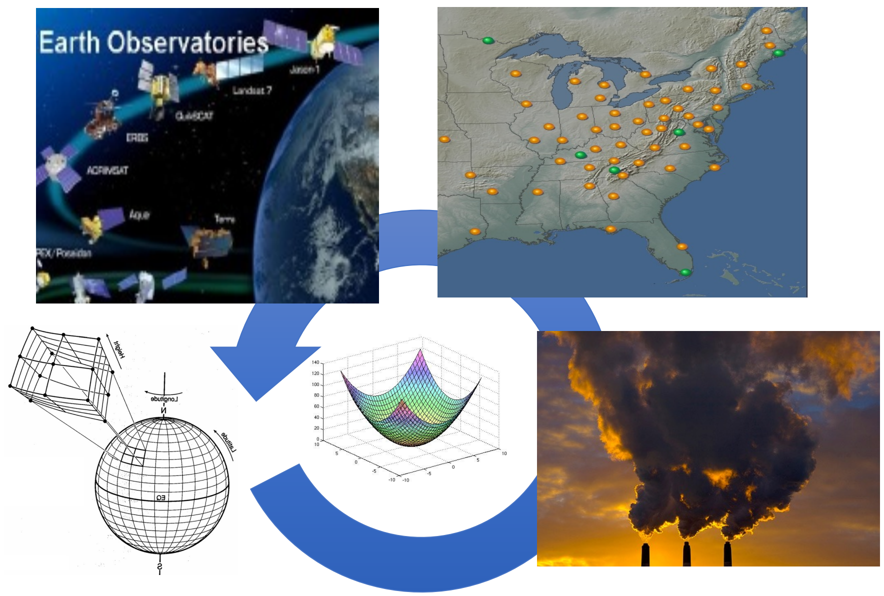
Mathematics of chemical data assimilation and inverse modeling



Daven K. Henze, CU Boulder
Benjamin Worden-Sapper, CU Boulder
Xueying Yu, Stanford University
Nicolas Bousserez, ECMWF
Zhen Qu, NC State

Bayesian data assimilation (DA) methods

- **Model:**

$$\mathbf{x}_k = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}, \boldsymbol{\lambda}) + \boldsymbol{\eta}_k$$

state model param error

- **Observation operator:**

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \boldsymbol{\epsilon}_k$$

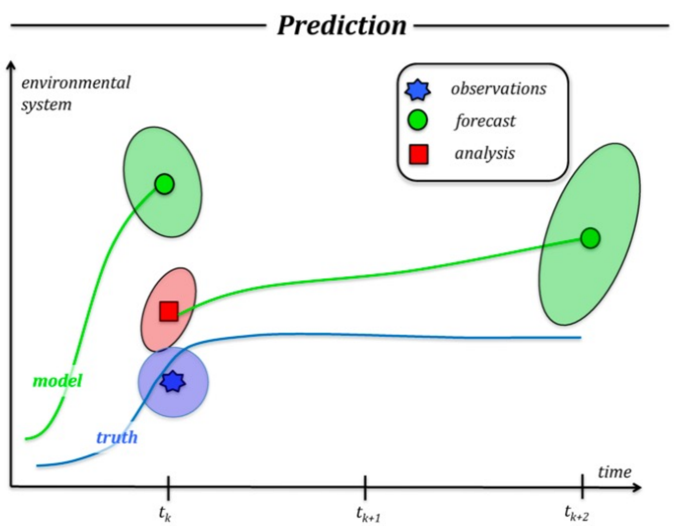
obs state noise

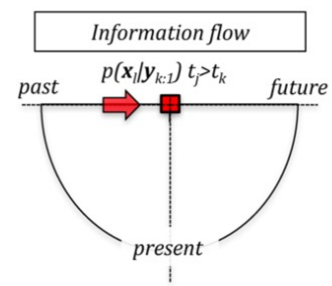
- **Bayes' rule:**

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}$$

- **These 3 components used to derive a wide range of DA and inverse modeling methods**

- **These can be categorized as 3 types of estimates (Wiener, 1949): prediction, filtering, smoothing**





Adopted from Carrassi et al. (2018).

For chemical DA reviews see:

- Sandu and Chai, 10.3390/atmos2030426, 2010
- Bocquet et al., 10.5194/acp-15-5325-2015, 2015
- Elbern et al., 10.1093/acprof:oso/9780198723844.003.0022, 2014

Bayesian data assimilation (DA) methods

- **Model:**

$$\mathbf{x}_k = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}, \boldsymbol{\lambda}) + \boldsymbol{\eta}_k$$

state model param error

- **Observation operator:**

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \boldsymbol{\epsilon}_k$$

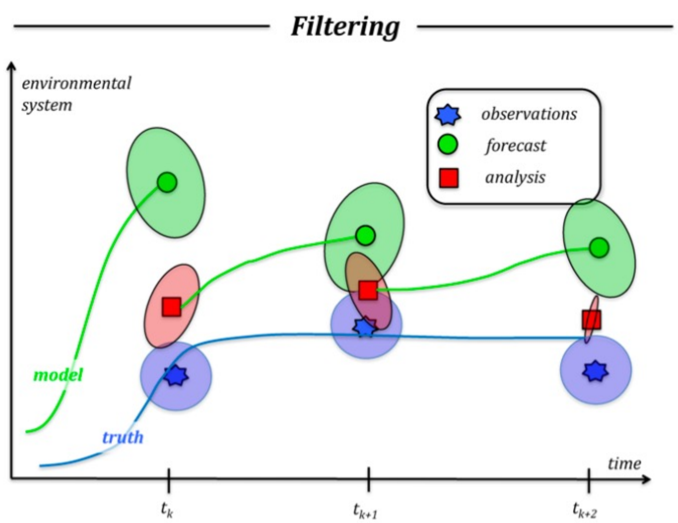
obs state noise

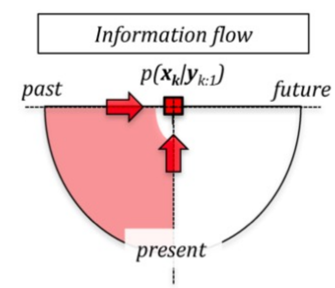
- **Bayes' rule:**

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}$$

- **These 3 components used to derive a wide range of DA and inverse modeling methods**

- **These can be categorized as 3 types of estimates (Wiener, 1949): prediction, filtering, smoothing**





Adopted from Carrassi et al. (2018).

For chemical DA reviews see:

- Sandu and Chai, 10.3390/atmos2030426, 2010
- Bocquet et al., 10.5194/acp-15-5325-2015, 2015
- Elbern et al., 10.1093/acprof:oso/9780198723844.003.0022, 2014

Bayesian data assimilation (DA) methods

- **Model:**

$$\mathbf{x}_k = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}, \boldsymbol{\lambda}) + \boldsymbol{\eta}_k$$

state model param error

- **Observation operator:**

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \boldsymbol{\epsilon}_k$$

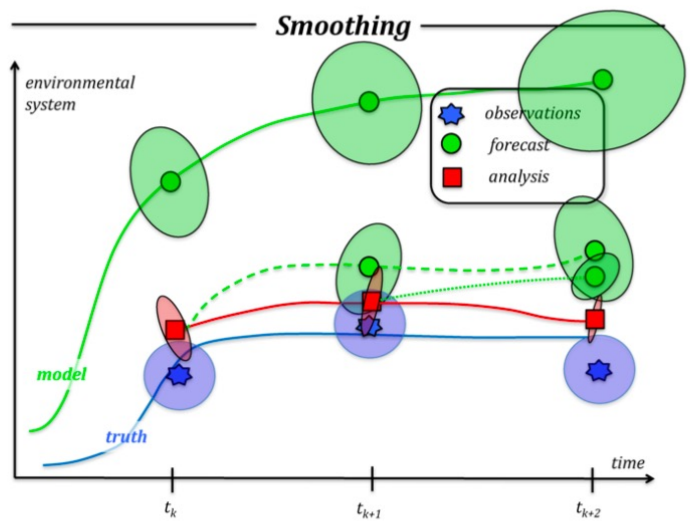
obs state noise

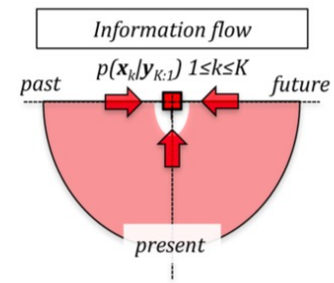
- **Bayes' rule:**

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}$$

- **These 3 components used to derive a wide range of DA and inverse modeling methods**

- **These can be categorized as 3 types of estimates (Wiener, 1949): prediction, filtering, smoothing**





Adopted from Carrassi et al. (2018).

For chemical DA reviews see:

- Sandu and Chai, 10.3390/atmos2030426, 2010
- Bocquet et al., 10.5194/acp-15-5325-2015, 2015
- Elbern et al., 10.1093/acprof:oso/9780198723844.003.0022, 2014

Solving for fully Bayesian solutions (i.e., pdfs) to nonlinear problems

- **Calculations of the complete posterior distribution (not just the mean and covariance)**
  - Particle filters (sequential Monte Carlo)
  - Markov Chain Monte Carlo (MCMC)
  - While most DA approaches require x10 to x100 model runs, these approaches require thousands to millions more simulations, and have thus been limited to low-dimensional systems.

Adopted from Carrassi et al. (2018)

Bayesian data assimilation (DA) methods: solve for features of the solution

- **Kalman filter: sequential update**

$$\mathbf{K}_k = \mathbf{P}^{\mathrm{f}}_k \mathbf{H}^{\mathrm{T}}_k (\mathbf{H}_k \mathbf{P}^{\mathrm{f}}_k \mathbf{H}^{\mathrm{T}}_k + \mathbf{R}_k)^{-1},$$
$$\mathbf{x}^{\mathrm{a}}_k = \mathbf{x}^{\mathrm{f}}_k + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k \mathbf{x}^{\mathrm{f}}_k),$$
$$\mathbf{P}^{\mathrm{a}}_k = (\mathbf{I}_k - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}^{\mathrm{f}}_k.$$

  – Uses prior (P^f) and observation (R) errors, the model forecast (x^f), and linearized (here, though not always) observation operator (H)
  – Explicitly estimates the analysis (x^a) and its error (P^a)
  – Optimal Bayesian solution when the forecast model and observation operator are linear (i.e., B.L.U.E.)
  – Equations above are for Kalman filter, there is also Kalman smoother..

- **Variational: nonlinear minimization to find maximum likelihood estimate**

$$\mathbf{x}^{\mathrm{a}}_{K:0} = \mathrm{argmin}(\mathcal{J}(\mathbf{x}_{K:0})) \qquad k = 1, \ldots, K.$$

$$\mathcal{J}(\mathbf{x}_{K:0}) = \frac{1}{2}\sum_{k=0}^{K} \|\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}} + \frac{1}{2}\sum_{k=1}^{K} \|\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\|^2_{\mathbf{Q}_k^{-1}} + \frac{1}{2}\left\|\mathbf{x}_0 - \mathbf{x}^{\mathrm{b}}\right\|^2_{\mathbf{B}^{-1}}$$

  – Also includes background (B), and model error (Q, i.e. model is a "weak constraint")
  – Posterior error can be estimated as Hessian at minimum, though some care required for numerical accuracy (Bousserez et al., 2015)

Adopted from Carrassi et al. (2018)

Solving for Bayesian features: implementation methods

For most systems, directly calculating all terms in the KF or the analytic solution to minimum of $J(x)$ (possible for linear systems) is not computationally feasible

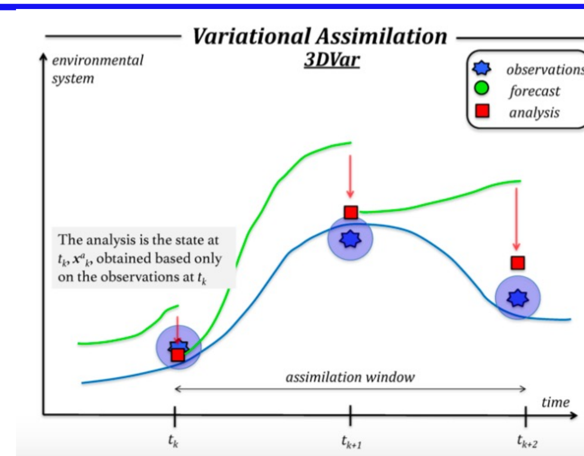
- **Ensemble methods for Kalman filters and smoothers**
  - Stochastic (e.g., Evensen, 2003) and deterministic square-root (e.g., Bishop et al., 2001; Anderson (2001) formulations
  - Many variations (ETKF, LETKF, MLEF, EAKF,…)
  - Require careful localization and inflation
  - Readily implemented in parallel with little modification to the standard forward model

- **Adjoint-based methods for variation solutions**
  - Tangent linear and adjoint models used to calculate dJ/dx
  - Deriving and maintaining these models is non-trivial
  - Use strong-constraint formulation to keep dimension of x manageable (plus Q is hard to specify)

- **Hybrid methods**
  - EnKF with hybridization of static and dynamic error covariance
  - Ensemble of variational systems (EDA)
  - Use of ensemble to estimate the tangent linear and adjoint models of 4D-Var (4D-EnVar)

Data assimilation and inverse modeling: considerations for chemical systems

- **Diversity of scales and sources of uncertainty**
  - Short-lived, highly non-linear species (e.g. NOx, secondary aerosol), strongly impacted by chemistry and emissions
  - Short-lived, linear (e.g., primary aerosol), strongly impacted by microphysics, sub-grid and BL dynamics
  - Long-lived, linear (e.g., CO_2, CH_4) strongly impacted by initial / boundary conditions and large-scale model transport error

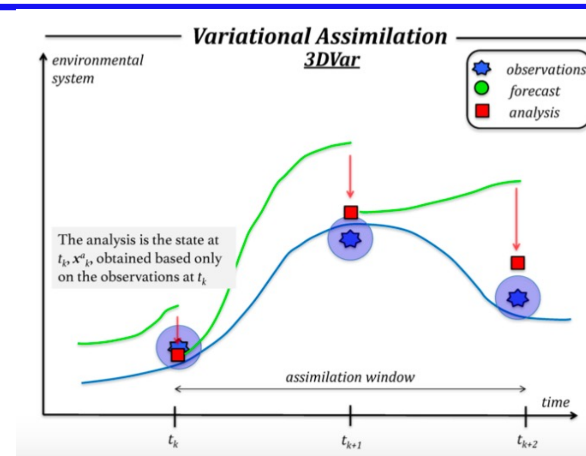


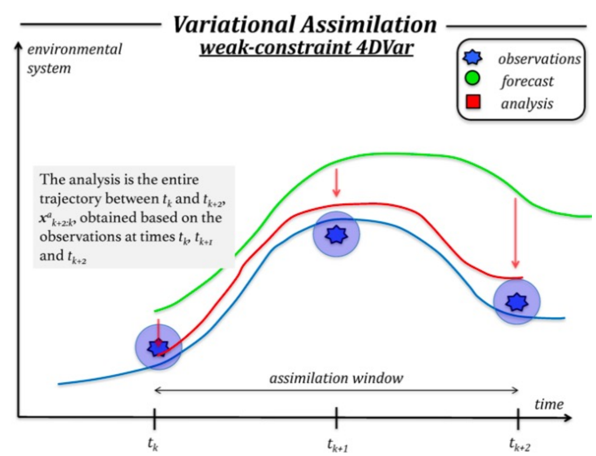
Carrassi et al. (2018)

Data assimilation and inverse modeling: considerations for chemical systems

- **Diversity of scales and sources of uncertainty**
  - Short-lived, highly non-linear species (e.g. NOx, secondary aerosol), strongly impacted by chemistry and emissions
  - Short-lived, linear (e.g., primary aerosol), strongly impacted by microphysics, sub-grid and BL dynamics
  - Long-lived, linear (e.g., CO_2, CH_4) strongly impacted by initial / boundary conditions and large-scale model transport error

- **Importance of emissions**
  - Recent emissions may nearly completely govern the observed state for short-lived species
  - Recent emissions may only barely govern the observed state for long-lived species, but this is still what we care about most
  - Often not normally distributed
  - Most not evolved in time with the atmospheric model
  - **What do we actually seek for when we "solve for emissions"?**

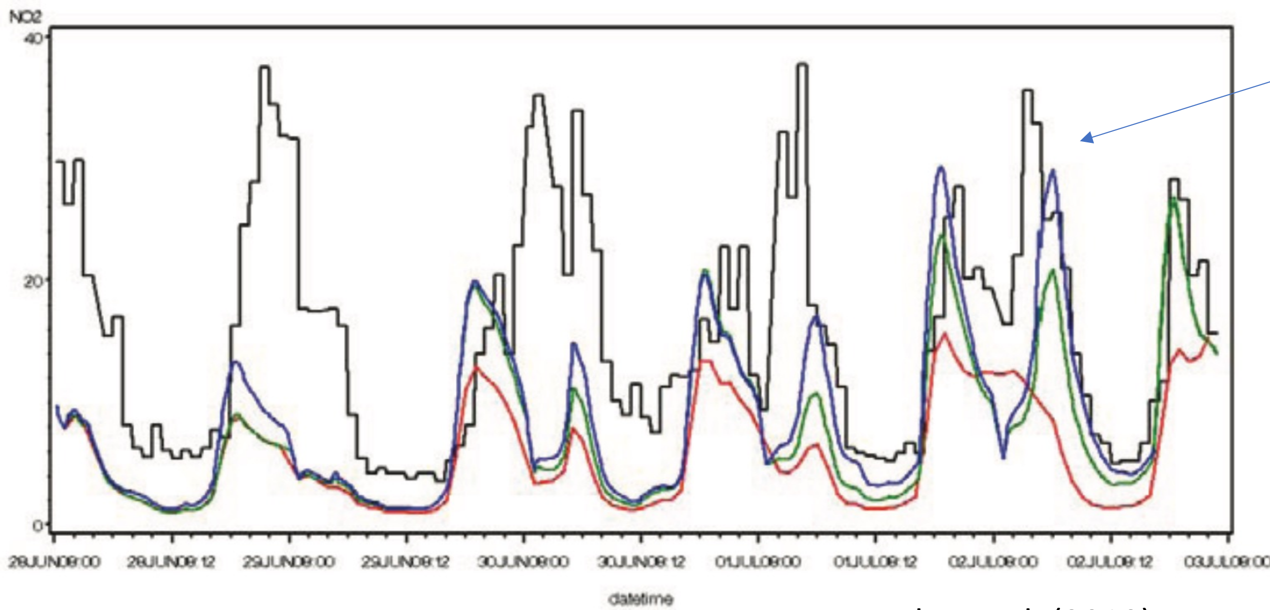




Carrassi et al. (2018)

Estimating emissions: short-lived species

For short-lived species, adjusting emissions at high time-resolution provides performance similar to what might be expected from weak constraint 4D-Var



Adjustments to diurnal variability of NOx emissions allows for this peak to be better simulated

- observed NO_2 [ug/m3]
- prior model
- posterior model

Resler et al. (2010)

Care needs to be taken to not over-optimize emissions to compensate for transport error.

Relatedly, high time-resolution observations (e.g., TEMPO) could help improve transport (e.g., Liu et al., 2021).

Estimating emissions: long-lived species

Remote sensing observations of column CO_2, CH_4 are often not dominated by local / immediate sources → this creates challenges for using these datasets for constraining emissions in short (i.e., less than decadal) or regional inversions

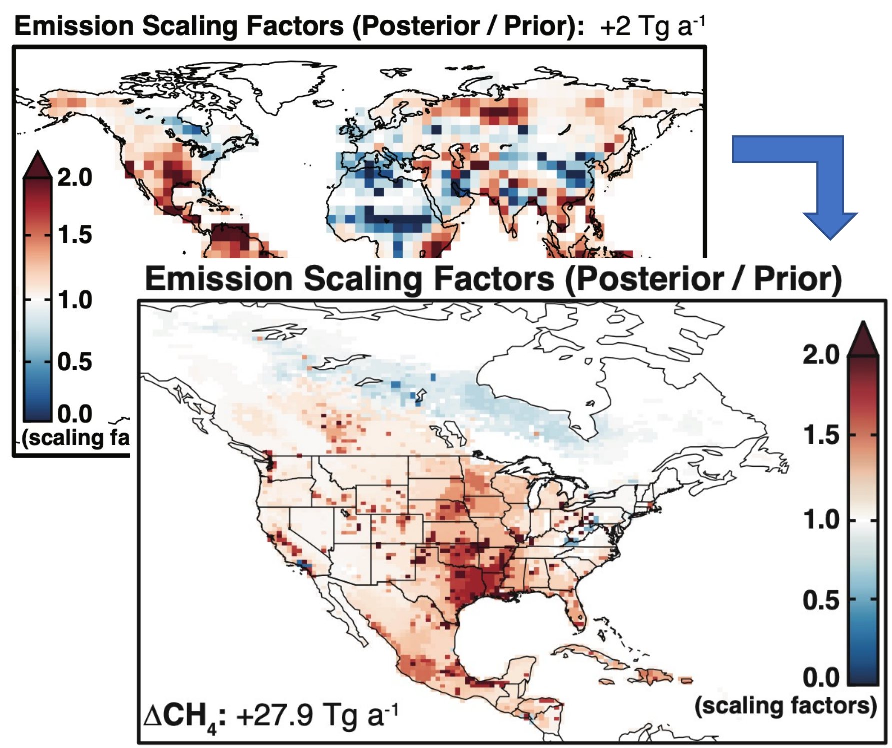
Approaches for regional CH_4 inversions:

- Iteratively optimize emissions vs boundary conditions (e.g., Wecht et al., 2014)
- Begin with prior emissions estimates from a previous global-scale inversion (Turner et al., 2015)

Approaches for global CO_2 inversions:

- Begin with initial conditions constrained by simplified method (e.g., 3D-Var) or based on in situ measurements (Deng et al., 2014)
- Begin from posterior concentration ***and*** error estimates from prior inversion with PvKF as starting point for 4D-Var (Voshtani et al., PhD Thesis, 2022)

Tracer correlations: observed co-emitted plumes of short-lived species (e.g., Kuhlmann et al., 2021)



Turner et al. (2015)

Estimating emissions: updates to the 4D-Var methods

- **Can we reduce the dimension of emissions vector for linear problems for increased efficiency and operational implementation? → Randomization methods (Bousserez et al., 2018; 2020)**
- **Can we estimate the error in our solution when using sub-optimal methods? → Worden-Sapper et al., in prep.**
- **Can we solve problems where our prior emissions are biased? → Yu et al. (2021)**
- Can we speed up highly non-linear inversions? →Hybrid inversions (Qu et al., 2017; Choi et al., 2022)
- **Can we make the inversion results for nonlinear systems more useful? → Sector based inversions (Qu et al., 2022)**

Dimension reduction

Motivation:

- Most inverse problem solutions lie in a small subspace (dimension $k<<n$) where the data are informative.
- Solving for smaller problems allows **fast computation** of the solution (sometimes analytically).
- Small problems allow us to **relax assumptions** on linearity and/or distributions and explore full posterior distributions (e.g., MCMC algorithm)

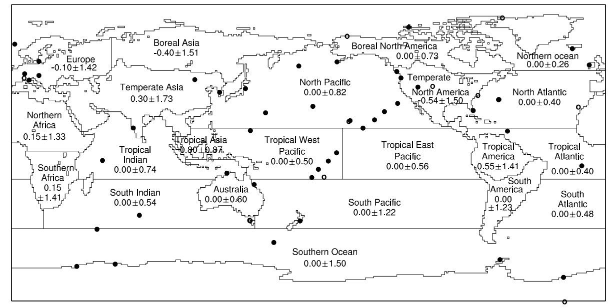
Problem:

- Optimizing the reduction requires access to **second-order information** (e.g., posterior error covariance)
- Difficulty is finding **scalable** algorithms.

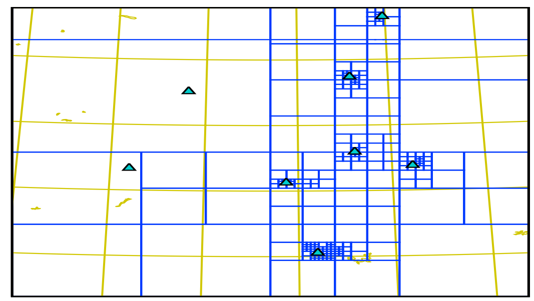
Bousserez

Previous studies

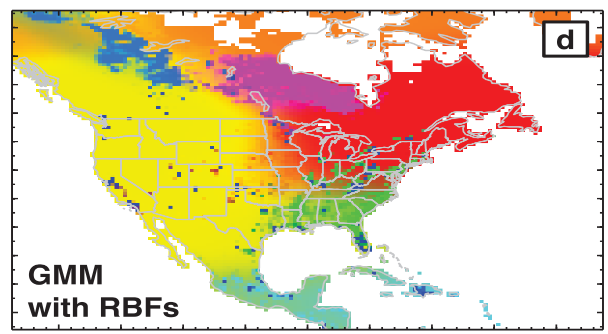
<u>TransCom</u> (Gurney et al., 2003):
- Geographical criteria.
- Allows analytical inversion.
- **Suboptimal**: no optimization w.r.t information content.



<u>Multi-scale</u> (Bocquet et al., 2011):
- Aggregation-based method.
- **Optimal**: maximizes DOFS of the inversion.
- **Not scalable**: requires the model Jacobian.



<u>Clustering</u> (Turner and Jacob, 2015):
- Gaussian-Mixture model exploiting prior information to construct bases.
- **Optimal**: <u>but</u> optimality obtained only for this specific class of basis functions.
- **Weak scalability**: incremental method where posterior errors are reevaluated for each added basis.

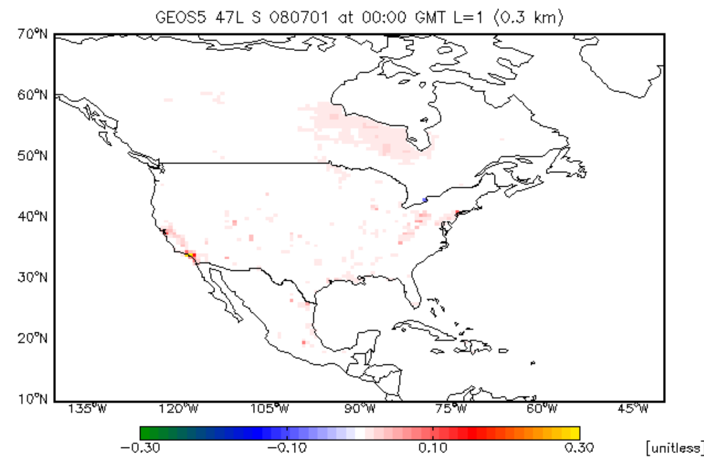


Bousserez

Previous studies

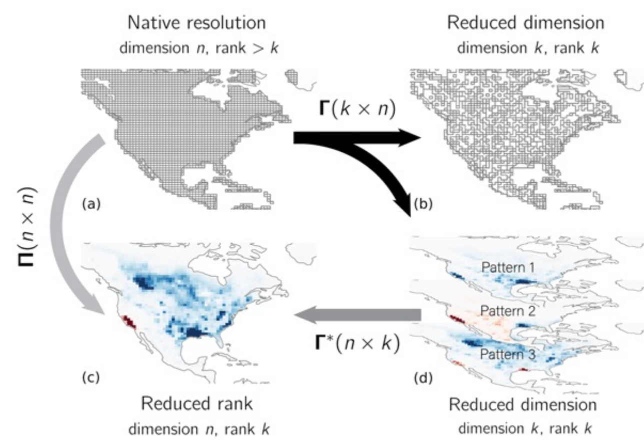
<u>Optimal projections scalably calculated for high-dimensional problems</u> (Bousserez and Henze, 2018):

- Maximizes DOFs of the inversion (Bocquet et al., 2011)
- Optimal basis from SVD of the prior-preconditioned Hessian, $\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$ (Spantini et al., 2015). Note: H from here out is Jacobian of linearized model
- SVD calculated scalably using randomization techniques (Halko et al., 2011; Bui-Thanh et al., 2012)



<u>Reduced-cost consruction of Jacobian matrices</u> (Nesser et al., 2021):

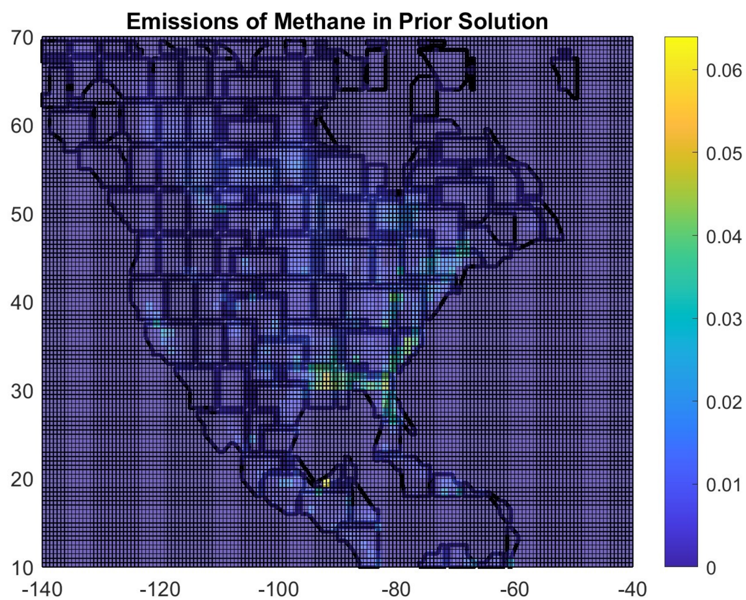
- Approximate Jacobian calculated with adjoint-free methods
- SVD of prior-preconditioned Hessian computed directly
- Error in the “optimal” projection solution owing to the approximate Jacobian not well known



Error estimation for sub-optimal aggregation schemes

Consider an inversion conducted on a non-optimally aggregated grid (or using a reduced-rank Jacobian):

non-optimal aggregation (P) for *k* = 125



Prior emission, full Jacobian from Nesser et al. (2021)

How much error is there in our solution ($x^a_{proj,k}$) compared to using the optimal rank *k* solution ($x^a_{opt,k}$)?

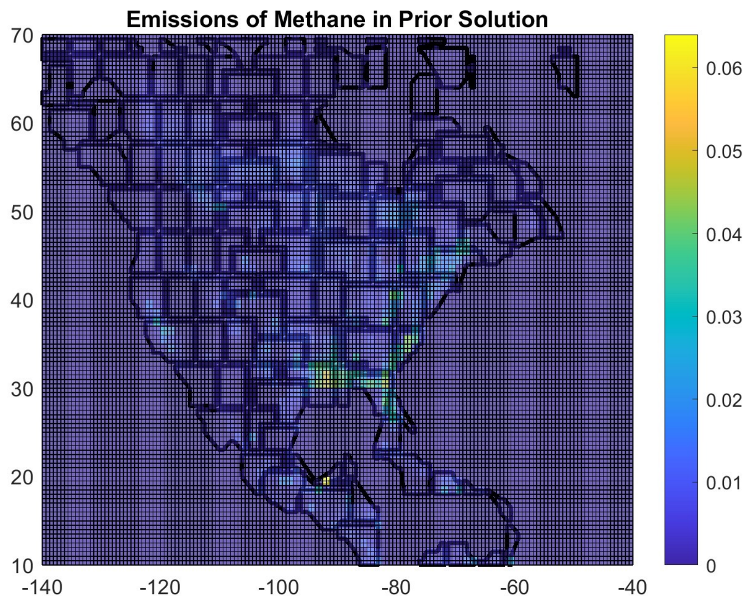
$$\mathbb{E}||x^a_{opt,k} - x^a_{proj,k}||^2_{B^{-1}} = ?$$

Benjamin Worden-Sapper (in prep.)

Error estimation for sub-optimal aggregation schemes

Consider an inversion conducted on a non-optimally aggregated grid (or using a reduced-rank Jacobian):

non-optimal aggregation (P) for *k* = 125



Prior emission, full Jacobian from Nesser et al. (2021)

How much error is there in our solution ($x^a_{proj,k}$) compared to using the optimal rank *k* solution ($x^a_{opt,k}$)?

$$\mathbb{E}||x^a_{opt,k} - x^a_{proj,k}||^2_{B^{-1}} = ?$$

Considering the symmetric projection with similar range, $\tilde{P}$

$$P = B\Gamma^T(\Gamma B\Gamma^T)^{-1}\Gamma$$
$$\tilde{P} = B^{1/2}\Gamma^T(\Gamma B\Gamma^T)^{-1}\Gamma B^{1/2}$$

we can estimate the error owing to sub-optimality as:

$$= ||(I - \tilde{P})B^{1/2}H^T(HBH^T + R)^{-1}HB^{1/2}||_F$$
$$= \sum_{i=1}^{n} s_i^2(1 - u_i^T\tilde{P}u_i)$$

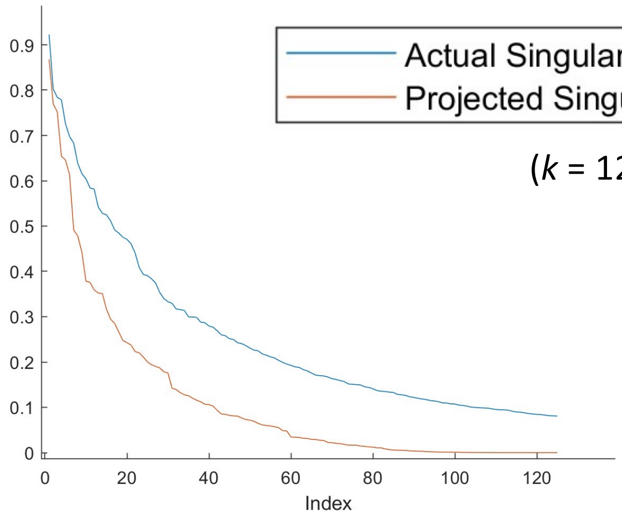
where u_i are left singular vectors of Q

Benjamin Worden-Sapper (in prep.)

Error estimation for sub-optimal aggregation schemes

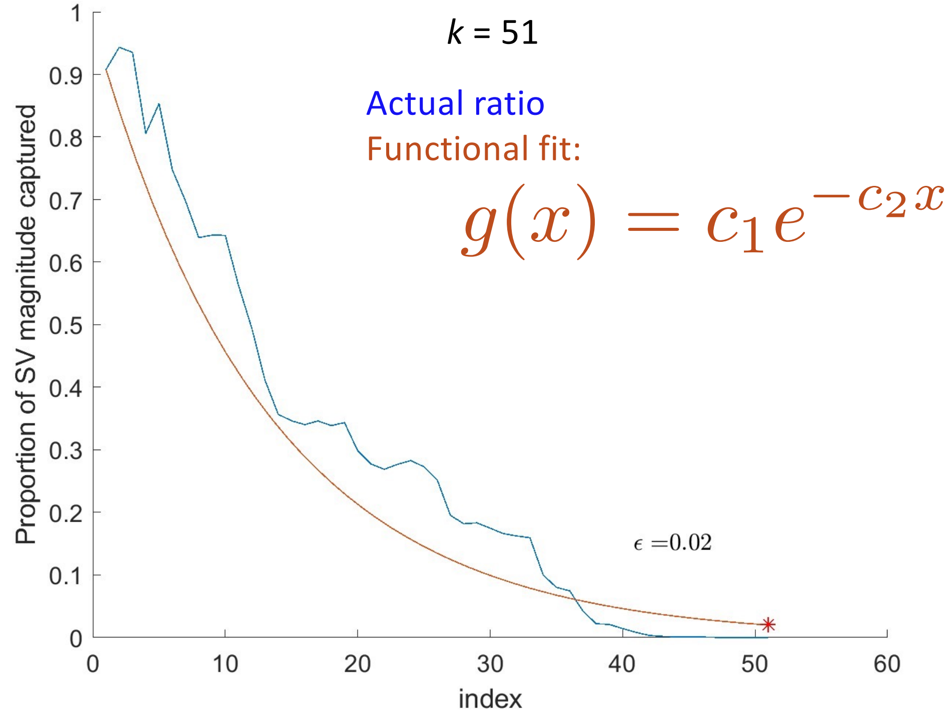
We may not know the SVD of the prior preconditioned Hessian (if we did, we would use it…).

But we do know the SVD of our projected problem. Compared to the full SVD, it is lower. But functionally lower in a way that depends only weakly on *k*.



From this ratio, we can estimate the full SVD from our SVD of the low-rank problem.

$$\frac{s(\tilde{P}Q)}{s(Q)} = g(x)$$

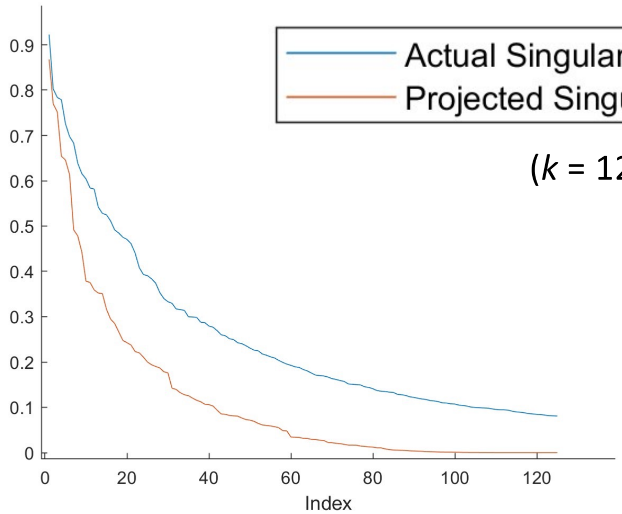


Benjamin Worden-Sapper (in prep.)

Error estimation for sub-optimal aggregation schemes

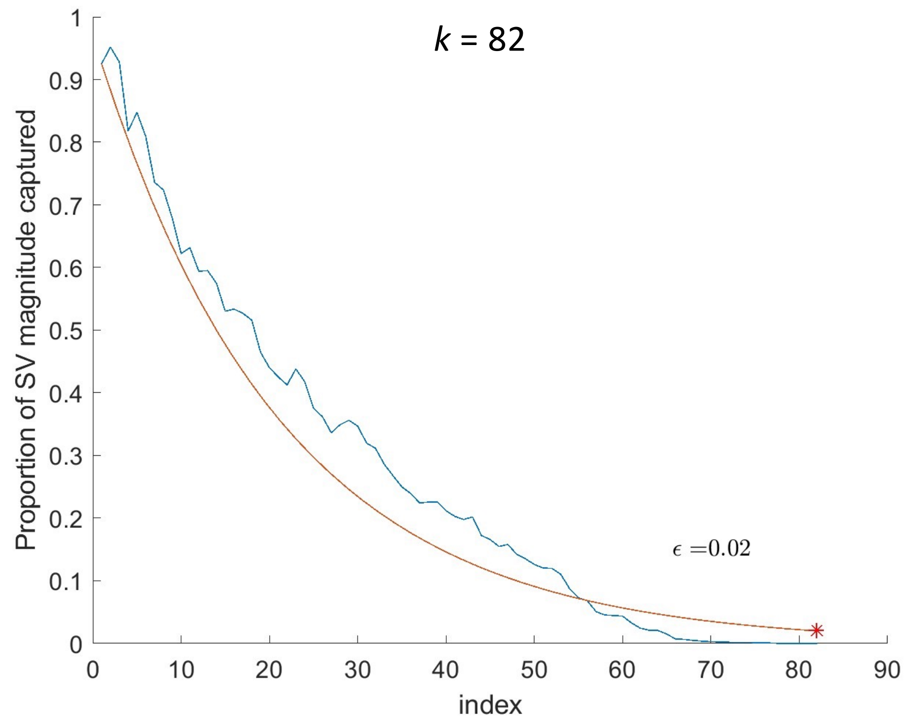
We may not know the SVD of the prior preconditioned Hessian (if we did, we would use it…).

But we do know the SVD of our projected problem. Compared to the full SVD, it is lower. But functionally lower in a way that depends only weakly on *k*.



From this ratio, we can estimate the full SVD from our SVD of the low-rank problem.

$$\frac{s(\tilde{P}Q)}{s(Q)} = g(x)$$

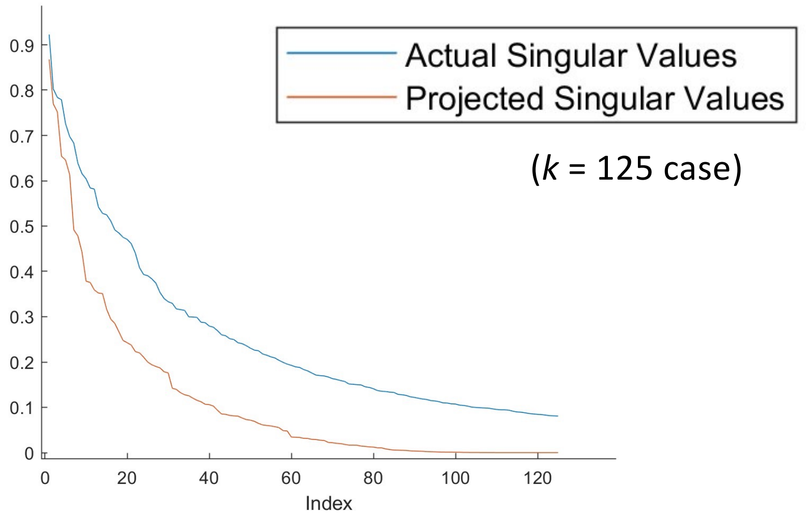


Benjamin Worden-Sapper (in prep.)

Error estimation for sub-optimal aggregation schemes

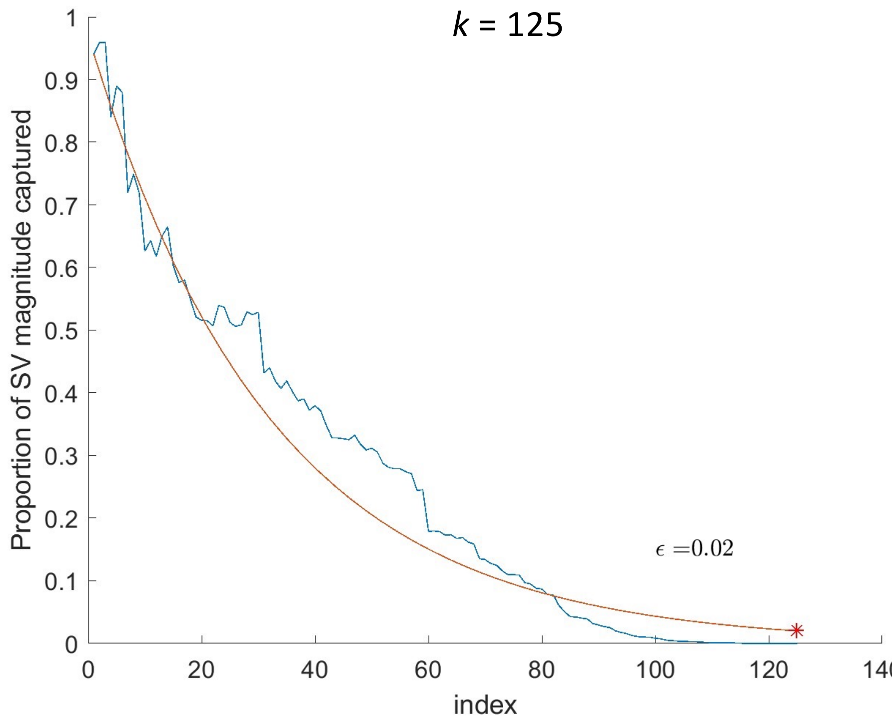
We may not know the SVD of the prior preconditioned Hessian (if we did, we would use it…).

But we do know the SVD of our projected problem. Compared to the full SVD, it is lower. But functionally lower in a way that depends only weakly on *k*.



From this ratio, we can estimate the full SVD from our SVD of the low-rank problem.

$$\frac{s(\tilde{P}Q)}{s(Q)} = g(x)$$

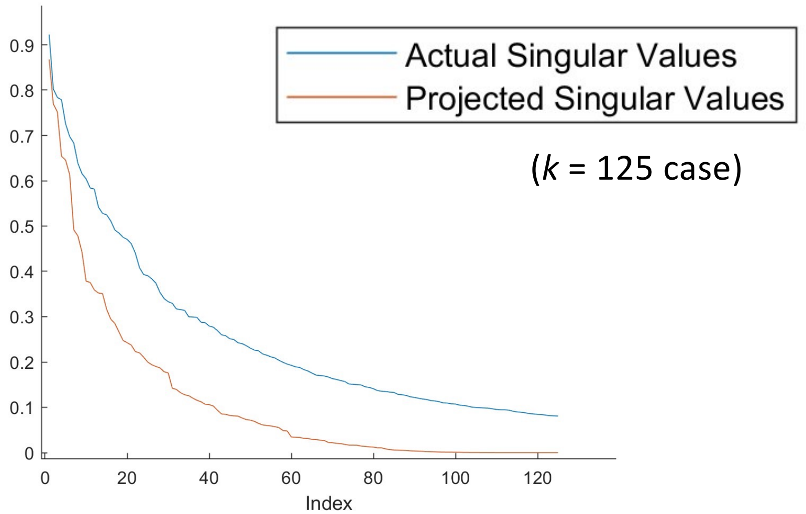


Benjamin Worden-Sapper (in prep.)

Error estimation for sub-optimal aggregation schemes

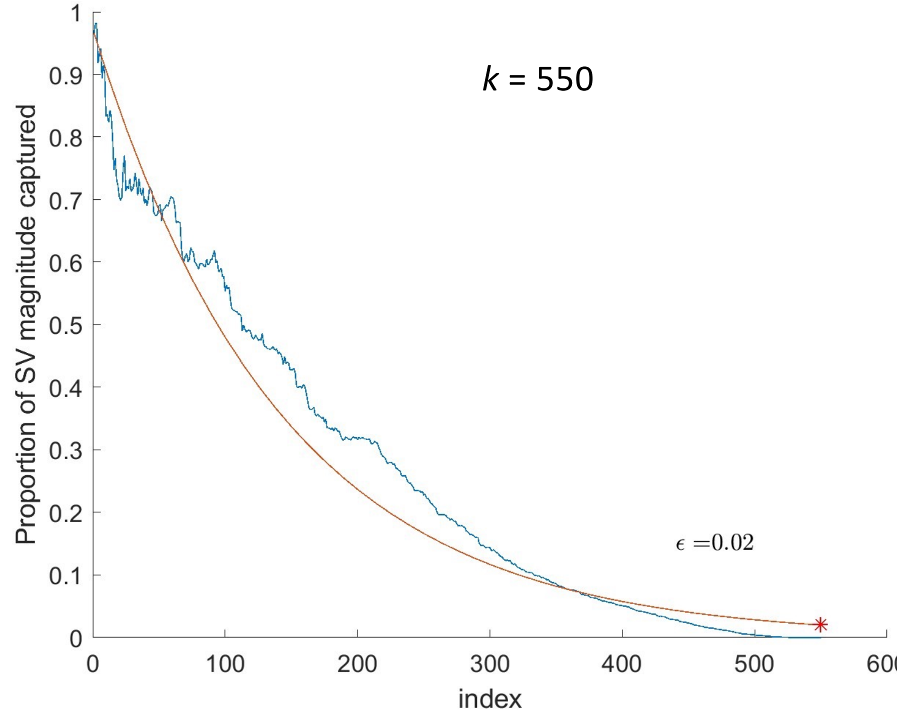
We may not know the SVD of the prior preconditioned Hessian (if we did, we would use it…).

But we do know the SVD of our projected problem. Compared to the full SVD, it is lower. But functionally lower in a way that depends only weakly on *k*.



From this ratio, we can estimate the full SVD from our SVD of the low-rank problem.

$$\frac{s(\tilde{P}Q)}{s(Q)} = g(x)$$



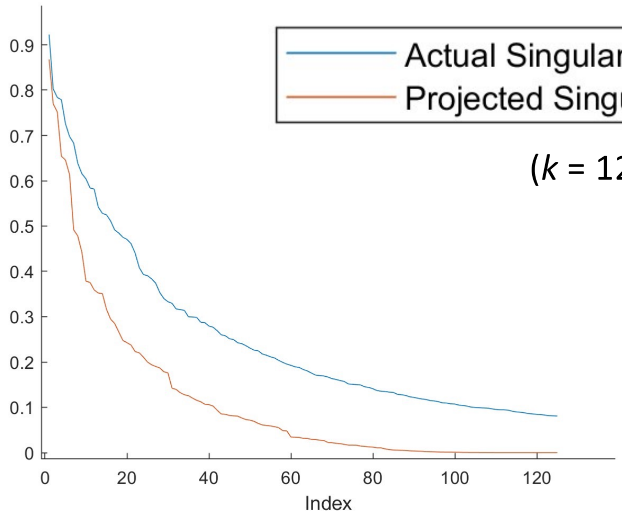
Benjamin Worden-Sapper (in prep.)

Error estimation for sub-optimal aggregation schemes

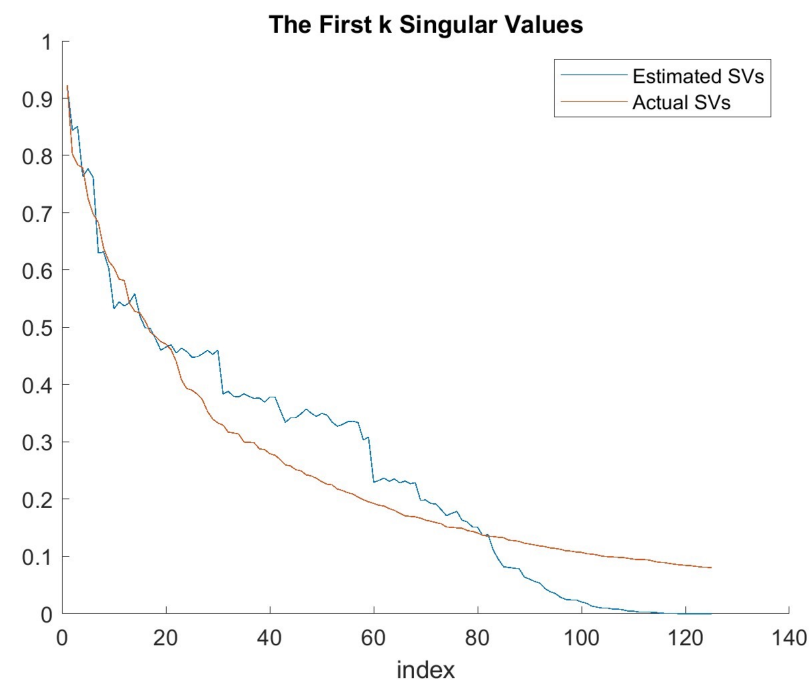
We may not know the SVD of the prior preconditioned Hessian (if we did, we would use it…).

$$\frac{s(\tilde{P}Q)}{s(Q)} = g(x)$$

But we do know the SVD of our projected problem. Compared to the full SVD, it is lower. But functionally lower in a way that depends only weakly on *k*.



(*k* = 125 case)



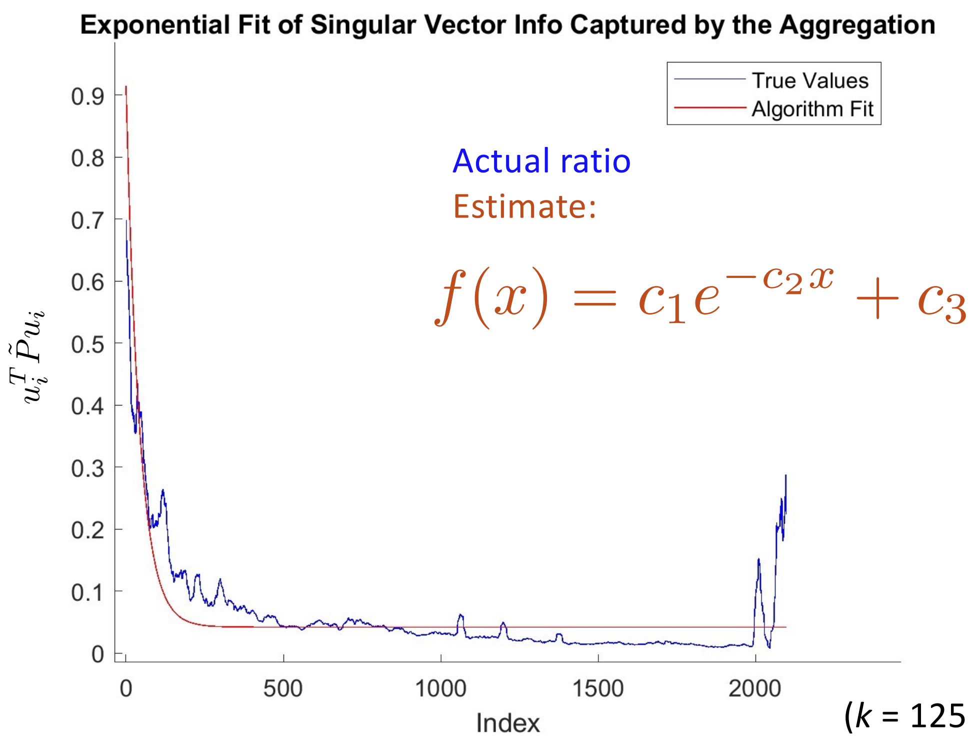
From this ratio, we can estimate the full SVD from our SVD of the low-rank problem.

Benjamin Worden-Sapper (in prep.)

Error estimation for sub-optimal aggregation schemes

We can also functionally estimate the $u_i^T \tilde{P} u_i$ terms using the following:

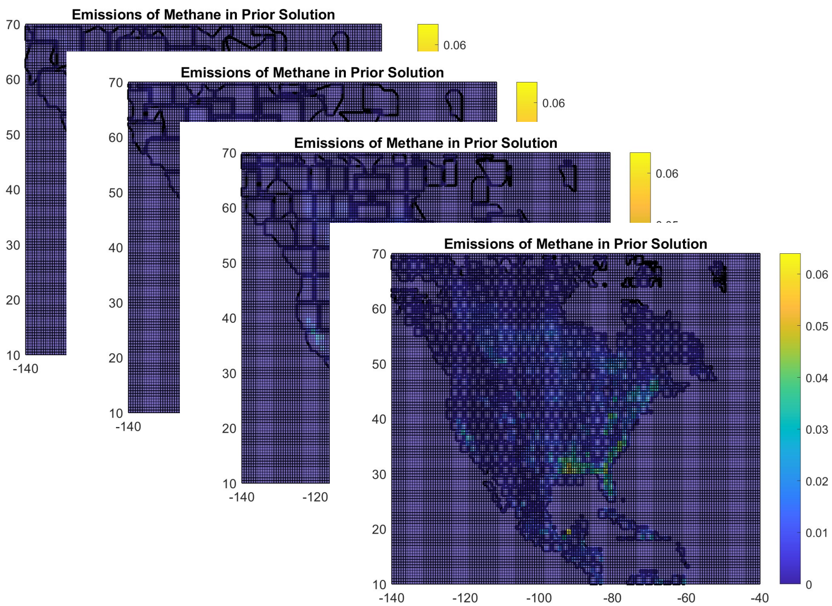
- $\int_0^n f(x)dx = k$
- upper bound for c_3 is k/n (assuming u_i and P uncorrelated)
- estimate leading singular vector and value (e.g., Liao and Sandu) or average of $u_i^T \tilde{P} u_i$

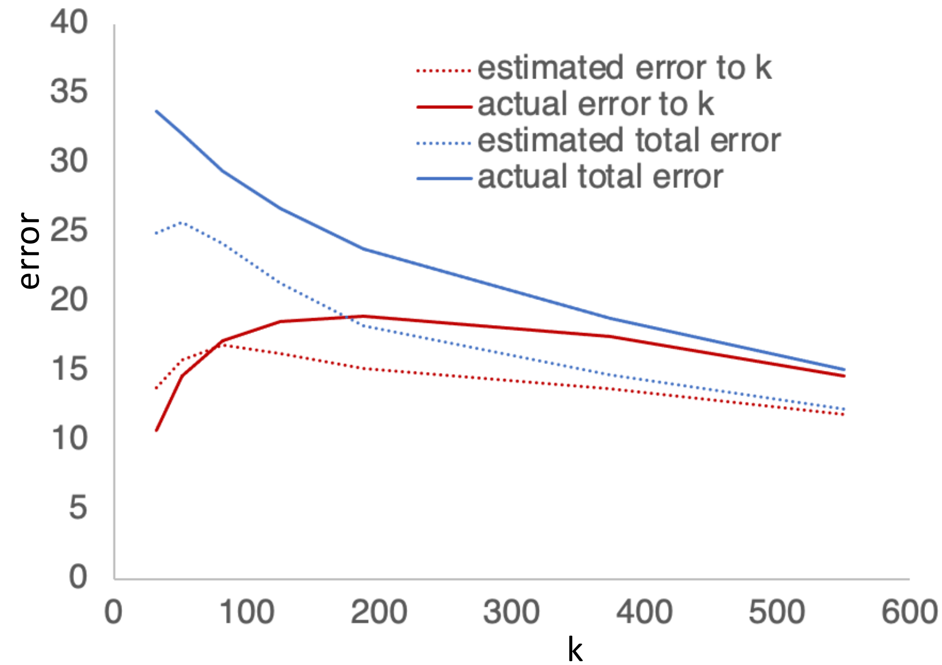


(k = 125 case)

Benjamin Worden-Sapper (in prep.)

Error estimation for sub-optimal aggregation schemes

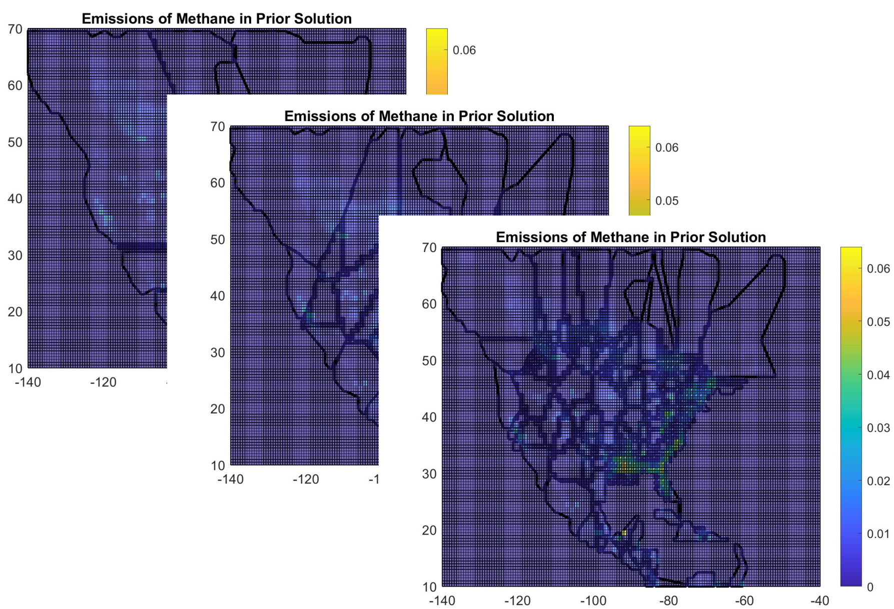


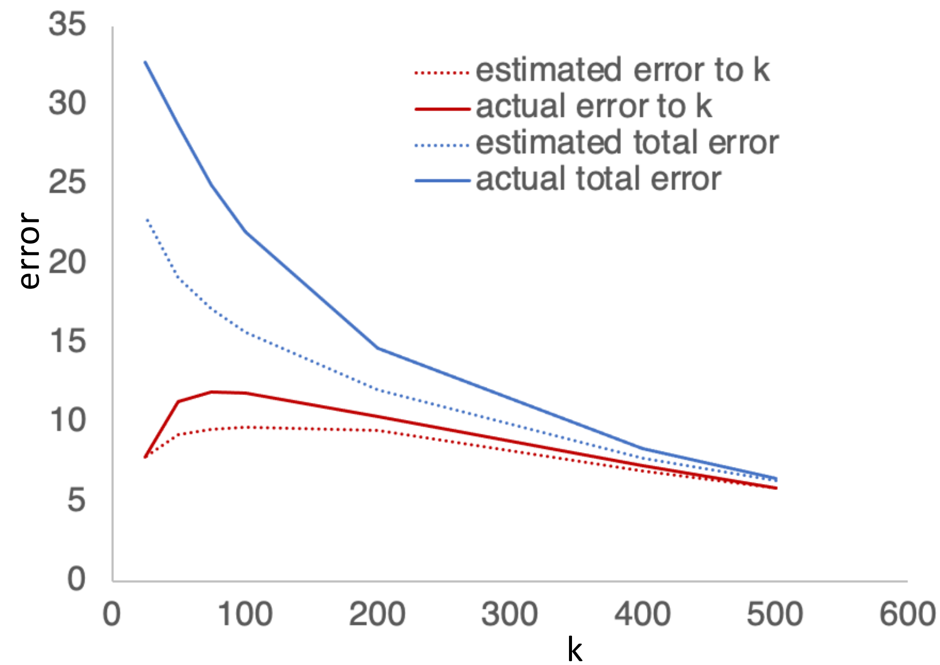


- approach closely estimates the error up to *k* as well as total
- tested for a variety of sub-optimal aggregation schemes, and those which use prior (e.g., GMM model of Turner and Jacob (2015)) have lower error
- Note: projection error is also equal to the DOF "missed" owing to the sub-optimal projection

Benjamin Worden-Sapper (in prep.)

Error estimation for sub-optimal aggregation schemes



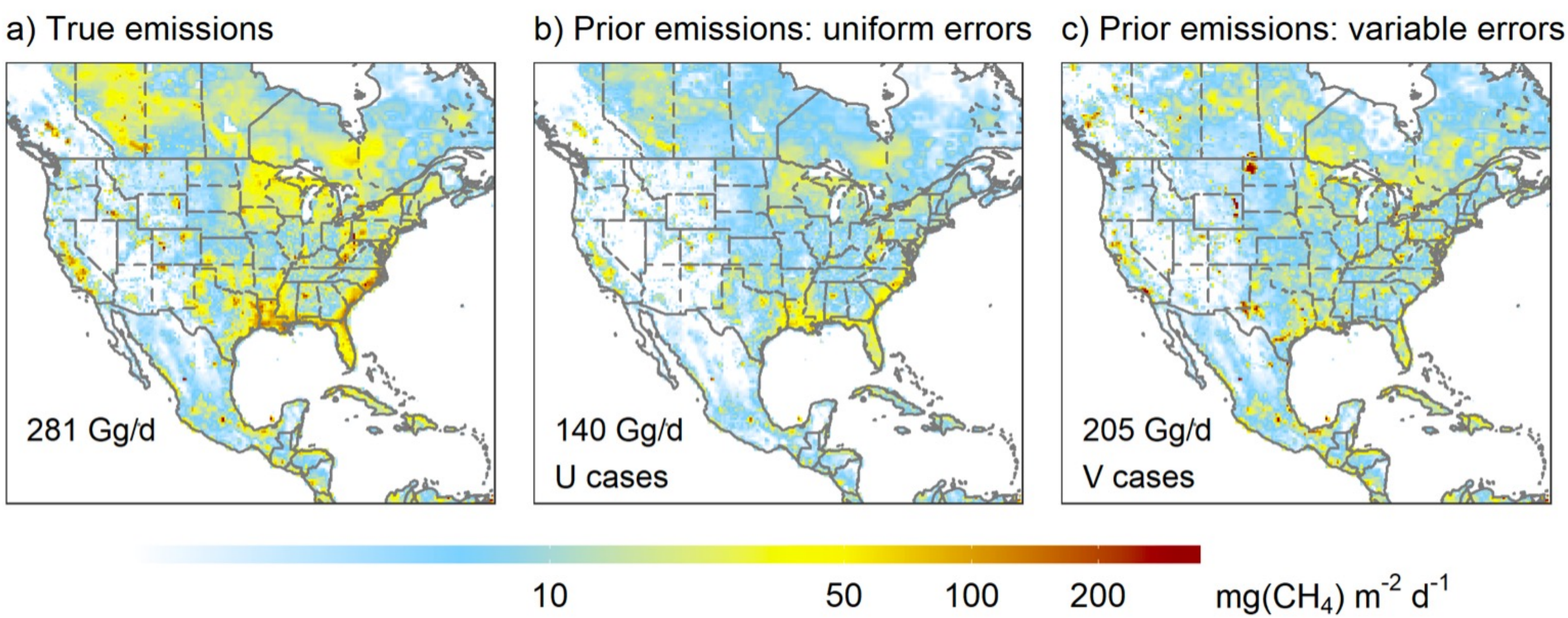


- approach closely estimates the error up to *k* as well as total
- tested for a variety of sub-optimal aggregation schemes, and those which use prior (e.g., GMM model of Turner and Jacob (2015)) have lower error
- Note: projection error is also equal to the DOF "missed" owing to the sub-optimal projection

$$\mathbb{E}||x^a_{opt,k} - x^a_{proj,k}||^2_{B^{-1}} = Tr(A_{opt,k} - A_{proj,k})$$

Benjamin Worden-Sapper (in prep.)

Inverse modeling when the prior emissions are biased

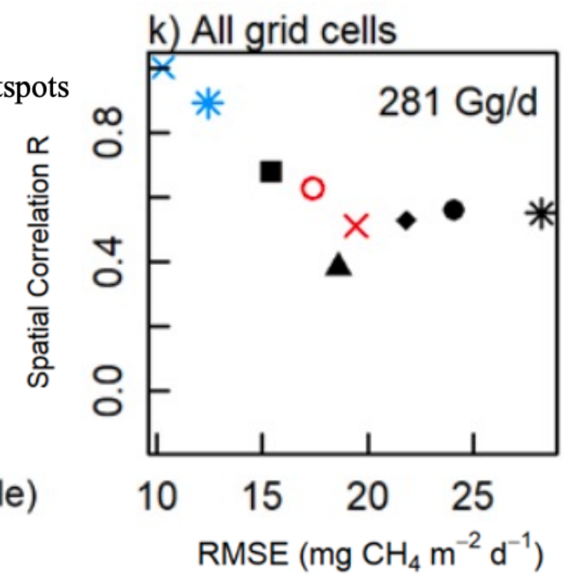


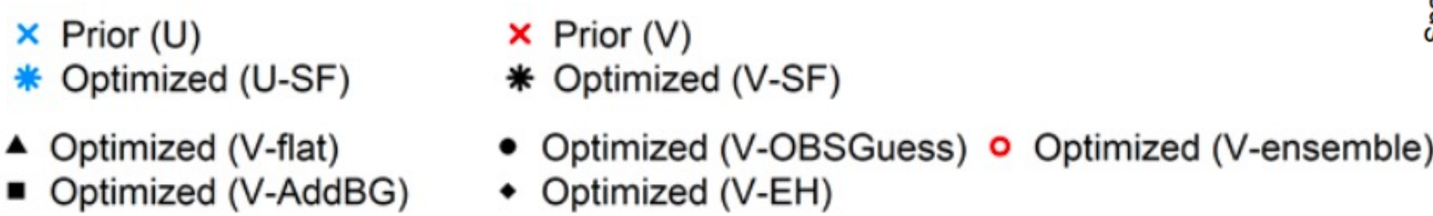
Yu et al. (2021)

Inverse modeling when the prior emissions are biased

Consider multiple ways of defining the control vector (x) for emissions adjustments in TROPOMI CH4 OSSE

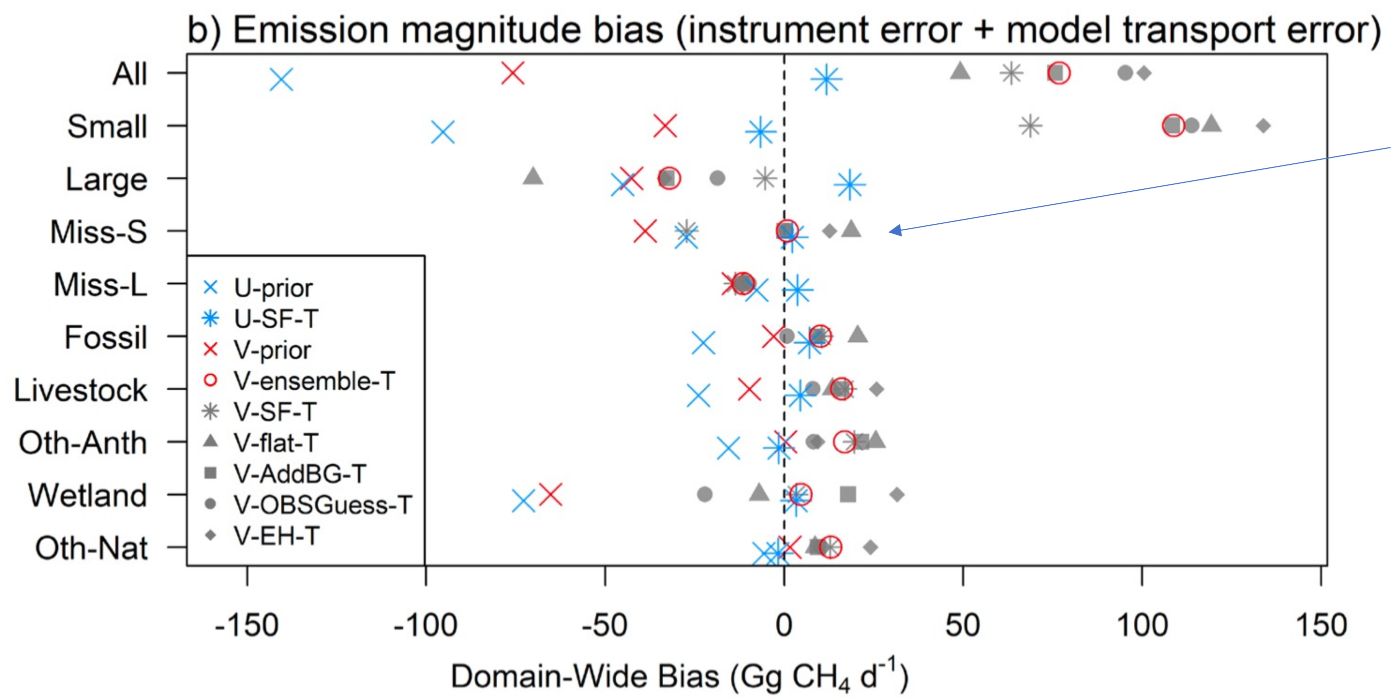
V-SF	Base-case SF	$\boldsymbol{x} = \boldsymbol{s} \circ \boldsymbol{x_a}$	Explore influence of spatial emission errors on base-case SF inversion
V-flat	Flat prior	$\boldsymbol{x} = x_{a\_ave}\, \boldsymbol{s}$	Identify constraints solely from TROPOMI without bottom-up knowledge
V-AddBG	Background increment	$\boldsymbol{x} = \boldsymbol{s} \circ (0.5\, \boldsymbol{x_a} + 0.5\, x_{a\_ave})$	Identify missing sources
V-OBSGuess	Observational guess	$\boldsymbol{x} = \boldsymbol{s} \circ (\boldsymbol{x_a} + \boldsymbol{x_{ObsGuess}})$	Resolve and optimize emission hotspots
V-EH	Enhancement	$\boldsymbol{x} = x_{inc}\, \boldsymbol{s} + \boldsymbol{x_a}$	Identify missing sources





Yu et al. (2021)

Inverse modeling when the prior emissions are biased



Consider the case where the ensemble (○) does better than the traditional (✳) approach – missing small emissions.

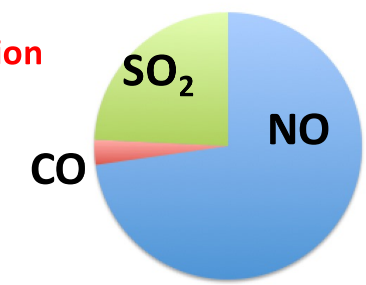
Xueying Yu et al. (2021)

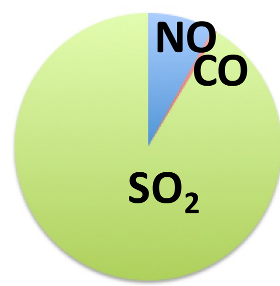
Accounting for correlated co-emitted pollutants (using 4D-Var)

Transportation

Energy

Sector-based emission scaling factor





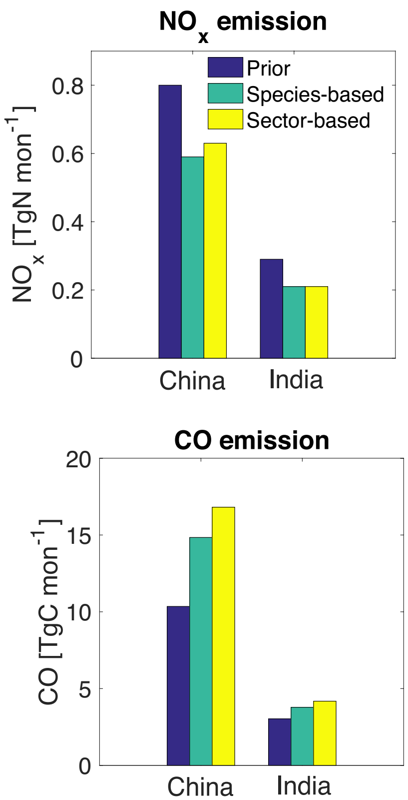
Assimilate:
MOPITT CO
OMI NO_2
OMI SO_2

Similar ratio of NO_x, SO_2 and CO emissions in the same sector, yet very different across sectors
→ Formulate inversion to adjust emissions by sector, rather than species

Zhen Qu, Daven K. Henze, Helen M. Worden, Zhe Jiang, Benjamin Gaubert, Nicolas Theys, Wei Wang
Geophys. Res. Lett., 2022, https://doi.org/10.1029/2021GL096009

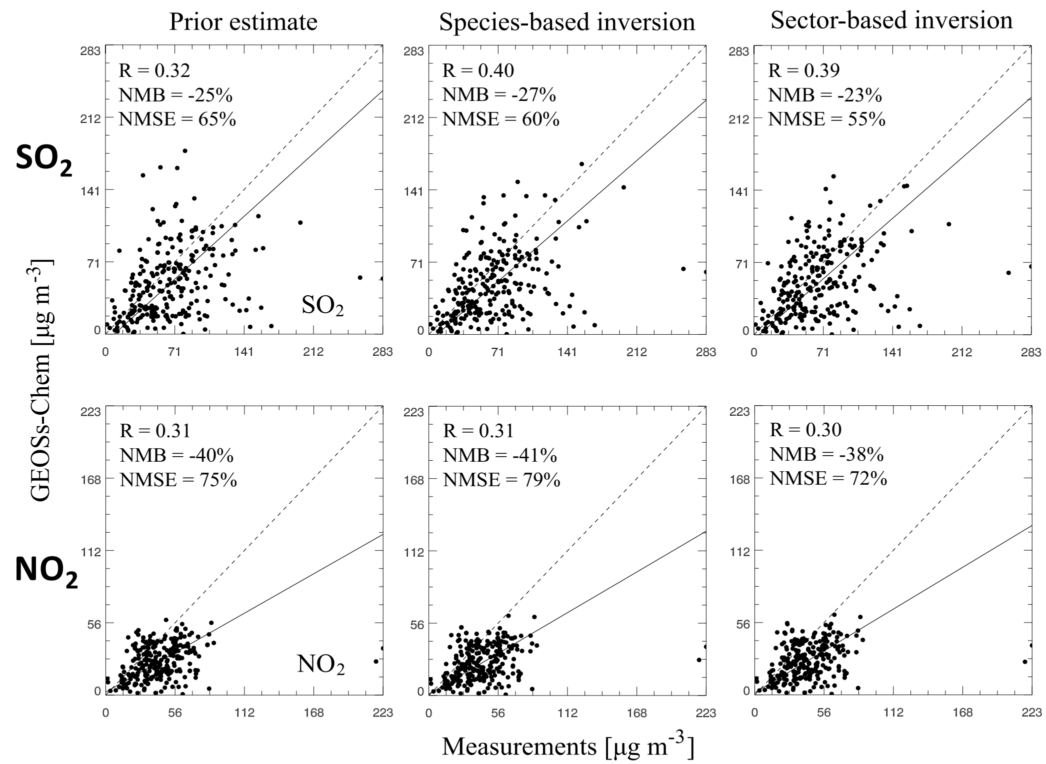
Compared to species-based inversions, sector-based constraints can be:

Different



National budgets
January, 2010

More accurate



(*Qu et al.*, GRL, 2022)

Summary

- Forward and inverse modeling methods to constrain emissions are being adapted to handle large data volumes and high-resolution capabilities of geostationary satellites.
- Multi-species, multi-instrument approaches can yield insight into air pollution processes and sector-specific source activity.
- Focus on emissions adjustments leads to important considerations regarding aggregation or emissions reduction, optimization of scaling factors or sector activity rates.
- Hybrid approaches that blend 4D-Var, mass-balance, and ensemble approaches exist within a variety of modeling frameworks (GEOS-Chem, WRF-Chem, CMAQ….) and could be made tractable in operational settings.