

Sparse optimization for inverse problems in atmospheric modelling[☆]



L. Adam^{*}, M. Branda

Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodárenskou věží 4, Prague, 18208, Czech republic

ARTICLE INFO

Article history:

Received 26 March 2015

Received in revised form

1 February 2016

Accepted 1 February 2016

Available online xxx

Keywords:

Inverse modelling

Sparse optimization

Integer optimization

Least squares

European tracer experiment

Free Matlab codes

ABSTRACT

We consider inverse problems in atmospheric modelling represented by a linear system which is based on a source-receptor sensitivity matrix and measurements. Instead of using the ordinary least squares, we add a weighting matrix based on the topology of measurement points and show the connection with Bayesian modelling. Since the source-receptor sensitivity matrix is usually ill-conditioned, the problem is often regularized, either by perturbing the objective function or by modifying the sensitivity matrix. However, both these approaches may be heavily dependent on specified parameters. To ease this burden, we propose to use techniques looking for a sparse solution with a small number of positive elements. Finally, we compare all these methods on the European Tracer Experiment (ETEX) data where there is no a priori information apart from the release position and some measurements.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

In the past decade, it often happened that a large amount of harmful substance was released into the atmosphere over a short period of time. As the probably most known examples we can mention the nuclear power accidents in Fukushima in 2011 and Chernobyl in 1986 or the Eyjafjallajökull volcanic eruption in 2010. When such an accident happens, there are two major questions to be asked. Where is the substance going to be in a day or two? And how much of the substance was released into the atmosphere?

The problematic part is that in such cases only a certain amount of measurements is known. And these two imposed questions have to be answered based on this information from these measurements. The first question is usually simpler to answer: from the measurements we apply a model and observe in which direction and how fast the substance is transferred. The second question is generally more demanding, besides applying a model, one usually has to consider a certain optimization problem which needs to be solved. Since in this case we are interested in computation of a quantity which happened in the past, this field is often called inverse modelling. In this paper, we will focus purely on the second question.

There has been much progress in the field of inverse modelling. For a general introduction to inverse modelling in environmental modelling and protection, see books Aster et al. (2012) and Tipping (2002). Furthermore, we mention works related to power plant accidents, for instance Martinez-Camara et al. (2013) and Stohl et al. (2012) focused on the Fukushima accident. Also volcanic ash, aerosol and gas emissions can influence climate significantly and pose hazards to human health and ocean productivity. Therefore, the impact of Eyjafjallajökull volcanic eruption was investigated by Stohl et al. (2011). Another application of the inverse modelling lies in the monitoring of emission limits, e.g. Pétron et al. (2002) focused on the emission of carbon monoxide. Miller et al. (2014) applied several inverse methods to US anthropogenic methane emissions modelling.

The problem described above gives rise to minimizing the discrepancy between the measurements and the model predictions. As we have already mentioned, one usually knows a vector of measurements $(y_1, \dots, y_m) \in \mathbb{R}^m$ and tries to identify the unknown release profile $(x_1, \dots, x_n) \in \mathbb{R}^n$. Both these vectors are aggregated over spatial and/or temporal domain. As an example, (y_1, \dots, y_m) may consist of measurements at different places and times while (x_1, \dots, x_n) may represent release profile at one place but for different times.

To be able to identify the release profile, one has to know a mapping $M : \mathbb{R}^n \rightarrow \mathbb{R}^m$ which assigns expected measurements to a release profile. In tracer transport modelling, this mapping is often linear, represented by a matrix $M \in \mathbb{R}^{m \times n}$. This leads to an

[☆] This research is supported by EEA/Norwegian Financial Mechanism under project 7F14287 STRADI.

^{*} Corresponding author.

E-mail addresses: adam@utia.cas.cz (L. Adam), branda@utia.cas.cz (M. Branda).

URL: <http://staff.utia.cas.cz/adam>

optimization problem

$$\begin{aligned} & \underset{x}{\text{minimize}} && \|Mx - y\|_2^2 \\ & \text{subject to} && x \geq 0, \end{aligned} \quad (1)$$

where $\|\cdot\|_2$ is the Euclidean norm. In the objective, we minimize the discrepancy between the measured and predicted values. The constraint $x \geq 0$ is natural here because release cannot be negative.

Matrix M is usually denoted as the source-receptor sensitivity matrix, see [Seibert and Frank \(2004\)](#). The reason for this is that element m_{ij} of matrix M represents the sensitivity of the measurement at point i to the release at point j . Thus, in an ideal world where we know M precisely, we would obtain $m_{ij}x_j = y_i$. This means that m_{ij} denotes the expected change in y_i upon a unit change in x_j . For matrix M , we will often use the shortened notation sensitivity matrix.

To obtain a column of the sensitivity matrix, one usually employs a transportation model and simulates a release of a large number of particles at the position corresponding to this column, see again [Seibert and Frank \(2004\)](#). After doing so, predicted values are computed at all places where a measurement is available. This means that the number of rows of M is fixed and is equal to the number of measurements. On the other hand, the number of columns of M may be chosen arbitrarily. For example, it may be increased by decreasing the time interval between two subsequent possible release instances. Since the computation of one column of M is very time consuming, M usually has many more rows than columns and thus the system is overdetermined.

There are two main approaches to finding a solution to (1). The first one is a deterministic approach and makes use directly of formulation (1) and tries to solve it by optimization techniques. Since the problem is often ill-conditioned, various regularizations are used to make the problem more tractable. This approach usually results in the necessity of solving a constrained quadratic problem. The second approach is a stochastic one and instead of solving (1), it assumes that

$$y = Mx + \varepsilon, \quad (2)$$

where ε is a random vector. Provided that ε has normal distribution with independent components having zero mean and the same variance, then applying the maximal likelihood estimate reduces precisely to solving (1). Naturally, this problem can be seen as ordinary least squares with nonnegativity constraint.

The main aim of this paper is twofold. The first goal is to propose a modification of the deterministic approach for solving [problem \(1\)](#) by adding nondiagonal weighting matrix W . This approach is closely connected with Bayesian modelling, where the weighting matrix enters as a covariance matrix of the measurements. However, we base the weighting matrix purely on the topology of the measurement points and not on the measurement values as is the case of Bayesian modelling.

The second goal is to show a new approach of dealing with ill-conditioned sensitivity matrix M . To the best of our knowledge, the usual approach is either to use some regularization or to ignore certain measurements, which reduces the number of rows in M . This, however, may lead to a suboptimal solution when the solution of the reduced problem is not a solution to the original problem. Moreover, if this approach is applied, a certain parameter, such as the maximal matrix row norm, has to be specified. This parameter usually has huge impact on the optimal solution and thus, the problem is not very stable with respect to a choice of this parameter. This is especially true if the parameter takes form of apriori information.

We try to prevent this behaviour and suggest to look for a sparse solution x , which means that x should contain as many zeros as possible. This problem may be formulated as a multiobjective (two-objective) optimization: we try to minimize the measurement error $\|Mx - y\|_2^2$ and at the same time, we try to minimize the number of nonzeros, which is denoted by $\|x\|_0$. To obtain the so called efficient solutions of such multiobjective problem, we might rewrite it as a single-objective one, in which one of the criteria is minimized and the other one is constrained by a prescribed maximal bound.

It is important to note that looking for a sparse solution has a natural effect of ignoring columns of M with small norm, and thus only the goal value enters the problem as a parameter. This parameter prescribes the maximum number of nonzeros $x_0 \leq k_{tol}$. Note that the parameter k_{tol} is an integer and has a natural explanation. Moreover, it is invariant to the scaling of variables, thus it is not necessary to rescale k_{tol} whenever y are measured in different units. As there are only finite number of possible values for k_{tol} , this parameter is relatively simple to choose. This is in contrast with most of the currently used regularization techniques.

The main assumption of following this approach is that the solution x is indeed sparse. For example, it can be used for both problems of identification of the source location and of the time process of the source release. In the first case, it is usually assumed that there is only one or few release points and the task is to find them. In the second case, the considered time window is usually much longer than the duration of the actual release. In both cases, the optimal solution should contain a large number of zeros, justifying the use of sparsity techniques. Clearly, there are also instances where the sparse solution approach cannot be used. One example of this possibility is the CO₂ emission monitoring. This is caused by the fact that CO₂ is released globally and thus does not have a localized point source.

This paper is organized as follows: in [Section 2](#) we comment on properties of M and suggest a way of enhancing model (1) by multiplying both term in the objective Mx and y by a certain matrix W . Relations to stochastic Bayesian inverse modelling techniques are discussed.

[Section 3](#) summarizes the current state of the art for inverse modelling of tracer transport and proposes alternative approaches. After a short motivational part showing that the sensitivity matrix M is often ill-conditioned, two subsections follow which present ways of solving [problem \(1\)](#) while dealing with this ill-conditioning. The first one summarizes the most widely used deterministic optimization techniques used for identifying point releases while the second one proposes a new way of handling the problem by looking for sparse solutions instead of regularizing the problem or modifying the sensitivity matrix.

In [Section 4](#) we apply the described techniques to the European Tracer Experiment (ETEX). After providing some information about the sensitivity matrix M , we show the validity of using the weighting matrix W . After doing so, we depict both the known and the proposed solutions and compare them.

Finally, we would like to emphasize that all codes and data are available online.¹

2. Spatial and temporal locations weighting

In the introduction we have already described that the sensitivity matrix M is obtained by repeatedly running an underlying tracer transport model. We will shortly discuss the possible shortcomings of applying these dispersive models. Unfortunately, it may happen that the model does not describe the real situation

¹ <http://staff.utia.cas.cz/adam/research.html>.

precisely because of certain reasons. There are inherent uncertainties related to the variability of the physical phenomena (turbulence, for instance), which would prevent the models to be fully realistic. This appears together with the specific problems of input, initialization, boundary conditions, advanced or less advanced parameterizations of all phenomena. Among specific problems we can include for example imprecise weather description, inability to incorporate mountain ranges, or omitting some meteorological events. Novel approaches for designing a source-receptor relationship for air pollution modelling were recently introduced by Clappier et al. (2015) or Vedrenne et al. (2014).

In this section, we introduce spatial and temporal location weighting to handle these possible deficiencies. We start with a simple example showing that relatively small changes in the sensitivity matrix M can lead to big changes in solution x of (1).

Example 1. Assume that we have the following data

$$M = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Thus, we have two measurements and only one potential release, leading to $m = 2$ and $n = 1$. By solving problem (1), we obtain trivial solution $x = 0$.

Matrix M suggests that if a release occurred, all content of this release was transmitted to the second measurement point and nothing to the first one. Assume now that the locations of these two measurement points are close to each other and that a wrong wind direction was considered and all release should have been transmitted to the first measurement point instead of the second one. This corresponds to using $\tilde{M} = (1, 0)^T$ instead of M . If we indeed consider \tilde{M} instead of M , then we obtain optimal release $\tilde{x} = 1$ instead of $x = 0$.

We can observe a highly undesirable situation when a small change in the real situation may cause a big change in the sensitivity matrix M and in the obtained solution x . In the rest of this subsection, we will suggest a way of dealing with this problem. Unfortunately, this situation cannot be handled by adding a possibility to perturb M with a penalization of such perturbation.

The potential problem with Example 1 is that model (1) may not be entirely suitable for modelling phenomena where additional information such as the spatial and temporal location of measurements is known. In our opinion, it could be advantageous not to compare Mx and y componentwise but to take into account their spatial and temporal locations and compare the sum on a neighbourhood of every component. To present this idea in a more concise way, assume that for every measurement y_j we know additional data $z_j = (z_j^x, z_j^y, z_j^t)$, where pair (z_j^x, z_j^y) represents the longitude and latitude of a measurement point and z_j^t the measurement time. We would like to define weights related to the distance between measurement points z_i and z_j . First, we define the space and time weights as follows

$$w_S(z_i, z_j) := \begin{cases} \exp(-\alpha_S \| (z_i^x, z_i^y) - (z_j^x, z_j^y) \|) & \text{if } \| (z_i^x, z_i^y) - (z_j^x, z_j^y) \| \leq s_{max}, \\ 0 & \text{otherwise,} \end{cases} \quad (3a)$$

$$w_T(z_i, z_j) := \begin{cases} \exp(-\alpha_T \| z_i^t - z_j^t \|) & \text{if } \| z_i^t - z_j^t \| \leq t_{max}, \\ 0 & \text{otherwise,} \end{cases} \quad (3b)$$

where $\alpha_S \geq 0$, $\alpha_T \geq 0$ and $s_{max} \in [0, \infty]$, $t_{max} \in [0, \infty]$ are given parameters; the last two are known as cutoff distances. Since both quantities in (3) lie in interval $[0, 1]$, we may define the (non-

normalized) weights related to the distance between z_i and z_j as

$$w(z_i, z_j) := w_S(z_i, z_j) w_T(z_i, z_j) \quad (4)$$

Note that this weight is zero if the measurements are performed at distant places (as specified by s_{max}) or at distant times (as specified by t_{max}). Moreover, the weight decreases with increasing spatial or temporal distance. This rate of decrease is determined by α_S and α_T .

When considering ordinary least squares, we try to minimize for every $j = 1, \dots, m$ the following discrepancy

$$\left((Mx)_j - y_j \right)^2. \quad (5)$$

Following the idea from Example 1, instead of minimizing quantity (5) at given point j , we will minimize the difference between Mx and y on a neighbourhood of point j . If we relate this neighbourhood to weight w defined in (4), for every measurement $j = 1, \dots, m$ we try to minimize the following quantity

$$\left(\sum_{i=1}^m \frac{w(z_j, z_i)}{\sum_{k=1}^m w(z_j, z_k)} (Mx)_i - \sum_{i=1}^m \frac{w(z_j, z_i)}{\sum_{k=1}^m w(z_j, z_k)} y_i \right)^2, \quad (6)$$

where the denominator is used for weight normalization. Formula (6) corresponds precisely to (5) provided that $w(z_i, z_i) = 1$ and $w(z_i, z_j) = 0$ for $i \neq j$. When we combine components (6) into one vector, we arrive at minimizing

$$\sum_{j=1}^m \left(\sum_{i=1}^m \frac{w(z_j, z_i)}{\sum_{k=1}^m w(z_j, z_k)} ((Mx)_i - y_i) \right)^2,$$

which is equal to

$$\| W(Mx - y) \|_2^2, \quad (7)$$

where matrix W consists of elements

$$w_{ij} = \frac{w(z_i, z_j)}{\sum_{k=1}^m w(z_i, z_k)}. \quad (8)$$

For other possibilities of choosing matrix W , we refer the reader to Gaspari and Cohn (1999) or Hamill et al. (2001).

To conclude this approach, we propose to solve the (weighted least squares) problem

$$\begin{aligned} & \underset{x}{\text{minimize}} \quad \| W M x - W y \|_2^2 \\ & \text{subject to} \quad x \geq 0 \end{aligned} \quad (9)$$

instead of the (ordinary least squares) problem (1), where matrix W has been defined in (3), (4) and (8). Now we present a continuation of Example 1 which strongly supports our arguments for using matrix W .

Example 1 (continued). Consider the same data as in Example 1. Denoting $w := w(z_1, z_2)$, we have

$$W = \frac{1}{1+w} \begin{pmatrix} 1 & w \\ w & 1 \end{pmatrix}, \quad W M x - W y = \frac{1}{1+w} \begin{pmatrix} -1 + w x \\ -w + x \end{pmatrix},$$

and thus problem (9) reads

$$\begin{aligned} & \underset{x}{\text{minimize}} \quad \frac{1}{(1+w)^2} \left[(1+w^2)x^2 - 4wx + 1 + w^2 \right] \\ & \text{subject to} \quad x \geq 0. \end{aligned}$$

Since $w \geq 0$, for the optimal solution we obtain

$$x = \frac{2w}{1+w^2}.$$

We note that for $w = 0$, matrix W is equal to the identity matrix. Then [problem \(9\)](#) amounts to solving [\(1\)](#), and that the solution is indeed $x = 0$ for this case. However, if w is close to 1, then the measurement points are at almost identical point, which means that they should not be distinguished as separate measurement points. Indeed, we obtain that x is close to 1, which is in accordance to our expectation.

Even though this whole paper is based on deterministic formulation, [problem \(9\)](#) has close connection with stochastic (random) solution methods, where one make use of [formula \(2\)](#) and employs Bayesian modelling, see papers by [Bocquet \(2008\)](#), [Osorio-Murillo et al. \(2015\)](#) with open source software package MAD# or [Smídl and Hofman \(2014\)](#), where Monte Carlo methods are proposed.

By computing the optimal solution to [\(9\)](#) without the non-negativity constraint imposed on x , we obtain

$$x = (M^T W^T W M)^+ M^T W^T W y,$$

where $(M^T W^T W M)^+$ is a Moore–Penrose pseudoinverse matrix, see [Albert \(1972\)](#). If the matrix in question is invertible, then the pseudoinverse reduces to the classical inverse matrix. If $W^T W$ is a regular matrix, then solution x corresponds to the best linear unbiased estimator (BLUE) for model $y = Mx + \varepsilon$, where ε is random vector with zero mean and covariance matrix $(W^T W)^{-1}$, see [Stewart et al. \(2008\)](#). However, there is a major difference between our approach and with that of the Bayesian modelling. In our approach, there will be a strong negative correlation between close points, which may seem slightly counterintuitive from the point of Bayesian modelling. However, there is a simple explanation for such event: the closeness of points is penalized in the Bayesian approach to prevent similar measurements to have great impact on the solution. However, in our approach we want to perform a sum with respect to neighbouring points, thus we want to promote and not penalize closeness of points. We show this behaviour in [Example 1](#).

Example 1 (continued). It is not difficult to compute

$$(W^T W)^{-1} = \frac{(1+w)^2}{(1-w^2)^2} \begin{pmatrix} 1+w^2 & -2w \\ -2w & 1+w^2 \end{pmatrix}.$$

We see that we have obtain nonnegative large off-diagonal elements for close points (w is close to one), which agrees with the discussion in the previous paragraph.

3. Sparse optimization

In this section we first provide reasons why the sensitivity matrix M may be ill-conditioned. The rest of the section is divided into two subsections and presents several possible ways of dealing with this problem. In the first one, we show several current methods and in the second one, we present the concept of sparse solution, thus a solution which contains many zero elements.

We are fully aware that the true release does not have to be sparse in general. In this case, our methods should not be used.

However, looking for a sparse solution makes sense in the following widely used applications: identifying release points and identifying release time profile. When we know that the release was performed only from one or few places and the goal is to find these places, sparsity is naturally present. On the other hand, if we know the release place and want to identify the release time profile, the sparsity is present as well because the examined window of potential release is usually much longer than the true release window.

As we have mentioned in the introduction, the sensitivity matrix is computed by simulating a release of huge number of particles. Since the underlying model is stochastic, the trajectory of every particle is governed by a different realization of the random vector. Thus, every particle follows a different trajectory. The resulting concentration y is then based on the number of particles in the vicinity of a measurement point.

To explain why the sensitivity matrix M may be ill-conditioned, we consider identification of a time profile of a release where the true release point is known. Then x_j denotes the released amount at some time t_j and the sum of j -th column of M denotes the total predicted sum of measurements at all measurement places per unit release at t_j . It is very likely that there will be some time instant t_j which has no influence on the predicted values at the measurement places, for example if t_j occurred much earlier than the first measurement was taken. But since the underlying process is stochastic and there is a large amount of considered particles, it may happen that some particles reached the measurement place. This would result in very small entries in the sensitivity matrix M . Unfortunately, this is a very undesirable behaviour as we show in the following example.

Example 2. Consider the following data

$$M = \begin{pmatrix} 1 & 0 \\ 1 & 10^{-10} \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Then it is clear that matrix M is ill-conditioned and that the unique solution is equal to $x = (1, 0)$. However, if there is slight error in the measurements, for example if we consider $\tilde{y} = (1, 1 + 10^{-5})$, then the original solution x changes by a big margin into $\tilde{x} = (1, 10^5)$.

Let us impose now a requirement that at most one component of x can be nonzero. Then since we want to minimize the discrepancy between Mx and y , the solution will always be close to $(1, 0)$ for any small perturbation of y . This is precisely the case of sparse optimization.

There have been some attempts to deal with this problem. In [Martinez-Camara et al. \(2014a, 2014b\)](#) outliers were detected and removed from the model. In the first paper, an extension of the Random Sample Consensus (RANSAC) algorithm was proposed while in the second one, two matrices M_1 and M_2 were constructed and outliers were detected and removed based on differences in these matrices. Finally, in [Martinez-Camara et al. \(2013\)](#) it was suggested to keep removing rows from M until the resulting matrix has a better condition number. However, we believe that the rows of M should stay mostly untouched because removing rows of M means ignoring valuable data (for example in the latter paper approximately 60% of all measurements were removed).

On the other hand, sparse optimization tries to find a sparse (approximate) solution of a linear system, where only a limited number of nonzero elements appears. There exist many underlying practical applications which make use of the developed theory. One of the well-known examples is the image compressing into special formats such as JPEG. During such compressing, the pixel information is transformed using special transforms such as Fourier transform. These transforms are constructed in such a way that only the first several coefficients are large (and thus important) and thus, the remaining ones can be ignored without losing much

information. Hence, the image has a sparse representation. Another application from image processing is the image denoising of image deblurring. Applications from different fields amount for magnetic resonance imaging (MRI) or target location via radars. For more information about this field see [Bruckstein et al. \(2009\)](#) and [Foucart and Rauhut \(2013\)](#) for annotated bibliography and references, and [Mairal et al. \(2010\)](#) for implementation of selected algorithms in SparseLab or Sparse Modeling Software (SPAMS).

3.1. Traditional – approximate – approaches to inverse modelling

In this subsection, we formulate and shortly discuss methods which were already used for dealing with badly scaled matrices. These methods will be then compared with the sparse optimization technique in the empirical part in Section 4. In particular, we focus on removing low-influential columns of M and two traditional regularization techniques.

3.1.1. Columns removal

The most obvious strategy is to remove columns of the sensitivity matrix. We can imagine that we solve [problem \(9\)](#) where the components of x corresponding to the omitted columns are fixed to zero. However, it may happen that the optimal solution of this problem is far from an optimal solution to the original [problem \(9\)](#). Specifically, denoting a column of M by m_j , we remove all columns of M which satisfy

$$\|m_j\|_2 \leq c_{column} \max_{k=1, \dots, m} \|m_k\|_2, \quad (10)$$

where $c_{column} > 0$ is a given constant.

3.1.2. l_2 -regularization: Tikhonov approach

Another possibility is to use the Tikhonov regularization and to solve problem

$$\begin{aligned} & \underset{x}{\text{minimize}} \quad \|WMx - Wy\|_2^2 + \alpha \|x\|_2^2 \\ & \text{subject to} \quad x \geq 0 \end{aligned} \quad (11)$$

with fixed penalization parameter $\alpha > 0$. The second term in the objective penalizes the Euclidean norm of the release. Note that the optimal solutions usually contain many small nonzero values, thus they are not sparse.

A similar approach was used in [Stohl et al. \(2012\)](#) where the following: $\underset{x \geq 0}{\text{minimize}} \|W_1(Mx - y)\|_2^2 + \|W_2x\|_2^2 + \|W_3Dx\|_2^2$. Matrices W_1 , W_2 and W_3 are diagonal with positive entries on their diagonal and thus allow for weighting of individual elements. Matrix D represents discretized differential operator, see Chapter 8 in [Nocedal and Wright \(2006\)](#), and thus the last term forces x to be as smooth as possible. Another possibility is to add term $\|W_4(x - x_a)\|_2^2$ to the objective function, where $x_a \neq 0$ is either an expert or apriori estimate. Since optimal solution x would strongly depend on the choice of x_a in this case, we try to avoid this possibility.

3.1.3. l_1 -regularization: LASSO

To the best of our knowledge, there have not been many attempts to use sparse optimization for atmospheric modelling. One of the attempts was performed in [Martinez-Camara et al. \(2013\)](#), where problem

$$\begin{aligned} & \underset{x}{\text{minimize}} \quad \|WMx - Wy\|_2^2 + \alpha \|x\|_1 \\ & \text{subject to} \quad x \geq 0 \end{aligned} \quad (12)$$

was considered. This can be understood as the Least absolute shrinkage and selection operator (LASSO) method, which was introduced in [Tibshirani \(1996\)](#) and later thoroughly studied in [Donoho and Tsai \(2008\)](#). Compared with (11), this method penalizes l_1 norm of the release instead of l_2 norm. Since x is nonnegative, [problem \(12\)](#) is smooth.

3.2. Exact approaches to sparse optimization

All approaches from the previous subsection somehow regularized the problem and depended heavily on a real parameter. In this subsection we present another approach. Instead of modifying the objective or the data, it assumes that the solution is sparse and tries to search for a solution with only a limited number of nonzeros. Even though a (regularization) parameter is present as well in this case, it is an integer and thus much easier to select properly.

There are two possibilities in [problem \(9\)](#). Either system $Mx = y$ is underdetermined, which usually corresponds to the case of $m < n$. Then there exist multiple solutions and the task of sparse optimization is to select the one with the lowest number of nonzero components. However, in our case it usually happens that the system is overdetermined, which usually corresponds to $m > n$. Then the solution of (9) is uniquely determined but the solution may be dense. In such cases it is possible to trade lower density for a slightly worse error $\|WMx - Wy\|_2$. A big advantage of sparse solutions is that columns corresponding to zero components of a solution are ignored. This means that sparse solutions naturally deal with ill-conditioned matrices as described earlier in this section.

The basic concept used is the l_0 “norm”, which is defined as

$$\|x\|_0 := \#\{i \mid x_i \neq 0\},$$

where $\#A$ denotes the number of elements in a set A . The term l_0 norm was coined because $\|x\|_0 = \lim_{p \rightarrow 0} \|x\|_p$, see Section 2.1 in [Foucart and Rauhut \(2013\)](#). Thus, sparse optimization tries to minimize $\|x\|_0$, together with criterion (7); a problem which can be understood as a multiobjective optimization, see [Miettinen \(1999\)](#). This gives rise to problem

$$\begin{aligned} & \underset{x}{\text{minimize}} \quad \|WMx - Wy\|_2^2 \\ & \text{subject to} \quad \|x\|_0 \leq k_{tot}, \\ & \quad \quad \quad x \geq 0, \end{aligned} \quad (13)$$

where $k_{tot} \in \mathbb{N}$ is a natural number which denotes the maximal number of nonzeros in x .

Because the feasible set of [problem \(13\)](#) is highly nonconvex (for example for $n = 2$ and $k_{tot} = 1$ it consists of two halflines), this problem is usually very difficult to solve efficiently. One of the possibilities of handling these constraints is to use artificial binary variables $z_i \in \{0, 1\}$ such that

$$\begin{aligned} z_i = 0 & \Leftrightarrow x_i = 0, \\ z_i = 1 & \Leftrightarrow x_i > 0, \end{aligned} \quad (14)$$

which leads to exact expression

$$\|x\|_0 = \sum_{i=1}^n z_i.$$

Relation (14) is equivalent to

$$0 \leq x_i \leq z_i \cdot ub_i, \quad (15)$$

where $ub_i > 0$ is a given upper bound for the release (or a sufficiently large number). Instead of this relation, we will consider

$$z_i \cdot lb_i \leq x_i \leq z_i \cdot ub_i, \quad (16)$$

where $lb_i \geq 0$ are lower bounds for release. Even though lb_i and ub_i may seem as additional penalization parameters, this is not the case. They serve only to improve the properties of the following integer programming. If $lb_i = 0$, then constraints (15) and (16) coincide. Moreover, if $ub_i = \infty$ then the upper bounds in (15) and (16) are eliminated from the system.

Concerning the interpretation of lb_i : if this constant is strictly positive, thus $lb_i > 0$, then either there is no release ($z_i = 0$ and $x_i = 0$) or if there is any positive release $x_i > 0$, then this release must be at least lb_i ($z_i = 1$ and $x_i \geq lb_i$). Hence, this constraint basically says that if there is any release, then it has to be automatically greater than lb_i .

By the analysis in the previous paragraph, we have derived the following problem related to (13)

$$\begin{aligned} & \underset{x,z}{\text{minimize}} && \|WMx - Wy\|_2^2 \\ & \text{subject to} && \sum_{i=1}^n z_i \leq k_{tol}, \\ & && z_i \cdot lb_i \leq x_i \leq z_i \cdot ub_i, \quad i = 1, \dots, n, \\ & && z_i \in \{0, 1\}. \end{aligned} \quad (17)$$

Note that the constraints imply $x_i \geq 0$ and thus, it is not necessary to state this constraint explicitly. For a solution x of this problem, we always have that at most k_{tol} components x_i are positive and if this is the case, then they are greater than lb_i .

Solvers available in Matlab employ a matrix representation of the problems, thus we present it for our problem (17). It is not difficult to see that the problem may be written in the following form

$$\begin{aligned} & \underset{u}{\text{minimize}} && u^T H u + h^T u \\ & \text{subject to} && A u \leq b, \\ & && u_z \in \{0, 1\}, \end{aligned} \quad (18)$$

where $u = (x, z)$, u_z refers to the second part of u and the data read

$$H = \begin{pmatrix} M^T W^T W M & 0_{n \times n} \\ 0_{n \times n} & 0_{n \times n} \end{pmatrix}, \quad h = \begin{pmatrix} -2M^T W^T W y \\ 0_{n \times 1} \end{pmatrix}$$

and

$$A = \begin{pmatrix} 0_{1 \times n} & 1_{1 \times n} \\ -\text{diag}(1_{n \times 1}) & \text{diag}(lb) \\ \text{diag}(1_{n \times 1}) & -\text{diag}(ub) \end{pmatrix}, \quad b = \begin{pmatrix} k_{tol} \\ 0_{n \times 1} \\ 0_{n \times 1} \end{pmatrix}.$$

Here $0_{m \times n}$ stands for matrix (vector for $n = 1$ or $m = 1$) of zeros, $1_{m \times 1}$ and $1_{1 \times n}$ stand for vectors of ones and $\text{diag}(v)$ is a diagonal matrix formed from vector v .

Thus, we managed to recast problem (13) into mixed-integer convex quadratic problem (18). Unfortunately, mixed-integer quadratic programming problems are highly computationally demanding. Thus, a special purpose software for solving such problems is necessary. Such software often works with a relaxed problem, where the integer requirements $u_z \in \{0, 1\}$ are relaxed into simple box constraints $u_z \in [0, 1]$. Since the relaxed problem is convex, then the branch-and-bound algorithm, well known from mixed-integer linear problems, works well for such problems. Moreover, cutting planes can further improve its performance, see



Fig. 1. Locations of measurement stations for the ETEX experiment.

Bertsimas and Shioda (2009) and Sun et al. (2013).

For the algorithmic implementation we have used one of the best available commercial solvers for mixed-integer linear and quadratic problems, CPLEX.² Solver CPLEX is delivered with IBM ILOG CPLEX Optimization Studio,³ which is free for academic purposes and can be easily connected to Matlab. Other possibility is to employ the free optimization library OPTI Toolbox delivered together with noncommercial solver SCIP, see Achterberg (2009). For the convex problems we have used CVX Grant and Boyd (2008).

4. Application to ETEX

In this section we present numerical results and comparison of methods presented in this paper. The ETEX (European Tracer Experiment) is one of the few controlled tracer experiments with detailed information about the release. Similar experiments are for example CAPTEX (Cross-Appalachian Tracer Experiment), ANATEX (Across North America Tracer Experiment) or the not so well-known KATREX (Karlsruhe Tracer Experiment), see Anfossi and Castelli (2014). Among these three experiments, ETEX is the newest one and was performed in 1994 near Rennes in France. This experiment was performed twice, for the first time on 23 October 1994 and for the second time on 14 November 1994. Here, we will consider only the first experiment.

A total mass of 340 kg of PMCH (Perfluoro-Methyl-Cyclo-Hexane) was released into the atmosphere during the course of 12 hours. After that, this substance got carried over Europe and its concentration was measured over the following several days. The sampling network consisted of 168 stations. The locations of these stations are depicted in Fig. 1.⁴ Each station was supposed to sample over the period of 72 hours with the time difference between two subsequent measurements being 3 hours. The starting point for sampling was different for all stations: stations closest to the release point started to sample 3 hours before the release while the stations far away from the release ended their sampling activity 90 hours after the release had started. Unfortunately, large part of the data was either corrupted

² <http://www-03.ibm.com/software/products/en/cpleoptiforzos>.

³ <http://www-03.ibm.com/software/products/en/category/decision-optimization>.

⁴ For plotting this and all further maps we have used package borders, which is freely available at <http://www.mathworks.com/matlabcentral/fileexchange/50390-borders>.

or not usable which resulted in having only 3104 instead of the intended number of 4032 measurements.

The sensitivity matrix was of dimension 3104×112 , thus $m = 3104$ and $n = 112$ and the system was overdetermined. Every component of vector x denoted a potential time when a release could have occurred. The time difference between two subsequent components was 1 hour. Since the true release was performed over 12 hours, the true solution had sparsity of 12, which is approximately 10% of the whole vector. The sensitivity matrix was generated by Hysplit 4. As we have mentioned several times, this matrix is badly scaled. This can be seen in the ratio of the greatest and smallest nonzero element, which is equal to $1.83 \cdot 10^7$. For the sensitivity matrix from a different experiment in [Martinez-Camara et al. \(2013\)](#), this ratio was much worse and reached $8.16 \cdot 10^{14}$.

We have divided this section into three subsections. In the first one, we provide numerical justification for introducing the weighting matrix W . In the second part, we briefly comment on the dependence of optimal solutions on parameters. Finally, in the last part we summarize results obtained by different optimization methods and compare them with each other.

4.1. Justification of using weighting matrix W

Recall that the weighting matrix has been constructed via relations (3), (4) and (8). To obtain it, we need to specify parameters α_S , α_T , s_{max} and t_{max} . We chose $t_{max} = 0$, which means that we allowed no measurement aggregation with respect to different times. In other words, we have $w_T(z_i, z_j) = 1$ when z_i and z_j were taken at the same time instants and 0 otherwise. This also made parameter α_T obsolete. The space distance $\| (z_i^x, z_i^y) - (z_j^x, z_j^y) \|$ was measured as the Euclidean norm applied on degrees of latitude and longitude. Parameter s_{max} was chosen in such a way that we aggregated data from approximately 5% of all stations. Having $n = 3104$ measurement, we chose $s_{max} = 2.5$ to obtain

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \chi_{w_{ij} > 0} = 8.32,$$

where

$$\chi_{w_{ij} > 0} = \begin{cases} 1 & \text{if } w_{ij} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

stands for the characteristic function. This indeed forms

approximately 5% from the 159 stations which were able to provide at least one measurement. For the remaining parameter, we have chosen $\alpha_S = 1.15$, which resulted in

$$\frac{1}{n} \sum_{i=1}^n w_{ii} = 0.50.$$

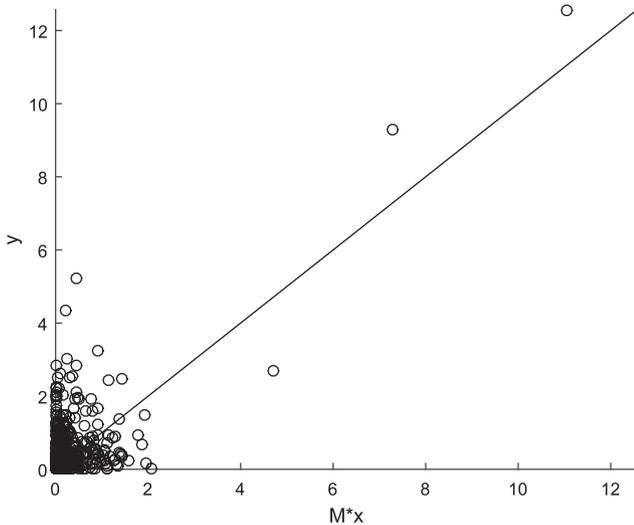
In other words, since the sum of every row of W equals to 1, this means that approximately one half of the objective function of (9) is based on the distance between the measured value y_i and the predicted value $(Mx)_i$ at the same point. On the other hand, the part consisting of the comparison of the measured value y_i at a point and the predicted values $(Mx)_j$ at a neighbourhood amounts approximately to the same amount of the objective. We find this to be a good balance between the ordinary least squares (which would correspond to only for the first part and none of the second one) and considering also some measurements on a neighbourhood of every point.

In the rest of the text, we will denote the OLS (ordinary least squares) and WLS (weighted least squares) solution by the solutions of [problems \(1\) and \(9\)](#), respectively. In [Fig. 2](#) we show the relation between the predicted and actual measurements. [Fig. 2a](#) shows the OLS solution and depicts the relation between y and Mx . Similarly, [Fig. 2b](#) shows that WLS solution and depicts the relation between Wy and WMx . For the best fit, all data would have to lie on the central line. Even though the fit is not perfect, we see that [Fig. 2b](#) shows better results than [Fig. 2a](#), and thus WLS solution has much better approximation capability than OLS solution. This is connected with the fact that adding weighting matrix W moves uncertainties from a point to its neighbourhood. This visual hypothesis is supported by the coefficient of determination R^2 , which tells how many percent of the measurements were explained by the model. We obtain the following data

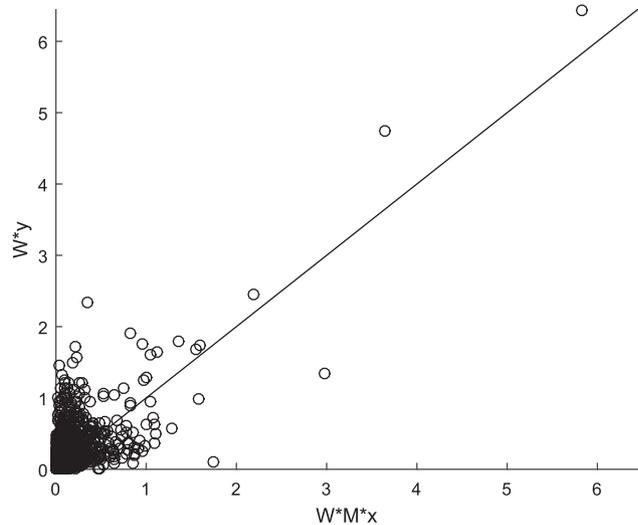
$$R^2_{OLS} := 1 - \frac{\|y - Mx\|^2}{\|y - \text{mean}(y)\|^2} = 0.497,$$

$$R^2_{WLS} := 1 - \frac{\|Wy - WMx\|^2}{\|Wy - \text{mean}(Wy)\|^2} = 0.515.$$

In other words, by using the weighting matrix, we were able to



(a) OLS solution with OLS fit $y \sim Mx$



(b) WLS solution with WLS fit $Wy \sim WMx$

Fig. 2. Comparison of OLS and WLS fits.

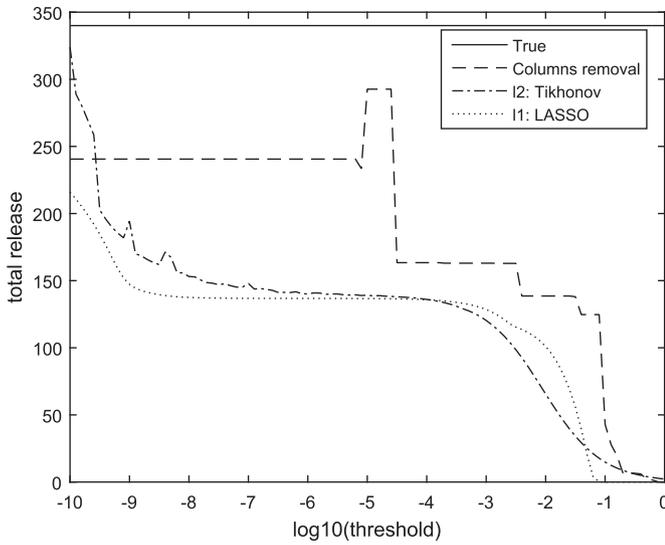


Fig. 3. Dependence of the total release $\sum_{i=1}^n x_i$ on parameters: either $\log_{10}(c_{column})$ (for matrix column reduction) or on $\log_{10}(\alpha)$ (for Tikhonov regularization and LASSO).

explain approximately 1.8% more of the measurements.

4.2. Parameter dependence of current methods

In this subsection, we show that the dependence of the current methods, described in Subsection 3.1, on the involved parameters is rather strong. We can see the results in Fig. 3, where the horizontal axis depicts parameter value, either $\log_{10}(c_{column})$ for column removal or $\log_{10}(\alpha)$ for Tikhonov regularization and LASSO. The vertical axis then depicts the total release $\sum_{i=1}^n x_i$ for the corresponding x . The full line depicts the true release which is to be estimated, the dashed line corresponds to the solution where we removed certain amount of columns, the dash-dotted line to the Tikhonov regularization (11) and finally the dotted line to the LASSO (12). It can be seen that if the parameter is chosen in a bad way, the computed solution is far off from the true solution. This shows the principal difficulty of the regularized ordinary least squared approach: it is not simple to choose the correct value of parameter α .

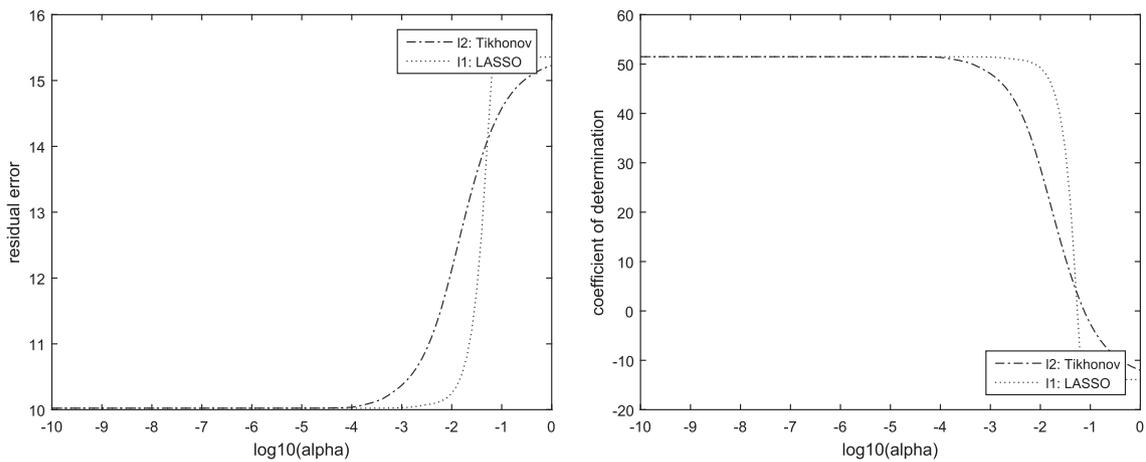
One of the possibilities is to choose the maximal α such that the residual error $\|WMx - Wy\|_2$ stays rather constant. As we can see in Fig. 4, this happened for approximately $\alpha = 10^{-4}$. However, it may be difficult to find such α because it is a real parameter. This may not be the case of sparse optimization and k_{tol} which may take only finite number of values. Moreover, in our opinion, it is much more meaningful to work with parameter k_{tol} rather than with α or c_{column} : while it is simple to explain the meaning of the former parameter, this is not the case of the latter two.

4.3. Sparse optimization and method comparison

We will start with Fig. 5 which depicts variant of Fig. 4 for sparse solutions and shows the dependence of error $\|WMx - Wy\|_2$ and the coefficient of determination R^2_{WLS} on the solution sparsity, thus the number of nonzeros. For a good comparison, we have included l_2 and l_1 regularizations as well, thus solutions of problems (11) and (12), respectively and their sparsities. Bennett et al. (2013) reviews other techniques available across various fields for characterizing the performance of environmental models. Constant α was chosen as in the previous section, thus $\alpha = 10^{-4}$. The remaining markers correspond to WLS solution and l_2 and l_1 regularizations. We see that for approximately $k_{tol} = 10$, the residual error got stable. For this reason, we would recommend this parameter.

In Fig. 6 we plot comparison of selected solutions which were presented in this paper. In the left column, we can see the true release, the WLS solution and finally the LASSO regularization with $\alpha = 10^{-4}$. On the right hand side, we can see the sparse solution with $k_{tol} = 5$, $k_{tol} = 10$ and $k_{tol} = 15$, respectively. On the horizontal axis the time scale is located. On the vertical axis we can see the release at the corresponding time instant. The grey dashed line on the background depicts the scaled column sum of WM . We did not include the Tikhonov regularization because it provided very similar results to LASSO.

All solutions provided rather good approximation of the true release. However, we may observe several interesting differences between various solution types. The WLS solution contains one large predicted release near the end of the time window. Note that the corresponding column of M contains only one nonzero element (out of 3104) and the corresponding column of WM contains six nonzero elements. What probably happened in this time is that the “best” approximation was found at the remaining coordinates and



(a) The dependence of residual error $\|WMx - Wy\|$ on α

(b) Dependence of the coefficient of determination R^2_{WLS} on α

Fig. 4. Dependence of Tikhonov and LASSO regularizations on parameter α .

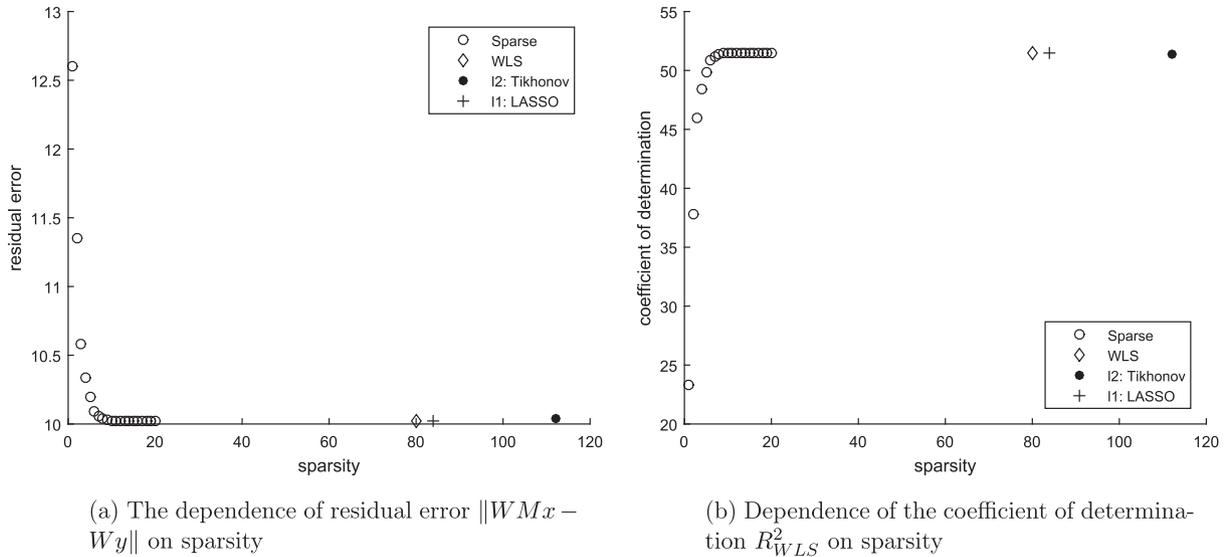


Fig. 5. Dependence of computed solutions on sparsity.

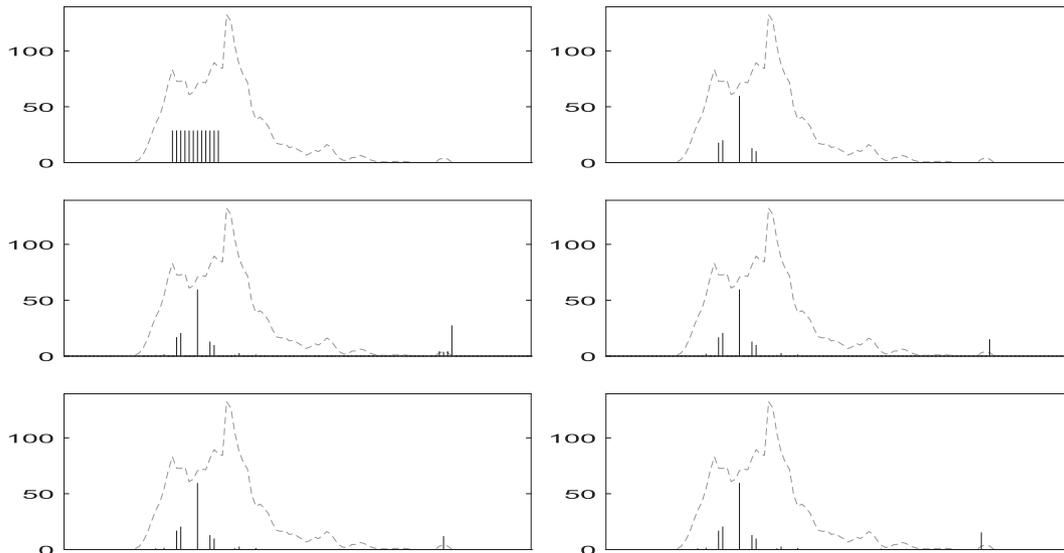


Fig. 6. Values of selected solutions x . The left column depicts the true solution, the WLS solution and the LASSO regularization. The right column depicts the sparse solutions with $k_{tol} = 5$, $k_{tol} = 10$ and $k_{tol} = 15$. The grey dashed line on the background depicts the scaled column sum of WM .

this coordinate provided only a small correction to the final solution. Unfortunately, this small correction in the residuals caused to increase the predicted total release by approximately 25%. It seems that the LASSO regularization has produced a sparse solution, but this is not true. From Fig. 5 we can see that this solution contains 84

nonzero coordinates, many of them are very small. This raises a question whether these small elements should really be present in the solution or if they just provide some small corrections in the solution and are present only because of model or measurement errors.

Table 1
Numerical evidence to Fig. 6. The negative value of the coefficient of determination is possible because the coefficient of determination is nonnegative only for the optimal solution (WLS in this case) and for other values it may be negative. Even though it may seem that the model is inappropriate because of this value, based on Fig. 7 we explain why this happened and that the model is still good.

	Total release	Mean square error	Coef. of deter.	Sparsity
True	340.0	32.62	-4.139	12
WLS	163.0	10.02	0.515	80
LASSO	135.8	10.02	0.515	84
Integer with $k_{tol} = 5$	118.5	10.20	0.500	5
Integer with $k_{tol} = 10$	138.0	10.02	0.515	10
Integer with $k_{tol} = 15$	139.5	10.02	0.515	15

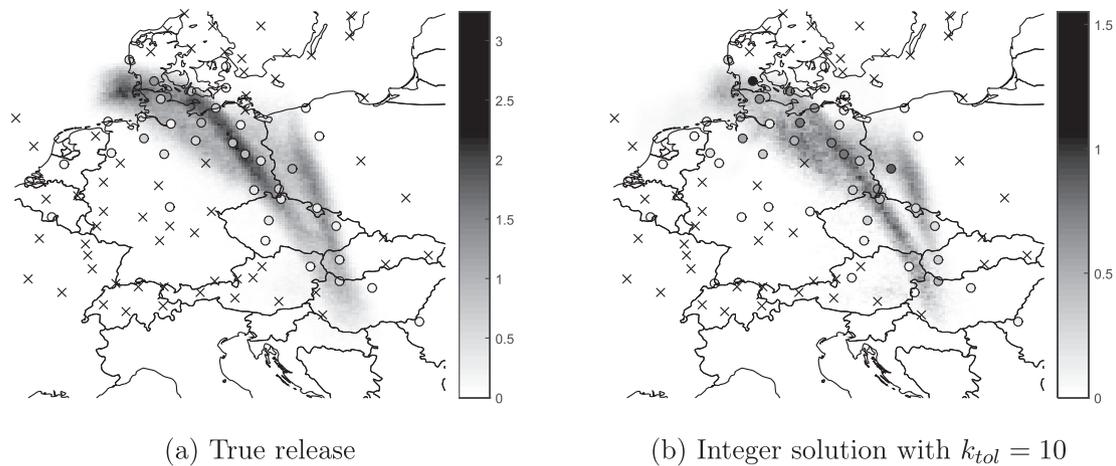


Fig. 7. The predicted plumes corresponding to the true release and the integer solution with $k_{tol} = 10$. The circles are the measurements with the value denoted by their shade of gray. Similarly, the crosses are zero (or very small) measurements.

Table 2
Review of models and optimization techniques.

Method	Solved problem	Variables	Optimization type
OLS	(1)	continuous only	continuous quadratic
WLS	(9)	continuous only	continuous quadratic
Tikhonov reg.	(11)	continuous only	continuous quadratic
LASSO	(12)	continuous only	continuous quadratic
Sparse opt.	(13)	integer and continuous	mixed-integer quadratic

This behaviour is suppressed for the sparse solutions which are presented in the right column of Fig. 6. We can see that the nonzero elements of the solution with $k_{tol} = 5$ are subset of nonzero elements of the solution with $k_{tol} = 10$ and similarly when k_{tol} is increased to 15. We would like to stress again that when using the sparse solutions, we do not have to perform any procedure of regularizing the sensitivity matrix or taking care of release elements with small or huge value.

We provide numerical summary of Fig. 6 in Table 1. We can see that the release of the solutions of LASSO and all integer problems are very similar and approximately half the true release. The release of the WLS solution is almost exactly as the true solution but this is caused mainly by the huge release at one of the later time instant. Since the mean square error is reciprocal to the coefficient of determination R_{WLS}^2 , we will comment only on the latter one. All the solutions besides the true have very similar coefficient of determination around 51%. Based on the coefficient of determination for the true solution, one may think that the model is totally inappropriate. However, this is not true as we show in Fig. 7. The last column shows the sparsity. It is not surprising that only the solutions obtained by the sparse optimization are indeed sparse.

In Fig. 7, we depict the position of the plume (released substance) at one particular time instant as predicted from the atmospheric model, on which the computation of the sensitivity matrix is based. The plume is showed in gray scale. The crosses denote measurements with zero or negligible value. The circles denote measurements with positive values, the value is depicted by the shade of the circle. In Fig. 7a, we can find the plume corresponding to the true solution while in Fig. 7b, we see plume corresponding to the solution of sparse optimization with $k_{tol} = 10$. Even though the coefficient of determination suggested that the true solution is not good, this figure tells that by slightly moving the plume (for example via an elastic transformation of the underlying mesh), we could obtain a very good fit as well.

5. Conclusion

In this paper, we have considered the identification of a source term in atmospheric modelling. First, we have mentioned the well-known technique of the ordinary least squares and showed the standard regularization techniques of the Tikhonov and LASSO regularization. After doing so, we have proposed to look for sparse solutions instead of using classical the above-mentioned methods. We can see the basic information about these methods in Table 2. As suggested in this table, all methods besides the sparse optimization are rather simple to perform and amount to solving a quadratic program. Besides suggesting to use the sparse optimization, we have also proposed to extend the model by adding a weighting matrix W which is based purely on the topology of the measurements and not on the measurement values as is the case in Bayesian modelling.

To verify the proposed methods numerically, we have applied them on the European Tracer Experiment. The obtained results show good results of our techniques. Even though they are computationally more demanding, it is much simpler to choose the parameters in our case. Moreover, it is not necessary to distinguish between small or high values of the release vector.

For the future, we see several possible extensions. As we have already mentioned, the problem may be computationally demanding. Thus, instead of solving the integer sparse problem, it may be possible to use its continuous approximations, which still provide a sparse solution and may be much easier to solve. This holds true especially for large problems. For a second possible extension, we intend to add additional constraints into the model, which may represent for example the maximal ratio of two substances when there are more of them in the model.

References

- Achterberg, T., 2009. SCIP: solving constraint integer programs. *Math. Program. Comput.* 1 (1), 1–41.

- Albert, A., 1972. Regression and the Moore–Penrose Pseudoinverse. Academic Press.
- Anfossi, D., Castelli, S.T., 2014. Atmospheric tracer experiment uncertainties related to model evaluation. *Environ. Model. Softw.* 51, 166–172.
- Aster, R.C., Borchers, B., Thurber, C.H., 2012. Parameter Estimation and Inverse Problems. Academic Press.
- Bennett, N.D., Croke, B.F., Guariso, G., Guillaume, J.H., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. *Environ. Model. Softw.* 40, 1–20.
- Bertsimas, D., Shioda, R., 2009. Algorithm for cardinality-constrained quadratic optimization. *Comput. Optim. Appl.* 43 (1), 1–22.
- Bocquet, M., 2008. Inverse modelling of atmospheric tracers: non-gaussian methods and second-order sensitivity analysis. *Nonlinear Process. Geophys.* 15, 127–143.
- Bruckstein, A.M., Donoho, D.L., Elad, M., 2009. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.* 51 (1), 34–81.
- Clappier, A., Pisoni, E., Thunis, P., 2015. A new approach to design source–receptor relationships for air quality modelling. *Environ. Model. Softw.* 74, 66–74.
- Donoho, D.L., Tsai, Y., 2008. Fast solution of l_1 -norm minimization problems when the solution may be sparse. *IEEE Trans. Inf. Theory* 54 (11), 4789–4812.
- Foucart, S., Rauhut, H., 2013. A Mathematical Introduction to Compressive Sensing. Springer.
- Gaspari, G., Cohn, S.E., 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* 125 (554), 723–757.
- Grant, M., Boyd, S., 2008. Graph implementations for nonsmooth convex programs. In: Blondel, V., Boyd, S., Kimura, H. (Eds.), *Recent Advances in Learning and Control. Lecture Notes in Control and Information Sciences*. Springer-Verlag Limited, pp. 95–110.
- Hamill, T.M., Whitaker, J.S., Snyder, C., 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.* 129, 2776–2790.
- Mairal, J., Bach, F., Ponce, J., Sapiro, G., 2010. Online learning for matrix factorization and sparse coding. *J. Mach. Learn. Res.* 11, 19–60.
- Martinez-Camara, M., Béjar Haro, B., Stohl, A., Vetterli, M., 2014a. A robust method for inverse transport modeling of atmospheric emissions using blind outlier detection. *Geosci. Model Dev.* 7 (5), 2303–2311.
- Martinez-Camara, M., Dokmanić, I., Ranieri, J., Scheibler, R., Vetterli, M., Stohl, A., 2013. The Fukushima inverse problem. In: 38th Int. Conf. Acoust. Speech, Signal Process, pp. 4330–4334.
- Martinez-Camara, M., Vetterli, M., Stohl, A., 2014b. Outlier removal for improved source estimation in atmospheric inverse problems. In: *IEEE Int. Conf. Acoust. Speech Signal Process*, 2, pp. 6820–6824.
- Miettinen, K., 1999. *Nonlinear Multiobjective Optimization*, 12. Springer.
- Miller, S.M., Michalak, A.M., Levi, P.J., 2014. Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions. *Geosci. Model Dev.* 7, 303–315.
- Nocedal, J., Wright, S.J., 2006. *Numerical Optimization*. Springer.
- Osorio-Murillo, C.A., Over, M.W., Savoy, H., Ames, D.P., Rubin, Y., 2015. Software framework for inverse modeling and uncertainty characterization. *Environ. Model. Softw.* 66, 98–109.
- Pétron, G., Granier, C., Khatatov, B., Lamarque, J.-F., Yudin, V., Müller, J.-F., Gille, J., 2002. Inverse modeling of carbon monoxide surface emissions using climate monitoring and diagnostics laboratory network observations. *J. Geophys. Res. Atmos.* 107 (D24), 10–23. ACH 10–1–ACH.
- Seibert, P., Frank, A., 2004. Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode. *Atmos. Chem. Phys.* 4, 51–63.
- Šmídl, V., Hofman, R., 2014. Efficient sequential Monte Carlo sampling for continuous monitoring of a radiation situation. *Technometrics* 56 (4), 514–528.
- Stewart, L.M., Dance, S.L., Nichols, N.K., 2008. Correlated observation errors in data assimilation. *Int. J. Numer. Meth. Fluids* 56, 1521–1527.
- Stohl, A., Prata, A.J., Eckhardt, S., Clarisse, L., Durant, A., Henne, S., Kristiansen, N.I., Minikin, A., Schumann, U., Seibert, P., Stebel, K., Thomas, H.E., Thorsteinsson, T., Tørseth, K., Weinzierl, B., 2011. Determination of time- and height-resolved volcanic ash emissions and their use for quantitative ash dispersion modeling: the 2010 Eyjafjallajökull eruption. *Atmos. Chem. Phys.* 11 (9), 4333–4351.
- Stohl, A., Seibert, P., Wotawa, G., Arnold, D., Burkhart, J.F., Eckhardt, S., Tapia, C., Vargas, A., Yasunari, T.J., 2012. 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 (5), 2313–2343.
- Sun, X., Zheng, X., Li, D., 2013. Recent advances in mathematical programming with semi-continuous variables and cardinality constraint. *J. Oper. Res. Soc. China* 1 (1), 55–77.
- Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B* 58 (1), 267–288.
- Tipping, E., 2002. *Inverse Problems in Atmospheric Constituent Transport*. Cambridge University Press.
- Vedrenne, M., Borge, R., Lumbreras, J., Rodríguez, M.E., 2014. Advancements in the design and validation of an air pollution integrated assessment model for Spain. *Environ. Model. Softw.* 57, 177–191.