Uncertainty quantification for CH_4 emission estimates in TM5-4DVAR

Mihai Alexe[1], Peter Bergamaschi[1], Arjo Segers[2]

[1]EC Joint Research Centre

[2]TNO

November 5, 2015

Motivation

TM5-4D-Var is a *deterministic* inversion framework.

- We estimate the CH_4 emissions that *best fit* the model and observations under some error assumptions (an optimal solution, also called the *analysis*).

However, error covariance information for the optimal solution is not easy to obtain.

- If the model is linear, and (model and observation) errors are Gaussian, then TM5-CONGRAD can estimate the leading components of the analysis error covariance matrix.
- **What can we do if the model is nonlinear and/or prior emission statistics are non-Gaussian?**

4D-Var

Joint Research Centre

We are given:

- the background state vector $\mathbf{x}_b$ with error covariance $\mathbf{B} = \mathbf{DCD} := \mathbf{LL}^T$ ($\mathbf{L} := \mathbf{DC}^{1/2}$),
- the (potentially nonlinear) model operator $\mathbf{H}$ (forward model + observation operator), and
- the measurement data vector $\mathbf{y}$ with error covariance $\mathbf{R}$.

Solve for the *analysis* state:

$$\mathbf{x}_a = \arg\min_{\mathbf{x}} \mathcal{J} = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{Hx} - \mathbf{y})^T \mathbf{R}^{-1}(\mathbf{Hx} - \mathbf{y})$$

CONGRAD UQ (Meirink et al., ACP, 2008)

Joint Research Centre

CONGRAD

Assumptions: **H** is **linear**. The prior and observation errors are **Gaussian**. Then:

$$\nabla^2_{\mathbf{x}} \mathcal{J} = \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$$

We optimize the preconditioned state vector $\boldsymbol{\xi} := \mathbf{L}^{-1}\mathbf{x}$. The **covariance matrix of the analysis errors** reads:

$$\mathcal{P}^{\mathrm{apos}}(\mathbf{x}_a) = \left(\nabla^2 \mathcal{J}(\mathbf{x}_a)\right)^{-1} \approx \mathbf{B} + \sum_{i=1}^{K} \left(\frac{1}{\lambda_i} - 1\right) (\mathbf{L}\mathbf{v}_i)(\mathbf{L}\mathbf{v}_i)^T$$

$(\lambda_i, \mathbf{v}_i)_{i=1\dots K}$ are the **leading eigen-pairs** of the preconditioned Hessian matrix.

The **uncertainty reduction factor**: $\mathbf{1} - \frac{\sigma_{\mathrm{apos}}}{\sigma_{\mathrm{apri}}}$, with $\sigma_{\mathrm{apos}} = \sqrt{\mathrm{diag}\left(\mathcal{P}^{\mathrm{apos}}(\mathbf{x}_{\mathrm{a}})\right)}$.

CONGRAD UQ (Meirink et al., ACP, 2008)

Joint Research Centre

CONGRAD

Assumptions: **H** is **linear**. The prior and observation errors are **Gaussian**. Then:

$$\nabla^2_{\mathbf{x}}\mathcal{J} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$$

We optimize the preconditioned state vector $\boldsymbol{\xi} := \mathbf{L}^{-1}\mathbf{x}$. The **covariance matrix of the analysis errors** reads:

$$\mathcal{P}^{\mathrm{apos}}(\mathbf{x}_a) = \left(\nabla^2\mathcal{J}(\mathbf{x}_a)\right)^{-1} \approx \mathbf{B} + \sum_{i=1}^{K}\left(\frac{1}{\lambda_i} - 1\right)(\mathbf{L}\mathbf{v}_i)(\mathbf{L}\mathbf{v}_i)^T$$

$(\lambda_i, \mathbf{v}_i)_{i=1\ldots K}$ are the **leading eigen-pairs** of the preconditioned Hessian matrix.

The **uncertainty reduction factor**: $\mathbf{1} - \frac{\sigma_{\mathrm{apos}}}{\sigma_{\mathrm{apri}}}$, with $\sigma_{\mathrm{apos}} = \sqrt{\mathrm{diag}\left(\mathcal{P}^{\mathrm{apos}}(\mathbf{x}_{\mathrm{a}})\right)}$.

☹ We may get **negative emissions** in regions that are not well constrained by observations (we can fit anything with a Gaussian...).

TM5-4DVAR: (semi-)lognormal emissions prior

Joint Research Centre

We optimize the *emission factors* $\mathbf{x}$, not the absolute emissions $\mathbf{z}$:

Semi-lognormal emissions, e.g. Bergamaschi et. al, 2009

$$\mathbf{z}^i_{\mathrm{emis}} := \begin{cases} E^i_{\mathrm{apri}}(1 + \mathbf{x}^i_{\mathrm{emis}}) & \mathbf{x}^i_{\mathrm{emis}} \geq 0 \\ E^i_{\mathrm{apri}} \exp \mathbf{x}^i_{\mathrm{emis}} & \mathbf{x}^i_{\mathrm{emis}} < 0. \end{cases}$$

If $\mathbf{x} \sim \mathcal{N}(0, \mathbf{B})$, then $\mathcal{E}\{\mathbf{z}^i_{\mathrm{emis}}\} = E^i_{\mathrm{apri}} \left(\frac{1}{2} + \frac{1}{\sqrt{2\pi}}\sigma_i + \frac{1}{2} \exp \frac{\sigma_i^2}{2} \operatorname{erfc} \frac{\sigma_i}{\sqrt{2}} \right)$.

Lognormal emissions

$$\mathbf{z}^i_{\mathrm{emis}} := E^i_{\mathrm{apri}} \exp \mathbf{x}^i_{\mathrm{emis}}$$

If $\mathbf{x} \sim \mathcal{N}(0, \mathbf{B})$, then $\mathbf{z} \sim \log \mathcal{N}(E^i_{\mathrm{apri}}, \mathbf{B})$ and $\mathcal{E}\{\mathbf{z}^i_{\mathrm{emis}}\} = E^i_{\mathrm{apri}} \exp \frac{\sigma_i^2}{2}$.

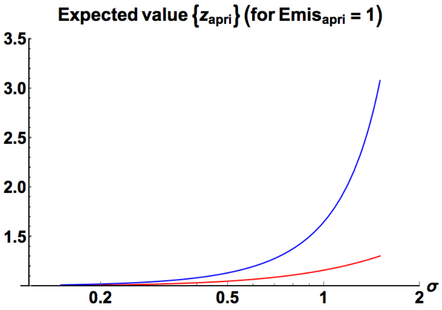
The diagonal elements of $\mathbf{B}$ are now the *relative* per-category emission uncertainties (variances), as defined in the TM5 `rc` file.
NB: $\mathcal{J}$ is defined in terms of $\mathbf{x}$, remains Gaussian (quadratic in $\mathbf{x}$).

TM5-4DVAR: (semi-)lognormal emissions prior

☺ The emission uncertainty space has positive support - a posteriori emissions cannot be negative.

☹ We invert for the mean of the (lognormal) a posteriori PDF. But this mean is **unbounded** wrt the (per-category) emission variance. Our choice of PDF pushes us away from the prior $\mathbf{z}_{\mathrm{apri}}$.

Semi-lognormal vs. lognormal



☹ The cost function $\mathcal{J}$ is no longer quadratic in the *absolute* emissions $\mathbf{z}$. We have to use M1QN3 instead of CONGRAD.

UQ with a nonlinear model: M1QN3 Hessian approximation

Joint Research Centre

Main idea

- What we want: the **leading eigen-pairs** $(\lambda_i, \mathbf{v}_i)$ of the Hessian matrix $\nabla^2_{\boldsymbol{\xi}} \mathcal{J}(\boldsymbol{\xi}_a)$.
- We cannot form the Hessian matrix explicitly (it can have more than $\sim 10^{12}$ elements for a zoomed model run). But we don't need to!
- Let's use a *black-box matrix-free eigenvalue solver* (such as ARPACK, available in `scipy`). We just need to calculate the action of the (approximate) Hessian on a given vector $\mathbf{q}$.
- TM5-4DVAR implementation: a post-inversion step (after we have computed $\mathbf{x}_a$).
- **How can we approximate the Hessian?**

The M1QN3 Hessian approximation

Joint Research Centre

M1QN3

- A limited-memory quasi-Newton optimization routine.
- At iteration k, the algorithm uses the state and gradient vectors from iterations $k-1, \ldots, k-m$ ($m \sim 10$) to build an approximation to $\nabla^2 J(\xi)$.
- ☺ We can efficiently compute the product $\nabla^2_{\mathrm{M1QN3}} J(\xi) \times \mathbf{q}$ ($\mathcal{O}(mn)$ FLOPs).

Finite-difference (FD) Hessian-vector products

Joint Research Centre

Finite difference of adjoints: $\epsilon > 0$ (sufficiently small)

$$\begin{aligned} \boldsymbol{\xi}_a &= \mathbf{L}^{-1}\mathbf{x}_a \\ \nabla^2_{\boldsymbol{\xi}}\mathcal{J}(\boldsymbol{\xi}_a) \times \mathbf{q} &= \frac{1}{2\epsilon}\left[\nabla_{\boldsymbol{\xi}}\mathcal{J}(\boldsymbol{\xi}_a + \epsilon\,\mathbf{q}) - \nabla_{\boldsymbol{\xi}}\mathcal{J}(\boldsymbol{\xi}_a - \epsilon\,\mathbf{q})\right] + \mathcal{O}(\epsilon^2) \end{aligned}$$

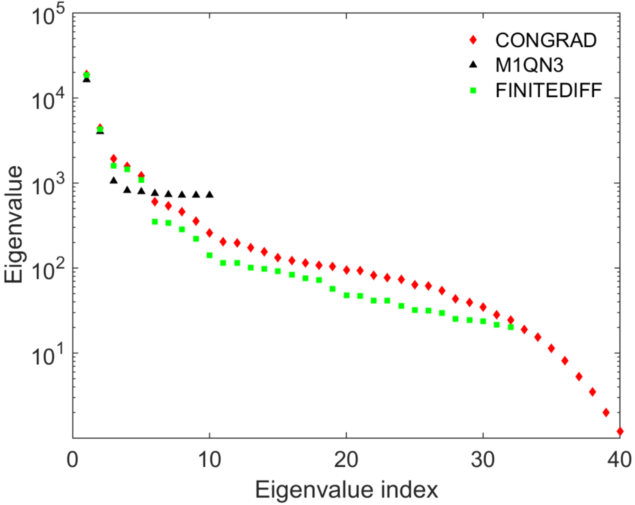
- ☹ Cost: two adjoint integrations per Hessian-vector product.
- ☺ The adjoint integrations can be performed in parallel.

The analysis error covariance

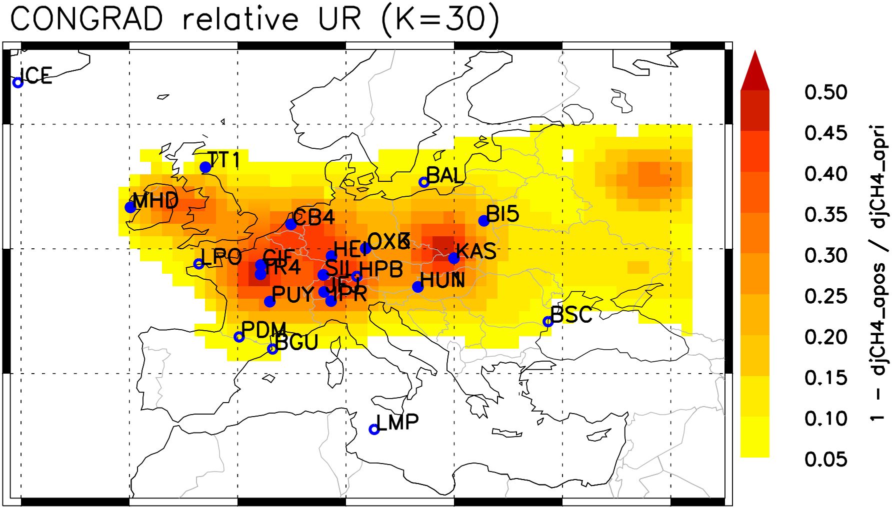
$$\begin{aligned} \mathbf{z}_a &:= h(\mathbf{x}_a) \\ \mathcal{P}^{\mathrm{apos}}(\mathbf{z}_a) &\approx \mathbf{B}_{\mathrm{abs}} + \left(\frac{\mathrm{d}h}{\mathrm{d}\mathbf{x}}\right)\left(\sum_{i=1}^{K}\left(\frac{1}{\lambda_i} - 1\right)(\mathbf{L}\mathbf{v}_i)(\mathbf{L}\mathbf{v}_i)^T\right)\left(\frac{\mathrm{d}h}{\mathrm{d}\mathbf{x}}\right)^T \end{aligned}$$

Hessian-based UQ: largest eigenvalues

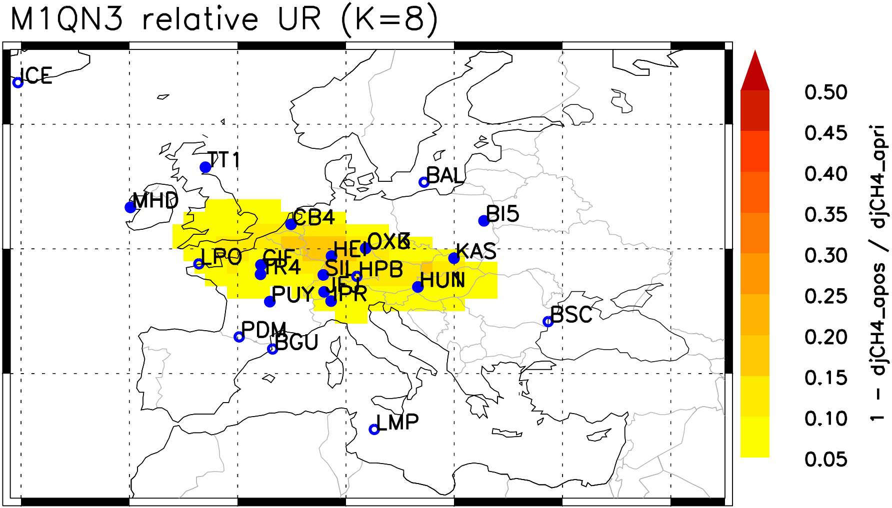
Joint Research Centre



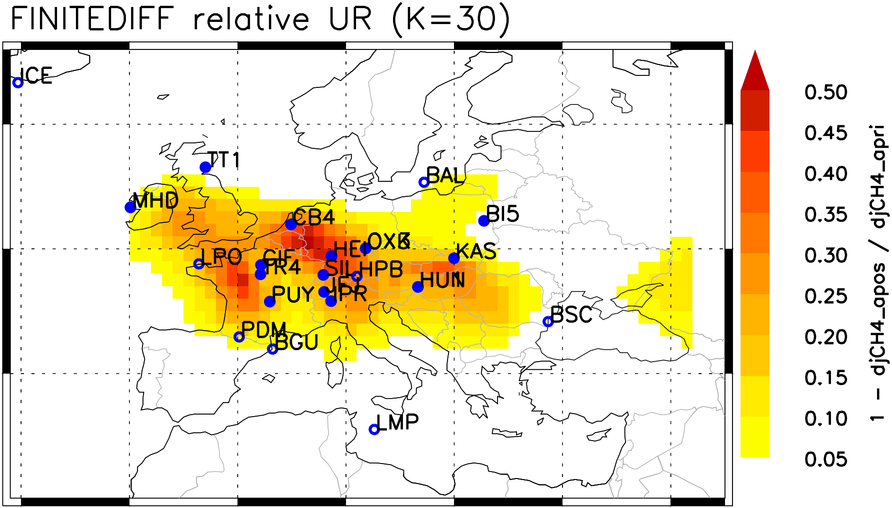
Hessian-based UQ: uncertainty reduction



Hessian-based UQ: uncertainty reduction



Hessian-based UQ: uncertainty reduction



Monte Carlo 4DVAR

Joint Research Centre

The algorithm

1. Generate an ensemble of prior states $\mathbf{x}_b^{(k)} \sim \mathcal{N}(0, \mathbf{B})$, $k = 1 \dots K$.
2. Perturb point observations and satellite total columns:

$$\mathbf{y}^{(k)} = \mathbf{y} + \mathbf{R}^{1/2} \rho^{(k)}, \quad \rho^{(k)} \sim \mathcal{N}(0, \mathbf{I}).$$

3. Run K inversions: each starts from $\mathbf{x}_b^{(k)}$ and assimilates the data vector $\mathbf{y}^{(k)}$. We get the a posteriori ensemble $\mathbf{x}_a^{(k)}$.
4. Calculate a posteriori statistics (mean, variances) from $\mathbf{x}_a^{(k)}$, aggregate (absolute) errors with KPP.

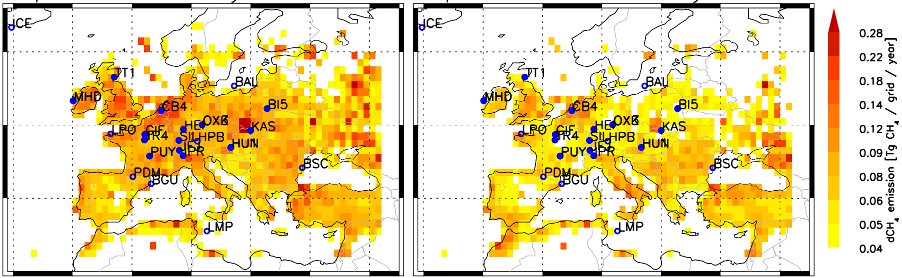
☺ Works with any prior/observation PDF, easy to implement (at least in theory), embarrassingly parallel.

☹ Computationally expensive, converges slowly (rate $\sim \frac{1}{\sqrt{K}}$).

Monte-Carlo UQ: numerical results ($K = 26$)

a priori uncertainty

a posteriori uncertainty



Wrap-up

Joint Research Centre

Computational cost:

Monte Carlo	$K \times \mathrm{Cost}_{\mathrm{4DVAR}}$
M1QN3	very small overhead
Finite differences	$\sim 3 \times K \times [\mathrm{Cost}_{\mathrm{FWD}} + \mathrm{Cost}_{\mathrm{ADJ}}]$

Wrap-up

Computational cost:

Monte Carlo	$K \times \mathrm{Cost}_{\mathrm{4DVAR}}$
M1QN3	very small overhead
Finite differences	$\sim 3 \times K \times [\mathrm{Cost}_{\mathrm{FWD}} + \mathrm{Cost}_{\mathrm{ADJ}}]$

Code status and future work:

- Monte-Carlo and Hessian-based UQ code now in the TM5-JRC SVN tree, to be merged into the SF .hg branch.
- Calculate and compare uncertainty estimates for a multi-year inversion.
- Make Monte-Carlo more HPC-friendly (use MPI with K processes instead of K single-core inversion jobs).
- Invert for the median/mode of the lognormal PDF (previous work by Fletcher and Zupanski, not trivial).

Questions?

The nitty-gritty...

Joint Research Centre

The effect of variable transformations

We are optimizing the emission coefficients $\mathbf{x}$.

What we really want is (the eigen-pairs of) the Hessian wrt the absolute emission variables $\mathbf{z}_{\mathrm{emis}} = h(\mathbf{x})$. After some math...

$$\frac{\partial^2 \mathcal{J}}{\partial \mathbf{z}^2}\mathbf{q} = \left(\frac{\mathrm{d}h}{\mathrm{d}\mathbf{x}}\right)^{-T} \frac{\partial^2 \mathcal{J}}{\partial \mathbf{x}}^2 \left(\frac{\mathrm{d}h}{\mathrm{d}\mathbf{x}}\right)^{-1} \mathbf{q} - \left(\frac{\mathrm{d}h}{\mathrm{d}\mathbf{x}}\right)^{-T} \left(\frac{\mathrm{d}^2 h}{\mathrm{d}\mathbf{x}^2} \otimes \left(\frac{\mathrm{d}h}{\mathrm{d}\mathbf{x}}\right)^{-1} \mathbf{q}\right)^T \left(\frac{\partial \mathcal{J}}{\partial \mathbf{z}}\right)^T$$

Ignoring the second order term (Hessian of h), we get:

$$\mathcal{P}^{\mathrm{apos}}(\mathbf{z}_a) \approx \left(\nabla^2 \mathcal{J}(\mathbf{z}_a)\right)^{-1} \approx \mathbf{B}_{\mathrm{abs}} + \left(\frac{\mathrm{d}h}{\mathrm{d}\mathbf{x}}\right) \left(\sum_{i=1}^{K} \left(\frac{1}{\lambda_i} - 1\right) (\mathbf{L}\mathbf{v}_i)(\mathbf{L}\mathbf{v}_i)^T\right) \left(\frac{\mathrm{d}h}{\mathrm{d}\mathbf{x}}\right)^T$$

Test inversion details

- TM5 version: JRC-M05
- Time period: 12/2009 - 3/2010
- 40 M1QN3 iterations, 40 CONGRAD iterations
- Prior emission uncertainties: 100% for all categories ($\sigma_{\mathbf{x}} = 1$).
- We assimilated surface observations (NOAA) and GOSAT OCPR v52 retrievals.
- Uncertainties aggregated with VPP.