Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions

Atmospheric Inverse Modeling via Sparse Reconstruction

Nils Hase[1], Scot M. Miller[2], Peter Maaß[1], Justus Notholt[3], Mathias Palm[3], and Thorsten Warneke[3]

[1]Center for Industrial Mathematics, University of Bremen, Bremen, Germany
[2]Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA
[3]Institute of Environmental Physics, University of Bremen, Bremen, Germany

Correspondence to: N. Hase (nilshase@math.uni-bremen.de)

Abstract.

Many applications in atmospheric science involve ill-posed inverse problems. A crucial component of many inverse problems is the proper formulation of a priori knowledge about the unknown parameters. In most cases, this knowledge is expressed as a Gaussian prior. This formulation often performs well at capturing smoothed, large-scale processes but is often ill-equipped to capture localized structures like large point sources or localized hot spots.

Over the last decade, scientists from a diverse array of applied math and engineering fields have developped sparse reconstruction techniques to identify localized structures. In this study we present a new regularization approach for ill-posed inverse problems in atmospheric science. It is based on Tikhonov regularization with sparsity constraint and allows bounds on the parameters. We enforce sparsity using a dictionary representation system. We analyze its performance in an atmospheric inverse modeling scenario by estimating anthropogenic US methane emissions from simulated atmospheric measurements.

Different measures indicate that our sparse reconstruction approach is better able to capture large point sources or localized hot spots than other methods commonly used in atmospheric inversions. It captures the overall signal equally well, but adds details on the grid scale. This can be of great value in many research projects. We show an example for source estimation of synthetic methane emissions from the Barnett shale formation.

1 Introduction

Inverse problems are widespread in atmospheric sciences. The estimation of greenhouse gas sources and sinks is a prime example. Numerous studies combine observations of greenhouse gas concentrations in the atmosphere and inverse modeling to infer sources and sinks at the Earth's surface. Existing studies have applied these techniques at municipal (e.g. Saide et al., 2011), regional (e.g. Zhao et al., 2009), continental (e.g. Miller et al., 2013) and global scale (e.g. Stohl et al., 2009). Inverse modeling estimates of greenhouse gas emissions are not only of scientific interest (e.g., to assess biospheric fluxes or improve process-based models). These estimates are also key for monitoring and evaluating greenhouse gas emissions regulations (U.S. National Research Council, 2010).

In almost all cases, these parameter estimation problems are ill-posed. Ill-posed means that small noise on the measurements can be largely amplified by the inversion leading to unrealistic estimates. Thus, special techniques are required for a stable inversion.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

A common tool in atmospheric sciences to handle the ill-posed nature of these problems is Bayesian inversion (e.g. Rodgers, 2000). In Bayesian inversion, the unknown parameters are assumed to follow an a priori distribution. The observations are used to calculate an update of these parameters. The final estimate is thus based on balanced information from the observations and the prior.

A classical approach is the use of a Gaussian prior, which allows rapid calculations via analytical expressions. However, the Gaussian prior is known to return smoothed versions of the true solution. It is well-suited to detect the overall process, but local structures such as large point sources are often smoothed out for ill-posed problems.

Other research areas solve inverse problems using Tikhonov regularization. Tikhonov regularization is formulated as an optimization problem. The functional to be minimized consists of a data fitting term and a penalty term that prevents overfitting. The classical choice of these terms is analogous to Bayesian inversion with Gaussian prior.

Recently, Tikhonov regularization with sparsity constraint has become a popular alternative to these classical inverse methods within a number of engineering fields. It has been used successfully in many different applications, including medical imaging, signal analysis and compressed sensing (see e.g. Hämäläinen et al., 2013; Knopp and Weber, 2013; Candès et al., 2011). All of these applications make use of the fact that the underlying process can be described as a localized signal in a suitable representation system. While the classical approach tends to smooth the true process (in any representation system), sparse reconstruction is designed to find such localized structures. Jin and Maass (2012) give a detailed summary of the mathematical advances with the sparsity constraint.

These modern inversion techniques have only been applied to atmospheric sciences in a small number of cases. Martinez-Camara et al. (2013) used a sparse reconstruction approach to estimate emissions of radioactive substances for the Fukushima accident, and Ray et al. (2015) analyzed fossil fuel carbondioxid emissions in an idealized, synthetic data setup.

The goal of this paper is to show how sparse reconstruction techniques can improve the flux estimates in an atmospheric inverse modeling scenario. We use a synthetic case study from Miller et al. (2014), where different inversion methods that force positive surface fluxes were analyzed. The setup considers anthropogenic methane emissions in the United States. We couple a sparse reconstruction approach with a positivity constraint.

Our study is organized as follows: First, we briefly introduce the atmospheric inverse modeling problem (Sect. 2). Section 3 gives an overview on inverse problems and introduces the concept of sparse reconstruction. We use a redundant dictionary representation system to sparsify the flux signal. The setup of the synthetic case study and our sparse dictionary reconstruction method are presented in Sect. 4. Estimates, error analysis and a comparison with state-of-the-art methods are shown in the results Sect. 5. We also analyze the sensitivity to emissions from an oil and gas drilling region before drawing conclusions.

Additional graphics, source code and a pseudocode of our sparse dictionary reconstruction method are included in the supplementary information.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

2 Surface flux estimation using atmospheric inverse modeling

Different techniques have been used to quantify greenhouse gas surface fluxes (Hensen et al., 2013). Atmospheric inverse modeling (AIM) is an approach, that relies on the knowledge of a proper atmospheric transport model to link surface sources and sinks to enhancements in atmospheric greenhouse gas concentrations. The idea is to invert the transport model and thus map atmospheric measurements to surface fluxes.

We use the WRF-STILT (Weather Research and Forecasting - Stochastic Time-Inverted Lagrangian Transport) model (Nehrkorn et al., 2010) to simulate atmospheric transport in this study, the same simulations used in Miller et al. (2013, 2014). WRF is a meteorology model (e.g. Skamarock et al., 2005) and STILT is a back-trajectory model (e.g. Lin et al., 2003; Gerbig et al., 2003). STILT will release an ensemble of imaginary particles at the time and location of an atmospheric measurement. The particles then travel backward in time and indicate where air masses were located before reaching the measurement location. STILT then uses the distribution of these particles to compute an upwind influence on the measurement, called footprint. The footprint quantitatively relates the surface fluxes to the atmospheric measurement.

For a given emission field, $\boldsymbol{x}$, the enhancement of the measurement above a known background level, y_k, can be simulated by integrating the product of footprint, A_k, and emissions over the Earth's surface area of interest, Ω, which gives

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x}, \quad \text{where} \quad [Ax]_k := \int_{\Omega} A_k(s)x(s)ds \tag{1}$$

and $s \in \Omega$ is the integration variable of location. The central question in AIM is how to determine a realistic flux field $\boldsymbol{x}$ given (noisy) atmospheric measurements $\boldsymbol{y_\delta}$ and footprints $\mathbf{A}$, which means solving the inverse problem of Eq. (1). Apart from the inverse problem itself, challenges involve estimation of background concentrations and proper modeling of atmospheric transport and chemistry, but this article only adresses the solution of the inverse problem.

3 Mathematical background of inverse problems

In this section we give some mathematical background for inverse problems and how our approach is related to commonly used inverse methods. We will formulate the AIM problem as a parameter optimization problem, which is based on norm notation. Thus, we define

$$\|\boldsymbol{z}\|_2 := \sqrt{\sum_k |z_k|^2} \quad \text{and} \quad \|\boldsymbol{z}\|_1 := \sum_k |z_k| \,.$$

Both norms measure the length of a vector $\boldsymbol{z}$, where $\boldsymbol{z}$ is a vector of any quantity. The 2-norm is the standard norm in most fields of study. The 1-norm is a central concept in sparse reconstruction, as we will see in the following.

3.1 Ill-posed inverse problems

Inverse problems arise, when the quantity of interest cannot be measured directly. Instead, another quantity $\boldsymbol{y}$ is measured that is related to the unknown parameters $\boldsymbol{x}$ by a forward model $\mathbf{F}$. The forward model maps from parameter space X to measurement

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

space Y. For most problems we have $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$. The forward problem is to calculate simulated measurement data $\boldsymbol{y}$ from known parameters $\boldsymbol{x}$ by evaluating the potentially nonlinear forward model, $\boldsymbol{y} = \mathbf{F}(\boldsymbol{x})$. Estimating realistic parameters that explain the given measurements means solving the inverse problem.

The problem of finding parameters that best explain noisy measurements $\boldsymbol{y_\delta}$ in a least squares sense is equivalent to solving

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in X} \frac{1}{2}\|\mathbf{F}(\boldsymbol{x}) - \boldsymbol{y_\delta}\|_2^2. \tag{2}$$

Often the inverse problem is ill-posed. This means that the minimizer $\boldsymbol{x}^*$ might be nonunique and particularly that the inversion is unstable. Unstable denotes that small changes in the measurements result in large changes in estimated parameters. In the real world, measurements are never exact. A common assumption is that $\boldsymbol{y_\delta}$ can be split up into a part that can be explained exactly by the true parameters $\boldsymbol{x}^+$ via the underlying forward model, so $\boldsymbol{y} = \mathbf{F}(\boldsymbol{x}^+)$, and some noise $\boldsymbol{\delta} \in Y$, such that $\boldsymbol{y_\delta} = \boldsymbol{y} + \boldsymbol{\delta}$. In this definition the noise includes errors from the measurements, the forward model as well as numerical approximations. Solving Eq. (2) means fitting the parameters to the data and the noise. The retrieved parameters are very sensitive to the data in ill-posed problems, so the true solution $\boldsymbol{x}^+$ is typically far away from the least squares solution $\boldsymbol{x}^*$. Thus, the inverse mapping using Eq. (2) is not suitable for ill-posed problems (see the supplementary information).

3.2 Tikhonov regularization

The inversion using Eq. (2) is unstable for ill-posed problems. Tikhonov regularization stabilizes the inversion by adding a convex penalty function $\phi : X \to \mathbb{R}$ (see e.g. Hansen, 2010; Louis, 1989)

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in X} \frac{1}{2}\|\mathbf{F}(\boldsymbol{x}) - \boldsymbol{y_\delta}\|_2^2 + \alpha\phi(\boldsymbol{x}). \tag{3}$$

Classical Tikhonov regularization uses $\phi(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x_a}\|_2^2$ with $\boldsymbol{x_a} = \mathbf{0}$. The regularization parameter α, with $\alpha > 0$, weights data fitting and penalty term. A greater value forces the solution to stay close to the a priori solution $\boldsymbol{x_a}$, trusting less in the data, while a small value results in less discrepancy in the data fitting term. A number of methods are available to automatically choose a balancing regularization parameter (see e.g. Reichel and Rodriguez, 2012). We use Morozov's discrepancy principle (see Eq. (13) in Sect. 4.2), which requires knowledge of the noise level $\|\boldsymbol{\delta}\|$. A suitable regularization parameter prevents overfitting of the estimated parameters $\boldsymbol{x}^*$ to the noisy data via the forward model.

If more detailed noise characteristics are known, these can be introduced by adaptation of the data fitting term. In case of Gaussian noise, penalized weighted least squares

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in X} \frac{1}{2}\|\mathbf{L}_\delta(\mathbf{F}(\boldsymbol{x}) - \boldsymbol{y_\delta})\|_2^2 + \alpha\phi(\boldsymbol{x}) \tag{4}$$

with an noise covariance $\mathbf{R}$, with $\mathbf{R}^{-1} = \mathbf{L}_\delta^t\mathbf{L}_\delta$, best include this information. The covariance is especially useful to weight measurements of different uncertainties. For a suitable penalty function, this approach translates to Bayesian inverse modeling (see Sect. 3.6).

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

3.3 Choice of the penalty function

We decided to solve the inverse problem using Tikhonov regularization, Eq. (4). Information about the measurement noise is introduced in the data fitting term of the Tikhonov functional, and prior information about the unknown parameters is formulated in the penalty function ϕ. If nothing is known, the classical Tikhonov penalty is the 2-norm penalty, $\phi(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x}\|_2^2$, also called zero a priori. Among all possible solutions, it will choose the solution that is closest to the origin but still reproduces the data. In particular, it prevents oszillations as encountered when the inversion is unstable. If an a priori estimate x_a of the parameters is available, it can be included by $\phi(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x_a}\|_2^2$.

Sometimes it can be useful to penalize the components of the parameter vector $\boldsymbol{x}$ differently. This results in a weighted 2-norm penalty, so $\phi(\boldsymbol{x}) = \frac{1}{2}\|\mathbf{L_a}\boldsymbol{x}\|_2^2$. In a Bayesian inversion setup $\mathbf{Q} := \frac{1}{\alpha}(\mathbf{L_a^t L_a})^{-1}$ gives the covariance of a Gaussian a priori distribution. Diagonal elements in $\mathbf{L_a}$ weight the parameters while off-diagonal entries correlate parameters resulting in smoother estimates in case of positive correlation and vice versa.

A large number of methods is available to solve optimization problems of the type

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in X} \frac{1}{2}\|\mathbf{L}_\delta(\mathbf{F}(\boldsymbol{x}) - \boldsymbol{y_\delta})\|_2^2 + \frac{\alpha}{2}\|\mathbf{L_a}(\boldsymbol{x} - \boldsymbol{x_a})\|_2^2. \tag{5}$$

The minimizer $\boldsymbol{x}^*$, if it exists (see Sect. 3.4), can also be interpreted as a maximum a posteriori solution to a Gaussian prior with Gaussian noise in a Bayesian inversion framework (see Rodgers, 2000, ch. 3).

Localized structures like point sources or edges in the true solution $\boldsymbol{x}^+$ are smoothed out by regularization with the 2-norm and thus disappear in the estimate $\boldsymbol{x}^*$.

Over the last decade, the sparsity contraint has become very popular for regularization of inverse problems. Instead of taking the 2-norm as a penalty function, the 1-norm is used to constrain the parameters. This results in the optimization problem

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in X} \frac{1}{2}\|\mathbf{L}_\delta(\mathbf{F}(\boldsymbol{x}) - \boldsymbol{y_\delta})\|_2^2 + \alpha\|\boldsymbol{x}\|_1. \tag{6}$$

Replacing the 2-norm by the 1-norm adds more penalty on small components in the solution while bigger components are included. The constraint aims at solutions of only few nonzero components, which are called sparse. Figure 1 illustrates why parameters that are not sufficiently constrained by the data are set to zero when using the sparsity constraint. Thus the method in Eq. (6) is called sparse reconstruction.

However, if the true solution $\boldsymbol{x}^+$ is not sparse, reconstruction by Eq. (6) will still seek a sparse approximation of the solution. Of course, the same is true for other penalty functions promoting certain properties of the estimate. Whether or not a signal is sparse is a matter of the representation system used. The solution might be non-sparse in the natural parameter space but have a sparse representation when transformed into a different space. To make use of the sparsity constraint, we have to find a representation system that allows for sparse representation for all possible solutions.

Signal and image processing offer a variety of transforms designed for sparse representation of oscillations, localized signals, edges and the like. Options include regular basis transforms, Fourier transforms, wavelets, shearlets, curvelets to among other options (see e.g. Diniz et al., 2010; Elad, 2010). However, finding a sparsifying transform for a given application is not

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

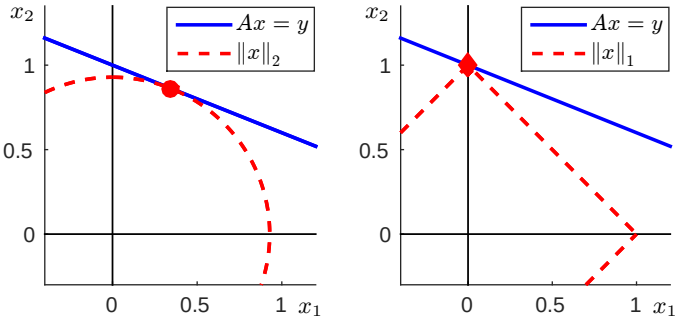

Figure 1. Illustration of 2-norm and 1-norm regularization for an underdetermined problem $\mathbf{A}\boldsymbol{x} = \boldsymbol{y}$ in $\mathbb{R}^2$. Generally, the minimum 1-norm solution is zero in one of the two components. This almost never happens when considering the 2-norm penalty.

straightforward, and often a basis with its unique representation of the state is too restrictive. We consider a more flexible representation system called dictionary, described in detail in the next paragraph.

A dictionary is a collection of N elementary functions $\boldsymbol{d_k} \in X$, called atoms, that can be combined linearly to represent the parameters, so $\boldsymbol{x} = \sum_{k=1}^{N} c_k \boldsymbol{d_k}$. By the choice of the atoms, it can be ensured, that there is at least one representation for each state $\boldsymbol{x} \in X$. Dictionaries are typically redundant representation systems, meaning that $N > \dim(X)$. This implicates that there are infinitely many representations $\boldsymbol{c}$ for the same parameter vector $\boldsymbol{x}$. From a suitable dictionary we expect that at least one of these represenations is sparse, which means that $\boldsymbol{x}^+$ is a linear combination of only a few atoms.

We give an example for $X = \mathbb{R}^3$. Consider the dictionary

$$\mathbf{D} = (\boldsymbol{d_1}, \boldsymbol{d_2}, \boldsymbol{d_3}, \boldsymbol{d_4}) = \begin{pmatrix} \frac{1}{\sqrt{2}} & 1 & 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Each column of $\mathbf{D}$ is an atom of norm one ($\|\boldsymbol{d_k}\|_2 = 1$). The vector $\boldsymbol{x} = (1,1,1)^t$ can be represented in the dictionary by coefficients $\boldsymbol{c} \in \mathbb{R}^4$ in infinitely many different ways as the dictionary is redundant. Some possible represenations are

$$\begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -\sqrt{2} \\ 2 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} \sqrt{2} \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix}, \ldots$$

The first representation is a somewhat natural choice as it represents each dimension with a different atom. The third representation has minimal 1-norm and the fourth minimal 2-norm. All other choices have more complicated structures. The third representation is also the sparsest possible representation. This example illustrates not only that the sparsest solution often coincides with the minimal 1-norm solution, but also how a redundant representation system is able to sparsify the signal with fewer nonzero entries than the vector it represents.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions

We assume that the estimated state in our AIM problem can be sparsely represented in a given dictionary $\mathbf{D}$, leading to the following optimization problem

$$\begin{aligned} \boldsymbol{c}^* &= \arg\min_{\boldsymbol{c}\in C} \frac{1}{2}\|\mathbf{L}_\delta(\mathbf{F}(\mathbf{D}\boldsymbol{c}+\boldsymbol{x_a}) - \boldsymbol{y_\delta})\|_2^2 + \alpha\|\boldsymbol{c}\|_1 \\ \boldsymbol{x}^* &= \mathbf{D}\boldsymbol{c}^*, \end{aligned} \tag{7}$$

where $\boldsymbol{x_a}$ is an a priori estimate of the state. Again, the assumption when solving Eq. (7) is that the difference between true solution and a priori, $\boldsymbol{x}^+ - \boldsymbol{x_a}$, can be approximated by a linear combination of a small number of dictionary atoms $\boldsymbol{d_k}$. We refer to this approach as sparse dictionary reconstruction.

3.4 Solving Tikhonov regularized inverse problems

We have formulated the AIM problem as optimization problems using Tikhonov functionals. In the following paragraphs, we focus on efficient methods to solve problems (5), (6) and (7).

Henceforth, we only consider linear forward models $\mathbf{F}(\boldsymbol{x}) := \mathbf{A}\boldsymbol{x}$. Nonlinear forward models require additional properties for the existence of a minimizer and might have local minima. They are typically addressed by solving a sequence of linearized problems. Theory for nonlinear inverse problems is still an active field of study. Jin and Maass (2012) summarize the basic results.

For linear forward models the classical Tikhonov functional, Eq. (5), is strictly convex and thus a unique global minimizer exists. The optimization problem can be solved by exerting the necessary conditions of first order, i.e. setting its derivative equal to zero. This leads to the linear equation

$$(\mathbf{A^t}\mathbf{L}_\delta^\mathbf{t}\mathbf{L}_\delta\mathbf{A} + \alpha\mathbf{L}_\mathbf{a}^\mathbf{t}\mathbf{L}_\mathbf{a})\boldsymbol{x} = \mathbf{A^t}\mathbf{L}_\delta^\mathbf{t}\mathbf{L}_\delta\boldsymbol{y_\delta} + \alpha\mathbf{L}_\mathbf{a}^\mathbf{t}\mathbf{L}_\mathbf{a}\boldsymbol{x_a}, \tag{8}$$

whose solution is the minimizer of problem (5). A variety of methods exist to solve this linear equation. Our choice is a conjugate gradient method. Note the similarity to Bayesian inversion when $\mathbf{L}_\delta^\mathbf{t}\mathbf{L}_\delta = \mathbf{R}^{-1}$ and $\alpha\mathbf{L}_\mathbf{a}^\mathbf{t}\mathbf{L}_\mathbf{a} = \mathbf{Q}^{-1}$.

For the sparse reconstruction problem, Eq. (6), the functional to minimize is only convex and no longer differentiable everywhere. For this kind of problem, subgradient methods can be applied to find a minimizer. A fundamental contribution was the Iterative Shrinkage Thresholding Algorithm (ISTA) (Daubechies et al., 2004), which is a simple iterative scheme consisting of a gradient and a shrinkage step

$$\begin{aligned} \boldsymbol{x_{k+1}} &= \mathbf{S}_{\alpha\beta}\left(\boldsymbol{x_k} - \beta\mathbf{A^t}\mathbf{L}_\delta^\mathbf{t}\mathbf{L}_\delta(\mathbf{A}\boldsymbol{x_k} - \boldsymbol{y_\delta})\right) \\ \mathbf{S}_\lambda(\boldsymbol{x})_i &:= \max(|x_i| - \lambda, 0)\operatorname{sign}(x_i). \end{aligned} \tag{9}$$

The stepsize β has to be chosen such that $0 < \beta < 2/\|\mathbf{L}_\delta\mathbf{A}\|^2$. The gradient step adds a non-sparse update to the current iterate. Then, the shrinkage operator $\mathbf{S}_\lambda : X \to X$ shrinks the updated parameters componentwise by λ towards zero. This way, only dominant components may increase and differ from zero. The algorithm converges rather slowly, but by now more sophisticated algorithms have been developed (Elad, 2010; Loris, 2009). We use the Fast Iterative Shrinkage Thresholding Algorithm (FISTA) (Beck and Teboulle, 2009).

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

The sparse dictionary reconstruction problem (7) translates into the sparse reconstruction problem (6) by defining $\tilde{\mathbf{A}} := \mathbf{AD}$ and $\tilde{\boldsymbol{y}}_{\boldsymbol{\delta}} := \boldsymbol{y}_{\boldsymbol{\delta}} - \mathbf{A}\boldsymbol{x}_{\boldsymbol{a}}$. This reformulation allows the use of the methods stated.

3.5 Bounds on the parameters

Some problems require a bound on the parameter space, so $\boldsymbol{x} \in M \subset X$. Examples are nonnegative physical quantities like concentrations. We will only consider a positivity constraint here, but our approach works for general closed convex subsets. When enforcing positivity, we use iterative methods to solve the optimization problems (5), (6) and (7) and couple the update scheme $\mathbf{T}$ with a projection step $\mathbf{P_M}$ onto the set of permitted parameters M;

$$\boldsymbol{x_{k+1}} = \mathbf{P_M}(\mathbf{T}(\boldsymbol{x_k})).$$

For a positivity constraint, the projection is straightforward when the iteration is carried out in the state space X by setting all negative parameters to zero. However, when sparsity is assumed in a different representation system (e.g. a dictionary), it can be complicated to translate these constraints to the corresponding space. For our sparse dictionary reconstruction, we have to calculate the corresponding parameters $\boldsymbol{x_k}$ to the current iterate $\boldsymbol{c_k}$ before projecting, so $\boldsymbol{x_k} = \mathbf{D}\boldsymbol{c_k}$. Then the projection can be performed, $\boldsymbol{x_k^+} = \mathbf{P_M}(\boldsymbol{x_k})$, and we have to translate back into dictionary space. Note, that there are infinitely many representations for the same state $\boldsymbol{x}$. Choosing the sparsest one means solving

$$\boldsymbol{c_{k+1}} = \arg\min_{\boldsymbol{c}\in C} \frac{1}{2}\|\mathbf{D}\boldsymbol{c} - \boldsymbol{x_k^+}\|_2^2 + \alpha\|\boldsymbol{c}\|_1. \tag{10}$$

The problem can be solved using the iterative shrinkage algorithm (see Eq. (9)). We can speed up the convergence with a good inital value, which is given by the current iterate $\boldsymbol{c_k}$. Moreover, the iteration does not need to run until convergence is reached as the outcome will be changed in the next update step anyway. Still, solving Eq. (10) for each iteration of the update scheme is a costly operation.

The projection step for the dictionary is so difficult because the dictionary $\mathbf{D}$ is not invertible. We suggest the following heuristic approach; we select a subset of atoms from the dictionary that form a basis. The projection update is then only calculated for these components. This might damage the sparsity of the current iterate. However, sparsity will be created by the next shrinkage step, if the update by the projection was not too large. It is important to note that this idea is heuristic, meaning that the algorithm may not converge against a minimizer of problem (7) in some cases.

3.6 Link to Bayesian inversion

The methods presented in this paper are formulated as Tikhonov regularizations. The inverse modeling community might be more familiar with the statistical formulation, namely Bayesian inversion. In the following, we briefly describe how both formulations overlap.

In a Bayesian inversion setup, noise and unknown parameters are assumed to be realizations of known probability distributions. Given these distributions and the forward model, Bayes' theorem is used to infer the a posteriori distribution. The

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

maximizer of the posterior probability density function, called maximum a posteriori solution, is often presented as a best estimate. Further evaluation of the posterior distribution also yields uncertainty bounds for the estimate.

We previously mentioned that covariance matrices for noise or prior translate into weighting matrices for the norms in the Tikhonov formulation. For proper weighting matrices Tikhonov regularization with 2-norm penalty as in Eq. (5) is equivalent to a Gaussian prior and a Gaussian noise model. By contrast, Tikhonov regularization with 1-norm penalty from Eq. (6) translates into a Laplacian prior and Gaussian noise. The probability density functions for Gaussian and Laplacian distributions are shown in Fig. 5. The Tikhonov approach only aims at the calculation of a best estimate, which compares to the maximum a posteriori solution in the Bayesian approach. Uncertainties can be assessed by additional calculations, which we will present in Sect. 3.7.

Inversions with non-Gaussian priors, like the Laplacian in Eqs. (6) and (7), rarely have an analytical solution simplifying the calculation of the posterior distribution. Thus, the computational cost for Bayesian inversion methods might become intractable. Tikhonov methods calculate a best estimate without further information about the underlying distribution. This property makes them more suitable for computationally demanding nonlinear or large scale problems, if no uncertainty analysis is carried out.

3.7 Error and sensitivity analysis

To judge the quality of an estimate, it is necessary to know the uncertainty associated with each estimated parameter. For Bayesian methods, these uncertainties come along with the calculation of a best estimate. For the Tikhonov methods used in this work, uncertainty estimates are an extra calculation performed after the retrieval of a best estimate. In this section, we present an uncertainty analysis for Tikhonov methods based on Rodgers (2000, ch. 3).

We call the true parameters $\boldsymbol{x}^+$ and the a priori $\boldsymbol{x_a}$ (we had $\boldsymbol{x_a} = \mathbf{0}$). Let $\mathbf{R}$ denote a general inversion method, $\mathbf{R} : Y \to X$, and $\mathbf{F}$ a general forward model, $\mathbf{F} : X \to Y$. Then, the best estimate is given by

$$\boldsymbol{x}^* = \boldsymbol{x_a} + \mathbf{R}(\boldsymbol{y_\delta} - \mathbf{F}(\boldsymbol{x_a})) = \boldsymbol{x_a} + \mathbf{R}(\mathbf{F}(\boldsymbol{x}^+) - \mathbf{F}(\boldsymbol{x_a}) + \boldsymbol{\delta}).$$

If we linearize the forward model and reconstruction method, we have in first order

$$\boldsymbol{x}^* = \boldsymbol{x_a} + \frac{\partial \mathbf{R}}{\partial \mathbf{y}} \frac{\partial \mathbf{F}}{\partial \mathbf{x}} (\boldsymbol{x}^+ - \boldsymbol{x_a}) + \frac{\partial \mathbf{R}}{\partial \mathbf{y}} \boldsymbol{\delta}. \tag{11}$$

For linear forward models we have $\frac{\partial \mathbf{F}}{\partial \mathbf{x}} = \mathbf{A}$, but the reconstruction methods remain nonlinear. Thus the analysis depends on the point of linearization. The total error can be differentiated between smoothing and (total) measurement error:

$$\boldsymbol{x}^* - \boldsymbol{x}^+ = \underbrace{\left(\frac{\partial \mathbf{R}}{\partial \mathbf{y}} \frac{\partial \mathbf{F}}{\partial \mathbf{x}} - \mathbf{I} \right) (\boldsymbol{x}^+ - \boldsymbol{x_a})}_{\text{smoothing error}} + \underbrace{\frac{\partial \mathbf{R}}{\partial \mathbf{y}} \boldsymbol{\delta}}_{\text{measurement error}}. \tag{12}$$

The measurement error describes how noise on the measurement data propagates to errors in the estimated parameters. Recall that our definition of noise includes errors in the measurement, the forward model and numerical approximations. Reconstruction methods try to suppress the effect of noise on the parameters by stabilizing the unstable inversion (e.g. via a penalty term or an a priori). This modification introduces the smoothing error. For ill-posed problems, a smaller smoothing error results in a greater measurement error and vice versa. The smallest total error is expected when both terms are balanced.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions

Generally, the terms in Eq. (12) are impossible to calculate as they require knowledge of the true solution $\boldsymbol{x}^+$ as well as the noise $\boldsymbol{\delta}$. We assume to have access only to the noise characteristics, which is common case in real data problems. In a Bayesian inversion, additional assumptions on the true solution are made to derive a posteriori uncertainties.

For our Tikhonov based methods, the measurement error can be estimated by sampling the noise and calculating its effect on the parameters. The smoothing error can be assessed via the sensitivity matrix $\mathbf{S} := \frac{\partial \mathbf{R}}{\partial \mathbf{y}} \frac{\partial \mathbf{F}}{\partial \mathbf{x}}$, sometimes called the averaging kernel matrix. The smoothing error vanishes for an ideal sensitivity matrix equal to the identity. Thus, the closer the sensitivity matrix is to the identity, the smaller smoothing error can be expected. We explain further details on how we analyze the sensitivity matrix in Sect. 4.6.

4 Case study: Methane emissions in the United States

Our goal is to apply the sparse dictionary reconstruction method in an atmospheric inverse modeling setup. We use anthropogenic methane emissions in the United states as a synthetic case study, the same case study used in Miller et al. (2014). This setup gives us the opportunity to compare our flux estimates not only to the known synthetic fluxes but also to the estimates by state-of-the-art methods already used in atmospheric inverse modeling. This section describes the details of the case study and the methods we use. Before we specify how the sparse dictionary reconstruction method is set up, we compare Tikhonov regularization with 2-norm and 1-norm penalty to visualize the effect of the 1-norm.

4.1 Case study details

Our study area is North American mainland at $25 - 55^\circ$ N and $145 - 51^\circ$ W. Emissions are estimated on a 1° by 1° grid and limited to land grid cells. We use a combination of synthetic in-situ aircraft and tall tower measurements, that were available during May to September 2008 from operations by the NOAA Earth Systems Research Laboratory (NOAA; Andrews et al., 2014), the United States Department of Energy (Biraud et al., 2013) and the START08 aircraft campaign (Pan et al., 2010). Footprints for these measurements, that define the forward model, are calculated using the WRF-STILT model (see Sect. 2).

Realistic synthetic methane emissions are generated from the Emission Database for Global Atmospheric Research (EDGAR). We project anthropogenic methane emissions from the EDGAR v3.2 FT2000 inventory (Olivier and Peters, 2005) onto our model grid. These emissions are constant in time during our observation period. As discussed in Miller et al. (2014), inventories have been updated, but we use this version for reasons of comparison to the previous study. Moreover, we work with simulated data only, so there is no strict need to use the most recent inventory version. Rather the solution should comprise typical features, which we assume holds for this version as well.

The simulated noisy measurements are calculated by employing the linear WRF-STILT forward model on the EDGAR fluxes and adding Gaussian noise of realistic magnitude. The noise vector is sampled from the multivariate Gaussian distribution with diagonal covariance matrix, that was estimated for the real data of these measurements by restricted maximum likelihood estimation in Miller et al. (2013).

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

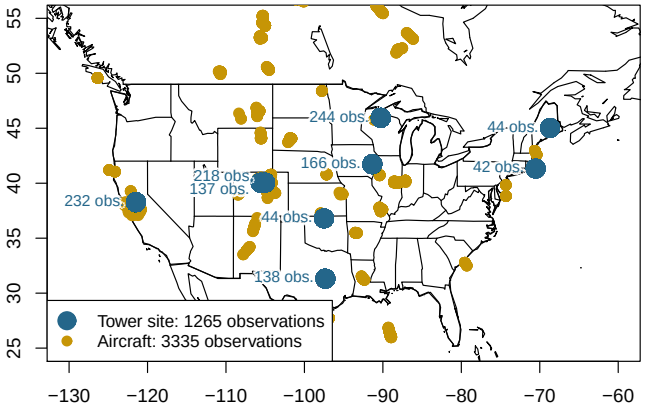

Figure 2. Available in-situ methane measurements for May to September 2008

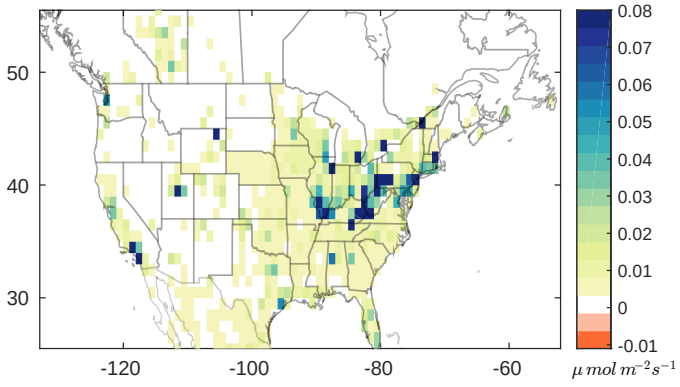

Figure 3. US methane emissions from the EDGAR v3.2 FT2000 in the native 1° by 1° resolution. The largest source regions, New York and Eastern Kentucky, have fluxes higher than three times the limit of the colormap.

4.2 Classical Tikhonov regularization vs. sparse reconstruction

We have one unknown emission parameter per land grid cell in our setup because we use temporally constant fluxes, so $\boldsymbol{x} \in \mathbb{R}^{1469}$. On the other hand, we have a total of 4600 measurements, so $\boldsymbol{y_\delta} \in \mathbb{R}^{4600}$. Despite the fact that there are more measurements than unknowns, the inverse problem is still ill-posed as many measurements yield similar pieces of information. We estimate the surface fluxes x only by knowledge of the forward model $\mathbf{A}$ defined by the footprints, the noisy measurements $\boldsymbol{y_\delta}$ and the noise characteristics determined by the noise covariance matrix $\mathbf{R}$. In this scenario $\mathbf{R}$ does not have any off-diagonal entries, thus it is easy to calculate $\mathbf{L}_\delta$ from $\mathbf{L}_\delta^t \mathbf{L}_\delta = \mathbf{R}^{-1}$. In general cases we recommend using the Cholesky factorization. For all problems we use a zero a priori, so $\boldsymbol{x_a} = \mathbf{0}$. In real data scenarios more advanced a priori models should be considered.

We start by comparing Tikhonov regularization with classical 2-norm penalty

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{L}_\delta(\mathbf{A}\boldsymbol{x} - \boldsymbol{y_\delta})\|_2^2 + \frac{\alpha^*}{2}\|\boldsymbol{x}\|_2^2 \tag{L2}$$

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions

and Tikhonov regularization with sparsity constraint

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{L}_\delta(\mathbf{A}\boldsymbol{x} - \boldsymbol{y_\delta})\|_2^2 + \alpha^*\|\boldsymbol{x}\|_1. \tag{L1}$$

The optimal regularization parameter α^* is approximated by Morozov's discrepancy principle for each problem; we start with a value α_0 that is certainly too large, so the corresponding minimizer is the a priori $\boldsymbol{x_a}$, here $\boldsymbol{x_a} = \mathbf{0}$. Then, the regularization parameter is reduced iteratively by $\alpha_k = q\alpha_{k-1}$ with $0 < q < 1$ and the corresponding minimizer $\boldsymbol{x_{\alpha_k}}$ is calculated. For each minimizer we check whether

$$\|\mathbf{L}_\delta(\mathbf{A}\boldsymbol{x_{\alpha_k}} - \boldsymbol{y_\delta})\|_2 < \tau\bar{\delta}, \tag{13}$$

where $\bar{\delta} = \|\delta\|_2$ is the expected noise-level and $\tau > 1$. If it holds, an appropriate regularization parameter $\alpha^* := \alpha_k$ is found and the corresponding minimizer $\boldsymbol{x_{\alpha^*}}$ is our best estimate $\boldsymbol{x}^*$. In Eq. (13) we have $\bar{\delta} = \sqrt{m}$, $m = 4600$ as the noise is normalized by $\mathbf{L}_\delta$.

For a fixed α-value we use a conjugate gradient method to solve problem (L2) via Eq. (8). We included a speed up by Frommer and Maass (1999), which detects in early iterations whether or not Morozov's criteria, Eq. (13), can be reached for the given parameter. This allows to continue with the next smaller regularization parameter α before convergence is reached.

To determine the sparse reconstruction solution, Eq. (L1), to a fixed α-value we use FISTA, which is an accelerated version of ISTA (see Sect. 3.4). Instead of calculating the update based on the last iteration only, it uses a weighted combination of the last two iterates.

4.3 Preliminary results: Classical Tikhonov regularization vs. sparse reconstruction

Figure 4 shows the methane emission estimates by the Tikhonov methods L2 and L1. Small sinks appear in both estimates because both methods are not restricted to positive emissions. We should mention that our measurements include some negative concentrations, which is due to the fact that they are enhancements above a background level and were perturbed by noise. But even with nonnegative measurements only, we could not expect to have no sinks in the estimates, if not explicitly enforced. An overestimation in one grid cell and an underestimation in another might still be consistent with the data because the data are not sufficient to fully constrain all flux parameters in an ill-posed inverse problem.

Furthermore, we observe that methane emissions are more localized, but greater in magnitude for L1. This reflects the properties of sparse reconstruction: Parameters that are estimated to be small in L2 receive more penalization by the sparsity constraint and are thus forced to zero. The data has to be explained by the remaining nonzero paramaters.

A comparison against the true fluxes shows that especially large emitters such as cities are much better reconstructed by L1. At the same time, L1 estimates regional sources as observed in the Southeast and Midwest as point sources and neglects many small sources. A grid cell by grid cell comparison slightly favors the L1 estimate but the total emissions are better estimated by L2 (see Table 1).

The histogram of the EDGAR fluxes in Fig. 5 supports the use of the sparsity constraint. As fluxes are nonnegative, in contrast to the a priori models of L2 and L1, we assigned a random sign to each flux after cutting off the largest 5% to avoid empty bins. The resulting empirical distribution agrees much better with a Laplacian than with a Gaussian prior.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

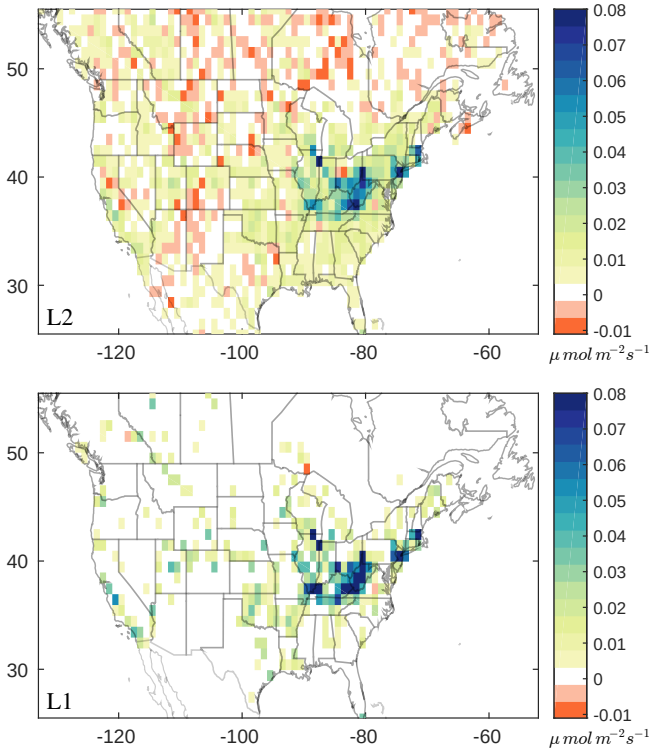

Figure 4. Emission estimates using Tikhonov regularization with classical 2-norm penalty (see Eq. (L2)) and sparsifying 1-norm penalty (see Eq. (L1)) inverted from noisy simulated methane measurements. The true flux field is shown in Fig. 3. While L2 shows the typical smoothing effect, L1 concentrates the signal, which results in better estimates of large sources, but also tends to explain regional emissions by larger point sources.

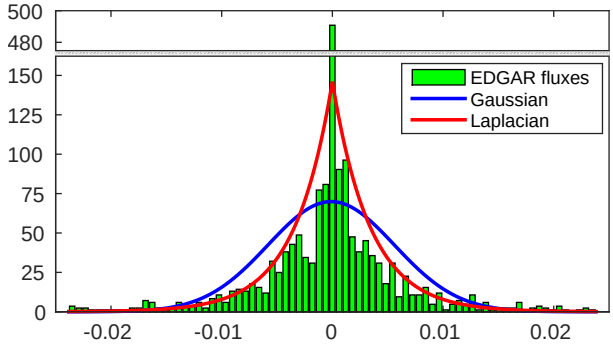

Figure 5. Normalized histogram of randomly signed EDGAR fluxes, ignoring the largest 5% of emissions. The histogram data has been used to estimate the parameters of corresponding Gaussian and Laplacian distributions. Even though the center bin is largely populated only less than 10% of all fluxes are zero.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

4.4 Sparse dictionary reconstruction

Sparse reconstruction in the standard representation system showed some improvements when looking at the estimate of large sources. However, we also observed that regional sources are likely incorrectly represented as point sources and that the total emissions are underestimated. Our solution is not naturally sparse because we expect methane emissions in most grid cells. However, L1 always heads for sparse solutions. Consequently, we have to find a representation system that is able to sparsely approximate all realistic emission fields.

We decided to use a dictionary. We therefore need to select atoms, such that the dictionary can sparsely approximate all methane emission patterns. Various dictionary learning algorithms are available, but a set of training data is required (see e.g Mairal et al., 2014). We could extract training data from the EDGAR inventory for other regions, learn a dictionary and use it for the United States setup, but results could be too optimistic as this will not be an option for real data scenarios. Our approach will be to identify typical source shapes and include these shape functions as atoms in our dictionary.

The $1°$ by $1°$ model grid is too coarse to identify individual sources. Many typical methane sources such as cities, landfills and waste, industrial facilities and mining all do not extend more than a grid cell. To be able to represent these efficiently, we include the pixel basis in our dictionary. Metropolitan areas, livestock areas, oil and gas fields might extend over several pixels though. Thus, we also add circular peak shape functions (see Fig. 6). We could add more functions with bigger and more complicated shapes, but this does not only require more computational time to solve the problem, but also leads to more redundancy. It is our intention to create some redundancy but only to the degree that it helps sparsify the representation of our possible emission fields. From numerical experiments we found that including bigger shape functions does not add value to the reconstruction.

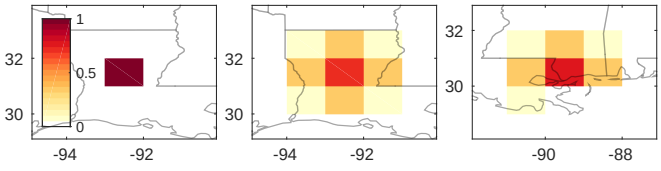

Figure 6. A selection of atoms from the dictionary used for the sparse dictionary reconstruction method. These atoms are used to represent the state vector via linear combination. The left and the middle element are the basic shapes centered in each grid cell of the domain. At coasts and lakes these shapes are limited to land grids. All atoms are normalized in the 2-norm. The dictionary chosen here also holds a constant background function.

Another idea to sparsify the representation is to look at the whole domain. A background is best represented by a constant function. With the same argument we could add regional background functions. Actually, we found that a division into regions as shown in Fig. 9 would improve the estimate, but the placement of those regions was partly inspired by looking at the true EDGAR fluxes.

Michalak et al. (2004) suggested a geostatistical inversion as an extension of L2, where a key concept is the model of the mean. It spatially correlates regions based on geostatistical information such as population density or agricultural use.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

This basically results in shape functions, which we could include in our dictionary as well. The difference between both approaches is that our parameters receive penalization, whereas they are unconstrained in the geostatistical inversion. Weighting the parameters accordingly would translate from one approach to the other.

For our experiments we decided not to include atoms that were constructed from EDGAR or geostatistical data. We will use a pixel basis, a basis with peaks that extend into the direct neighbors (see Fig. 6) and a background function for the entire domain. Thus, we have $\mathbf{D} \in \mathbb{R}^{n \times N}$ with $N \approx 2n+1$. All atoms are normalized in the 2-norm. Generally speaking, our dictionary holds functions to represent processes at different spatial scales.

To estimate the flux parameters x with sparse dictionary reconstruction we solve

$$\boldsymbol{c}^* = \arg\min_{\boldsymbol{c}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{L}_\delta(\mathbf{A}(\mathbf{D}\boldsymbol{c}+\boldsymbol{x_a}) - \boldsymbol{y_\delta}\|_2^2 + \alpha^*\|\boldsymbol{c}\|_1$$

$$\boldsymbol{x}^* = \mathbf{D}\boldsymbol{c}^*. \qquad \text{(L1 DIC)}$$

As before, we use a zero a priori, $\boldsymbol{x_a} = \mathbf{0}$, the discrepancy principle, Eq. (13), to determine the optimal regularization parameter α^* and FISTA to solve problem (L1 DIC) for a given α. A pseudocode is included in the supplementary information.

4.5 Enforcing positive fluxes

For further analysis, we add a positivity constraint on the flux parameters, $\boldsymbol{x} \in R_+^n$, which we denote by the suffix POS. We use the projection approach described in Sect. 3.5 for all methods to enforce positivity. Projecting the iterates of a conjugate gradient method may lead to poor performance, as the special structure of the search directions is lost. Thus, we also use FISTA with a shrinkage operator for the 2-norm penalty when solving L2 POS. Note that the projection step for L1 DIC POS is more demanding as it involves the transition from parameter to dictionary space.

4.6 Error and sensitivity analysis

We carry out the uncertainty analysis based on knowledge about the noise characteristics but without knowledge about the true fluxes as would be the case for real data scenarios. As discussed in Sect. 3.7, we assess smoothing and measurement error separately.

The smoothing error is only known if additional assumptions on the true fluxes are made. The best we can do is to analyze the sensitivity matrix; the k-th column of this matrix expresses how the estimate reacts on a perturbation in the k-th parameter of the true fluxes $\boldsymbol{\Delta x_k}$, so

$$\mathbf{S}_{:,\mathbf{k}} := \frac{\mathbf{R}(\boldsymbol{y_\delta} + \mathbf{A}\boldsymbol{\Delta x_k}) - \mathbf{R}(\boldsymbol{y_\delta})}{\|\boldsymbol{\Delta x_k}\|}. \qquad (14)$$

Note that the calculation of the sensitivity matrix can be carried out without knowledge of the true fluxes. The numerical computation requires to solve one reconstruction problem per parameter, which can be done in parallel, but might still be infeasible for large scale problems.

As mentioned before, an ideal sensitivity matrix is equal to the identity. Due to the nonlinearity of the reconstruction methods, interpretation of the sensitivity matrix is difficult. The information is only local and cannot predict reactions of the estimate for

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

perturbations different from $\boldsymbol{\Delta x_k}$ via Eq. (11). Even the normalization in Eq. (14) might be misleading, if the amplitude of the perturbation $\|\boldsymbol{\Delta x_k}\|$ is not provided. However, sensitivities are still meaningful for assessing the accuracy of the estimate.

We address two measures: The column sum of the sensitivity matrix should be close to one. Otherwise, it indicates that the method over- (> 1) or under estimates (< 1) in that region. The diagonal of the sensitivity matrix shows the sensitivity of the parameter that is perturbed. Values close to one indicate high confidence in the reconstruction. Smaller values are either a consequence of smoothing or of not being sensitive at all. The latter is captured by looking at the column sum as well.

The measurement error shows the influence of the noise on the estimated parameters. Statisical quantities like the standard deviation can be approximated via resampling of the noise and recalculation of the estimate under this noise for a sufficient number of samples. This numerical approach is computationally demanding, but can also be run in parallel. If the noise characteristics were unknown, resampling could be achieved by bootstrapping of the residual of the estimate (see Banks et al., 2010).

4.7 Comparison to other methods

We compare our approaches to state-of-the-art methods studied in Miller et al. (2014). The scope of the article was to analyze different formulations to enforce positive parameters. The methods are:

- Standard inversion: This is a geostatistical approach following Michalak et al. (2004). It does not include a positivity constraint and was taken as a benchmark method in Miller et al. (2014).
- Transform inversion: Flux parameters are enforced to be positive by a power transformation (see Snodgrass and Kitanidis, 1997). This technique can also be used to straighten skewed parameter distributions.
- Lagrange multiplier method: Positivity is enforced by formulating an optimization problem with inequality constraint, which is solved via the Lagrangian function. As a deterministic method, no direct uncertainty estimates are given, but they can be approximated using the approaches from Sect. 3.7.
- Gibbs sampler: The Gibbs sampler belongs to the group of Markov Chain Monte Carlo (MCMC) methods. These methods can generate realizations of complicated probability distributions such as the posterior distribution to non-Gaussian priors in a Bayesian inversion framework. They differ in the way these realizations are calculated. One can estimate statistical quantities such as mean and standard deviation given a sufficient number of realizations. Positivity is formulated in the prior distribution. In theory, one could implement one of several MCMC algorithms (see Miller et al., 2014), but we focus on the Gibbs sampler here (e.g. Michalak and Kitanidis, 2003).

All methods have been discussed in a Bayesian inversion framework. Further details and references are given in Miller et al. (2014).

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

5 Results

In this section, we analyze the performance of our suggested sparse dictionary reconstruction method L1 DIC POS in the AIM scenario described in the previous section. First, we compare it with the methods L2 POS and L1 POS and carry out an error analysis. Then, we include the methods from Miller et al. (2014) into the comparison. Finally, we analyze the ability of each method to reproduce spatially discrete emissions from oil and gas extraction in the Barnett Shale region of Texas.

5.1 Methane emission estimates

Figure 7 shows the estimated emissions using L2 POS, L1 POS and L1 DIC POS. Our first observation is that there are no sinks anymore as positive fluxes are enforced by all methods. The results for L2 POS and L1 POS are close to the ones obtained by L2 and L1 (see Fig. 4), setting negative fluxes to zero. However, projecting the final estimate (see Sect. 3.5) should be avoided as the mismatch between modeled and measured data increases and thus does not make full use of the information in the data. A projection step in every iteration allows the algorithm to correct for this in the next update. Especially for L2, we measure significant improvement excluding sinks (see Table 1).

In contrast to L1 POS, the solution of L1 DIC POS does not look sparse, as sparsity is enforced on the coefficients of the dictionary. The background function in the dictionary has been selected to represent a base level of small emissions (not visible in the color map). It improves the inversion's ability to accurately estimate total US emissions. Regionally, other atoms have been added and subtracted from this background level. The smooth character of the estimate is a result of the broader dictionary functions (see Fig. 6). L1 DIC POS shows significant improvement in the estimate of localized sources against L2 POS (e.g. when looking at West Coast emissions), but a slight setback against L1 POS. Moreover, the emissions estimate is too spatially diffuse across regions of Canada with point sources. These diffuse patterns may be caused by uninformative footprints in that region.

We observe that the locations of significant sources agree much better with both sparsity methods L1 POS and L1 DIC POS than with the classical L2 approach. This result can be explained by the fact that the sparse schemes look for the dominant sources. Even if the magnitude is not captured exactly, the method L1 might be used in applications to identify source locations. All nonzero components are refered to as the support of the solution. The FISTA method we use iteratively approximates the sparse minimizer by non-sparse iterates. Other methods that solve the sparse reconstruction problem, Eq. (6), iteratively extend their support and calculate a least squares fit on this selected set of parameters to capture the magnitude. Therefore, sparse reconstruction methods tend to better locate sources, when the support is limited.

5.2 Error and sensitivity analysis

As described in Sects. 3.7 and 4.6 we assess smoothing and measurement error separately.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

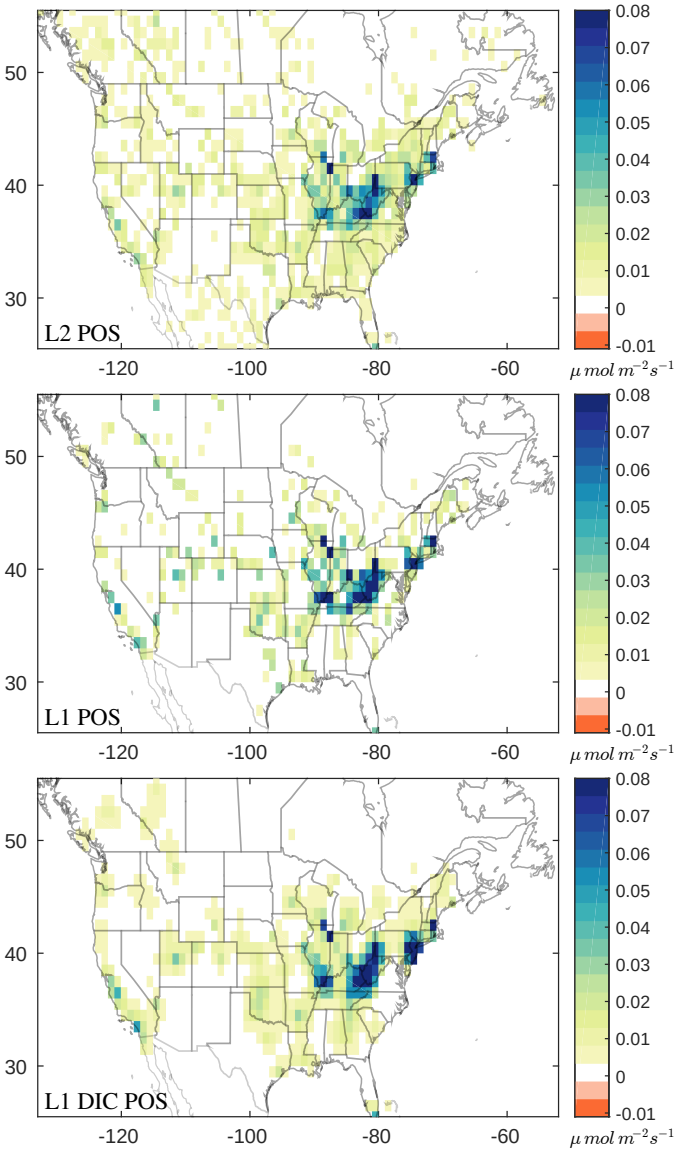

Figure 7. Emission estimates from the methods L2 POS, L1 POS and L1 DIC POS inverted from noisy simulated methane measurements. The true flux field is shown in figure 3.

5.2.1 Smoothing error

Without knowledge of the true fluxes, the sensitivity matrix gives the best insight to the smoothing error. Each column describes, how an additional pixel source would change the flux estimate. A perfect sensitivity matrix is thus equal to the identity. The column sum indicates regions that are over- or underestimated. This can only be observed in regions with small footprint information outside the main study area, namely Florida, Mexico and Central and Eastern Canada. The locations are similar

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

for all methods, but L1 POS is far more biased in those regions. Table 1 shows, that L1 POS indeed poorly estimates the total emissions.

The most valuable information is contained on the diagonal of the sensitivity matrix, plotted in Fig. 8 (left). The diagonal shows the sensitivity of the parameter that is perturbed. Unsurprisingly, all methods are most sensitive in the vicinity upwind of tower observation sites. Also, we observe that both sparse reconstruction schemes are less sensitive than L2 in regions that are poorly constrained by the data and have an increased sensitivity in regions of greater footprint values. The interpretation is a slightly different for both types of methods. If we excluded the positivity constraint, L2 (POS) would be a linear method, meaning that the sensitivity matrix is independent of the parameters and could be used to predict how additional sources would be reconstructed using Eq. (11). For positive parameters, nonlinearity due to the positivity constraint should have little influence. In contrast, the sparsity constraint adds to the nonlinearity. The sensitivity for small sources might be small or even zero in some regions, but large sources or sources in several neighboring grid cells could still be reconstructed. Figure 8 (left) shows the reconstruction ability of a single larger pixel source, as we used a rather large perturbation to approximate the sensitivity matrices.

5.2.2 Measurement error

The uncertainties in the estimates originating from the noise on the data are described by the meausurement error, which we approximated numerically by resampling the noise. Two standard deviations are plotted in Fig. 8 (right). The reason for regularization is to limit the influence of noise on the estimate. However, no influence at all would also mean that the method is not sensitive to data at all. Significant measurement uncertainties are present in all regions for the L2 POS approach, while these uncertainties are more limited to regions with large emissions for the L1 POS and L1 DIC POS approaches.

5.3 Comparison to other methods

In this section, we evaluate the estimates of our Tikhonov based methods by comparing them to the estimates of the methods studied in Miller et al. (2014) (see Sect. 4.7).

5.3.1 General results

We examine the reconstruction quality using several measures, each of which focuses on aspects or qualities. Those are:

$$\text{relative total error} := \frac{\int_\Omega x^+ - \int_\Omega x^*}{\int_\Omega x^+} \tag{15}$$

$$\text{relative regional error} := \frac{\int_\Omega |x^+ * \bigcirc - x^* * \bigcirc|}{\int_\Omega (x^+ * \bigcirc)} \tag{16}$$

$$\text{relative local error} := \frac{\int_\Omega |x^+ - x^*|}{\int_\Omega x^+}. \tag{17}$$

The local error compares estimate and truth grid cell by grid cell while the total error sums up all grid cell emissions to a North American flux before comparison. We chose to include the regional error as an intermediate measure between those two.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

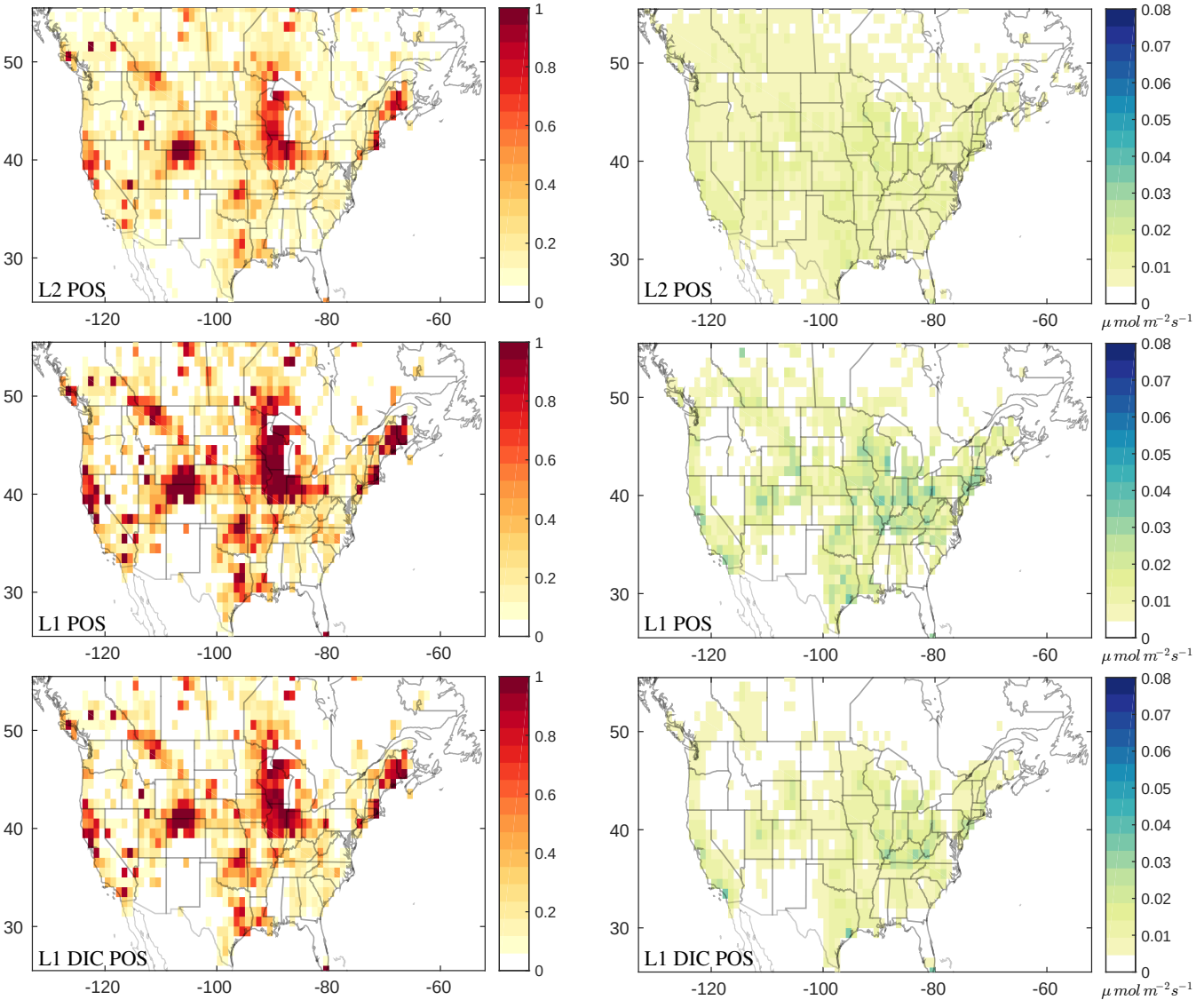

Figure 8. Left: Diagonal of the numerically calculated sensitivity matrices for large deviations of $\|\boldsymbol{\Delta x_k}\| = 0.1\mu$ mol m^{-2} s^{-1}. Right: Two standard deviation uncertainties due to noise on the measurements.

The idea is to see if the total emissions over a region around the grid cell is estimated correctly. This approach relativizes the smoothing effect of the reconstruction methods. Here, $\bigcirc$ is a circular filter that gives a weighted sum of the two neighbouring grid cells; in other words this measure compares smoothed versions of the solution and the estimate.

The results for all estimates are listed in Table 1. We observe that our Tikhonov methods typically underestimate the total emissions, which is expected when taking a zero a priori. Surprisingly, all methods from Miller et al. (2014) overestimate the

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

total emissions, even though $\approx 5\%$ difference can be considered a good result in this setup. Estimates of the total flux from L1 POS and the Gibbs sampler are poor in this scenario.

In the local error measure, which compares grid cell by grid cell, L1 DIC POS and the transform inversion perform best. These methods come closer at addressing questions on the grid cell level but errors are still too high for accurate answers. Reasonable estimates can only be made on a coarser scale by spatially integrating grid cells. The regional measure suggests that L2 POS and the Lagrange Multiplier method also perform well on a coarser grid.

From a modeling perspective the standard inversion is comparable to our method L2, whereas the Lagrange multiplier method and Gibbs sampler include positivity constraints and compare to L2 POS. The estimates show similar features to our estimates for L2 and L2 POS (see the supplementary information), namely rather smooth emission estimates. The spatial correlation between parameters used by Miller et al. (2014) adds to the smoothness. As already discussed for our methods, large pixel sources such as cities appear more as regional sources in the estimate. That is why these methods do not perform well on a grid cell level.

Our sparse dictionary reconstruction method and the transform inversion both estimate parameters in a different space, but the transforms are fundamentally different. For L1 DIC POS, the sparsity constraint and the dictionary with the pixel elements promote the estimation of pixel sources. For the tranform inversion the nonlinear mapping between coefficient space and parameter space allows larger pixel emissions than the smoothing methods.

Table 1. Reconstruction errors measured on a local, regional and total scale (see Eqs. (15) - (17)). For the total, a negative sign means overestimation. The regional and local meausures are always positive. All measures are relative and thus without unit.

method	rel. total	rel. regional	rel. local
L2	0.113	0.398	0.945
L2 POS	0.076	0.351	0.815
L1	0.295	0.411	0.862
L1 POS	0.299	0.423	0.864
L1 DIC POS	0.055	0.353	0.727
Standard Inv.	−0.052	0.508	1.004
Transform Inv.	−0.057	0.384	0.683
Lagrange Mult.	−0.053	0.376	0.827
Gibbs Sampler	−0.273	0.520	0.957

5.3.2 Regional emission estimates

A common task is to determine the total emissions for a political or geographic region. Thus, we divided the domain into ten regions, mainly along political borders. The flux estimates for these regions are shown in Fig. 9. We see that the estimates of all methods agree for the central regions, while there are large differences in regions like Mexico and Canada. This is a consequence of data availability. As many measurement stations are located in the central regions the associated parameters are

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

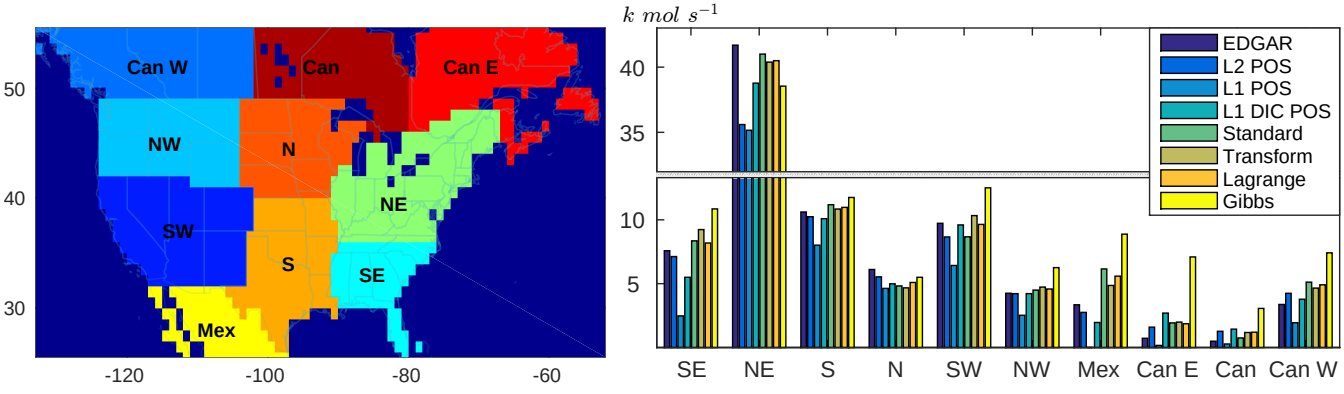

Figure 9. Regional EDGAR emissions and emissions estimates for the methods studied in Miller et al. (2014) and the Tikhonov reconstruction methods studied here.

rather well constrained by the data, whereas the formulation of the a priori knowledge determines the parameters in regions with fewer observations. This results in underestimation for L2 POS and L1 POS as parameters are forced to be small respectively equal to zero. For L1 DIC POS, the background function and the broader peak shape functions included in the dictionary are used to describe the emissions in poorly-constrained regions and allow a proper estimate of the regional fluxes. Along with the transform inversion, L1 DIC POS best estimates the total flux of a region.

5.3.3 Case study: Methane emissions from the Barnett

In a final scenario, we test the reconstruction quality of our methods for methane emissions from unconventional gas wells. We chose the Barnett shale formation in Texas because it had the highest production of any US reservoirs in summer 2008. We add a small synthetic source on top of the EDGAR fluxes and simulate noisy measurements. The synthetic emissions are inspired by the location of the formation and a recent map of well distribution (see Karion et al., 2015). The magnitude of the emissions is roughly calculated from the 2008 production rate of 85 Million $\mathrm{m}^3\ \mathrm{y}^{-1}$ and a leakage of about 1.5% as estimated by Zavala-Araiza et al. (2015). It is not our aim to have the most accurate emissions but to analyze the potential of the methods.

The plots in Fig. 10 show the change in the estimates induced by this additional source. Table 2 states the numbers for the spatially integrated flux change over the Barnett and the overall flux change. First, we observe that all methods underestimate the Barnett emissions. The reason for this underestimate is that all methods have low to middle sensitivity in that region (see Fig. 8), which is a consequence of data availability. However, methods differ largely in the estimated magnitude of these emissions. Best results are achieved by the transform inversion and L1 DIC POS. The other methods might adjust the total emissions adequatly but have problems attributing these emissions to the Barnett. This result suggests that classical approaches lead to excessive smoothing for this application. We should add that this scenario has not been designed to favor one of these methods. The source shape or magnitude cannot be represented by a single atom in the dictionary. However, if parameter correlating source shapes like the distribution of wells are available, L1 DIC POS will benefit from this knowledge.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

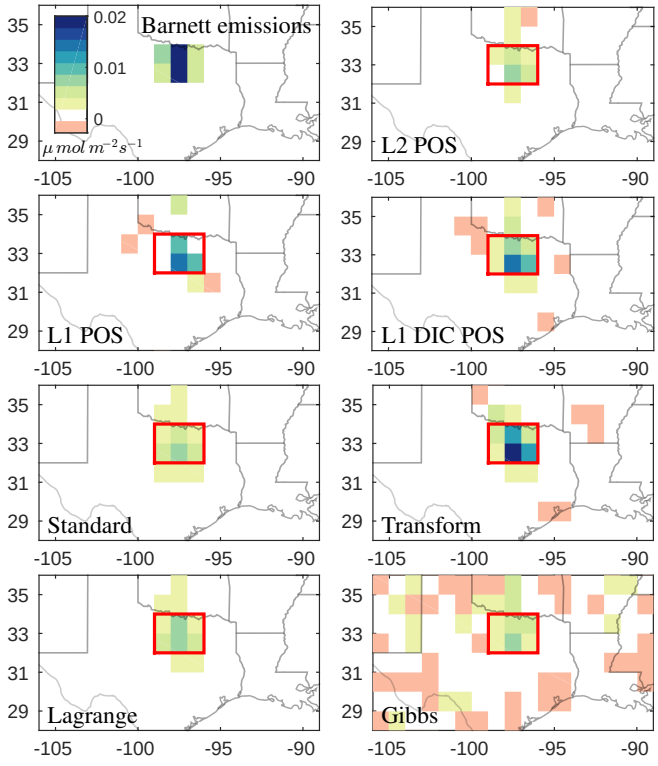

Figure 10. Additional methane sources in the Barnett shale gas reservoir (upper left) are added to the EDGAR emissions (see Fig. 3) and noisy data is simulated. Differences to previous reconstructions from simulated EDGAR data are shown for each method. Emissions in the red box are attributed to the Barnett.

6 Conclusions

This study analyzes different methods to solve inverse problems. We introduce Tikhonov regularization with the commonly used 2-norm and the sparsifying 1-norm penalty function. We show how these approaches translate to a Gaussian and a Laplacian prior, respectively, in a Bayesian inversion framework. We present a new sparse reconstruction method, that enforces sparsity in a redundant dictionary representation system taylored to this application. A simple heuristic approach is applied to all methods to force nonnegative parameters. To test our methods we consider an atmospheric inverse modeling scenario, in which we estimate methane surface fluxes for the United States from atmospheric in-situ measurements.

We find that the choice of the penalty term has a significant influence on the estimate and is thus a crucial step when solving inverse problems. Gaussian-like priors such as the 2-norm penalty in Tikhonov regularization produce a smoothing effect. In our scenario this means that large localized sources such as emissions from cities cannot be estimated accurately. Instead, they appear more as regional sources. In constrast, the sparse reconstruction approach can reproduce these large emitters, but it also suppresses too many small emissions to properly estimate the total flux. However, we find a simple dictionary representation

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

Table 2. Results for the Barnett scenario: Estimated emissions in the Barnett region (red boxes in Fig. 10) and total flux change induced by the additional source for the methods of this study and the previous study by Miller et al. (2014).

method	Barnett [mol s^{-1}]	tot. flux [mol s^{-1}]
L2 POS	245.69 (37.3%)	587.55 (89.3%)
L1 POS	340.06 (51.7%)	562.73 (85.5%)
L1 DIC POS	448.32 (68.1%)	647.93 (98.5%)
Standard Inv.	323.52 (49.2%)	625.19 (95.0%)
Transform Inv.	580.59 (88.4%)	631.76 (96.0%)
Lagrange Mult.	318.63 (48.4%)	624.35 (94.9%)
Gibbs Sampler	262.26 (39.9%)	809.23 (123.0%)*
true fluxes	657.99 (100.0%)	657.99 (100.0%)

* The mean is approximated from a limited number of random samples from the a posteriori distribution. Thus, it slightly differs with every restart of the Gibbs sampler.

system that is able to sparsely approximate the emission field. Our resulting sparse dictionary reconstruction method works equally well as established methods in determining the overall flux field and adds information on the local scale.

The Barnett case study shows the importance of such local information: While the smoothing methods recognize the additional emissions in the total flux, they cannot attribute these to the Barnett. Our sparse dictionary reconstruction method and the transform inversion studied in Miller et al. (2014) perform much better in localizing these emissions. This suggests that the standard Gaussian prior is too prohibitive towards large emitters in this application and more sophisticated models are required.

As concluded in the previous study by Miller et al. (2014), we can confirm that the positivity constraint on the flux parameters further improves the estimate. For our Tikhonov based methods, we find a heuristic approach to meet these constraints by using an iterative solver in combination with a projection step. Our iterative methods are also well-suited for large scale problems as they avoid costly numerical operations. However, an error analysis might be intractable for very large problems.

Obviously, the sparsity constraint works best when the underlying signal is sparse or can be sparsely approximated. The representation of the signal in a dictionary is very flexible and can create a sparse signal for many applications. Our sparse reconstruction method is thus applicable to any inverse problem, but the dictionary would need to be adapted to suit the application. For some applications, sparsifying transforms or training data to learn a dictionary might be available. In others, finding a sparsifying dictionary might be a challenge on its own. We construct the dictionary using a simple approach. The estimate can be further improved by using spatial information about sources encoded in shape functions.

In summary, the sparse reconstruction approach here is a good alternative to commonly used Gaussian priors when the emission field has many point sources or heterogeneous spatial structure. The combination of a sparsifying dictionary representation system and sparse reconstruction is a powerful tool for many inverse modeling applications.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

7 Code and data availability

The numerical methods and the case study data are available for download. The code is written in Matlab 2014b. See the supplement for more information.

Author contributions. N. H. carried out the numerical experiments, prepared and finalized the article. S. M. provided the experimental framework and was involved in the finalization of the paper. P. M., J. N., M. P. and T. W. supervised the project from a mathematical respectively an environmental physics point of view.

Acknowledgements. The work was funded by the Center for Industrial Mathematics of the University of Bremen. The collaboration was supported by a research scholarship by the Deutscher Akademischer Austauschdienst (DAAD).

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-256, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 14 November 2016

References

Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., Chao, D. Y., Kitzis, D. R., Novelli, P. C., Zhao, C. L., Dlugokencky, E. J., Lang, P. M., Crotwell, M. J., Fischer, M. L., Parker, M. J., Lee, J. T., Baumann, D. D., Desai, A. R., Stanier, C. O., De Wekker, S. F. J., Wolfe, D. E., Munger, J. W., and Tans, P. P.: CO_2, CO, and CH_4 measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts, Atmospheric Measurement Techniques, 7, 647–687, doi:10.5194/amt-7-647-2014, 2014.

Banks, H., Holm, K., and Robbins, D.: Standard error computations for uncertainty quantification in inverse problems: Asymptotic theory vs. bootstrapping, Mathematical and Computer Modelling, 52, 1610 – 1625, doi:10.1016/j.mcm.2010.06.026, 2010.

Beck, A. and Teboulle, M.: A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems, SIAM J. Img. Sci., 2, 183–202, doi:10.1137/080716542, 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, Atmospheric Measurement Techniques, 6, 751–763, doi:10.5194/amt-6-751-2013, 2013.

Candès, E. J., Eldar, Y. C., Needell, D., and Randall, P.: Compressed sensing with coherent and redundant dictionaries, Applied and Computational Harmonic Analysis, 31, 59 – 73, doi:10.1016/j.acha.2010.10.002, 2011.

Daubechies, I., Defrise, M., and De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Communications on Pure and Applied Mathematics, 57, 1413–1457, doi:10.1002/cpa.20042, 2004.

Diniz, P. S. R., da Silva, E. A. B., and Netto, S. L.: Digital signal processing: system analysis and design, Cambridge Univ. Press, Cambridge [u.a.], 2. ed. edn., 2010.

Elad, M.: Sparse and redundant representations: from theory to applications in signal and image processing, Mathematics, Springer, New York, NY [u.a.], 2010.

Frommer, A. and Maass, P.: Fast CG-Based Methods for Tikhonov–Phillips Regularization, SIAM Journal on Scientific Computing, 20, 1831–1850, doi:10.1137/S1064827596313310, 1999.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of CO2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, Journal of Geophysical Research: Atmospheres, 108, n/a–n/a, doi:10.1029/2003JD003770, 4757, 2003.

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

Hensen, A., Skiba, U., and Famulari, D.: Low cost and state of the art methods to measure nitrous oxide emissions, Environmental Research Letters, 8, 025 022, 2013.

Hämäläinen, K., Kallonen, A., Kolehmainen, V., Lassas, M., Niinimäki, K., and Siltanen, S.: Sparse Tomography, SIAM Journal on Scientific Computing, 35, B644–B665, doi:10.1137/120876277, 2013.

Jin, B. and Maass, P.: Sparsity regularization for parameter identification problems, Inverse Problems, 28, 123 001, 2012.

Karion, A., Sweeney, C., Kort, E. A., Shepson, P. B., Brewer, A., Cambaliza, M., Conley, S. A., Davis, K., Deng, A., Hardesty, M., Herndon, S. C., Lauvaux, T., Lavoie, T., Lyon, D., Newberger, T., Pétron, G., Rella, C., Smith, M., Wolter, S., Yacovitch, T. I., and Tans, P.: Aircraft-Based Estimate of Total Methane Emissions from the Barnett Shale Region, Environmental Science & Technology, 49, 8124–8131, doi:10.1021/acs.est.5b00217, 2015.

Knopp, T. and Weber, A.: Sparse Reconstruction of the Magnetic Particle Imaging System Matrix, IEEE Transactions on Medical Imaging, 32, 1473–1480, 2013.

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

Loris, I.: On the performance of algorithms for the minimization of l1 -penalized functionals, Inverse Problems, 25, 035 008, 2009.

Louis, A. K.: Inverse und schlecht gestellte Probleme, Teubner-Studienb"ucher, Mathematik, Teubner, Stuttgart, 1989.

Mairal, J., Bach, F., and Ponce, J.: Sparse Modeling for Image and Vision Processing, Foundations and Trends in Computer Graphics and Vision, 8, 85–283, doi:10.1561/0600000058, 2014.

Martinez-Camara, M., Dokmanic, I., Ranieri, J., Scheibler, R., Vetterli, M., and Stohl, A.: The Fukushima inverse problem, in: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, pp. 4330–4334, doi:10.1109/ICASSP.2013.6638477, 2013.

Michalak, A. M. and Kitanidis, P. K.: A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification, Water Resources Research, 39, n/a–n/a, doi:10.1029/2002WR001480, 2003.

Michalak, A. M., Bruhwiler, L., and Tans, P. P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, Journal of Geophysical Research: Atmospheres, 109, n/a–n/a, doi:10.1029/2003JD004422, 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, Proceedings of the National Academy of Sciences, 110, 20 018–20 022, doi:10.1073/pnas.1314392110, 2013.

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

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, Meteorology and Atmospheric Physics, 107, 51–64, doi:10.1007/s00703-010-0068-x, 2010.

NOAA: NOAA Earth System Research Laboratory, Global Monitoring Division, Aircraft Program, http://www.esrl.noaa.gov/gmd/ccgg/aircraft/index.html, accessed: 2016-04-15.

Olivier, J. G. and Peters, J. A.: CO2 from non-energy use of fuels: A global, regional and national perspective based on the IPCC Tier 1 approach, Resources, Conservation and Recycling, 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, Bulletin of the American Meteorological Society, 91, 327–342, doi:10.1175/2009BAMS2865.1, 2010.

Ray, J., Lee, J., Yadav, V., Lefantzi, S., Michalak, A. M., and van Bloemen Waanders, B.: A sparse reconstruction method for the estimation of multi-resolution emission fields via atmospheric inversion, Geoscientific Model Development, 8, 1259–1273, doi:10.5194/gmd-8-1259-2015, 2015.

Reichel, L. and Rodriguez, G.: Old and new parameter choice rules for discrete ill-posed problems, Numerical Algorithms, 63, 65–87, doi:10.1007/s11075-012-9612-8, 2012.

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

Saide, P., Bocquet, M., Osses, A., and Gallardo, L.: Constraining surface emissions of air pollutants using inverse modelling: method intercomparison and a new two-step two-scale regularization approach, Tellus B, 63, doi:10.1111/j.1600-0889.2011.00529.x, 2011.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., , W. W., and Powers, J. G.: A description of the advanced research WRF version 2, Tech. rep., University Corporation for Atmospheric Research, 2005.

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

Stohl, A., Seibert, P., Arduini, J., Eckhardt, S., Fraser, P., Greally, B. R., Lunder, C., Maione, M., Mühle, J., O'Doherty, S., Prinn, R. G., Reimann, S., Saito, T., Schmidbauer, N., Simmonds, P. G., Vollmer, M. K., Weiss, R. F., and Yokouchi, Y.: An analytical inversion method for determining regional and global emissions of greenhouse gases: Sensitivity studies and application to halocarbons, Atmospheric Chemistry and Physics, 9, 1597–1620, doi:10.5194/acp-9-1597-2009, 2009.

U.S. National Research Council: Verifying Greenhouse Gas Emissions : Methods to Support International Climate Agreements., National Academies Press, 2010.

Zavala-Araiza, D., Lyon, D. R., Alvarez, R. A., Davis, K. J., Harriss, R., Herndon, S. C., Karion, A., Kort, E. A., Lamb, B. K., Lan, X., Marchese, A. J., Pacala, S. W., Robinson, A. L., Shepson, P. B., Sweeney, C., Talbot, R., Townsend-Small, A., Yacovitch, T. I., Zimmerle, D. J., and Hamburg, S. P.: Reconciling divergent estimates of oil and gas methane emissions, Proceedings of the National Academy of Sciences, 112, 15 597–15 602, doi:10.1073/pnas.1522126112, 2015.

Zhao, C., Andrews, A. E., Bianco, L., Eluszkiewicz, J., Hirsch, A., MacDonald, C., Nehrkorn, T., and Fischer, M. L.: Atmospheric inverse estimates of methane emissions from Central California, Journal of Geophysical Research - Atmospheres, 114, n/a–n/a, doi:10.1029/2008JD011671, 2009.