Technical note: Posterior Uncertainty Estimation via a Monte Carlo Procedure Specialized for Data Assimilation

Michael Stanley[1], Mikael Kuusela[2], Brendan Byrne[3], and Junjie Liu[4]

[1, 2]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA, USA
[3, 4]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Correspondence: Michael Stanley (mcstanle@andrew.cmu.edu)

Abstract. Through the Bayesian lens of data assimilation, uncertainty on model parameters is traditionally quantified through the posterior covariance matrix. However, in modern settings involving high-dimensional and computationally expensive forward models, posterior covariance knowledge must be relaxed to deterministic or stochastic approximations. In the carbon flux inversion literature, Chevallier et al. (2007) proposed a stochastic method capable of approximating posterior variances of

5   linear functionals of the model parameters that is particularly well-suited for large-scale Earth-system data assimilation tasks. This note formalizes this algorithm and clarifies its properties. We provide a formal statement of the algorithm, demonstrate why it converges to the desired posterior variance quantity of interest, and provide additional uncertainty quantification allowing incorporation of the Monte Carlo sampling uncertainty into the method's Bayesian credible intervals. The methodology is demonstrated using toy simulations and a realistic carbon flux inversion observing system simulation experiment.

10   **Keywords** — Bayesian inference; data assimilation; uncertainty quantification; carbon flux inversion; uncertainty on uncertainty; observing system simulation experiment

1 Introduction

Uncertainty quantification (UQ) for data assimilation (DA) tasks is often non-trivial, but scientifically paramount to their understanding and interpretation. Since DA broadly describes methods combining observations with a computational model of a physical system, a Bayesian framework is often sensible for inference on the model parameters, as the posterior distribution

15   quantifies knowledge resulting from this combination. As such, Bayesian statistical models are regularly used as the UQ framework. For example, Bayesian procedures play a central role in the general idea of optimal estimation (Rogers, 2000), the broad field of DA (Kalnay, 2003), and the more specific field of carbon flux estimation (Deng et al., 2014; Liu et al., 2016). Inference for DA tasks using this statistical framework is typically challenging due to high-dimensional settings (e.g., high-

20   resolution spatiotemporal grids) and the computer model's implicit numerical definition of the physical system of interest, often requiring supercomputers and long compute times. Prior and observation error distributions are often assumed to be Gaussian, yielding a Gaussian posterior distribution under a linear forward model. Although a Gaussian posterior can be exactly characterized by its mean vector and covariance matrix, the high-dimensionality makes dealing directly with the posterior covariance matrix intractable and the implicit computationally demanding forward model makes infeasible standard traditional

25 Bayesian computational techniques, such as Markov Chain Monte Carlo (MCMC). The implicit posterior necessitates the development of computational methods that implicitly access it.

CO$_2$ flux inversion is a representative example of a high-dimensional DA task to which Bayesian modeling is applied and used to compute estimated flux fields (Crowell et al., 2019; Enting et al., 1995; Gurney et al., 2002). In this problem, estimates of net surface-atmosphere CO$_2$ fluxes are inferred from atmospheric CO$_2$ measurements, with fluxes and atmospheric

30 measurements being related by a chemical transport model (the computational forward model). However, the relatively sparse atmospheric CO$_2$ observations underconstrain surface fluxes of CO$_2$, and regularization with prior information is the Bayesian approach to making the problem well-posed. These analyses have historically assimilated measurements of atmospheric CO$_2$ from a global network of flask and in situ measurements (Enting et al., 1995), but more recent work (Byrne et al., 2023; Crowell et al., 2019; Deng et al., 2014; Houweling et al., 2015) has shifted to assimilating space-based column-averaged

35 dry-air mole fractions, denoted X$_{CO_2}$, as observations availability has expanded since 2009. In these analyses, the prior and error distributions are typically assumed to be Gaussian and the forward model can be reasonably assumed linear in the net surface-atmosphere fluxes.

When the number of model parameters is low and a forward model run is inexpensive, it is possible to explicitly construct the posterior covariance matrix. Successful examples of this approach date back at least to Vasco et al. (1993) in seismic to-

40 mography, where inversion is performed on 12,496 model parameters. However, more contemporary problems typically have orders of magnitude more parameters and substantially more expensive forward models, requiring other approaches to access posterior covariance matrix information. Once the discretization of the computational model is set, the dimensionality problem can be handled either by defining an approximate statistical model on a lower dimensional problem, or by working in some subspace of the full-dimensional problem. A recent example of the first strategy is seen in Zammit-Mangion et al.

45 (2022) in the WOMBAT inversion system which lowers the dimension of the statistical model via an intelligently chosen set of basis functions, facilitating MCMC. Alternatively, Petra et al. (2014) propose with Stochastic Newton MCMC (SN-MCMC) the possibility for MCMC in the full parameter space by using a low-rank approximation to the posterior covariance within the proposal distribution of a Metropolis–Hasting algorithm. Although WOMBAT and SN-MCMC are both MCMC-based, WOMBAT assumes a linear forward model, while SN-MCMC does not, allowing it to characterize non-Gaussian posteriors.

50 Staying with a linear forward model assumption, other approaches leverage low-rank posterior covariance approximations. Flath et al. (2011) develop a low-rank algorithm for approximating the posterior covariance by computing the leading eigenvalues and eigenvectors of a prior-conditioned Hessian matrix of the associated objective function (i.e., the log posterior). In a similar spirit, Kalmikov and Heimbach (2014) provide a derivative-based algorithm to compute leading Hessian eigenvalues and eigenvectors and extend the uncertainty quantification to quantities of interest in global ocean state estimation. The

55 algorithms in both Flath et al. (2011) and Kalmikov and Heimbach (2014) rely upon the Lanczos method (Lanczos, 1950) for matrix-free computation of the low-rank approximation. Alternatively, Bousserez and Henze (2018) more recently proposed a low-rank approximation algorithm dependent upon the randomized SVD algorithm (Halko et al., 2011). All of the aforementioned methods can be grouped by their reliance upon some low-dimensional deterministic approximation.

In contrast, stochastic approximations of the posterior distribution rely neither upon pre-inversion dimension reductions nor
60   low-rank matrix approximations, but rather generate ensembles of inversions using random generators. In carbon flux inversion,
Chevallier et al. (2007) developed such a method to estimate the posterior variance of *functionals* of the flux field (i.e., maps
from the flux field to the reals). The method uses the forward model, specified prior, and known observation error distributions
in a particularly efficient manner. Broadly, the algorithm creates an ensemble of prior means and observation errors, sampling
according to their respective distributions. For each ensemble member, it finds the maximum a posteriori (MAP) estimator, to
65   which the functional is applied. Finally, it finds the empirical variance across the ensemble members to estimate the posterior
variance of the functional. This method is well-suited for carbon flux estimation and DA UQ more generally for a few key
reasons. First, each ensemble member is computationally independent, making the method parallelizable and hence offering
a substantial computational benefit compared to sequential methods, such as MCMC. Second, although in general prior mis-
specification biases the posterior, the prior mean does not need to be correctly specified in order for the procedure to produce
70   an unbiased estimator of the posterior variance. Third, the ensemble of inversions can flexibly produce UQ estimates for ar-
bitrary functionals post hoc, as opposed to requiring the specification of a functional ahead of the analysis. Finally, since this
method is more generally a Monte Carlo (MC) method for a Gaussian statistical model, the method's sampling uncertainty can
be analytically characterized and accounted for in the final UQ estimate. The ability to easily characterize this uncertainty of
the uncertainty stands in contrast to the difficulty in characterizing deterministic error of the aforementioned low-dimensional
75   approaches.

Although Chevallier et al. (2007) appear to have been the first to develop this method, which was later applied in Liu et al.
(2016), we are unaware of a formal statement or analysis of this algorithm. These previous works also did not quantify the
algorithm's MC uncertainty. As such, the primary contributions of this paper are a rigorous formal statement of the algorithm,
an analysis showing the convergence of its output to the true posterior quantity of interest, and uncertainty quantification of the
80   algorithm itself so that the algorithm's sampling uncertainty can be accounted for in the final inference.

The rest of this paper is structured as follows. In Section 2, we fully describe the algorithm, present mathematical results
proving its correctness, and derive deflation and inflation factors to apply to the estimated posterior uncertainty to quantify the
MC uncertainty. Proofs of the mathematical results can be found in Appendix A. In Section 3, we provide two experimental
demonstrations: the first is a low-dimensional problem in which we explicitly know the linear forward model and the second
85   is a carbon flux observing system simulation experiment (OSSE) to which we applied this method to compute global monthly
flux credible intervals along their MC uncertainty. Finally, we provide some concluding remarks in Section 4. For reference,
all mathematical notation in order of appearance is collected in Table 1.

2   Monte Carlo Method Exposition, Analysis, and Uncertainty Quantification

2.1   The Bayesian 4D-Var Setup

90   Following along with the mathematical setup of Henze et al. (2007), the prior and posterior distributions are defined in a scaling
factor space and hence the prior and posterior distributions on the physical quantity of interest are obtained by multiplying the

Table 1. Mathematical symbols and notation used herein (in order of appearance).

$\mathbf{c} \in \mathbb{R}^m$	Scaling factors	$\tilde{\mathbf{y}} \in \mathbb{R}^n$	X_{CO_2} Observations
$\boldsymbol{\mu} \in \mathbb{R}^m$	Control fluxes	f	Forward model
$\boldsymbol{\epsilon} \in \mathbb{R}^n$	Observation Noise	$\mathbf{R}$	Observation Noise Covariance
$\mathbf{A} \in \mathbb{R}^{n \times m}$	Linear forward model	$\mathbf{a} \circ \mathbf{b}$	Element-wise multiplication of $\mathbf{a}$ and $\mathbf{b}$
$\mathbf{z} \in \mathbb{R}^n$	Non-biospheric X_{CO_2} component	$\mathbf{y} \in \mathbb{R}^n$	Biospheric component of X_{CO_2}
$\mathbf{c}^b \in \mathbb{R}^m$	Prior scaling factor expectation	$\mathbf{B} \in \mathbb{R}^{m \times m}$	Prior scaling factor covariance
$\pi(\mathbf{c} \mid \mathbf{y})$	Posterior scaling factor density	$b^2 \in \mathbb{R}_+$	Prior variance parameter
$\mathbf{I}_m \in \mathbb{R}^{m \times m}$	$m \times m$ identity matrix	$\boldsymbol{\theta} = \mathbf{c} \circ \boldsymbol{\mu}$	Flux vector
$\mathbf{A}_{\boldsymbol{\mu}} \in \mathbb{R}^{n \times m}$	Forward model with control flux $\boldsymbol{\mu}$	$\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$	Posterior scaling factor covariance
$\boldsymbol{\alpha} \in \mathbb{R}^m$	Posterior scaling factor expectation	$\boldsymbol{\delta} \in \mathbb{R}^m$	Posterior flux expectation
$\boldsymbol{\Gamma} \in \mathbb{R}^{m \times m}$	Posterior flux covariance	$(\mathbf{c}_k, \mathbf{y}_k) \in \mathbb{R}^m \times \mathbb{R}^n$	kth MC sample
$\boldsymbol{\Sigma}_{\mathbf{c}_{MAP}^k} \in \mathbb{R}^{m \times m}$	MC MAP estimator covariance	$\mathbf{h} \in \mathbb{R}^m$	Functional of interest
$\overline{\varphi} \in \mathbb{R}$	Mean MC functional value	$\hat{\sigma}_\varphi^2 \in \mathbb{R}_+$	Empirical functional variance
χ_{M-1}^2	Chi-squared distribution with $M-1$ dof	$\chi_{M-1,\alpha/2}^2$	Chi-squared $(\alpha/2)$-quantile
$\alpha \in (0,1)$	Frequentist confidence level	$\gamma \in (0,1)$	Bayesian credible interval level
L	Deflation factor for MC variance	R	Inflation factor for MC variance

respective scaling factor by a control quantity. In carbon flux estimation, the control quantity is a control flux, typically an ansatz CO_2 flux between the Earth's surface and the atmosphere. Note that if the prior distribution mean in scaling factor space is unity, then the control flux is also the prior mean in the physical quantity of interest space. Mathematically, fix the scaling
95 factor vector $\mathbf{c} \in \mathbb{R}^m$ and let $\tilde{\mathbf{y}} \in \mathbb{R}^n$ be the observation vector and $\boldsymbol{\mu} \in \mathbb{R}^m$ the control physical quantity. The following model generally describes the relationship between the scaling factors and the observations:

$$\tilde{\mathbf{y}} = f(\mathbf{c}; \boldsymbol{\mu}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \tag{1}$$

where $\mathbf{R} \in \mathbb{R}^{n \times n}$ is the observation covariance matrix. The observation vector $\tilde{\mathbf{y}}$ is a sequence of X_{CO_2} observations produced from a remote sensing satellite, e.g., GOSAT or OCO-2 (O'Dell et al., 2012). This model expression is a composition of an
100 atmospheric transport model mapping scaling factors to atmospheric CO_2 concentrations composed with a remote sensing observation operator mapping CO_2 concentrations to X_{CO_2} scalar values. The atmospheric transport model is known to be affine due to the physics of CO_2 atmospheric transport. In reality, the true mapping from atmospheric CO_2 concentrations to X_{CO_2} is non-linear, but in line with Liu et al. (2016), we use an affine form involving the known GOSAT averaging kernel. As such, the affine composed function f is of the form $f(\mathbf{c}; \boldsymbol{\mu}) = \mathbf{A}(\mathbf{c} \circ \boldsymbol{\mu}) + \mathbf{z}$, where $\mathbf{A} \in \mathbb{R}^{n \times m}$ is the linear forward model
105 matrix, $\mathbf{c} \circ \boldsymbol{\mu}$ denotes the component-wise scaling of $\boldsymbol{\mu}$ by the scaling factors $\mathbf{c}$, and $\mathbf{z}$ is comprised of the non-biospheric CO_2 contribution to the observations along with the prior mean of the X_{CO_2} retrieval algorithm. As such, we define $\mathbf{y} := \tilde{\mathbf{y}} - \mathbf{z}$, giving the linear model

$$\mathbf{y} = \mathbf{A}(\mathbf{c} \circ \boldsymbol{\mu}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \tag{2}$$

on which the following analysis is performed. For ease of reference, we herein refer to the linear map $\mathbf{A}$ as the forward model.
110 We emphasize that the matrix $\mathbf{A}$ is not explicitly available to us, but is implicitly defined by the atmospheric transport model (and the satellite observation operator, which is usually explicitly available).

To regularize the problem and provide uncertainty quantification on the estimated scaling factors and fluxes, $\mathbf{c}$ in Equation (2) is given a Gaussian prior distribution, yielding the following Bayesian generative model:

$$\mathbf{c} \sim \mathcal{N}(\mathbf{c}^b, \mathbf{B}), \tag{3}$$

115 $$\mathbf{y} \mid \mathbf{c} \sim \mathcal{N}(\mathbf{A}(\mathbf{c} \circ \boldsymbol{\mu}), \mathbf{R}), \tag{4}$$

where $\mathbf{c}^b \in \mathbb{R}^m$ is the scaling factor prior mean and $\mathbf{B}$ is the prior covariance matrix.

Finding the posterior mean (or, equivalently, the posterior mode) defined by Equations (3) and (4) characterizes a common DA problem of interest. A typical DA approach in carbon flux estimation is four dimensional variational data assimilation (4D-Var) (Deng et al., 2014; Liu et al., 2016), a method optimizing carbon fluxes simultaneously over all time steps. 4D-Var can
120 be regarded as a least-squares optimization with an ℓ_2 regularizer (ridge regression), or, equivalently, as maximum a posteriori (MAP) estimation in the Bayesian paradigm. This connection means that the 4D-Var optimization is connected to the posterior resulting from the prior and likelihood in Equations (3) and (4). From the Bayesian perspective, the 4D-Var cost function $F(\mathbf{c})$ is the negative log-posterior density of the scaling factors given the observations,

$$F(\mathbf{c}) = -\log(\pi(\mathbf{c} \mid \mathbf{y})) = \frac{1}{2} \left(\mathbf{c} - \mathbf{c}^b \right)^\top \mathbf{B}^{-1} \left(\mathbf{c} - \mathbf{c}^b \right) + \frac{1}{2} \left(\mathbf{y} - \mathbf{A}(\mathbf{c} \circ \boldsymbol{\mu}) \right)^\top \mathbf{R}^{-1} \left(\mathbf{y} - \mathbf{A}(\mathbf{c} \circ \boldsymbol{\mu}) \right) + C, \tag{5}$$

125 where $C \in \mathbb{R}$ is a normalizing constant for the posterior distribution and $\pi(\mathbf{c} \mid \mathbf{y})$ denotes the posterior density. Thus, finding the MAP estimator, i.e., the $\mathbf{c}$ that maximizes the posterior density, is equivalent to finding the vector $\mathbf{c}$ that minimizes the 4D-Var cost function.

In this study (Sect. 3.2), the prior covariance $\mathbf{B}$ is parameterized with a single real value, $\mathbf{B} := b^2 \mathbf{I}_m$, where $b \in \mathbb{R}$. This is inline with several published studies (Deng et al., 2014; Liu et al., 2016), and implies that all prior spatio-temporal indices are
130 statistically independent. Similarly, the noise covariance $\mathbf{R}$ is assumed to be a diagonal matrix where each diagonal element is simply the variance of the corresponding X_{CO_2} observation. As such, each diagonal element depends on the uncertainty of its corresponding X_{CO_2} retrieval and the observations are assumed statistically independent given the scaling factors.

If the forward model $\mathbf{A}$ is known explicitly, the posterior mean and covariance of $\mathbf{c}$ are analytically tractable. We can then find the posterior uncertainty of the physical quantity $\boldsymbol{\theta} = \mathbf{c} \circ \boldsymbol{\mu}$ given the observations. We rewrite Equation (4) using the
135 short-hand notation for the $\circ$ operation (see Appendix A), i.e., $\mathbf{A}(\mathbf{c} \circ \boldsymbol{\mu}) = \mathbf{A}_{\boldsymbol{\mu}} \mathbf{c}$:

$$\mathbf{y} \mid \mathbf{c} \sim \mathcal{N}(\mathbf{A}_{\boldsymbol{\mu}} \mathbf{c}, \mathbf{R}). \tag{6}$$

Hence, $\mathbf{c} \mid \mathbf{y} \sim \mathcal{N}(\boldsymbol{\alpha}, \boldsymbol{\Sigma})$, where by the posterior mean and covariance Equations 4.3 and 4.7 in Rogers (2000) we have:

$$\boldsymbol{\Sigma} = \left(\frac{1}{b^2} \mathbf{I}_m + \mathbf{A}_{\boldsymbol{\mu}}^\top \mathbf{R}^{-1} \mathbf{A}_{\boldsymbol{\mu}} \right)^{-1} = \left(\left(\mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A} \right) \circ \boldsymbol{\mu} \boldsymbol{\mu}^\top + \frac{1}{b^2} \mathbf{I}_m \right)^{-1}, \tag{7}$$

$$\boldsymbol{\alpha} = \boldsymbol{\Sigma} \left(\mathbf{A}_{\boldsymbol{\mu}}^\top \mathbf{R}^{-1} \mathbf{y} + \frac{1}{b^2} \mathbf{c}^b \right) = \boldsymbol{\Sigma} \left(\left(\mathbf{A}^\top \mathbf{R}^{-1} \mathbf{y} \right) \circ \boldsymbol{\mu} + \frac{1}{b^2} \mathbf{c}^b \right), \tag{8}$$

where Equation (7) follows from Corollary 4 and Equation (8) follows from Lemma 5. Note, $\boldsymbol{\alpha}$ is also the MAP estimator of $\mathbf{c}$. Furthermore, the posterior distribution for the physical quantity, $\boldsymbol{\theta} = \mathbf{c} \circ \boldsymbol{\mu}$, is $\boldsymbol{\theta} \mid \mathbf{y} \sim \mathcal{N}(\boldsymbol{\delta}, \boldsymbol{\Gamma})$, where

$$\boldsymbol{\delta} = \mathbb{E}[\boldsymbol{\theta} \mid \mathbf{y}] = \mathbb{E}[\mathbf{c} \circ \boldsymbol{\mu} \mid \mathbf{y}] = \mathbb{E}[\mathbf{c} \mid \mathbf{y}] \circ \boldsymbol{\mu} = \boldsymbol{\alpha} \circ \boldsymbol{\mu}, \tag{9}$$

$$\boldsymbol{\Gamma} = \mathrm{Cov}[\boldsymbol{\theta} \mid \mathbf{y}] = \mathrm{Cov}[\mathbf{c} \circ \boldsymbol{\mu} \mid \mathbf{y}] = \mathrm{Cov}[\mathbf{c} \mid \mathbf{y}] \circ \boldsymbol{\mu}\boldsymbol{\mu}^\top = \boldsymbol{\Sigma} \circ \boldsymbol{\mu}\boldsymbol{\mu}^\top, \tag{10}$$

where we have used Lemmas 1 and 2 from Appendix A.

In practice, the forward model $\mathbf{A}$ is only known implicitly via a computer simulator, making direct use of Equations (7)–(10) intractable. Instead, the 4D-Var cost function in Equation (5) is minimized using the L-BFGS-B algorithm (Byrd et al., 1995) with the cost function gradient computed numerically using the adjoint method (Henze et al., 2007). After a handful of iterations, L-BFGS-B finds a reasonable approximation of the posterior mean / mode $\boldsymbol{\alpha}$, which yields a point estimator of $\boldsymbol{\theta}$ using Equation (9). We now describe a procedure that provides an approach for uncertainty quantification of $\boldsymbol{\theta}$ despite the intractability of the posterior covariance $\boldsymbol{\Sigma}$.

2.2 The Monte Carlo Procedure

To execute the Monte Carlo procedure introduced in Chevallier et al. (2007), we generate M ensemble members. For each $k = 1, 2, \ldots, M$, we sample a new prior mean $\mathbf{c}_k$ and new observation $\mathbf{y}_k$ as follows:

$$\mathbf{c}_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{1}, b^2 \mathbf{I}_m), \tag{11}$$

$$\mathbf{y}_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{A}\boldsymbol{\mu}, \mathbf{R}), \tag{12}$$

where b^2 is the prior uncertainty mentioned in Section 2.1. Notice that $\mathbf{A}\boldsymbol{\mu}$ is known after a single forward model run, and hence Equation (12) is more illuminatingly seen as sampling Gaussian noise for each Monte Carlo sample: letting $\boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, for $k = 1, \ldots, M$, each Monte Carlo iteration involves sampling a pair, $(\mathbf{c}_k, \mathbf{y}_k) \in \mathbb{R}^m \times \mathbb{R}^n$, where $\mathbf{y}_k := \mathbf{A}\boldsymbol{\mu} + \boldsymbol{\epsilon}_k$ for each $k = 1, \ldots, M$.

The MAP estimator from Equation (8) corresponding to prior mean $\mathbf{c}_k$ and observation $\mathbf{y}_k$ is analytically tractable for each ensemble member when $\mathbf{A}$ is explicitly known:

$$\mathbf{c}_{MAP}^k = \boldsymbol{\Sigma}\left(\left(\mathbf{A}^\top \mathbf{R}^{-1} \mathbf{y}_k \right) \circ \boldsymbol{\mu} + \frac{1}{b^2} \mathbf{c}_k \right). \tag{13}$$

Similarly, the covariance matrix of this MAP estimator, henceforth denoted as $\boldsymbol{\Sigma}_{\mathbf{c}_{MAP}^k}$, is

$$\boldsymbol{\Sigma}_{\mathbf{c}_{MAP}^k} = \boldsymbol{\Sigma} \, \mathrm{Cov}\left[\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{y}_k \circ \boldsymbol{\mu} + \frac{1}{b^2}\mathbf{c}_k\right]\boldsymbol{\Sigma}^\top \tag{14}$$

165
$$= \boldsymbol{\Sigma}\left(\mathrm{Cov}\left[\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{y}_k \circ \boldsymbol{\mu}\right] + \frac{b^2}{b^4}\mathbf{I}_m\right)\boldsymbol{\Sigma}$$

$$= \boldsymbol{\Sigma}\left(\mathrm{Cov}\left[\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{y}_k\right] \circ \boldsymbol{\mu}\boldsymbol{\mu}^\top + \frac{1}{b^2}\mathbf{I}_m\right)\boldsymbol{\Sigma}$$

$$= \boldsymbol{\Sigma}\left(\left(\mathbf{A}^\top \mathbf{R}^{-1}\mathrm{Cov}\left[\mathbf{y}_k\right]\mathbf{R}^{-1}\mathbf{A}\right) \circ \boldsymbol{\mu}\boldsymbol{\mu}^\top + \frac{1}{b^2}\mathbf{I}_m\right)\boldsymbol{\Sigma}$$

$$= \boldsymbol{\Sigma}\left(\left(\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{A}\right) \circ \boldsymbol{\mu}\boldsymbol{\mu}^\top + \frac{1}{b^2}\mathbf{I}_m\right)\boldsymbol{\Sigma} = \boldsymbol{\Sigma},$$

since $\boldsymbol{\Sigma}^{-1} = \left(\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{A} \circ \boldsymbol{\mu}\boldsymbol{\mu}^\top + \frac{1}{b^2}\mathbf{I}_m\right)$. This shows that the covariance matrix of the Monte Carlo ensemble of scaling factor

170 MAP estimators is equal to the desired posterior covariance $\boldsymbol{\Sigma}$. Note, the $\circ$ operation step on the third line of derivation (14) follows from Lemma 2 in Appendix A. To the best of our knowledge, proof of this equality has not appeared in previous literature on this method.

This covariance equality also exists in the physical quantity space, e.g., carbon flux space. The estimator of the physical quantity corresponding to $\mathbf{c}_{MAP}^k$ is

175 $$\boldsymbol{\theta}_k = \mathbf{c}_{MAP}^k \circ \boldsymbol{\mu}. \tag{15}$$

Using the result from Lemma 2 in Appendix A, the covariance matrix $\mathrm{Cov}[\boldsymbol{\theta}_k]$ of this estimator is

$$\mathrm{Cov}[\boldsymbol{\theta}_k] = \mathrm{Cov}[\mathbf{c}_{MAP}^k \circ \boldsymbol{\mu}] = \mathrm{Cov}[\mathbf{c}_{MAP}^k] \circ \boldsymbol{\mu}\boldsymbol{\mu}^\top = \boldsymbol{\Sigma} \circ \boldsymbol{\mu}\boldsymbol{\mu}^\top = \boldsymbol{\Gamma}. \tag{16}$$

Hence, the covariance matrix of the Monte Carlo physical quantity estimator is equal to the posterior covariance matrix of that physical quantity.

180 However, for most DA tasks the forward model is not explicitly available, so each ensemble member MAP estimator $\mathbf{c}_{MAP}^k$ must be obtained with an iterative optimization algorithm minimizing (5) with $\mathbf{c}_k$ and $\mathbf{y}_k$ as the prior mean and observation vectors. Once these ensemble members are obtained, we could in principle estimate the posterior scaling factor covariance matrix $\boldsymbol{\Sigma}$ with the empirical covariance estimator $\widehat{\boldsymbol{\Sigma}}$ based on the Monte Carlo ensemble as follows,

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{M-1}\sum_{k=1}^M \left(\mathbf{c}_{MAP}^k - \overline{\mathbf{c}}\right)\left(\mathbf{c}_{MAP}^k - \overline{\mathbf{c}}\right)^\top, \tag{17}$$

185 where $\overline{\mathbf{c}} = \frac{1}{M}\sum_{k=1}^M \mathbf{c}_{MAP}^k$. To translate Equation (17) to the physical quantity space, we simply plug in the empirical covariance estimator to Equation (16), i.e.,

$$\hat{\boldsymbol{\Gamma}} = \widehat{\mathrm{Cov}[\boldsymbol{\theta}_k]} = \widehat{\boldsymbol{\Sigma}} \circ \boldsymbol{\mu}\boldsymbol{\mu}^\top. \tag{18}$$

In practice, DA scenarios like carbon flux inversion are typically high dimensional, making direct interaction with these covariance matrices difficult. Indeed, accurate estimation of $\boldsymbol{\Sigma}$ using Equation (17) would require an enormously large Monte

190   Carlo ensemble and would require storing and working with an $m \times m$ matrix, where $m \sim 10^5$ or larger. Fortunately, we often care about the variance of one-dimensional summaries of $\boldsymbol{\theta}$, such as the posterior flux variance for a specific region during some time period, as opposed to the full posterior covariance matrix. For instance, we might wish to estimate North American fluxes over a month. For the remaining presentation of these ideas, we only discuss the posterior scaling factor vector, $\mathbf{c} \mid \mathbf{y}$, but note that obtaining the posterior flux is achieved by a simple component-wise scaling with the control flux $\boldsymbol{\mu}$.

195   Obtaining quantities of the above type is mathematically implemented using a linear functional of the underlying high-dimensional parameter. That is, we wish to characterize the posterior of $\varphi(\mathbf{c}) = \boldsymbol{h}^{\top}\mathbf{c}$, where $\boldsymbol{h} \in \mathbb{R}^m$ contains weights necessary to aggregate the desired scaling factors. Hence, building off Equations (8) and (7), we obtain the posterior distribution for the functional of interest:

$$\varphi(\mathbf{c}) \mid \mathbf{y} \sim \mathcal{N}(\boldsymbol{h}^{\top}\boldsymbol{\alpha}, \boldsymbol{h}^{\top}\boldsymbol{\Sigma}\boldsymbol{h}). \tag{19}$$

200   We wish to obtain the posterior variance of this functional. Define $\sigma_\varphi^2 = \mathrm{Var}(\varphi(\mathbf{c}) \mid \mathbf{y}) = \boldsymbol{h}^{\top}\boldsymbol{\Sigma}\boldsymbol{h}$. We could inefficiently estimate this using $\hat{\sigma}_\varphi^2 = \boldsymbol{h}^{\top}\hat{\boldsymbol{\Sigma}}\boldsymbol{h}$, but we wish to avoid working directly with the full empirical covariance matrix. The following algebraic steps provide a better alternative:

$$\hat{\sigma}_\varphi^2 = \boldsymbol{h}^{\top} \left(\frac{1}{M-1} \sum_{k=1}^{M} \left(\mathbf{c}_{MAP}^k - \overline{\mathbf{c}} \right) \left(\mathbf{c}_{MAP}^k - \overline{\mathbf{c}} \right)^{\top} \right) \boldsymbol{h} \tag{20}$$

$$= \frac{1}{M-1} \sum_{k=1}^{M} \boldsymbol{h}^{\top} \left(\mathbf{c}_{MAP}^k - \overline{\mathbf{c}} \right) \left(\mathbf{c}_{MAP}^k - \overline{\mathbf{c}} \right)^{\top} \boldsymbol{h} \tag{21}$$

205
$$= \frac{1}{M-1} \sum_{k=1}^{M} \left[\boldsymbol{h}^{\top} \left(\mathbf{c}_{MAP}^k - \overline{\mathbf{c}} \right) \right]^2 \tag{22}$$

$$= \frac{1}{M-1} \sum_{k=1}^{M} \left(\varphi_k - \overline{\varphi} \right)^2, \tag{23}$$

where $\varphi_k = \boldsymbol{h}^{\top}\mathbf{c}_{MAP}^k$ and $\overline{\varphi} = \boldsymbol{h}^{\top}\overline{\mathbf{c}} = \frac{1}{M}\sum_{k=1}^{M}\varphi_k$. The above algebra shows that the posterior variance of the functional can be computed using the functionals of the Monte Carlo samples without having to form the full empirical covariance matrix. See Algorithm 1 for a succinct exposition of the above procedure. Note, the control flux can be built into the definition of $\boldsymbol{h}$

210   so that the functional has the desired units. Notice also that the functional does not need to be specified when creating the Monte Carlo ensemble. As long as the ensemble $\{\mathbf{c}_{MAP}^k\}_{k=1}^{M}$ is stored and made available to the end users, they may evaluate post-hoc the uncertainty of any functional that is of interest in their specific use-case.

2.3   Quantifying the Monte Carlo Uncertainty

Although Section 2.2 establishes the equality of the Monte Carlo MAP estimator ensemble member covariance to the posterior
215   covariance (and therefore the equality of the Monte Carlo ensemble member functional variance to the posterior functional variance), we have not yet established that the empirical covariance matrix (and functional variance) converges in probability to the true posterior covariance matrix (and functional variance). There are consistency results showing the empirical covariance

---

Algorithm 1 Monte Carlo Algorithm to Estimate Posterior Uncertainty in 4D-Var Data Assimilation

Inputs:

- $M \in \mathbb{N}$: Number of Monte Carlo samples.

- $\mathbf{R} \in \mathbb{R}^{n \times n}$: Observation error covariance built from retrieval uncertainties.

- $b^2 \in \mathbb{R}_+$: Scaling factor prior variance.

- $\mathbf{A} \in \mathbb{R}^{n \times m}$: Forward model mapping fluxes to X_{CO_2} satellite observations (note, this can be known either explicitly via a matrix or implicitly via a computer simulator).

- $\boldsymbol{\mu} \in \mathbb{R}^m$: Control flux.

- $\boldsymbol{h} \in \mathbb{R}^m$: Vector defining the functional of interest.

Steps:

1. Let $\mathcal{S}$ denote an array of length M that will store the MAP estimators for each Monte Carlo sample.

2. For $k = 1, \ldots, M$:

    (a) Simulate $\mathbf{c}_k \sim \mathcal{N}(\mathbf{1}, b^2 \mathbf{I}_m)$.

    (b) Simulate $\boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ and construct $\mathbf{y}_k = \mathbf{A}\boldsymbol{\mu} + \boldsymbol{\epsilon}_k$.

    (c) Find MAP estimator $\mathbf{c}_{MAP}^k$. If $\mathbf{A}$ is known explicitly, it can be found using Equation (13). If $\mathbf{A}$ is known implicitly through a computational model, use a numerical optimizer (e.g., L-BFGS-B) to optimize a the 4D-Var cost function as defined in Equation (5).

    (d) $\mathcal{S}[k] \leftarrow \mathbf{c}_{MAP}^k$.

3. Estimate Posterior Functional Variance:

    (a) Compute the mean Monte Carlo sample functional: $\overline{\varphi} = \frac{1}{M} \sum_{k=1}^{M} \varphi_k$, where $\varphi_k = \boldsymbol{h}^\top \mathbf{c}_{MAP}^k$.

    (b) Compute the empirical posterior functional variance:

$$\hat{\sigma}_\varphi^2 = \frac{1}{M-1} \sum_{k=1}^{M} (\varphi_k - \overline{\varphi})^2.$$

---

matrix converging in probability to the true covariance matrix (see, for instance, Chapter 6 in Wainwright (2019)). However, since this application is primarily concerned with linear functionals of the form $\varphi(\mathbf{c}) = \boldsymbol{h}^\top \mathbf{c}$ as described in Section 2.2, we

220 can appeal directly to the consistency of the sample variance as shown, for example, in Chapter 5 in Casella and Berger (2002).

Additionally, using the above algorithm, we would like to know either the uncertainty of the variance estimate given the number of Monte Carlo samples, or the number of samples required to obtain a particular level of Monte Carlo uncertainty on the variance. In essence, we would like to quantify the uncertainty of our uncertainty. To do so, we take a frequentist

approach and construct confidence intervals on $\hat{\sigma}^2_\varphi$. The confidence intervals can be constructed by recognizing that the ratio
of the Monte Carlo functional empirical posterior variance to the true functional posterior variance scaled by $(M-1)$ follows
a χ^2_{M-1} distribution.

Since each sampled $\mathbf{c}^k_{MAP}$ is a linear function of other Gaussian samples (see Equation (13)), $\mathbf{c}^k_{MAP}$ is also Gaussian,
and thus the random variables $\varphi_k = \boldsymbol{h}^\top \mathbf{c}^k_{MAP}$ $(k=1,\ldots,M)$ are sampled independently and identically from a Gaussian
distribution with some mean and variance $\sigma^2_\varphi = \boldsymbol{h}^\top \boldsymbol{\Sigma} \boldsymbol{h}$. By Theorem 5.3.1 of Casella and Berger (2002), we have the following
distributional result,

$$\frac{(M-1)\hat{\sigma}^2_\varphi}{\sigma^2_\varphi} \sim \chi^2_{M-1}. \tag{24}$$

Thus, for $\alpha \in (0,1)$, the distribution in Equation (24) enables creating a $1-\alpha$ confidence interval for the true posterior variance,
σ^2_φ, as a function of the empirical posterior variance, $\hat{\sigma}^2_\varphi$. Using the exact distribution in Equation (24), we can create either
one- or two-sided confidence intervals. Focusing on the two-sided case, we have,

$$\mathbb{P}\left\{ \chi^2_{M-1,\alpha/2} \leq \frac{(M-1)\hat{\sigma}^2_\varphi}{\sigma^2_\varphi} \leq \chi^2_{M-1,1-\alpha/2} \right\} = 1-\alpha, \tag{25}$$

where $\chi^2_{M-1,\alpha/2}$ is the $\alpha/2$-quantile of a chi-squared distribution with $M-1$ degrees of freedom. Hence, with some algebraic
manipulation we arrive at the confidence interval of the posterior variance,

$$\mathbb{P}\left\{ \frac{(M-1)\hat{\sigma}^2_\varphi}{\chi^2_{M-1,1-\alpha/2}} \leq \sigma^2_\varphi \leq \frac{(M-1)\hat{\sigma}^2_\varphi}{\chi^2_{M-1,\alpha/2}} \right\} = 1-\alpha. \tag{26}$$

Since in practice we would like to characterize uncertainty in the same units as the flux estimate, we can provide an analogous
confidence interval for the posterior standard deviation by taking square roots of all the terms within the probability statement
in Equation (26), giving

$$\mathbb{P}\left\{ \hat{\sigma}_\varphi \sqrt{\frac{M-1}{\chi^2_{M-1,1-\alpha/2}}} \leq \sigma_\varphi \leq \hat{\sigma}_\varphi \sqrt{\frac{M-1}{\chi^2_{M-1,\alpha/2}}} \right\} = 1-\alpha. \tag{27}$$

Equation (27) facilitates the computation of a $(1-\alpha) \times 100\%$ frequentist interval estimator of the Bayesian credible interval
for the functional of interest φ. For each endpoint of the true Bayesian credible interval, we find a confidence interval such
that the probability that both endpoint confidence intervals simultaneously cover the true credible interval endpoints is $1-\alpha$.
Let $\gamma \in (0,1)$ and $\varphi_{MAP} = \boldsymbol{h}^\top \mathbf{c}_{MAP}$ be the functional MAP estimator as described in Section 2.1. Because the posterior is
Gaussian and φ_{MAP} is a linear functional, it is a one-dimensional Gaussian. Hence, the Bayesian $(1-\gamma) \times 100\%$ credible
interval is computed as follows,

$$\left[\underline{\varphi}^*, \overline{\varphi}^* \right] = \left[\varphi_{MAP} - z_{1-\gamma/2} \cdot \sigma_\varphi, \varphi_{MAP} + z_{1-\gamma/2} \cdot \sigma_\varphi \right], \tag{28}$$

where $z_{1-\gamma/2}$ is the $1-\gamma/2$ quantile of a standard Gaussian distribution. Equation (27) allows us to construct the aforementioned endpoint confidence intervals as follows. For readability, define $L^2 := \frac{M-1}{\chi^2_{M-1,1-\alpha/2}}$ and $R^2 := \frac{M-1}{\chi^2_{M-1,\alpha/2}}$. Thus, we have

the following,

$$
\begin{aligned}
1 - \alpha &= \mathbb{P}\left\{z_{1-\gamma/2}\hat{\sigma}_\varphi L \le z_{1-\gamma/2}\sigma_\varphi \le z_{1-\gamma/2}\hat{\sigma}_\varphi R\right\} \\
&= \mathbb{P}\{-z_{1-\gamma/2}\hat{\sigma}_\varphi R \le -z_{1-\gamma/2}\sigma_\varphi \le -z_{1-\gamma/2}\hat{\sigma}_\varphi L \text{ and} \\
&\qquad z_{1-\gamma/2}\hat{\sigma}_\varphi L \le z_{1-\gamma/2}\sigma_\varphi \le z_{1-\gamma/2}\hat{\sigma}_\varphi R\} \\
&= \mathbb{P}\{\varphi_{MAP} - z_{1-\gamma/2}\hat{\sigma}_\varphi R \le \varphi_{MAP} - z_{1-\gamma/2}\sigma_\varphi \le \varphi_{MAP} - z_{1-\gamma/2}\hat{\sigma}_\varphi L \text{ and} \\
&\qquad \varphi_{MAP} + z_{1-\gamma/2}\hat{\sigma}_\varphi L \le \varphi_{MAP} + z_{1-\gamma/2}\sigma_\varphi \le \varphi_{MAP} + z_{1-\gamma/2}\hat{\sigma}_\varphi R\} \\
&= \mathbb{P}\{\varphi_{MAP} - z_{1-\gamma/2}\hat{\sigma}_\varphi R \le \underline{\varphi}^* \le \varphi_{MAP} - z_{1-\gamma/2}\hat{\sigma}_\varphi L \text{ and} \\
&\qquad \varphi_{MAP} + z_{1-\gamma/2}\hat{\sigma}_\varphi L \le \overline{\varphi}^* \le \varphi_{MAP} + z_{1-\gamma/2}\hat{\sigma}_\varphi R\}.
\end{aligned}
$$

More concisely, defining

$$
\underline{I} := \left[\varphi_{MAP} - z_{1-\gamma/2}\hat{\sigma}_\varphi R, \varphi_{MAP} - z_{1-\gamma/2}\hat{\sigma}_\varphi L\right], \tag{29}
$$

$$
\overline{I} := \left[\varphi_{MAP} + z_{1-\gamma/2}\hat{\sigma}_\varphi L, \varphi_{MAP} + z_{1-\gamma/2}\hat{\sigma}_\varphi R\right], \tag{30}
$$

it follows that

$$
\mathbb{P}\left\{\underline{\varphi}^* \in \underline{I} \text{ and } \overline{\varphi}^* \in \overline{I}\right\} = 1 - \alpha. \tag{31}
$$

The intervals $\underline{I}$ and $\overline{I}$ quantify uncertainty on uncertainty, and provide a rigorous probabilistic characterization of the Monte Carlo procedure's uncertainty.

In practice, the original Bayesian credible interval in Equation (28) can be modified to account for the Monte Carlo uncertainty. To obtain an *upper bound* on the Bayesian credible interval, we apply an *inflation* factor (defined above as R), and thus obtain the interval $\left[\underline{\varphi}_u, \overline{\varphi}_u\right] = \left[\varphi_{MAP} - z_{1-\gamma/2}\hat{\sigma}_\varphi R, \varphi_{MAP} + z_{1-\gamma/2}\hat{\sigma}_\varphi R\right]$, such that $\mathbb{P}\left\{\left[\underline{\varphi}_u, \overline{\varphi}_u\right] \supset \left[\underline{\varphi}^*, \overline{\varphi}^*\right]\right\} = 1 - \alpha/2$. This probability is $(1 - \alpha/2)$ instead of $(1 - \alpha)$ since the probability in Equation (25) evaluated with only the lower bound yields a probability of $(1 - \alpha/2)$. Following the same steps as above, we obtain lower and upper bounds on the lower and upper endpoints of credible interval (28), respectively, holding with probability exactly $(1 - \alpha/2)$. Similarly, to obtain a *lower bound* on the Bayesian credible interval, we apply a *deflation* factor (defined above as L) and thus obtain the interval $\left[\underline{\varphi}_l, \overline{\varphi}_l\right] = \left[\varphi_{MAP} - z_{1-\gamma/2}\hat{\sigma}_\varphi L, \varphi_{MAP} + z_{1-\gamma/2}\hat{\sigma}_\varphi L\right]$ such that $\mathbb{P}\left\{\left[\underline{\varphi}_l, \overline{\varphi}_l\right] \subset \left[\underline{\varphi}^*, \overline{\varphi}^*\right]\right\} = 1 - \alpha/2$, holding with equality by the same logic as that used for the inflation factor.

Observing that the aforementioned inflation and deflation factors monotonically asymptote to one as the number of Monte Carlo samples M gets large, the effect of the Monte Carlo procedure wanes as the number of samples grows. As shown in Table 2, the deflation factor monotonically approaches one from below as M gets large while the inflation factor monotonically approaches one from above as M gets large. As is characteristic of DA methods, each Monte Carlo iteration requires a non-trivial amount of computation, which practically restricts the number of Monte Carlo samples that can be obtained. As such, the inflated interval protects against underestimating the uncertainty, while the deflated interval provides a lower bound or "best-case" scenario for the uncertainty of the Bayesian procedure.

Table 2. Inflation and deflation factors for Monte Carlo (MC) estimated posterior standard deviation with $\alpha = 0.05$. When $M = 100$, by inflating the MC estimated posterior standard deviation by a factor of 1.1607 (inflating by 16.07%), the extra uncertainty resulting from the MC procedure is accounted for with 97.5% confidence. Similarly, when $M = 100$, deflating the MC estimated posterior standard deviation by a factor of 0.8785 provides a lower bound on the true underlying Bayesian uncertainty with 97.5% confidence. When considered simultaneously, the inflation and deflation factors bracket the true uncertainty with 95% confidence.

# Monte Carlo samples, M	Deflation: $L = \sqrt{\frac{M-1}{\chi^2_{M-1,1-\alpha/2}}}$	Inflation: $R = \sqrt{\frac{M-1}{\chi^2_{M-1,\alpha/2}}}$
10	0.6987	1.7549
100	0.8785	1.1607
1,000	0.9580	1.0458
10,000	0.9863	1.0141
100,000	0.9956	1.0044
1,000,000	0.9986	1.0014

3 Numerical Examples

3.1 Low-Dimensional Example

285 We construct a two-dimensional toy example to provide a numerical demonstration that this MC procedure computes a consistent estimate of the posterior covariance, and is numerically close in practice. We define a linear forward model with the following matrix:

$$\mathbf{A} = \begin{bmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{bmatrix}, \tag{32}$$

where $\epsilon > 0$. Let $\boldsymbol{\theta} \in \mathbb{R}^2$ be the true state of some physical quantity and $\boldsymbol{\mu} \in \mathbb{R}^2$ be the control state. We use the values in Table 3

290 to demonstrate the agreement between the analytical equations for the Bayesian procedure and the Monte Carlo procedure in addition to showing their agreement with the empirical covariance computed from the Monte Carlo ensemble members. Using the Table 3 parameters and Equation (7), we obtain the following posterior covariance matrix:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2.10838562 & -0.0867085 \\ -0.0867085 & 0.8693668 \end{bmatrix}. \tag{33}$$

For the analytical covariance of the MAP estimator, we obtain the following matrix using Equation (14):

295 $$\boldsymbol{\Sigma}_{\mathbf{c}^k_{MAP}} = \begin{bmatrix} 2.10838562 & -0.0867085 \\ -0.0867085 & 0.8693668 \end{bmatrix}. \tag{34}$$

Table 3. Parameter settings for the low-dimensional example.

Parameter	Value	Description
θ	$\begin{bmatrix} 1 & 2 \end{bmatrix}^\top$	True state of the system
$\boldsymbol{\mu}$	$\begin{bmatrix} 1/2 & 1 \end{bmatrix}^\top$	Control state
n	2	The number of observations in each ensemble member (also the dimension of $\mathbf{y}$ in Equation (4))
M	10^6	The number of MC ensemble members
b^2	4	The prior variance for each element in the scaling factor vector
ϵ	0.05	Parameter of the matrix $\mathbf{A}$ defined in Equation (32)
σ^2	1	Observation error variance

Indeed, these matrices are expected to be the same. Using simulated ensemble members, we obtain the following empirical covariance matrix using Equation (17):

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 2.09697751 & -0.08748562 \\ -0.08748562 & 0.87053267 \end{bmatrix}. \tag{35}$$

This empirical covariance matrix is numerically very close to the analytical matrices. Using Equation (6.12) from Wainwright (2019), we can compute an upper bound on the relative error between the empirical covariance matrix and the true posterior covariance matrix with respect to the ℓ_2 operator norm that holds with at least some desired probability. Concretely, with $n = 2$ and $M = 10^6$, this deviation bound implies that the relative error is at most 0.01784 (1.784%) with probability at least 95%. This tight error bound matches the above numerical closeness.

All of the above quantities exist in the scaling factor space. For the physical quantity space covariance matrix, we obtain the following from Equations (10) and (16):

$$\boldsymbol{\Gamma} = \begin{bmatrix} 0.52709641 & -0.04335425 \\ -0.04335425 & 0.8693668 \end{bmatrix} \tag{36}$$

$$\text{Cov}[\boldsymbol{\theta}_k] = \begin{bmatrix} 0.52709641 & -0.04335425 \\ -0.04335425 & 0.8693668 \end{bmatrix} \tag{37}$$

$$\frac{1}{M-1}\sum_{k=1}^{M}(\theta_k - \overline{\theta})(\theta_k - \overline{\theta})^\top = \begin{bmatrix} 0.52424438 & -0.04374281 \\ -0.04374281 & 0.87053267 \end{bmatrix} \tag{38}$$

Again, we see that the agreement between the analytical forms and the empirical covariance matrix is very close.

This toy example demonstrates the correctness of the Monte Carlo procedure in terms of the covariance matrix. However, as seen in Algorithm 1, we are actually interested in scenarios in which we are considering a functional of high-dimensional parameters, and hence a variance. Such an example is considered in the following section.

3.2 Carbon Flux Inversion OSSE

We show here an example of this Monte Carlo procedure being used to compute posterior uncertainties for global carbon fluxes. We follow the flux inversion setup used by Byrne et al. (2019) (see Section 2.3 of that study). This setup uses the GEOS-Chem Adjoint model (Henze et al., 2007) to estimate scaling factors on a $4° \times 5°$ surface grid from January 2010 up to and including August. For each spatial point, there is one scaling factor parameter for each month, totaling $m = 72 \times 46 \times 8 = 26,492$ scaling factors parameters. This model is linear in terms of realistic fluxes (e.g., not including abnormally large negative fluxes), and hence amenable to this uncertainty quantification procedure. The OSSE defines ground-truth fluxes from the Joint UK Land Environment Simulator (JULES) (Clark et al., 2011; Harper et al., 2018) and uses Net Ecosystem Exchange (NEE) fluxes from NOAA's CarbonTracker version CT2016 (Peters et al. (2007), with updates documented at https://www.esrl.noaa.gov/gmd/ccgg/carbontracker/) as the control fluxes. The satellite X_{CO_2} observations for the assimilation are generated from the JULES fluxes by running a forward GEOS-Chem simulation and sampling the model with the GOSAT observational coverage and observation operator (O'Dell et al., 2012).

The prior uncertainty, as described in Equation (3), is set to $b = 1.5$ (where $\mathbf{B} := b^2\mathbf{I}_M$). To perform the Monte Carlo procedure, we draw $M = 60$ ensemble members, as described in Sec. 2.2. The X_{CO_2} observation uncertainty $\boldsymbol{\Sigma}$ (a diagonal matrix, as the observations are assumed to be independent) comes directly from the GOSAT data product and varies between observations. For each ensemble member k, the output of the GEOS-Chem Adjoint optimization provides monthly scaling factor MAP estimators $\mathbf{c}_{MAP}^k$ according to the ensemble member inputs as described by Equation (13). Each ensemble MAP estimator is then multiplied by the control flux to obtain a MAP estimator in flux space.

The functionals of interest φ are monthly global fluxes. The flux values on the 3-hour $4° \times 5°$ spatial-temporal grid are mapped to a global monthly flux using a weighted average with weights proportional to the surface area of each grid cell and uniform time weighting. The global flux posterior variance is computed for each month by finding the empirical variance of the Monte Carlo global flux members, as shown in Equation (23). To get a sense of how the DA is reducing prior uncertainty, for each month, we compute a % uncertainty reduction as follows:

$$\% \text{ Uncertainty Reduction} = 1 - \frac{\sigma_{posterior}}{\sigma_{prior}}. \tag{39}$$

Since we do not precisely know the posterior standard deviation in Equation (39), we consider the reduction both in terms of the raw Monte Carlo point estimate of the posterior standard deviation and its inflated version (i.e., R as defined in Section 2.3).

The left side of Figure 1 shows the timeseries of global mean functionals and their credible intervals. The posterior flux is shown to have reduced error against the true flux, especially during the boreal summer months. Similarly, the Monte Carlo posterior uncertainty estimate shows considerable reduction relative to the prior. The uncertainty estimates with inflated endpoints, increase the posterior uncertainty by 22% while the deflated endpoints decrease the posterior uncertainty by 15%, resulting in credible interval endpoint bounds that capture the true credible interval endpoints with 95% probability. The right side of Figure 1 further emphasizes the prior to posterior uncertainty reduction that we mathematically expect. However, we notice that the inflated uncertainty is only reduced during boreal summer months. In January, February, March, and April the inflated Monte Carlo estimated posterior uncertainty is actually larger than the prior uncertainty. There is a logical explanation for this:
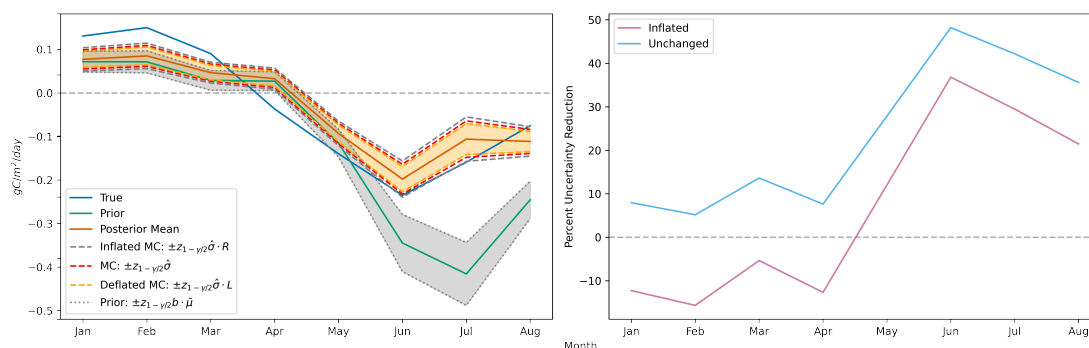
Figure 1. (Left) Estimated posterior $(1 - \gamma) \times 100\% = 95\%$ credible intervals around the monthly global flux functionals show markedly improved uncertainty over the prior during boreal summer months. The three interval types shown are the unchanged MC estimated intervals (red), the inflated MC estimated intervals (gray), and the deflated MC estimated intervals (orange). As described in Section 2.3, for each month, the true upper and lower credible interval endpoints are contained within the inflated and deflated endpoints with probability $1 - \alpha = 0.95$. Note, $\hat{\sigma}$ is a shorthand notation for the empirical functional standard deviation as defined in Step (3b) of Algorithm 1, $\overline{\mu}$ is the globally averaged control flux to which the prior uncertainty, b, is applied for each month, and R and L are the inflation and deflation factors, respectively, as defined in Table 2. We observe that with even as few as $M = 60$ ensemble members, at the monthly/global scale, the magnitude of the Monte Carlo sampling uncertainty is small in comparison to the posterior uncertainty. **(Right)** Percent reduction in uncertainty from prior to posterior for the monthly global fluxes is most significant during the boreal summer. The light blue curve shows the percent reduction estimated with the unchanged MC estimated posterior standard deviation, while the magenta curve shows the percent reduction estimated with the inflated MC estimated posterior standard deviation. The true reduction is larger than the reduction shown by the latter curve with 97.5% confidence.

since most of the landmass generating NEE fluxes is in the Northern Hemisphere and GOSAT requires sunlight to measure
350   X_{CO_2}, the satellite observations impose much weaker constraints on the fluxes during boreal winter. Furthermore, since the prior uncertainty is defined as a percentage, the prior is more concentrated during the boreal winter months when the absolute magnitude of the CarbonTracker fluxes is smaller. As a result of these two effects, the actual posterior uncertainty during the winter months is only slightly smaller than the prior uncertainty. Since we are obtaining a noisy Monte Carlo estimate of this uncertainty from using 60 ensemble members, the inflated value accounting for the Monte Carlo uncertainty of the posterior
355   uncertainty is slightly larger than the prior uncertainty.

4   Conclusions

For Bayesian uncertainty quantification in which the forward model is only available as a simulator, the carbon flux estimation community has proposed a useful Monte Carlo method to compute posterior uncertainties. This method is especially well-

suited to DA tasks since it is parallelizable, works with computationally intensive physical simulators, and allows for flexible

360 post-hoc uncertainty quantification on any desired functional of the model parameters. In this note, we analytically established the mathematical correctness of this procedure in the case of Gaussian prior and error distributions and provided additional uncertainty quantification to account for the Monte Carlo sampling variability in the final estimated credible interval. We also provided two numerical examples. In the first, we demonstrated the agreement between the analytical equations and empirical results for an explicitly known linear forward model. In the second, we showed that this procedure applies to a large-scale DA

365 problem in the form of a carbon flux inversion OSSE, and reasoned that the uncertainty quantification results are mathematically and practically sensible. Future investigations of this method could be based on an exploration of how many ensemble members must be sampled before the Monte Carlo uncertainty is sufficiently less than the posterior uncertainty.

Appendix A: Supporting Algebraic Results

There are a few key properties of the element-wise multiplication operation that must be stated in order to support the derivation

370 of the equations presented in this paper.

For the following, let $\mathbf{x} \in \mathbb{R}^m$ be a random vector such that $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}$ and $\mathrm{Cov}[\mathbf{x}] = \boldsymbol{\Sigma}$. Additionally, suppose $\mathbf{a} \in \mathbb{R}^m$ and $\mathbf{A} \in \mathbb{R}^{n \times m}$.

Lemma 1. $\mathbb{E}[\mathbf{x} \circ \mathbf{a}] = \mathbb{E}[\mathbf{x}] \circ \mathbf{a}$.

Proof. By definition, we have

375 $$\mathbb{E}[\mathbf{x} \circ \mathbf{a}] = [\mathbb{E}[x_i a_i]]_i = [a_i \mathbb{E}[x_i]]_i = \mathbb{E}[\mathbf{x}] \circ \mathbf{a}. \tag{A1}$$

$\square$

Lemma 2. $\mathrm{Cov}[\mathbf{x} \circ \mathbf{a}] = \mathrm{Cov}[\mathbf{x}] \circ \mathbf{a}\mathbf{a}^\top$

Proof. There are two terms that need to be computed: (1) $\mathrm{Var}[x_i a_i]$ and (2) $\mathrm{Cov}[x_i a_i, x_j a_j]$. (1) is straightforward by properties of variance, namely, $\mathrm{Var}[x_i a_i] = a_i^2 \mathrm{Var}[x_i]$. (2) simply requires the definition of covariance, i.e.,

380 $$\mathrm{Cov}[x_i a_i, x_j a_j] = \mathbb{E}\big[\big(x_i a_i - \mathbb{E}[x_i a_i]\big)\big(x_j a_j - \mathbb{E}[x_j a_j]\big)\big] = a_i a_j \mathrm{Cov}[x_i, x_j]. \tag{A2}$$

Hence, it follows that $\mathrm{Cov}[\mathbf{x} \circ \mathbf{a}] = \mathrm{Cov}[\mathbf{x}] \circ \mathbf{a}\mathbf{a}^\top$. $\square$

Lemma 3. *Let* $\boldsymbol{\mu} \in \mathbb{R}^m$ *and let* $\mathbf{A}_{\boldsymbol{\mu}}$ *be such that the ith row of* $\mathbf{A}_{\boldsymbol{\mu}}$ *is equal to* $[A_{ij}\mu_j]_j$. *Then the following equation holds:*

$$\mathbf{A}_{\boldsymbol{\mu}}^\top \mathbf{A}_{\boldsymbol{\mu}} = \mathbf{A}^\top \mathbf{A} \circ \boldsymbol{\mu}\boldsymbol{\mu}^\top. \tag{A3}$$

Proof. To prove Equation (A3), first note that $\boldsymbol{\mu}\boldsymbol{\mu}^\top = [\mu_i \mu_j]_{ij}$. Let $i \in [m]$ and $j \in [n]$. By definition, we have

$$\big[\mathbf{A}_{\boldsymbol{\mu}}^\top \mathbf{A}_{\boldsymbol{\mu}}\big]_{ij} = \sum_{l=1}^m A_{li} A_{lj} \mu_i \mu_j = \mu_i \mu_j \sum_{l=1}^m A_{li} A_{lj}.$$

Hence, we can see that

$$\left[\mathbf{A}_{\boldsymbol{\mu}}^{\top}\mathbf{A}_{\boldsymbol{\mu}}\right]_{ij} = \left[\sum_{l=1}^{m} A_{li}A_{lj}\right]_{ij} \cdot [\mu_i\mu_j]_{ij}$$

and we have the desired result. □

Corollary 4. *Let* $\mathbf{M} \in \mathbb{R}^{n \times n}$ *be a positive-definite matrix and let* $\mathbf{A}_{\boldsymbol{\mu}}$ *be defined as in Lemma 3. It follows that,*

$$\mathbf{A}_{\boldsymbol{\mu}}^{\top}\mathbf{M}\mathbf{A}_{\boldsymbol{\mu}} = \mathbf{A}^{\top}\mathbf{M}\mathbf{A} \circ \boldsymbol{\mu}\boldsymbol{\mu}^{\top}. \tag{A4}$$

Proof. Since $\mathbf{M}$ is positive-definite, it has a lower-triangular Cholesky decomposition, $\mathbf{M} = \mathbf{L}\mathbf{L}^{\top}$. For all $i \in [m]$ and $j \in [n]$, we have the following equivalence:

$$\left[\mathbf{L}^{\top}\mathbf{A}_{\boldsymbol{\mu}}\right]_{ij} = \sum_{l=1}^{n} A_{lj}L_{li}\mu_j = \left[\left(\mathbf{L}^{\top}\mathbf{A}\right)_{\boldsymbol{\mu}}\right]_{ij}. \tag{A5}$$

Therefore, $\mathbf{L}^{\top}\mathbf{A}_{\boldsymbol{\mu}} = \left(\mathbf{L}^{\top}\mathbf{A}\right)_{\boldsymbol{\mu}}$. Thus, Lemma 3 implies,

$$\mathbf{A}_{\boldsymbol{\mu}}^{\top}\mathbf{M}\mathbf{A}_{\boldsymbol{\mu}} = \left(\mathbf{L}^{\top}\mathbf{A}_{\boldsymbol{\mu}}\right)^{\top}\left(\mathbf{L}^{\top}\mathbf{A}_{\boldsymbol{\mu}}\right) = \left(\mathbf{L}^{\top}\mathbf{A}\right)_{\boldsymbol{\mu}}^{\top}\left(\mathbf{L}^{\top}\mathbf{A}\right)_{\boldsymbol{\mu}} = \left(\mathbf{L}^{\top}\mathbf{A}\right)^{\top}\left(\mathbf{L}^{\top}\mathbf{A}\right) \circ \boldsymbol{\mu}\boldsymbol{\mu}^{\top}, \tag{A6}$$

and the result follows since $\left(\mathbf{L}^{\top}\mathbf{A}\right)^{\top}\left(\mathbf{L}^{\top}\mathbf{A}\right) = \mathbf{A}^{\top}\mathbf{M}\mathbf{A}$. □

Lemma 5. *Let* $\mathbf{A}_{\boldsymbol{\mu}} \in \mathbb{R}^{n \times m}$ *be defined as in the above,* $\mathbf{M} \in \mathbb{R}^{n \times n}$ *and* $\mathbf{y} \in \mathbb{R}^n$. *Then,*

$$\mathbf{A}_{\boldsymbol{\mu}}^{\top}\mathbf{M}\mathbf{y} = \mathbf{A}^{\top}\mathbf{M}\mathbf{y} \circ \boldsymbol{\mu}. \tag{A7}$$

Proof. It is sufficient to show the case when $\mathbf{M} = \mathbf{I}_n$, as otherwise we can simply define a new vector $\tilde{\mathbf{y}} = \mathbf{M}\mathbf{y}$. By matrix multiplication, for all $j \in [m]$,

$$\left[\mathbf{A}_{\boldsymbol{\mu}}^{\top}\mathbf{y}\right]_j = \sum_{i=1}^{n} A_{ij}y_i\mu_j = \left[\mathbf{A}^{\top}\mathbf{y} \circ \boldsymbol{\mu}\right]_j. \tag{A8}$$

□

Lemma 6. *Let* $\mathbf{A}_{\boldsymbol{\mu}} \in \mathbb{R}^{n \times m}$ *be defined as in the above. Then,*

$$\mathbf{A}(\mathbf{a} \circ \boldsymbol{\mu}) = \mathbf{A}_{\boldsymbol{\mu}}\mathbf{a}. \tag{A9}$$

Proof. This property follows simply from the definition of $\mathbf{A}_{\boldsymbol{\mu}}$ and matrix multiplication. □

Corollary 7. *Define* $\mathbf{A} \in \mathbb{R}^{n \times m}$ *as above, let* $\alpha, \beta \in \mathbb{R}$, *and let* $\mathbf{a}, \mathbf{b}, \boldsymbol{\mu} \in \mathbb{R}^m$. *Then* $\mathbf{A}(\mathbf{a} \circ \boldsymbol{\mu})$ *is linear in* $\mathbf{a}$.

Proof. By Lemma 6, we have,

$$\mathbf{A}\left((\alpha\mathbf{a} + \beta\mathbf{b}) \circ \boldsymbol{\mu}\right) = \mathbf{A}_{\boldsymbol{\mu}}(\alpha\mathbf{a} + \beta\mathbf{b}) = \alpha\mathbf{A}(\mathbf{a} \circ \boldsymbol{\mu}) + \beta\mathbf{A}(\mathbf{b} \circ \boldsymbol{\mu}). \tag{A10}$$

□

Author contributions. MS and MK derived all mathematical and statistical results. BB and JL provided scientific expertise. BB provided the OSSE used for the numerical experiment. MS prepared the manuscript and ran all experiments with contributions from all co-authors. All authors participated in reviewing and editing the manuscript.

Competing interests. The authors declare that they have no competing interests.

410 *Acknowledgements.* This work was supported by NSF grant DMS-2053804, JPL RSA No. 1670375 and 1689177 and a grant from the C3.AI Digital Transformation Institute. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (grant no. 80NM0018D004). Liu and Byrne would like to acknowledge the funding support from NASA Orbiting Carbon Observatory Science Team program (17-OCO2-17-0013). Finally, we would like to thank Anna Harper for providing the JULES fluxes used in this study's OSSE, the STAMPS research group at Carnegie Mellon University for
415 supporting this work, and the Uncertainty Quantification group at the Jet Propulsion Laboratory for facilitating this collaboration. We would like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation.

References

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

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C.: A limited memory algorithm for bound constrained optimization, SIAM Journal on Scientific
     Computing, 16, 1190 – 1208, 1995.

Byrne, B., Jones, D., Strong, K., et al.: On what scales can GOSAT flux inversions constrain anomalies in terrestrial ecosystems?, Atmo-
     spheric Chemistry and Physics, 19, 13 017 – 13 035, 2019.

425   Byrne, B., Baker, D. F., Basu, S., Bertolacci, M., Bowman, K. W., Carroll, D., Chatterjee, A., Chevallier, F., Ciais, P., Cressie, N., Crisp, D.,
     Crowell, S., Deng, F., Deng, Z., Deutscher, N. M., Dubey, M. K., Feng, S., García, O. E., Griffith, D. W. T., Herkommer, B., Hu, L., Jacob-
     son, A. R., Janardanan, R., Jeong, S., Johnson, M. S., Jones, D. B. A., Kivi, R., Liu, J., Liu, Z., Maksyutov, S., Miller, J. B., Miller, S. M.,
     Morino, I., Notholt, J., Oda, T., O'Dell, C. W., Oh, Y.-S., Ohyama, H., Patra, P. K., Peiro, H., Petri, C., Philip, S., Pollard, D. F., Poulter,
     B., Remaud, M., Schuh, A., Sha, M. K., Shiomi, K., Strong, K., Sweeney, C., Té, Y., Tian, H., Velazco, V. A., Vrekoussis, M., Warneke,
430   T., Worden, J. R., Wunch, D., Yao, Y., Yun, J., Zammit-Mangion, A., and Zeng, N.: National CO_2 budgets (2015–2020) inferred from
     atmospheric CO_2 observations in support of the global stocktake, Earth System Science Data, 15, 963–1004, https://doi.org/10.5194/essd-
     15-963-2023, 2023.

Casella, G. and Berger, R. L.: Statistical Inference, Thomson Learning Inc, 2 edn., 2002.

Chevallier, F., Breon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO_2 sources and sinks:
435   Theoretical study in a variational data assimilation framework, Journal of Geophysical Research, 112, 2007.

Clark, D., Mercado, L., Sitch, S., Jones, C., Gedney, N., Best, M., Pryor, M., Rooney, G., Essery, R., Blyth, E., et al.: The Joint UK Land
     Environment Simulator (JULES), model description–Part 2: carbon fluxes and vegetation dynamics, Geoscientific Model Development,
     4, 701–722, 2011.

Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., Liu, J., Deng, F., Feng, L., McKain, K., Chatterjee, A., Miller,
440   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.

Deng, F., Jones, D., Henze, D., et al.: Inferring regional sources and sinks of atmospheric CO_2 from GOSAT X_{CO_2} data, Atmospheric
     Chemistry and Physics, 14, 3703 – 3727, 2014.

445   Enting, I., Trudinger, C., and Francey, R.: A synthesis inversion of the concentration and δ^{13} C of atmospheric CO_2, Tellus B, 47, 35–52,
     1995.

Flath, H. P., Wilcox, L., Akçelik, V., Hill, J., Waanders, B. V. B., and Ghattas, O.: Fast algorithms for Bayesian uncertainty quantification in
     large-scale linear inverse problems based on low-rank partial Hessian approximations, SIAM Journal on Scientific Computing, 33, 407 –
     432, 2011.

450   Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., et al.: Towards
     robust regional estimates of CO_2 sources and sinks using atmospheric transport models, Nature, 415, 626–630, 2002.

Halko, N., Martinsson, P., and Tropp, J.: Finding structure with Randomness: Probabilistic algorithms for constructing approximate matrix
     decompositions, SIAM Review, 53, 217 – 288, 2011.

Harper, A. B., Wiltshire, A. J., Cox, P. M., Friedlingstein, P., Jones, C. D., Mercado, L. M., Sitch, S., Williams, K., and Duran-Rojas,
455   C.: Vegetation distribution and terrestrial carbon cycle in a carbon cycle configuration of JULES4.6 with new plant functional types,
Geoscientific Model Development, 11, 2857–2873, 2018.

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.

Houweling, S., Baker, D., Basu, S., Boesch, H., Butz, A., Chevallier, F., Deng, F., Dlugokencky, E., Feng, L., Ganshin, A., et al.: An inter-
460   comparison of inverse models for estimating sources and sinks of CO_2 using GOSAT measurements, Journal of Geophysical Research:
Atmospheres, 120, 5253–5266, 2015.

Kalmikov, A. G. and Heimbach, P.: A Hessian-based method for uncertainty quantification in global ocean state estimation, SIAM Journal
on Scientific Computing, 36, S267 – S295, 2014.

Kalnay, E.: Atmospheric modeling, data assimilation and predictability, Cambridge University Press, 2003.

465 Lanczos, C.: An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, Journal of Research
of the National Bureau of Standards, 45, 255 – 282, 1950.

Liu, J., Bowman, K., Meemong, L., et al.: Carbon Monitoring system flux estimation and attribution: impact of ACOS-GOSAT
X_{CO_2} sampling on the inference of terrestrial biospheric sources and sinks, Tellus B: Chemical and Physical Meteorology, 66,
https://doi.org/https://doi.org/10.3402/tellusb.v66.22486, 2016.

470 O'Dell, C., Connor, B., Bosch, H., et al.: The ACOS CO_2 retrieval algorithm - Part 1: Description and validation against synthetic observa-
tions, Atmospheric Measurement Techniques, 5, 99 – 121, 2012.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M., Pétron, G., Hirsch, A. I.,
et al.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, Proceedings of the National Academy
of Sciences, 104, 18 925–18 930, 2007.

475 Petra, N., Martin, J., Stadler, G., and Ghattas, O.: A computational framework for infinite-dimensional Bayesian inverse problems, part
II: stochastic newton MCMC with application to ice sheet flow inverse problems, SIAM Journal on Scientific Computing, 36, A1525 –
A1555, 2014.

Rogers, C. D.: Inverse Methods for Atmospheric Sounding, World Scientific Publishing, 2000.

Vasco, D., Pulliam, R. J., and Johnson, L. R.: Formal inversion of ISC arrival times for mantle P-velocity structure, Geophysical Journal
480   International, 113, 586 – 606, 1993.

Wainwright, M. J.: High-Dimensional Statistics, Cambridge University Press, 2019.

Zammit-Mangion, A., Bertolacci, M., Fisher, J., Stavert, A., Rigby, M., Cao, Y., and Cressie, N.: WOMBAT v1.0: a fully Bayesian global
flux-inversion framework, Geoscientific Model Development, 15, 45 – 73, 2022.