**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.

**Section 1: Introduction**

Satellite retrievals of atmospheric trace gases can be used in inverse models to infer the emissions that best explain the observed concentrations. As the observing capacity of satellites increases, inversions will constrain surface emissions at increasing resolution. The standard inverse approach uses an adjoint model to optimize emissions given the observations. The computational cost of such variational approaches is independent of the resolution at which emissions are optimized and the approach can be used in both linear and nonlinear systems. However, the solution provides incomplete characterization of errors and information content. Characterizing the information content of the inverse solution is necessary to prevent over-interpretation of high-resolution inverse results. In linear systems with normally distributed errors, as in the problem of methane, an analytic solution exists that characterizes both the error and information content of the optimized emissions. But, the computational cost of analytic inversions is limited by the resolution at which surface emissions are constrained. In this paper, we define a method to decrease the dominant computational cost in an analytic inversion, the characterization of the linear relationship between emissions and observations, given by the Jacobian matrix. We iteratively update a low-cost initial estimate of the Jacobian matrix by applying a finite difference scheme to the dominant patterns of information content in the system. The resulting low-rank approximation of the Jacobian is demonstrated in a high-resolution (1º x 1.25º) inversion of atmospheric methane column retrievals from the TANSO-FTS instrument on board the Greenhouse Gases Observing Satellite (GOSAT) for July 2009 over the North American domain.

The Tropospheric Monitoring Instrument (TROPOMI) aboard the Sentinel-5 precursor launched in October 2017 now provides daily, global retrievals of total column methane concentrations at 7 x 7 km2 nadir pixel resolution. Even at a ~3% retrieval rate, TROPOMI produces orders of magnitude more observations at higher density than the other satellite providing global retrievals of column methane concentrations, GOSAT. Maasakkers et al. (2019) conducted a global inversion of ~1.2 x 106 GOSAT retrievals over a six-year period. TROPOMI returned almost 6.4 x 106 successful retrievals in August 2018. TROPOMI also provides denser observations in areas where retrievals are successful. TROPOMI is one of many existing and planned earth-observing satellites that retrieve concentrations of methane and other atmospheric trace gases at high spatial and temporal resolution. The significant increase in current and planned observing capacity supports an increase in the resolution at which inversions optimize emissions. However, satellites do not observe all grid cells equally. TROPOMI’s ~3% retrieval rate reflects its inability to observe in areas with high cloud coverage, high aerosol loading, and high albedo. As a result, the information content of the observations may not support a dramatic increase in resolution everywhere. Defining the information content of the inverse solution along with the optimal emissions requires analytic solution of the inversion.

Inverse models describe the dependence of emissions on atmospheric concentrations, inverting a forward model that simulates atmospheric concentrations given input emissions fields. Significant errors in both the observations and model require that the inverse solution be a statistical optimization of the emissions given the observations. In a Bayesian inversion, errors in the model and observations and in the prior are assumed to be normally distributed and are summarized by the observational and prior error covariance matrices **SO** and **SA**, respectively. Bayes’ theorem allows the explicit formation of a cost function **J**(**x**) that, when minimized over all **x**, maximizes the probability of the emissions given the observations. The cost function can be minimized in one of two ways. An adjoint model can be used to iteratively update an initial estimate for the emissions (i.e. the prior) until convergence. Or, when the forward model is linear, the emissions that minimize the cost function can be found by analytic solution of . The analytic solution yields the posterior emissions as well as the associated error and the information content of posterior solution, given by the averaging kernel **A**.

Variational approaches to minimizing the cost function can be applied to linear and non-linear systems alike, and the computational cost is not limited by the dimension of the emissions vector **x**, i.e. by the number of grid boxes. However, variational approaches have a number of disadvantages. First, in high-dimensional systems, the cost function **J** is often shallow, and the variational approach may converge before the true minimum is reached. Second, variational approaches do not characterize the error or information content of the posterior solution. While ensemble approaches can approximate error, these estimates are only as good as the number of ensemble members. Third, each inverse solution requires an independent application of the adjoint model. The additional computational cost of sensitivity tests therefore mitigates the initial computational benefit. Finally, variational approaches require the continued development of adjoint models, which often lag behind state-of-the-science forward models. Efficient automatic differentiation may reduce the lag.

Analytic solutions to inversions find the true minimum of the cost function, fully characterize the error and information content of the posterior solution, support numerous sensitivity tests at only small additional computational cost, and do not require use of the adjoint model. However, analytic solutions exist only in linear or approximately linear systems and the computational cost is, as a result, limited by the resolution at which emissions are optimized. The computational cost is dominated by the cost of characterizing the linear relationship between observations and emissions, given by the Jacobian **K**. The Jacobian is generated using a finite difference scheme to find the model response to a perturbation of every optimized grid-cell. Constructing the Jacobian therefore requires *n* + 1 forward model runs. As the resolution of an inversion increases, the number of forward model runs increases exponentially and the cost of a single model run increases. Even with highly parallelized forward model runs and significant computational resources, constructing a Jacobian at such a high resolution is challenging.

Past attempts to reduce the computational cost of high-resolution analytic inversions fall into two general categories: dimension reduction and rank reduction. In the first case, the dimension of the inverse problem is reduced by aggregating grid cells. Bocquet et al. (2011) defined a method to find the optimal reduced-dimension grid over a dictionary of all allowed reduced-dimension grids. The resulting grid was optimal only within the class of allowed grids and required the storage of inverse variables for all listed grids, incurring significant memory use and mitigating the computational benefit of grid aggregation. Turner and Jacob (2015) reduced the dimension of an analytic inversion by creating a *k*-member Gaussian mixture model. While the choice of *k* minimized the sum of aggregation and smoothing error to find the optimal dimension, the Gaussians were defined using the sub-optimally on the basis of the subjectively-determined similarity of grid cells to each other. The method also relies on prior emissions estimates to allocate the reduced-dimension solution to the original grid. Reduced rank approaches defined first by Spantini et al. (2015) and elaborated on by Bousserez and Henze (2018) take advantage of the variability of information content in an inverse system to solve the inversion only in the directions with most information content. The resulting low-rank approximations for the posterior solution decrease the computational cost of inverting dense *n* x *n* matrices but not the cost of constructing the Jacobian. Bousserez and Henze (2018) define also a random matrix approach for constructing the rank-reducing projection. This approach employs both the forward model and the adjoint, increasing the computational cost and requiring continued development of the adjoint.

Here we define a method for constructing a rank *k* Jacobian that uses only ~*k* forward model runs and entirely avoids the use of the adjoint. The method converges to an approximation of the Jacobian that is most accurate in areas with high information content. We define an a posteriori filter that sets the Jacobian to zero in areas without sufficient information content. The resulting posterior solution optimizes emissions only in those areas with high information content and defaults to the prior estimate elsewhere. Section 2 describes a method by we iteratively update a low-cost initial estimate of the Jacobian matrix by applying a finite difference scheme to the dominant patterns of information content in the system. Section 3 applies the method to an inversion of atmospheric methane column retrievals from GOSAT for July 2009 over the North American domain at 1º x 1.25º resolution.

**Section 2: Methods**

Inverse models define the dependence of state variables, such as emissions, on observations. Inversions account for errors in the observations, model, and prior emissions estimates by solving for the most likely set of *n* state variables given *m* atmospheric observations **y**. In a Bayesian inversion, errors in the model and observations and in the prior estimates are assumed to be normally distributed and are summarized by the observational and prior error covariance matrices **SO** and **SA**, respectively. Bayes’ theorem allows the explicit formation of a cost function **J**(**x**) that, when minimized over all **x**, maximizes the probability of the state variables given the observations:

where **xA** is a vector of prior state variable estimates that is introduced to regularize the under-constrained problem. If the forward model is linear so that **F**(**x**) = **Kx** + **c**, where **K** = d**y**/d**x** is the Jacobian matrix, an analytic solution to the cost function minimum exists that yields the posterior mean and error and the information content. The information content is given by the averaging kernel **A**, which represents the sensitivity of the posterior mean to the truth. The sum of the diagonal elements gives the degrees of freedom for signal (DOFS), the number of pieces of information an inversion can constrain. Most inversions constrain fewer pieces of information than the number of state variables: DOFS << *n*.

The computational cost of an analytic inversion where the number of observations is much larger than the number of state vector elements, *m* >> *n*, is limited by the cost of constructing the Jacobian matrix **K**. The Jacobian is typically constructed column-wise by finding the model response to a perturbation of each state variable, requiring *n* + 1 forward model runs. However, not all state variables are equally well constrained by inversion. The computational cost of constructing the Jacobian matrix can therefore be reduced by (1) decreasing the dimension of the state variable vector by aggregating together state variables or (2) decreasing the number of model runs needed to construct the Jacobian matrix at the original state variable resolution, creating a low-rank approximation of the Jacobian. In both cases, the decrease in dimension or rank should maximize the information content of the resulting inversion.

Bousserez and Henze (2018) define a sequence of linear transformations that decrease and restore the dimension of the state space while maximizing the information content of the inversion. The transformations occur in the directions of maximum information content, which are given by the eigendecomposition

where the columns of **W** are the eigenvectors of **Q** and the diagonal elements of are the eigenvalues of **Q**.[[1]](#footnote-1) **Q** is symmetric semi-positive definite, so the eigenvectors form an orthonormal basis. Defining as the matrix of the first *k* columns of **W**, the linear transformation decreases the dimension of the state space and restores the original dimension, both while maximizing information content. Following the formalism of Bousserez and Henze (2018), we refer to as a reduction operator and as a prolongation operator. Applying these transformations in sequence yields a projection transforms the original rank *p*, dimension *n* state space to a rank *k* << *p*, dimension *n* subspace. This subspace maximizes the information content of the inversion relative to all other rank *k* subspaces.

The computational cost of constructing the Jacobian matrix can be decreased by building the matrix within the optimal decreased-dimension or low-rank subspaces. However, the optimal transformations are a function of the Jacobian matrix. In the case in which the state vector elements are surface emissions over a discretized grid, a mass-balance approach can generate an initial estimate of the full dimension Jacobian matrix at low computational cost. This approach assumes that the fine structure of the Jacobian is given by the prior. Given a perturbation to a prior emissions element , the change in observed atmospheric concentrations over that grid cell can be estimated as

where Mair and Mx are the molecular weights of dry air and the optimized species, respectively, g is gravity, U is the wind speed, W is the distance the wind travels across the grid box, and P is the surface pressure (source). The wind speed, length scale, and pressure can either be assumed constant or generated by a single model run. The Jacobian is then constructed column-wise with the *i*th column given by . The resulting matrix is sparse and generates an eigendecomposition of **Q** that does not accurately approximate the eigendecomposition of **Q**true. To decrease the sparsity, the perturbation can be assumed to influence all observed atmospheric concentrations within some radius of the perturbed grid box, with the influence decreasing with distance from the perturbation.

Given an initial estimate of the Jacobian matrix and the resulting approximations of the optimal dimension transformations and , we define methods to decrease the computational cost of constructing the Jacobian matrix by (1) decreasing the dimension of the state vector by aggregating together state variables and (2) decreasing the rank of the Jacobian. In what follows, we consider the case in which the state vector elements are surface emissions, discretized over a grid. The results can be generalized to temporal as well as spatial state variables.

**Section 2.1: Dimension Reduction**

Prior dimension reduction decreases the computational cost of inversions with large state vectors. A multi-scale grid can be defined that preserves resolution where high information content is available and goes to coarser resolution elsewhere. The information content of a grid cell is given by the diagonal elements of the averaging kernel **A**. [Define the significance vector , where is the averaging kernel associated with the initial estimate of the Jacobian.] [However, this definition neglects the off-diagonal structure of the averaging kernel. The row-wise magnitude of the prolongation matrix, the columns of which span the averaging kernel, is a measure of the sensitivity of the leading directions of information content to each grid cell, giving a significance vector

]

An optimal multi-scale grid preserves resolution where is largest and aggregates grid cells where is smaller. Grid cells corresponding to smaller values of can be consolidated using, for example, k-means clustering. K-means clustering aggregates together grid cells based only on their spatial proximity and therefore is likely to aggregate together grid cells that yield similar model responses. It also can easily be adapted to generate clusters that satisfy the desired final state vector dimension. However, k-means clustering neglects the variation in emissions magnitudes or profiles in different grid cells. An algorithm that considers the similarity of emissions, such as the Gaussian mixture model implemented by Turner and Jacob (2015), could be used instead.

The reduced-dimension Jacobian can be constructed on the basis of this multi-scale grid. However, the initial estimate of the Jacobian may not generate information content that accurately reproduces the “truth.” The additional information content generated by running the model and the resulting reduced-dimension Jacobian can be used as the basis for a second update to the initial estimate. can be re-gridded to the original grid, using a prior estimate to allocate the model responses to individual grid cells, and the significance vector calculated. To avoid superfluous model runs, the multi-scale grid can be adjusted only in places where the relative difference between and , both re-gridded to the first multi-scale grid, is sufficiently large. The adjusted multi-scale grid can be used as the basis of the updated reduced-dimension Jacobian .

The reduced-dimension Jacobian can be iteratively updated in this manner until convergence. Convergence can be defined using the relative difference between the previous and current reduced-dimension Jacobians or between the previous and current significance vectors and , respectively. In either case, convergence is achieved when the relative differences are sufficiently small. Notably, using the significance vectors rather than the reduced-dimension Jacobians avoids unnecessary model runs. Box 1 describes the full algorithm for developing a multi-scale Jacobian.

**Multi-scale Jacobian algorithm**

1. Initialize the Jacobian . Define the significance vector . Generate a multi-scale grid that preserves the highest resolution where is largest and aggregates grid cells elsewhere.
2. Perturb each of the elements of the multi-scale grid in the forward model to produce a reduced-dimension Jacobian .
3. Re-grid to the original grid according to the prior and calculate the updated significance vector . Where the relative difference between and is sufficiently large, update the multi-scale grid.
4. Repeat steps (ii) and (iii) until the relative difference between all elements of and is sufficiently small.

Box 1

Iteratively updating the multi-scale grid ensures high resolution in areas where the model, observations, and prior produce high information content. It also allows for exact solution of the analytic inversion on the multi-scale grid. If higher resolution results are needed, additional information must be introduced to allocate the posterior solution to the finer grid. Alternatively, a low-rank Jacobian can preserve everywhere the highest resolution without requiring additional information.

**Section 2.2: Rank Reduction**

A low-rank Jacobian represents the linear relationship between emissions and observations in a low-rank space that neglects insignificant components of information content. The patterns of information content, given by the eigenvectors of the averaging kernel,form an orthonormal basis of the inverse system in the same way as discrete grid boxes. Ranked according to the corresponding eigenvalue, the eigenvectors explain decreasing fractions of the variance in information content. Often, the first *k* << *n* eigenvectors explain most of the variance. To decrease computational cost, a low-rank Jacobian is built from the model response to perturbations of the first *k* dominant patterns of information content rather than the *n* individual grid cells. As in the case of the reduced-dimension Jacobian, the initial estimate of the Jacobian may not generate information content that accurately reproduces the “truth.” The low-rank Jacobian is therefore iteratively updated, using patterns of information content that increasingly reflect contributions from the prior, observations, and model.

The number of patterns *k*, given by the eigenvectors of , should be chosen to maximize the information content spanned by the eigenvectors and minimize the number of model runs. The first *p* descending eigenvalues of give the fraction of information content explained by the corresponding first *p* eigenvectors. Because the information content represented by lacks contributions from the model and observations, the tailing eigenvectors may not span the same information content space as the “true” tailing eigenvectors. As a result, the initial choice of *k* should encompass between ~50% and ~80% of the information content. Subsequent choices of *k* should encompass either between ~80% and ~95% of information content or, if the spectrum exhibits a discontinuity, *k* should be twice the previous choice so as to encompass the additional information introduced by the most recent model runs. An illustration of the latter case will be shown in the results.

The response of a forward model F to the *j*th eigenvector , given by the *j*th column of , is

In effect, the eigenvector is used as a perturbation to the prior emissions, scaled by a factor to ensure numerical stability. The model responses form the columns of the matrix. If the forward model is linear, F can be written as and as . is the Jacobian of a reduced-dimension space spanned by the first *k* eigenvectors; this Jacobian must be transformed to the original state dimension for use in analytic inversions. The resulting Jacobian will have dimension and rank . Bousserez et al. (2018) show that the low-rank Jacobian is given by The low-rank update is therefore given by .

**Low-rank Jacobian algorithm**

1. Initialize the Jacobian .
2. Calculate and determine *k*. Calculate the reduction operator and the prolongation operator .
3. Perturb the *k* columns of in the forward model to produce a reduced-dimension Jacobian .
4. Transform the reduced-dimension Jacobian to the original dimension by .
5. Repeat steps (ii) through (iv) until the relative difference between all elements of and is sufficiently small.

Box 2

The low-rank approximation can be updated by recalculating the eigenvectors and the model response. Convergence can be determined by comparing the relative difference between the averaging kernels associated with the previous and current low-rank Jacobian approximations, and , respectively. Box 2 describes the full algorithm for developing a low-rank Jacobian.

The resulting Jacobian is a low-rank approximation of the linear forward model. It accurately quantifies the forward model where the inverse system has high information content and loses accuracy in areas with lower information content. Consequently, the resulting posterior solution is accurate in areas with high information content and has increasing errors with decreasing information content. A filter can be applied that sets the posterior solution to the prior value in grid cells where the averaging kernel corresponding to the approximated Jacobian is small. [Insert exact threshold here.] This approach solves the inversion in a manner consistent with the information content of the inverse system.

**Section 3: Results and Discussion**

We demonstrate both the reduced-dimension and low-rank Jacobian approaches in an analytic Bayesian inversion of atmospheric methane columns observed by the GOSAT satellite over North America in July 2009. We construct a Jacobian at 1º x 1.25º resolution, a reduced-dimension Jacobian with ~300 state vector elements, and a rank 200 Jacobian. We also construct a reduced-dimension Jacobian following the Gaussian mixture model (GMM) method described by Turner and Jacob (2015). We use those Jacobians within the inverse framework described by Maasakkers et al. (2019), adapted from the global system to the North American domain. We solve for posterior scaling factors and compare results. [Insert summary of results.]

In all inversions, we use the nested North American GEOS-Chem CTM version 12.4.0 as forward model to simulate atmospheric methane column concentrations at 0.5º x 0.625º resolution. The model is driven with MERRA-2 meteorological fields (Bosilovich et al., 2016) from the NASA Global Modeling and Assimilation Office (GMAO). We use boundary conditions and initial conditions [check this] from a global GEOS-Chem 4º x 5º simulation for July 2009 driven by prior emissions with posterior scaling factors applied as described by Maasakkers et al. (2019). All inversions also use the prior emissions, prior error covariances, observations, and observational error covariances as described by Maasakkers et al. (2019). In particular, we use the University of Leicester version 7 CO2 proxy retrieval over land (Parker et al. 2011, 2015) for July 2009, excluding glint data and observations north of 60ºN. Figure 1 shows the GOSAT data used by all inversions.

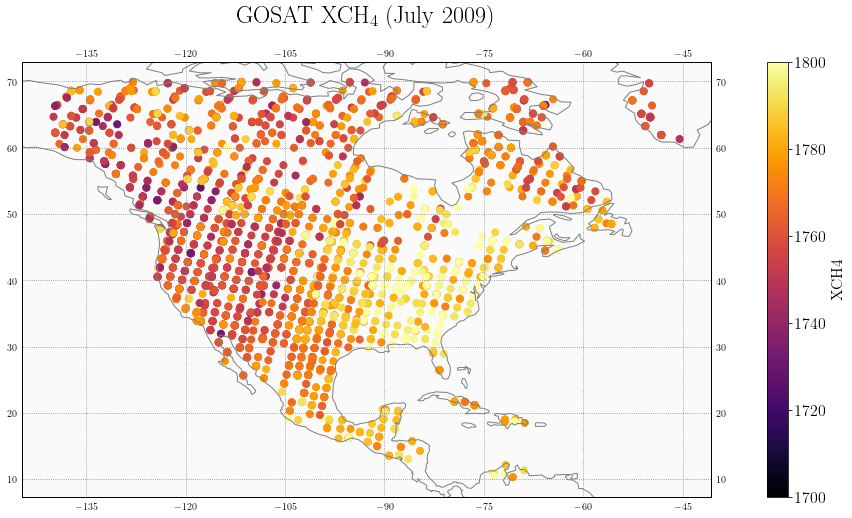


Figure 1: GOSAT atmospheric methane column retrievals for July 2009

Figure 2 shows the information content of the true system, as given by the diagonal elements of the averaging kernel. Grid boxes with large (close to one) values have more information content, or a stronger ability to constrain emissions. Grid boxes with small (close to zero) values have less ability to constrain emissions. The information content displays significant spatial variability. Notably, all grid boxes have relatively low averaging kernel values because of the limited number of observations incorporated into the inversion. Introducing more observations may change the distribution of the information content but the spatial variability would likely be preserved. This spatial variability justifies the reduced-dimension and low-rank Jacobian approaches.

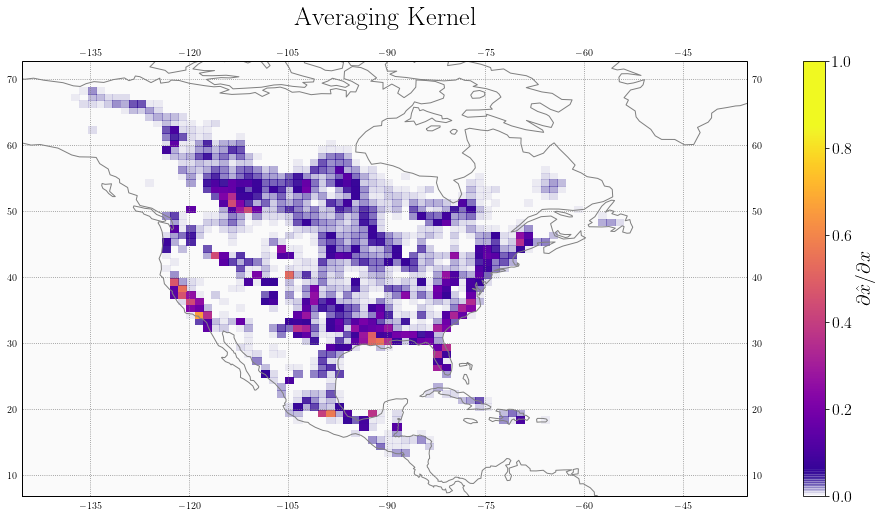


Figure 2: Information content of the true inverse system, as given by the diagonal elements of the averaging kernels. The inversion can better constrain emissions in grid boxes with large (closer to one) averaging kernel values.

We use a 1º x 1.25º grid as the base resolution for all Jacobians. The true Jacobian is constructed by perturbing aggregated 0.5º x 0.625º grid boxes, requiring 2,098 model runs. The initial estimate for the reduced-dimension and low-rank Jacobians is constructed following the mass balance approach described in Section 2. Figure 3 shows the initial estimate Jacobian elements plotted against the true Jacobian elements. While the elements of the initial estimate are of the same order of magnitude as the true Jacobian, significant scatter exists in the values. However, the patterns of information content, given by the eigenvectors of , are broadly consistent with the true patterns of information content. As illustration, the top row of figure 4 shows the first four patterns of information content for the true Jacobian, while the second row shows those patterns for the initial estimate. The third row shows the eigenvalue spectrum for the true and initial Jacobians, illustrating the similarity in the information content explained by each of the eigenvectors for both the true and initial inverse systems.

To demonstrate the efficacy of the reduced-dimension and low-rank Jacobian methods proposed, we construct both Jacobians with the goal of reducing the number of model runs by an order of magnitude, from ~2,000 to ~200-300. It is worth noting that increasing the number of model runs, particularly in the second iteration of both methods, will always increase the accuracy of the Jacobian relative to the true Jacobian. Generally, then, the number of model runs should be set by computational limits. We consider first the reduced-dimension and then the low-rank Jacobians.

[Paragraph about reduced-dimension Jacobian

Note: I don’t currently have this paragraph completed. I need to think of a way to evaluate the accuracy of the reduced-dimension Jacobian. While the posterior solution is exact on the multi-scale grid, it loses significant accuracy when regridded to the original resolution.



Figure 3: Mass-balance estimated Jacobian plotted element-wise against the true Jacobian. [Note: I’ll get rid of the text in the upper left in the next version of this plot.]



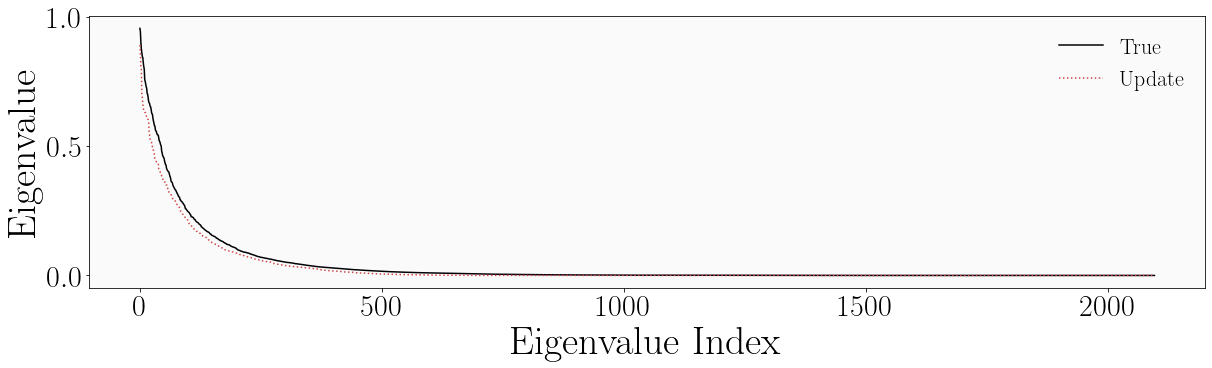


Figure 4: Eigenvectors and eigenvalues of the mass-balance estimated Jacobian and true Jacobian. The top row shows the first four eigenvectors of and the second row the first four eigenvectors of . While significant differences exist in the exact patterns, the estimated Jacobian generates patterns of information content that capture the broad regions contained in the true patterns. These similarities persist beyond the first four eigenvectors. The bottom row shows the spectra of the eigenvalues associated with the true Jacobian (black) and the estimated Jacobian (red dashes) [Note: I’ll change the label “Update” to “Initial Estimate.”] The two spectra are similar, demonstrating that similar fractions of information content are captured by each eigenvector. [Note: I will also make the fonts a consistent size.]

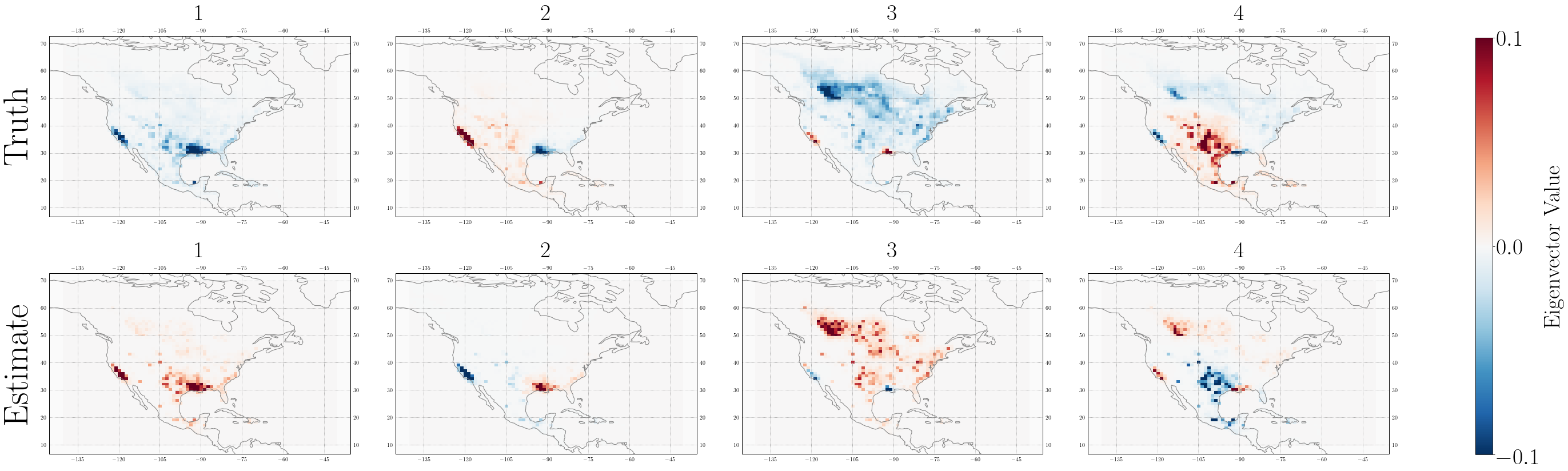
Paragraph about reduced-dimension Jacobian, cont.

Figure 5: Plot showing the delineations of the final multi-scale grid

Figure 6: Plot showing posterior solution on multi-scale grid [left] and the posterior solution – true solution on multi-scale grid [center] and the posterior solution vs. the true solution [right]

? Figure 7: Plot showing scatter of Jacobian, posterior mean, and posterior variance?]

The low-rank Jacobian is constructed in two iterations. Because the averaging kernel corresponding to the mass balance Jacobian lacks contributions from the model and observations, the tailing patterns of information content may not span the same information content space as the “true” tailing patterns. We therefore update the initial Jacobian by perturbing the patterns that correspond to 80% of the information content, requiring 102 model runs. Figure 7 shows the first four eigenvectors and eigenvalue spectrum for the updated and true Jacobian, following the same format as figure 4. The eigenvectors associated with the updated Jacobian better capture the true patterns of information content for approximately the first 100 eigenvectors (the first four are shown here as a demonstration). The eigenvalue spectrum exhibits a discontinuity near index 100, consistent with a rank 100 approximation. We update the second Jacobian by perturbing the first 204 patterns to ensure that we capture the additional patterns of information content introduced by the first update. Visual inspection shows that the first several hundred patterns of information content associated with the first and second updates are similar, demonstrating convergence. [Delete previous sentence and insert improved convergence statement: compare averaging kernels of previous and current update.]



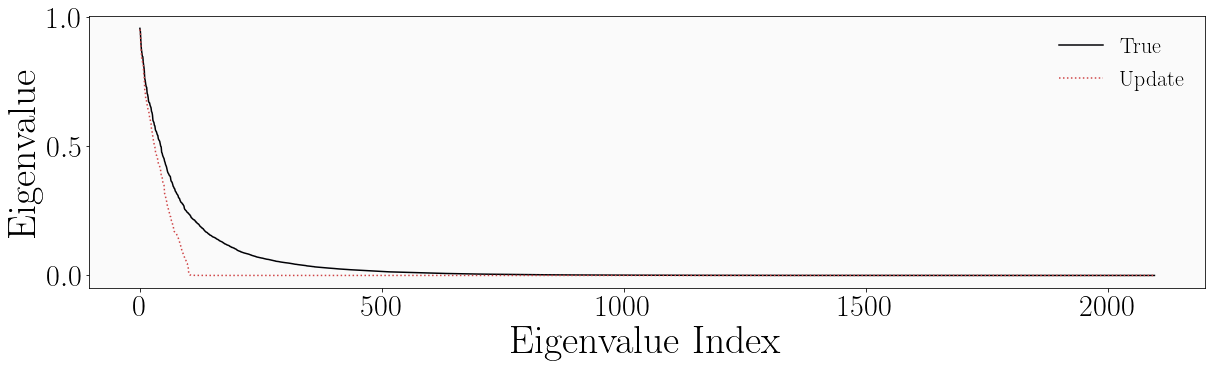
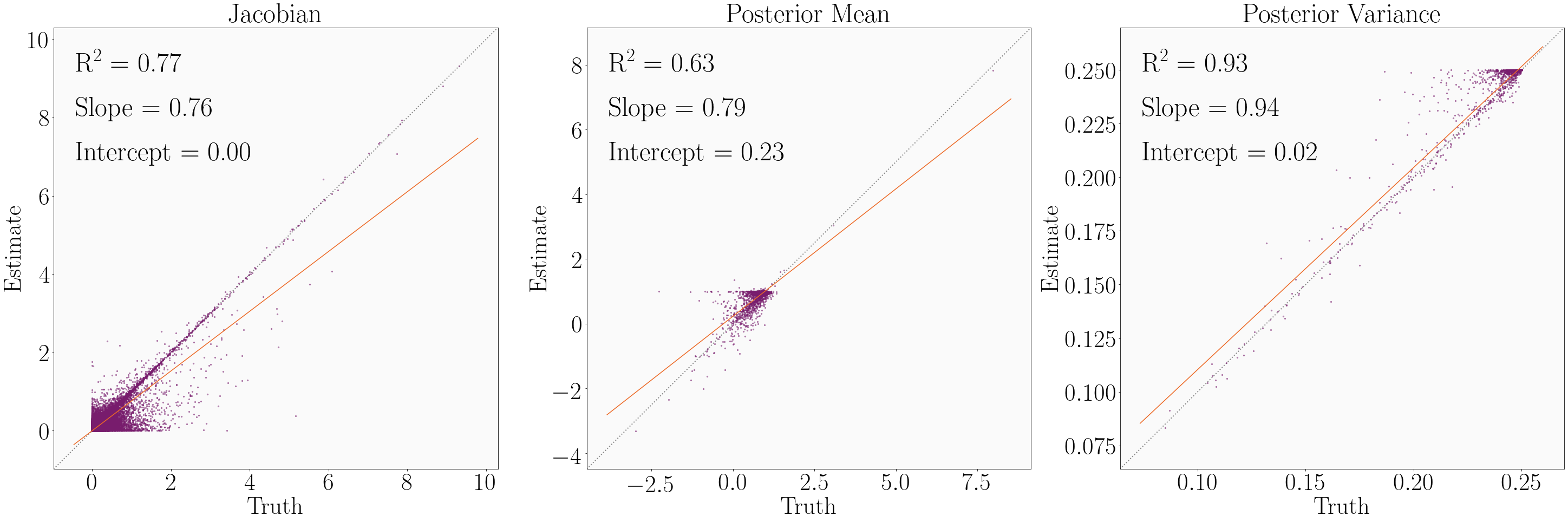


Figure 7: Eigenvectors and eigenvalues of the first update Jacobian and true Jacobian. The eigenvectors associated with the first update Jacobian better capture the patterns given by the true eigenvectors. Variations in sign are unimportant and variations in order are negligible. The eigenvalue spectrum corresponding to the first update Jacobian exhibits a discontinuity around index 100, consistent with its rank.

The first row of figure 8 shows the resulting low-rank Jacobian and the corresponding posterior mean and posterior variance plotted against the true Jacobian, posterior mean, and posterior variance, from left to right. The resulting posterior mean, shown in the center panel, corresponds well with the true posterior mean, with an r2 of 0.63. The fit is negatively affected by the cluster of values where the posterior mean associated with the low-rank Jacobian is equal to one. These values occur in grid boxes where the Jacobian does not optimize the posterior and instead maintains the prior value. These non-optimized grid boxes correspond to those grid boxes with low information content, as given by the diagonal elements of the averaging kernel associated with the low-rank Jacobian. In the second row of figure 8, we apply a filter and consider only those grid boxes with averaging kernel values greater than 0.005. [Note: in the future, I will try simply filtering out the areas where the posterior mean is one (i.e. the prior) rather than using an averaging kernel threshold. And/or come up with a better explanation for 0.005.] The filter improves the fit of the approximate mean and variance to the true mean and variance, respectively. While this filter decreases the number of optimized grid boxes from 2,098 to 465, the excluded grid boxes had such limited information content that even the true Jacobian would have limited ability to constrain emissions therein. Figure 9 shows the approximate posterior mean (left) and the difference between the approximate and true posterior mean (right) plotted over the North American domain. The non-optimized values are greyed out. [Note: in the future, I will use stippling.]



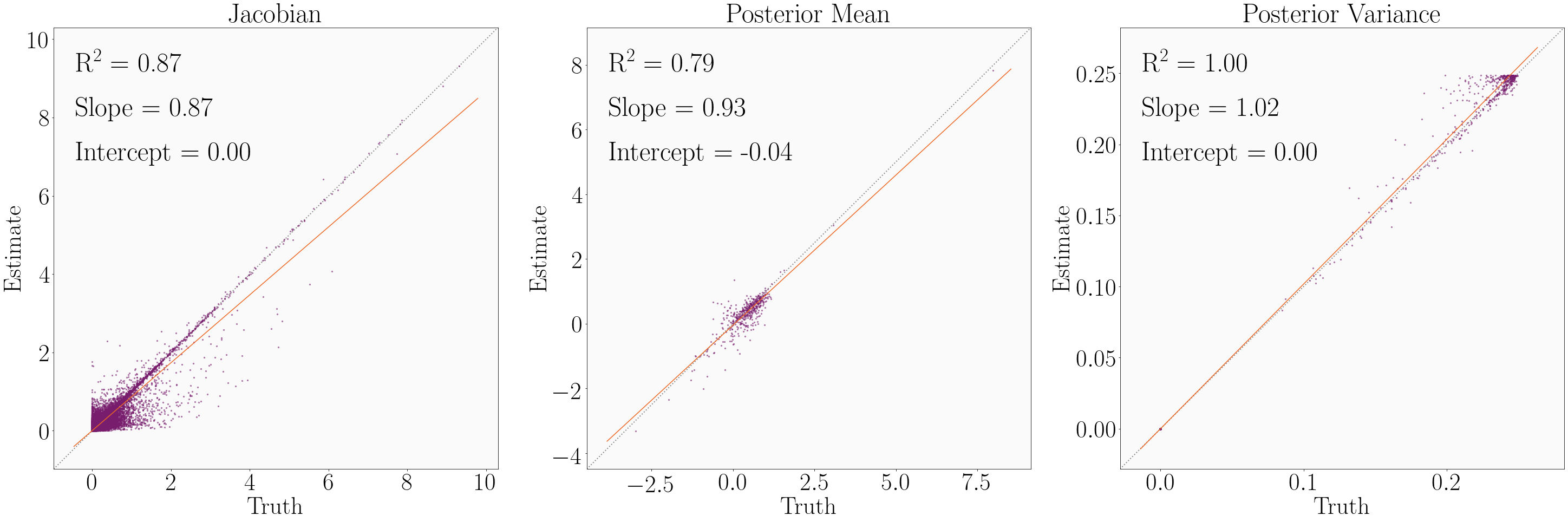
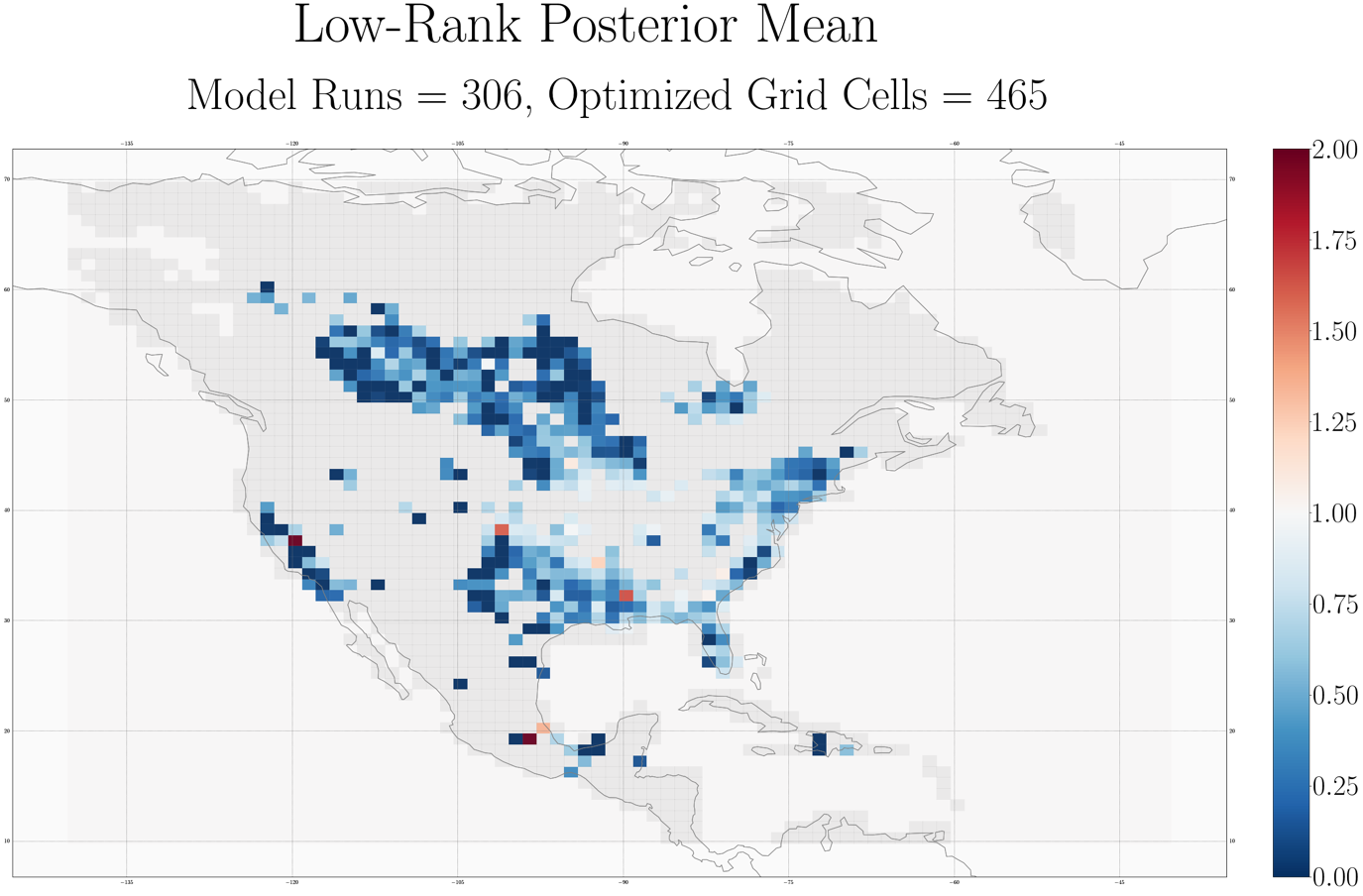


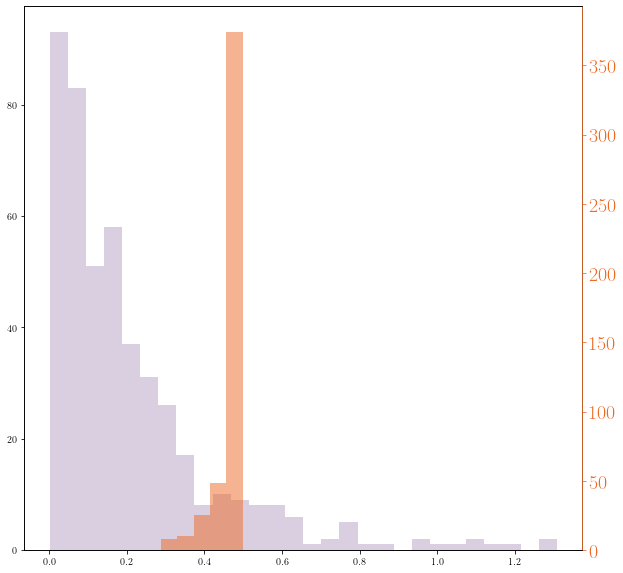
Figure 8: The second update low-rank Jacobian and the corresponding posterior mean and variance plotted element-wise against the true Jacobian, posterior mean, and posterior variance, from left to right. The first row shows all elements. The second row applies a filter that excludes grid boxes with averaging kernel diagonal values less than 0.005, improving the fit in all cases but reducing the number of optimized grid cells from 2,098 to 465.



[Note: missing difference plot here.]

Figure 9: The posterior mean associated with the low-rank Jacobian (left) and the difference between the approximate posterior mean and true posterior mean (right).

The errors in the filtered approximate posterior mean are mostly insignificant. Figure 10 shows in purple the distribution of the absolute error in the approximate posterior mean, as measured against the true posterior mean, and in orange the distribution of the posterior error, as given by the posterior error variance. Most of the errors in the approximate posterior mean are significantly smaller than the posterior error. There are some errors that are larger than the posterior errors; it is worth noting that (1) following the inverse framework of Maasakers et al. (2019), prior errors are capped at 0.5, resulting in a similar cap on posterior errors [in the furture: check the math here] and (2) posterior errors calculated in an analytic inversion often underestimate the true posterior errors (insert citation). It is therefore likely that even in the limited cases when the error in the approximate mean is larger than the calculated posterior error, the error in the mean is likely negligible.



[Note: I need to change a lot on this plot. The left axis corresponds to the purple histogram, which shows the histogram of error in the approximate posterior mean as measured against the true posterior mean. The right axis corresponds to the orange histogram, which shows the histogram of error as given by the posterior error variance. The x-axis shows the error values.]

Figure 10: Distribution of errors in the approximate posterior mean as given by the difference between the approximate and true means (purple) and the posterior variance (orange). Most of the error in the approximate mean is less than the error given by the posterior variance.

We also conduct an inversion with a Jacobian constructed using the Gaussian mixture model (GMM) approach described by Turner and Jacob (2015) following Maasakkers et al. (update once published). [Insert description of GMM and GMM results here.]

**Section 4: Conclusions**

We suggest two methods of decreasing the computational cost of analytic Bayesian inversions of linear systems by an order of magnitude. We reduce the number of model numbers necessary to characterize the linear relationship between modeled observations and emissions, given by the Jacobian matrix. In the standard approach, constructing the Jacobian requires a model run for each state vector element, here taken to be a grid cell, optimized. We demonstrate two methods that reduce the number of model runs by an order of magnitude. Both methods take advantage of the spatial variability in information content in the inverse system by constraining at either highest resolution or highest accuracy those grid cells that are most informed. The methods iteratively update a low-cost initial estimate of the Jacobian built using a mass balance approach on the prior emissions estimate. The first method constructs a reduced-dimension Jacobian by iteratively developing a multiscale grid that preserves resolution where information content is highest. The second method constructs a low-rank Jacobian by perturbing the dominant patters of information content. We demonstrate both methods in an inversion of GOSAT atmospheric methane column observations over the North American domain for July 2009 at 1º x 1.25º resolution. We also construct a reduced-dimension Jacobian following the Gaussian mixture model (GMM) approach described by Turner and Jacob (2015). In all cases, we reduce the number of model runs from 2,098 to ~300.

The reduced-dimension, low-rank, and GMM Jacobians all reduce the number of model runs necessary to construct a Jacobian, and therefore the computational cost of an analytic inversion, by an order of magnitude. The three Jacobians produce posterior means and variances that approximate the true posterior mean and variance to varying degrees. The reduced-dimension Jacobian produces everywhere on the multi-scale grid an exact, but lacks accuracy when regridded to the original resolution. The low-rank Jacobian produces an approximate solution at the original resolution. However, it is most accurate in the subset of grid cells with highest information content. Still, it constrains more grid cells per model run than a Jacobian constructed using a full finite-difference scheme. It also makes explicit that the inversion is solved only where there is sufficient information content. [Insert GMM Jacobian conclusion.] Both of our approaches introduce additional error by reducing the number of model runs. However, most of these errors are smaller than the posterior errors and it is likely that the errors are smaller than the true posterior errors. Increasing the number of model runs will decrease these errors.

While the proposed methods are evaluated in the context of an inversion of atmospheric methane column observations, they are applicable in any linear system, including many long-lived climate forcers. As satellite observations of atmospheric constituents continue to improve, it will be possible to conduct inversions at increasingly high resolution. While analytic inversions provide characterization of the information content and can help avoid over-interpretation of inverse results, their computational cost is limited by the number of grid cells constrained by the inversion. Our approaches allow analytic solution of the inverse system at significantly reduced computational cost. Future work could consider the expansion of both approaches to the temporal dimension. And, as cloud computing becomes increasingly available, the computational benefits of performing large, parallel computations in cloud environments should be considered.

[Note: I kept the conclusion short, anticipating significant changes to the rest of the text that will change the tone/content of the conclusion.]

**References**

[Insert references]

1. The eigendecomposition of the prior-preconditioned Hessian

   yields the same eigenvectors **V** = **W** and eigenvalues while avoiding the inversion of a large, non-dense matrix. These eigenvalues correspond to the eigenvalues of the pre-whitened Jacobian. [↑](#footnote-ref-1)