**Reduced Cost Construction of Jacobian Matrices for High-Resolution Inverse Modeling: An Application to Optimizing North American Methane Sources from TROPOMI Satellite Data**

**Authors**

Hannah Nesser1, Daniel J. Jacob1, Joannes D. Maasakkers2, Melissa P. Sulprizio1, Yuzhong Zhang1, Tia Scarpelli1

**Affiliations**

**1** Harvard University, Cambridge, Massachusetts, USA.

2 SRON Netherlands Institute for Space Research, Utrecht, the Netherlands.

**Abstract**

Global high-resolution observations of atmospheric trace gas concentrations from satellites can greatly improve our understanding of surface emissions through inverse analyses. For example, the new TROPOspheric Monitoring Instrument (TROPOMI) retrieves daily global observations of atmospheric methane concentrations at 7x7 km2 pixel resolution. Variational inverse methods can optimize surface emissions globally at this resolution but do not readily provide error characterization, including information content, for the posterior solution. In fact, the information content of the satellite data may be considerably lower than the data density would suggest because of limited retrieval success rate, instrument noise, and error correlations that propagate through the inversion. This could lead to smoothing errors in variational methods. An analytic inverse solution provides closed-form characterization of the posterior error statistics and information content but requires the construction of the Jacobian matrix relating emissions to atmospheric concentrations. Building the Jacobian matrix is computationally expensive at fine resolution because it involves perturbing each emission element, typically individual grid cells, in the atmospheric transport model. We propose a method to greatly decrease the computational cost of analytic inversions by constructing the Jacobian matrix using only the emission elements with sufficient information content from the observations. Starting from an initial estimate of the Jacobian matrix that assumes simple transport, we iteratively apply perturbations to the leading patterns of information content rather than to the individual model grid cells. The resulting matrix optimizes emissions only in areas spanned by these leading patterns. We demonstrate the method in an analytic Bayesian inversion of TROPOMI data over North America in July 2018. We confirm that the estimated Jacobian matrix produces posterior emission estimates and error covariances that are similar to an inversion conducted with the Jacobian matrix for the original model grid. Our method enables computationally efficient, high-resolution analytic inversions of high-density satellite data.

**Introduction**

Satellites retrieve concentrations of methane and other atmospheric trace gases at increasingly high spatial and temporal resolution. Past studies applied inverse modelling techniques to low-density retrievals of trace gas concentrations to improve constraints on emissions. Variational approaches optimize emissions at low computational cost, but provide no or poor error characterization. Analytic approaches require significant computational investment, but provide full characterization of the error and information content of the optimized emissions. However, the computational cost of analytic inversions is limited by the resolution at which surface emissions are constrained. As the observing capacity of satellites increases, so will the resolution at which inversions constrain surface emissions, resulting in a corresponding increase in computational cost. In this paper, we define a method to reduce the computational cost of analytic Bayesian inversions by generating a low-rank approximation of the Jacobian. This approx

1. Paragraph 2: Satellite overview (TROPOMI, future instruments) (and uncertainty)
2. Paragraph 3: Methane uncertainty
3. Paragraph 4: Inversion summary
   1. Statistical optimization that accounts for the errors in both satellite observations and bottom up inventories and finds the most likely set of emissions, given the observations
   2. A Bayesian inversion assumes that the errors are normally distributed; we can then write a cost function, the solution to which minimizes the distance between the prior/posterior, observations/modeled observations, weighted by errors.
   3. Can be solved by two methods: variational, analytic
4. Paragraph 5: Solving inversions (variational)
   1. Variational approaches: minimize cost function by iteratively updating the posterior
      1. Disadvantages:
         1. Shallow cost function --> may converge before it reaches the true minimum
         2. Posterior error is not characterized. (Ensemble approaches may be used; these have severe rank deficiencies)
         3. Computational cost grows significantly with sensitivity tests
         4. In the case that the adjoint is used to characterize the relationship between observations and emissions, model development cost (and the adjoint typically lags behind model developments)
5. Paragraph 6: Solving inversions (analytic)
   1. Analytic approaches: in the case that the forward model is linear (i.e. a linear relationship between emissions and the modeled observations), analytic solution to the cost function exists: posterior mean, posterior error, and information content of the solution.
      1. Disadvantages:
         1. Computational cost is limited by the cost of constructing the Jacobian, the matrix that represents the linear relationship between emissions and observations. The Jacobian is constructed via a finite differencing scheme: the dependence of observations on emissions is found by perturbing each of the grid cells for which methane emissions are constrained.
         2. As resolution increases (1) number of model runs needed increases, (2) computational cost of each model run increases.
6. Paragraph 7: Reducing computational cost of high resolution inversions
   1. Dimension reduction via grid cell aggregation (Turner, Bocquet)
   2. Reduced rank approach (Bousserez and Henze)
      1. This approach relied on the adjoint of the model and random matrix methods to construct the Jacobian
      2. For a rank n matrix, they ran the forward model n times and the adjoint n times
         1. The adjoint being computationally expensive (check IGC9 notes here--approximately 3x cost of one forward model runs), this is roughly equivalent to 4\*n model runs for a rank n Jacobian.
      3. Other approaches?
7. Paragraph 8: Proposal
   1. Here, we attempt to develop a method for constructing a Jacobian at reduced computational cost that avoids the use of the adjoint, thereby (a) reducing model development costs and adjoint lag problems and (b) reducing the total computational cost of constructing the Jacobian.
   2. We use the reduced-rank approach developed by Bousserez and Henze to propose a reduced rank Jacobian construction scheme: Rather than perturbing individual grid cells, we perturb instead the dominant patterns of information content. Because information content is a function of the forward model, this approach is iterative. We propose an iteration scheme and convergence criteria.
   3. We then demonstrate the method in the context of an inversion of TROPOMI observations over North America for July 2018 (?)