**Reduced Cost Construction of Jacobian Matrices for High-Resolution Inverse Modeling: An Application to Optimizing North American Methane Sources from GOSAT 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 GOSAT data over North America in July 2009. 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 observations of atmospheric composition provide an increasingly powerful resource to improve our understanding of emissions (Streets et al., 2013). Satellites provide dense observations at a global scale but are subject to large errors, both from the measurement retrieval and from the chemical transport models (CTMs) used to relate the observed atmospheric concentrations to emissions (Jacob et al., 2016). Exploiting satellite data to quantify emissions at high resolution with reliable error characterization is of considerable interest but is also computationally expensive and may be limited by the actual information content of the satellite data. Here we present a method to achieve high-resolution inversions of satellite data to infer emissions with full error characterization, limited solely by the information content of the satellite data and by the availability of computational resources.

The standard inversion approach to infer emissions from observations of atmospheric composition is to fit the data to a CTM that simulates atmospheric concentrations as a function of emissions. The CTM represents the forward model for the inverse problem. The fit is generally done by Bayesian optimization, minimizing a cost function that is regularized with prior information on the emissions. The optimal (posterior) estimate of emissions corresponds to the minimum of the cost function. This minimum is typically determined using a numerical (variational) method, often employing the adjoint of the CTM to compute the cost function gradient (Henze et al. 2007). However, this numerical solution provides no explicit characterization of the error or information content in the solution. The error can be estimated from numerical construction of the inverse Hessian of the cost function but this is computationally expensive (Bousserez ref)

In the frequent case where the observed atmospheric concentrations have a linear dependence on emissions and error statistics can be assumed normal, an analytic solution to the Bayesian optimization problem exists that provides closed-form expressions for posterior emissions estimates, error statistics on these emissions, and information content (Rodgers, 2000). The analytic solution can be extended to log-normal error statistics and to non-linear problems with an iterative approach (Maasakkers et al. 2019; Rodgers 2000). The analytic solution requires explicit construction of the Jacobian matrix of the forward model, which represents the sensitivity of the simulated concentrations to the optimized emission state vector (Brasseur and Jacob 2017). The Jacobian is constructed column-wise by conducting *n* CTM simulations perturbing individual state vector components *xi­* and obtaining the corresponding column . The construction of the Jacobian matrix is a parallel computational problem that can be executed efficiently on large computing clusters (Maasakkers et al., 2019). The computational cost of conducting the *n* simulations still limits the size of the state vector **x** and therefore the resolution at which the inversions can be conducted (Turner and Jacob 2015). However, once the Jacobian matrix is constructed, repeated inversions can be conducted at essentially no additional computational cost to investigate the sensitivity of the solution to changes in the specification of errors, prior assumptions, or the number and type of observations included.

Here we will focus on methane as a case in point. Methane is a major greenhouse gas but its emissions have large uncertainty (Saunois GCP reference). Satellite observations of atmospheric methane can be used to improve emission estimates (Jacob et al., 2016). This was first done with the SCIAMACHY satellite instrument that provided high-quality data though with coarse spatial resolution (30x60 km2 in nadir) for 2005-2007 (Bergamaschi et al., 2009, 2013; Wecht et al., 2014). The longest satellite record of atmospheric methane (2009-present) has been from GOSAT (Kuze et al.), which measures atmospheric methane columns with 10-km diameter pixels approximately 250 km apart along-track and cross-track. The GOSAT data were used by Maasakkers et al. (2019) in an analytic inversion of methane emissions and trends at 4º x 5º spatial resolution, commensurate with the GOSAT data and corresponding to ~2,000 state vector elements. The launch of the Tropospheric Monitoring Instrument (TROPOMI) aboard the Sentinel-5 precursor in October 2017 changes the magnitude of the computation problem. TROPOMI now provides daily, global retrievals of atmospheric methane columns at 7 x 7 km2 nadir pixel resolution, increasing coverage by orders of magnitude compared to GOSAT (Veefkind et al.). However, TROPOMI’s methane retrieval has only a ~3% success rate limited by clouds, high aerosol loadings, and variable surface albedo and topography (Hu et a., 2018). While analytic solution of the inverse problem at the scale of the TROPOMI observations is computationally prohibitive, this solution still has the advantage of providing full error characterization, and the low spatially-dependent success rate of the TROPOMI retrievals suggests that one could effectively define a variable spatial resolution that enables analytic solution while preserving most of the information content of the observaions.

Several methods have been developed to decrease the spatial resolution of the emission state vector while preserving the information content of the observing system. Bocquet et al. (2011) developed a method to findthe optimal multiscale grid for the state vector given an observing system . However, the method requires a search through all possible combinations of multiscale grids and that by itself is computationally expensive.. Turner and Jacob (2015) reduced the dimension of an analytic inversion using a *k*-member Gaussian mixture model with *k* < *n,* but the Gaussian groupings of high-resolution grid cells were defined by subjective similarity criteria including the prior distribution of emissions. Spantini et al. (2015) and Bousserez and Henze (2018) presented reduced-rank methods to identify the dominant pieces of information in an observing system. However, this required explicit construction of the original (*m*x*n*) Jacobian matrix (Spantini et al., 2015) to calculate the directions of information content, or expectation that the information content is limited to a few elements as is the case for severely underconstrained observing systems (Bousserez and Henze, 2018).

Here we present two methods to construct the Jacobian matrix for a native *n*-dimensional state vector that preserve the information content of the observation system while requiring *k* < *n* forward model simulations. These methods allow for analytic inversion of satellite data without pre-judging information content. The first method uses an initial estimate of the Jacobian matrix to generate a multiscale grid that preserves resolution in the areas with highest information content. The second system populates an initial estimate of the native-resolution Jacobian matrix along the dominant patterns of information content in the system. In both cases, the number of model simulations *k* used to construct the Jacobian is elected by the user and represents a trade-off between computational tractability and information content. We apply both methods to a 1-month inversion of satellite data as demonstration.

**Section 2: Methods**

*Section 2.1: Analytic Solution to the Inverse Problem*

The optimal estimate **xhat** of a state vector **x** given an observation vector **y**, prior estimate **xA,** and normal error statistics is given from Bayes’ Bayes’ theorem by the minimization of a scalar cost function *J*(**x**): :

Here **SA** is the prior error covariance matrix, **SO**is the observational error covariance matrix including contributions from both measurement and forward model errors, and **F**(**x**) is the forward model, i.e., the prediction **y** = **F**(**x**) from the CTM. If the forward model is linear so that **F**(**x**) = **Kx** + **c**, where **K** = d**y**/d**x** is the Jacobian matrix and **c** is a constant, then an analytic solution to the cost function minimum exists that yields the posterior emissions estimate and its error covariance matrix :

Comparison of  and defines the information content of the inverse system, quantified by the averaging kernel matrix **A**, which represents the sensitivity of the posterior emissions estimate to the true state **x**. **A** can be calculated as **A = I -**  ,, or alternatively as (Rodgers, 2000)

The trace of **A** measures the number of pieces of information that can be independently constrained by the observing systemand is called the degrees of freedom for signal or DOFS (Rodgers, 2000).

In an analytic inversion of satellite observations, where the number of observations is much larger than the number of state vector elements, *m* >> *n*, the computational cost is determined by the cost of constructing the Jacobian matrix **K**. The Jacobian matrix is constructed column-wise by computing the model response to a perturbation of each state variable, requiring *n* + 1 forward model simulationscovering the observation period. This construction is computationally expensive. The observation system may also place a limit on the useful size of *n* as given by the DOFS. The computational cost of constructing the Jacobian matrix can be reduced by (1) decreasing the dimension *n* or (2) retaining the original dimension but decreasing the number of model simulations to generate a low-rank approximation of the Jacobian. We consider both of these approaches in what follows.. We discuss first optimal reductions in both dimension and rank for an inverse system with a known Jacobian matrix (Section 2.2). We then discuss a two-step update method to approximate the Jacobian matrix using the defined reductions in dimension and rank (Sections 2.3 through 2.5). For the sake of example, the state vector defining the dimension *n* will be taken here as a gridded field of emissions, where *n* sets the spatial resolution which may vary across the field. The method is readily generalizable to a temporally resolved state vector, or to another spatial representation of emissions such as the GMM (Turner and Jacob, 2015).

*Section 2.2: Optimal Reductions in Dimension and Rank of Jacobian Matrix*

In an inverse system with a known Jacobian matrix, reductions in dimension and rank can decrease the memory needed to store and manipulate the large, non-sparse matrices needed for the analytic solution. Figure 1 illustrates dimension and rank reductions for an emission grid over North America. The left panel represents the original *n*-dimensional state space, i.e., the native-resolution grid. A linear transformation reduces the dimension of the state space from *n* to *k*. This transformation may reduce dimension discretely, as in the case of grid cell aggregation, or non-discretely, as in the case of a GMM. A second linear transformation can extend the dimension of the state space from *k* back to the original *n*. The resulting space, depicted in the middle panel, is a low-rank approximation of the original state space. The projection transforms the original state space to the low-rank subspace. The inverse problem can be solved in any of these three spaces.

We would like to find optimal and matrices to minimize the information loss in going from the original state vector to the reduced-dimension (or reduced-rank) state vector. This is given by the projection that preserves best the probability density function (pdf) of the original state vector **x** given the reduced dimension state vector . That pdf is defined by the prior error covariance matrix **SA.** Bousserez and Henze (2018) show that this is achieved when where . Within this class of projections over the range of the linear operators , information loss is minimized when the DOFS of the resulting reduced-rank space is maximized. Bousserez and Henze (2018) find that for a projection of this form, where **A** is the averaging kernel matrix at the original resolution (equation (4)). Define

where the columns of **W** are the eigenvectors of and the diagonal elements of are the corresponding eigenvalues of ranked in descending order. Because is a symmetric semi-positive definite matrix, the eigenvectors form an orthonormal basis for the space spanned by . As a result, and is maximized when where is the matrix of the first *k* columns of **W**, with *k* corresponding to the rank of the projected subspace. The projection that maximizes the information content of the resulting subspace is then

This projection applies a dimension-reducing transformation followed by a dimension-restoring transformation . The columns of and give the *k* leading left and right eigenvectors, respectively, of the averaging kernel matrix **A**, suggesting a method for selecting *k* (Rodgers 2000). The eigenvalues of **Q** give the eigenvalues of the averaging kernel matrix. Bousserez and Henze (2018) show that the sum of the first *k* largest eigenvalues gives the DOFS of the rank *k* subspace given by equation (6). In any inverse system, then, the fraction of information content explained by the first *k* right eigenvectors of **A** is the sum of the first *k* largest eigenvalues divided by the total DOFS. We can then select *k* so that most of the information content is explained by the *k* corresponding eigenvectors. Furthermore, the diagonal of

gives the singular values of the pre-whitened Jacobian matrix , representing the signal-to-noise ratio of each eigenvector (Rodgers 2000). The rank *k* can then be chosen so that the *k* leading eigenvectors explain most of the information content in the inverse system or so that all eigenvectors have a sufficiently large signal-to-noise ratio.

*Section 2.3: Iterative Construction of Jacobian Matrix*

Section 2.2 described the optimal reduction in dimension and rank of the Jacobian matrix, assuming knowledge of the original full-rank matrix **K**. But our goal is to avoid explicit construction of that original full-rank matrix. We present a two-step update method to construct **K** together with its reduced-dimension and reduced-rank representations. We start from an initial estimate constructed at no computational cost using a trivial assumption for atmospheric transport, as described in the next paragraph.. We calculate the averaging kernel matrix corresponding to **K(0)** and select the *k* eigenvectors that explain most of the information content in the initial system. We then propose two methods for updating the Jacobian matrix on the basis of these eigenvectors. First, we construct a multiscale grid that maintains resolution in areas of highest information content and generate the updated Jacobian matrix on the resulting grid (Section 2.4). Second, we construct on the basis of the dominant eigenvectors, generating a reduced-rank approximation (Section 2.5). In both cases, improves the estimate of the averaging kernel matrix and its eigenvectors by incorporating information content from forward model. We use these improved eigenvectors to calculate a final Jacobian matrix update. Further iterations do not significantly improve the estimate of the dominant eigenvectors. [Insert transition?]

*Section 2.3: Initializing the Jacobian Matrix*

A first guess of the structure of the original-resolution Jacobian matrix relating concentration enhancements to emissions can be made by assuming that emissions produce local concentration enhancements as determined by the local wind, with scaled enhancements in neighboring gridboxes. In the case of satellite observations where the measured quantity is a column concentration [molecules cm-2], the sensitivity of the column concentration *yj* in gridsquare *j* to emissions in gridbox *i,* representing the Jacobian matrix element *kij* = ∂ *yj*/∂*xi* , is given by

where is a factor that decreases with the distance of the *j*th observation from the *i*th grid cell, *M*air and *M*x are the molecular weights of dry air and the optimized species, respectively, *g* is gravity, *U* is the wind speed, *W* is the grid square dimension, and *P* is the surface pressure 016). The Jacobian is then constructed column-wise with the *i*th column given by .

*Section 2.4: Constructing the Jacobian Matrix on a Multiscale Grid*

Reducing the dimension of the state vector by aggregating grid cells lowers the computational cost of inversions with large state vectors by decreasing the number of model runs needed to construct the Jacobian matrix (Bocquet and Wu 2011; Turner and Jacob 2015). A multiscale grid that preserves resolution where information content is high and goes to coarser resolution elsewhere will allow solution of the inverse problem in areas with sufficient information at minimal computational cost. In an inverse system with a known Jacobian matrix **K**, a multiscale grid can be constructed on the basis of the eigenvectors of the averaging kernel matrix **A** that best explain the system’s information content, given by the columns of the dimension-restoring transformation . The rows of describe the contribution of each grid cell to each eigenvector. Then, the row-wise magnitude of , as measured by the L2 norm and described by the vector , is a measure of each grid cell’s significance in the low-rank information content. Where is largest, we preserve the original state vector resolution. Elsewhere, we consolidate grid cells to achieve the desired dimension. K-means clustering aggregates together spatially proximate grid cells that are likely to yield similar model responses but 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 also be used. The reduced-dimension Jacobian matrix can then be constructed on the multiscale grid by calculating the model response to perturbations of each grid element.

Without a fully characterized Jacobian matrix, the multiscale grid can be constructed in a two-step update that iteratively improves an initial estimate of the inverse system’s information content. We use the initial estimate of the Jacobian matrix to calculate an averaging kernel matrix and its eigenvectors . We select the eigenvectors that have a signal-to-noise ratio greater than or equal to one and generate the significance vector . We construct an initial multiscale grid and the associated reduced-dimension Jacobian , introducing information content from the forward model to the inverse system. We regrid to the original grid with weights given by the prior emissions estimate and repeat the process of generating a multiscale grid. On the second update, we use the eigenvectors that span most of the information content from the initial estimate. While the eigenvalues of underestimate the DOFS relative to the “true” inverse system, incorporating information content from the forward model will not substantially change the rate at which the eigenvalues decrease. Then, we generate the updated significance vector . To avoid superfluous model runs, the multiscale grid is adjusted only where the relative difference between and is sufficiently large. We then construct the reduced-dimension Jacobian on the updated multiscale grid.

The information content associated with both and includes contributions from prior emissions estimates, the observations, and the forward model. As a result, the leading directions of information content associated with the are largely the same as those associated with . Further iterations would improve the characterization of the patterns of low information content relative to the true patterns, but these eigenvectors would not change the multiscale grid. We therefore take as our Jacobian matrix. The analytic inversion can the be solved exactly on the multiscale grid. In order to interpret the inverse results at the original state vector resolution, additional information must be introduced to allocate the posterior solution to the original grid. The weights could, for example, be given by the prior emissions estimate. However, this introduces additional error. To solve the inversion at the original state vector resolution, it is necessary to reduce the rank and not the dimension of the inverse system.

*Section 2.5: Constructing a Reduced-Rank Approximation of the Jacobian Matrix*

A reduced-rank Jacobian matrix characterizes the linear relationship between emissions and observations in the directions that best represent the information content of the inverse system. To understand what a reduced-rank Jacobian matrix represents, consider first an inverse system with a known Jacobian matrix **K** and the corresponding averaging kernel matrix **A**. We showed earlier that the patterns of information content, given by the eigenvectors of **A** or, equivalently, by the columns of the dimension-restoring transformation ,form an orthogonal basis for the information content of the inverse system. The eigenvectors with the largest corresponding eigenvalues explain the largest fractions of the variance in the information content. For any *k*, the *k* leading eigenvectors span a rank *k*, dimension *n* subspace of the original information content space. A Jacobian matrix can be constructed within this space by calculating the model response to perturbations given by these eigenvectors. The response of a forward model F to the *j*th eigenvector , given by the *j*th column of , is

where is a scaling factor applied to ensure numerical stability. The model responses form the columns of the matrix, which is the Jacobian for an inverse system with a reduced-dimension state space spanned by the first *k* eigenvectors of the information content. This reduced-dimension Jacobian must be transformed to the original state dimension for use in analytic inversions. If the forward model is linear, F can be written as and as . Bousserez et al. (2018) show that the reduced-rank Jacobian is given by Then, the reduced-rank Jacobian can be calculated from the reduced-dimension Jacobian by . The resulting Jacobian has dimension and rank .

In an inverse system without a known Jacobian matrix, the reduced-rank Jacobian matrix approximation can be constructed in a two-step update that iteratively improves the patterns of information content used as perturbations. We use the initial estimate of the Jacobian matrix to calculate an averaging kernel matrix and its eigenvectors . We select the eigenvectors that have a signal-to-noise ratio greater than or equal to one and calculate the model response to each of the eigenvectors using equation (9). We transform the resulting reduced-dimension Jacobian to the full-dimension state space with , generating a rank approximation of the Jacobian on the basis of the initial eigenvectors of information content.

The reduced-rank Jacobian matrix approximation introduces information from the forward model to the inverse system. We calculate the associated averaging kernel matrix and its eigenvectors . We use the eigenvectors that span most of the information content from the initial estimate, as defined by the eigenvalues of . We cannot use the updated information content to determine *k* because we reduced the rank of the inverse system, so most of the information content described by is contained in the first eigenvectors. However, underestimates the true rank of the system because information is contained in eigenvectors even when their signal-to-noise ratio is less than one. In contrast, the eigenvalues of may underestimate the DOFS relative to the “true” inverse system but are likely to accurately capture the spectrum of information content; information content from the forward model is unlikely to change the rate at which information content decreases with increasing eigenvector index. Then, on the basis of the first eigenvectors of information content, we construct an updated reduced-rank Jacobian matrix approximation as above.

The resulting Jacobian matrix is a rank approximation of the linear forward model. It accurately quantifies the forward model where the inverse system has high information content as and loses accuracy in areas with lower information content. The resulting posterior emissions and error are similarly biased. In areas with lower information content, the posterior emissions estimate tends toward the prior emissions estimate. To eliminate outliers and reduce error in the inverse solution, we set the posterior emissions estimate to the prior value in grid cells where the trace of the averaging kernel matrix is small. We discuss the selection of the threshold for information content in Section 3, where we demonstrate the reduced-dimension and reduced-rank approaches as they apply to an inversion of atmospheric methane columns observed by the GOSAT satellite over North America in July 2009.

**Section 3: Results and Discussion**

We demonstrate both the reduced-dimension and reduced-rank Jacobian matrix construction approaches in an analytic Bayesian inversion of atmospheric methane columns observed by the GOSAT satellite over North America in July 2009. We construct a true Jacobian matrix for 2,098 grid boxes at 1º x 1.25º resolution, a reduced-dimension Jacobian matrix with ~300 state vector elements, and a rank 200 Jacobian matrix. We also construct a reduced-dimension Jacobian matrix following the Gaussian mixture model (GMM) method described by Turner and Jacob (2015). We use those Jacobian matrices within the inverse framework described by Maasakkers et al. (2019), adapted from the global system to the North American domain and with artificially increased information content. We solve for and compare posterior scaling factors, error covariances, and averaging kernels. [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. We aggregate native resolution grid boxes to generate a state vector composed of 2,098 1º x 1.25º grid boxes. 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 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. Unlike Maasakkers et al. (2019), we use observations north of 60ºN. Any bias introduced will exist equally in both the true inverse system and the systems with estimated Jacobians. Figure 2 shows the 2,582 GOSAT observations used by all inversions.

Solving the analytic inversion with the true Jacobian yields 81 DOFS for the 2,098 constrained grid boxes. In order to demonstrate that the proposed methods are both computationally affordable and accurate in inverse systems with higher information content, we artificially increase the DOFS in our system. We introduce a regularizing factor that increases the weight of the observational term relative to the emissions term in the cost function:

The regularizing factor functionally decreases the observational error covariance, increasing the DOFS. We set , which increases the DOFS in the true inverse system from 81 to 195. Further increasing results in posterior emissions that fit the noise in the observations and that are not representative of typical inverse results. The true posterior emissions associated with are not physical, but produce patterns of scaling factors that are consistent with typical inverse results, as shown in the left panel of Figure 3.

The right panel of Figure 3 shows the diagonal elements of the true averaging kernel matrix. Grid boxes with large (close to one) values have more information content and can constrain emissions well. The information content displays significant spatial variability, justifying reducing the dimension or rank of the inverse system. In what follows, we demonstrate both the reduced-dimension and reduced-rank approaches to approximating the Jacobian matrix. To demonstrate the potential computational savings of the proposed methods, we reduce the number of model simulations needed by an order of magnitude, from 2,098 to ~300. We compare the resulting Jacobian matrices and posterior solutions, including emissions, error covariances, and averaging kernels, to the values for the true inverse system. [insert transition?]

We first construct an initial estimate for the Jacobian matrix following the mass balance approach described in Section 2.3. Figure 4 shows the elements of the estimated Jacobian matrix plotted against the elements of the true Jacobian matrix. While the elements of the initial estimate are of the same order of magnitude as the true Jacobian, large errors exist. However, the averaging kernel also includes contributions from the prior and observational error covariances. As a result, the leading eigenvectors of correctly identify the regions with highest information content. Moreover, despite errors in the exact patterns, the fraction of information content explained by each eigenvector is consistent with the true values, as shown in Figure 5. Introducing additional information from the forward model is unlikely to significantly change the rate at which information content accumulates when the prior and observational error covariances remain constant. We can therefore use the spectrum of information content associated with the initial estimate to identify the optimal dimension or rank of the first update to the approximated Jacobian matrix. Consider first the reduced-dimension approach.

[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.

Paragraph about reduced-dimension Jacobian, cont.

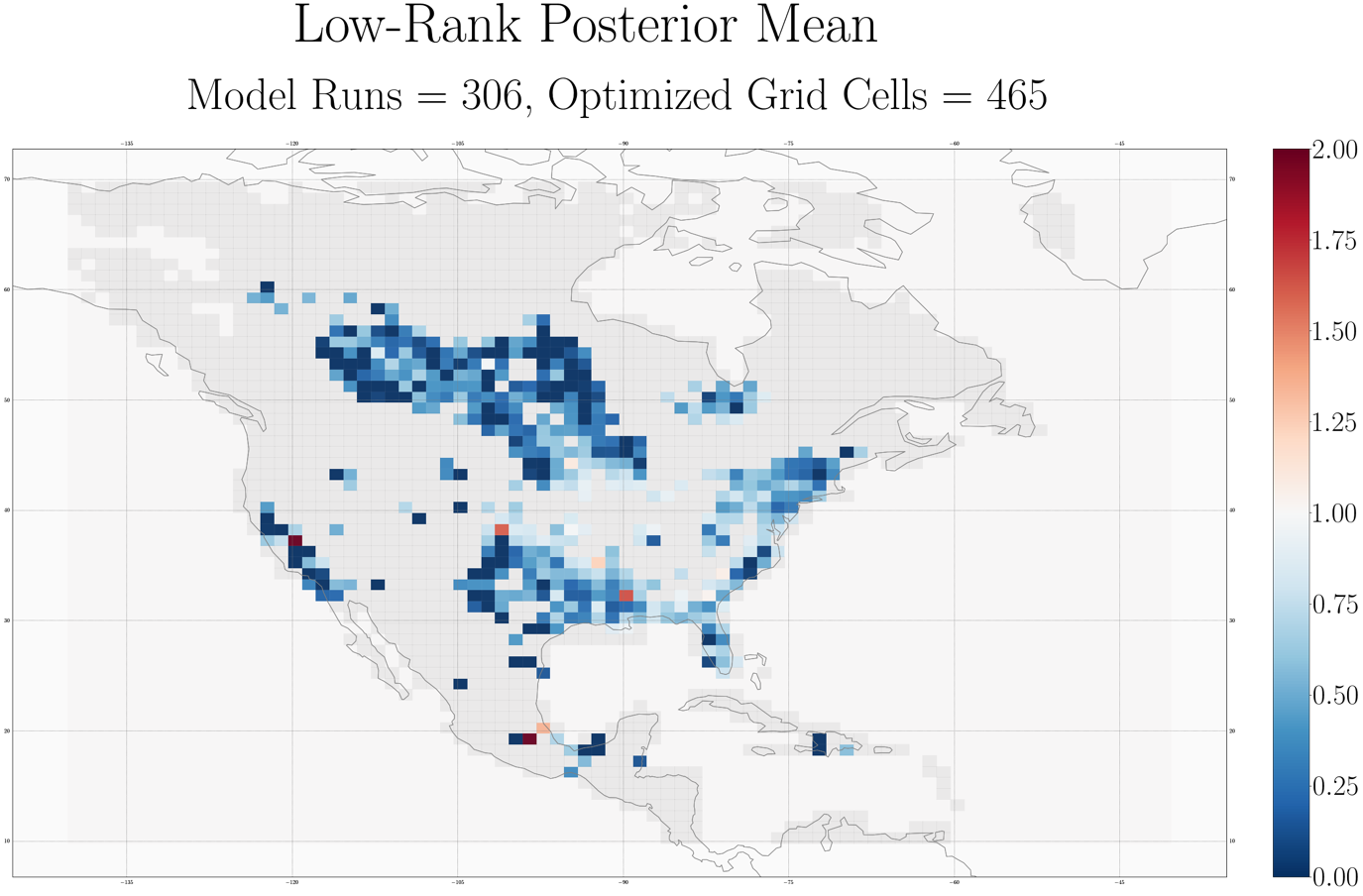
Figure 5: Plot showing the delineations of the final multi-scale grid

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.]

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.]



[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.

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]