Computationally efficient methods for large-scale atmospheric inverse modeling

Taewon Cho[1], Julianne Chung[2], Scot M. Miller[3], and Arvind K. Saibaba[4]

[1]Department of Mathematics, Virginia Tech, Blacksburg, VA
[2]Department of Mathematics and Computational Modeling and Data Analytics Division, Academy of Integrated Science, Virginia Tech, Blacksburg, VA
[3]Department of Environmental Health and Engineering, Johns Hopkins University, Baltimore, MD
[4]Department of Mathematics, North Carolina State University, Raleigh, NC

Correspondence: Julianne Chung (jmchung@vt.edu)

Abstract. Atmospheric inverse modeling describes the process of estimating greenhouse gas fluxes or air pollution emissions at the Earth's surface using observations of these gases collected in the atmosphere. The launch of new satellites, the expansion of surface observation networks, and a desire for more detailed maps of surface fluxes has yielded numerous computational and statistical challenges for standard inverse modeling frameworks that were often originally designed with much smaller data sets in mind. In this article, we discuss computationally efficient methods for large-scale atmospheric inverse modeling and focus on addressing some of the main computational and practical challenges. We develop generalized hybrid projection methods, which are iterative methods for solving large-scale inverse problems, and specifically we focus on the case of estimating surface fluxes. These algorithms confer several advantages. They are *efficient*, in part because they converge quickly, they exploit efficient matrix-vector multiplications, and do not require inverting any matrices. These methods are also *robust* because they can accurately reconstruct surface fluxes, they are *automatic* since regularization or covariance matrix parameters and stopping criteria can be determined as part of the iterative algorithm, and they are *flexible* because they can be paired with many different types of atmospheric models. We demonstrate the benefits of generalized hybrid methods with a case study from NASA's Orbiting Carbon Observatory 2 (OCO-2) satellite. We then address the more challenging problem of solving the inverse model when the mean of the surface fluxes is not known a priori; we do so by reformulating the problem, thereby extending the applicability of hybrid projection methods to include hierarchical priors. We further show that by exploiting mathematical relations provided by the generalized hybrid method, we can efficiently calculate an approximate posterior variance, thereby providing uncertainty information.

1   Introduction

Numerous satellites and ground-based sensors observe greenhouse gas and air pollution mixing ratios in the atmosphere. A primary goal of atmospheric inverse modeling (AIM) is to estimate emissions or fluxes at the Earth's surface using these observations (Brasseur and Jacob, 2017; Enting, 2002; Michalak et al., 2004; Tarantola, 2005).

The number of greenhouse gas and air pollution measurements has greatly expanded in the past decade, enabling investigations of surface fluxes across larger regions, longer time periods, and/or at finer spatial and temporal detail. Carbon dioxide (CO_2) offers an illustrative example. The Greenhouse Gas Observing Satellite (GOSAT), the first satellite dedicated to monitoring CO_2 from space, launched in 2009 and collects $\sim 1 \times 10^3$ high quality observations globally each day (e.g., Nakajima et al., 2012). NASA's Orbiting Carbon Observatory 2 (OCO-2 satellite) launched in 2014 and collects ~ 100 times more high quality observations (Crisp, 2015; Eldering et al., 2017), and upcoming satellites like the Geostationary Carbon Observatory (GeoCarb) could collect up to $\sim 1 \times 10^7$ each day (though some of these observations will likely be unusable due to cloud cover) (Buis, 2018). These new observations are complemented by an expanding ground-based network of observations (NOAA Global Monitoring Laboratory) and expanded aircraft observations, including partnerships with several airlines to measure atmospheric CO_2 from regular commercial flights (Machida et al., 01 Oct. 2008; Petzold et al., 2015). In addition, numerous cities are now monitored by dense urban networks of ground based air sensors (e.g., Shusterman et al., 2016; Davis et al., 2017; Mitchell et al., 2018).

The growing spatial coverage and sheer number of atmospheric GHG observations makes it increasing possible to estimate GHG fluxes in greater spatial and temporal detail; it is key to develop inverse models that can represent all of the information in the observations by estimating emissions at sufficiently high spatiotemporal resolution. In addition, the increasing number of observation networks opens the possibility of rapid monitoring of the carbon cycle and monitoring of changes in anthropogenic emissions – from local to global scales. There is a concomitant need to report realistic uncertainty bounds on the these emissions. Arguably, scientists require inverse modeling systems that can be used to estimate emissions and associated uncertainties quickly and at sufficient spatiotemporal resolution to facilitate this broad goal.

However, the computational challenges of achieving these goals are many. First, large-scale inverse models based on Bayesian statistics often require formulating very large covariance matrices, calculating matrix-matrix products with those covariance matrices, and/or solving linear systems with those matrices. Second, existing inverse models often assume a Gaussian prior distribution for use with Bayes' theorem, where the prior mean vector and covariance matrix are required. Statistical approaches to estimating these covariance matrix parameters (e.g., restricted maximum likelihood estimation or Markov Chain Monte Carlo methods) are often difficult to implement for extremely large inverse problems (Ganesan et al., 2014; Michalak et al., 2005), and a common approach is to populate the covariance matrices using expert knowledge. Third, fluxes often need to be estimated using iterative optimization algorithms for very large problems, and convergence of these algorithms can be slow (Miller et al., 2020). Fourth, calculating uncertainties in the estimated fluxes can be computationally prohibitive. The design of new methods to improve the computational feasibility of large atmospheric inverse problems has been the focus of numerous recent publications, see e.g., (Baker et al., 2006; Bousserez and Henze, 2018; Chatterjee and Michalak, 2013; Chatterjee et al., 2012; Chen et al., 2021a, b; Gourdji et al., 2012; Henze et al., 2007; Liu et al., 2020; Meirink et al., 2008; Miller et al., 2020, 2014; Yadav and Michalak, 2013; Zammit-Mangion et al., 2021).

Overview of features and contributions. The purpose of this study is to integrate several state-of-the-art computational and mathematical tools with AIM—tools that have been developed for and have had considerable success in other scientific fields (e.g., passive seismic tomography, medical imaging). Specifically, we investigate the use of generalized hybrid (genHyBR)

projection methods for surface flux estimation and extend their use for inverse problems where the mean of the fluxes is not known *a priori* (sometimes referred to as geostatistical inverse modeling) (Chung and Saibaba, 2017; Saibaba et al., 2020). We address these challenges in our present work.

60   Building on prior work in Miller et al. (2020), we propose a unified computational framework for large-scale AIM with the following features:

1. We describe iterative genHyBR methods that are computationally efficient since they typically converge in a few iterations, are efficient in terms of storage, and work for very large satellite-based inverse problems. For example, we demonstrate the performance of genHyBR methods on two case studies previously considered in Miller et al. (2020). In the larger case study, we solve an inverse problem with 9×10^6 unknown CO_2 fluxes and 1×10^5 CO_2 observations.

2. We extend these methods to handle the case where the mean of the prior distribution is unknown, making genHyBR applicable to a broader range of inverse modeling applications that are common in the atmospheric science community (e.g., geostatistical inverse modeling).

3. Our approach is flexible in that it can be combined with any atmospheric transport model (e.g., either Lagrangian, particle-following models or the adjoint of an Eulerian model), and it can be used with a wide variety of covariance matrices for the unknown parameters and the noise.

4. Our framework also allows for efficiently estimating regularization parameters as part of the reconstruction, thus making it easier to objectively estimate the hyperparameters or covariance matrix parameters as part of the inverse model. We focus on the discrepancy principle (DP), which requires prior knowledge of an estimate of the noise, but provide alternate methods such as the unbiased predictive risk estimator and the generalized cross validation, the latter of which does not require prior information regarding the noise level.

5. During the solution of the estimates, our solver stores information about the Krylov subspaces that can be used to estimate the posterior variance (at minimal computational cost), which gives insight into the uncertainty in the reconstructed solution. More precisely, evaluating uncertainties does not require additional model evaluations.

80   An overview of the paper is as follows. In Section 2, we describe the problem setup from a Bayesian perspective. In Section 3, we describe generalized hybrid methods for atmospheric inverse modeling. We show how to efficiently compute the maximum a posterior (MAP) estimate and uncertainty estimates (e.g., posterior variance). The focus of this section is on the fixed mean case. We briefly mention an extension to the unknown mean case, but defer most of the details to the Appendix B. Numerical results are provided in Section 4, and conclusions can be found in Section 5.

2 Bayesian approach to inverse modeling

An AIM will estimate greenhouse gas fluxes or air pollution emissions that match atmospheric observations, given an atmospheric transport model. It can be represented as an inverse problem of the form

$$z = \mathbf{H}s + \epsilon \tag{1}$$

where $z \in \mathbb{R}^m$ is a vector of atmospheric observations, $\mathbf{H} \in \mathbb{R}^{m \times n}$ represents the forward atmospheric transport model, $s \in \mathbb{R}^n$ is a vector of the unknown surface fluxes or emissions, and $\epsilon \in \mathbb{R}^m$ represents noise or errors, including errors in the atmospheric observations (z) and in the atmospheric transport model ($\mathbf{H}$). Note that s represents a vector containing spatial or spatio-temporal fluxes. Also, we assume $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ where $\mathbf{R} \in \mathbb{R}^{m \times m}$ is a positive definite matrix whose inverse and square root are inexpensive (e.g., a diagonal matrix with positive diagonal entries). The goal of the inverse problem is to estimate s given z and $\mathbf{H}$. The inverse problem may be ill-posed or under-constrained by available observations. Therefore, it is common to include prior information to mitigate the ill-posedness, which is often referred to as variational regularization (Scherzer et al., 2009; Benning and Burger, 2018). We describe two different priors: fixed mean and unknown mean.

Fixed mean. A common approach in Bayesian inverse modeling is to model s as a Gaussian random variable with a fixed, known mean $\boldsymbol{\mu} \in \mathbb{R}^n$ and prior covariance matrix $\lambda^{-2} \mathbf{Q}_s \in \mathbb{R}^{n \times n}$. In many cases, this known mean ($\boldsymbol{\mu}$) is an emissions inventory, a bottom-up flux model, or a process-based model of CO_2 fluxes (Brasseur and Jacob, 2017). This approach is also known in the literature as *Bayesian synthesis inversion*. Using this framework, the prior distribution of s is given as follows:

$$s \sim \mathcal{N}(\boldsymbol{\mu}, \lambda^{-2} \mathbf{Q}_s). \tag{2}$$

We assume that matrix $\mathbf{Q}_s$ is defined by a covariance kernel that describes the spatial and temporal variance and covariance in the prior distribution (Rasmussen and Williams, 2006). Furthermore, λ is a scaling parameter that is known *a priori* or has to be determined prior to the inversion process. The posterior distribution can be obtained by applying Bayes' theorem $\pi(s|z) \propto \pi(z|s)\pi(s)$, which takes the form

$$\pi(s|z) \propto \exp\left(-\frac{1}{2}\|\mathbf{H}s - z\|^2_{\mathbf{R}^{-1}} - \frac{\lambda^2}{2}\|s - \boldsymbol{\mu}\|^2_{\mathbf{Q}_s^{-1}}\right), \tag{3}$$

where $\|x\|^2_{\mathbf{M}} = x^\top \mathbf{M} x$ for any symmetric positive definite matrix $\mathbf{M}$, and '$\propto$' denotes a proportionality constant. The maximum a posterior (MAP) estimate corresponding to this posterior distribution can be obtained by solving the optimization problem

$$s_{\text{post}} := \arg\min_{s \in \mathbb{R}^n} \frac{1}{2}\|\mathbf{H}s - z\|^2_{\mathbf{R}^{-1}} + \frac{\lambda^2}{2}\|s - \boldsymbol{\mu}\|^2_{\mathbf{Q}_s^{-1}}. \tag{4}$$

Alternatively, it can be computed by solving the system of equations

$$(\mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H} + \lambda^2 \mathbf{Q}_s^{-1})s_{\text{post}} = \mathbf{H}^\top \mathbf{R}^{-1} z + \lambda^2 \mathbf{Q}_s^{-1} \boldsymbol{\mu}.$$

It is worth mentioning that the resulting posterior distribution is also Gaussian, with mean s_{post} and covariance $\mathbf{Q}_{\text{post}}$, denoted as $s|z \sim \mathcal{N}(s_{\text{post}}, \mathbf{Q}_{\text{post}})$.

The reconstruction quality of (1) depends crucially on choosing appropriate covariance matrix parameters, or hyperparameters, that govern this prior (2) and the noise distribution of ϵ. In Section 3.1 we describe genHyBR methods for AIM where μ is fixed but λ is not known in advance.

In many applications, the prior mean μ may also not be known in advance and must be estimated as a part of the inversion process. Some inverse models (commonly referred to as geostatistical inverse models) directly assimilate environmental data or data on emitting activities directly into the inverse model, and the relationships between the surface fluxes and these data are rarely known a priori. In other cases, an emissions inventory or bottom-up flux model may be biased too high or too low. In these cases, Eq. (2) no longer holds, violating the statistical assumptions of the inverse model. One workaround is to scale the inventory or flux model as part of the inverse modeling process, which we now describe.

Unknown mean. In cases where the prior mean is unknown, we can represent the prior information in the form of the hierarchical model

$$s|\beta \sim \mathcal{N}(\mathbf{X}\beta, \lambda^{-2}\mathbf{Q}_s), \qquad \beta \sim \mathcal{N}(\mu_\beta, \lambda_\beta^{-2}\mathbf{Q}_\beta). \tag{5}$$

Here $\mathbf{X} \in \mathbb{R}^{n \times p}$ is a fixed matrix that includes covariates (e.g., environmental data or activity data) or a bottom-up inventory/flux model, $\mathbf{Q}_s \in \mathbb{R}^{n \times n}$ is the prior covariance matrix, and λ is a scaling parameter. A set of unknown coefficients $\beta \in \mathbb{R}^p$ scale the columns of $\mathbf{X}$ and are estimated as part of the inverse model. These coefficients are assumed to follow a Gaussian distribution with given mean $\mu_\beta \in \mathbb{R}^p$, covariance matrix $\mathbf{Q}_\beta \in \mathbb{R}^{p \times p}$, and scaling parameter λ_β.

Given the assumptions in (1) and (5), from Bayes' theorem the posterior probability density function for the unknown mean case can be written as

$$\pi(s, \beta|z) \propto \pi(z|s, \beta)\pi(s|\beta)\pi(\beta)$$
$$\propto \exp\left(-\frac{1}{2}\|\mathbf{H}s - z\|_{\mathbf{R}^{-1}}^2 - \frac{\lambda^2}{2}\|s - \mathbf{X}\beta\|_{\mathbf{Q}_s^{-1}}^2 - \frac{\lambda_\beta^2}{2}\|\beta - \mu_\beta\|_{\mathbf{Q}_\beta^{-1}}^2\right). \tag{6}$$

The MAP estimate can be written as the solution of the optimization problem

$$\gamma_{\text{post}} = \underset{\gamma=[s^\top, \beta^\top]^\top}{\arg\min} \frac{1}{2}\|\mathbf{H}s - z\|_{\mathbf{R}^{-1}}^2 + \frac{\lambda^2}{2}\|s - \mathbf{X}\beta\|_{\mathbf{Q}_s^{-1}}^2 + \frac{\lambda_\beta^2}{2}\|\beta - \mu_\beta\|_{\mathbf{Q}_\beta^{-1}}^2. \tag{7}$$

The posterior distribution in (6) is Gaussian; therefore, the mean of the posterior distribution is also the MAP estimate and the covariance is the inverse of the Hessian matrix of (7) and is given by

$$\Gamma_{\text{post}} = \begin{bmatrix} \lambda^2\mathbf{Q}_s^{-1} + \mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H} & -\lambda^2\mathbf{Q}_s^{-1}\mathbf{X} \\ -\lambda^2\mathbf{X}^\top\mathbf{Q}_s^{-1} & \lambda_\beta^2\mathbf{Q}_\beta^{-1} + \lambda^2\mathbf{X}^\top\mathbf{Q}_s^{-1}\mathbf{X} \end{bmatrix}^{-1}. \tag{8}$$

Therefore, the resulting posterior distribution is

$$\gamma \,|\, z \sim \mathcal{N}\left(\gamma_{\text{post}}, \Gamma_{\text{post}}\right).$$

Here, λ and λ_β are scaling parameters that may not be known in advance, but we assume that $\lambda_\beta = \alpha\lambda$ with a constant $\alpha > 0$, where α is set in advance. We describe genHyBR methods for the unknown mean case in Section 3.3.

Note that previous works (Miller et al., 2020; Saibaba and Kitanidis, 2015) assume an improper prior for $\boldsymbol{\beta}$ (i.e., $p(\boldsymbol{\beta}) \propto 1$), in which case, a solution estimate can be obtained as $\widehat{\boldsymbol{s}} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Q}_s \mathbf{H}^\top \widehat{\boldsymbol{\xi}}$, where

$$
\begin{bmatrix}
\mathbf{H}\mathbf{Q}_s\mathbf{H}^\top + \mathbf{R} & \mathbf{H}\mathbf{X} \\
(\mathbf{H}\mathbf{X})^\top & \mathbf{0}
\end{bmatrix}
\begin{bmatrix}
\widehat{\boldsymbol{\xi}} \\
\widehat{\boldsymbol{\beta}}
\end{bmatrix}
=
\begin{bmatrix}
\boldsymbol{z} \\
\mathbf{0}
\end{bmatrix}.
\tag{9}
$$

The system in (9) is often referred to as the dual function form, and there are several equivalent formulations of these equations (Michalak et al., 2004). The size of the resulting system of equations is $(m + p) \times (m + p)$, where m is the number of measurements and p is the number of unknown parameters in $\boldsymbol{\beta}$, so forming or inverting the matrix in (9) is infeasible in many applications. The approach taken in Saibaba and Kitanidis (2012) and Miller et al. (2020) uses a matrix-free iterative method to solve (9); however, the number of required iterations can be very large, especially for problems with many measurements, and even with the use of a preconditioner to speed convergence.

In this paper, we follow a different approach to handle the unknown mean case by using iterative hybrid approaches on a reformulated problem. Since these methods work directly on the least-squares problem (7), the number of unknown parameters is $n + p$. However, the size of the linear system that defines the MAP estimate is independent of the number of observations, making it attractive for large datasets. Furthermore, our framework can handle a wide class of prior covariance operators, where the resulting prior covariance matrices are large and dense and explicitly forming and factorizing these matrices is prohibitively expensive. These include, for example, prior covariance matrices that arise from nonseparable, spatiotemporal covariance kernels and parameterized kernels on non-uniform grids. Our approach only relies on forming matrix-vector products with the covariance matrices, and is compatible with acceleration techniques using Fast Fourier Transform (FFT) or hierarchical matrix approaches. See Chung and Saibaba (2017); Chung et al. (2018) for a detailed discussion. Thus, as we show in Section 4, our approach can incorporate various prior models and can scale to very large data sets.

3   Generalized hybrid projection methods for AIM

In this section, we describe generalized hybrid projection methods, dubbed genHyBR methods, for AIM. Hybrid projection methods were first developed in the 1980's as a way to combine iterative projection methods (e.g., Krylov subspace methods) and variational regularization methods (e.g., Tikhonov regularization) for solving very large inverse problems. These are iterative methods, where each iteration requires the expansion of the solution subspace, the estimation of the regularization parameter(s), and the solution of a projected, regularized problem. We point the interested reader to survey papers (Chung and Gazzola, 2021; Gazzola and Sabaté Landman, 2020). In Chung and Saibaba (2017), genHyBR methods were developed for computing Tikhonov regularized solutions to problems where explicit computations of the square root and inverse of the prior covariance matrix are not feasible. This work enabled hybrid projection methods for more general regularization terms. The main benefits of genHyBR methods that make them ideal for large large-scale AIM are *efficiency*, due to fast convergence to an accurate reconstruction of surface fluxes where efficient matrix-vector multiplications are exploited at each iteration, *automatic* estimation of parameters (e.g., hyperparameters and algorithmic parameters), and *flexibility* because they can be paired with many different atmospheric models.

We describe genHyBR methods for both the fixed mean case (Section 3.1) and the unknown mean case (Section 3.3), with particular emphasis on the associated challenges for large datasets and subsequent UQ (Section 3.2). We provide a general overview of our approach, including the main components of genHyBR methods, in the flowchart in Figure 1.
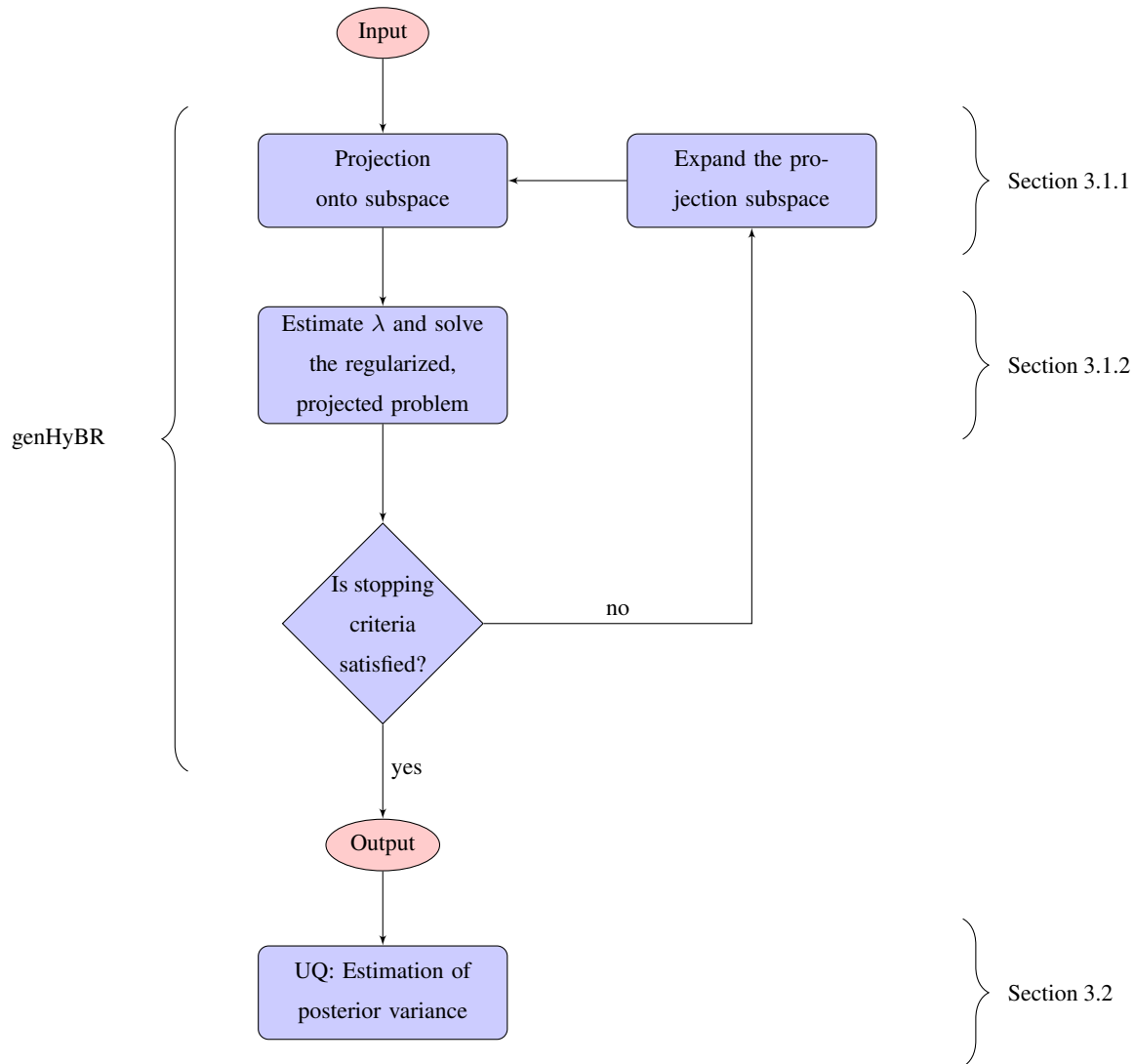
Figure 1. This flowchart provides a general overview of using genHyBR methods for AIM and subsequent UQ. Given input (corresponding to the observations and details of the problem setup), genHyBR is an iterative approach to approximate the MAP estimate. Each iteration of genHyBR consists of expanding the solution subspace, projecting the problem, estimating a regularization parameter, and solving a projected, regularized problem. After obtaining the MAP estimate, information computed from genHyBR can be used to efficiently estimate the posterior variance for UQ.

3.1 Generalized hybrid methods with fixed mean

To introduce genHyBR methods, we begin with the fixed mean case described in Section 2. If symmetric decompositions $\mathbf{R}^{-1} = \mathbf{L}_{\mathbf{R}}^{\top} \mathbf{L}_{\mathbf{R}}$ and $\mathbf{Q}_s^{-1} = \mathbf{L}_s^{\top} \mathbf{L}_s$ are available, then optimization problem (4) can be rewritten in the standard least-squares form

$$\boldsymbol{s}_{\text{post}} = \operatorname*{arg\,min}_{\boldsymbol{s} \in \mathbb{R}^n} \frac{1}{2} \left\| \begin{bmatrix} \mathbf{L}_{\mathbf{R}} \mathbf{H} \\ \lambda \mathbf{L}_s \end{bmatrix} \boldsymbol{s} - \begin{bmatrix} \mathbf{L}_{\mathbf{R}} \boldsymbol{z} \\ \lambda \mathbf{L}_s \boldsymbol{\mu} \end{bmatrix} \right\|_2^2 .$$

However, computing $\mathbf{L}_s$ can be computationally infeasible for large n and λ may not be known *a priori*. This motivates us to use the following change of variables,

$$\boldsymbol{x} \leftarrow \mathbf{Q}_s^{-1}(\boldsymbol{s} - \boldsymbol{\mu}), \quad \boldsymbol{b} \leftarrow \boldsymbol{z} - \mathbf{H}\boldsymbol{\mu}, \tag{10}$$

in which case, the solution to problem (4) is given by $\boldsymbol{s}_{\text{post}} = \boldsymbol{\mu} + \mathbf{Q}_s \boldsymbol{x}$ where $\boldsymbol{x}$ solves

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{H}\mathbf{Q}_s \boldsymbol{x} - \boldsymbol{b}\|_{\mathbf{R}^{-1}}^2 + \frac{\lambda^2}{2} \|\boldsymbol{x}\|_{\mathbf{Q}_s}^2 . \tag{11}$$

Note that, with this reformulation, we avoid $\mathbf{L}_s$, $\mathbf{L}_s^{-1}$, and $\mathbf{Q}_s^{-1}$ and only require matrix-vector products with $\mathbf{Q}_s$. Furthermore, for iterative methods for (11), the matrix $\mathbf{H}$ does not need to be formed explicitly, as we only need access to matrix-vector and matrix-transpose-vector products.

There are two main ingredients in the genHyBR approach: (1) the generalized Golub-Kahan bidiagonalization approach for constructing a solution subspace and (2) regularization parameter estimation methods for computing a suitable regularization parameter in the projected space.

3.1.1 Generalized Golub-Kahan bidiagonalization

We now describe the generalized Golub-Kahan bidiagonalization approach that is the backbone of the genHyBR method. Given matrices $\mathbf{H}$, $\mathbf{R}$, $\mathbf{Q}_s$ and vector $\boldsymbol{b}$ from (11), the basic idea is to generate a set of basis vectors contained in $\mathbf{V}_k$ for the Krylov subspace,

$$\mathcal{S}_k \equiv \mathcal{R}(\mathbf{V}_k) = \mathcal{K}_k(\mathbf{H}^{\top} \mathbf{R}^{-1} \mathbf{H} \mathbf{Q}_s, \mathbf{H}^{\top} \mathbf{R}^{-1} \boldsymbol{b})$$

where $\mathcal{R}(\cdot)$ denotes the column space and the Krylov subspace is $\mathcal{K}_k(\mathbf{M}, \boldsymbol{f}) = \text{Span}\{\boldsymbol{f}, \mathbf{M}\boldsymbol{f}, \ldots, \mathbf{M}^{k-1}\boldsymbol{f}\}$. The generated basis vectors span a low-dimensional subspace that is rich in information about important directions; thus, solutions to the (smaller) projected problem often provide good approximations to the solution of the high-dimensional problem. With initializations $\delta_1 = \|\boldsymbol{b}\|_{\mathbf{R}^{-1}}$, $\boldsymbol{u}_1 = \boldsymbol{b}/\delta_1$ and $\gamma_1 \boldsymbol{v}_1 = \mathbf{H}^{\top} \mathbf{R}^{-1} \boldsymbol{u}_1$, the kth iteration of the generalized Golub-Kahan bidiagonalization procedure generates vectors $\boldsymbol{u}_{k+1}$ and $\boldsymbol{v}_{k+1}$ such that

$$\gamma_{k+1} \boldsymbol{u}_{k+1} = \mathbf{H}\mathbf{Q}_s \boldsymbol{v}_k - \gamma_k \boldsymbol{u}_k$$

$$\delta_{k+1} \boldsymbol{v}_{k+1} = \mathbf{H}^{\top} \mathbf{R}^{-1} \boldsymbol{u}_{k+1} - \delta_{k+1} \boldsymbol{v}_k$$

205   where scalars $\gamma_k, \delta_k \geq 0$ are computed such that $\|\boldsymbol{u}_k\|_{\mathbf{R}^{-1}} = \|\boldsymbol{v}_k\|_{\mathbf{Q}_s} = 1$. At the end of k iterations, we have

$$\mathbf{B}_k \equiv \begin{bmatrix} \gamma_1 & 0 & \cdots & 0 \\ \delta_2 & \gamma_2 & \ddots & \vdots \\ 0 & \delta_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \gamma_k \\ 0 & \cdots & 0 & \delta_{k+1} \end{bmatrix}, \quad \mathbf{U}_{k+1} \equiv \begin{bmatrix} \boldsymbol{u}_1, \ldots, \boldsymbol{u}_{k+1} \end{bmatrix}, \quad \text{and} \quad \mathbf{V}_k \equiv \begin{bmatrix} \boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \end{bmatrix},$$

where the following relations hold

$$\mathbf{H}\mathbf{Q}_s \mathbf{V}_k = \mathbf{U}_{k+1}\mathbf{B}_k \quad \text{and} \quad \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{U}_{k+1} = \mathbf{V}_k \mathbf{B}_k^\top + \gamma_{k+1}\boldsymbol{v}_{k+1}\boldsymbol{e}_{k+1}^\top, \tag{12}$$

where $\boldsymbol{e}_j$ is the jth column of an identity matrix with the appropriate dimensions. Also, matrices $\mathbf{U}_{k+1}$ and $\mathbf{V}_k$ satisfy the
210   following orthogonality conditions

$$\mathbf{U}_{k+1}^\top \mathbf{R}^{-1}\mathbf{U}_{k+1} = \mathbf{I}_{k+1} \quad \text{and} \quad \mathbf{V}_k^\top \mathbf{Q}_s \mathbf{V}_k = \mathbf{I}_k, \tag{13}$$

with $\mathbf{U}_{k+1}\delta_1 \boldsymbol{e}_1 = \boldsymbol{b}$. Then, for given $\lambda > 0$ the solution to (11) is recovered by $\boldsymbol{x}_{k,\lambda} = \mathbf{Q}_s \mathbf{V}_k \boldsymbol{y}_{k,\lambda}$ where $\boldsymbol{y}_{k,\lambda}$ is the solution to
the regularized, projected problem

$$\min_{\boldsymbol{y} \in \mathbb{R}^k} \frac{1}{2}\|\mathbf{B}_k \boldsymbol{y} - \delta_1 \boldsymbol{e}_1\|_2^2 + \frac{\lambda^2}{2}\|\boldsymbol{y}\|_2^2. \tag{14}$$

215   Notice that (14) is a standard least-squares problem with Tikhonov regularization, and since the coefficient matrix $\mathbf{B}_k$ is
of size $(k+1) \times k$, the solution can be computed efficiently (Björck, 1996). Each iteration of the generalized Golub-Kahan
bidiagonalization process requires one matrix-vector product with $\mathbf{H}$ and its adjoint (suppose we denote its cost by $T_{\mathbf{H}}$), two
matrix-vector products with $\mathbf{Q}_s$ (similarly, denoted $T_{\mathbf{Q}_s}$) and additional $\mathcal{O}(m+n)$ floating point operations (flops). To compute
the solution of the least-squares problem (14), the cost is $\mathcal{O}(k^3)$ flops, and the cost of forming $\boldsymbol{x}_{k,\lambda}$ is $\mathcal{O}(nk)$ flops. Thus, the
220   overall cost of the algorithm is

$$T_{\text{genGK}} = 2k(T_{\mathbf{H}} + T_{\mathbf{Q}_s}) + \mathcal{O}(k(m+n)) \text{ flops.}$$

In practice, the vectors $\{\boldsymbol{u}_k\}$ and $\{\boldsymbol{v}_k\}$ lose orthogonality in floating point arithmetic, so full or partial re-orthogonalization
(Barlow, 2013) can be used to ensure orthogonality. This costs an additional $\mathcal{O}(k^2(m+n))$ flops. Thus far we have described
an iterative method for approximating the MAP estimate (4), where the kth iterate is given by $\boldsymbol{s}_{k,\lambda} = \boldsymbol{\mu} + \mathbf{Q}_s \boldsymbol{x}_{k,\lambda}$ for fixed
225   regularization parameter λ.

3.1.2   Regularization parameter estimation methods

In this subsection, we highlight one of the main computational benefits of hybrid projection methods, which is the ability
to estimate regularization parameters efficiently and adaptively, while still ensuring robustness of the solution. For genHyBR

approaches, we use the generalized Golub-Kahan bidiagonalization to generate a projection subspace, and solve the projected
230 problem (14), while simultaneously estimating the regularization parameter λ. Notice that the regularization term in the projected system (14) is standard Tikhonov regularization, and a plethora of parameter estimation methods exist for Tikhonov regularization (Bardsley, 2018; Hansen, 2010). Here we focus on the discrepancy principle (DP) and point the interested reader to Appendix A for further details on other parameter estimation methods can be incorporated within genHyBR methods for AIM.

235 The discrepancy principle (DP) is a common approach for estimating a regularization parameter, where the main goal is to determine λ such that the residual norm for the regularized reconstruction matches a given estimate of the noise level in the observations. That is, the DP method selects the largest parameter value λ for which the reconstructed fluxes $\boldsymbol{s}_\lambda$ satisfy

$$\mathcal{D}_{\text{full}}(\lambda) = \|\mathbf{H}\boldsymbol{s}_\lambda - \boldsymbol{z}\|_{\mathbf{R}^{-1}}^2 \leq \tau m, \tag{15}$$

where $\tau \geq 1$ is a user-defined parameter and m is the expected value of $\|\boldsymbol{\epsilon}\|_{\mathbf{R}^{-1}}^2$. Typical choices for safety factor τ are in the
240 range $1 \leq \tau \leq 2$.

The DP has been used in AIM (Hase et al., 2017), as well as more generally in inverse problems (Groetsch, 1983; Hansen, 2010). For a given λ, evaluating $\mathcal{D}_{\text{full}}(\lambda)$ requires computing $\boldsymbol{s}_\lambda$ and matrix-vector multiplication with $\mathbf{H}$, which can get costly if many different values of λ are desired. However, a major distinction here is that by using a iterative hybrid formulation, we can exploit relationships in Eq. (12) for more efficient parameter selection. In particular, the residual norm can be simplified as

245

$$\|\mathbf{H}\boldsymbol{s}_{k,\lambda} - \boldsymbol{z}\|_{\mathbf{R}^{-1}}^2 = \|\mathbf{B}_k\boldsymbol{y}_{k,\lambda} - \delta_1\boldsymbol{e}_1\|_2^2 \equiv \mathcal{D}_{\text{proj}}(\lambda). \tag{16}$$

Thus, we let λ_k be the regularization parameter estimated for the projected problem at the kth iteration, such that $\mathcal{D}_{\text{proj}}(\lambda_k) \leq \tau m$. Then as the number of iterations k increases, the estimated DP regularization parameter for the projected problem becomes a better approximation of the DP parameter for the original problem.

250 The advantage of this approach is two-fold: the regularization parameter is selected adaptively (i.e., each iteration can have a different regularization parameter), and the cost of parameter selection is cheap ($\mathcal{O}(k^3)$ flops) since we work with small matrices of size $(k+1) \times k$ and k is much smaller than m and n. Furthermore, there are various theoretical results that show that selecting the regularization parameter for the projected problem (i.e., project-then-regularize) is equivalent to first estimating the regularization parameter and then using an iterative projection method (i.e., regularize-then-project) (Chung and Gazzola,
255 2021).

3.2 Approximation to the posterior covariance matrices

In the Bayesian approach for fixed parameter λ, the posterior distribution (3) is Gaussian and, thus, is fully specified by the mean and covariance matrix. However, neither computing nor storing the covariance matrix is feasible, making further uncertainty estimation challenging. Instead, we follow the approach described in Chung et al. (2018) and Saibaba et al. (2020) for the fixed
260 mean case, where an approximation to the posterior covariance matrix is obtained using the computed vectors generated during

the generalized Golub-Kahan bidiagonalization process. An advantage of this approach is that, by storing partial information while computing the MAP estimate, we can approximately compute the uncertainty associated with the MAP estimate (e.g., posterior variance) with minimal additional cost, and no further accesses to the forward and adjoint models. For the fixed mean case, we refer to this approach as `genHyBRs` with UQ and provide a summary in Algorithm 1.

Algorithm 1 AIM with fixed mean—`genHyBRs` with UQ

Require: Matrices $\mathbf{H}$, $\mathbf{R}$ and $\mathbf{Q}$, and vector $\boldsymbol{b}$.

 1: { Compute MAP estimate }

 2: initialize $\boldsymbol{u}_1 = \boldsymbol{b}/\|\boldsymbol{b}\|_{\mathbf{R}^{-1}}$

 3: **for** $j = 1, \ldots, k$ **do**

 4:     one iteration of generalized Golub-Kahan bidiagonalization to obtain $\mathbf{B}_j, \mathbf{U}_{j+1}$, and $\mathbf{V}_j$

 5:     estimate regularization parameter λ and compute $\boldsymbol{x}_{j,\lambda} = \mathbf{Q}_s \mathbf{V}_j \boldsymbol{y}_{j,\lambda}$ where $\boldsymbol{y}_{j,\lambda}$ solves (14)

 6: **end for**

 7: compute the MAP estimate $\boldsymbol{s}_{k,\lambda} = \boldsymbol{\mu} + \mathbf{Q}_s \boldsymbol{x}_{k,\lambda}$

 8: { Compute the approximation to the posterior variance }

 9: compute the eigendecomposition $\mathbf{B}_k^\top \mathbf{B}_k = \mathbf{W}_k \mathbf{\Theta}_k \mathbf{W}_k^\top$

10: compute $\mathbf{Z}_k = \mathbf{Q}_s \mathbf{V}_k \mathbf{W}_k$ and $\mathbf{\Delta}_k = \lambda^{-2} \mathrm{diag}\left(\frac{\theta_1}{\theta_1 + \lambda^2}, \ldots, \frac{\theta_k}{\theta_k + \lambda^2}\right)$

11: compute $\boldsymbol{d}_{\mathrm{LR}} = \mathrm{LowRankDiag}(\mathbf{Z}_k \mathbf{\Delta}_k, \mathbf{Z}_k)$ using Algorithm 2 and $\boldsymbol{d}_s = \mathrm{diag}(\mathbf{Q}_s)$

12: estimate diagonal of approximate posterior covariance matrix $\boldsymbol{d}_{k,\lambda} = \lambda^{-2} \boldsymbol{d}_s - \boldsymbol{d}_{\mathrm{LR}}$

13: **return** MAP estimate $\boldsymbol{s}_{k,\lambda}$ and variance estimate $\boldsymbol{d}_{k,\lambda}$

---

265   In the following, we provide some details regarding the estimation of the posterior variance in the known mean case. This material has previously appeared in Chung et al. (2018, Section 4.1), but we provide a brief description here for completeness. An alternative expression for the posterior covariance is

$$\mathbf{Q}_{\mathrm{post}} = \mathbf{Q}_s(\lambda^2 \mathbf{Q}_s + \mathbf{Q}_s \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H} \mathbf{Q}_s)^{-1} \mathbf{Q}_s,$$

which is obtained by factoring out $\mathbf{Q}_s$. This expression is not computationally feasible for large inverse problems but can be
270 approximated using the outputs of the generalized Golub-Kahan bidiagonalization (described in Subsection 3.1.1). After k steps of the generalized Golub-Kahan bidiagonalization approach, we have matrices $\mathbf{U}_{k+1}, \mathbf{V}_k$, and $\mathbf{B}_k$. Let $\mathbf{B}_k^\top \mathbf{B}_k = \mathbf{W}_k \mathbf{\Theta}_k \mathbf{W}_k^\top$ be the eigenvalue decomposition with eigenvalues $\theta_1, \ldots, \theta_k$. Next, we compute the matrix $\mathbf{Z}_k = \mathbf{Q}_s \mathbf{V}_k \mathbf{W}_k$ and the diagonal matrix

$$\mathbf{\Delta}_k = \lambda^{-2} \begin{bmatrix} \frac{\theta_1}{\theta_1 + \lambda^2} & & \\ & \ddots & \\ & & \frac{\theta_k}{\theta_k + \lambda^2} \end{bmatrix} \in \mathbb{R}^{k \times k}.$$

275 Then we can approximate the posterior covariance matrix as

$$\widetilde{\mathbf{Q}}_{\mathrm{post}} = \mathbf{Q}_s(\lambda^2 \mathbf{Q}_s + \mathbf{Z}_k \mathbf{\Theta}_k \mathbf{Z}_k^\top)^{-1} \mathbf{Q}_s = \lambda^{-2} \mathbf{Q}_s - \mathbf{Z}_k \mathbf{\Delta}_k \mathbf{Z}_k^\top,$$

providing an efficient representation of the posterior covariance matrix (as a low-rank perturbation of the prior covariance matrix) that can be used for efficient uncertainty quantification. The accuracy of the approximate posterior covariance matrix and of the resulting posterior distribution can be monitored using the information available from the generalized Golub-Kahan bidiagonalization (Saibaba et al., 2020). Note that in the hybrid approach, we estimate the regularization parameters typically at every iteration, the results in (Saibaba et al., 2020) are to be applied after the iterations have been terminated and the regularization parameter has been estimated. The uncertainty estimates therefore depend on the value of the regularization parameter. In general, the approximate posterior variance overestimates the uncertainty and decreases monotonically with the number of iterations k.

To visualize the uncertainty, it is common to compute the diagonals of the posterior covariance (known as the posterior variance). The diagonals of $\mathbf{Q}_s$ are typically known analytically; the diagonals of the low-rank term $\mathbf{Z}_k \boldsymbol{\Delta}_k \mathbf{Z}_k^\top$ can be computed efficiently using Algorithm 2 with input $\mathbf{Y} = \mathbf{Z}_k \boldsymbol{\Delta}_k$ and $\mathbf{Z} = \mathbf{Z}_k$. An important point worth emphasizing is that the approximation to the posterior covariance need not be computed explicitly. More precisely, in addition to storing the information required for storing $\mathbf{Q}_s$, we only need to store $nk + k$ additional entries corresponding to the matrices $\mathbf{Z}_k$ and $\boldsymbol{\Delta}_k$.

In addition, one can also use genHyBRs to compute the posterior variance of the sum of the fluxes (or analogously, the variance of the mean). To do so, let $\mathbf{1}$ denote an $n \times 1$ vector of ones and multiply the components of $\widetilde{\mathbf{Q}}_{\text{post}}$ as

$$\mathbf{1}^\top \widetilde{\mathbf{Q}}_{\text{post}} \mathbf{1} = \lambda^{-2} (\mathbf{1}^\top \mathbf{Q_s} \mathbf{1}) - (\mathbf{1}^\top \mathbf{Z_k}) \boldsymbol{\Delta_k} (\mathbf{1}^\top \mathbf{Z_k})^\top.$$

Several existing studies (Yadav and Michalak, 2013; Miller et al., 2020) describe how to efficiently compute $\mathbf{1}^\top \mathbf{Q_s} \mathbf{1}$ using Kronecker products.

Note that in previous works (Chung et al., 2018; Saibaba et al., 2020), we found that additional reorthogonalization of the generalized Golub-Kahan basis vectors yielded more accurate results, so we perform them in the numerical experiments.

Algorithm 2 Compute the diagonals of the low-rank term $\mathbf{YZ}^\top$. Call as $[\boldsymbol{d}] = \text{LowRankDiag}(\mathbf{Y}, \mathbf{Z})$

Require: Matrices $\mathbf{Y}, \mathbf{Z} \in \mathbb{R}^{n \times k}$ defining the outer product $\mathbf{YZ}^\top$

  1: **for** $i = 1, \ldots, n$ **do**

  2:     $d_i = \sum_{j=1}^{k} \mathbf{Y}_{ij} \mathbf{Z}_{ij}$

  3: **end for**

  4: **return** vector $\boldsymbol{d} \in \mathbb{R}^n$ containing the diagonals of $\mathbf{YZ}^\top$

3.3 Hierarchical Gaussian priors: Reformulation for mean estimation

Next we describe genHyBR methods for AIM with unknown mean as described in Section 2 with assumptions given in Eq. (5). First, we reformulate the problem for simultaneous estimation of the surface fluxes in $\boldsymbol{s}$ and the covariate parameters in $\boldsymbol{\beta}$. Then, we describe how to use genHyBR methods for computing the corresponding MAP estimate (7) and for subsequent UQ. We refer to this approach as genHyBRmean with UQ, and since the derivations follow those in Sections 3.1 and 3.2, specific

details have been relegated to Appendix B. However, we would like to emphasize that this derivation is new and a contribution of this work.

Notice that optimization problem (7) can be rewritten in standard least-squares form

$$\gamma_{\text{post}} = \operatorname*{arg\,min}_{\gamma=[s^\top, \beta^\top]^\top} \frac{1}{2} \left\| \begin{bmatrix} \mathbf{L_R H} & \mathbf{0} \\ \lambda \mathbf{L}_s & -\lambda \mathbf{L}_s \mathbf{X} \\ \mathbf{0} & \lambda_\beta \mathbf{L}_\beta \end{bmatrix} \begin{bmatrix} s \\ \beta \end{bmatrix} - \begin{bmatrix} \mathbf{L_R} z \\ \mathbf{0} \\ \lambda_\beta \mathbf{L}_\beta \boldsymbol{\mu}_\beta \end{bmatrix} \right\|_2^2,$$

if symmetric decompositions $\mathbf{R}^{-1} = \mathbf{L_R}^\top \mathbf{L_R}, \mathbf{Q}_s^{-1} = \mathbf{L}_s^\top \mathbf{L}_s$, and $\mathbf{Q}_\beta^{-1} = \mathbf{L}_\beta^\top \mathbf{L}_\beta$ are available. Since computing $\mathbf{L}_s$ can be computationally infeasible (e.g., for spatiotemporal covariance matrices where n is large), we propose a similar change of variables to avoid $\mathbf{L}_s$,

$$\widetilde{s} \leftarrow s - \mathbf{X}\boldsymbol{\mu}_\beta, \quad \widetilde{\beta} \leftarrow \beta - \boldsymbol{\mu}_\beta, \quad \widetilde{z} \leftarrow z - \mathbf{H}\mathbf{X}\boldsymbol{\mu}_\beta. \tag{17}$$

We define the concatenated vector $\widetilde{\gamma} = [\widetilde{s}^\top, \widetilde{\beta}^\top]^\top$ and let $\mathbf{K} = \begin{bmatrix} \mathbf{H} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{m \times (n+p)}$, where $\mathbf{0} \in \mathbb{R}^{m \times p}$ is included to ensure proper matrix multiplication. Then, optimization problem (7) can be written as

$$\min_{\widetilde{\gamma} \in \mathbb{R}^{n+p}} \frac{1}{2} \| \mathbf{K}\widetilde{\gamma} - \widetilde{z} \|_{\mathbf{R}^{-1}}^2 + \frac{\lambda^2}{2} \| \widetilde{\gamma} \|_{\mathbf{Q}^{-1}}^2 \tag{18}$$

where

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_s + \frac{1}{\alpha^2} \mathbf{X}\mathbf{Q}_\beta \mathbf{X}^\top & \frac{1}{\alpha^2} \mathbf{X}\mathbf{Q}_\beta \\ \frac{1}{\alpha^2} (\mathbf{X}\mathbf{Q}_\beta)^\top & \frac{1}{\alpha^2} \mathbf{Q}_\beta \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \frac{1}{\alpha^2} \begin{bmatrix} \mathbf{X} \\ \mathbf{I} \end{bmatrix} \mathbf{Q}_\beta \begin{bmatrix} \mathbf{X}^\top & \mathbf{I} \end{bmatrix}. \tag{19}$$

Derivations are provided in Appendix B1. Note that, in practice, neither of the matrices $\mathbf{H}$ nor $\mathbf{K}$ need to be formed explicitly since we only need access to matrix-vector products with these matrices and their transposes. Also we do not explicitly construct $\mathbf{Q}$, but instead provide an efficient way to form matrix-vector products with $\mathbf{Q}$.

In summary, to handle AIM with unknown mean, the genHyBR method can be used to solve Eq. (18) (as described in Section 3.1 with $\mathbf{Q}$ instead of $\mathbf{Q}_s$, $\mathbf{K}$ instead of $\mathbf{H}$, and $\widetilde{z}$ instead of $\boldsymbol{b}$) to efficiently obtain the solution $\widetilde{\gamma}_{k,\lambda} = [\widetilde{s}_{k,\lambda}^\top, \widetilde{\beta}_{k,\lambda}^\top]^\top$. Then, we recover the MAP estimate for s and β as

$$\gamma_{\text{post}} \approx \gamma_{k,\lambda} := \begin{bmatrix} \widetilde{s}_{k,\lambda} + \mathbf{X}\boldsymbol{\mu}_\beta \\ \widetilde{\beta}_{k,\lambda} + \boldsymbol{\mu}_\beta \end{bmatrix}. \tag{20}$$

Similar to the fixed mean case, we can efficiently approximate the posterior covariance matrix and its diagonals using elements of the generalized Golub-Kahan bidiagonalization algorithm, as described in Appendix B3. We remark that efficient UQ approaches for the unknown mean case were considered in Saibaba and Kitanidis (2015), but our approach differs in that we reuse information contained in the subspaces generated during the iterative method, rather than randomization techniques, making these derivations straightforward and the approaches widely applicable.

4 Numerical results

We evaluate the inverse modeling algorithms described in this paper using the two case studies described in Section 4.1. For the numerical experiments, we denote the two methods we test as

- `genHyBRs` refers to the fixed mean case, and involves solving the optimization problem (4), using the approach described in Sections 3.1.1 and 3.1.2,

- `genHyBRmean` refers to the unknown mean case, and involves solving the optimization problem (18), using the approach described in Section 3.3.

For comparison, we use a direct inversion method, which solves Eq. (9) using MATLAB's "backslash" operator. Numerical experiments presented here were obtained using MATLAB on a compute server with four Intel 15 core 2.8 GHz processors and 1 TB of RAM.

4.1 Overview of the case studies

We explore two case studies on estimating CO_2 fluxes across North America using observations from NASA's OCO-2 satellite. In the first case study, we estimate CO_2 fluxes for 6 weeks (late June through July 2015), an inverse problem that is small enough to estimate using the direct method. The second case study using one year of observations (Sept. 2014 − Aug. 2015) is too large to estimate directly on many or most computer systems. The goal of these experiments is to demonstrate the performance of the generalized hybrid methods for solving the inverse problem with automatic parameter selection.

The case studies explored here parallel those in Miller et al. (2020) and Liu et al. (2020). We provide an overview of these case studies, but refer to Miller et al. (2020) for additional detail on the specific setup. Both of the case studies use synthetic OCO-2 observations that are generated using CO_2 fluxes from NOAA's CarbonTracker product (version 2019b). As a result of this setup, the true CO_2 fluxes (*s*) are known, making it easier to evaluate the accuracy of the algorithms tested here. All atmospheric transport simulations are from the Weather Research and Forecasting (WRF) Stochastic Time-Inverted Lagrangian Transport Model (STILT) modeling system (Lin et al., 2003; Nehrkorn et al., 2010). These simulations were generated as part of NOAA's CarbonTracker-Lagrange program (Hu et al., 2019; Miller et al., 2020). Note that the WRF-STILT outputs can be used to explicitly construct $\mathbf{H}$, making it straightforward to calculate the direct solution to the inverse problem in the 6 week case study. By contrast, many inverse modeling studies use the adjoint of an Eulerian model. These modeling frameworks rarely produce an explicit $\mathbf{H}$ but instead output the product of $\mathbf{H}$ or $\mathbf{H}^\top$ and a vector. Though we use WRF-STILT for the case studies presented here, the genHyBR algorithms could also be paired with the the adjoint of an Eulerian model.

The goal is to estimate CO_2 fluxes at a 3-hourly temporal resolution and a $1° \times 1°$ latitude-longitude spatial resolution. Using this modeling framework, synthetic observations were obtained as in Eq. (1) by adding white Gaussian noise (representing measurement and model errors) to the output from the atmospheric transport model. Note that WRF-STILT footprints from CarbonTracker-Lagrange are available at two second intervals along the OCO-2 flight track, meaning that there are fewer footprints available than there are observations in the original OCO-2 data files. Indeed, many inverse modeling studies to date

have thinned or averaged the OCO-2 observations before assimilating them within an inverse model (e.g., Peiro et al., 2022).

We generate the synthetic observations by multiplying CO_2 fluxes from CarbonTracker by these footprints, described in greater detail in Miller et al. (2020) and Liu et al. (2020). In total, there are $m = 1.92 \times 10^4$ synthetic observations and $n = 1.06 \times 10^6$ unknown CO_2 fluxes in the 6 week case study. By contrast, for the much larger one year case study, there are $m = 9.9 \times 10^4$ synthetic observations and $n = 9.4 \times 10^6$ unknown CO_2 fluxes to be estimated.

The noise covariance matrix is structured as $\mathbf{R} = \sigma^2 \mathbf{I}$ where σ^2 represents the noise variance. In this case, the discrepancy principle formula simplifies to

$$\mathcal{D}_{\text{full}}(\lambda) = \|\mathbf{H}\boldsymbol{s}_\lambda - \boldsymbol{z}\|_2^2 \leq \tau m \sigma^2.$$

Notice that the DP approach requires *a priori* knowledge or an estimate of the noise variance σ^2. In Miller et al. (2020), $\sigma = 2$ was used, which leads to a relatively large amount of noise in the observations. We test different values of σ, corresponding to different noise levels (referred to as nlevel), as shown in Table 1. More specifically, let $\boldsymbol{n}$ be a realization of the Gaussian process $\sim \mathcal{N}(\mathbf{0}, \mathbf{I})$; then the amount of noise added in Eq. (1) is $\boldsymbol{\epsilon} = \sigma \boldsymbol{n}$ where $\sigma = \texttt{nlevel} \cdot \frac{\|\mathbf{H}\boldsymbol{s}\|_2}{\|\boldsymbol{n}\|_2}$. We note that some of the considered noise levels, although very high compared to examples in the inverse problems literature, are lower than typically observed in practice for atmospheric inverse problems. However, recent studies Miller et al. (2018); O'Dell et al. (2018); Crowell et al. (2019); and Miller and Michalak (2020) show that errors in OCO-2 observations have been gradually decreasing with regular improvements in the satellite retrieval algorithms and bias corrections. Some of the values in Table 1 are low, even considering these recent improvements. With that said, these values are aspirational and may become more realistic in the future as observational and atmospheric modeling errors decline. Furthermore, they provide an opportunity to explore the behavior of the proposed inverse modeling algorithms at many different error levels.

Next, we describe the prior used in both case studies. The prior flux estimate is set to a constant value for the case studies explored here. As a result of this setup, the prior flux estimate does not contain any spatiotemporal patterns, and the patterns in the posterior fluxes solely reflect the information content of the atmospheric observations. For the cases with a fixed mean (genHyBRs), we set $\boldsymbol{\mu} = \mathbf{0}$, as has been done in several existing studies on inverse modeling algorithms (Rodgers, 2000; Chung and Saibaba, 2017). For the unknown mean case (genHyBRmean), the columns of $\mathbf{X}$ contain ones and zeros, denoting fluxes in a given time period. Specifically, for the 6 week case study, $\mathbf{X}$ has eight columns, and each column corresponds to a different 3-hourly time period of the day. A given column of $\mathbf{X}$ contains values of one for all flux elements that correspond to a given 3-hourly time of day and zero for all other elements. CO_2 fluxes have a large diurnal cycle, and this setup accounts for

nlevel	σ (μ mol $m^{-2}s^{-1}$)
5%	0.0565
10%	0.1134
50%	0.5648

Table 1. Noise level and corresponding noise standard deviation σ used in the 6-weeks case study experiments.
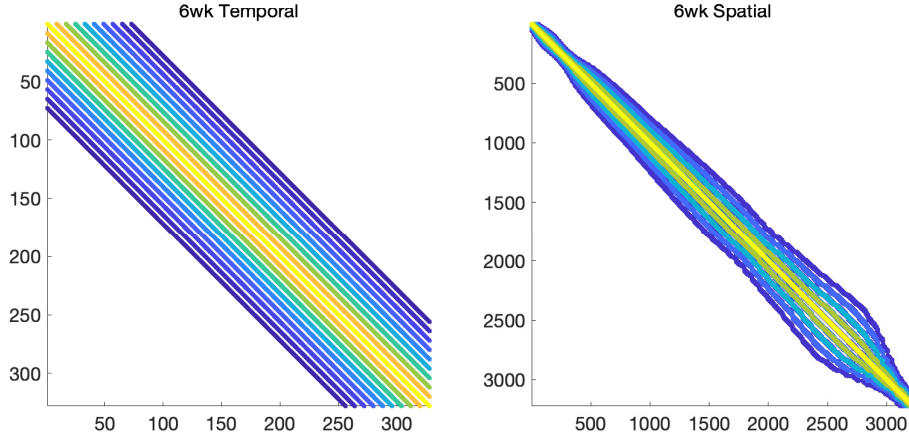
Figure 2. Sparsity pattern of the prior covariance matrices $\mathbf{Q}_t$ and $\mathbf{Q}_g$ for the 6 weeks case study.

the fact that CO_2 fluxes at different times of day will have a different mean. In the 1 year case study, $\mathbf{X}$ has 12 columns, and each column corresponds to fluxes in a different month of the year, a setup identical to that used in Miller et al. (2020).

For the prior covariance matrix of unknown fluxes, $\mathbf{Q}_s = \mathbf{Q}_t \otimes \mathbf{Q}_g$ where $\mathbf{Q}_t$ represents the temporal covariance and $\mathbf{Q}_g$ represents the spatial covariance in the fluxes. The symbol $\otimes$ denotes the Kronecker product. We use a spherical covariance model for the spatial and temporal covariance. A spherical model is ideal because it decays to zero at the correlation length or time, and the resulting matrices are usually sparse:

$$k_t(d_t;\theta_t) = \begin{cases} 1 - \frac{3}{2}\left(\frac{d_t}{\theta_t}\right) + \frac{1}{2}\left(\frac{d_t}{\theta_t}\right)^3 & \text{if } d_t \leq \theta_t, \\ 0 & \text{if } d_t > \theta_t, \end{cases} \tag{21}$$

$$k_g(d_g;\theta_g) = \begin{cases} 1 - \frac{3}{2}\left(\frac{d_g}{\theta_g}\right) + \frac{1}{2}\left(\frac{d_g}{\theta_g}\right)^3 & \text{if } d_g \leq \theta_g, \\ 0 & \text{if } d_g > \theta_g, \end{cases} \tag{22}$$

where d_t is the temporal difference, d_g is the spherical distance, and θ_t, θ_g represent the decorrelation time which has units of days and decorrelation length which has units of km, respectively. For the 6 week case study, we set $\theta_t = 9.854$ days and $\theta_g = 555.42$ km, as in Miller et al. (2020). The sparsity patterns of these covariance matrices are provided in Figure 2. We further set the diagonal elements of $\mathbf{Q}_s$ equal to one, and the regularization parameter in the inverse model (λ) will ultimately scale $\mathbf{Q}_s$ (e.g., Eq. 2). For the one year study, we use slightly different parameters, as listed in full detail in Miller et al. (2020). Notably, the variance is different in each month to better capture the impact of seasonal changes on the variability of CO_2 fluxes. Note that for the fixed mean case, the covariance matrices are sparse and can be efficiently represented in factored form. However, the hybrid approaches proposed here can handle much more complicated cases (see e.g., Chung and Saibaba, 2017; Chung et al., 2018).

For the unknown mean case, the covariance matrix $\mathbf{Q}_\beta$ is set to be the identity matrix, and $\alpha = 10$ is used in the numerical experiments. We experimented with various choices for α and observed consistently good results with $\alpha = 10$. In all of the numerical experiments, we use the DP approach to select the regularization parameter within genHyBR methods with $\tau = 1$.

In subsequent discussion of the case study results, we provide relative reconstruction error norms computed as $\|\boldsymbol{s}_{k,\lambda} - \boldsymbol{s}\|_2 / \|\boldsymbol{s}\|_2$, where $\boldsymbol{s}$ denotes the true fluxes and $\boldsymbol{s}_{k,\lambda}$ contains the reconstructed spatiotemporal fluxes at the kth iteration.

4.2 Results of the case studies

Both the `genHyBRs` and `genHyBRmean` methods converge quickly and yield accurate estimates of the CO_2 fluxes relative to other inverse modeling methods. For the 6 weeks case study, Fig. 3 shows the relative reconstruction error norms for both `genHyBRs` and `genHyBRmean` for three different noise levels and for different options for selecting the regularization parameter. DP corresponds to the discrepancy principle and is an automatic approach that depends on the data and noise level. The optimal regularization parameter, which corresponds to selecting the regularization parameter at each iteration that minimizes the reconstruction error, is provided for comparison, although it is not obtainable in practice. All of the plots with "none" correspond to $\lambda = 0$ and show *semiconvergent behavior*, which is revealed in the "U"-shape of the relative error plot. That is, the relative reconstruction error norms decrease in early iterations, but increase in later iterations due to noise contamination in the reconstructions.

We observe that for all noise levels, genHyBR methods with regularization parameter estimation (`genHyBRs-opt` and `genHyBRs-dp`) result in reconstruction error norms that decrease and flatten, thereby overcoming the semiconvergent behavior of genHyBR methods with no regularization (`genHyBRs-none`). For reference, we mark with a horizontal line the relative reconstruction error norm for a direct reconstruction. Recall that the direct method and L-BFGS typically require λ to be fixed in advance, and a poor choice of λ can yield poor reconstructions of the CO_2 fluxes. Using the optimal regularization parameter computed from `genHyBRmean` (these values are provided in Table 2), we show that a good reconstruction can be obtained with the direct method if a good regularization parameter is available, but obtaining this result may require careful and expensive tuning.

We also find that the algorithm that simultaneously estimates the mean (`genHyBRmean`) yields lower errors than the algorithm with a fixed mean (`genHyBRs`). Notably, the difference in performance between these two algorithms grows as the noise level increases. In other words, the comparative advantage of `genHyBRmean` is even larger at higher noise levels. This result implies that mean estimation becomes critical for problems with large noise levels.

For comparison, we show the convergence behavior of the Limited Memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm that is commonly used in AIM (Fig. 3). One line (with diamond symbols) shows results using L-BFGS to minimize the inverse model with an unknown mean (with $\lambda = 1$), and another line (shown with plus symbols) shows a variant of L-BFGS with a data transformation to speed convergence (also with $\lambda = 1$). The specific approach used here is described in detail in Miller et al. (2020). Both algorithms converge more slowly than the genHyBR algorithms. This result is significant because the main computational cost per iteration (i.e., one matrix-vector multiplication with $\mathbf{H}$ and its adjoint) is the same among the algorithms in Fig. 3. In addition, the relative errors increase for L-BFGS at later iterations, though the inverse modeling

optimization function continues to decrease at these iterations; the inverse model minimizes the errors with respect to the observations (z) but is not guaranteed to minimize errors with respect to the true fluxes (s).

The estimated fluxes using genHyBR also exhibit spatial patterns that mirror fluxes estimated using a direct (e.g., analytical) approach. Maps of CO_2 fluxes, averaged over 6 weeks and corresponding to 50% noise level, are shown in Fig. 4. We provide the true average flux, the `genHyBRs` and `genHyBRmean` reconstructions for various parameter choices, and the direct method reconstruction using the optimal regularization parameter from `genHyBRmean`. Notice that these relative reconstruction error values are different than those provided in Fig. 3 because they represent error norms computed on the average image rather than on the native 3-hour resolution of the estimated fluxes. Also, to better highlight broad spatial patterns, the colormap has been constrained so that average flux estimates above 2 are set to 2 and estimates below -5 are set to -5.

Perhaps surprisingly, the genHyBR algorithms require less computing time than other algorithms tested, including the direct or analytical method. Table 2 displays the measured turnaround time for each case, along with the number of iterations and
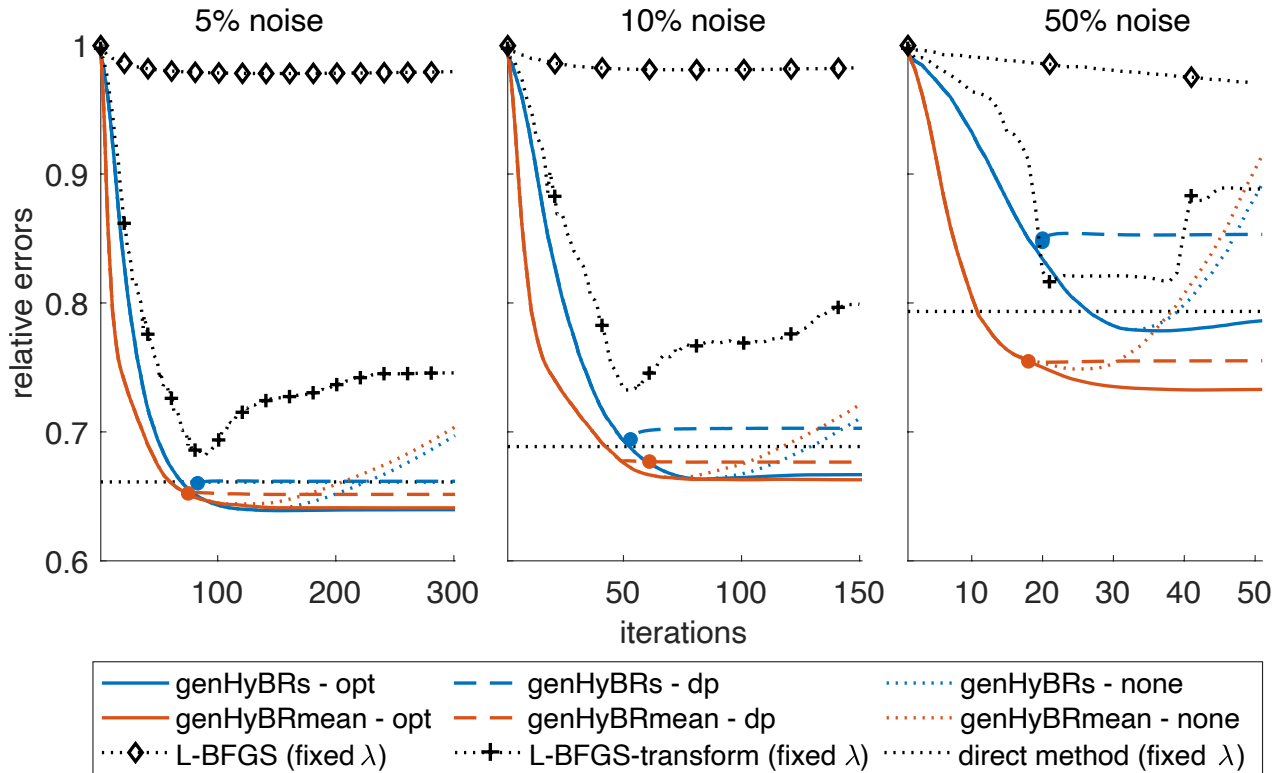
Figure 3. 6 weeks case study: Relative reconstruction error norms per iteration of `genHyBRs` and `genHyBRmean` with 5%, 10%, and 50% noise levels. We compare results for the optimal regularization parameter, the automatically selected DP parameter, $\lambda = 0$, and for the L-BFGS with a fixed regularization parameter ($\lambda = 1$). The filled circle indicates the stopping iteration for the genHyBR methods with DP. In addition, the horizontal line denotes to the relative error for the reconstruction obtained using the direct method.
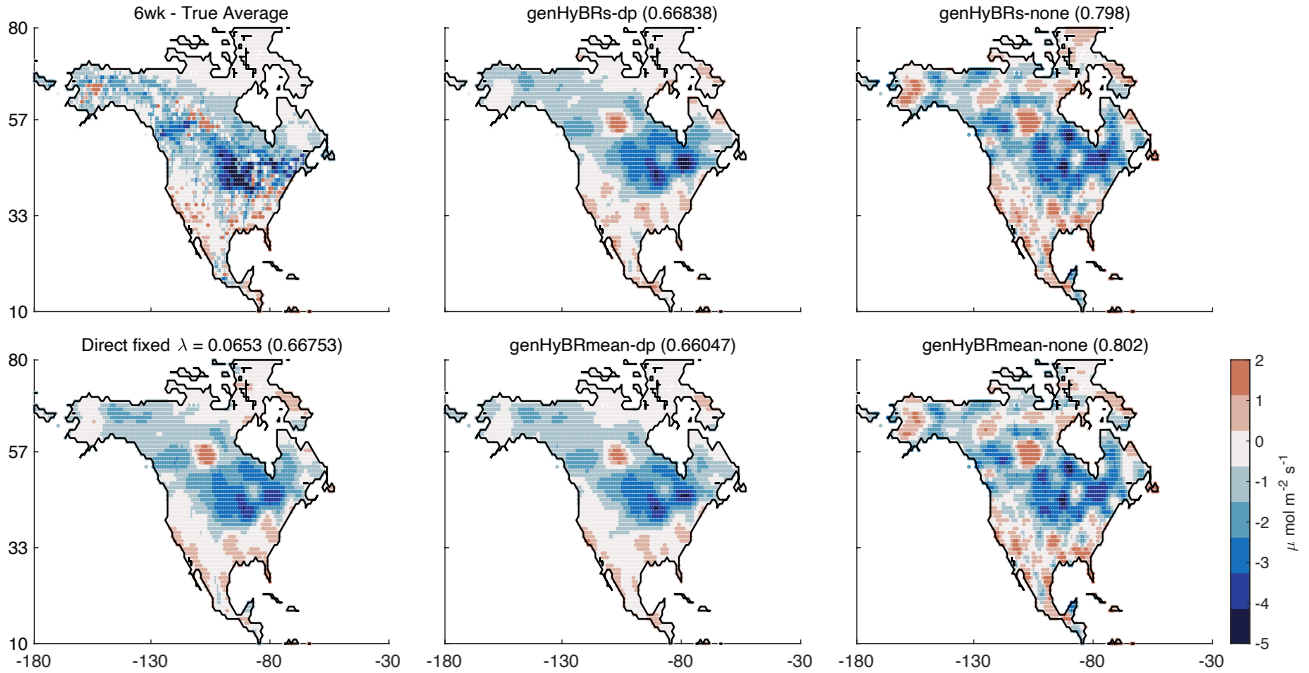
Figure 4. 6 weeks case study: Reconstructed fluxes, averaged over 6 weeks, are provided for `genHyBRs` and `genHyBRmean` for the automatically selected DP parameter and for $\lambda = 0$. The true average fluxes and the reconstruction using a direct method with the optimal regularization parameter computed from `genHyBRmean` are provided for comparison. These results correspond to 50% noise level and relative error norms of average fluxes over 6 weeks are provided in the titles.

the relative reconstruction error norms for each method. Compared to the direct method with fixed λ, hybrid methods require less time to compute the estimated CO_2 fluxes. Moreover, since the regularization parameter can be selected automatically,
450 genHyBR methods can obtain results with smaller reconstruction errors.

Finally, we demonstrate the ability to perform UQ for AIM. In the last columns of Table 2, we provide the times needed to compute uncertainties. We remark that the additional time and the difference in the number of iterations can be attributed to the need to perform reorthogonalization of the Krylov basis vectors, which is not as critical for obtaining the MAP estimate. Furthermore, uncertainty estimation for genHyBR does not require any additional forward or adjoint model runs beyond what
455 is required to estimate the fluxes (s). Hence, the additional computing time is modest, given the size of the problem and the ability to obtain solution variance estimates. The estimated posterior variance is also similar across both methods (`genHyBRs` and `genHyBRmean`).

GenHyBR also requires relatively few iterations (i.e., Krylov basis vectors) to converge on a reasonable uncertainty estimate. Figure 5 shows the posterior uncertainty in the total Continental US CO_2 flux by iteration for each algorithm. This figure
460 compares the direct approach to uncertainty estimation, GenHyBR, and the low rank approach described in Saibaba and Kitanidis (2015) and Miller et al. (2020). In this figure, we use the same regularization parameter (λ) across all algorithms to

Noise			recons			recons + uncert		
Level	Methods	Selection of λ	Iter.	Time(s)	Δs	Iter.	Time(s)	Δs
	direct	0.15	-	8,714	0.661	-	-	-
5%	genHyBRs-dp	0.24	99	**4,450**	0.662	83	10,114	0.661
	genHyBRmean-dp	0.18	102	**3,836**	0.652	75	8,025	0.651
	direct	0.14	-	8,268	0.689	-	-	-
10%	genHyBRs-dp	0.16	67	**3,125**	0.703	53	4,667	0.703
	genHyBRmean-dp	0.27	60	**2,334**	0.677	61	5,572	0.677
	direct	0.11	-	8,77	0.7934	-	-	-
50%	genHyBRs-dp	0.13	20	**922**	0.853	20	1,200	0.853
	genHyBRmean-dp	0.15	19	**773**	0.755	18	981	0.755

Table 2. 6 weeks case study: For various noise levels, we provide comparisons of `genHyBRs` and `genHyBRmean` (with DP selected regularization parameter) to standard direct and iterative methods (with fixed regularization parameter). We provide the number of iterations, the CPU timing in seconds, and the relative reconstruction error norms for the computed spatiotemporal fluxes denoted by Δs. Note that when computing the reconstructions and approximating the posterior variance, additional reorthogonalization is performed which explains the slight difference in the number of iterations and run time.

make the results from different algorithms more easily comparable to one another. The low rank approach described in Saibaba and Kitanidis (2015) and Miller et al. (2020) starts at very high values and converges more slowly than the uncertainties estimated using GenHyBR. Furthermore, the low rank approach requires new forward and adjoint runs for the uncertainty calculations, while GenHyBR uses forward and adjoint runs already generated when calculating the best estimate of the fluxes.

We finalize this section with results for the 1 year case study. Since this case study has approximately 9 times the number of unknown CO_2 fluxes and 5 times the number of observations compared to the 6 weeks case study, iterative methods are computationally more appealing than direct methods for obtaining reconstructions. The previous case study already explored the behavior of the algorithms at different noise levels, so here we only consider the 50% noise level, which corresponds to $\sigma = 0.4076$. Figure 6 provides relative reconstruction error norms for `genHyBRs` and `genHyBRmean`. With the regularization parameter automatically selected using DP, both methods result in reconstructions with relative errors smaller than 0.85. Since it is difficult to show spatiotemporal flux reconstructions over the entire year, we provide the annual average of the reconstructed CO_2 fluxes in Fig. 7. Compared to the reconstructions in Fig. 4, these average maps are much smoother. This inability to resolve fine details can be attributed to the significantly fewer observations compared to the number of unknowns in this case
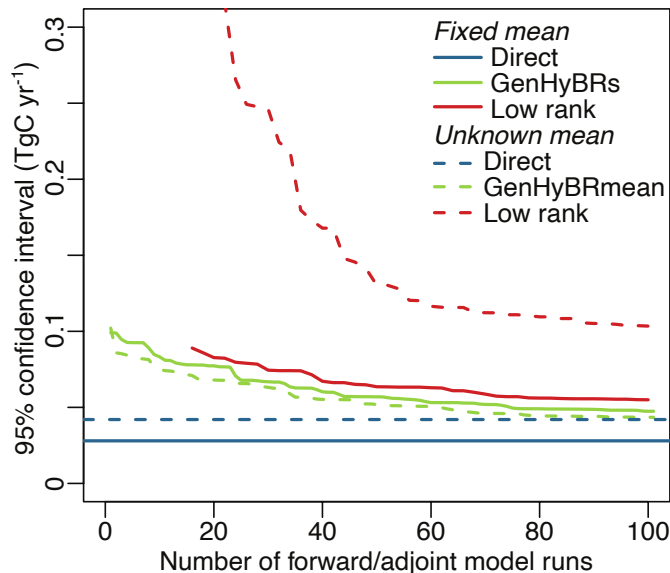
Figure 5. Estimated uncertainties for the total CO_2 flux from the Continental US over the 6 weeks case study (50% noise levels). The figure compares posterior uncertainties calculated using the direct approach, using GenHyBR, and using the low rank approach described in (Miller et al., 2020). Solid lines show uncertainties estimated for simulations with a fixed mean and dashed lines for simulations with an unknown mean. In addition, we use the same regularization parameter (λ) for the estimates shown here to make the convergence behavior of the different approaches easier to compare ($\lambda = 0.13$ for fixed mean and $\lambda = 0.15$ for unknown mean, as in Table 2). GenHyBR reaches uncertainty estimates that are close to the direct estimate more quickly than the low rank approach, particularly for the unknown mean case. In addition, GenHyBR uses forward and adjoint model runs previously generated as part of the best estimate calculations, yielding large computational savings over the low rank approach, which requires new model runs.

study. Nevertheless, these results show that the algorithms described in this study can be scaled to very large inverse problems – problems where the direct method is either computationally prohibitive or time consuming.

5 Conclusions

This article describes a mathematically advanced iterative method for AIM with large datasets. Specifically, we discuss generalized hybrid methods for inverse models with a fixed prior mean (e.g., Bayesian synthesis inverse modeling) and an unknown prior mean (e.g., geostatistical inverse modeling). We also describe a means of obtaining posterior variance estimates at very little additional computational cost. Compared to standard inverse modeling procedures (e.g., direct and iterative methods), genHyBR methods are computationally cheaper and exhibit faster convergence. One of the main advantages of genHyBR methods is the ability to efficiently and adaptively estimate the regularization parameter during the inversion process, and we described various regularization parameter estimation methods for Tikhonov regularization. Numerical experiments for case studies for 6 weeks and 1 year demonstrate that genHyBR methods provide an efficient, flexible, robust, and automatic ap-
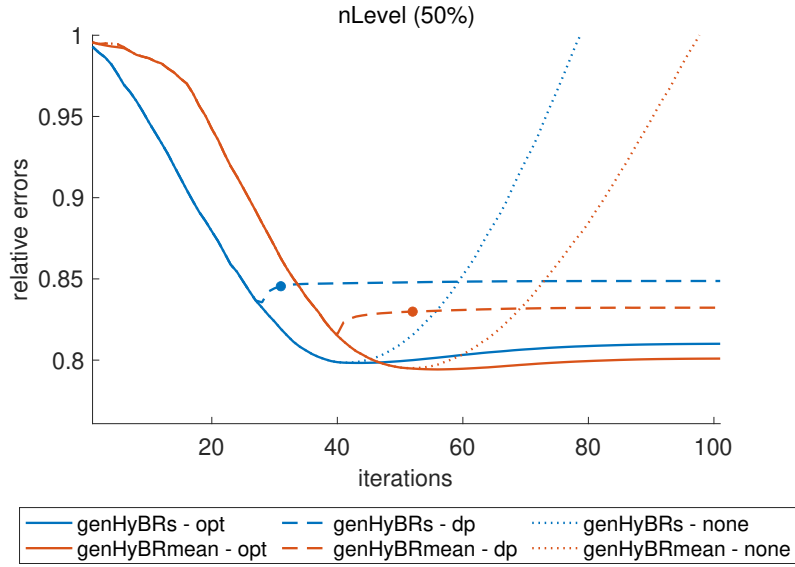
Figure 6. 1 year case study: For 50% noise level, we provide relative reconstruction error norms per iteration of `genHyBRs` and `genHyBRmean` and compare results for the optimal regularization parameter, the automatically selected DP parameter, and $\lambda = 0$.

proach for AIM with very large spatiotemporal fluxes. Furthermore, since these methods only require forward and adjoint model evaluations, these methods can be paired with different types of atmospheric transport models.

Code availability. The Matlab codes for the 6 weeks case study that were used to generate the results in Section 4 are available at https:
490 //doi.org/10.5281/zenodo.5772660.

Appendix A: Regularization parameter estimation methods for genHyBR

One of the main advantages of hybrid projection methods is the ability to adaptively and automatically estimate the regularization parameter during the iterative process. We described the discrepancy principle (DP), but other common parameter estimation techniques in the context of hybrid projection methods include the generalized cross validation (GCV) method and
495 the weighted-GCV (WGCV) method (Chung et al., 2008; Renaut et al., 2017). A summary of methods with respective functions used to compute the parameters based on the original problem and the projected problem are summarized in Table A1, where for simplicity we have assumed that $\mathbf{R} = \sigma^2 \mathbf{I}$. We have used the notation $\mathbf{B}_{k,\lambda}^{\dagger} = (\mathbf{B}_k^{\top} \mathbf{B}_k + \lambda^2 \mathbf{I})^{-1} \mathbf{B}_k^{\top}$ for given $\lambda > 0$ and $\mathbf{y}_{k,\lambda}$ is the solution to the projected problem (14).
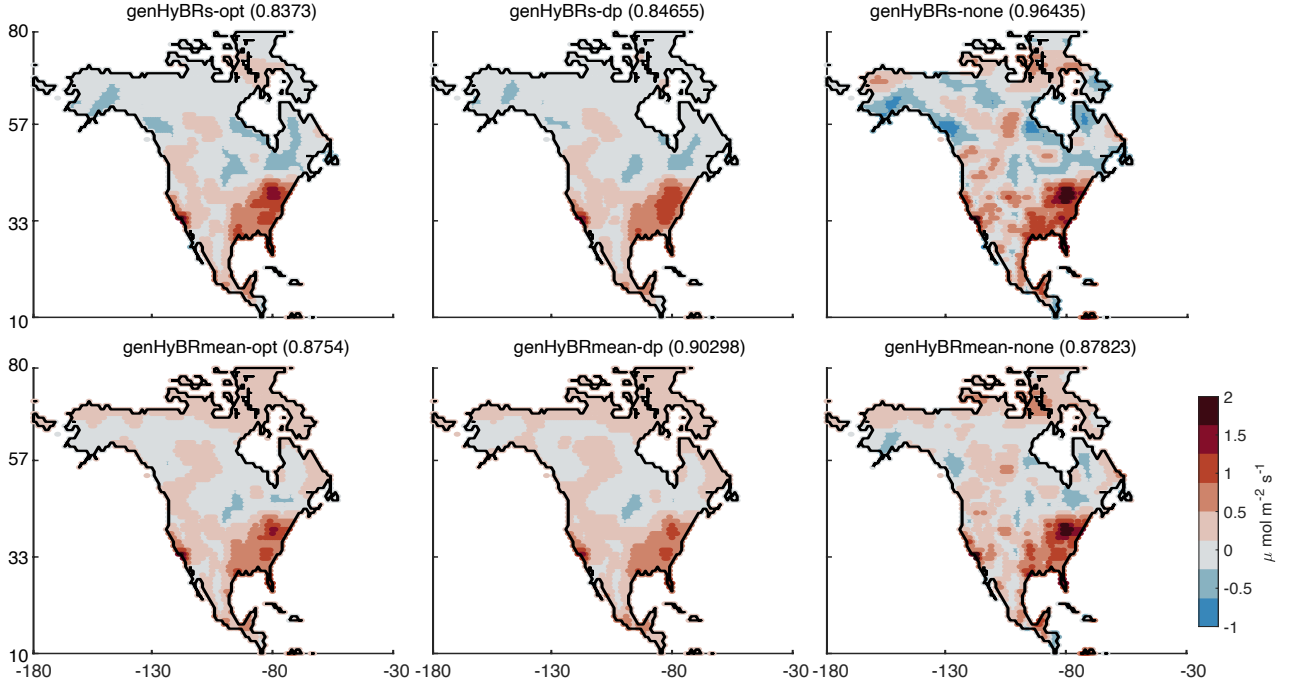
Figure 7. 1 year case study: Reconstructed fluxes, averaged over 1 year, are provided for `genHyBRs` and `genHyBRmean` for various parameter choices. These results correspond to 50% noise level and relative error norms of average fluxes over 1 year are provided in the titles.

Methods	Original problem (11)	Projected Problem (14)
DP	$\mathcal{D}_{\mathrm{full}}(\lambda) = \|\mathbf{H}\boldsymbol{s}_\lambda - \boldsymbol{z}\|^2_{\mathbf{R}^{-1}}$	$\mathcal{D}_{\mathrm{proj}}(\lambda) = \|\mathbf{B}_k\boldsymbol{y}_{k,\lambda} - \delta_1\boldsymbol{e}_1\|^2_2$
GCV	$\mathcal{G}_{\mathrm{full}}(\lambda;\omega) = \dfrac{m\|\mathbf{H}\boldsymbol{x}_\lambda - \boldsymbol{b}\|^2_{\mathbf{R}^{-1}}}{(\mathrm{tr}(\mathbf{I} - \omega\mathbf{H}\mathbf{H}^\dagger_\lambda))^2}$	$\mathcal{G}_{\mathrm{proj}}(\lambda;\omega) = \dfrac{k\|\mathbf{B}_k\boldsymbol{y}_{k,\lambda} - \delta_1\boldsymbol{e}_1\|^2_2}{(\mathrm{tr}(\mathbf{I} - \omega\mathbf{B}_k\mathbf{B}^\dagger_{k,\lambda}))^2}$

Table A1. Regularization parameter selection methods for use within genHyBR methods.

More specifically, DP selects the largest parameter value λ for which $\mathcal{D}_{\mathrm{full}}(\lambda) \leq \tau m\sigma^2$, where $\tau \geq 1$ is a user-defined parameter. Note that $m\sigma^2$ is the expected value of $\|\boldsymbol{\epsilon}\|^2_{\mathbf{R}^{-1}}$. For the projected problem, we choose the largest λ such that $\mathcal{D}_{\mathrm{proj}}(\lambda) \leq \tau m\sigma^2$. The WGCV method selects λ by minimizing the objective function $\mathcal{G}_{\mathrm{full}}(\lambda;\omega)$. Note that if $\omega = 1$ then WGCV becomes GCV. In the projected problem, we minimize $\mathcal{G}_{\mathrm{proj}}(\lambda;\omega)$ at each iteration. The parameter ω can be chosen automatically, as described in (Chung et al., 2008; Renaut et al., 2017). Note that the DP approach requires a priori knowledge of the noise variance σ^2, whereas the GCV approaches do not require prior knowledge about the noise level.

Appendix B: Extension to unknown mean: Hierarchical Bayes

In this section, we provide details for the derivation of genHyBR methods for AIM with unknown mean. We begin in Appendix B1 with the problem reformulation to simultaneously estimate the unknown fluxes and the covariate parameters. Then in B2 we provide the details of the genHyBR approach for the unknown mean case, which closely follows the derivation in Section 3.1.1. Finally, in Appendix B3 we show how to approximate the posterior variance using the generalized Golub-Kahan bidiagonalization.

B1 Reformulation for simultaneous estimation

In order to apply genHyBR methods to the unknown mean estimation problem, we first reformulation the MAP estimate from (7) to (18) as follows. For the data fit term, consider

$$
\begin{aligned}
\frac{1}{2}\|\mathbf{H}s - z\|^2_{\mathbf{R}^{-1}} &= \frac{1}{2}\left\|\begin{bmatrix}\mathbf{H} & \mathbf{0}\end{bmatrix}\begin{bmatrix}s \\ \beta\end{bmatrix} - z\right\|^2_{\mathbf{R}^{-1}} \\
&= \frac{1}{2}\left\|\mathbf{K}\begin{bmatrix}\widetilde{s} + \mathbf{X}\mu_\beta \\ \widetilde{\beta} + \mu_\beta\end{bmatrix} - z\right\|^2_{\mathbf{R}^{-1}} \\
&= \frac{1}{2}\left\|\mathbf{K}\begin{bmatrix}\widetilde{s} \\ \widetilde{\beta}\end{bmatrix} + \mathbf{H}\mathbf{X}\mu_\beta - z\right\|^2_{\mathbf{R}^{-1}} \\
&= \frac{1}{2}\left\|\mathbf{K}\begin{bmatrix}\widetilde{s} \\ \widetilde{\beta}\end{bmatrix} - \widetilde{z}\right\|^2_{\mathbf{R}^{-1}}
\end{aligned}
$$

and for the regularization terms in (7), and we have

$$
\begin{aligned}
\frac{\lambda^2}{2}\|s - \mathbf{X}\beta\|^2_{\mathbf{Q}_s^{-1}} + \frac{\lambda_\beta^2}{2}\|\beta - \mu_\beta\|^2_{\mathbf{Q}_\beta^{-1}} &= \frac{\lambda^2}{2}\|\widetilde{s} - \mathbf{X}\widetilde{\beta}\|^2_{\mathbf{Q}_s^{-1}} + \frac{\lambda_\beta^2}{2}\|\widetilde{\beta}\|^2_{\mathbf{Q}_\beta^{-1}} \\
&= \frac{1}{2}(\lambda^2\widetilde{s}^\top\mathbf{Q}_s^{-1}\widetilde{s} - 2\lambda^2\widetilde{s}^\top\mathbf{Q}_s^{-1}\mathbf{X}\widetilde{\beta} + \widetilde{\beta}^\top(\lambda^2\mathbf{X}^\top\mathbf{Q}_s^{-1}\mathbf{X} + \lambda_\beta^2\mathbf{Q}_\beta^{-1})\widetilde{\beta}) \\
&= \frac{1}{2}(\lambda^2\widetilde{s}^\top\mathbf{Q}_s^{-1}\widetilde{s} - 2\lambda^2\widetilde{s}^\top\mathbf{Q}_s^{-1}\mathbf{X}\widetilde{\beta} + \widetilde{\beta}^\top(\lambda^2\mathbf{X}^\top\mathbf{Q}_s^{-1}\mathbf{X} + (\alpha\lambda)^2\mathbf{Q}_\beta^{-1})\widetilde{\beta}) \\
&= \frac{1}{2}\begin{bmatrix}\widetilde{s}^\top & \widetilde{\beta}^\top\end{bmatrix}\begin{bmatrix}\lambda^2\mathbf{Q}_s^{-1} & -\lambda^2\mathbf{Q}_s^{-1}\mathbf{X} \\ -\lambda^2\mathbf{X}^\top\mathbf{Q}_s^{-1} & \lambda^2\mathbf{X}^\top\mathbf{Q}_s^{-1}\mathbf{X} + (\alpha\lambda)^2\mathbf{Q}_\beta^{-1}\end{bmatrix}\begin{bmatrix}\widetilde{s} \\ \widetilde{\beta}\end{bmatrix} \\
&= \frac{\lambda^2}{2}\begin{bmatrix}\widetilde{s}^\top & \widetilde{\beta}^\top\end{bmatrix}\underbrace{\begin{bmatrix}\mathbf{Q}_s^{-1} & -\mathbf{Q}_s^{-1}\mathbf{X} \\ -\mathbf{X}^\top\mathbf{Q}_s^{-1} & \mathbf{X}^\top\mathbf{Q}_s^{-1}\mathbf{X} + \alpha^2\mathbf{Q}_\beta^{-1}\end{bmatrix}}_{=\mathbf{Q}^{-1}}\begin{bmatrix}\widetilde{s} \\ \widetilde{\beta}\end{bmatrix} \\
&= \frac{\lambda^2}{2}\|\widetilde{\gamma}\|^2_{\mathbf{Q}^{-1}}.
\end{aligned}
$$

We identify $\widetilde{\gamma} = [\widetilde{s}^\top, \widetilde{\beta}^\top]^\top$ and this completes the derivation of (18).

Next we provide the derivation of the augmented prior covariance matrix (19). Since we need $\mathbf{Q}$ and not $\mathbf{Q}^{-1}$ in genHyBR, we use the formula for the inverse of a 2×2 block matrix

520
$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} + \mathbf{D}^{-1} \end{bmatrix}$$

which is defined if $\mathbf{D}$ and $\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ are invertible. Using the above formula, we get

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_s^{-1} & -\mathbf{Q}_s^{-1}\mathbf{X} \\ -\mathbf{X}^\top\mathbf{Q}_s^{-1} & \mathbf{X}^\top\mathbf{Q}_s^{-1}\mathbf{X} + \alpha^2\mathbf{Q}_\beta^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{Q}_s + \frac{1}{\alpha^2}\mathbf{X}\mathbf{Q}_\beta\mathbf{X}^\top & \frac{1}{\alpha^2}\mathbf{X}\mathbf{Q}_\beta \\ \frac{1}{\alpha^2}\mathbf{Q}_\beta\mathbf{X}^\top & \frac{1}{\alpha^2}\mathbf{Q}_\beta \end{bmatrix}.$$

This simplifies to

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \frac{1}{\alpha^2}\begin{bmatrix} \mathbf{X} \\ \mathbf{I} \end{bmatrix}\mathbf{Q}_\beta\begin{bmatrix} \mathbf{X}^\top & \mathbf{I} \end{bmatrix},$$

525   which completes the derivation of (19).

B2   genHyBR approach for AIM with unknown mean

In this subsection, we derive the genHyBR approach for the unknown mean case. We initialize $\delta_1^K = \|\widetilde{z}\|_{\mathbf{R}^{-1}}$, $\boldsymbol{u}_1^K = \widetilde{z}/\delta_1^K$ and $\gamma_1^K\boldsymbol{v}_1^K = \mathbf{K}^\top\mathbf{R}^{-1}\boldsymbol{u}_1^K$, then the kth iteration of the generalized Golub-Kahan bidiagonalization procedure generates vectors $\boldsymbol{u}_{k+1}^K$ and $\boldsymbol{v}_{k+1}^K$ such that

530   $$\gamma_{k+1}^K\boldsymbol{u}_{k+1}^K = \mathbf{K}\mathbf{Q}\boldsymbol{v}_k^K - \gamma_k^K\boldsymbol{u}_k^K$$
$$\delta_{k+1}^K\boldsymbol{v}_{k+1}^K = \mathbf{K}^\top\mathbf{R}^{-1}\boldsymbol{u}_{k+1}^K - \delta_{k+1}^K\boldsymbol{v}_k^K$$

where scalars $\gamma_k^K, \delta_k^K \geq 0$ are computed such that $\|\boldsymbol{u}_k^K\|_{\mathbf{R}^{-1}} = \|\boldsymbol{v}_k^K\|_{\mathbf{Q}} = 1$. At the end of k iterations, we have

$$\mathbf{B}_k^K \equiv \begin{bmatrix} \gamma_1^K & & & & \\ \delta_2^K & \gamma_2^K & & & \\ & \delta_3^K & \ddots & & \\ & & \ddots & \gamma_k^K & \\ & & & \delta_{k+1}^K \end{bmatrix}, \quad \mathbf{U}_{k+1}^K \equiv \begin{bmatrix} \boldsymbol{u}_1^K, \dots, \boldsymbol{u}_{k+1}^K \end{bmatrix}, \quad \text{and} \quad \mathbf{V}_k^K \equiv \begin{bmatrix} \boldsymbol{v}_1^K, \dots, \boldsymbol{v}_k^K \end{bmatrix}.$$

The above matrices satisfy the following relations

$$\mathbf{U}_{k+1}^K\delta_1^K\boldsymbol{e}_1 = \widetilde{z}$$
535   $$\mathbf{K}\mathbf{Q}\mathbf{V}_k^K = \mathbf{U}_{k+1}^K\mathbf{B}_k^K \tag{B1}$$
$$\mathbf{K}^\top\mathbf{R}^{-1}\mathbf{U}_{k+1}^K = \mathbf{V}_k^K(\mathbf{B}_k^K)^\top + \gamma_{k+1}^K\boldsymbol{v}_{k+1}^K\boldsymbol{e}_{k+1}^\top.$$

Also, matrices $\mathbf{U}_{k+1}^K$ and $\mathbf{V}_k^K$ satisfy the following orthogonality conditions:

$$(\mathbf{U}_{k+1}^K)^\top \mathbf{R}^{-1} \mathbf{U}_{k+1}^K = \mathbf{I}_{k+1} \quad \text{and} \quad (\mathbf{V}_k^K)^\top \mathbf{Q} \mathbf{V}_k^K = \mathbf{I}_k. \tag{B2}$$

The columns of the matrix $\mathbf{V}_k^K$ form a basis for the Krylov subspace $\mathcal{K}_k(\mathbf{K}^\top \mathbf{R}^{-1} \mathbf{K} \mathbf{Q}, \mathbf{K}^\top \mathbf{R}^{-1} \widetilde{z})$, which we use to search for approximate solutions.

540 To obtain the approximate solution, we solve the least-squares problem

$$\min_{\mathbf{y} \in \mathbb{R}^k} \|\mathbf{B}_k^K \mathbf{y} - \delta_1^K \boldsymbol{e}_1\|_2^2 + \lambda^2 \|\mathbf{y}\|_2^2, \tag{B3}$$

to obtain the optimizer $\mathbf{y}_{k,\lambda}^K$ and to compute the approximate solution $\boldsymbol{\gamma}_{k,\lambda} = \mathbf{Q} \mathbf{V}_k^K \mathbf{y}_{k,\lambda}^K$. We can extract the approximations $\widetilde{s}_{k,\lambda}$ and $\widetilde{\boldsymbol{\beta}}_{k,\lambda}$ as $\boldsymbol{\gamma}_{k,\lambda} = \begin{bmatrix} \widetilde{s}_{k,\lambda} \\ \widetilde{\boldsymbol{\beta}}_{k,\lambda} \end{bmatrix}$. To estimate the regularization parameter, we can adapt the techniques in Section 3.1.2; for example, using the Discrepancy principle, we pick a regularization parameter λ such that

545 $$\mathcal{D}_{\mathrm{proj}}^K(\lambda) = \|\mathbf{B}_k^K \mathbf{y}_{k,\lambda}^K - \delta_1^K \boldsymbol{e}_1\|_2^2 \leq \tau m,$$

where $\tau \geq 1$ is a user-defined parameter and m is the expected value of $\|\boldsymbol{\epsilon}\|_{\mathbf{R}^{-1}}^2$. Other parameter selection techniques such as GCV and WGCV can also be adapted to the unknown mean case with similar expressions as in Table A1 but we omit the details.

B3 Approximation to the posterior variance

550 We propose an approximation to the posterior covariance matrix $\boldsymbol{\Gamma}_{\mathrm{post}}$ corresponding to the posterior distribution $\pi(\boldsymbol{s}, \boldsymbol{\beta}|\boldsymbol{z})$ in (6). First notice that from (19) and (8), we get the following expression of the posterior covariance matrix

$$\boldsymbol{\Gamma}_{\mathrm{post}} = \left(\lambda^2 \mathbf{Q}^{-1} + \mathbf{K}^\top \mathbf{R}^{-1} \mathbf{K}\right)^{-1} = \mathbf{Q} \left(\lambda^2 \mathbf{Q} + \mathbf{Q}(\mathbf{K}^\top \mathbf{R}^{-1} \mathbf{K})\mathbf{Q}\right)^{-1} \mathbf{Q}, \tag{B4}$$

where the last expression is obtained by factoring out $\mathbf{Q}$. Assume that k iterations of the generalized Golub-Kahan bidiagonalization have been performed to solve (18). Let $(\mathbf{B}_k^K)^\top \mathbf{B}_k^K = \mathbf{W}_k^K \boldsymbol{\Theta}_k^K (\mathbf{W}_k^K)^\top$ be an eigenvalue decomposition with 555 eigenvalues $\theta_1^K, \ldots, \theta_k^K$ and let $\mathbf{Z}_k^K = \mathbf{Q} \mathbf{V}_k^K \mathbf{W}_k^K$. Then consider the following low-rank approximation,

$$\mathbf{Q}(\mathbf{K}^\top \mathbf{R}^{-1} \mathbf{K})\mathbf{Q} \approx \mathbf{Q}(\mathbf{V}_k^K (\mathbf{B}_k^K)^\top \mathbf{B}_k^K \mathbf{V}_k^\top)\mathbf{Q} = \mathbf{Z}_k^K \boldsymbol{\Theta}_k^K (\mathbf{Z}_k^K)^\top. \tag{B5}$$

Using (B5) in (B4), we can define an approximate covariance matrix as

$$\widetilde{\boldsymbol{\Gamma}}_{\mathrm{post}} = \mathbf{Q}(\lambda^2 \mathbf{Q} + \mathbf{Z}_k^K \boldsymbol{\Theta}_k^K (\mathbf{Z}_k^K)^\top)^{-1} \mathbf{Q} = \lambda^{-2} \mathbf{Q} - \mathbf{Z}_k^K \boldsymbol{\Delta}_k^K (\mathbf{Z}_k^K)^\top, \tag{B6}$$

where $\boldsymbol{\Delta}_k^K$ is diagonal matrix,

560 $$\boldsymbol{\Delta}_k^K \equiv \lambda^{-2} \begin{bmatrix} \frac{\theta_1^K}{\theta_1^K + \lambda^2} & & \\ & \ddots & \\ & & \frac{\theta_k^K}{\theta_k^K + \lambda^2} \end{bmatrix} \in \mathbb{R}^{k \times k}.$$

The last expression was obtained using the Woodbury formula.

Therefore, we have an efficient representation of the approximate posterior covariance matrix as a low-rank perturbation of the prior covariance matrix $\mathbf{Q}$. It is important to note that as with the prior covariance matrix $\mathbf{Q}$, we do not need to store $\widetilde{\boldsymbol{\Gamma}}_{\text{post}}$ explicitly. More precisely, in addition to storing the information required for storing $\mathbf{Q}$, we only need to store $nk+k$ additional entries corresponding to the matrices $\mathbf{Z}_k^K$ and $\boldsymbol{\Delta}_k^K$. Furthermore, the error in the low-rank approximation can be analyzed using similar techniques as in (Saibaba et al., 2020). Similar to the approach described in Section 3.2, the posterior variance, which corresponds to the diagonal entries of $\boldsymbol{\Gamma}_{\text{post}}$, can be approximated using the diagonal entries of (B6). First, note the diagonals of $\mathbf{Q}$ are obtained from the diagonals of the block matrices $\mathbf{Q}_s + \alpha^{-2}\mathbf{X}\mathbf{Q}_\beta\mathbf{X}^\top$ and $\alpha^{-2}\mathbf{Q}_\beta$. The diagonals of $\mathbf{Q}_s$ and $\mathbf{Q}_\beta$ are typically known analytically. The diagonals of $\mathbf{X}\mathbf{Q}_s\mathbf{X}^\top$ and $\mathbf{Z}_k^K\boldsymbol{\Delta}_k^K(\mathbf{Z}_k^K)^\top$ are easy to compute in $\mathcal{O}((n+p)k^2)$ flops since they are low-rank matrices. Therefore, given the approximate representation of the covariance matrix (B6), we can estimate the posterior variance (i.e., the diagonals of the posterior covariance). A complete description of the method is given in Algorithm 3.

Algorithm 3 AIM with unknown mean—`genHyBRmean` with UQ

Require: Matrices $\mathbf{K}$, $\mathbf{R}$ and $\mathbf{Q}$, and vector $\widetilde{z}$.

1: $\{/*$ Compute MAP estimate $*/\}$

2: initialize $\boldsymbol{u}_1^K = \widetilde{z}/\|\widetilde{z}\|_{\mathbf{R}^{-1}}$

3: **for** $j = 1, \ldots, k$ **do**

4:    one iteration of generalized Golub-Kahan bidiagonalization to obtain $\mathbf{B}_j^K$, $\mathbf{U}_{j+1}^K$, and $\mathbf{V}_j^K$

5:    estimate regularization parameter λ and compute $\boldsymbol{y}_{j,\lambda}^K$ by solving (B3)

6: **end for**

7: compute $\begin{bmatrix} \widetilde{\boldsymbol{s}}_{k,\lambda} \\ \widetilde{\boldsymbol{\beta}}_{k,\lambda} \end{bmatrix} = \mathbf{Q}\mathbf{V}_k^K\boldsymbol{y}_{k,\lambda}^K$ and $\boldsymbol{\gamma}_{k,\lambda} = \begin{bmatrix} \widetilde{\boldsymbol{s}}_{k,\lambda} + \mathbf{X}\boldsymbol{\mu}_\beta \\ \widetilde{\boldsymbol{\beta}}_{k,\lambda} + \boldsymbol{\mu}_\beta \end{bmatrix}$

8: $\{/*$ Compute the approximation to the posterior variance $*/\}$

9: compute the eigendecomposition $(\mathbf{B}_k^K)^\top\mathbf{B}_k^K = \mathbf{W}_k^K\boldsymbol{\Theta}_k^K(\mathbf{W}_k^K)^\top$

10: compute $\mathbf{Z}_k^K = \mathbf{Q}\mathbf{V}_k^K\mathbf{W}_k^K$ and $\boldsymbol{\Delta}_k^K = \text{diag}(\frac{\theta_1^K}{\theta_1^K+\lambda^2}, \ldots, \frac{\theta_k^K}{\theta_k^K+\lambda^2})$

11: compute $\boldsymbol{d}_{\text{LR}} = \text{LowRankDiag}(\mathbf{Z}_k^K\boldsymbol{\Delta}_k^K, \mathbf{Z}_k^K)$ using Algorithm 2

12: compute $[\boldsymbol{d}_\beta] = \text{LowRankDiag}(\mathbf{X}\mathbf{Q}_\beta, \mathbf{X})$, $\boldsymbol{d}_{\mathbf{Q}} = [\text{diag}(\mathbf{Q}_s) + \alpha^{-2}\boldsymbol{d}_\beta; \alpha^{-2}\text{diag}(\mathbf{Q}_\beta)]$

13: estimate posterior variance $\boldsymbol{d}_{k,\lambda} = \lambda^{-2}\boldsymbol{d}_{\mathbf{Q}} - \boldsymbol{d}_{\text{LR}}$

14: **return** MAP estimate $\boldsymbol{\gamma}_{k,\lambda}$ and variance estimate $\boldsymbol{d}_{k,\lambda}$

Author contributions. JC, SMM and AKS designed the study. TC conducted the numerical simulations. All authors contributed to the writing and presentation of results.

575   *Competing interests.* The authors declare that they have no conflict of interest.

Acknowledgements. This work was partially supported by the National Science Foundation ATD program under grants DMS-2026841, 2026830, and 2026835. This material was also based upon work partially supported by the National Science Foundation under Grant DMS-1638521 to the Statistical and Applied Mathematical Sciences Institute. The OCO-2 CarbonTracker-Lagrange footprint library was produced by NOAA/GML and AER Inc with support from NASA Carbon Monitoring System project "Andrews (CMS 2014) Regional Inverse Mod-
580   eling in North and South America for the NASA Carbon Monitoring System." We especially thank Arlyn Andrews, Michael Trudeau, and Marikate Mountain for generating and assisting with the footprint library. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Baker, D. F., Doney, S. C., and Schimel, D. S.: Variational data assimilation for atmospheric CO_2, Tellus B: Chemical and Physical Meteorology, 58, 359–365, https://doi.org/10.1111/j.1600-0889.2006.00218.x, 2006.

Bardsley, J.: Computational Uncertainty Quantification for Inverse Problems, Computer Science and Engineering, SIAM, https://books.google.com/books?id=mfV1DwAAQBAJ, 2018.

Barlow, J. L.: Reorthogonalization for the Golub–Kahan–Lanczos bidiagonal reduction, Numerische Mathematik, 124, 237–278, 2013.

Benning, M. and Burger, M.: Modern regularization methods for inverse problems, Acta Numerica, 27, 1–111, 2018.

Björck, Å.: Numerical Methods for Least Squares Problems, SIAM, 1996.

Bousserez, N. and Henze, D. K.: Optimal and scalable methods to approximate the solutions of large-scale Bayesian problems: theory and application to atmospheric inversion and data assimilation, Quarterly Journal of the Royal Meteorological Society, 144, 365–390, https://doi.org/10.1002/qj.3209, 2018.

Brasseur, G. and Jacob, D.: Modeling of Atmospheric Chemistry, Cambridge University Press, https://books.google.com/books?id=UNzFDgAAQBAJ, 2017.

Buis, A.: GeoCarb: A New View of Carbon Over the Americas, ExploreEarth, https://www.nasa.gov/feature/jpl/geocarb-a-new-view-of-carbon-over-the-americas, 2018.

Chatterjee, A. and Michalak, A. M.: Technical Note: Comparison of ensemble Kalman filter and variational approaches for CO_2 data assimilation, Atmospheric Chemistry and Physics, 13, 11 643–11 660, https://doi.org/10.5194/acp-13-11643-2013, 2013.

Chatterjee, A., Michalak, A. M., Anderson, J. L., Mueller, K. L., and Yadav, V.: Toward reliable ensemble Kalman filter estimates of CO_2 fluxes, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/https://doi.org/10.1029/2012JD018176, 2012.

Chen, Z., Liu, J., Henze, D. K., Huntzinger, D. N., Wells, K. C., Sitch, S., Friedlingstein, P., Joetzjer, E., Bastrikov, V., Goll, D. S., Haverd, V., Jain, A. K., Kato, E., Lienert, S., Lombardozzi, D. L., McGuire, P. C., Melton, J. R., Nabel, J. E. M. S., Poulter, B., Tian, H., Wiltshire, A. J., Zaehle, S., and Miller, S. M.: Linking global terrestrial CO_2 fluxes and environmental drivers: inferences from the Orbiting Carbon Observatory 2 satellite and terrestrial biospheric models, Atmospheric Chemistry and Physics, 21, 6663–6680, https://doi.org/10.5194/acp-21-6663-2021, 2021a.

Chen, Z., Huntzinger, D. N., Liu, J., Piao, S., Wang, X., Sitch, S., Friedlingstein, P., Anthoni, P., Arneth, A., Bastrikov, V., Goll, D. S., Haverd, V., Jain, A. K., Joetzjer, E., Kato, E., Lienert, S., Lombardozzi, D. L., McGuire, P. C., Melton, J. R., Nabel, J. E. M. S., Pongratz, J., Poulter, B., Tian, H., Wiltshire, A. J., Zaehle, S., and Miller, S. M.: Five years of variability in the global carbon cycle: comparing an estimate from the Orbiting Carbon Observatory-2 and process-based models, Environmental Research Letters, 16, 054 041, https://doi.org/10.1088/1748-9326/abfac1, 2021b.

Chung, J. and Gazzola, S.: Computational methods for large-scale inverse problems: a survey on hybrid projection methods, arXiv preprint arXiv:2105.07221, 2021.

Chung, J. and Saibaba, A. K.: Generalized hybrid iterative methods for large-scale Bayesian inverse problems, SIAM Journal on Scientific Computing, 39, S24–S46, 2017.

Chung, J., Nagy, J. G., and O'Leary, D. P.: A weighted GCV method for Lanczos hybrid regularization, Electronic Transactions on Numerical Analysis, 28, 149–167, 2008.

Chung, J., Saibaba, A. K., Brown, M., and Westman, E.: Efficient generalized Golub–Kahan based methods for dynamic inverse problems, Inverse Problems, 34, 024 005, 2018.

Crisp, D.: Measuring atmospheric carbon dioxide from space with the Orbiting Carbon Observatory-2 (OCO-2), in: Earth Observing Systems XX, edited by Butler, J. J., Xiong, X. J., and Gu, X., vol. 9607, pp. 1 – 7, International Society for Optics and Photonics, SPIE, https://doi.org/10.1117/12.2187291, 2015.

Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., Liu, J., Deng, F., Feng, L., McKain, K., Chatterjee, A., Miller, J. B., Stephens, B. B., Eldering, A., Crisp, D., Schimel, D., Nassar, R., O'Dell, C. W., Oda, T., Sweeney, C., Palmer, P. I., and Jones, D. B. A.: The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network, Atmospheric Chemistry and Physics, 19, 9797–9831, https://doi.org/10.5194/acp-19-9797-2019, 2019.

Davis, K. J., Deng, A., Lauvaux, T., Miles, N. L., Richardson, S. J., Sarmiento, D. P., Gurney, K. R., Hardesty, R. M., Bonin, T. A., Brewer, W. A., Lamb, B. K., Shepson, P. B., Harvey, R. M., Cambaliza, M. O., Sweeney, C., Turnbull, J. C., Whetstone, J., and Karion, A.: The Indianapolis Flux Experiment (INFLUX): A test-bed for developing urban greenhouse gas emission measurements, Elementa: Science of the Anthropocene, 5, https://doi.org/10.1525/elementa.188, 21, 2017.

Eldering, A., Wennberg, P. O., Crisp, D., Schimel, D. S., Gunson, M. R., Chatterjee, A., Liu, J., Schwandner, F. M., Sun, Y., O'Dell, C. W., Frankenberg, C., Taylor, T., Fisher, B., Osterman, G. B., Wunch, D., Hakkarainen, J., Tamminen, J., and Weir, B.: The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes, Science, 358, https://doi.org/10.1126/science.aam5745, 2017.

Enting, I.: Inverse Problems in Atmospheric Constituent Transport, Cambridge Atmospheric and Space Science Series, Cambridge University Press, https://books.google.com/books?id=b61nA7D1XXUC, 2002.

Ganesan, A. L., Rigby, M., Zammit-Mangion, A., Manning, A. J., Prinn, R. G., Fraser, P. J., Harth, C. M., Kim, K.-R., Krummel, P. B., Li, S., Mühle, J., O'Doherty, S. J., Park, S., Salameh, P. K., Steele, L. P., and Weiss, R. F.: Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods, Atmos. Chem. Phys., 14, 3855–3864, https://doi.org/10.5194/acp-14-3855-2014, 2014.

Gazzola, S. and Sabaté Landman, M.: Krylov methods for inverse problems: Surveying classical, and introducing new, algorithmic approaches, GAMM-Mitteilungen, 43, e202000 017, 2020.

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., and Michalak, A. M.: North American CO_2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosciences, 9, 457–475, https://doi.org/10.5194/bg-9-457-2012, 2012.

Groetsch, C.: Comments on Morozov's discrepancy principle, in: Improperly posed problems and their numerical treatment, pp. 97–104, Springer, 1983.

Hansen, P. C.: Discrete Inverse Problems: Insight and Algorithms, SIAM, Philadelphia, 2010.

Hase, N., Miller, S. M., Maaß, P., Notholt, J., Palm, M., and Warneke, T.: Atmospheric inverse modeling via sparse reconstruction, Geoscientific Model Development, 10, 3695–3713, 2017.

Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem, Atmospheric Chemistry and Physics, 7, 2413–2433, https://doi.org/10.5194/acp-7-2413-2007, 2007.

Hu, L., Andrews, A. E., Thoning, K. W., Sweeney, C., Miller, J. B., Michalak, A. M., Dlugokencky, E., Tans, P. P., Shiga, Y. P., Mountain, M., Nehrkorn, T., Montzka, S. A., McKain, K., Kofler, J., Trudeau, M., Michel, S. E., Biraud, S. C., Fischer, M. L., Worthy, D. E. J., Vaughn, B. H., White, J. W. C., Yadav, V., Basu, S., and van der Velde, I. R.: Enhanced North American carbon uptake associated with El Niño, Science Advances, 5, https://doi.org/10.1126/sciadv.aaw0076, 2019.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the up-stream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/https://doi.org/10.1029/2002JD003161, 2003.

660 Liu, X., Weinbren, A. L., Chang, H., Tadić, J., Mountain, M. E., Trudeau, M. E., Andrews, A. E., Chen, Z., and Miller, S. M.: Data reduction for inverse modeling: an adaptive approach v1.0, Geoscientific Model Development Discussions, 2020, 1–21, https://doi.org/10.5194/gmd-2020-246, 2020.

Machida, T., Matsueda, H., Sawa, Y., Nakagawa, Y., Hirotani, K., Kondo, N., Goto, K., Nakazawa, T., Ishikawa, K., and Ogawa, T.: World-wide Measurements of Atmospheric CO_2 and Other Trace Gas Species Using Commercial Airlines, Journal of Atmospheric and Oceanic
665 Technology, 25, 1744 – 1754, https://doi.org/10.1175/2008JTECHA1082.1, 01 Oct. 2008.

Meirink, J. F., Bergamaschi, P., and Krol, M. C.: Four-dimensional variational data assimilation for inverse modelling of atmo-spheric methane emissions: method and comparison with synthesis inversion, Atmospheric Chemistry and Physics, 8, 6341–6353, https://doi.org/10.5194/acp-8-6341-2008, 2008.

Michalak, A. M., Bruhwiler, L., and Tans, P. P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, Journal of
670 Geophysical Research: Atmospheres, 109, https://doi.org/10.1029/2003JD004422, 2004.

Michalak, A. M., Hirsch, A., Bruhwiler, L., Gurney, K. R., Peters, W., and Tans, P. P.: Maximum likelihood estimation of covari-ance parameters for Bayesian atmospheric trace gas surface flux inversions, Journal of Geophysical Research: Atmospheres, 110, https://doi.org/10.1029/2005JD005970, 2005.

Miller, S. M. and Michalak, A. M.: The impact of improved satellite retrievals on estimates of biospheric carbon balance, Atmospheric
675 Chemistry and Physics, 20, 323–331, https://doi.org/10.5194/acp-20-323-2020, 2020.

Miller, S. M., Michalak, A. M., and Levi, P. J.: Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions, Geoscientific Model Development, 7, 303–315, https://doi.org/10.5194/gmd-7-303-2014, 2014.

Miller, S. M., Michalak, A. M., Yadav, V., and Tadić, J. M.: Characterizing biospheric carbon balance using CO_2 observations from the OCO-2 satellite, Atmospheric Chemistry and Physics, 18, 6785–6799, https://doi.org/10.5194/acp-18-6785-2018, 2018.

680 Miller, S. M., Saibaba, A. K., Trudeau, M. E., Mountain, M. E., and Andrews, A. E.: Geostatistical inverse modeling with very large datasets: an example from the Orbiting Carbon Observatory 2 (OCO-2) satellite, Geoscientific Model Development, 13, 1771–1785, https://doi.org/10.5194/gmd-13-1771-2020, 2020.

Mitchell, L. E., Crosman, E. T., Jacques, A. A., Fasoli, B., Leclair-Marzolf, L., Horel, J., Bowling, D. R., Ehleringer, J. R., and Lin, J. C.: Monitoring of greenhouse gases and pollutants across an urban area using a light-rail public transit platform, Atmospheric Environment,
685 187, 9–23, https://doi.org/https://doi.org/10.1016/j.atmosenv.2018.05.044, 2018.

Nakajima, M., Kuze, A., and Suto, H.: The current status of GOSAT and the concept of GOSAT-2, in: Sensors, Systems, and Next-Generation Satellites XVI, edited by Meynart, R., Neeck, S. P., and Shimoda, H., vol. 8533, pp. 21 – 30, International Society for Optics and Photonics, SPIE, https://doi.org/10.1117/12.974954, 2012.

Nehrkorn, T., Eluszkiewicz, J., Wofsy, S., Lin, J., Gerbig, C., Longo, M., and Freitas, S.: Coupled weather research
690 and forecasting–stochastic time-inverted Lagrangian transport (WRF–STILT) model, Meteorol Atmos Phys, 107, 51–64, https://doi.org/https://doi.org/10.1007/s00703-010-0068-x, 2010.

NOAA Global Monitoring Laboratory: Observation sites, https://gml.noaa.gov/dv/site/.

O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., Frankenberg, C., Kiel, M., Lindqvist, H., Mandrake, L., Merrelli, A., Natraj, V., Nelson, R. R., Osterman, G. B., Payne, V. H., Taylor, T. E., Wunch, D., Drouin, B. J., Oyafuso, F., Chang, A.,

695   McDuffie, J., Smyth, M., Baker, D. F., Basu, S., Chevallier, F., Crowell, S. M. R., Feng, L., Palmer, P. I., Dubey, M., García, O. E., Griffith, D. W. T., Hase, F., Iraci, L. T., Kivi, R., Morino, I., Notholt, J., Ohyama, H., Petri, C., Roehl, C. M., Sha, M. K., Strong, K., Sussmann, R., Te, Y., Uchino, O., and Velazco, V. A.: Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm, Atmospheric Measurement Techniques, 11, 6539–6576, https://doi.org/10.5194/amt-11-6539-2018, 2018.

Peiro, H., Crowell, S., Schuh, A., Baker, D. F., O'Dell, C., Jacobson, A. R., Chevallier, F., Liu, J., Eldering, A., Crisp, D., Deng, F., Weir,

700   B., Basu, S., Johnson, M. S., Philip, S., and Baker, I.: Four years of global carbon cycle observed from the Orbiting Carbon Observatory 2 (OCO-2) version 9 and in situ data and comparison to OCO-2 version 7, Atmospheric Chemistry and Physics, 22, 1097–1130, https://doi.org/10.5194/acp-22-1097-2022, 2022.

Petzold, A., Thouret, V., Gerbig, C., Zahn, A., Brenninkmeijer, C. A. M., Gallagher, M., Hermann, M., Pontaud, M., Ziereis, H., Boulanger, D., Marshall, J., Nèdèlec, P., Smit, H. G. J., Friess, U., Flaud, J.-M., Wahner, A., Cammas, J.-P., Volz-Thomas, A., and TEAM, I.: Global-

705   scale atmosphere monitoring by in-service aircraft - current achievements and future prospects of the European Research Infrastructure IAGOS, Tellus B: Chemical and Physical Meteorology, 67, 28 452, https://doi.org/10.3402/tellusb.v67.28452, 2015.

Rasmussen, C. E. and Williams, C.: Gaussian Processes for Machine Learning, the MIT Press, 2, 4, 2006.

Renaut, R. A., Vatankhah, S., and Ardestani, V. E.: Hybrid and iteratively reweighted regularization by unbiased predictive risk and weighted GCV for projected systems, SIAM Journal on Scientific Computing, 39, B221–B243, 2017.

710   Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, vol. 2, World scientific, 2000.

Saibaba, A. K. and Kitanidis, P. K.: Efficient methods for large-scale linear inversion using a geostatistical approach, Water Resources Research, 48, 2012.

Saibaba, A. K. and Kitanidis, P. K.: Fast computation of uncertainty quantification measures in the geostatistical approach to solve inverse problems, Advances in Water Resources, 82, 124 – 138, https://doi.org/http://dx.doi.org/10.1016/j.advwatres.2015.04.012, 2015.

715   Saibaba, A. K., Chung, J., and Petroske, K.: Efficient Krylov subspace methods for uncertainty quantification in large Bayesian linear inverse problems, Numerical Linear Algebra with Applications, 27, e2325, https://doi.org/https://doi.org/10.1002/nla.2325, 2020.

Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M., and Lenzen, F.: Variational methods in imaging, 2009.

Shusterman, A. A., Teige, V. E., Turner, A. J., Newman, C., Kim, J., and Cohen, R. C.: The BErkeley Atmospheric CO_2 Observation Network: initial evaluation, Atmospheric Chemistry and Physics, 16, 13 449–13 463, https://doi.org/10.5194/acp-16-13449-2016, 2016.

720   Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, Other Titles in Applied Mathematics, SIAM, 2005.

Yadav, V. and Michalak, A. M.: Improving computational efficiency in large linear inverse problems: an example from carbon dioxide flux estimation, Geoscientific Model Development, 6, 583–590, https://doi.org/10.5194/gmd-6-583-2013, 2013.

Zammit-Mangion, A., Bertolacci, M., Fisher, J., Stavert, A., Rigby, M. L., Cao, Y., and Cressie, N.: WOMBAT: A fully Bayesian global flux-inversion framework, arXiv preprint arXiv:2102.04004, 2021.