Geoscientific
Model Development

Open Access

Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions

S. M. Miller[1], A. M. Michalak[2], and P. J. Levi[1]

[1]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA
[2]Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA

Correspondence to: S. M. Miller (scot.m.miller@gmail.com)

Abstract. Many inverse problems in the atmospheric sciences involve parameters with known physical constraints. Examples include nonnegativity (e.g., emissions of some urban air pollutants) or upward limits implied by reaction or solubility constants. However, probabilistic inverse modeling approaches based on Gaussian assumptions cannot incorporate such bounds and thus often produce unrealistic results. The atmospheric literature lacks consensus on the best means to overcome this problem, and existing atmospheric studies rely on a limited number of the possible methods with little examination of the relative merits of each.

This paper investigates the applicability of several approaches to bounded inverse problems. A common method of data transformations is found to unrealistically skew estimates for the examined example application. The method of Lagrange multipliers and two Markov chain Monte Carlo (MCMC) methods yield more realistic and accurate results. In general, the examined MCMC approaches produce the most realistic result but can require substantial computational time. Lagrange multipliers offer an appealing option for large, computationally intensive problems when exact uncertainty bounds are less central to the analysis. A synthetic data inversion of US anthropogenic methane emissions illustrates the strengths and weaknesses of each approach.

1 Introduction

Inverse modeling and data assimilation have become ubiquitous in the atmospheric sciences, and one of the most common applications is the estimation of trace gas surface fluxes. These top-down approaches optimize emissions or

flux estimates such that modeled atmospheric concentrations reproduce observed concentrations. Most methods are based on Bayesian statistical principles and assumptions of Gaussian probability density functions (pdfs), implemented in a variety of ways (e.g., Gurney et al., 2002; Michalak et al., 2004; Henze et al., 2007; Peters et al., 2007; Gourdji et al., 2008; Stohl et al., 2012).

Many applications require estimating emissions or fluxes that have known physical limits, often referred to simply as inequality constraints. For example, there are few anthropogenic sinks of carbon dioxide or methane, and the release history of air toxins from an industrial hazard site is never negative. In many cases, predicted sources that violate inequality constraints are not only meaningless but distort prediction in surrounding regions or times. For example, if an inversion estimates an unrealistic negative emissions region, emissions in adjacent regions may become larger than expected due to mass conservation (e.g., Michalak, 2008). Hence, it would not be sufficient to simply reset negative emissions to zero. Doing so would not correct for distorted sources elsewhere and would erroneously increase the overall estimated emissions budget (i.e., would violate the mass balance or budget as constrained by the atmospheric observations).

Additionally, enforcing inequality constraints is often necessary for obtaining realistic uncertainty estimates. Even if the posterior emissions themselves do not violate the inequality constraints, their confidence intervals could very well extend beyond known limits under Gaussian assumptions. In such cases, an unconstrained inversion will produce both upper and lower confidence intervals that are unrealistically large (e.g., Snodgrass and Kitanidis, 1997;

Michalak and Kitanidis, 2003). The problem occurs because unrealistically low emissions within the lower confidence interval must be balanced by larger emissions elsewhere in the upper confidence interval, or vice versa.

In response to the problems associated with unconstrained inversions, existing trace gas flux estimation studies typically use one of four methods to apply inequality constraints. One method employed in previous studies is a data transformation (refer to Sect. 3.1, e.g., Muller and Stavrakou, 2005; Bergamaschi et al., 2009). A second method decreases the uncertainty assigned to many of the prior fluxes until the posterior fluxes obey the known bounds (e.g., Eckhardt et al., 2008; Stohl et al., 2012). This adjustment may run counter to the modeler's physical understanding of the prior estimate or associated uncertainties and therefore is not discussed in great detail here. A third method is that of Lagrange multipliers (refer to Sect. 3.2, e.g., Henze et al., 2007; Kopacz et al., 2009; Göckede et al., 2010). Finally, two recent studies use a class of Markov chain Monte Carlo (MCMC) methods known as Metropolis–Hastings (Rigby et al., 2011; Burrows et al., 2013), implemented in a manner that enforces nonnegativity. Existing atmospheric studies provide little guidance on the merits of one approach over another.

The objective of this study is thus to investigate the merits of the above approaches and two additional MCMC implementations. MCMC algorithms are ubiquitous in Bayesian statistics but are not commonly applied to atmospheric studies. The remainder of this paper is organized as follows: Sect. 2 examines the statistical assumptions of common inversion methods that are incompatible with inequality constraints. Section 3 discusses several possible alternatives to mitigate these statistical assumptions, including data transformations, Lagrange multipliers, and two specific MCMC implementations – a multiple-try Metropolis–Hastings algorithm and a Gibbs sampler. Finally, Sects. 4 and 5 discuss the costs and benefits of each approach in the context of a synthetic case study estimating North American anthropogenic methane emissions.

2   Common Bayesian approaches to inverse modeling

This section describes common approaches to inverse modeling and indicates which statistical assumptions are incompatible with known bounds.

In a typical inverse problem, the unknown quantity to be estimated (s, dimensions $m \times 1$) is different from the quantity actually observed (z, dimensions $n \times 1$), and the two are related to one another by a function $h(s)$. In the case of trace gas inversions, s are the true, unknown emissions or fluxes; z are observations of atmospheric concentration; and the function h is often defined by an atmospheric transport and/or chemistry model:

$$z \sim h(s) + \mathcal{N}(\mathbf{0}, \mathbf{R}), \tag{1}$$

where $\mathcal{N}(\mathbf{0}, \mathbf{R})$, in this case, represents the combined model, measurement, representation, and spatial/temporal aggregation errors, collectively termed model–data mismatch. These errors are most commonly assumed to be random and normally distributed with a mean of zero and an $n \times n$ covariance matrix $\mathbf{R}$.

Any a priori information on the spatial or temporal distribution of s can be incorporated into a model of the mean, $E[s]$:

$$s \sim E[s] + \mathcal{N}(\mathbf{0}, \mathbf{Q}). \tag{2}$$

This model, $E[s]$, rarely matches the unknown s, and the $m \times m$ covariance matrix $\mathbf{Q}$ describes the magnitude and structure of the residuals between s and $E[s]$. As with the model–data mismatch, these residuals are also typically assumed to be normally distributed with a mean of zero.

The model of the mean can be formulated in a number of ways, but one common method, used in this paper, is as follows:

$$E[s] = \mathbf{X}\boldsymbol{\beta}, \tag{3}$$

where the $m \times p$ matrix $\mathbf{X}$ includes p different covariates, and the unknown $p \times 1$ drift coefficients ($\boldsymbol{\beta}$) adjust the magnitude of these covariates to best match the observations. The model of the mean could be uninformative (e.g., $\mathbf{X}$ is an $m \times 1$ vector of ones as in Mueller et al., 2008) or could include any number of covariates, including climatological information or an existing emissions inventory (e.g., Gourdji et al., 2012; Miller et al., 2013, 2014). Some inversion approaches assume that the drift coefficients are known, in which case $E[s]$ becomes an $m \times 1$ vector (e.g., Rodgers, 2000; Enting, 2002; Tarantola, 2005). An inversion with unknown coefficients has typically been used within the context of a "geostatistical" representation of the inverse problem (used in this study), while the coefficients are usually assumed in the "synthesis Bayesian" approach, though both approaches are Bayesian in nature.

Equation (1) can be expanded using the formulation of s described in Eq. (3):

$$z \sim \mathbf{H}\,\mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{Q}) + \mathcal{N}(\mathbf{0}, \mathbf{R}), \tag{4}$$

where $\mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{Q})$ represents the distribution of s relative to the prior or model of the mean ($\mathbf{X}\boldsymbol{\beta}$). The $n \times m$ sensitivity matrix, $\mathbf{H}$, is a linearized form of h. This setup assumes, as in most existing studies, that the measurement residuals ($z - \mathbf{H}s$) and flux residuals ($s - \mathbf{X}\boldsymbol{\beta}$) follow a multivariate normal distribution, as will the posterior probabilities of s and $\boldsymbol{\beta}$.

The optimal estimate of unknown s can be obtained by minimizing the sum of squared residuals subject to the covariances:

$$L_{s,\boldsymbol{\beta}} = \frac{1}{2}(z - \mathbf{H}s)^T \mathbf{R}^{-1}(z - \mathbf{H}s)$$
$$+ \frac{1}{2}(s - \mathbf{X}\boldsymbol{\beta})^T \mathbf{Q}^{-1}(s - \mathbf{X}\boldsymbol{\beta}). \tag{5}$$

If $\mathbf{H}$ does not depend on the unknown value of s, then s can typically be estimated by solving a system of linear equations (refer to Michalak et al. (2004) and Tarantola (2005) for more in-depth discussion on estimating s and the associated posterior uncertainties). Otherwise, the algorithm is usually iterative.

If s has known bounds, then Eq. (4) must be reformulated in a way that honors the inequality constraints. Some deterministic methods permanently remove elements of s from the optimization if they violate the bounds (e.g., Lagrange multipliers). In a purely stochastic approach, the first term in Eq. (4) will instead follow some multivariate distribution (f_l^u) that is restricted to within the lower and upper constraints (l and u, respectively):

$$z \sim \mathbf{H} \, f_l^u(s|\mathbf{X}, \mathbf{Q}) + \mathcal{N}(\mathbf{0}, \mathbf{R}). \tag{6}$$

This alteration modifies the prior probability of the fluxes; the fluxes (s) will deviate from the prior or expected value ($E[s]$) only to the extent that the fluxes remain within the bounds l and u. In contrast, the distribution of the model-measurement residuals, $\mathcal{N}(\mathbf{0}, \mathbf{R})$, remains the same as in the case without inequality constraints. In Eq. (6), f_l^u could be formulated as a multivariate truncated normal, exponential, or gamma distribution, among many other choices. Most formulations of f_l^u result in a cost function that does not have a straightforward analytical minimum, unlike the multivariate normal case in Eq. (5). Instead, the unknown quantity (s) must be estimated using an algorithm that samples the posterior probability density. Even then, it can be difficult to sample this density efficiently. The next section describes example deterministic and stochastic approaches in greater detail.

3 Strategies for enforcing inequality constraints

3.1 Data transformations

Data transformations can enforce inequality constraints with relatively easy implementation, but transformations typically render a linear inverse problem nonlinear and therefore require an iterative solution. A number of different data transformations exist, but the power transformation is a common approach because it is defined at zero (unlike log transformations; e.g., Snodgrass and Kitanidis, 1997; Wilks, 2011, Chap. 3):

$$\tilde{s} = \alpha \left(s^{1/\alpha} - 1 \right), \tag{7}$$

where s are the fluxes in normal space and α can be any scalar value such that $\tilde{s} > -\alpha$, though larger values of α cause more extreme transformations. This formulation approaches the natural logarithm for large values of α.

For the power transform, the Jacobian or sensitivity matrix ($\mathbf{H}$) is not linear in the transformed space. The algorithm, as a result, becomes iterative and requires updating $\mathbf{H}$ at every iteration until both $\mathbf{H}$ and the best estimate in the transform space ($\tilde{s}$) converge (described in detail by Snodgrass and Kitanidis (1997) and Fienen et al. (2004), among others). Most transformations assume a skewed pdf and therefore lead to skewed posterior uncertainty estimates, and such asymmetry can have a number of implications as discussed in Sect. 5. Furthermore, most common transformations can only enforce a single upper or lower bound that is the same for all elements of s.

3.2 Lagrange multipliers and the trust region algorithm

The method of Lagrange multipliers is commonly used in deterministic optimization problems to enforce equality or inequality constraints. The approach has also been adapted to a number of stochastic inverse problems in hydrology (Barnes and You, 1992; Walvoort and de Gruijter, 2001; Michalak and Kitanidis, 2004) and more recently in an atmospheric context (Henze et al., 2007; Kopacz et al., 2009; Göckede et al., 2010). The method of Lagrange multipliers can be applied to an inversion by modifying the original cost function $L_{s,\beta}$:

$$L_{s,\beta,\lambda} = L_{s,\beta} - \lambda^T (s - l), \tag{8}$$

where l in this case is a lower bound on s, where the bound can be spatially and temporally variable, and λ are the unknown Lagrange multipliers.

A number of implementations exist, but all methods share many similarities. Any element of s that would otherwise violate the inequality constraints becomes fixed on one of the bounds. Most algorithms are iterative and add or remove these elements from the "active" set at each iteration. The optimization proceeds only on the active set and ignores all other elements that have been fixed (e.g., Gill et al., 1981). A large difference among algorithms is the way in which elements are removed or added to the active set.

One result of this setup is that elements in the fixed set are not modeled as continuous random variables. Estimated emissions in these regions have no associated posterior uncertainty. In other words, Lagrange multipliers compromise the stochastic nature of the inverse problem in order to enforce the desired constraints.

Several numerical methods are available for solving constrained optimization problems via the method of Lagrange multipliers, but many are restricted to small or medium-sized problems (e.g., s has fewer than 1000 elements). These include the method of Theil and Van de Panne (Theil and Panne, 1960; Snyman, 2005, Chap. 3.4) and the active set method (e.g., Gill et al., 1981, or Antoniou and Lu, 2007, Chap. 13.3).

A number of algorithms can efficiently solve large, bounded optimization problems, including the trust region method and the bounded, limited-memory Broyden–Fletcher–Goldfarb–Shanno approach (L-BFGS-B; e.g., Byrd et al., 1995). This paper implements a trust region algorithm. The method can efficiently handle large, bounded optimization problems because it adds and/or subtracts multiple elements from the fixed set at each iteration (e.g., More, 1988; Lin and More, 1999). A trust region algorithm approximates the objective function at each iteration and estimates the range over which this approximation can be trusted (referred to as the trust region). The algorithm optimizes s within the trust region and compares the approximated improvement to the actual reduction in the cost function (in this case Eq. 5). If the cost function approximation performs well, the algorithm is allowed to make more aggressive moves at each iteration. In other words, the algorithm may increase the size of the trust region if the approximation does well and vice versa. Though it was originally developed for unconstrained problems, Gay (1984) extended the trust region method to constrained optimization. For additional discussion of general trust region algorithms, see Sorensen (1982), Lin and More (1999), Conn et al. (2000, Chap. 1, Chap. 6), or Yuan (2000).

This paper adopts a general algorithm outlined in Lin and More (1999). The reader is referred to a review article (Yuan, 2000) for a broad discussion of possible trust region implementations. Most require the gradient (∇L) and Hessian ($\nabla^2 L$) of the cost function. For reference, these equations are listed below for the geostatistical approach:

$$
\begin{aligned}
\nabla L_{s,\beta} &= -\frac{1}{2}\mathbf{H}^T \mathbf{R}^{-1}(z - \mathbf{H}s) + \frac{1}{2}\mathbf{G}s \\
\nabla^2 L_{s,\beta} &= \frac{1}{2}\mathbf{G} + \frac{1}{2}\mathbf{H}^T \mathbf{R}^{-1}\mathbf{H} \\
\mathbf{G} &= \mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Q}^{-1}.
\end{aligned}
\tag{9}
$$

To construct these equations, we take the derivative of the original cost function (Eq. 5) with respect to β and set this derivative equal to zero (Kitanidis, 1986). We then rearrange the cost function to omit the unknown drift coefficients β. The resulting Hessian and gradient above are written only in terms of the unknown vector s. Rodgers (2000) presents analogous equations for a prior model setup that has predetermined coefficients (β).

3.3 MCMC algorithms applied to bounded inversions

The following sections discuss two possible MCMC implementations for inequality-constrained problems. In general, MCMC algorithms make it possible to generate realizations of the unknown quantity from high-dimensional probability density functions. These algorithms make problems with non-Gaussian distributions and/or complex joint pdfs tractable (e.g., Andrieu et al., 2003).

MCMC algorithms simulate a Markov chain with an equilibrium distribution that matches the distributions of the quantities being estimated. The methods rely on the generation of conditional realizations; each realization is a guess of the unknown (e.g., s) that should represent a random draw from the posterior probability distribution. The algorithms create a new realization based only upon the previous one, and the means of doing so differentiate the various MCMC methods. Many conditional realizations are typically generated to adequately sample or represent the equilibrium distribution (Geyer, 2011). The point-wise properties of the equilibrium probability density (e.g., mean, median, percentiles, standard deviation) can be used to represent the statistics of the unknown state, including its uncertainties. A thorough introduction to MCMC approaches is given by Geyer (2011).

MCMC methods can also be used for the solution of bounded problems. Each individual realization of the unknown quantity is restricted by the inequality constraints (Gelfand et al., 1992), ensuring that both the posterior best estimate and associated uncertainties will honor known physical limits.

3.3.1 Metropolis–Hastings

Metropolis–Hastings algorithms have become widespread in Bayesian statistics (see Chib and Greenberg, 1995; Bolstad, 2012, for in-depth discussion). The modeler uses an existing, accepted realization of the unknown quantity (in this case s) to generate a new proposed realization with a Markov chain whose properties are defined by the modeler. One possible approach might generate many realizations of s by using slightly modified inputs for Eq. (5). Instead of using z in Eq. (5), sample randomly from $\mathcal{N}(z, \mathbf{R})$. Instead of using $E[s]$, use random, sequentially correlated samples from $\mathcal{N}(E[s], \mathbf{Q})$ (see the Supplement for further discussion).

Each subsequent proposed realization is accepted or discarded based on its likelihood relative to the previous, accepted realization. Realizations with a relative likelihood greater than one are always accepted, while those with a relatively likelihood less than one are only sometimes accepted. A large number of realizations is sequentially generated in this way to sample across the probability space of the unknown (in this case the posterior probability distribution of s).

The modeler must carefully balance two considerations when proposing new realizations. If each proposed realization is too close to the previous accepted realization, the algorithm will sample the probability space very slowly. However, if the proposed realization is too far from the previous accepted realization, it will likely have a low relative likelihood and be rejected (e.g., Chib and Greenberg, 1995).

A number of studies have implemented the general algorithm with an adaptation for inequality constraints, both in hydrology (e.g., Michalak and Kitanidis, 2004; Wang and Zabaras, 2006; Zanini and Kitanidis, 2009) and more

recently in the atmospheric sciences (e.g., Rigby et al., 2011; Burrows et al., 2013).

The cited hydrology studies use an implementation suitable for large problems (refer to the Supplement). For the specific implementation in these studies, each proposed realization is first constrained to be nonnegative with Lagrange multipliers before being tested for acceptance. This implementation is a compromise between a purely stochastic approach that would represent all elements as continuous random variables and the method of Lagrange multipliers that completely removes some elements from the optimization.

Though ideal for small problems, Metropolis–Hastings algorithms can often become stuck in local regions of high probability when there are many quantities being estimated (i.e., when m is large). The acceptance rate can become so small as to make implementation impractical (Liu et al., 2000). This study implements a multiple-try Metropolis–Hastings algorithm (Liu et al., 2000) suitable for larger-scale inverse problems, described fully in the Supplement.

3.3.2 Gibbs sampler

Unlike the Metropolis–Hastings algorithm, the Gibbs sampler calculates a new realization for each element of the unknown state sequentially (in this case, each of m elements in s). This method involves calculating a probability distribution for an individual element of s conditional on the current realization for all other elements. The algorithm takes a random sample from the element-wise conditional probability density, and this sample becomes the new guess for the given element. Using this method, the Gibbs sampler sequentially calculates a conditional distribution and random sample for each of m elements in s until an entire new, full conditional realization has been formed (see the Supplement). Like the Metropolis–Hastings algorithm, the Gibbs sampler requires generating a large number of conditional realizations, and the statistics of these realizations can be used to define a best estimate and associated uncertainties. For an in-depth review of the Gibbs sampler, refer to Casella and George (1992) or Bolstad (2012, Chap. 10).

Several studies in hydrology apply Gibbs sampler methods to constrained inverse problems (e.g., Michalak and Kitanidis, 2003; Wang and Zabaras, 2005; Fienen et al., 2006; Michalak, 2008). Michalak (2008) describes a flexible implementation in the context of groundwater problems that can incorporate any kind of spatial or temporal correlation in the a priori covariance matrix $\mathbf{Q}$, and this implementation is adapted for the case study here. The implementation uses a multidimensional truncated normal as the prior pdf (Michalak, 2008). However, because the Gibbs sampler uses element-wise pdfs, it avoids the mathematical challenge of explicitly maximizing a high-dimensional truncated pdf. The approach thereby enforces the inequality constraints in a computationally tractable manner.

The implementation in this study differs from Michalak (2008) in one important way. Some regions of the United States and Canada have zero anthropogenic methane emissions, and we alter the shape of the pdfs to allow a high probability at zero. The implementation here draws a random sample from a Gaussian conditional distribution for sequential elements in s. If the sample is positive, it becomes the new realization for that element of s. If the sample is negative, we use zero as the realization for that element. The approach is equivalent to modeling the prior, and subsequently the posterior, distributions as truncated Gaussian with an added Dirac delta (a function that is zero at every point except zero). This modification relative to Michalak (2008) results in a peak in the posterior densities at zero.

4 Methane case study setup

A synthetic case study of estimating US anthropogenic methane emissions illustrates the comparative drawbacks and benefits of the approaches described above: the power transformation, Lagrange multipliers, and two MCMC implementations, with an unconstrained inversion for comparison. The synthetic study setup uses an existing methane emissions inventory and a model of atmospheric transport to create an estimation problem with known true emissions. The prescribed methane emissions are always nonnegative, so the constraints on this inversion are simple; the estimated emissions must also be nonnegative ($l = 0$). The remainder of this section describes the case study setup in detail.

4.1 Model and synthetic data setup

This study employs a regional-scale, particle-following model known as STILT, the Stochastic Time-Inverted Lagrangian Transport model (Lin et al., 2003), to quantify the sensitivity of atmospheric observations to surface sources, and thereby to estimate the sensitivity matrix $\mathbf{H}$. STILT simulations are driven by Weather Research and Forecasting (WRF) wind fields, version 2.2 (Skamarock et al., 2005). In other words, the combination of WRF and STILT serves as the forward model for the methane case study. Nehrkorn et al. (2010) provide a detailed description of the WRF fields used here.

We use WRF-STILT to generate synthetic methane mixing ratios in the same locations as aircraft and tall tower observation sites in the United States (4600 total observations). The tower sites are those in the NOAA Earth Systems Research Laboratory and DOE monitoring network and are displayed in Fig. 1. Aircraft data include methane measurements from the NOAA Earth Systems Research Laboratory aircraft program at a variety of locations over North America, DOE flights over the US southern Great Plains (Biraud et al., 2013), and observations from the START08 measurement campaign (Pan et al., 2010). This study includes only

Table 1. Covariance matrix parameters.

Parameter	Value
$\sigma_{R,\text{tower}}$ (ppb)[a]	13.1–68.9
$\sigma_{R,\text{aircraft}}$	19.8
Untransformed space	
σ_Q (μmol m^{-2}s^{-1})	0.017
l (km)	101
Transformed space ($\alpha = 6$)[b]	
σ_Q	0.81
l	261

[a] Each tower has a different estimated σ_R.
Refer to Miller et al. (2013).
[b] In the case of the power transform, we set
$\alpha = 6$. This value brings the posterior
emissions estimate closest to being normally
distributed in the transformed space.

aircraft measurements within 2500 m of the ground – measurements that are consistently sensitive to surface fluxes (see Fig. 1, Miller et al., 2013).

The study generates synthetic methane measurements using the EDGAR v3.2 FT2000 anthropogenic inventory (Olivier and Peters, 2005). Newer EDGAR inventories are available (e.g., EDGAR v4.2), but top-down studies suggest that version 3.2 best captures the magnitude of anthropogenic sources over the United States (Kort et al., 2008; Miller et al., 2013, see Fig. 1).

We add noise to each synthetic measurement, randomly sampled from the model–data mismatch covariance matrix (**R** with diagonal elements σ_R^2). The companion study Miller et al. (2013) statistically infers this information from in situ methane measurements using restricted maximum likelihood estimation (REML) (e.g., Kitanidis and Lane, 1985; Michalak et al., 2004). Table 1 summarizes the model–data mismatch values inferred for the towers and aircraft.

4.2 The inversion setup

The inversion covers much of North America (25–55° N latitude, 145–51° W longitude) on a 1° × 1° spatial resolution over the months May–September 2008. Anthropogenic methane sources do not change markedly from one season to another (Miller et al., 2013). Therefore, the synthetic data study here estimates a single set of emissions over the entire five-month period.

All inversions presented here utilize an uninformative prior (e.g., Michalak et al., 2004; Mueller et al., 2008). In other words, the inversion prior is a single unknown constant across the entire geographic inversion domain. This method makes as few a priori assumptions as possible and relies on the atmospheric data to the fullest extent to infer information about the emissions. This framework is particularly well

suited to a synthetic data study; any a priori inventory would be arbitrary since the true emissions are already known.

Despite the lack of information in the prior itself, the inversion incorporates important structural information about the fluxes in the a priori covariance matrix (**Q**). Specifically, the diagonal elements of **Q** describe the total variability of the fluxes (σ_Q^2 – the variance over long spatial scales), and the off-diagonal elements describe the degree of spatial correlation in the posterior flux field, assuming an exponential covariance function. The spatial characteristics of the known emissions field are listed in Table 1 and are used to construct **Q** (σ_Q^2 and l, the decorrelation length parameter). The parameters for the untransformed space are used in the unconstrained, Metropolis–Hastings, and Gibbs sampler inversions.

5 Results and discussion

The inversion implementations discussed in this study produce variable results. All methods place large methane emissions in Kentucky, West Virginia, and along the eastern seaboard, similar to the true synthetic fluxes (see Fig. 2), but the methods differ in many other regards. The remainder of this section highlights these differences to illustrate the relative merits of each approach.

5.1 Unconstrained inversion

The unconstrained case causes several undesirable side effects, including but not limited to negative emissions estimates (Fig. 2). It may appear counterintuitive that the emissions estimate ($\hat{s}$) could have negative components when the observations (z) and model transport (**H**) contain only positive elements. However, these negative emissions are not necessarily caused by any violation of the statistical assumptions in the inversion. Rather, the estimate ($\hat{s}$) may contain negative elements due to the effect of model–data mismatch errors. When these errors are present, the gradients in the observations can be consistent with adjacent positive and negative sources. In the methane case study, we synthetically generate model–data mismatch errors, so these errors are guaranteed to obey all assumptions of the statistical model. Negative emissions, in this case, could not be caused by biases in wind fields, chemistry, or by a biased prior. In this study, the true winds are known, chemistry is absent, and the geostatistical inverse model does not assume any prior value for the emissions.

Aside from unrealistic negative emissions, the uncertainties are also too large. Conditional realizations and confidence intervals based on multi-Gaussian probabilities extend well beyond the known bounds on the problem (i.e., are not strictly nonnegative), even in regions where the best estimate itself falls within these bounds (Fig. 3). Figure 4 visualizes this problem in terms of the marginal probability
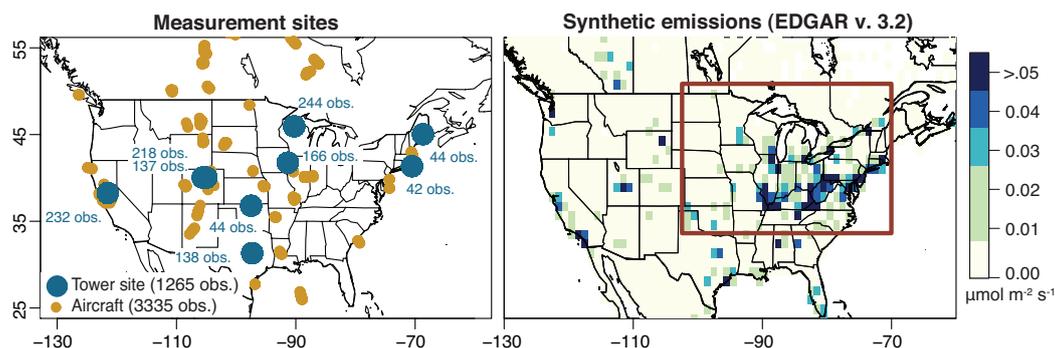
Fig. 1. The synthetic measurements and synthetic emissions used in this study. Blue numbers (left) indicate the observation count at each tower site. The red box (right) displays the region represented in Table 2.

Table 2. Eastern US/Canada anthropogenic budgets and 95 % confidence intervals (Tg C month^{-1}) for the true synthetic emissions and inversion estimates.

Type	Budget	% of true emiss. encapsulated in the given confidence interval	
		68.2 %	95 %
True emissions	1.61		
Unconstrained inversion	1.60 ± 0.13	90	97
Transform	1.59 ± 0.20	64	87
Lagrange multipliers	1.60		
Metropolis–Hastings	1.60 ± 0.08	86	97
Gibbs sampler	1.58 ± 0.08	86	96

distributions – the probability of an individual element in the emissions field integrating over all possible values of the remaining elements. Even emissions estimated over source regions (Fig. 4b) include negative values in the 95% confidence interval.

Additionally, the unconstrained confidence intervals and conditional realizations extend too high for reasons noted in Sect. 1. Figure 5 shows sample conditional realizations from each method. Emissions in the unconstrained realization extend both lower and higher than the realizations estimated by either MCMC algorithm.

5.2 Data transformations

Transformations can be straightforward to implement, but this class of methods skews the probability distributions in the inversion: three of the most important implications are discussed here. First, the posterior covariances (e.g., prior and posterior uncertainties) cannot be directly transformed back to normal space; instead, upper and lower estimation bounds (i.e., percentiles) must be back-transformed to produce posterior confidence intervals. In other words, the covariances become central-value-dependent and are otherwise difficult to physically interpret in back-transformed units. Second, because the covariances are central-value-dependent, it can be difficult to estimate the a priori

covariance matrix (**Q**) in the transformed space, particularly for two of the most common estimation methods. One could use existing knowledge of the emissions to estimate the covariances, but this approach becomes difficult when the covariance matrix has little physical meaning in the transformed space. The covariance matrices can also be inferred from the data and model itself using statistical approaches such as REML. The transformation necessitates iterating between covariance parameter estimation and flux estimation until both converge (Snodgrass and Kitanidis, 1997). The nonlinearities created by the transformation often hinder convergence.

Third, the skewness implied by the power transformation is, in many cases, not representative of actual uncertainties in the emissions best estimate. The uncertainties can become too large in regions of high emissions and too small in regions of low emissions (Fig. 2; e.g., Snodgrass and Kitanidis, 1997; Fienen et al., 2004; Muller and Stavrakou, 2005). For example, conditional realizations follow the lower bounds in the methane case study but produce estimates of the sources that are too large in some high-emissions areas. As a result, the back-transformed conditional realizations have an average eastern US budget of 2.1 ± 0.2 Tg C month^{-1}. The mean of these realizations is much higher than the emissions best estimate for the transform inversion (Table 2). In this case,
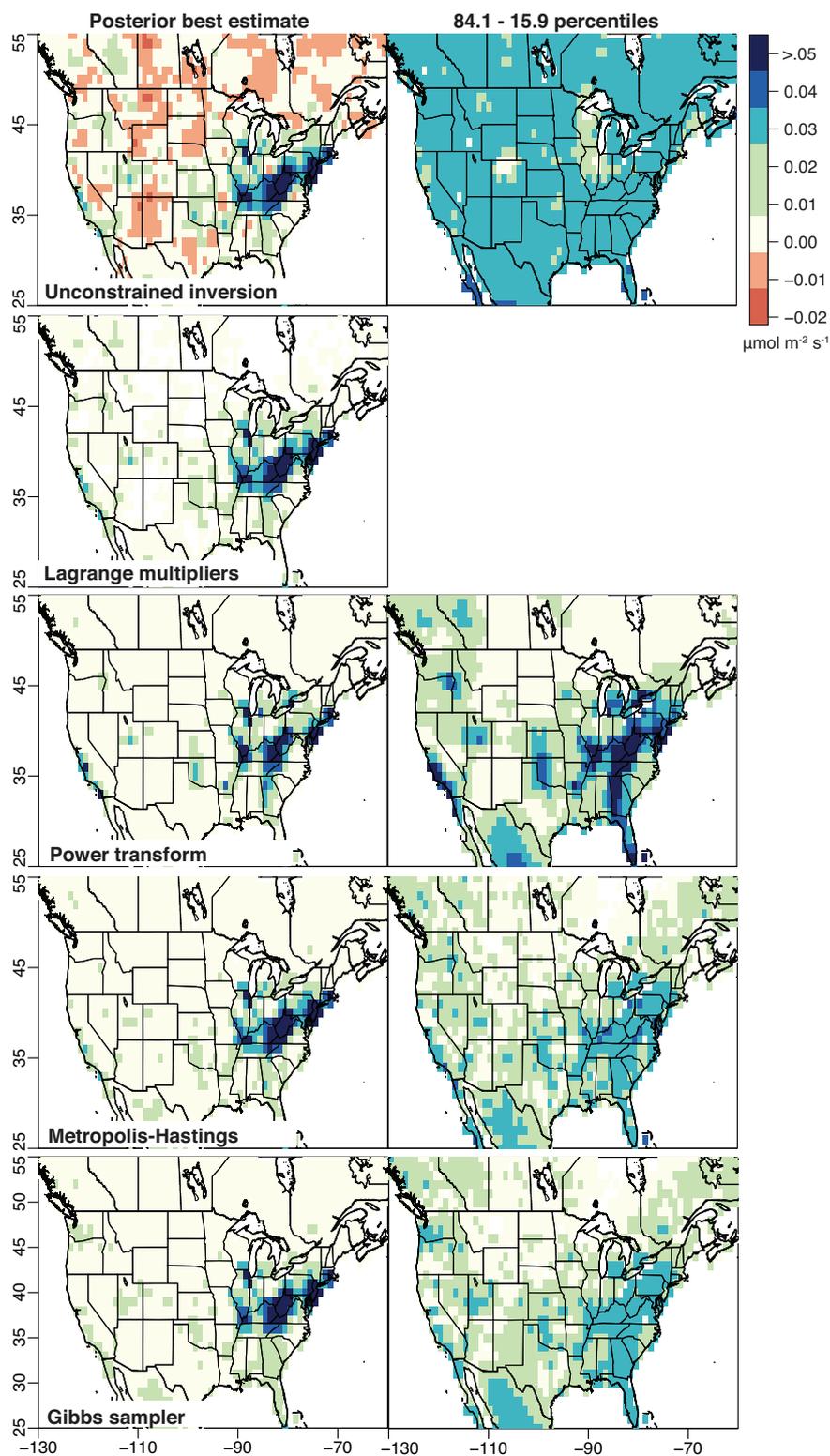
Fig. 2. The posterior best estimate of the emissions and uncertainties associated with each methodological approach. The method of Lagrange multipliers does not support a direct means of estimating uncertainties.
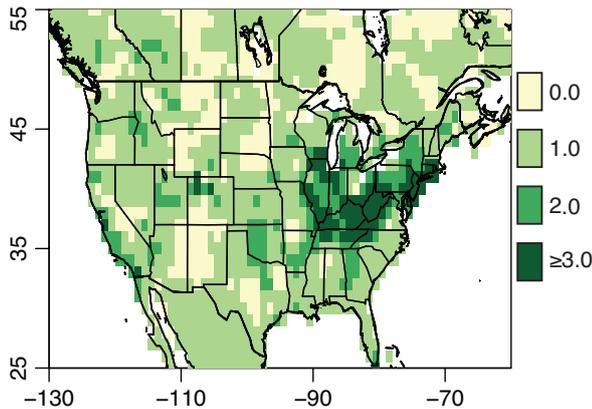
Fig. 3. The number of posterior uncertainty standard deviations before the methane emissions become negative in the unconstrained estimate.

the best estimate is computed by minimizing the inversion cost function directly. In contrast, the mean of the conditional realizations is identical to the emissions best estimate for all other methods discussed in this paper. In addition, the average uncertainties in Table 2 are larger than any other method, yet these uncertainties capture a lower percentage of the synthetic fluxes at grid scale than other methods. These pitfalls illustrate difficulties in interpreting uncertainties or realizations in the power transform case. Snodgrass and Kitanidis (1997), Fienen et al. (2004) and Muller and Stavrakou (2005) provide further discussion on several of the above challenges associated with data transformations.

5.3 Lagrange multipliers

The emissions estimated via Lagrange multipliers reproduce the magnitude and distribution of the true sources. This method is not truly stochastic, however, and removes many elements from the optimization entirely (e.g., emissions over most of Manitoba, Ontario, and Quebec, Canada, in Fig. 2). As such, there is no way to calculate either uncertainty bounds or conditional realizations using this approach. The uncertainties assigned to the posterior emissions are typically borrowed from the unconstrained case, though the uncertainties could be borrowed from any other method. Hence, estimation via Lagrange multipliers resolves the problem of unrealistic emissions, but it does not address the challenge of estimating bounded posterior uncertainties or confidence intervals.

5.4 MCMC implementations

The MCMC implementations discussed here provide an appealing option when robust uncertainty bounds are a priority in the analysis. Both of the explored implementations ensure that the best estimate (Fig. 2), conditional realizations (Fig. 5), and confidence intervals respect the known bounds.
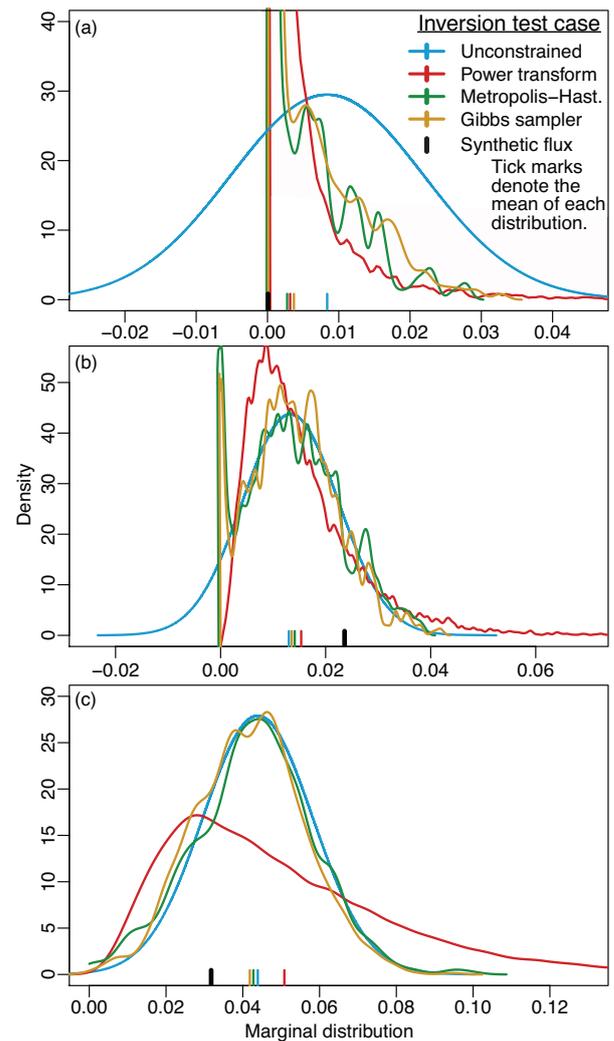

Fig. 4. The marginal posterior density for the estimate of methane emissions at three individual locations. Case **(a)** is an estimate of emissions north of Thunder Bay, Ontario, Canada; case **(b)** is over Indianapolis, Indiana; and case **(c)** is over eastern Kentucky. The unconstrained case is plotted as a normal distribution, and the other plotted probability densities are produced by applying a kernel smoother to the histogram of realizations. Note that this figure does not include the Lagrange multipliers case because this deterministic approach produces only a best estimate with no associated marginal densities.

Both MCMC implementations produce much narrower uncertainty bounds relative to the other methods (Fig. 2, Table 2). As discussed in the introduction, the reason for this is twofold. First, the confidence intervals must be smaller because they cannot include values outside the inequality constraints. Second, if the lower range of the confidence intervals is limited, then the maximum emissions values in the interval will also be less extreme (and vice versa; see Sect. 1). The uncertainties are smaller, and yet 96–97 % of the synthetic fluxes still fall within the inversion's 95 % confidence
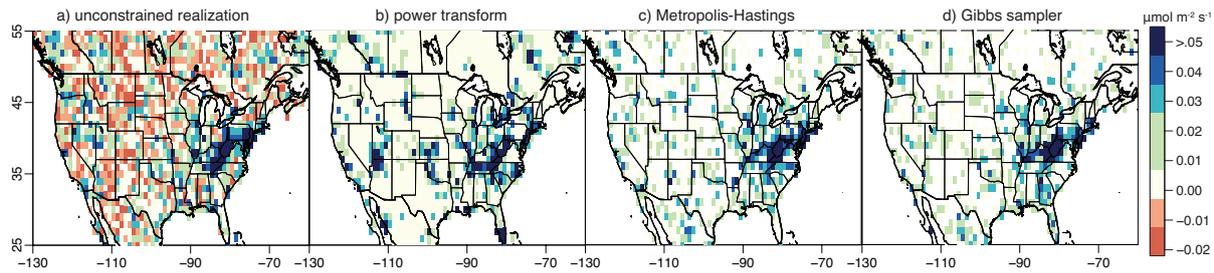
Fig. 5. Example conditional realizations from each different optimization approach.

interval (Table 2). For these reasons, the smaller confidence intervals estimated by the MCMC implementations are most realistic.

The estimated emissions and marginal distributions look very similar between the two MCMC implementations, but the methods show several subtle differences. Unlike the Gibbs sampler, the implementation of the Metropolis–Hastings algorithm here uses Lagrange multipliers and therefore does not explicitly model every element of s in every realization as a continuous random variable. As a result, this Metropolis–Hastings method will always produce a high probability at the inequality constraints (e.g., Fig. 4). For smaller problems, in contrast, it may be feasible to use any shape of prior pdf that is defined only within the inequality constraints. That setup would model every element of s as a continuous random variable, but the approach would likely become computationally intractable for larger problems (see the supplement for further discussion). Instead, the Metropolis–Hastings implementation used here is an extension of Lagrange multipliers to circumstances that require bounded confidence intervals. The Gibbs sampler implementation produces a similar peak in the pdf at zero due to the Dirac delta. In general, however, the Gibbs sampler allows more flexibility in setting the shape of the marginal distributions near the bounds.

Appropriate distributional assumptions are important for any type of inversion, and the inversion with inequality constraints is no different. The Gibbs sampler in this case study models the marginal distributions as a truncated Gaussian with a Dirac delta function (see Sect. 3.3.2). If the fluxes or emissions are unlikely to be zero, an implementation without the Dirac delta would be more suitable.

Furthermore, the choice of a truncated normal distribution may not always be appropriate. If the total budget is poorly constrained by the data, this distributional choice could increase estimated emissions in remote regions far from measurement sites. A Gaussian pdf that has been truncated at zero will have a higher mean than the equivalent, full Gaussian distribution, and this effect can shift the posterior mean in poorly constrained problems. One solution could be to fix the drift coefficients (β) at predetermined values, but these coefficients are rarely known in practice. In contrast, if

measurement sites are sensitive to emissions across the entire geographic domain (as indicated by $\mathbf{H}$), then either distributional assumption will produce the same trace gas budget.

The MCMC implementations produce the most realistic best estimate, conditional realizations, and uncertainty bounds, but one drawback can be computational cost. The generation of j conditional realizations using the Gibbs sampler requires a for loop with jm iterations, and j is usually 1000 or greater to adequately sample the posterior probability space. The computational time of the multiple-try Metropolis–Hastings depends on the convergence rate of the Lagrange multiplier algorithm and on the number of trial realizations (denoted k; see the Supplement) computed in each step of the multiple-try implementation. The often large ratio of trial to accepted realizations means that the multiple-try Metropolis–Hastings may be less efficient than the Gibbs implementation. This comparison may seem counterintuitive because Metropolis–Hastings does not require the posterior pdf to be sampled one element at a time, as is the case for the Gibbs sampler.

In general, the computational cost of MCMC algorithms increases both with the number of unknown fluxes (m) and the number of realizations (j) required to sample the posterior probability space. Furthermore, the recommended number of realizations changes depending on the size of the problem and the degree of correlation among successive iterations; Gelman (2004) provides a number of guidelines for choosing this number. One approach requires initiating multiple independent Markov chains. The MCMC algorithm reaches convergence when each individual chain has a similar distribution to the combination of all chains. Refer to Gelman (2004, Chap. 11) for more detail. Parallelization can also alleviate some time expense for MCMC algorithms.

In summary, the Gibbs and Metropolis–Hastings implementations produce similar results, but the Gibbs sampler can afford two advantages: greater flexibility in determining the shape of the marginal distributions at the bounds and reduced computational time for the case examined here.

6 Conclusions

For inverse problems with parameters that have known physical limits, an unconstrained inversion presents difficulties that go beyond just an unrealistic estimate, and the common remedy of using data transformations can have many undesirable side effects. This study uses anthropogenic methane emissions as a lens to evaluate this approach, as well as several less common alternative approaches.

Inverse problems can be constructed to honor known bounds without compromising the integrity of the estimate. Lagrange multipliers are a viable approach for large problems in which computational time is paramount. However, this method does not provide an explicit means for calculating uncertainty bounds. Uncertainties are usually borrowed from the unconstrained case instead, and these are generally unrealistically large.

Markov chain Monte Carlo (MCMC) methods show the most promise but are rarely applied in the existing atmospheric literature. Both MCMC implementations here produce similar results for the methane case study, but the Gibbs sampler offers better computational efficiency and more flexibility in determining the shape of posterior probability at the bounds. In general, MCMC algorithms can be applied to inverse problems with known bounds to produce the most realistic best estimates, confidence intervals, and conditional realizations of any of the aforementioned approaches.

Supplementary material related to this article is available online at http://www.geosci-model-dev.net/7/303/2014/gmd-7-303-2014-supplement.pdf.

References

Andrieu, C., de Freitas, N., Doucet, A., and Jordan, M.: An Introduction to MCMC for machine learning, Mach. Learn., 50, 5–43, doi:10.1023/A:1020281327116, 2003.

Antoniou, A. and Lu, W.: Practical Optimization: Algorithms and Engineering Applications, Springer, New York, NY, 2007.

Barnes, R. and You, K.: Adding bounds to kriging, Math. Geol., 24, 171–176, doi:10.1007/BF00897030, 1992.

Bergamaschi, P., Frankenberg, C., Meirink, J. F., Krol, M., Villani, M. G., Houweling, S., Dentener, F., Dlugokencky, E. J., Miller, J. B., Gatti, L. V., Engel, A., and Levin, I.: Inverse modeling of global and regional CH_4 emissions using SCIAMACHY satellite retrievals, J. Geophys. Res., 114, D22301, doi:10.1029/2009JD012287, 2009.

Biraud, S. C., Torn, M. S., Smith, J. R., Sweeney, C., Riley, W. J., and Tans, P. P.: A multi-year record of airborne CO_2 observations in the US Southern Great Plains, Atmos. Meas. Tech., 6, 751–763, doi:10.5194/amt-6-751-2013, 2013.

Bolstad, W.: Understanding Computational Bayesian Statistics, Wiley Series in Computational Statistics, John Wiley & Sons, Hoboken, New Jersey, 2012.

Burrows, S. M., Rayner, P. J., Butler, T., and Lawrence, M. G.: Estimating bacteria emissions from inversion of atmospheric transport: sensitivity to modelled particle characteristics, Atmos. Chem. Phys., 13, 5473–5488, doi:10.5194/acp-13-5473-2013, 2013.

Byrd, R., Lu, P., Nocedal, J., and Zhu, C.: A Limited Memory Algorithm for Bound Constrained Optimization, SIAM J. Sci. Comput., 16, 1190–1208, doi:10.1137/0916069, 1995.

Casella, G. and George, E. I.: Explaining the Gibbs Sampler, Am. Stat., 46, 167–174, 1992.

Chib, S. and Greenberg, E.: Understanding the Metropolis–Hastings Algorithm, Am. Stat., 49, 327–335, 1995.

Conn, A., Gould, N., and Toint, P.: Trust-Region Methods, Mps-Siam Series on Optimization, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2000.

Eckhardt, S., Prata, A. J., Seibert, P., Stebel, K., and Stohl, A.: Estimation of the vertical profile of sulfur dioxide injection into the atmosphere by a volcanic eruption using satellite column measurements and inverse transport modeling, Atmos. Chem. Phys., 8, 3881–3897, doi:10.5194/acp-8-3881-2008, 2008.

Enting, I.: Inverse Problems in Atmospheric Constituent Transport, Cambridge Atmospheric and Space Science Series, Cambridge University Press, Cambridge, UK, 2002.

Fienen, M., Kitanidis, P., Watson, D., and Jardine, P.: An application of Bayesian inverse methods to vertical deconvolution of hydraulic conductivity in a heterogeneous aquifer at Oak Ridge National Laboratory, Math. Geol., 36, 101–126, doi:10.1023/B:MATG.0000016232.71993.bd, 2004.

Fienen, M. N., Luo, J., and Kitanidis, P. K.: A Bayesian geostatistical transfer function approach to tracer test analysis, Water Resour. Res., 42, W07426, doi:10.1029/2005WR004576, 2006.

Gay, D.: A trust-region approach to linearly constrained optimization, in: Numerical Analysis, edited by: Griffiths, D., Springer, Berlin, Heidelberg, Lect. Notes Math., 1066, 72–105, doi:10.1007/BFb0099519, 1984.

Gelfand, A. E., Smith, A. F. M., and Lee, T.-M.: Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling, J. Am. Stat. Assoc., 87, 523–532, 1992.

Gelman, A.: Bayesian Data Analysis., Texts in Statistical Science, Chapman & Hall/CRC, Boca Raton, Florida, 2nd Edn., 2004.

Geyer, C.: Introduction to Markov Chain Monte Carlo, in: Handbook of Markov Chain Monte Carlo, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, Taylor & Francis, London, 3–48, 2011.

Gill, P. E., Murray, W., and Wright, M. H.: Practical Optimization, Academic Press, London, 1981.

Göckede, M., Turner, D. P., Michalak, A. M., Vickers, D., and Law, B. E.: Sensitivity of a subregional scale atmospheric inverse CO_2 modeling framework to boundary conditions, J. Geophys. Res., 115, D24112, doi:10.1029/2010JD014443, 2010.

Gourdji, S. M., Mueller, K. L., Schaefer, K., and Michalak, A. M.: Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data, J. Geophys. Res., 113, D21115, doi:10.1029/2007JD009733, 2008.

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, doi:10.5194/bg-9-457-2012, 2012.

Gurney, K., Law, R., Denning, A., Rayner, P., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fan, S., Fung, I., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C.: Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, Nature, 415, 626–630, doi:10.1038/415626a, 2002.

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

Kitanidis, P. K.: Parameter uncertainty in estimation of spatial functions: Bayesian analysis, Water Resour. Res., 22, 499–507, 1986.

Kitanidis, P. K. and Lane, R. W.: Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method, J. Hydrol., 79, 53–71, doi:10.1016/0022-1694(85)90181-7, 1985.

Kopacz, M., Jacob, D. J., Henze, D. K., Heald, C. L., Streets, D. G., and Zhang, Q.: Comparison of adjoint and analytical Bayesian inversion methods for constraining Asian sources of carbon monoxide using satellite (MOPITT) measurements of CO columns, J. Geophys. Res., 114, D04305, doi:10.1029/2007JD009264, 2009.

Kort, E. A., Eluszkiewicz, J., Stephens, B. B., Miller, J. B., Gerbig, C., Nehrkorn, T., Daube, B. C., Kaplan, J. O., Houweling, S., and Wofsy, S. C.: Emissions of CH_4 and N_2O over the United States and Canada based on a receptor-oriented modeling framework and COBRA-NA atmospheric observations, Geophys. Res. Lett., 35, L18808, doi:10.1029/2008GL034031, 2008.

Lin, C. and More, J.: Newton's method for large bound-constrained optimization problems, SIAM J. Optimiz., 9, 1100–1127, doi:10.1137/S1052623498345075, 1999.

Lin, J., Gerbig, C., Wofsy, S., Andrews, A., Daube, B., Davis, K., and Grainger, C.: A near-field tool for simulating the upstream influence of atmospheric observations: the Stochastic Time-Inverted Lagrangian Transport (STILT) model, J. Geophys. Res.-Atmos., 108, 4493, doi:10.1029/2002JD003161, 2003.

Liu, J. S., Liang, F., and Wong, W. H.: The multiple-try method and local optimization in metropolis sampling, J. Am. Stat. Assoc., 95, 121–134, doi:10.1080/01621459.2000.10473908, 2000.

Michalak, A. M.: A Gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling, Water Resour. Res., 44, W09437, doi:10.1029/2007WR006645, 2008.

Michalak, A. M. and Kitanidis, P.: A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification, Water Resour. Res., 39, 1033, doi:10.1029/2002WR001480, 2003.

Michalak, A. M., Bruhwiler, L., and Tans, P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, J. Geophys. Res., 109, D14109, doi:10.1029/2003JD004422, 2004.

Michalak, A. M. and Kitanidis, P. K.: Application of geostatistical inverse modeling to contaminant source identification at Dover AFB, Delaware, J. Hydraul. Res., 42, 9–18, doi:10.1080/00221680409500042, 2004.

Miller, S. M., Wofsy, S. C., Michalak, A. M., Kort, E. A., Andrews, A. E., Biraud, S. C., Dlugokencky, E. J., Eluszkiewicz, J., Fischer, M. L., Janssens-Maenhout, G., Miller, B. R., Miller, J. B., Montzka, S. A., Nehrkorn, T., and Sweeney, C.: Anthropogenic emissions of methane in the United States, P. Natl. Acad. Sci. USA, doi:10.1073/pnas.1314392110, 2013.

Miller, S. M., Worthy, D. E. J., Michalak, A. M., Wofsy, S. C., Kort, E. A., Havice, T. C., Andrews, A. E., Dlugokencky, E. J., Kaplan, J. O., Levi, P. J., Tian, H., and Zhang, B.: Top-down controls on the distribution, seasonality, and environmental predictors of North American boreal methane emissions, Global Biogeochem. Cy., doi:10.1002/2013GB004580, 2014.

More, J.: Trust regions and projected gradients, Lect. Notes Contr. Inf., 113, 1–13, 1988.

Mueller, K. L., Gourdji, S. M., and Michalak, A. M.: Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 1. Results using atmospheric measurements, J. Geophys. Res., 113, D21114, doi:10.1029/2007JD009734, 2008.

Müller, J.-F. and Stavrakou, T.: Inversion of CO and NO_x emissions using the adjoint of the IMAGES model, Atmos. Chem. Phys., 5, 1157–1186, doi:10.5194/acp-5-1157-2005, 2005.

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

Olivier, J. and Peters, J.: CO2 from non-energy use of fuels: a global, regional and national perspective based on the IPCC Tier 1 approach, Resour. Conserv. Recy., 45, 210–225, doi:10.1016/j.resconrec.2005.05.008, 2005.

Pan, L. L., Bowman, K. P., Atlas, E. L., Wofsy, S. C., Zhang, F., Bresch, J. F., Ridley, B. A., Pittman, J. V., Homeyer, C. R., Romashkin, P., and Cooper, W. A.: The stratosphere–troposphere analyses of regional transport 2008 experiment, B. Am. Meteorol. Soc., 91, 327–342, doi:10.1175/2009BAMS2865.1, 2010.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Pétron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An

atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, P. Natl. Acad. Sci. USA, 104, 18925–18930, doi:10.1073/pnas.0708986104, 2007.

Rigby, M., Manning, A. J., and Prinn, R. G.: Inversion of long-lived trace gas emissions using combined Eulerian and Lagrangian chemical transport models, Atmos. Chem. Phys., 11, 9887–9898, doi:10.5194/acp-11-9887-2011, 2011.

Rodgers, C.: Inverse Methods for Atmospheric Sounding: Theory and Practice, Series on Atmospheric, Oceanic and Planetary Physics, World Scientific, Singapore, 2000.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Wang, W., and Powers, J.: A description of the advanced research WRF version 2, available at: www.mmm.ucar.edu/wrf/users/docs/arw_v2.pdf (last access: 16 August 2013), 2005.

Snodgrass, M. and Kitanidis, P.: A geostatistical approach to contaminant source identification, Water Resour. Res., 33, 537–546, doi:10.1029/96WR03753, 1997.

Snyman, J.: Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms, Applied Optimization, Springer, Boston, MA, 2005.

Sorensen, D. C.: Newton's method with a model trust region modification, SIAM J. Numer. Anal., 19, 409–426, 1982.

Stohl, A., Seibert, P., Wotawa, G., Arnold, D., Burkhart, J. F., Eckhardt, S., Tapia, C., Vargas, A., and Yasunari, T. J.: Xenon-133 and caesium-137 releases into the atmosphere from the Fukushima Dai-ichi nuclear power plant: determination of the source term, atmospheric dispersion, and deposition, Atmos. Chem. Phys., 12, 2313–2343, doi:10.5194/acp-12-2313-2012, 2012.

Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2005.

Theil, H. and Panne, C. V. D.: Quadratic programming as an extension of classical quadratic maximization, Manage. Sci., 7, 1–20, 1960.

Walvoort, D. and de Gruijter, J.: Compositional kriging: a spatial interpolation method for compositional data, Math. Geol., 33, 951–966, doi:10.1023/A:1012250107121, 2001.

Wang, J. and Zabaras, N.: Hierarchical Bayesian models for inverse probl. in heat conduction, Inverse Probl., 21, 183–206, doi:10.1088/0266-5611/21/1/012, 2005.

Wang, J. and Zabaras, N.: A Markov random field model of contamination source identification in porous media flow, Int. J. Heat Mass. Tran., 49, 939–950, doi:10.1016/j.ijheatmasstransfer.2005.09.016, 2006.

Wilks, D.: Statistical Methods in the Atmospheric Sciences, Academic Press, Oxford, UK, 3rd Edn., 2011.

Yuan, Y.-X.: A review of trust region algorithms for optimization in: Proceedings of the Fourth International Congress on Industrial &Applied Mathematics, ICIAM 99, Edinburgh, 5–9 July 1999, Oxford Univ. Press, Oxford, UK, 271–282, 2000.

Zanini, A. and Kitanidis, P. K.: Geostatistical inversing for large-contrast transmissivity fields, Stoch. Env. Res. Risk. A., 23, 565–577, doi:10.1007/s00477-008-0241-7, 2009.