Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Geoscientific Model Development Discussions

Optimization of an Urban Monitoring Network for Retrieving an Unknown Point Source Emission

Hamza Kouichi[1], Pierre Ngae[1], Pramod Kumar[1], Amir-Ali Feiz[1], and Nadir Bekka[1,2]

[1]LMEE, Université d'Evry Val-d'Essonne, 40 Rue du Pelvoux 91020 Courcouronnes, France
[2]LSA, Université Saad Dahlab-Blida, 09130 Blida, Algérie

Correspondence: Pramod KUMAR (pramod.kumar@univ-evry.fr)

Abstract. This study presents a methodology for the optimization of a monitoring network of sensors measuring the polluting substances in an urban environment with a view to estimate an unknown emission source. The methodology was presented by coupling the Simulated Annealing algorithm with the renormalization inversion technique and the Computational Fluid Dynamics (CFD) modeling approach. Performance of a network was analyzed by reconstructing the unknown continuous point emissions using the concentration measurements from the sensors in that optimized network. This approach was successfully applied and validated with 20 trials of the Mock Urban Setting Test (MUST) tracer field experiment in an urban-like environment. The optimal networks in the MUST urban region enabled to reduce the size of original network (40-sensors) to $\sim 1/3^{rd}$ (13-sensors) and to $1/4^{th}$ (10-sensors). The 10 and 13 sensors optimal networks have estimated the averaged location errors of 19.20 m and 17.42 m, respectively, which are comparable to 14.62 m from the original 40-sensors network. In 80% trials, emission rates with the 10 and 13 sensors networks were estimated within a factor of two which are also comparable to 75% from the original network. This study presents the first application of the renormalization data-assimilation approach for the optimal network design to estimate a continuous point source emission in an urban-like environment.

1 Introduction

In case of an accidental or deliberated release of a hazardous contaminant in the densely populated urban or industrial regions, it is important to accurately retrieve the location and the intensity of that unknown emission source for the risk assessment, emergency response and mitigation strategies by the concern authority. This retrieval of an unknown source in various source reconstruction methodologies is completely dependent on the contaminant's concentrations detected by some pre-deployed sensors in that affected or a nearby region. However, pre-deployment of these limited number of sensors in that region required an optimal strategy for the establishment of an optimized monitoring network to achieve the maximum information from a limited and noisy set of the concentration measurements. The optimal monitoring networks for the characterization of the unknown emission sources in complex urban or industrial regions is a challenging problem.

The problem to optimize a monitoring network is common and consists in reducing the size of a network of sensors at the level of a city, county, or a neighborhood while retaining its properties. The positions of a small number of sensors are thus optimally determined so as to preserve the objectives of the initial monitoring network. These objectives are generally diverse,

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Geoscientific Model Development Discussions

e.g., detection of an emitting source, analysis of the air quality, triggering of an alert, etc. This study will be focused on the detection of an unknown continuous point source's parameters in an urban-like environment. Using the concentration measurements from an optimally monitoring sensors network, the determination of an unknown or fabricated pollutant emissions from some industrial and accidental releases can be useful for mitigation strategies and also to impose strict actions on such pollutant sources.

This study presents a methodology for the sensor's locations choice, leading to the best network for the estimation of an unknown point point source in a geometrically complex urban environment. This type of network is of great interest in case of an accidental or intentional pollutants release because this makes it possible to estimate the sources of pollution with limited number of measurements from an optimal sensors network. In these conditions, it is necessary to know the location, and the evolution of the spatial extent of the contaminant for an emergency response. The intensity, location and time of the release are often unknown and should be inferred from sensor measurements. The source term estimation (STE) from the measurements is an inverse modeling problem. The establishment of an optimal network requires the availability of sensor concentration measurements, meteorological data, atmospheric dispersion model, choice of a STE procedure and an optimization algorithm.

Ko et al. (1995) showed that the optimization of sensors network is an NP-hard problem. Various optimization algorithms have been proposed to find the best solution, but these methods are not applicable to all the cases especially for large size problems. To solve such problems, the metaheuristic algorithms are efficient. Some studies discussed the optimization of sensor distribution and number for gas emission monitoring, e.g. Ma et al. (2013). Ma et al. (2013) used a direct approach with the Gaussian dispersion model to optimize the sensors networks in homogeneous terrains. However, the present study describes an inverse approach by solving the adjoint transport-diffusion equation with the building-resolving Computational Fluid Dynamics (CFD) model for an urban environment. This approach includes the geometric and flow complexity inherent in the urban region for the optimization process. In this study, the Simulated Annealing (SA) stochastic optimization algorithm (Jiang et al., 2007; Abida et al., 2008; Abida and Bocquet, 2009; Saunier et al., 2009; Kouichi, 2017, etc.) is utilized. The SA algorithm was designed for the statistical physics. It incorporates a probabilistic approach to explore the search space and converges iteratively to the solution. This algorithm is often used and recommended to solve the problems of sensors network optimization (Abida, 2010). The network optimization process consists of finding the best set of sensors that leads to the minimum of a defined cost function. A cost function can be defined as a regularized norm square of the distance between the measurements and forecasts which is also used for the STE (Sharan et al., 2012).

The STE problem for atmospheric dispersion events has been an important topic of much consideration as reviewed in Rao (2007); Hutchinson et al. (2017). Often, the source term is estimated using a network of static sensors deployed in a region. In inverse modeling process, the adjoint source-receptor relationship and concentrations and meteorological datasets are required for the STE. The adjoint source-receptor relationship is defined by an inverse computation of the atmospheric transport dispersion model (Pudykiewicz, 1998). This relationship is often affected by the nonlinearities in the flow-field by building effects in complex scenarios arising in urban environments, where the backward and forward dispersion concentrations will not match. Various inversion methods can be classified in two major categories: probabilistic and deterministic. The probabilistic category treats source parameters as the random variables associated to the probabilities distribution. This includes the Bayesian

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Estimation Theory (Bocquet, 2005; Monache et al., 2008; Yee et al., 2014, etc.), Monte Carlo algorithms using Markov chains (MCMC) (Gamerman and Lopes, 2006; Keats, 2009, etc.) and various stochastic sampling algorithms (Zhang et al., 2014, 2015, etc.). Deterministic methods use cost functions to assess the difference between observed and modeled concentrations and are based on an iterative process to minimize this difference (Seibert, 2001; Penenko et al., 2002; Issartel, 2005; Sharan et al., 2009, 2012; Kumar et al., 2015b, etc.). Among the other approaches, advanced search algorithm like genetic algorithm (Haupt et al., 2006, etc.) or neural network algorithm (Wang et al., 2015, etc.) and other regularization methods (Ma et al., 2017; Zhang et al., 2017, etc.) have been used for the STE. In this study, we focused on the renormalization inversion method (Issartel, 2005), which is deterministic in nature and does not required any initial guess for the source parameters. Initially, the renormalization inversion method was proposed to estimate emission of the distributed sources (Issartel, 2005). Sharan et al. (2009) and other studies have shown that this technique is also effective for estimating the continuous point sources. For these applications, the hypothesis of a linear relationship between the receptors and the source was assumed. For homogeneous terrains, the adjoint functions can analytically be computed based on the Gaussian solution of the diffusion transport equation to estimate a continuous point release. However, the flow-field in urban or industrial environments is quite complex and the asymmetry of the flow and the dispersed plume in urban regions is generated mainly by the presence of buildings and other structures. The Gaussian models are unable to capture the effects of complex urban geometries on adjoint sensitivities between source and receptors and also if dense gases are involved, the Gaussian distribution hypothesis fails. Recently, Kumar et al. (2015b, 2016) have extended the renormalization inversion technique to retrieve an unknown emission source in the urban environments, where a CFD approach was used to generate the adjoint receptors-source relationship. In this process, a coupled CFD-renormalization source reconstruction approach was described for the identification of an unknown continuous point source located at the ground surface or at a horizontal plane corresponding to a known or predefined altitude above the ground surface, or an elevated release in an urban area.

The main objective of this study is to determine the sensors locations choice in an urban domain for an optimal monitoring network dedicated to estimate the location and intensity of a continuously polluting point source. A methodology was proposed to determine an optimal network formed by a predetermined number of sensors, to better characterize a source of pollutant in a complex urban environment. This study deals with a case of reducing the number of sensors in order to obtain an optimal network from an existing network. For this purpose, a predefined network of sensors deployed in an area of interest is considered to determine an optimized network with smaller number of sensors, but, with comparable information. This work explores with two requirements of the optimal networks that modifies the spatial configuration of an existing network by moving the sensors and also reduces the number of sensors of an existing large network. The methodological approach to optimize the monitoring network in urban environment was presented by coupling the SA stochastic algorithm with the renormalization inversion technique and the CFD modeling approach. The concentration measurements from these optimized networks of sensors in 20 trials of the Mock Urban Setting Test (MUST) field tracer experiment were utilize to validate the methodology to retrieve an unknown continuous point source in an urban-like environment.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

2 Source Term Estimation Method: The Renormalization

In the context of an inversion approach, source parameters are often determined using the concentrations measurements at the sensor locations and a source-receptors relationship. The release is considered continuous from a point source located at the ground or at an horizontal plane corresponds to an altitude of a known source height. The renormalization inversion theory to estimate a continuous point release is briefly presented in following subsections.

2.1 Source-Receptor relationship

The source-receptor relationship is an important concept in source reconstruction process and it can be linear or nonlinear. This study deals with the linear relationship, as except from the nonlinear chemical reactions, most of the other processes occurring during the atmospheric transport of trace substances are linear: advection, diffusion, convective mixing, dry and wet deposition, and radioactive decay (Seibert and Frank, 2004). A source-receptor relationship between the measurements and the source function is defined based on a solution of the adjoint transport-diffusion equation that exploits the computed adjoint functions (retroplumes) corresponding to each receptor (Pudykiewicz, 1998; Issartel et al., 2007, etc.). These retroplumes provide a sensitivity information between the source position and the sensor locations. Let's consider a discretized domain of N grid cells in a 2-dimensional space $\mathbf{x} = (x, y)$, a vector of M concentration measurements $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_M)^T \in \mathbb{R}^M$, and an unknown source vector $\mathbf{s}(\mathbf{x}) \in \mathbb{R}^N$ to estimate. The measurements $\boldsymbol{\mu}$ are related to the source vector $\mathbf{s}$ by the use of sensitivity coefficients (also referred as adjoint functions) (Hourdin and Talagrand, 2006). The sensitivity coefficients describe the backward propagation of information from the receptors toward the unknown source. These vectors are related by the following linear relationship:

$$\boldsymbol{\mu} = \mathbf{A}\mathbf{s} + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^M$ is the total measurements error and $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the sensitivity matrix with $\mathbf{A}(\mathbf{x}) = [\mathbf{a}(\mathbf{x}_1), \mathbf{a}(\mathbf{x}_2), ..., \mathbf{a}(\mathbf{x}_N)]$. Here, each column vector $\mathbf{a}(\mathbf{x}_i) \in \mathbb{R}^M$ of the matrix $\mathbf{A}$ represents the potential sensitivity of a grid cell with respect to all M concentration measurements.

For a given set of the concentration measurements $\boldsymbol{\mu}$, the source estimate function $\mathbf{s}(\mathbf{x})$ in Eq. (1) can easily be estimated by formulating a constrained optimization problem. This optimization problem minimizes a cost function $J(\mathbf{s}) = \mathbf{s}^T\mathbf{s}$, subjected to a constraint $\boldsymbol{\epsilon} = \boldsymbol{\mu} - \mathbf{A}\mathbf{s} = 0$. Using the method of Lagrange multipliers, $\mathbf{s}(\mathbf{x})$ can be estimated as a least-norm solution (Kumar et al., 2016):

$$\mathbf{s} = \mathbf{A}^T\mathbf{H}^{-1}\boldsymbol{\mu} \tag{2}$$

where $\mathbf{H}^{-1}$ is inverse of the *Gram matrix* $\mathbf{H} = \mathbf{A}\mathbf{A}^T$. This estimate (Eq. (2)) is not satisfactory because it generates artifacts at the grid cells corresponding to the measurement points. Adjoint functions become singular at these points and have very large values. These large values do not represent a physical reality, but rather an artificial information. This was also highlighted by Issartel et al. (2007) which reduced this artificial information by a process of renormalization.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Geoscientific Model Development Discussions
Open Access
EGU

2.2 Renormalization process

This process involves a weight function in space $\mathbf{W}(\mathbf{x}) \in \mathbb{R}^{N\times N}$, which is purely a diagonal matrix with the diagonal elements $w_{jj} > 0$ such that $\sum_{i=1}^{N} w_{jj} = M$. Introduction of $\mathbf{W}$ transforms the source-receptor relationship in Eq. (1) to:

$$\boldsymbol{\mu} = \mathbf{A}_w \mathbf{W}\mathbf{s} + \boldsymbol{\epsilon}. \tag{3}$$

where the modified sensitivity matrix $\mathbf{A}_w$ is defined as $\mathbf{A}_w = \mathbf{A}\mathbf{W}^{-1} = [\mathbf{a}_w(\mathbf{x}_1), \mathbf{a}_w(\mathbf{x}_2), ..., \mathbf{a}_w(\mathbf{x}_N)]$ in which the column vector $\mathbf{a}_w(\mathbf{x}_i) = \mathbf{a}(\mathbf{x}_i)/w(\mathbf{x}_i)$ of $\mathbf{A}_w$ is the weighted sensitivity vector at $\mathbf{x}_i$. Considering a similar approach that outlined in previous subsection, a new constrained optimization problem can be formulated for Eq. (3) to estimate $\mathbf{s}(\mathbf{x})$. This optimization problem minimizes a cost function $J(\mathbf{s}) = \mathbf{s}^T\mathbf{W}\mathbf{s}$, subjected to a constraint $\boldsymbol{\epsilon} = \boldsymbol{\mu} - \mathbf{A}_w\mathbf{W}\mathbf{s} = 0$, and deduces the following expression $\mathbf{s}_w$ of $\mathbf{s}$ (Kumar et al., 2016):

$$\mathbf{s}_w = \mathbf{A}_w^T \mathbf{H}_w^{-1} \boldsymbol{\mu} \tag{4}$$

where $\mathbf{H}_w^{-1}$ is the inverse of $\mathbf{H}_w = \mathbf{A}_w\mathbf{W}\mathbf{A}_w^T$.

2.3 Weight function

Issartel et al. (2007) demonstrated that a weight function, which reduces the artifacts of the adjoint functions at the measurement points, must verify the following renormalization criterion:

$$\mathbf{a}_w^T(\mathbf{x})\mathbf{H}_w^{-1}\mathbf{a}_w(\mathbf{x}) \equiv 1 \tag{5}$$

Following an iterative algorithm by Issartel et al. (2007), $w(\mathbf{x})$ is determined as:

$$w_0(\mathbf{x}) = 1, \quad \text{and} \quad w_{k+1}(\mathbf{x}) = w_k(\mathbf{x})\sqrt{\mathbf{a}_{wk}^T(\mathbf{x})\mathbf{H}_{wk}^{-1}\mathbf{a}_{wk}(\mathbf{x})} \tag{6}$$

2.4 Identification of point source

Consider a point source of continuous release at a position $\mathbf{x_o} = (x_o, y_o)$ and with the intensity q_o. The point source is thus expressed as a function of the preceding parameters: $\mathbf{s}(\mathbf{x}) = q_o\delta(\mathbf{x} - \mathbf{x_o})$. The relationship between the source and the measurements (Eq. (3)) becomes: $\boldsymbol{\mu} = q_o\mathbf{a}_w(\mathbf{x_o})w(\mathbf{x_o}) + \boldsymbol{\epsilon}$. By replacing the measurement term in Eq. (4), one obtains:

$$\mathbf{s}_w = q_o w(\mathbf{x_o})\mathbf{A}_w^T\mathbf{H}_w^{-1}\mathbf{a}_w(\mathbf{x_o}). \tag{7}$$

$\mathbf{s}_w$ reaches its maximum at position $\mathbf{x_o}$ as the renormalization criterion (Eq. (5)) is satisfied only at this position $\mathbf{x_o}$. Thus, $\mathbf{s}_w(\mathbf{x})$ at $\mathbf{x_o}$ becomes:

$$\mathbf{s}_w(\mathbf{x_o}) = q_o w(\mathbf{x_o}), \tag{8}$$

which estimates the source intensity $q_o = \mathbf{s}_w(\mathbf{x_o})/w(\mathbf{x_o})$.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Geoscientific Model Development Discussions

3 The Combinatorial Optimization of a Monitoring Network

A predefined network of n sensors deployed in an area of interest is considered to determine an optimized network with smaller number of sensors, but with comparable information. For a given number of p sensors such that $p < n$, one determines an array of p sensors among n, which delivers maximum of the information. It is a combinatorial optimization problem that consists of choosing p sensors among n, and thus constituting an optimal network. The optimal network will consist of p sensors for which a defined cost function is minimum. The number of possible choices ${}^nC_p$ (number of combinations of p among n) is very high, when an initial network is sufficiently instrumented (n large) and p is small with respect to n. As the number of combinations to be tested is very large, minimum of a cost function will be evaluated by a stochastic algorithm, viz. simulated annealing (SA).

3.1 Cost function

A cost function is defined as a function that minimizes the quadratic distance between the observed and the simulated measurements according to the $\mathbf{H}_w$ norm. $\mathbf{H}_w$ is the *Gram matrix* defined in a previous section 2.2. As the cost function is convex, its minimum value must also correspond to the maximum intensity of the source. The quadratic distance between the real and the simulated concentration measurements according to the $\mathbf{H}_w$ norm is given by :

$$J_s = \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2_{\mathbf{H}_w^{-1}} = \frac{1}{2}\left[(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^T \mathbf{H}_w^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})\right] \tag{9}$$

When considering a point source, $\hat{\boldsymbol{\mu}}$ is written by $\hat{\boldsymbol{\mu}} = q\mathbf{a}_w(\mathbf{x})w(\mathbf{x})$, where q and $\mathbf{x}$ are respectively the intensity and the position vector of a source. By replacing $\hat{\boldsymbol{\mu}}$ in Eq. (9), one obtains:

$$J_s = J_s(q,\mathbf{x}) = \frac{1}{2}\left[(\boldsymbol{\mu} - q\mathbf{a}_w(\mathbf{x})w(\mathbf{x}))^T H_w^{-1} (\boldsymbol{\mu} - q\mathbf{a}_w(\mathbf{x})w(\mathbf{x}))\right] \tag{10}$$

The conditions of maximum intensity and convexity are expressed by two following equations:

$$\frac{\partial J_s(q,\mathbf{x})}{\partial q} = 0 \tag{11}$$

$$\frac{\partial^2 J_s(q,\mathbf{x})}{\partial q^2} > 0 \tag{12}$$

The second condition (Eq. (12)) is always satisfied as $\frac{\partial^2 J(q,\mathbf{x})}{\partial q^2} = w^2(\mathbf{x}) > 0, \forall \mathbf{x}$ (Sharan et al., 2012). Sharan et al. (2012) has also shown that the first condition (Eq. (11)) leads to following expression of J_s:

$$J_s(\mathbf{x}) = \frac{1}{2}\left[1 - \frac{\mathbf{s}_w^2}{\boldsymbol{\mu}^T \mathbf{H}_w^{-1} \boldsymbol{\mu}}\right] \tag{13}$$

Considering Eq. (13), it appears that the minimization of $J_s(\mathbf{x})$ also corresponds to the maximization of the term $\frac{\mathbf{s}_w^2}{\boldsymbol{\mu}^T \mathbf{H}_w^{-1} \boldsymbol{\mu}}$. A global minimum of the cost function is evaluated by the SA algorithm.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Geoscientific Model Development Discussions

3.2 Simulated Annealing (SA) algorithm for the Sensor's Network Optimization

The SA algorithm is a random optimization technique based on an analogy with thermodynamics. To find a global minimum, it incorporates the temperature into a minimization procedure. So at high temperature (starting temperature), the space of solution is widely explored, while at lower temperature the exploration is restricted. The algorithm is stopped when the cold temperature is reached. It is necessary to choose the law of decreasing temperature, called as cooling schedule. Different approaches to parameterize the SA are explored in Siarry (2016). Kirkpatrick et al. (1983) proposed an average probability to determine the initial (starting) temperature. Nourani and Andresen (1998) compared the most used cooling schedules (exponential, logarithmic, and linear). The SA algorithm starts minimization of an objective function at annealing temperature from a single stochastic point, then it searches for the minimal solutions by attempting all the points in search domain with respect to their value of the temperature. The algorithm is depicted in a flow diagram in Figure 2 and a step by step implementation of the SA procedure for an optimized monitoring network in an urban environment is described as follows:

Step 1. Parameters setting and initialization

Network parameters (n and p): n is the number of possible locations of the sensors and p is the optimal network number of sensors.

Starting temperature (T_0): T_0 is also called the highest temperature. It was determined from the Metropolis law: $T_0 = -\frac{\overline{(\Delta J_s)}}{log(P_0)}$, where $\overline{(\Delta J_s)}$ is an average of the difference of cost functions calculated for a large number of cases. P_0 is an acceptance probability and following the recommendations of Kirkpatrick et al. (1983), it was set to 0.8. Start iterations ($I_{tt} = 0$).

Length of the bearing (L_{max}): A length of the bearing is the number of iterations to be performed at each temperature level. An equilibrium is reached for this number of iterations and any significant improvement of the cost function can be expected. No general rule is proposed to determine a suitable length. This number is often constant and proportional to the size of the problem.

The temperature decay factor (θ): The temperature remains constant for L_{max} iterations corresponding to each bearing. We used the exponential schedule due to its efficiency as denoted by Nourani and Andresen (1998). Then, the temperature decreases law between two bearings varies as: $T_{b+1} = \theta T_b$, with $0 < \theta < 1$, where b represents a bearing. So, it was retained a decay pattern by the bearings.

The cold temperature (T_{cold}): T_{cold} is often called the stopping temperature. There is no clear rule to set this parameter. It is possible to stop calculations when no improvement in the cost function is observed during a large number of combinations. One can estimate this number and take into account the maximum length L_{max} of each bearing, thus the cold temperature can be expressed as a fraction of the starting temperature T_0.

Assigning the first best set of sensors, $\mathbf{x}_{Best} \leftarrow \mathbf{x}_{rand}(p,n)$: $\mathbf{x}_{rand}(p,n)$ corresponds to a vector of p sensors locations randomly chosen among the *n* possible locations. A new solution is randomly explored. This vector is assigned to the first 'best' set of sensors.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Step 2. Assigning a new set of sensors

$\mathbf{x}_{new} \leftarrow \mathbf{x}_{rand}(p,n)$, where $\mathbf{x}_{rand}(p,n)$ corresponds to a vector of p sensors locations randomly chosen among the n possible locations. This vector is assigned to a new set $(\mathbf{x}_{new})$ of the sensors.

Step 3. Cost difference

Given a sensor location $\mathbf{x}_{new}$, the cost function $J_s(\mathbf{x}_{new})$ is computed as follows:

- set $\boldsymbol{\mu}$ vector by using the measurements at the $\mathbf{x}_{new}$ locations,
- set rows of matrix $\mathbf{A}$ using the sensitivity at the $\mathbf{x}_{new}$ locations,
- determine $w(\mathbf{x})$, $\mathbf{H}_w$, and $\mathbf{a}_w$ iteratively using the algorithm in Eq. (6),
- compute the source term $\mathbf{s}_w(\mathbf{x})$ using Eq. (4),
- compute the cost function $J_s(\mathbf{x}_{new})$ using Eq. (13).

$J_s(\mathbf{x}_{best})$ is computed like $J_s(\mathbf{x}_{new})$ using the same precedent steps. A cost difference is then calculated using $\Delta J_s = J_s(\mathbf{x}_{new}) - J_s(\mathbf{x}_{best})$. Increment the iterations $(I_{tt} \leftarrow I_{tt} + 1)$.

Step 4. Test of sign of ΔJ_s

If $\Delta J_s < 0$, the error associated with $\mathbf{x}_{new}$ is less than that with $\mathbf{x}_{best}$ and thus $\mathbf{x}_{new}$ will become the next 'best network' (*Step* 6). If this condition is not satisfied, the algorithm can jump out of a local minimum (*Step* 5).

Step 5. Conditional jump

When $\Delta J_s > 0$, the algorithm has ability to jump out any local minima if condition: $P_{01} \leq \exp(-\frac{\Delta J_s}{T})$ is satisfied, where P_{01} is the acceptance probability (a random number between 0 and 1), and T is the current annealing temperature. It means that $\mathbf{x}_{new}$ will be the next 'best network' even if the associated error is greater than that of $\mathbf{x}_{best}$. If $P_{01} > \exp(-\frac{\Delta J_s}{T})$, go to *Step* 7.

Step 6. Update $\mathbf{x}_{best}$

In this step, $\mathbf{x}_{best}$ is updated by $\mathbf{x}_{new}$.

Step 7. Maximum iteration check

If the maximum number of iterations of a bearing (L_{max}) is reached, a state of equilibrium is then achieved for this temperature and one can cool the actual temperature (*Step* 8). If not, continue iterations (*Step* 2).

Step 8. Temperature cooling

Temperature is cooled using the cooling schedule and iteration variable is reset to zero.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Step 9. Cold temperature test

The cold temperature (T_{cold}) is also known as the stopping temperature. If this temperature is reached, the algorithm stopped. When T_{cold} is not reached, other temperature bearing are performed using the cooling schedule.

Step 10. Optimal network

At this step, the last best network $\mathbf{x}_{best}$ is the optimal network. Source parameters are then estimated using the concentration measurements and retroplumes only for sensors from the obtained optimal network as: (i) $\mathbf{x}_0$ is estimated at position of the maximum of the source estimate function $\mathbf{s}_w(\mathbf{x})$, and (ii) the intensity q_0 is given by $q_0 = \mathbf{s}_w(\mathbf{x}_0)/w(\mathbf{x}_0)$.

In stochastic optimization algorithms, especially in the SA, it was observed that there is no guarantee for the convergence of the algorithm with such a strong cooling (Cohn and Fielding, 1999; Abida et al., 2008). However, chances are that a near-optimal network configuration can be reached. Due to this, one or more near-optimal networks can be obtained from this methodology that satisfy the conditions of near overall optimum condition.

4 The Mock Urban Setting Test (MUST) Tracer Field Network

The MUST field experiment was conducted by the Defense Threat Reduction Agency (DTRA) in 2001. It was aimed to help developing and validating the numerical models for flow and dispersion in an idealized urban environment. The experimental design and observations are described in detail in Biltoft (2001) and Yee and Biltoft (2004). In this experiment, an urban canopy was represented by a grid of 120 containers. These containers were arranged along 12 rows and 10 columns on the army ground in the Utah desert, USA. Each container has dimensions of 2.54 m high, 12.2 m long and 2.42 m wide. The spacing between the horizontal lines is 12.9 m, while the columns are separated by a distance of 7.9 m. The total area thus formed is approximately 200×200 m^2. The experiment consists of 63 releases of a flammable gas (propylene C_3H_6) that is not dangerous or harmful in quantities and could be released through the dissemination system into the open atmosphere (Biltoft, 2001). Different wind conditions (direction, speed, atmospheric stability) as well as different positions for gas emissions (inside or outside the MUST urban canopy at different heights) were considered. These gas emissions were carried out under stable, very stable, and neutral stability conditions. In this study, 20 trials in various atmospheric stability conditions are selected and the meteorological variables are taken from an analysis of meteorological and micro-meteorological observations in Yee and Biltoft (2004) (Table 1). It is noted that the errors related to meteorological data can affects the accuracy of the source term estimation (Zhang et al., 2014, 2015), although this error is not considered in this study. In each trial, the gas was continuously released for ≈ 15 min, during which the concentration measurements were made. These concentration measurements were carried out by 48 photoionization detectors (PIDs). 40 sensors were positioned on four horizontal lines at 1.6 m height (Figure 1) and 8 sensors were deployed in vertical direction at a tower located approximately in center of the MUST array.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Geoscientific Model Development Discussions
Open Access
EGU

5 CFD Modelling for Retroplumes in an Urban Environment

The flow-field in atmospheric dispersion models in geometrically complex urban or industrial environments cannot be considered as homogeneous throughout the computational domain. This is because the buildings and other structures in that region influence and divert the flow into unexpected directions. Consequently, the dispersion of a pollutant and computations of the adjoint functions (retroplumes) are affected by the flow-field induced by these structures in an urban region. Recently, Kumar et al. (2015b) utilized a CFD model to compute the flow-field and then the retroplumes for all selected trials in the MUST experiment. A CFD model fluidyn-PANACHE was utilized to calculate the flow-field, considering a subdomain of calculation (whose dimensions are 250 $\times$ 225 m^2 with a height of 100 m) that consists the MUST urban array created by the containers, sources, receptors, and other instruments in this experiment. This subdomain is embedded in a larger computational domain (dimensions of 800 $\times$ 800 m^2 with a height of 200 m) to ensure a smooth transition of the flow between the edges of the domain and the obstacles zone. This extension of the outer domain far from the main experimental site is essential to reduce effects of the inflow boundary conditions imposed at inlet of the outer domain. A more detailed description about the CFD model and its simulations, e.g., boundary conditions, turbulence model, etc. is briefly presented in Supplementary Information (SI). An unstructured mesh was generated in both domains with more refinement in the urbanized area in inner subdomain and at the positions of receptors, thus generating 2849276 meshes.

The simulations results with fluidyn-PANACHE in each MUST trial were obtained with inflow boundary conditions from vertical profiles of the wind (U), the turbulent kinetic energy (k) and its dissipation rate (ϵ). These inflow profiles include: *(i) Wind profile*: Gryning et al. (2007) profiles in stable and neutral conditions and a profile based on the stability function by Beljaars and Holtslag (1991) in very stable conditions, *(ii) Temperature profile:* Monin-Obukhov similarity theory based logarithmic profiles, *(iii) Turbulence profiles:* k and ϵ profiles are based on an approximate analytical solution of one-dimensional $k-\epsilon$ prognostic equations (Yang et al., 2009). The atmospheric stability effects in the CFD model fluidyn-PANACHE are included through the inflow boundary condition (via advection). fluidyn-PANACHE includes a Planetary Boundary Layer (PBL) model that serves as an interface between the meteorological observations and the boundary conditions required by the CFD solver. The observed turbulence parameters, e.g. (i) sensible heat flux (Q_h), the Obukhov length (L), (iii) surface friction velocity (u_*) and the temperature scale (θ_*) were used to derive the vertical profiles of mean velocity and potential temperature. Since the released gas propylene is heavier than the air and would behave as a dense gas, a buoyancy model was used to model the body force term in the Navier-Stokes equations. The buoyancy model is suitable for the dispersion of heavy gases where density difference in the vertical direction drives the body force.

In order to compute the retroplumes in each MUST trial, firstly the CFD simulations were performed to compute the converged flow-field in computational domain, secondly the flow-field is reversed and used in the standard advection-diffusion equation to compute the adjoint functions $\mathbf{a}_i(\mathbf{x})$. In this computation of the retroplumes corresponding to each receptor in a selected trial, the advection-diffusion equation is solved by considering a receptor as a virtual point source with unit release rate at the height of that receptor. Also, the meteorological conditions remained invariant during the whole experimental period in a trial. The details about the retroplumes and the correlated theory of the duality verification (i.e. comparison of the con-

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Geoscientific Model Development Discussions

centrations predicted with the forward (direct) model and the adjoint model) for all 20 trials of the MUST field experiment are given in Kumar et al. (2015a). Since we are concerned to establish an optimized monitoring network in a domain that contains the MUST urban array, the retroplumes are computed in the inner subdomain only. Consequently all the computations for an optimized monitoring network were carried out in the inner subdomain only. The sensors in the optimized monitoring network are supposed to deploy on a fixed vertical height above the ground surface. Accordingly, the retroplumes corresponding to only 40 receptors at 1.6 m height were utilized in computations for the optimized monitoring networks in the MUST urban environment.

6 Results and Discussion

The calculations were performed by coupling the SA algorithm to a deterministic renormalization inversion algorithm and the CFD adjoint fields to optimize the monitoring network in an urban environment of the MUST field experiment. The network optimization process consists of finding the best set of sensors that leads to the lowest cost function. In this study, two optimal monitoring networks were formed to have the sizes reduced by one-third (~ 13 sensors) and one-fourth (10 sensors) respectively compared to the initial network of 40 sensors at 1.6 m above the ground surface. The optimization calculations were performed using Matlab on a computer with configuration "Intel® Core™ i7-4790 CPU @ 3.60 GHz and 16 GB RAM". The averaged computational time for optimization of one 10 sensors network was ≈ 2.5 hrs and ≈ 8.5 hrs for 13 sensors network. In computations, a value of parameter $T_0 = 10$ was fixed according to the scale of cost function and using the methodology described in *Step* 1 and $T_{cold} = 10^{-13}$ was used for both the optimal sensors networks. θ is a decay factor of the temperature for an exponential cooling schedule that describes a procedure of the temperature decrease. The best cooling schedule is the exponential decay as demonstrated by Nourani and Andresen (1998); Cohn and Fielding (1999). θ was fixed as 0.9 following the recommendation in literature (Siarry, 2014). This value allows a sufficiently slow cooling in order to give more chance to the algorithm to explore a large search space and to avoid the local minima. L_{max} is taken as 100 & 200 for 10 & 13 sensors networks, respectively, following the recommendation in Siarry (2014) and according to number of the possible combinations that increases with the number of sensors (8.5×10^8 for 10 sensors and 1.2×10^{10} for 13 sensors).

Figure 3 shows the optimal networks of 10 and 13 sensors respectively for three representative trials 5 (very stable), 11 (neutral), and 19 (stable) in the MUST urban array. These three trials correspond to one trial each in neutral, stable, and very stable atmospheric conditions during the release. The optimal monitoring networks of 10 and 13 sensors for all selected 20 MUST trials are shown in SI Figures S1.1&S1.2.

In order to analyze the performance of the optimal monitoring configurations, the source reconstructions were performed to estimate the unknown location and the intensity of a continuous point release. These source reconstruction results were obtained from the optimal monitoring networks formed by 10 and 13 sensors in each MUST trial. In this, the retroplumes and the concentration measurements were utilized from the sensors corresponding to these optimal networks. The retroplumes were computed using CFD simulations, considering the dispersion in a complex terrain. The source reconstruction results from

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Geoscientific Model Development Discussions

both the optimal monitoring networks were also compared with results computed from the initial MUST network formed by 40 sensors.

Source estimation results from the different monitoring networks are shown in Table 2 for all 20 selected trials of the MUST experiment. These results are presented in terms of the location error (E_l^p), which is an euclidean distance between the estimated and the true source location, and E_q^p, a ratio of the estimated to the true source intensity. The corresponding monitoring network is represented by a superscript p (representing the number of sensors in a optimal network) on E_l^p and E_q^p. For a given trial, a parameter *skeleton* represents the common sensors between two optimal networks of different sizes (with 10 and 13 sensors). These results exhibit that the SA algorithm coupled with renormalization inversion theory and CFD modeling approach has succeeded in proposing the good optimal monitoring networks to estimate the unknown emissions in an urban environment.

Figure 4 shows isopleths of the renormalized weight function (also called as the visibility function) and the normalized source estimate function $\mathbf{s}_w^n(\mathbf{x}) = \mathbf{s}_w(\mathbf{x})/\max(\mathbf{s}_w(\mathbf{x}))$ correspond to both optimal monitoring networks for three representative trials (e.g. 5, 11, and 19) of the MUST experiment. These isopleths for all selected 20 MUST trials are shown in SI Figures S2. A statistical parameter, factor of g (FAg), for the source reconstruction results from each monitoring network is presented in Table 3, where FAg represents the percentage no. of trials in which the source intensity is estimated within a factor of g. The statistics calculated with 40 sensors network show that the average location error for all 20 trials is 14.62 m, and in 75% of the trials, the intensity of the source is estimated within a factor of two. In 90% of the trials, intensity was estimated within a factor of three and within a factor of four in 95% trials (Table 3). If trial 2 is considered, large location errors (greater than 30 m) and the intensity values ranged between a factor of three to five, were observed independently of the number of sensors in the networks (Table 2). If we consider the trials 15, 16, & 20, it was noted that the larger location errors do not necessarily correspond to the high intensity errors (Table 2).

From distribution of the optimized sensors in networks in Figures 3 for trials 5, 11 & 19 and SI Figures S1.1&S1.2 for all selected trials, it was noted that a larger number of sensors are close to the source position in the optimal networks in most of the trials. This tendency makes it possible for sensors from an optimal network to monitor the region where a source can be located and it can be explained in terms of the visibility function. The visibility function includes the natural information associated with a monitoring network for the source retrieval in a domain and physically interprets the extent of regions seen by the network (Issartel, 2005; Sharan et al., 2009). The visibility function is independent of the effective values of the concentration measurements and depends only on geometry of the monitoring network. Hence, this leads to a priori information about the unknown source apparent to the monitoring network. It was observed that the visibility functions have significant levels at the source positions (Figures 4 & S2).

The source reconstruction results from the optimal monitoring networks formed by 10 sensors have an averaged location error (E_l^{10}) of 19.20 m for all 20 trials in the MUST experiment (Tables 2&3). In most of the trials, the location and the intensity of a continuous point emission are estimated accurately and close to the true source parameters. The location error is minimum in trial 14 (E_l^{10} = 5.50 m) and maximum in trial 2 (E_l^{10} = 56.88 m) (Table 2). For this configuration of the optimal

sensors network, the source intensity in 80% of the trials are estimated within a factor of two to their true release rates (Tables 2&3).

For all 20 trials, the averaged location error E_l^{13} is 17.42 m for the optimal networks formed by 13 sensors, which is smaller than the averaged E_l^{10} = 19.20 m obtained with 10 sensors (Tables 2&3). The location error is observed minimum in trial 5 (E_l^{13} = 2.13 m) and maximum in trial 16 (E_l^{13} = 63.04 m) (Table 2). For this optimal network, in 80% of the trials, the source intensity is estimated within a factor of two. It was noted that the increase in the number of sensors in a network has little influence on the accuracy of the estimated intensity (Tables 2&3).

In some trials, it was also noted that the distance of an estimated source to real source can decreases with a decrease in sensors number and are also increases with the number of sensors in some other cases. It is because the information added by a new sensor was not necessarily beneficial. As it is noticeable that in a particular meteorological condition (i.e. wind direction, speed and atmospheric stability), some of the sensors in a network may have little contribution to the STE. So, increasing the number of the sensors may not always provide the best estimation because with addition of the more no. of sensors, we also add more model and measurements errors in the estimation process. These errors can affect the source estimation results in some trials. In some cases, it may also depend on sensitiveness of the added sensor's position in an extended optimal network to the source estimation. It is also noted that for a monitoring network, not only the number of sensors but also the sensors distribution form (or sensor position) affect the information captured from network.

In fact, both optimal networks for each trial show a diversity of structures independently of the number of sensors considered. For this, the *skeleton* was used to analyze the heterogeneity of the structures of different optimal networks. A skeleton with 7 sensors is considered as a strong common base for the networks. This is the case for trials 3,6,14,15&20 (Table 2). It is noted that the overall results obtained are comparable (little differences between the results obtained by the networks). For these networks, a strong common base leads to a near global optimum. If we consider networks with a weak common base, the *skeleton* was formed of up to 3 sensors, particularly in trials 1 and 11. The performances do not systematically converge independently of the size of networks. Thus, for trial 1, better results were obtained with a network formed by 13 sensors compared to that by 10 sensors. This result may reflect the fact that the algorithm leading to a near global optimum is contained in the network formed by 13 sensors. For trial 11 also, it was noted that the performances obtained by the two networks are identical. This shows that the networks with different sensors configurations may lead to a near overall optimum.

Considering the networks of intermediate structures, with *skeletons* varying from 4 to 6 sensors, notedly for trials 2,4,5,7,8,9, 12,13,16,17,18,&19, no obvious trend is noticed. These results tend to show that for a given trial, one or more optimal networks can satisfy the conditions of a near overall optimum (to be minimized). The obtained optimal networks may have a more or less common structure (having a greater or lesser number of *skeleton*).

It should be noted that this study deals with the case of reducing number of sensors in order to obtain an optimal network from an existing large network. This optimization was carried out under the constraints of an existing network of the original 40 sensors in the MUST field experiment. If one compares the performances of the obtained optimal monitoring networks with the initial (original) network of 40 sensors in MUST environment, both optimal networks provide satisfactory estimations of unknown source parameters. The 40 sensors network gives an averaged location error of 14.62 m for all trials and the release

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Geoscientific Model Development Discussions
Open Access
EGU

rate were estimated within a factor of two in 75% trials. However, reducing the number of sensors to $1/3^{rd}$ from the original 40 sensors, the 13 sensors optimal networks also give comparable source estimations performance with an averaged location error of 17.42 m. Even with the 13 sensors optimal networks, source intensities in 80% trials were accurately estimated within a factor of two. Similarly for 10 sensors optimal networks, the averaged location error (=19.20 m) is slightly larger than that obtained from 13 and 40 sensors networks. However, reducing the number of sensors to $1/4^{th}$ gives extra advantages in case of the limited available sensors for a network in emergency scenarios of an accidental or deliberated releases in complex urban environments.

7 Conclusions

This study describes an approach for the optimization of a monitoring network of the sensors in a geometrically complex urban environment. It is a matter of reducing the size of networks while retaining the capabilities of detecting an unknown source in an urban region. Given an urban-like environment of the MUST field experiment, the renormalization inversion method was chosen for the Source Term Estimation. It was coupled with the CFD model fluidyn-PANACHE for generation of the adjoint fields. Combinatorial optimization by the simulated annealing consisted in choosing a set of sensors which leads to an optimal monitoring network and allows an accurate unknown source estimation. This study presents the first application of the renormalization data-assimilation approach for the optimal network design to estimate an unknown continuous point release in an urban-like environment.

The numerical calculations were performed by coupling the simulated annealing stochastic algorithm to the renormalization inversion technique and the CFD modeling approach to optimize the monitoring network in urban environment of the MUST field experiment. The optimal networks enabled to reduce size of the original networks (40 sensors) to one-third (13 sensors) and to one-fourth (10 sensors). The 10 and 13 sensors optimal networks have estimated the averaged location errors of 19.20 m and 17.43 m, respectively, and have comparable source estimations performance with an averaged location error of 14.62 m from the original 40 sensors network. In 80% of the trials, the emission rates with 10 and 13 sensors networks were estimated within a factor of two which are also comparable with the factor of two source intensities in 75% trials with the original network.

It was shown that in most of the MUST trials, the number of sensors in optimal networks slightly influences the location error of an estimated source and this error tends to increase as the number of sensors decreases. In 20 MUST trials, an analysis of the networks formed by 10 & 13 sensors revealed the heterogeneity of their structures in an urban domain. It was observed that for some trials, optimal networks had a strong common structure. This tends to prove that a certain number of sensors have a primordial role in determining an unknown source. It would reflect a fact that the disjoint sets of sensors can lead to the best estimate of an unknown source in an urban region. This opens enormous prospects for assessing the relative importance of each sensor in a source reconstruction process in an urban environment. Defining a global optimal network for all meteorological conditions is a complex problem, but of greater importance that one may want to pursue. This challenge consists to define an optimal static network able to reconstruct the sources in all varied meteorological conditions. This information can be of

great importance to determine an optimal monitoring network by reducing the number of sensors for characterization of the unknown emissions in the complex urban or industrial environments.

Data availability. The authors received access to the MUST field experiment dataset from Dr. Marcel Koñig of Leibniz Institute for Tropospheric Research. The MUST database was officially available from the Defense Threat Reduction Agency (DTRA) at https://must-dpg.dpg.army.mil/.

Acknowledgements. The authors would like to thank Dr. Marcel Koñig of Leibniz Institute for Tropospheric Research and the Defense Threat Reduction Agency (DTRA) for providing access to the MUST field experiment dataset. The authors gratefully acknowledge Fluidyn France for use of the CFD model fluidyn-PANACHE. We also thank to Dr. Claude Souprayen from Fluidyn France for useful discussions.

References

Abida, R.: Static and mobile networks design for atmospheric accidental releases monitoring, Theses, Ecole des Ponts ParisTech, https://pastel.archives-ouvertes.fr/pastel-00638050, 2010.

Abida, R. and Bocquet, M.: Targeting of observations for accidental atmospheric release monitoring, Atmospheric Environment, 43, 6312 – 6327, https://doi.org/http://dx.doi.org/10.1016/j.atmosenv.2009.09.029, http://www.sciencedirect.com/science/article/pii/S1352231009008048, 2009.

Abida, R., Bocquet, M., Vercauteren, N., and Isnard, O.: Design of a monitoring network over France in case of a radiological accidental release, Atmospheric Environment, 42, 5205 – 5219, https://doi.org/http://dx.doi.org/10.1016/j.atmosenv.2008.02.065, http://www.sciencedirect.com/science/article/pii/S1352231008002409, 2008.

Beljaars, A. and Holtslag, A.: Flux parameterization over land surfaces for atmospheric models, Journal of Applied Meteorology, 30, 327–341, https://doi.org/10.1175/1520-0450(1991)030<0327:FPOLSF>2.0.CO;2, http://dx.doi.org/10.1175/1520-0450(1991)030<0327:FPOLSF>2.0.CO;2, 1991.

Biltoft, C. A.: Customer Report for Mock Urban Setting Test, Tech. rep., West Desert Test Center, U.S. Army Dugway Proving Ground, Dugway, Utah, 2001.

Bocquet, M.: Reconstruction of an atmospheric tracer source using the principle of maximum entropy. I: Theory, Quarterly Journal of the Royal Meteorological Society, 131, 2191–2208, https://doi.org/10.1256/qj.04.67, http://dx.doi.org/10.1256/qj.04.67, 2005.

Cohn, H. and Fielding, M.: Simulated Annealing: Searching for an Optimal Temperature Schedule, SIAM Journal on Optimization, 9, 779–802, https://doi.org/10.1137/S1052623497329683, https://doi.org/10.1137/S1052623497329683, 1999.

Gamerman, D. and Lopes, H. F.: Markov chain Monte Carlo: stochastic simulation for Bayesian inference, CRC Press, 2006.

Gryning, S.-E., Batchvarova, E., Brummer, B., Jorgensen, H., and Larsen, S.: On the extension of the wind profile over homogeneous terrain beyond the surface boundary layer, Boundary-Layer Meteorology, 124, 251–268, https://doi.org/10.1007/s10546-007-9166-9, http://dx.doi.org/10.1007/s10546-007-9166-9, 2007.

Haupt, S. E., Young, G. S., and Allen, C. T.: Validation of a receptor-dispersion model coupled with a genetic algorithm using synthetic data, Journal of applied meteorology and climatology, 45, 476–490, https://doi.org/10.1175/JAM2359.1, http://dx.doi.org/10.1175/JAM2359.1, 2006.

Hourdin, F. and Talagrand, O.: Eulerian backtracking of atmospheric tracers. I: Adjoint derivation and parametrization of subgrid-scale transport, Quarterly Journal of the Royal Meteorological Society, 132, 567–583, https://doi.org/10.1256/qj.03.198.A, http://dx.doi.org/10.1256/qj.03.198.A, 2006.

Hutchinson, M., Oh, H., and Chen, W.-H.: A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors, Information Fusion, 36, 130 – 148, https://doi.org/https://doi.org/10.1016/j.inffus.2016.11.010, http://www.sciencedirect.com/science/article/pii/S156625351630152X, 2017.

Issartel, J.-P.: Emergence of a tracer source from air concentration measurements, a new strategy for linear assimilation, Atmospheric Chemistry and Physics, 5, 249–273, https://doi.org/10.5194/acp-5-249-2005, http://www.atmos-chem-phys.net/5/249/2005/, 2005.

Issartel, J.-P., Sharan, M., and Modani, M.: An inversion technique to retrieve the source of a tracer with an application to synthetic satellite measurements, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 463, 2863–2886, https://doi.org/10.1098/rspa.2007.1877, http://rspa.royalsocietypublishing.org/content/463/2087/2863.abstract, 2007.

Jiang, Z., de Bruin, S., Heuvelink, G., and Twenhöfel, C.: Optimization of mobile radioactivity monitoring networks, http://edepot.wur.nl/24174, 2007.

Keats, W. A.: Bayesian inference for source determination in the atmospheric environment, Ph.D. thesis, http://hdl.handle.net/10012/4317, 2009.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P.: Optimization by Simulated Annealing, Science, 220, 671–680, https://doi.org/10.1126/science.220.4598.671, http://science.sciencemag.org/content/220/4598/671, 1983.

Ko, C.-W., Lee, J., and Queyranne, M.: An Exact Algorithm for Maximum Entropy Sampling, Operations Research, 43, 684–691, https://doi.org/10.1287/opre.43.4.684, https://doi.org/10.1287/opre.43.4.684, 1995.

Kouichi, H.: Sensors networks optimization for the characterization of atmospheric releases source, Theses, Université Paris Saclay, France, https://hal.archives-ouvertes.fr/tel-01593834, 2017.

Kumar, P., Feiz, A.-A., Ngae, P., Singh, S. K., and Issartel, J.-P.: CFD simulation of short-range plume dispersion from a point release in an urban like environment, Atmospheric Environment, 122, 645 – 656, https://doi.org/http://dx.doi.org/10.1016/j.atmosenv.2015.10.027, http://www.sciencedirect.com/science/article/pii/S1352231015304465, 2015a.

Kumar, P., Feiz, A.-A., Singh, S. K., Ngae, P., and Turbelin, G.: Reconstruction of an atmospheric tracer source in an urban-like environment, Journal of Geophysical Research: Atmospheres, 120, 12 589–12 604, https://doi.org/10.1002/2015JD024110, http://dx.doi.org/10.1002/2015JD024110, 2015b.

Kumar, P., Singh, S. K., Feiz, A.-A., and Ngae, P.: An urban scale inverse modelling for retrieving unknown elevated emissions with building-resolving simulations, Atmospheric Environment, 140, 135 – 146, https://doi.org/http://dx.doi.org/10.1016/j.atmosenv.2016.05.050, http://www.sciencedirect.com/science/article/pii/S1352231016303983, 2016.

Ma, D., Deng, J., and Zhang, Z.: Comparison and improvements of optimization methods for gas emission source identification, Atmospheric Environment, 81, 188 – 198, https://doi.org/https://doi.org/10.1016/j.atmosenv.2013.09.012, http://www.sciencedirect.com/science/article/pii/S1352231013007000, 2013.

Ma, D., Tan, W., Zhang, Z., and Hu, J.: Parameter identification for continuous point emission source based on Tikhonov regularization method coupled with particle swarm optimization algorithm, Journal of Hazardous Materials, 325, 239 – 250, https://doi.org/https://doi.org/10.1016/j.jhazmat.2016.11.071, http://www.sciencedirect.com/science/article/pii/S0304389416311141, 2017.

Monache, L. D., Lundquist, J. K., Kosović, B., Johannesson, G., Dyer, K. M., Aines, R. D., Chow, F. K., Belles, R. D., Hanley, W. G., Larsen, S. C., Loosmore, G. A., Nitao, J. J., Sugiyama, G. A., and Vogt, P. J.: Bayesian Inference and Markov Chain Monte Carlo Sampling to Reconstruct a Contaminant Source on a Continental Scale, Journal of Applied Meteorology and Climatology, 47, 2600–2613, https://doi.org/10.1175/2008JAMC1766.1, https://doi.org/10.1175/2008JAMC1766.1, 2008.

Nourani, Y. and Andresen, B.: A comparison of simulated annealing cooling strategies, Journal of Physics A: Mathematical and General, 31, 8373, http://stacks.iop.org/0305-4470/31/i=41/a=011, 1998.

Penenko, V., Baklanov, A., and Tsvetova, E.: Methods of sensitivity theory and inverse modeling for estimation of source parameters, Future Generation Computer Systems, 18, 661 – 671, https://doi.org/http://dx.doi.org/10.1016/S0167-739X(02)00031-6, http://www.sciencedirect.com/science/article/pii/S0167739X02000316, {ICCS2001}, 2002.

Pudykiewicz, J. A.: Application of adjoint tracer transport equations for evaluating source parameters, Atmospheric Environment, 32, 3039–3050, https://doi.org/http://dx.doi.org/10.1016/S1352-2310(97)00480-9, http://www.sciencedirect.com/science/article/pii/S1352231097004809, 1998.

Rao, K. S.: Source estimation methods for atmospheric dispersion, Atmospheric Environment, 41, 6964 – 6973, https://doi.org/https://doi.org/10.1016/j.atmosenv.2007.04.064, http://www.sciencedirect.com/science/article/pii/S1352231007004086, 2007.

Saunier, O., Bocquet, M., Mathieu, A., and Isnard, O.: Model reduction via principal component truncation for the optimal design of atmospheric monitoring networks, Atmospheric Environment, 43, 4940 – 4950, https://doi.org/http://dx.doi.org/10.1016/j.atmosenv.2009.07.011, http://www.sciencedirect.com/science/article/pii/S1352231009005913, 2009.

Seibert, P.: Inverse Modelling with a Lagrangian Particle Disperion Model: Application to Point Releases Over Limited Time Intervals, pp. 381–389, Springer US, Boston, MA, https://doi.org/10.1007/0-306-47460-3_38, https://doi.org/10.1007/0-306-47460-3_38, 2001.

Seibert, P. and Frank, A.: Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode, Atmospheric Chemistry and Physics, 4, 51–63, 2004.

Sharan, M., Issartel, J.-P., Singh, S. K., and Kumar, P.: An inversion technique for the retrieval of single-point emissions from atmospheric concentration measurements, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 465, 2069–2088, https://doi.org/10.1098/rspa.2008.0402, http://rspa.royalsocietypublishing.org/content/early/2009/04/09/rspa.2008.0402.abstract, 2009.

Sharan, M., Singh, S. K., and Issartel, J.-P.: Least Square Data Assimilation for Identification of the Point Source Emissions, Pure and Applied Geophysics, 169, 483–497, https://doi.org/10.1007/s00024-011-0382-3, http://dx.doi.org/10.1007/s00024-011-0382-3, 2012.

Siarry, P.: Métaheuristiques: Recuits simulé, recherche avec tabous, recherche à voisinages variables, méthodes GRASP, algorithmes évolutionnaires, fourmis artificielles, essaims particulaires et autres méthodes d'optimisation, Editions Eyrolles, 2014.

Siarry, P.: Simulated Annealing, pp. 19–50, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-45403-0_2, http://dx.doi.org/10.1007/978-3-319-45403-0_2, 2016.

Wang, B., Chen, B., and Zhao, J.: The real-time estimation of hazardous gas dispersion by the integration of gas detectors, neural network and gas dispersion models, Journal of Hazardous Materials, 300, 433 – 442, https://doi.org/http://dx.doi.org/10.1016/j.jhazmat.2015.07.028, http://www.sciencedirect.com/science/article/pii/S0304389415005658, 2015.

Yang, Y., Gu, M., Chen, S., and Jin, X.: New inflow boundary conditions for modelling the neutral equilibrium atmospheric boundary layer in computational wind engineering, Journal of Wind Engineering and Industrial Aerodynamics, 97, 88–95, https://doi.org/http://dx.doi.org/10.1016/j.jweia.2008.12.001, http://www.sciencedirect.com/science/article/pii/S0167610508001815, 2009.

Yee, E. and Biltoft, C.: Concentration Fluctuation Measurements in a Plume Dispersing Through a Regular Array of Obstacles, Boundary-Layer Meteorology, 111, 363–415, https://doi.org/10.1023/B:BOUN.0000016496.83909.ee, http://dx.doi.org/10.1023/B%3ABOUN.0000016496.83909.ee, 2004.

Yee, E., Hoffman, I., and Ungar, K.: Bayesian Inference for Source Reconstruction: A Real-World Application, International Scholarly Research Notices, 2014, 1–12, https://doi.org/10.1155/2014/507634, http://dx.doi.org/10.1155/2014/507634, 2014.

Zhang, X., Raskob, W., Landman, C., Trybushnyi, D., and Li, Y.: Sequential multi-nuclide emission rate estimation method based on gamma dose rate measurement for nuclear emergency management, Journal of Hazardous Materials, 325, 288 – 300, https://doi.org/https://doi.org/10.1016/j.jhazmat.2016.10.072, http://www.sciencedirect.com/science/article/pii/S0304389416310032, 2017.

Zhang, X. L., Su, G. F., Yuan, H. Y., Chen, J. G., and Huang, Q. Y.: Modified ensemble Kalman filter for nuclear accident atmospheric dispersion: Prediction improved and source estimated, Journal of Hazardous Materials, 280, 143 – 155, https://doi.org/https://doi.org/10.1016/j.jhazmat.2014.07.064, http://www.sciencedirect.com/science/article/pii/S0304389414006396, 2014.

Zhang, X. L., Su, G. F., Chen, J. G., Raskob, W., Yuan, H. Y., and Huang, Q. Y.: Iterative ensemble Kalman filter for atmospheric dispersion in nuclear accidents: An application to Kincaid tracer experiment, Journal of Hazardous Materials, 297, 329 – 339, https://doi.org/https://doi.org/10.1016/j.jhazmat.2015.05.035, http://www.sciencedirect.com/science/article/pii/S0304389415004318, 2015.

Table 1. The values of the meteorological (wind speed (S_{04}), wind direction (α_{04}) at 4 m level of mast S), turbulence (the Obukhov length (L), friction velocity (u_*), turbulent kinetic energy (k) at 4 m level of tower T), and source parameters (source height (z_s), release duration (t_s), release rate (q_s)) in 20 selected trials of the MUST field experiment (Biltoft, 2001; Yee and Biltoft, 2004). Here, Trial Nos. 1-20 are assigned for just continuation and simplicity and these are not correspond to the same assigned trial no. for a given Trial name in the MUST experiment.

Trial No.	Trial Name (JJJhhmm)	q_s (l/min)	t_s (min)	z_s (m)	S_{04} (m/s)	α_{04} (deg)	u_* (m/s)	L (m)	k (m^2s^{-2})
1	2640138	175	21	0.15	2.35	17	0.26	91	0.359
2	2640246	200	15	0.15	2.01	30	0.25	62	0.306
3	2671852	200	22	0.15	3.06	-49	0.32	330	0.436
4	2671934	200	15	1.8	1.63	-48	0.08	5.8	0.148
5	2672033	200	15	1.8	2.69	-26	0.17	4.8	0.251
6	2672101	200	14	0.15	1.89	-10	0.16	7.7	0.218
7	2672150	200	16	0.15	2.30	36	0.35	150	0.409
8	2672213	200	15	1.8	2.68	30	0.35	150	0.428
9	2672235	200	15	2.6	2.32	36	0.26	48	0.387
10	2672303	200	19	1.8	2.56	17	0.25	74	0.367
11	2681829	225	15	1.8	7.93	-41	1.10	28000	1.46
12	2681849	225	16	0.15	7.26	-50	0.76	2500	0.877
13	2682256	225	15	0.15	5.02	-42	0.66	240	0.877
14	2682320	225	15	2.6	4.55	-39	0.50	170	0.718
15	2682353	225	15	5.2	4.49	-47	0.44	120	0.727
16	2692054	225	22	1.3	3.34	39	0.36	170	0.362
17	2692131	225	17	1.3	4.00	39	0.42	220	0.582
18	2692157	225	15	2.6	2.98	43	0.39	130	0.505
19	2692223	225	15	1.3	2.63	26	0.35	120	0.484
20	2692250	225	17	1.3	3.38	36	0.37	130	0.537

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Geoscientific Model Development Discussions

Table 2. Source estimation results from the different monitoring networks for each selected trial of the MUST field experiment. E_l^p and E_q^p respectively denote the location error (m) and ratio of the estimated to true source intensity with the corresponding monitoring network. Here, the superscript p on E_l^p & E_q^p represents the no. of sensors in an optimal network. *Skeleton* refers to the number of sensors common to the optimal networks of 10 and 13 sensors for a given MUST trial.

Run No.	E_l^{40} (m)	E_l^{13} (m)	E_l^{10} (m)	E_q^{40}	E_q^{13}	E_q^{10}	*Skeleton* sensors
1	3.3	19.6	33.76	0.92	1.04	1.24	3
2	42.9	31.91	56.88	4.01	3.21	5.12	4
3	10.8	9.01	9.01	1.17	0.71	0.71	7
4	22.8	18.07	18.07	0.27	0.83	0.83	6
5	21.9	2.13	11.56	0.57	0.95	0.67	6
6	5.0	6.96	6.96	2.14	1.04	1.04	7
7	12.4	18.85	12.99	0.41	3.11	1.06	4
8	15.8	12.86	15.79	2.22	1.32	1.76	6
9	7.7	8.20	8.08	1.37	3.06	7.55	5
10	8.8	8.00	8.00	1.08	1.08	1.08	8
11	19.8	17.19	17.19	1.67	1.62	1.62	3
12	7.4	5.43	10.22	0.95	0.85	0.2	4
13	7.7	8.63	8.63	0.97	0.78	0.78	4
14	2.2	5.50	5.50	1.42	0.88	0.88	7
15	1.1	30.23	37.98	1.88	0.57	0.17	7
16	26.7	63.04	29.80	1.70	0.29	0.67	5
17	7.0	14.07	23.05	0.90	1.10	1.52	6
18	14.3	12.83	12.83	1.15	1.15	1.15	6
19	22.3	10.77	13.46	1.76	0.99	0.83	6
20	32.5	45.23	44.29	0.83	1.68	1.56	7

Table 3. Statistics for the source reconstruction results from each monitoring network. Here, E_l^p is the averaged location error for all 20 trials corresponding to each network. FAg represents the percentage number of trials in which the source intensity is estimated within a factor of g.

Sensors (p)	E_l^p **(m)**	**FA4**	**FA3**	**FA2**
40	14.62	95	90	75
13	17.42	100	80	80
10	19.20	80	80	80

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

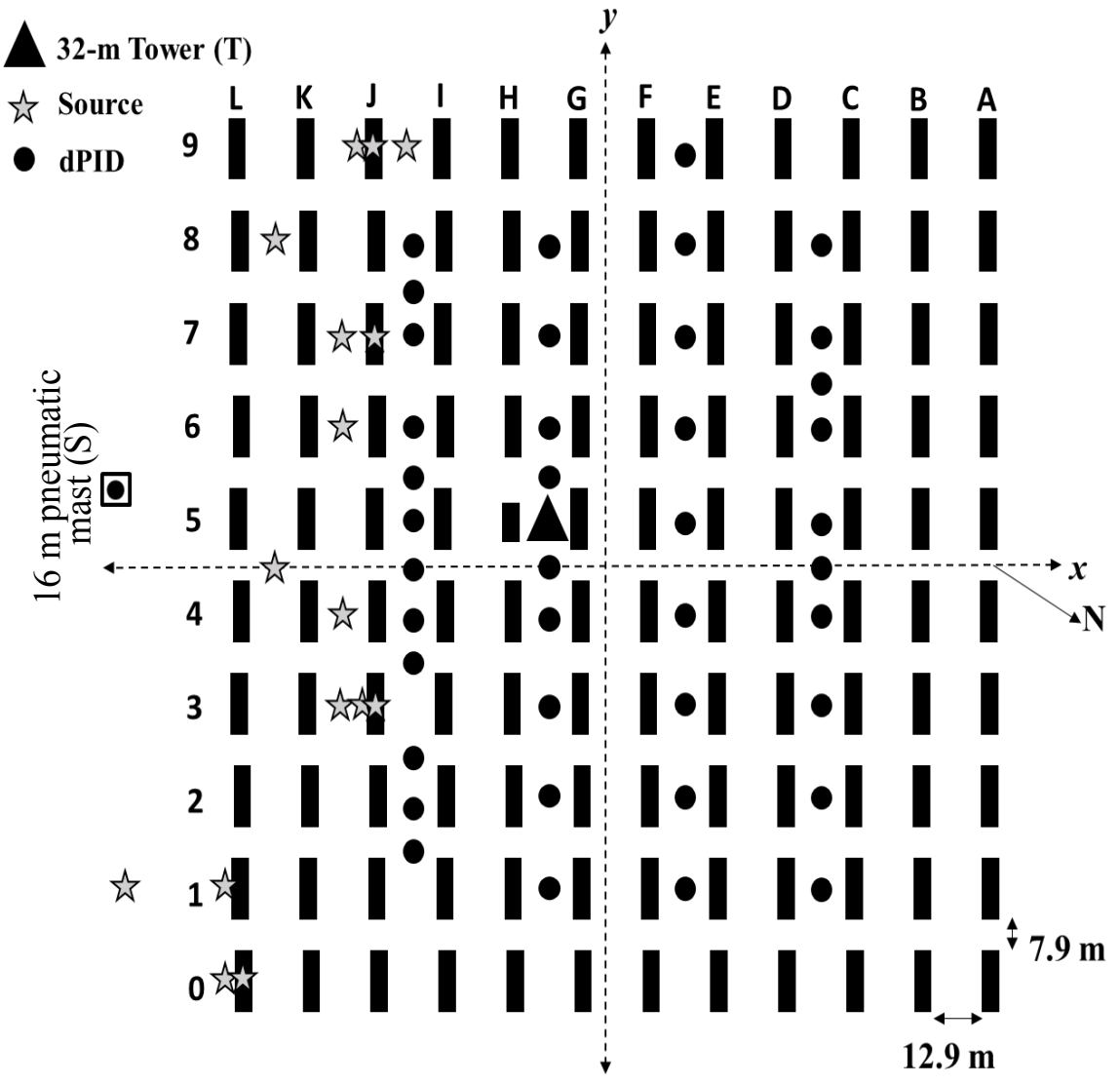

Figure 1. A schematic diagram of the MUST geometry showing 120 containers and source (stars) and receptors (black filled circles) locations. In a given trial - only one source was operational.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

Geoscientific Model Development Discussions

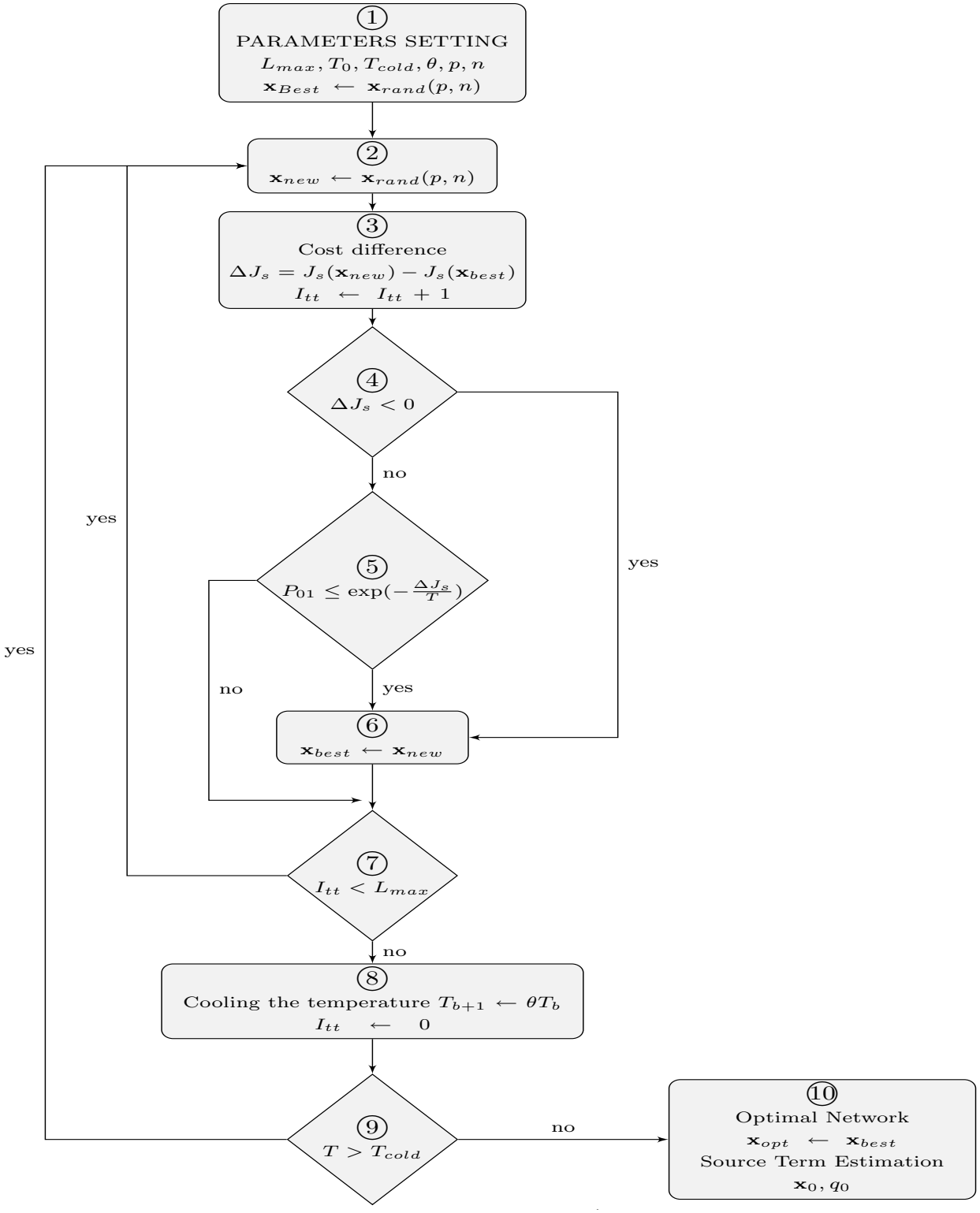

Figure 2. Flow diagram of the Simulated Annealing procedures to determine an optimized monitoring network.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

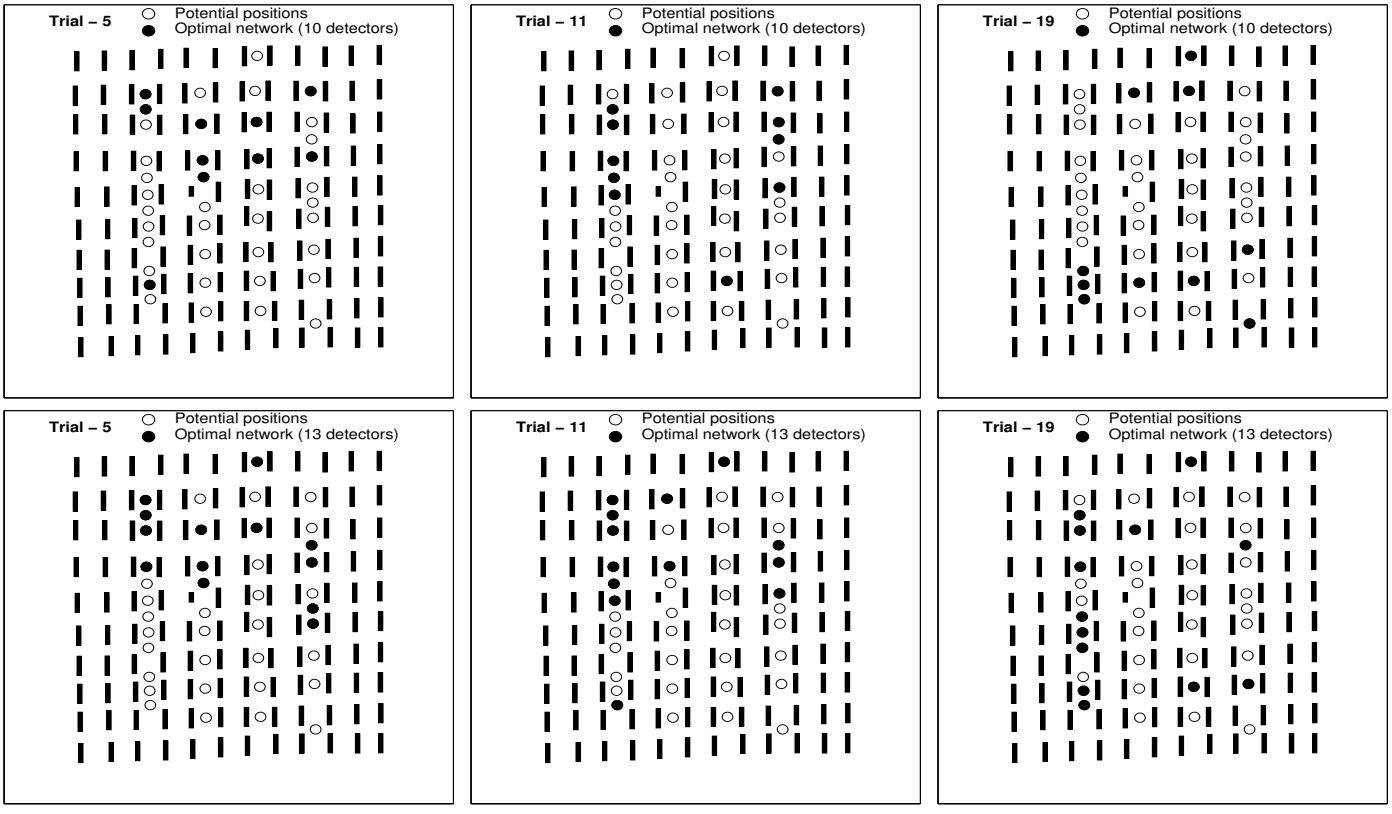

Figure 3. The optimal networks of 10 (first row) and 13 sensors (second row) respectively for trials 5 (very stable), 11 (neutral), and 19 (stable). Blank and filled black circles respectively represent the all (40) potential positions and the optimal positions of sensors.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-6
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 9 March 2018

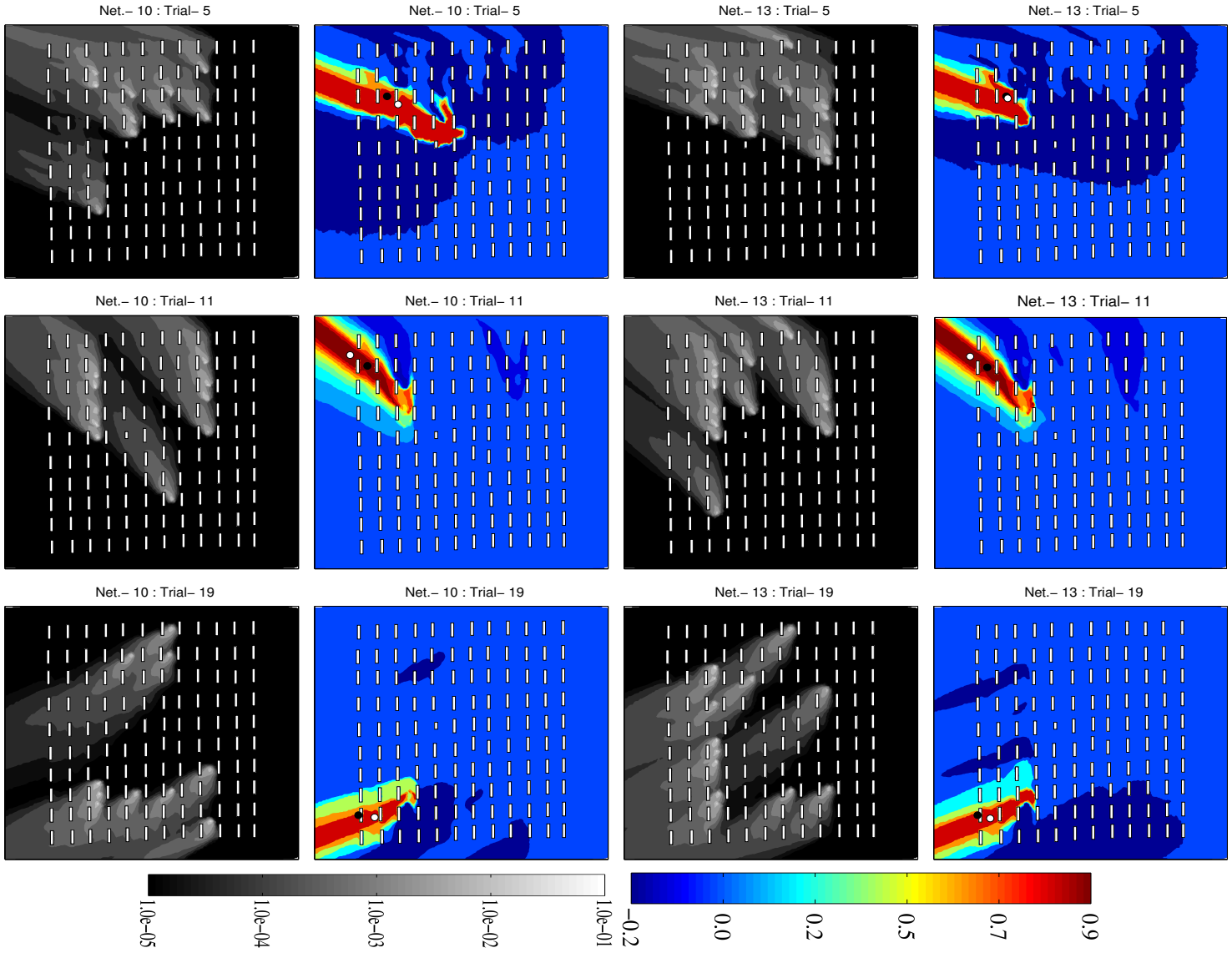

Figure 4. Isopleths of the renormalized weight function $(w(\mathbf{x}))$ (gray colored in first and third columns) and the normalized source estimate function $(\mathbf{s}_w^n(\mathbf{x}) = \mathbf{s}_w(\mathbf{x})/\max(\mathbf{s}_w(\mathbf{x})))$ (colored in second and fourth columns) for both optimal networks of 10 and 13 sensors respectively for trials 5 (very stable), 11 (neutral), and 19 (stable). The black and white filled circles respectively represent the true and estimated source locations.