ✓ | ✗ | ✗ |
| Type 3: | $J_3(x)=\frac{1}{2}\gamma(x-x_b)^T\mathbf{B}^{-1}(x-x_b)+\sum_{t=0}^{n}\frac{1}{2}\left(y_t^o-H_t(c_1^a,x)\right)^T\left(H^o\mathbf{P}_t^f(\mathbf{A}_1)H^{o\,T}+\mathbf{R}_t\right)^{-1}\left(y_t^o-H_t(c_1^a,x)\right)$ | Other variation | ✓ | ✓ | ✗ |

$c_1^a$ : optimal analysis field
$\mathbf{P}(\mathbf{A}_1)$: propagated analysis error covariance
$\mathbf{P}(\mathbf{Q})$ : propagated modelling (transport) error covariance

# **Part III**: Uniform Perturbations

This study

Classical form

Other variation

$c_1^a$ ✓ P(A$_1$) ✓ P(Q) ✓

$c_1^a$ ✗ P(A$_1$) ✗ P(Q) ✗

$c_1^a$ ✓ P(A$_1$) ✗ P(Q) ✗

$c_1^a$ ✓ P(A$_1$) ✓ P(Q) ✗

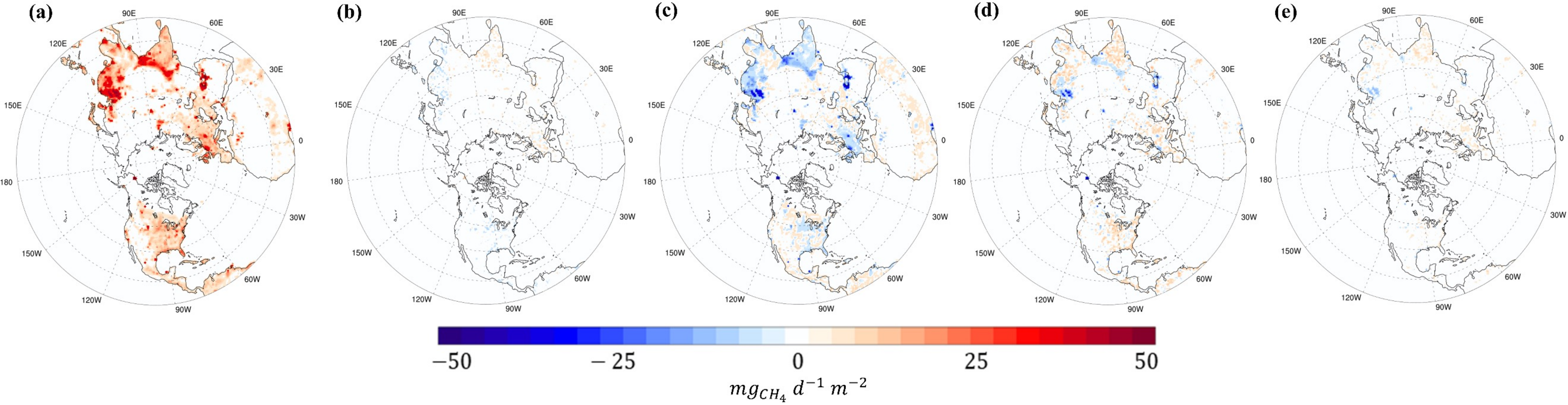
$\Delta e^{prior}$

Type 0: $\Delta e_{J_0}^{posterior}$

Type 1: $\Delta e_{J_1}^{posterior}$

Type 2: $\Delta e_{J_2}^{posterior}$

Type 3: $\Delta e_{J_3}^{posterior}$



# **Part III**: Uniform Perturbations

This study

$c_1^a$ ✓ $\mathbf{P}(\mathbf{A}_1)$ ✓ $\mathbf{P}(\mathbf{Q})$ ✓

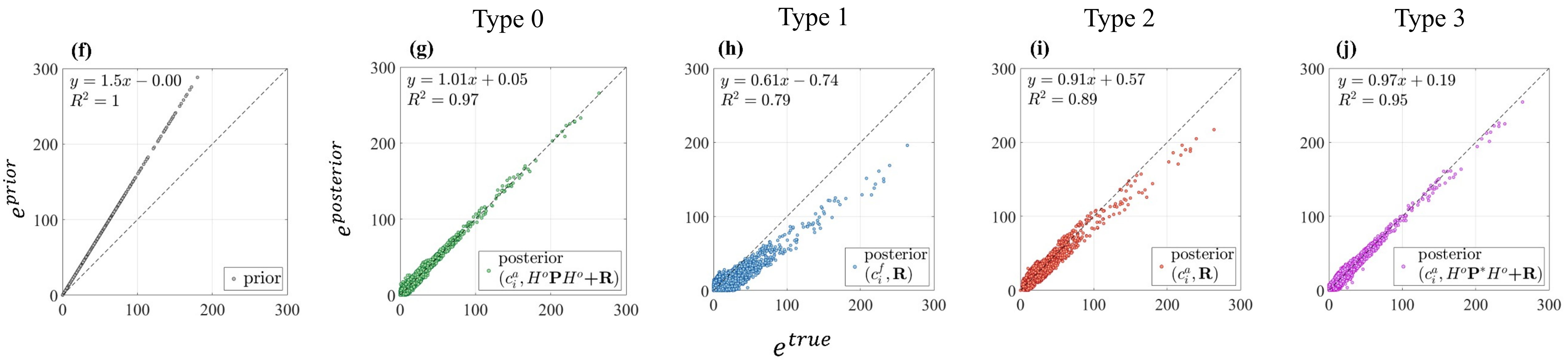
Classical form

$c_1^a$ ✗ $\mathbf{P}(\mathbf{A}_1)$ ✗ $\mathbf{P}(\mathbf{Q})$ ✗

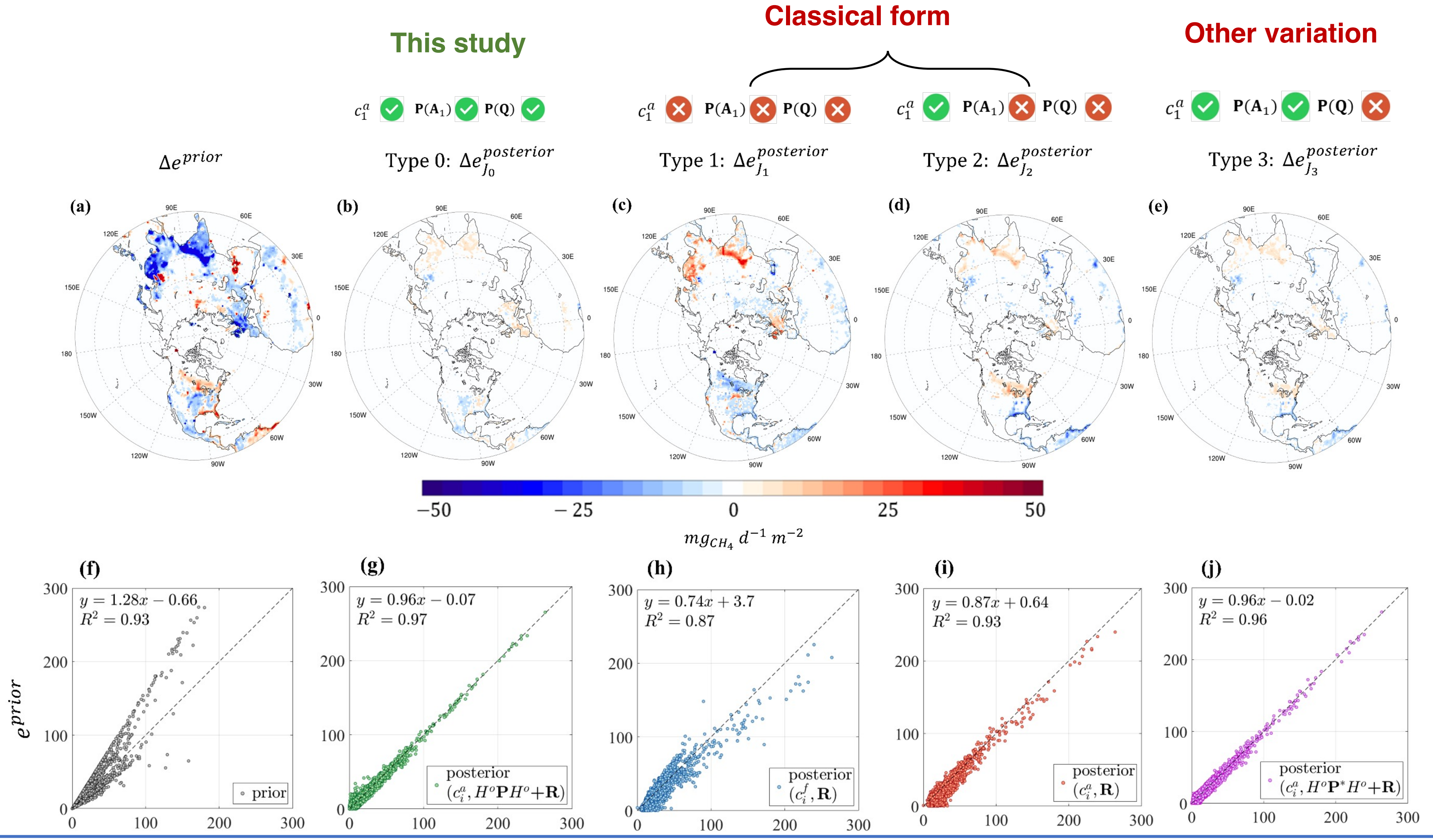
$c_1^a$ ✓ $\mathbf{P}(\mathbf{A}_1)$ ✗ $\mathbf{P}(\mathbf{Q})$ ✗

Other variation

$c_1^a$ ✓ $\mathbf{P}(\mathbf{A}_1)$ ✓ $\mathbf{P}(\mathbf{Q})$ ✗



# **Part III**: Non-uniform Perturbations



# Conclusions & Future Work

(**I**) PvKF assimilation is a stand-alone DA framework that improves our understanding of atmospheric methane estimation

- No need to assume a perfect model
- Provides an (continuous) **estimation of methane analysis and its uncertainties cost-effectively**

(**II**) Realistic error statistics and optimal analysis play a key role in PvKF assimilation

- Optimal analysis is obtained by **optimizing error statistics using cross-validation**
- Non-optimal error covariances can lead to an analysis even worse than the model forecast

(**III**) PvKF assimilation can be used in conjunction with an inversion system

- **Improve the typical 4D-Var inversion results** by providing more sophisticated form of error correlations and initial optimal analysis field

**Suggestions for future work**

- Extending PvKF assimilation framework to a jointly source-state estimation (i.e., emissions error will be estimated as part of the solution)
- Further development of PvKF assimilation for other species such as short-lived pollutants (likely requires evolving error correlations)
- Conducting PvKF with dense satellite observations (e.g., TROPOMI) for high-resolution inversion in regional domain (e.g., CONUS)
- Application of PvKF analysis to remove (measurement) biases over remote area such as oceans

# References:

Bergamaschi P, Karstens U, Manning AJ, Saunois M, Tsuruta A, Berchet A, Vermeulen AT, Arnold T, Janssens-Maenhout G, Hammer S et al. 2018. Inverse modelling of European ch4 emissions during 2006-2012 using different inverse models and reassessed atmospheric observations. Atmospheric Chemistry and Physics. 18(2):901-920.

Bousserez N, Henze DK, Rooney B, Perkins A, Wecht KJ, Turner AJ, Natraj V, Worden JR. 2016. Constraints on methane emissions in North America from future geostationary remote-sensing measurements. Atmospheric Chemistry and Physics (Online). 16(10).

CMAQ tutorials. 2022. [accessed 2 October 2022]. https://www.epa.gov/cmaq/cmaq-documentation.

Ganesan AL, Schwietzke S, Poulter B, Arnold T, Lan X, Rigby M, Vogel FR, van der Werf GR, Janssens-Maenhout G, Boesch H et al. 2019. Advancing scientific understanding of the global methane budget in support of the Paris agreement. Global Biogeochemical Cycles. 33(12):1475-1512.

Janardanan R, Maksyutov S, Tsuruta A, Wang FJ, Tiwari YK, Valsala V, Ito A, Yoshida Y, Kaiser JW, Janssens-Maenhout G et al. 2020. Country-scale analysis of methane emissions with a high-resolution inverse model using GOSAT and surface observations. Remote Sensing. 12(3):24.

Miller SM, Michalak AM, Detmers RG, Hasekamp OP, Bruhwiler LMP, Schwietzke S. 2019. China's coal mine methane regulations have not curbed growing emissions. Nature Communications. 10.

Menard R, Deshaies-Jacques M, Gasset N. 2016. A comparison of correlation-length estimation methods for the objective analysis of surface pollutants at environment and climate change canada. Journal of the Air & Waste Management Association. 66(9):874-895.

Menard R, Deshaies-Jacques M. 2018b. Evaluation of analysis by cross-validation. Part i: Using verification metrics. Atmosphere. 9(3):16.

Menard R, Deshaies-Jacques M. 2018a. Evaluation of analysis by cross-validation, part ii: Diagnostic and optimization of analysis error covariance. Atmosphere. 9(2):21.

Turner AJ, Jacob DJ, Wecht KJ, Maasakkers JD, Lundgren E, Andrews AE, Biraud SC, Boesch H, Bowman KW, Deutscher NM et al. 2015. Estimating global and North American methane emissions with high spatial resolution using GOSAT satellite data. Atmospheric Chemistry and Physics. 15(12):7049-7069.

Voshtani S, Menard R, Walker TW, Hakami A. 2022a. Assimilation of GOSAT methane in the hemispheric CMAQ; part i: Design of the assimilation system. Remote Sensing. 14(2):32.

Voshtani S, Menard R, Walker TW, Hakami A. 2022b. Assimilation of GOSAT methane in the hemispheric CMAQ; part ii: Results using optimal error statistics. Remote Sensing. 14(2):27.

Voshtani S, Menard R, Walker TW, Hakami A. 2023 (in review). Use of Assimilation Analysis in 4D-Var Source Inversion: Observing System Simulation Experiments (OSSEs) with GOSAT Methane and Hemispheric CMAQ. Atmosphere.

Yu K, Keller CA, Jacob DJ, Molod AM, Eastham SD, Long MS. 2018. Errors and improvements in the use of archived meteorological data for chemical transport modeling: An analysis using geos-chem v11-01 driven by geos-5 meteorology. Geoscientific Model Development. 11(1):305-319.

Zavala-Araiza D, Alvarez RA, Lyon DR, Allen DT, Marchese AJ, Zimmerle DJ, Hamburg SP. 2017. Super-emitters in natural gas infrastructure are caused by abnormal process conditions. Nature Communications. 8:10.

Zhang YZ, Jacob DJ, Lu X, Maasakkers JD, Scarpelli TR, Sheng JX, Shen L, Qu Z, Sulprizio MP, Chang JF et al. 2021. Attribution of the accelerating increase in atmospheric methane during 2010-2018 by inverse analysis of GOSAT observations. Atmospheric Chemistry and Physics. 21(5):3643-3666.

# Thank you!

sina.voshtani@utoronto.ca