https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

Geoscientific Model Development Discussions

Generalized spatiotemporally-decoupled framework for reconstructing the source of non-constant atmospheric radionuclide releases

Yuhan Xu[1], Sheng Fang [1,*], Xinwen Dong[1], Shuhan Zhuang[1]

[1]Institute of Nuclear and New Energy Technology, Collaborative Innovation Centre of Advanced Nuclear Energy Technology, Key Laboratory of Advanced Reactor Engineering and Safety of Ministry of Education, Tsinghua University, Beijing 100084, China

**Correspondence to*: Sheng Fang (fangsheng@tsinghua.edu.cn)

Abstract. Determining the source location and release rate are critical in assessing the environmental consequences of atmospheric radionuclide releases, but remain challenging because of the huge multi-dimensional solution space. We propose a generalized spatiotemporally-decoupled two-step framework to reduce the dimension of the solution space in each step and improves the reconstruction accuracy, which is applicable to non-constant releases. The decoupling process is conducted by applying a temporal sliding-window average filter to the observations, thereby reducing the influence of temporal variations in the release rate and ensuring that the features of the filtered data are dominated by the source location. A machine learning model is trained to link these features to the source location, enabling independent source localization. Then the release rate is determined using projected alternating minimization with the L1-norm and total variation regularization algorithm. Validation using SCK-CEN ^{41}Ar experimental data demonstrates that the localization error is less than 1%, and the temporal variations, peak release rate, and total release are reconstructed accurately. The proposed method exhibits higher accuracy and a smaller uncertainty range than the correlation-based and Bayesian methods. Furthermore, it achieves stable performance with different hyperparameters and produces low error levels even with only a single observation site.

1. Introduction

Atmospheric radionuclide release is a major environmental concern of nuclear industry, including nuclear energy and its heat applications, isotope production and post-processing of radioactive waste. These releases have been observed in the Chernobyl nuclear accident (Anspaugh et al., 1988), and Fukushima nuclear explosion (Katata et al., 2012) with partially-known source information, i.e. the location. Recently, there have been several atmospheric radionuclide leaks from unknown sources, such as the 2017 ^{106}Ru leakage (Masson et al., 2019) and 2020 $^{134/137}Cs$ detection in northern Europe (Ingremeau and Saunier, 2022), which have raised global concerns regarding the subsequent hazard to public health. Identification of source information in these events is critical for safe operation of nuclear facilities, consequence assessment, and emergency response.

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

Geoscientific Model Development
Discussions
Open Access
EGU

During these events, source information could not be directly measured or determined because of the lack of information on the source of the leak. Instead, source information could only be reconstructed through inversion methods that identified the optimal solution by comparing the environmental observations with atmospheric dispersion simulations using different estimates of the source location and release rate. Such reconstruction simultaneously identifies the source location and release rate because the observations are intuitively determined by both parameters. In this case, the reconstruction searches for the solution over a tremendous multi-dimensional space, where the dimension is the sum of the number of space coordinates and the length of the estimated release window. Therefore, the inversion is weakly constrained and can become ill-posed in the case of spatiotemporally limited observations and uncertainties in the atmospheric dispersion models. Unfortunately, this is quite often the case for atmospheric radionuclide releases.

To reduce the problem of ill-posedness, most previous studies have attempted to constrain the reconstruction by imposing assumptions on a certain feature of the solution, including both statistical and deterministic assumptions. Statistical assumptions are widely used in Bayesian methods to simultaneously reconstruct the posterior distributions of spatiotemporal source parameters (De Meutter et al., 2021; Meutter and Hoffman, 2020; Xue et al., 2017a). This assumes that the model–observation discrepancies follow a certain statistical distribution (i.e. the likelihood of Bayesian methods), with the normal (Eslinger and Schrom, 2016; Guo et al., 2009; Keats et al., 2007, 2010; Rajaona et al., 2015; Xue et al., 2017a, b; Yee, 2017; Yee et al., 2008; Zhao et al., 2021) and log-normal (Chow et al., 2008; Dumont Le Brazidec et al., 2020; KIM et al., 2011; Monache et al., 2008; Saunier et al., 2019; Senocak, 2010; Senocak et al., 2008) distributions being two popular choices. Other candidate distributions include Cauchy, log-Cauchy, and T3-10, which have been compared with normal and log-normal distributions in reconstructing the source parameters of the Prairie Grass field experiment (Wang et al., 2017). The results demonstrate that the likelihoods are sensitive to both the dataset and the targeted source parameters. Some studies have constructed the likelihood based on multiple metrics that measure the model–observation discrepancies in an attempt to better constrain the solution (Lucas et al., 2017; Jensen et al., 2019). More sophisticated methods involve the use of different statistical distributions for the likelihoods of non-detections and detections (De Meutter et al., 2021; Meutter and Hoffman, 2020). Recent studies have suggested the use of log-based distributions and tailored parameterization of the covariance matrix as a means of better quantifying the uncertainties in the reconstruction (Dumont Le Brazidec et al., 2021). These Bayesian methods have been applied to real atmospheric radionuclide releases, such as the 2017 106Ru event, and have provided important insights into the source and release process (Dumont Le Brazidec et al., 2020; Saunier et al., 2019; Dumont Le Brazidec et al., 2021; De Meutter et al., 2021). However, these studies have also revealed that the likelihood in Bayesian methods must be exquisitely designed and parameterized to achieve satisfactory spatiotemporal source reconstruction (Dumont Le Brazidec et al., 2021; Wang et al., 2017). With suboptimal design, the reconstruction may exhibit a bimodal posterior distribution (Meutter and Hoffman, 2020), which remains a challenge for robust applications in different scenarios.

Deterministic assumptions mainly involve entropy (Krysta and Bocquet, 2007; Bocquet, 2005b, a) and a constant release rate (Kovalets et al., 2020, 2018; Efthimiou et al., 2018, 2017; Tomas et al., 2021; Andronopoulos and Kovalets, 2021; Ma et al., 2018). Compared with entropy, the constant-release assumption is more popular and is embedded in many Bayesian

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

methods (Yee et al., 2008; Eslinger and Schrom, 2016; Meutter and Hoffman, 2020; Zhao et al., 2021; De Meutter et al., 2021), substantially reducing the dimension of the solution space to 5 or 6 (i.e. the two or three source location coordinates, the start and end time of the release, and the total release). Recently, the constant-release assumption has been found to enable separate reconstruction of the source location and release rate because the relative spatiotemporal distribution of the simulated atmospheric concentrations is independent of the release rate and is determined by the source location and meteorology (Efthimiou et al., 2017; Kovalets et al., 2018). Because the meteorology can be independently calculated, it is possible to efficiently identify the source location without knowing the specific release rate by simply comparing the correlation coefficients between the observations and simulations using different source locations. The constant release rate is then obtained by calculating a scale factor between the simulation using the identified source location and the observations. This method exhibits high accuracy for constant releases under stationary meteorological conditions, such as synthetic simulation experiments (Ma et al., 2018) and wind tunnel experiments (Kovalets et al., 2018; Efthimiou et al., 2017). However, real releases usually exhibit temporal variations and may experience non-stationary meteorological fields, both of which can lead to noticeable uncertainties in the source localization results of the correlation-based method (Tomas et al., 2021; Andronopoulos and Kovalets, 2021), thus limiting its practical usage. Despite this restriction of the constant-release assumption, the correlation-based method does efficiently reduce the solution space for the source localization and release rate estimation, suggesting the potential for reliable source reconstruction.

In this study, we relax the constant-release assumption and propose a more general spatiotemporally decoupled source reconstruction method for non-constant release scenarios. Our approach uses the simple facts that the source location is fixed during the atmospheric radionuclide release process and the peak amplitude of temporal observations is mainly affected by the release rate of a spatially fixed source (Li et al., 2019b). Based on these facts, the proposed method removes the influence of the release rate on the observations through a temporal sliding-window average filter, which approximates the constant-release case and enables decoupled source localization in the absence of release rate estimation. Because the peak amplitude is reduced, existing methods based on direct observation–simulation comparisons may be unable to localize the source. Thus, the response features of the filtered observations are extracted and mapped to the source location by training a data-driven machine learning model using the extreme gradient boosting (XGBoost) algorithm (Chen and Guestrin, 2016). To fully capture the response features at each observation site, tailored time- and frequency-domain features are designed and optimized using the feature selection technique of XGBoost. Using this optimized model, the source is localized based on the filtered observations. Once the source location has been retrieved, the non-constant release rate is determined using the Projected Alternating MInimization with L1-norm and Total variation regularization (PAMILT) algorithm (Fang et al., 2022), which is robust to model uncertainties. The sequential spatiotemporal reconstruction reduces the dimension of the solution space at each step, which helps to improve the accuracy and reliability of the reconstruction.

The performance of the proposed method is compared with the correlation-based method (constant-release assumption) for source localization and the Bayesian method (statistical assumption) for spatiotemporal accuracy. The sensitivity of the source

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

Geoscientific Model Development Discussions

localization to the spatial search range, size of the sliding window, feature type, and number and combination of sites is also investigated.

2. Material and Methods

2.1 Source reconstruction models

For an atmospheric radionuclide release, Eq. (1) relates the observations at each observation site to the source parameters:

$$\boldsymbol{\mu} = \mathbf{F}(\mathbf{r}, \mathbf{q}) + \boldsymbol{\varepsilon}\,, \tag{1}$$

where $\boldsymbol{\mu} = [\mu_1, \mu_2, \cdots, \mu_N] \in \mathbb{R}^N$ is an observation vector composed of observations at N sequential time steps, the function $\mathbf{F}$ maps the source parameters to the observations, i.e. an atmospheric dispersion model, $\mathbf{r}$ refers to the source location, $\mathbf{q} \in \mathbb{R}^N$ is the temporally varying release rate, and $\boldsymbol{\varepsilon} \in \mathbb{R}^N$ is a vector containing both model and measurement errors.

In most source reconstruction models, $\mathbf{F}$ is simplified to the product of $\mathbf{q}$ and a source–receptor matrix $\mathbf{A}$ that depends on the source location:

$$\boldsymbol{\mu} = \mathbf{A}(\mathbf{r})\mathbf{q} + \boldsymbol{\varepsilon}\,, \tag{2}$$

where $\mathbf{A}(\mathbf{r}) = [A_1(\mathbf{r}),\ A_2(\mathbf{r}), \cdots, A_N(\mathbf{r})]^T \in \mathbb{R}^{N\times N}$ and each row describes the sensitivity of an observation to the release rate $\mathbf{q}$ given the source location $\mathbf{r}$.

2.2 Spatiotemporal decoupling

A straightforward way to solve Eq. (2) is to simultaneously retrieve the source location and release rate, which the solution space is huge and difficult to constrain. Several studies have pointed out that the source location can be retrieved separately without knowledge of the exact release rate, on the condition that the release rate is constant (Efthimiou et al., 2018; Kovalets et al., 2018; Efthimiou et al., 2017; Ma et al., 2018). The key reason is that, in constant-release cases, the relative spatiotemporal distribution of radionuclides is determined by the meteorological conditions and the relative positions between the source and receptors, and the constant release rate only changes the absolute values. Although the release rate may counteract the influence of the meteorological conditions and relative position at a single observation site, it cannot change the whole spatiotemporal distribution at multiple observation sites. Therefore, by analysing the spatiotemporal distribution of radionuclides at multiple observation sites, it is possible to localize the source without knowing the release rate under the constant-release assumption.

To provide a more general method, we take advantage of the fact that the source location has been fixed during all known atmospheric radionuclide releases, such as the Chernobyl nuclear accident (Anspaugh et al., 1988), the Fukushima nuclear explosion (Katata et al., 2012), and the 2017 ^{106}Ru leakage (Masson et al., 2019). With a fixed source location, the release rate mainly influences the peak values of the temporal observations at each site, whereas the meteorology determines the timing of the peaks (Li et al., 2019b). Therefore, the constant-release case can be approximated by reducing the influence of the release

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

rate, enabling separate source localization and release rate estimation in a more general case and reducing the solution space at each step. For this purpose, we introduce an operator matrix $\mathbf{P} \in \mathbb{R}^{N\times N}$ to reduce the temporal variations of $\mathbf{A}(\mathbf{r})\mathbf{q}$:

$$\boldsymbol{\mu}_p = \mathbf{P}\boldsymbol{\mu} = \mathbf{P}\mathbf{A}(\mathbf{r})\mathbf{q} + \mathbf{P}\boldsymbol{\varepsilon}\,, \tag{3}$$

where $\boldsymbol{\mu}_p$ refers to the decoupled observations. In this study, the following operator matrix is constructed to impose a temporal sliding-window average filter (Eamonn Keogh, Selina Chu, 2004):

$$\mathbf{P} = \frac{1}{T}\begin{bmatrix} 1 & & & & & & & & & \\ 1 & 1 & & & & & & & & \\ & \vdots & & & & & & & & \\ 1 & 1 & \cdots & 1 & & & & & & \\ 1 & 1 & \cdots & 1 & 1 & & & & & \\ & 1 & 1 & \cdots & 1 & 1 & & & & \\ & & 1 & 1 & \cdots & 1 & 1 & & & \\ & & & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & & & 1 & 1 & \cdots & 1 & 1 & \\ & & & & & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \tag{4}$$

where T is the size of the sliding window. Although a sliding-window average filter is used in this study, Eq. (3) is compatible with more advanced processing methods, thereby providing a general framework for the spatiotemporal decoupling of $\boldsymbol{\mu}$.

2.3 Source localization without knowing the exact release rates

After applying the filter in Eq. (4), the local peak amplitude of the observations is smoothed out, but the influences of the source position and meteorology remain relatively unchanged, as they determine the long-term temporal trends of observations and are less affected by the filter. The meteorology is known, so it becomes possible to localize the source using the filtered observations. Nevertheless, the specificity of those localization methods that rely on direct observation–simulation comparisons may be substantially compromised because the peak amplitude is reduced. A better choice for source localization would be to use the response features of the filtered observations, which preserve most of the location information. Therefore, it is necessary to establish a link between the response features of the filtered observations and the source location. To achieve this, we train an XGBoost model that maps the response features of the filtered observations to the coordinates of the source.

XGBoost is an optimized distributed gradient boosting library. Suppose $D = \{(\mathbf{X}_i, \mathbf{r}_i)\}(|D| = n, \mathbf{X}_i \in \mathbb{R}^p, \mathbf{r}_i \in \mathbb{R}^2)$, where the number of samples is n and each sample contains p features. $\mathbf{X}_i$ is the given input feature vector of the i-th sample and $\mathbf{r}_i = (x_i, y_i)$ is the location vector. XGBoost typically uses multiple decision trees (Fig. 1) to fit the target, which can be formulated as:

$$\hat{\mathbf{r}}_i = G(\mathbf{X}) = \sum_{k=1}^{K} f_k(\mathbf{X}_i)\,, f_k \in \boldsymbol{\mathcal{F}}\,, \tag{5}$$

where K is the number of trees, $\boldsymbol{\mathcal{F}} = \{f(x) = \boldsymbol{\omega}_{Q(x)}\}(Q: \mathbb{R}^p \to M, \boldsymbol{\omega} \in \mathbb{R}^M)$ is the space of decision trees, and Q represents the structure of each tree, mapping the feature vector to M leaf nodes. Each f_k corresponds to an independent tree structure Q

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

Geoscientific Model Development Discussions
Open Access
EGU

with leaf node weight $\boldsymbol{\omega}$. Equation (5) is then used to predict $\hat{\mathbf{r}}_i = (\hat{x}_i, \hat{y}_i)$ for the i-th sample.

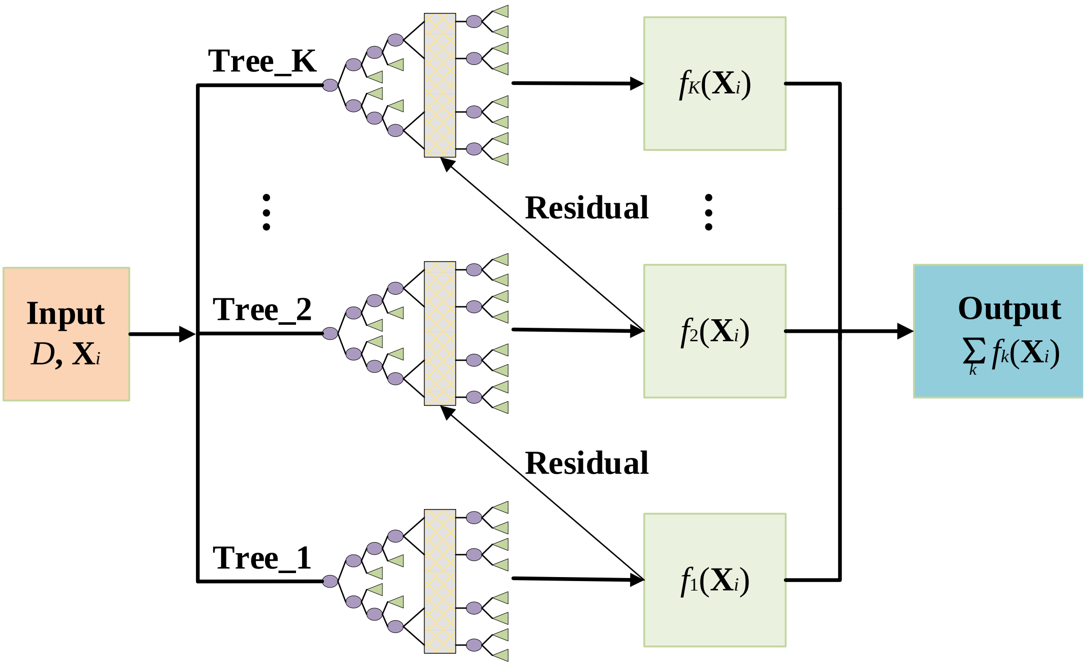


Figure 1. Flowchart of XGBoost for predicting $\hat{\mathbf{r}}_i$ based on decision tree model.

XGBoost trains $G(\mathbf{X})$ in Eq. (5) by continuously fitting the residual error until the following objective function is minimized:

$$Obj^{(t)} = \sum_{i=1}^{n} \left(\mathbf{r}_i - \left(\hat{\mathbf{r}}_i^{(t-1)} + f_t(\mathbf{X}_i) \right) \right)^2 + \sum_{i=1}^{t} \Omega(f_i) \,, \tag{6}$$

where t represents the training of the t-th tree and $\Omega(f_i)$ is the regularization term, given by:

$$\Omega(f) = \Upsilon T + \frac{1}{2} \lambda \sum_{j=1}^{T} \omega_j^2 \,, \tag{7}$$

The minimization of Eq. (6) provides the parametric model $G(\mathbf{X})$ that maps the feature ensemble $\mathbf{X}$ extracted from $\boldsymbol{\mu}_p$ to the source location $\mathbf{r}$.

To comprehensively evaluate the influence of the source location, both time- and frequency-domain features (as outlined in Table 1) are considered during the training process and mapped to the source location by $G(\mathbf{X})$. Among the time-domain features, the wave rate quantifies the amplitude of fluctuations in $\boldsymbol{\mu}_p$, while the temporal mean and median values represent the central moment of $\boldsymbol{\mu}_p$. Additionally, the sample entropy measures the complexity of $\boldsymbol{\mu}_p$, with a lower sample entropy indicating greater self-similarity and less randomness in $\boldsymbol{\mu}_p$. The frequency-domain features are calculated based on the fast Fourier transform (FFT). The FFT mean is the mean value of the Fourier coefficients for $\boldsymbol{\mu}_p$ and the FFT shape mean describes the shape of the Fourier coefficients. These quantities are formulated as follows:

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

$$\text{FFT mean} = \frac{1}{N}\sum_{k=1}^{N}|\mu_{ik}| \ , \tag{8}$$

$$\text{FFT shape mean} = \frac{1}{\sum_{k=1}^{N}|\mu_{ik}|}\sum_{k=1}^{N} k|\mu_{ik}| \ , \tag{9}$$

where μ_{ik} is the Fourier coefficient and N is the length of $\boldsymbol{\mu}_p$. These features are calculated from the simulated observations at each site and provided to XGBoost as initial inputs.

Table 1. Summary of the basic information on the observation series features.

Attribute	Feature	Description
Time domain	Wave rate	Difference between 90-th and 10-th quantile of normalized observation series
	Mean	Temporal mean value of observation series
	Median	Temporal median value of observation series
	Sample entropy	Complexity of observation series
Frequency domain	FFT mean	Amplitude of power spectral density by FFT
	FFT shape mean	Shape of power spectral density by FFT

2.4 Release rate estimation

Once the source location has been retrieved, many existing methods can be used to inversely estimate the release rate. In this study, we choose the recent PAMILT method (Fang et al., 2022) because it can correct the intrinsic model errors of the release rate estimation and reduce the propagation of localization errors into the release rate estimate.

2.5 Numerical implementation

2.5.1 Pre-screening of potential source locations

To reduce the computational cost and remove low-quality samples, the search range for the source location is pre-screened by evaluating the correlation coefficients between the observations and atmospheric dispersion model simulations, where the source locations are randomly sampled in the considered calculation domain. Because the release rate is unknown, it is assumed to be 1 Bq/h for all simulations. Those source locations corresponding to correlation coefficients above the 40th percentile are selected as the search range of the subsequent refined source localization using XGBoost.

2.5.2 Samples for training XGBoost

The samples for training $G(\mathbf{X})$ in Eq. (5) are generated based on the simulations described in Sect. 2.5.1, and the source

Geoscientific Model Development Discussions
Open Access EGU

locations of these simulations are within the search range as determined according to Sect. 2.5.1. The simulation data are scaled by a constant factor (the ratio between the median value of all observations and that of the simulations using a unit release rate), which ensures that the simulations and observations have the same order of magnitude. Gaussian noise is added to the simulation data to simulate the statistical fluctuations of radiation measurements. Those simulations between the first and last data points above the noise level are filtered by a temporal sliding-window average filter with a window size of 5, yielding samples for feature extraction as described in Sect. 2.3.

2.5.3 Automatic optimization of XGBoost model

The XGBoost model for source localization is automatically optimized with respect to the hyperparameters and feature selection. Specifically, the Bayesian optimization algorithm is used to optimize the hyperparameters by minimizing the following generalization coefficient (GC) defined under the five-fold cross-validation framework:

$$\mathrm{GC} = (1 - \mathrm{MCV})^2 + Var(R_k^2) \,, \tag{10}$$

$$\mathrm{MCV} = \frac{1}{5}\sum_k R_k^2 \,, \tag{11}$$

where R_k^2 is the goodness of fit and k is the index of each fold ($k = 1, 2, \ldots, 5$). MCV is the mean cross-validation score R_k^2 among the five folds and $Var(R_k^2)$ measures the variance of R_k^2. This function aims to balance the average and the variance of R_k^2, thus enhancing the generalization ability of the XGBoost model. In this study, the optimized hyperparameters include *max_depth* (maximum depth of a tree), *learning_rate* (step size shrinkage when updating), *n_estimators* (number of decision trees), *min_child_weight* (minimum sum of sample weight of a child node), *subsample* (subsample ratio of the training samples), *colsample_bytree* (subsample ratio of columns when constructing a tree), *reg_lambda* (L2 regularization term on weights), and *gamma* (minimum loss reduction required to split the tree).

The initial input features (Table 1) are optimized by recursive feature elimination with cross-validation (Akhtar et al., 2019), which sorts the features in order of importance and removes the least-important features based on the MCV results. The overall flowchart of the proposed spatiotemporally decoupled source reconstruction model is shown in Fig. S1.

2.6 Validation case

2.6.1 SCK-CEN ^{41}Ar field experiment

The proposed methodology was validated against the observations of the SCK-CEN ^{41}Ar field experiment, which was carried out at the BR1 research reactor in Mol, Belgium, in October 2001 as a collaboration between NKS and the Belgian Nuclear Research Centre (SCK-CEN) (Rojas-Palma et al., 2004). The major part of the experiment was conducted on 3–4 October, during which time ^{41}Ar was emitted from a 60-m stack with a release rate of approximately 1.5×10^{11} Bq/h. Meteorological data such as wind speed and direction were provided by the on-site weather mast. For most of the experimental period, the atmospheric stability was neutral, and the wind was blowing from the southwest. As illustrated in Fig. 2, the source coordinates

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

were (650 m, 210 m).

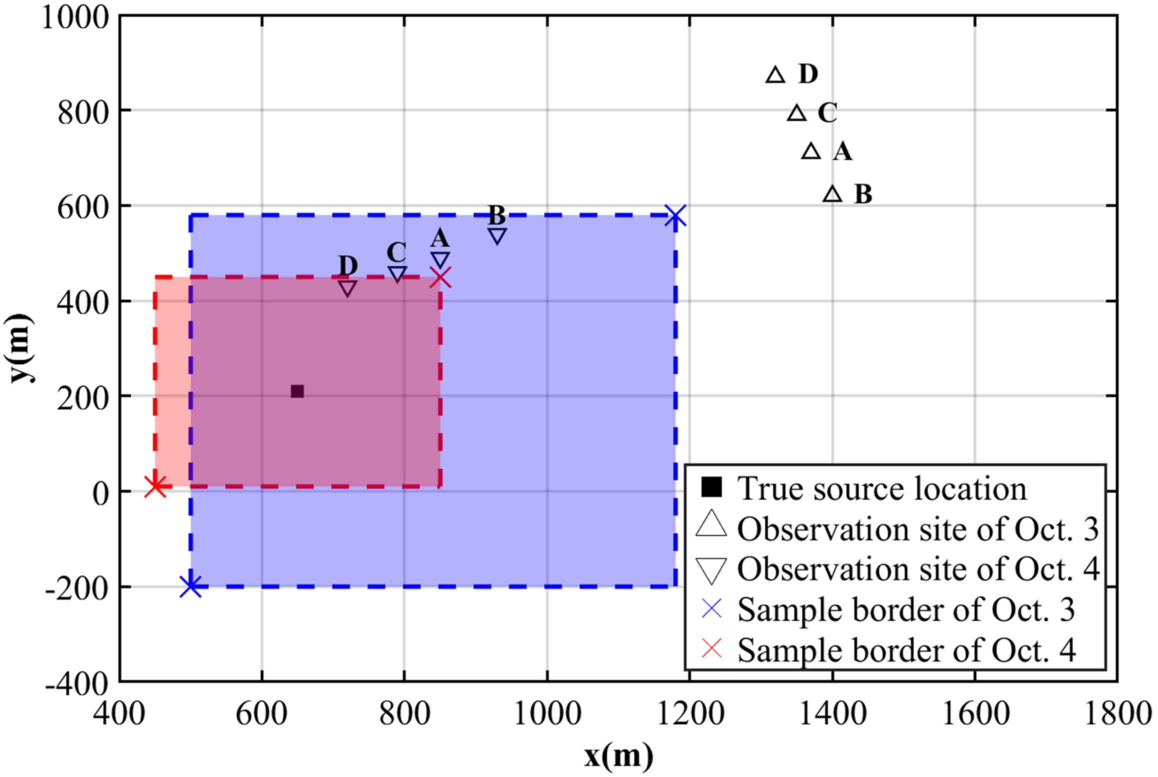


Figure 2. Release location and observation sites of SCK-CEN ^{41}Ar experiment. The map was created based on the relative positions of the release source and observation sites, as detailed in (Drews et al., 2002). It was plotted using MATLAB 2016b, instead of created by a map provider.

The 60-s-average fluence rates were continuously collected by an array of NaI (Tl) gamma detectors, with different observation sites used on the two days. To convert the measured fluence rates to gamma dose rates (mSv/h), we used the ^{41}Ar parameters of a previous study (Li et al., 2019a): $E_\gamma = 1.2938$ MeV, $f^n(E_\gamma) = 0.9921$, $\mu_a = 2.05 \times 10^{-3}\ \mathrm{m}^{-1}$, and $\omega = 7.3516 \times 10^{-1}\ \mathrm{Sv/Gy}$. More details of these measurements can be found in reference (Rojas-Palma et al., 2004).

2.6.2 Simulation settings of atmospheric dispersion model

The Risø Mesoscale PUFF (RIMPUFF) model was employed to simulate the dispersion of radionuclides and to calculate the dose rates at each observation site (Thykier-Nielsen et al., 1999). The simulations used on-site measured meteorological data and the modified Karlsruhe–Jülich diffusion coefficients. The calculation domain measured 1800 m×1800 m and the grid resolution was 10 m×10 m. Other RIMPUFF calculation settings followed those of a previous study (Li et al., 2019a), and have been validated against the observations.

To establish the datasets for the XGBoost model, 2000 samples and 1000 samples with different source locations were

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

Geoscientific Model Development
Discussions
Open Access

calculated by RIMPUFF for Oct. 3 and Oct. 4 respectively. The source locations were sampled from the shaded zones in Fig. 2, which were determined according to the positions of the observation sites and the upwind direction. As described in Sect. 2.5.1, we calculated the correlation coefficient for each sample and preserved samples with correlation coefficients greater than the 40th percentile of all correlation coefficients (i.e. 800 samples for Oct. 3 and 400 samples for Oct. 4). The constant factors mentioned in Sect. 2.5.2 are 1.53×10^{11} and 1.48×10^{11} for Oct. 3 and Oct. 4, respectively.

2.7 Sensitivity study

(1) Search range

The search range is controlled by the pre-screening threshold. The source localization is implemented with pre-screening thresholds determined by the 20th, 40th, 50th, 60th, 80th, and 100th percentiles of the correlation coefficients in the pre-screening step, where a lower percentile corresponds to a more refined search range.

(2) Size of the sliding window

Temporal filtering with different sliding-window sizes is applied to decouple the source localization from the release rate estimation. In this study, the size of the sliding window ranges from 3 to 10. With these decoupled data, the XGBoost model is trained using the same pattern for the source localization.

(3) Feature type

The XGBoost model is trained using only time-domain features and only frequency-domain features, respectively, to investigate the influence of these features on the source localization. The performance of the time-feature-only and frequency-feature-only models is compared with the all-features result.

(4) Number and combination of observation sites

The XGBoost model is trained and applied to the source localization with different numbers of observation sites, namely a single site, two sites, and three sites. For the two- and three-site cases, the model is trained using different combinations of sites and the source location is estimated accordingly.

In all the sensitivity tests, the source location is estimated 50 times with randomly initialized hyperparameters to demonstrate the uncertainty range of the proposed method under different circumstances. The performance of source localization is compared quantitatively using the metrics specified in Sect. 2.8.3.

2.8 Performance evaluation

2.8.1 Decoupling

The feasibility of decoupling was demonstrated using both the synthetic and real observations of the SCK-CEN ^{41}Ar field experiment. The former were generated by a simulation using a synthetic temporally varying release profile with sharp increase, stable, and gradual decrease phases (as illustrated in Fig. S2), which is typical for an atmospheric radionuclide release (Davoine and Bocquet, 2007). The simulations corresponding to the synthetic and real observations should first be processed following

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

Geoscientific Model Development Discussions

the procedure in Sect. 2.5.2. The decoupling performance is evaluated by comparing the simulation–observation differences before and after the decoupling step. Several statistical metrics can be used to quantify this difference, including the normalized mean square error (NMSE), Pearson's correlation coefficient (PCC), and the fraction of predictions within a factor of 2 and 5 of the observations (FAC 2 and FAC 5, respectively) (Chang and Hanna, 2004).

2.8.2 Optimization of the XGBoost model

The hyperparameters are optimized with respect to the GC in Eq. (10) and the features are optimized with respect to the MCV in Eq. (11). Larger values of MCV and smaller values of GC indicate better optimization performance. In addition, the importance of each feature to the XGBoost training is evaluated with the built-in *feature importance* measure of the XGBoost model.

2.8.3 Source reconstruction

The relative errors of source localization ($\delta_{\mathbf{r}}$) and total release (δ_Q) are calculated to evaluate the source reconstruction accuracy:

$$\delta_{\mathbf{r}} = \frac{|\mathbf{r}_{true} - \mathbf{r}_{est}|}{L_D} \times 100\% , \tag{12}$$

$$\delta_Q = \frac{|Q_{true} - Q_{est}|}{Q_{true}} \times 100\% , \tag{13}$$

where $\mathbf{r}_{true}$ and Q_{true} refer to the real source location and total release of the SCK-CEN ^{41}Ar field experiment and $\mathbf{r}_{est}$ and Q_{est} are the estimated location and total release, respectively. L_D represents the range of the source domain, which is the distance between the lower and upper borders of the sampled zone, equal to 1034.8 m and 565.7 m on Oct. 3 and Oct. 4, respectively. In addition to the total release, the reconstructed release rates are also compared with the true value in terms of the temporal profile.

2.8.4 Comparison with the Bayesian method

The proposed method is compared with the popular Bayesian method based on the SCK-CEN ^{41}Ar experiment, with the same search range used for source localization in both methods (Fig. 2). The Bayesian method is augmented with an in-loop inversion of the release rate at each iteration step of the Markov chain Monte Carlo sampling. The prior distribution of the Bayesian method is a uniform distribution and the likelihood is a log-Cauchy distribution. More detailed information is presented in Supplementary Note S1.

2.8.5 Uncertainty range

The uncertainty ranges are calculated and compared for the correlation-based method, the Bayesian method, and the proposed method. For the correlation-based method, the uncertainty range is calculated using the source locations with the top-50

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

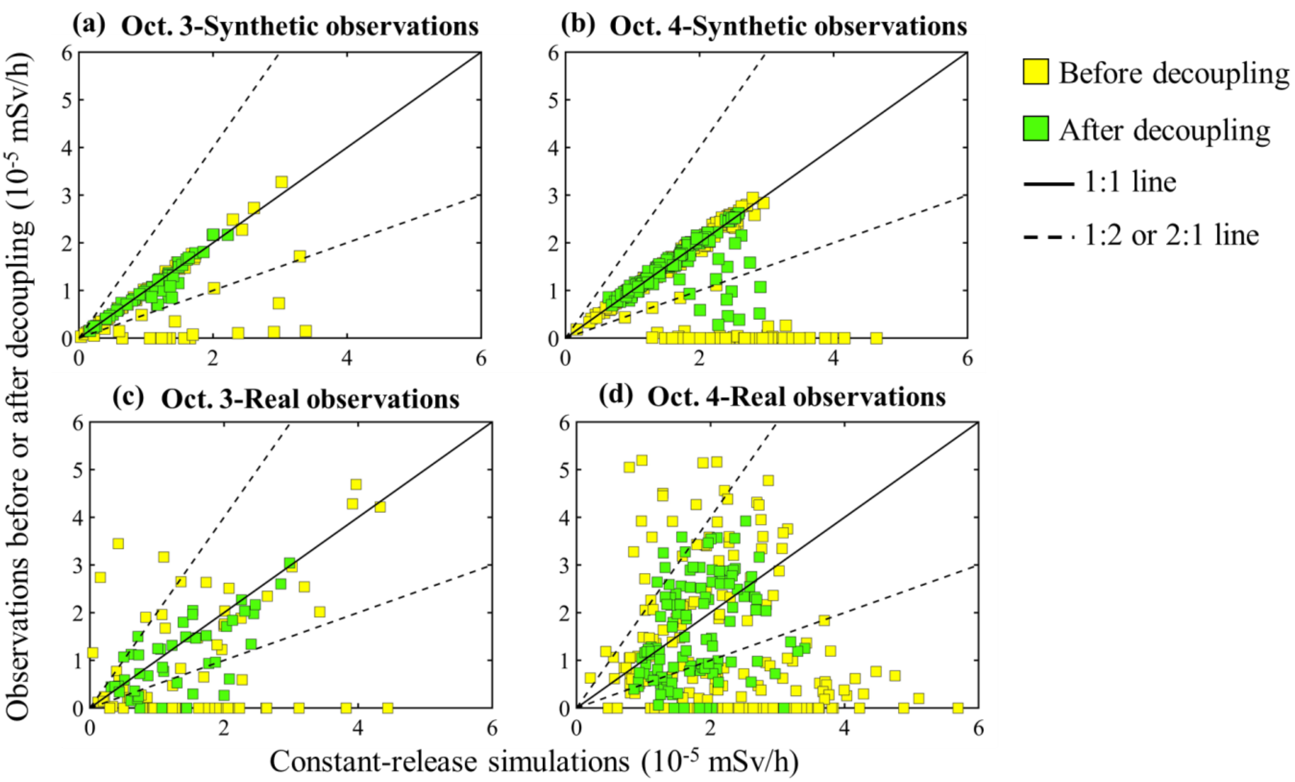
correlation coefficients. For the proposed method, the uncertainty range is calculated from 50 Monte Carlo runs with randomly initialized hyperparameters. The Bayesian method itself provides the uncertainty range through the posterior distribution. To be consistent with the other two methods, the results with the top-50 frequencies are selected for the comparison.

3. Results and Discussion

3.1 Decoupling performance

Figure S3 displays the original and filtered observations at different observation sites for both days. The results demonstrate that the peak values have been smoothed out and the long-term trends are preserved to a large degree.

Figure 3 compares the decoupling performance for both the synthetic and real observations, by plotting the constant-release simulations against the observations before and after decoupling. For the synthetic observations, the decoupled data are more concentrated along the 1:1 line for both days, and all the decoupled data fall within the 2-fold lines for Oct. 3. For real observations, the dots before decoupling in Fig. 3 have a dispersed distribution for both Oct. 3 and Oct. 4, indicating limited correlations with the simulations. After decoupling, the dots are more concentrated along the 1:1 and 1:2 (2:1) lines. These phenomena indicate a noticeably increased agreement between the decoupled observations and the constant-release simulations.



https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

Geoscientific Model Development
Discussions

Figure 3. Scatter plots of the original (yellow squares) and decoupled (green squares) observations versus the constant-release simulation results. (a) Oct. 3-Synthetic observations; (b) Oct. 4-Synthetic observations; (c) Oct. 3-Real observations; (d) Oct. 4-Real observations.

Table 2 quantitatively compares the agreements presented in Fig. 3. For both the synthetic and real observations, all metrics are greatly improved after decoupling, especially NMSE and PCC, confirming the better agreement between the decoupled observations and the constant-release simulations. The decoupling performs better with the synthetic observations than with the real observations. This is because the synthetic observations are free of measurement errors. The improved agreement indicates that the decoupling step significantly reduces the influence of temporal variations in release rates across the real observations.

Table 2. Quantitative metrics for the decoupling validation.

Experiment			NMSE	PCC	FAC2	FAC5
Synthetic observations	Oct. 3	Before decoupling	0.69704	0.5315	0.7647	0.8235
		After decoupling	0.0239	0.9514	1	1
	Oct. 4	Before decoupling	0.9290	-0.0267	0.7292	0.7292
		After decoupling	0.0956	0.6179	0.9412	0.9779
Real observations	Oct. 3	Before decoupling	1.4437	0.3572	0.3824	0.5147
		After decoupling	0.2730	0.6976	0.7273	0.8864
	Oct. 4	Before decoupling	1.9290	-0.2099	0.3073	0.4948
		After decoupling	0.3668	0.2802	0.6552	0.9310

3.2 Optimization of XGBoost model

3.2.1 Hyperparameters

Table S1 summarizes the optimal hyperparameters and corresponding GCs used for source localization in this study. The optimal GC on Oct. 3 is smaller than that on Oct. 4, indicating better fitting performance. This difference is possibly the result of the larger training dataset for Oct. 3.

3.2.2 Feature selection

Figure S4 shows the variation of MCV with the number of features for the x and y coordinates. The MCV first increases with the number of features, and then decreases slightly after reaching the maximum for both days. The optimal number of features for Oct. 4 is noticeably smaller than for Oct. 3. In addition, the selected features for Oct. 3 involve all four sites, whereas those for Oct. 4 involve three sites for x and two sites for y. The reduced feature and site numbers indicate a high level of redundancy in the observations acquired on Oct. 4. This is because the observation sites are parallel to the downwind direction and provide

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

similar location information in the crosswind direction.

Figure 4 compares the importance of the selected features at each site. For both days, the temporal features are dominant. For Oct. 3, Site B is the most important, possibly because it is farthest away in the crosswind direction. For Oct. 4, the four sites provide redundant feature information, and many features are removed. This is because the distribution of observation sites is almost parallel to the wind direction on this day. According to Fig. S3, the measurements from Site A and B have a high correlation, thus leading to the removal of features from Site A on Oct. 4. In summary, the feature selection process adapts XGBoost to different application scenarios.

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

Geoscientific Model Development
Discussions
Open Access
EGU

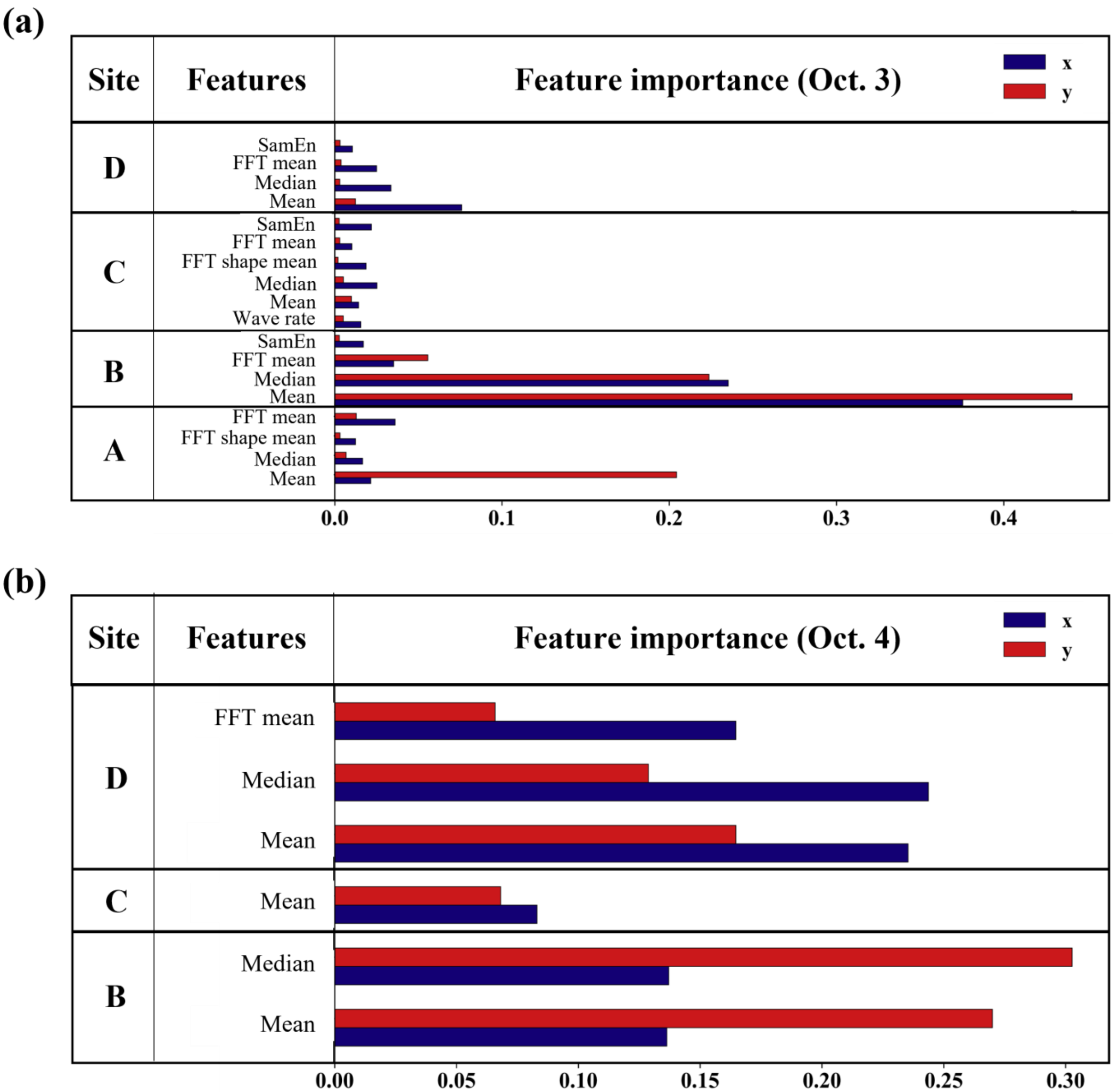


Figure 4. Feature importance. (a) Oct. 3; (b) Oct. 4.

3.3 Source reconstruction

3.3.1 Localization

Figure 5 compares the best-estimated source locations of the correlation-based method, the Bayesian method, and the proposed

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

method with the ground truth. The pre-screening zone covers the true source location for both days, but the areas with the highest correlation coefficients are still too large for the point source to be accurately located. The locations with the maximum correlation exhibit errors of 270.19 m and 36.06 m for Oct. 3 and Oct. 4, respectively, indicating that the correlation-based method may produce biased results in the case of non-constant releases. The Bayesian method estimates the location with errors of 19.62 m and 52.81 m for Oct. 3 and Oct. 4, respectively. In comparison, the proposed method achieves the best performance among all the methods. The estimates without feature selection are only 10.65 m (Oct. 3) and 20.62 m (Oct. 4) away from the true locations. Feature selection further reduces these errors to 6.19 m (Oct. 3) and 4.52 m (Oct. 4), which are below the grid size (10 m × 10 m) of the ATDM simulation. The proposed method gives a relative error of less than 0.9% for both days, whereas the Bayesian method produces a relative error of above 11% and that of the correlation-based method can be as high as 26%. The best estimates for Oct. 3 are more accurate than those for Oct. 4, possibly because of the better layout of observation sites (Fig. 2) and the better decoupling results (Fig. 3).

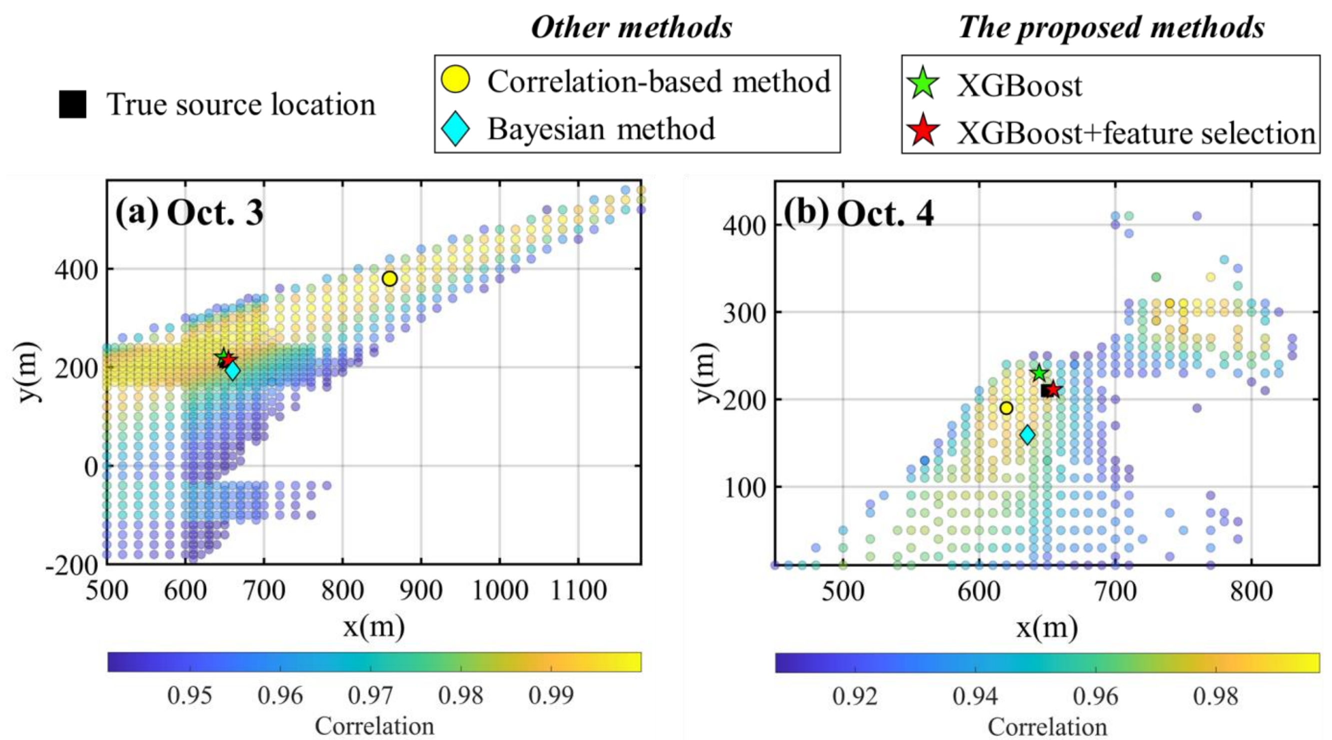


Figure 5. Source localization results. The yellow dots denote the maximum correlation points, which are the localization results of the correlation-based method. The green and red stars represent the localization results based on XGBoost before and after feature selection, respectively. The cyan diamonds represent the localization results based on the Bayesian method. (a) Oct. 3; (b) Oct. 4.

3.3.2 Release rates

Figure 6 displays the release rates estimated by the Bayesian and PAMILT methods based on the source localization results in Fig. 5. The release rates provided by the Bayesian method present several sharp peaks, corresponding to overestimates of up

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

to 269.03% (Oct. 3) and 532.35% (Oct. 4). Furthermore, the Bayesian estimates exhibit unrealistic oscillations in the stable release phase. In contrast, the PAMILT method successfully retrieves the peak releases without oscillations for both days. Both the Bayesian and PAMILT estimates give delayed start times of the release, but accurately estimate the end time, especially for Oct. 3. The PAMILT estimate underestimates the total release by 30.01% and 45.95% for Oct. 3 and Oct. 4, respectively; these values are reduced to about 23.83% and 30.60% after feature selection. The Bayesian method gives better total releases because of the overestimated peaks.

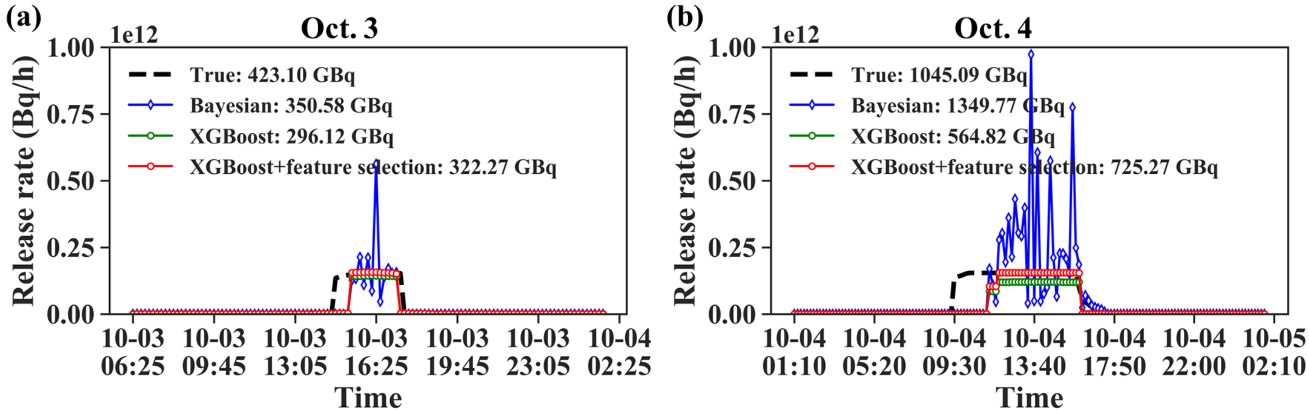


Figure 6. Release rate estimation results with different location estimates. (a) Oct. 3; (b) Oct. 4. The rectangles inside each figure present the location estimates used in the release rate estimations. The green and red stars denote the source locations estimated by XGBoost without and with feature selection, respectively.

3.3.3 Uncertainty range

Figure 7 compares the spatial distribution of 50 estimates produced by the different source localization methods. The estimates of the correlation-based method are highly spread for both days, leading to a highly uniform distribution of the x coordinate for Oct. 3 and two separate distributions of both the coordinates for Oct. 4. The Bayesian method produces a multimodal distribution for both days, in which the estimates are more concentrated than those of the correlation-based method. The corresponding full posteriori distributions in Fig. S5 better reveal the multimodal feature of the Bayesian method, with several peaks of similar probabilities in the estimates of both coordinates on Oct. 3 and the y coordinate on Oct. 4. The multimodal feature indicates the difficulty of constraining the solution in simultaneous spatiotemporal reconstruction, as reported in a previous study (Meutter and Hoffman, 2020). In comparison, the proposed method provides the most concentrated source location estimates. The feature selection moves the centre of the distribution closer to the true location and narrows the distribution of the estimates, especially for Oct. 4.

Geoscientific Model Development Discussions

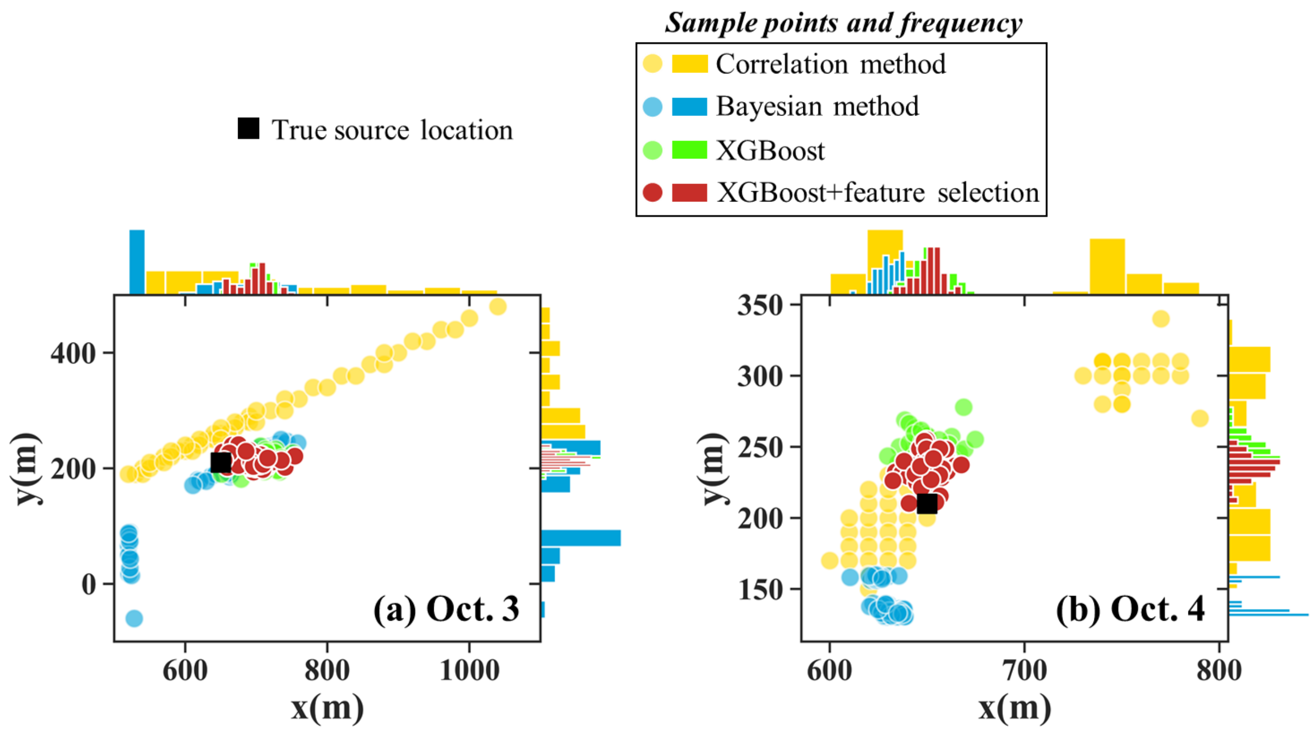


Figure 7. Distribution of source localization results over 50 calculations. (a) Oct. 3; (b) Oct. 4.

Figure 8 compares the uncertainty range and the mean total release of the release rate estimations. For Oct. 3, the Bayesian estimates significantly overestimate the mean values and have a large uncertainty range, whereas the mean PAMILT estimate is very close to the true release and the uncertainty range is smaller than that of the Bayesian method. For Oct. 4, the mean Bayesian estimate exhibits more deviations than the mean PAMILT estimate. Feature selection improves the mean estimate and reduces the uncertainty range of PAMILT because it improves the source localization, thus reducing the deviation in the inverse model of the release rate. On Oct. 3 and Oct. 4, the PAMILT method underestimates the total release by 18.30% and 47.42%, respectively, whereas the Bayesian method gives overestimations of 153.61% and 42.29%, respectively. These results demonstrate that the PAMILT method is robust to localization deviations and can reconstruct the timing, peaks, and total release with relatively high accuracy. This robustness reduces the propagation of localization errors to the release rate estimation, and improves the accuracy of spatiotemporally decoupled source reconstruction.

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

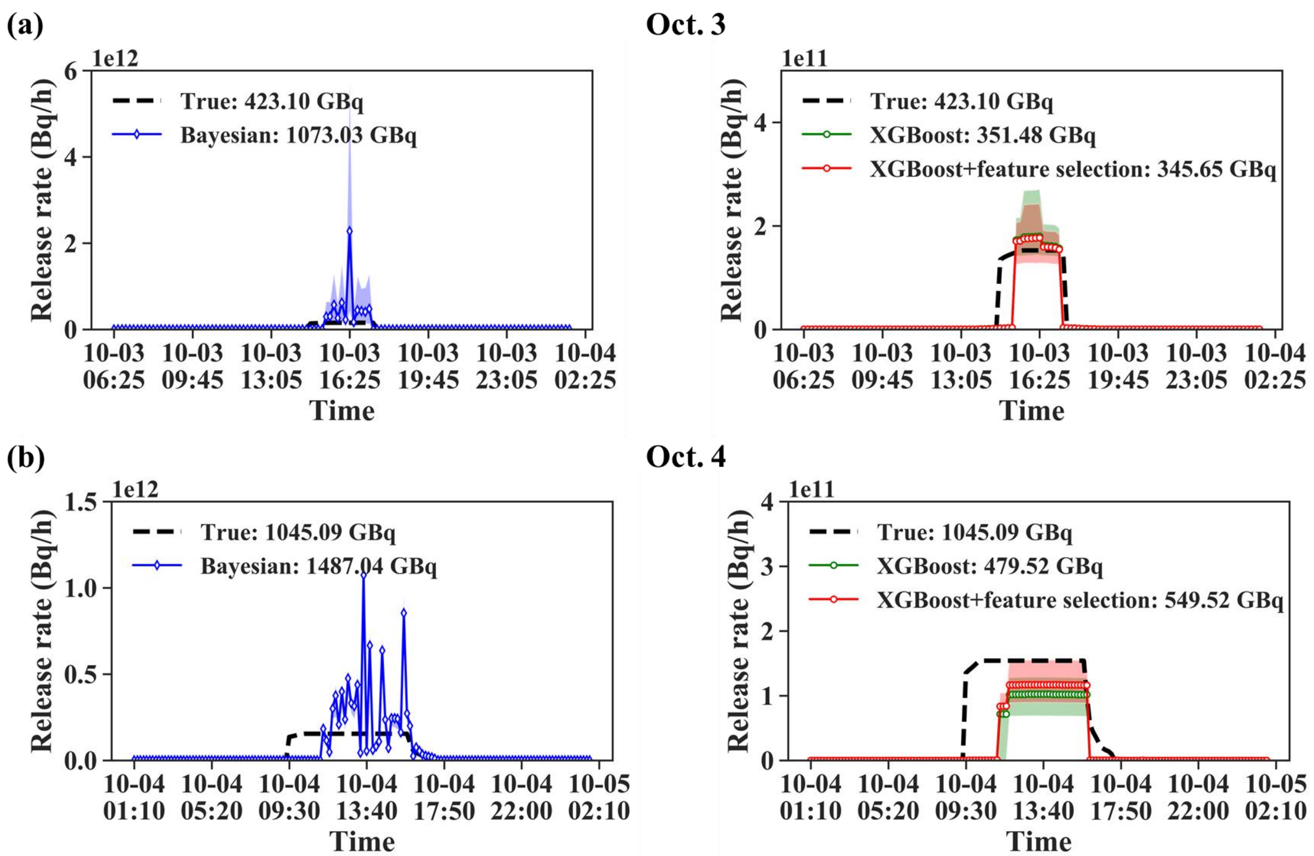


Figure 8. Release rate estimates over 50 calculations. (a) Oct. 3; (b) Oct. 4. The shadow represents the release rate range between the minimum and the maximum.

Table 3 lists the mean and standard deviation of the relative errors for the 50 estimates given by the various methods. In terms of source localization, the correlation-based method produces the largest mean relative error and standard deviation. The proposed method gives the smallest mean error, about half that of the Bayesian method. Its standard deviation is three-fold smaller than that of the Bayesian method for Oct. 3, but slightly larger for Oct. 4. For the total release, the PAMILT method gives a better standard deviation of the relative error for both days and a better mean relative error for Oct. 3, whereas the Bayesian method produces a better mean relative error for Oct. 4. Feature selection reduces the mean relative error, except for the total release for Oct. 3, and slightly increases the standard deviation of the source location and the total release results for Oct. 3.

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

Geoscientific Model Development Discussions Open Access EGU

400 **Table 3.** Relative errors of source reconstruction.

Experiments	Statistical parameters (Relative error)		Correlation-based method	Bayesian method	The proposed method	
					XGBoost	XGBoost+ feature selection
Oct. 3	Source location (δ_r)	Mean	14.10%	11.88%	5.18%	4.68%
		Std	11.37%	7.53%	1.79%	2.05%
	Total release (δ_Q)	Mean	-	177.46%	18.03%	18.45%
		Std	-	167.66%	7.13%	7.68%
Oct. 4	Source location (δ_r)	Mean	14.30%	12.83%	6.83%	4.71%
		Std	9.60%	1.68%	1.76%	1.53%
	Total release (δ_Q)	Mean	-	42.29%	54.12%	47.42%
		Std	-	15.05%	6.47%	5.85%

401 **3.4 Sensitivity analysis results**

402 **3.4.1 Sensitivity to the search range**

403 Figure 9 displays the localization error obtained using different pre-screening thresholds to determine the search range. The
404 error is smaller with a lower threshold, implying that a small pre-screening range helps reduce the mean and median errors. As
405 the threshold increases, the mean and median errors, as well as the error range, show an overall tendency to increase, but not
406 in a strictly monotonic way. The mean/median error is less than 12% for Oct. 3 and less than 22% for Oct. 4, indicating robust
407 performance in these tests. Feature selection reduces the mean/median, range, and lower bound of the errors in most tests,
408 demonstrating its efficiency.

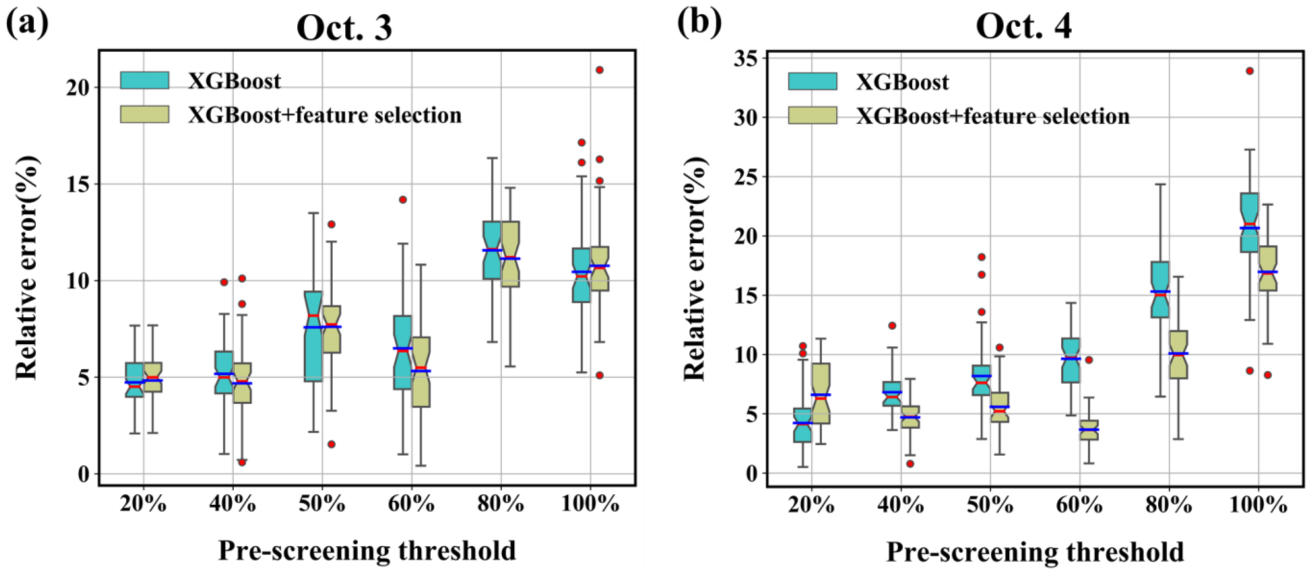


Figure 9. Distribution of relative error (%) over 50 runs with different search ranges. The blue and red solid lines denote average relative error (%) and median relative error (%), respectively. The upper and lower boundaries represent the upper and lower quartiles of relative error (%), respectively. The fences are 1.5 times the inter-quartile ranges of the upper/lower quartiles. The red circles denote data that are not included between the fences. (a) Oct. 3; (b) Oct. 4.

3.4.2 Sensitivity to the size of the sliding window

Figure 10 shows the localization errors obtained with different sliding-window sizes. The mean/median error is less than 8% for Oct. 3 and less than 11% for Oct. 4, both of which are smaller than for the various pre-screening ranges. This indicates that the proposed method is more robust to this parameter than to the pre-screening range. For both days, the lowest mean/median and error range occur with relatively large window sizes, i.e. window size of 9 for Oct. 3 and window size of 10 for Oct. 4. This is because a large window size increases the strength of the decoupling and removes the temporal variations of release rates more completely. However, a large window size leads to increased computational cost. Because the errors vary in a limited range, a medium window size provides a better balance between accuracy and computational cost. Feature selection improves the results for medium and small window sizes, but may have less effect with large window sizes. This tendency implies that it is more appropriate to apply feature selection with medium window sizes than with large window sizes, as is done in this study.

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

Geoscientific Model Development Discussions
Open Access
EGU

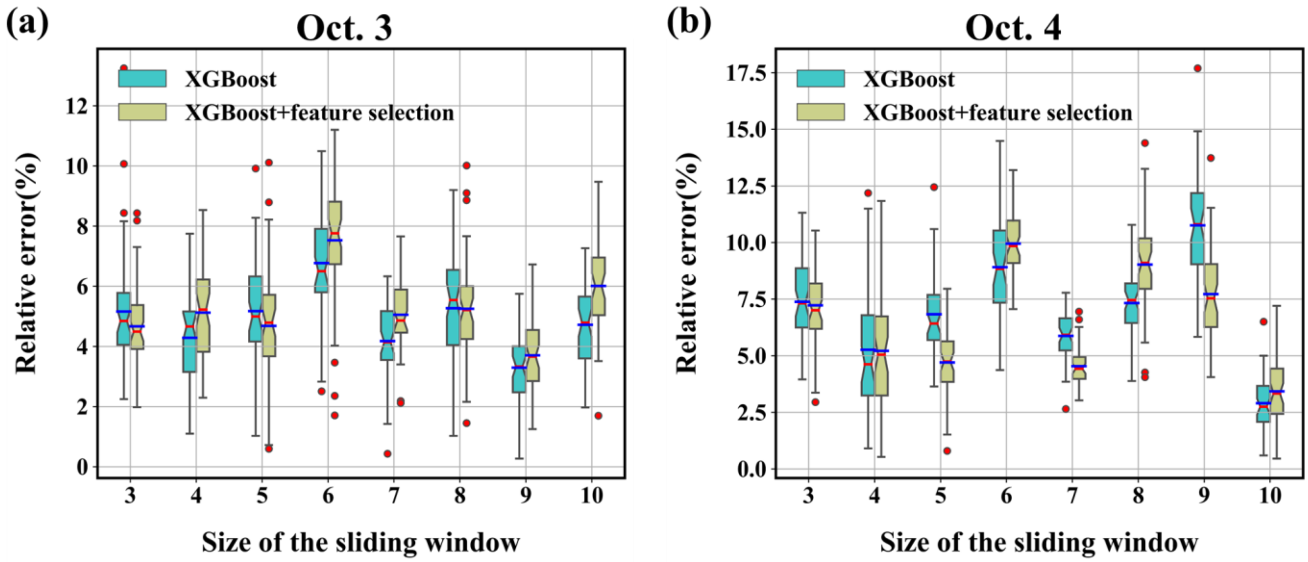


Figure 10. Sensitivity to the size of the sliding window. (a) Oct. 3; (b) Oct. 4.

3.4.3 Sensitivity to the feature type

Figure 11 compares the results obtained with different feature types. For Oct. 3, the localization errors are quite low when using only the time-domain features for the reconstruction; indeed, the errors are only slightly larger than when using all the features. In contrast, the results obtained using only the frequency-domain features exhibit noticeably larger errors, indicating that the time-domain features make a greater contribution to the results for Oct. 3. For Oct. 4, the mean localization errors are similar when using the features of either the time or frequency domain, but the error range is higher when the frequency-domain features are used. In addition, the errors of both single-domain-feature results are higher than those of the all-feature results, indicating that both feature types should be considered to ensure accurate and robust source localization.

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

Geoscientific Model Development
Discussions
Open Access

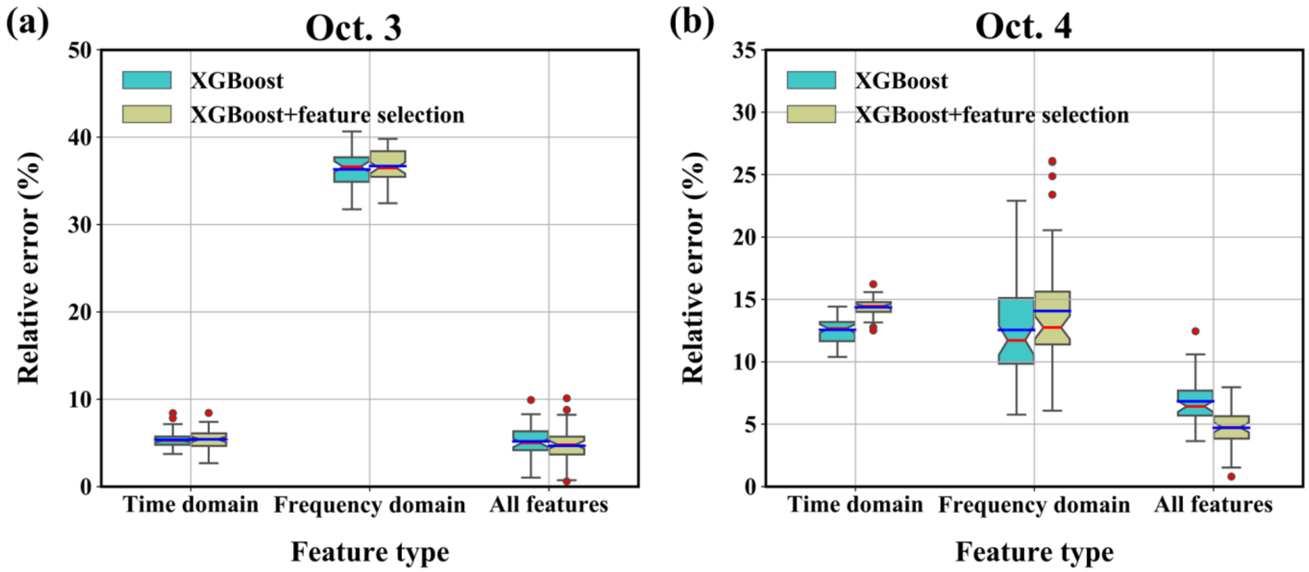


Figure 11. Sensitivity to feature type. (a) Oct. 3; (b) Oct. 4.

3.4.4 Sensitivity to the number and combination of observation sites

Figure 12 compares the results obtained with different numbers and combinations of observation sites. The results indicate that the localization error may be more sensitive to the position of the observation site than to the number of sites included. For both days, the results obtained using a subset of sites produce the lowest error level, i.e. Site ABD on Oct. 3 and Site BD on Oct. 4. These results are completely consistent with the feature importance results in Fig. 4. The implication is that sampling a plume at multiple locations with correspondingly different activity concentrations is more likely to result in a better reconstruction because the environmental variability is more fully captured through direct observations. For Oct. 3, source localization with Site B always produces lower error levels, though Site B is farthest away in the crosswind direction. However, the results on Oct. 4 do not exhibit the same phenomenon, mainly because almost all observation sites are parallel to the downwind direction. In addition, the single-site results can also achieve low error levels, such as Site B on Oct. 3 and Site A on Oct. 4. Feature selection reduces the mean error level in most test cases. These results indicate that the proposed method may achieve satisfactory results with different numbers and combinations of observation sites. It also remains effective with very few observation sites, on the condition that the observations sufficiently capture the plume.

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

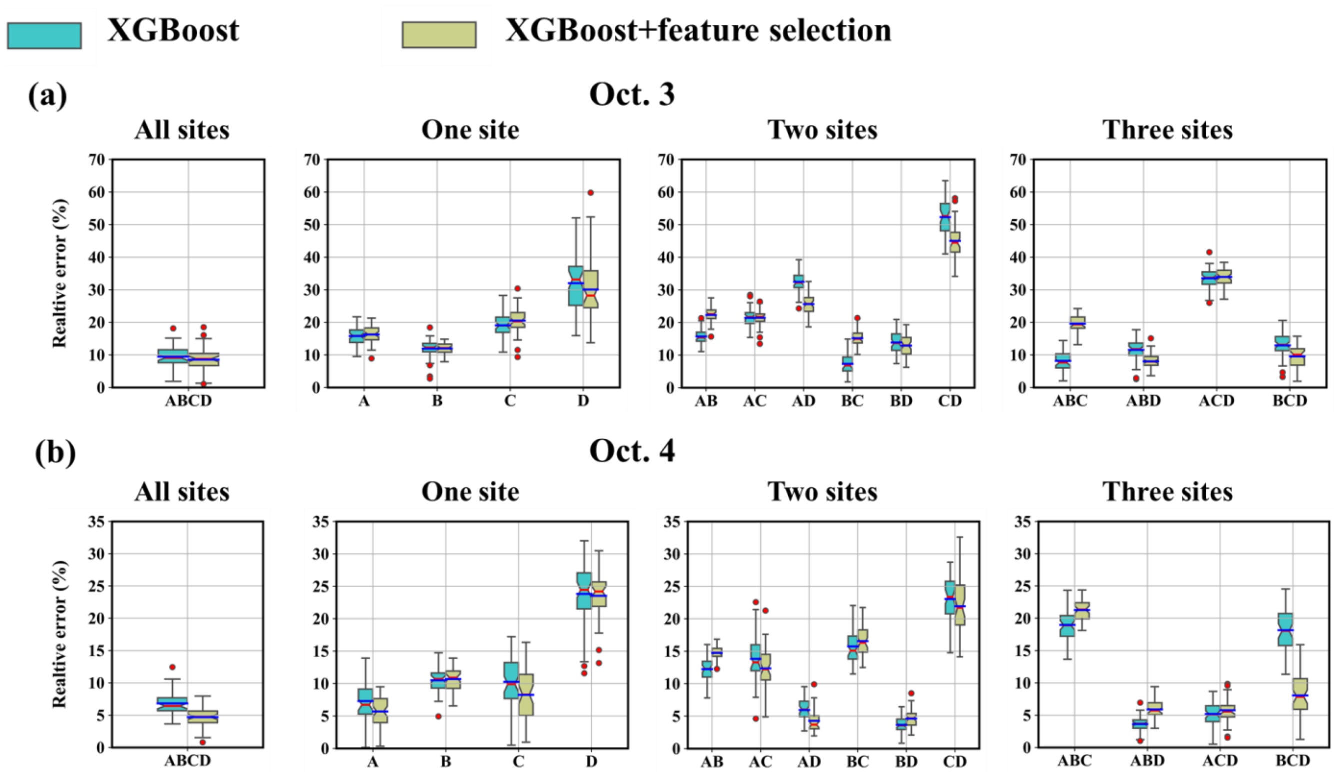


Figure 12. Sensitivity to the number and combination of observation sites. (a) Oct. 3; (b) Oct. 4.

4. Conclusions

In this study, we relaxed the unrealistic constant-release assumption of source reconstruction. Instead, we took advantage of the fact that most atmospheric radionuclide releases have a spatially fixed source and thus the release rate mainly influences the peak values in the temporal observations. Based on this, a more general spatiotemporally decoupled source reconstruction method was developed to estimate non-constant releases. Decoupling was achieved by applying a temporal sliding-window average filter to the observations. This filter reduces the influence of temporal variations in release rates on the observations, so that the relative spatiotemporal distribution of the filtered observations is dominated by the source location and known meteorology. A response feature vector was extracted to quantify the long-term temporal response trends at each observation site, involving tailored indicators of both the time and frequency domains. The XGBoost algorithm was used to train a machine learning model that links the source location to the feature vector, enabling independent source localization without knowing the release rates. With the retrieved source location, the detailed temporal variations of the release rate were determined using the PAMILT algorithm. Validation was performed against the two-day SCK-CEN ^{41}Ar field experimental data, and the results demonstrate that the proposed method successfully removes the influence of temporal variations in release rates across observations and accurately localizes the source. Source localization was achieved with deviations of only 4.68% and 4.71%

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

Geoscientific Model Development Discussions

on Oct. 3 and Oct. 4, respectively, representing reductions of 9.42% and 9.59%, respectively, compared with the results from a recent correlation-based method and 7.20% and 8.12%, respectively, compared with the results from the Bayesian method. In terms of the release rate, the PAMILT method reconstructed the temporal variations, peak, and total release with high accuracy, thus avoiding the unrealistic oscillations given by the Bayesian estimate. The proposed method also exhibited smaller uncertainty ranges in terms of the source location and total release than the Bayesian method, and avoided the multimodal distribution of the Bayesian method. Sensitivity analyses revealed that the proposed method exhibits stable decoupling and localization performance with different parameters and remains effective with only a single observation site, as long as the selected site is at an appropriate position. These results demonstrate that spatiotemporally decoupled source reconstruction is feasible and achieves satisfactory accuracy in the non-constant-release scenario, thereby providing a promising framework for reconstructing atmospheric radionuclide releases.

Code and data availability. The code and data for the proposed method can be downloaded from Zenodo (https://doi.org/10.5281/zenodo.10200141). More recent versions of the code and data will be published on GitHub.com (https://github.com/rocket1ab/Source-reconstruction-of-non-constant-atmospheric-radionuclide-releases, last access: 23 November 2023). The implementation is provided in Python, and the instruction file is also available in the provided link.

Author contributions. YX conducted the source reconstruction tests and wrote the manuscript draft; SF provided guidance on the RIMPUFF modeling and suggestions on source reconstruction tests; XD and SZ reviewed and edited the manuscript.

Competing interests. The authors have declared that they have no conflict of interest.

Acknowledgements. This work is supported by the National Natural Science Foundation of China (grant numbers 12275152 and 11875037), LingChuang Research Project of China National Nuclear Corporation, and International Atomic Energy Agency (TC project number CRP9053).

References

Akhtar, F., Li, J., Pei, Y., Xu, Y., Rajput, A., and Wang, Q.: Optimal Features Subset Selection for Large for Gestational Age Classification Using GridSearch Based Recursive Feature Elimination with Cross-Validation Scheme, in: International Conference on Frontier Computing, 63–71, https://doi.org/10.1007/978-981-15-3250-4_8, 2019.

Andronopoulos, S. and Kovalets, I. V.: Method of source identification following an accidental release at an unknown location using a lagrangian atmospheric dispersion model, Atmosphere (Basel)., 12, 7–12, https://doi.org/10.3390/atmos12101305, 2021.

Anspaugh, L. R., Catlin, R. J., and Goldman, M.: The global impact of the chernobyl reactor accident, Science (80-.)., 242, 1513–1519, https://doi.org/10.1126/science.3201240, 1988.

Bocquet, M.: Reconstruction of an atmospheric tracer source using the principle of maximum entropy. I: Theory, Q. J. R. Meteorol. Soc., 131, 2191–2208, https://doi.org/10.1256/qj.04.67, 2005a.

Bocquet, M.: Reconstruction of an atmospheric tracer source using the principle of maximum entropy. II: Applications, Q. J. R. Meteorol. Soc., 131, 2209–2223, https://doi.org/10.1256/qj.04.68, 2005b.

Chang, J. C. and Hanna, S. R.: Air quality model performance evaluation, Meteorol. Atmos. Phys., 87, 167–196, https://doi.org/10.1007/s00703-003-0070-7, 2004.

Chen, T. and Guestrin, C.: XGBoost: A scalable tree boosting system, Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 13-17-Augu, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Chow, F. K., Kosović, B., and Chan, S.: Source inversion for contaminant plume dispersionin urban environments using building-resolving simulations, J. Appl. Meteorol. Climatol., 47, 1533–1572, https://doi.org/10.1175/2007JAMC1733.1, 2008.

Davoine, X. and Bocquet, M.: Inverse modelling-based reconstruction of the Chernobyl source term available for long-range transport, Atmos. Chem. Phys., 7, 1549–1564, https://doi.org/10.5194/acp-7-1549-2007, 2007.

Drews, M., Aage, H. K., Bargholz, K., Ejsing Jørgensen, H., Korsbech, U., Lauritzen, B., Mikkelsen, T., Rojas-Palma, C., and Ammel, R. Van: Measurements of plume geometry and argon-41 radiation field at the BR1 reactor in Mol, Belgium, 1–43 pp., 2002.

Dumont Le Brazidec, J., Bocquet, M., Saunier, O., and Roustan, Y.: MCMC methods applied to the reconstruction of the autumn 2017 Ruthenium-106 atmospheric contamination source, Atmos. Environ. X, 6, 100071, https://doi.org/10.1016/j.aeaoa.2020.100071, 2020.

Dumont Le Brazidec, J., Bocquet, M., Saunier, O., and Roustan, Y.: Quantification of uncertainties in the assessment of an atmospheric release source applied to the autumn 2017 106Ru event, Atmos. Chem. Phys., 21, 13247–13267, https://doi.org/10.5194/acp-21-13247-2021, 2021.

Eamonn Keogh, Selina Chu, D. H. and M. P.: SEGMENTING TIME SERIES: A SURVEY AND NOVEL APPROACH, in: Data mining in time series databases, 1–21, https://doi.org/10.1142/9789812565402_0001, 2004.

Efthimiou, G. C., Kovalets, I. V., Venetsanos, A., Andronopoulos, S., Argyropoulos, C. D., and Kakosimos, K.: An optimized inverse modelling method for determining the location and strength of a point source releasing airborne material in urban environment, Atmos. Environ., 170, 118–129, https://doi.org/10.1016/j.atmosenv.2017.09.034, 2017.

Efthimiou, G. C., Kovalets, I. V, Argyropoulos, C. D., Venetsanos, A., Andronopoulos, S., and Kakosimos, K. E.: Evaluation of an inverse modelling methodology for the prediction of a stationary point pollutant source in complex urban environments, Build. Environ., 143, 107–119, https://doi.org/10.1016/j.buildenv.2018.07.003, 2018.

Eslinger, P. W. and Schrom, B. T.: Multi-detection events, probability density functions, and reduced location area, J. Radioanal. Nucl. Chem., 307, 1599–1605, https://doi.org/10.1007/s10967-015-4339-3, 2016.

Fang, S., Dong, X., Zhuang, S., Tian, Z., Chai, T., Xu, Y., Zhao, Y., Sheng, L., Ye, X., and Xiong, W.: Oscillation-free

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

source term inversion of atmospheric radionuclide releases with joint model bias corrections and non-smooth competing priors, J. Hazard. Mater., 440, https://doi.org/10.1016/j.jhazmat.2022.129806, 2022.

Guo, S., Yang, R., Zhang, H., Weng, W., and Fan, W.: Source identification for unsteady atmospheric dispersion of hazardous materials using Markov Chain Monte Carlo method, Int. J. Heat Mass Transf., 52, 3955–3962, https://doi.org/10.1016/j.ijheatmasstransfer.2009.03.028, 2009.

Ingremeau, J. and Saunier, O.: Investigations on the source term of the detection of radionuclides in North of Europe in June 2020, EPJ Nucl. Sci. Technol., 8, 10, https://doi.org/10.1051/epjn/2022003, 2022.

Jensen, D. D., Lucas, D. D., Lundquist, K. A., and Glascoe, L. G.: Sensitivity of a Bayesian source-term estimation model to spatiotemporal sensor resolution, Atmos. Environ. X, 3, https://doi.org/10.1016/j.aeaoa.2019.100045, 2019.

Katata, G., Ota, M., Terada, H., Chino, M., and Nagai, H.: Atmospheric discharge and dispersion of radionuclides during the Fukushima Dai-ichi Nuclear Power Plant accident. Part I: Source term estimation and local-scale atmospheric dispersion in early phase of the accident, J. Environ. Radioact., 109, 103–113, https://doi.org/10.1016/j.jenvrad.2012.02.006, 2012.

Keats, A., Yee, E., and Lien, F. S.: Bayesian inference for source determination with applications to a complex urban environment, Atmos. Environ., 41, 465–479, https://doi.org/10.1016/j.atmosenv.2006.08.044, 2007.

Keats, A., Yee, E., and Lien, F. S.: Information-driven receptor placement for contaminant source determination, Environ. Model. Softw., 25, 1000–1013, https://doi.org/10.1016/j.envsoft.2010.01.006, 2010.

KIM, J. Y., JANG, H.-K., and LEE, J. K.: Source Reconstruction of Unknown Model Parameters in Atmospheric Dispersion Using Dynamic Bayesian Inference, Prog. Nucl. Sci. Technol., 1, 460–463, https://doi.org/10.15669/pnst.1.460, 2011.

Kovalets, I. V., Efthimiou, G. C., Andronopoulos, S., Venetsanos, A. G., Argyropoulos, C. D., and Kakosimos, K. E.: Inverse identification of unknown finite-duration air pollutant release from a point source in urban environment, Atmos. Environ., 181, 82–96, https://doi.org/10.1016/j.atmosenv.2018.03.028, 2018.

Kovalets, I. V., Romanenko, O., and Synkevych, R.: Adaptation of the RODOS system for analysis of possible sources of Ru-106 detected in 2017, J. Environ. Radioact., 220–221, https://doi.org/10.1016/j.jenvrad.2020.106302, 2020.

Krysta, M. and Bocquet, M.: Source reconstruction of an accidental radionuclide release at European scale, Q. J. R. Meteorol. Soc., 133, 529–544, https://doi.org/10.1002/qj.3, 2007.

Li, X., Xiong, W., Hu, X., Sun, S., Li, H., Yang, X., Zhang, Q., Nibart, M., Albergel, A., and Fang, S.: An accurate and ultrafast method for estimating three-dimensional radiological dose rate fields from arbitrary atmospheric radionuclide distributions, Atmos. Environ., 199, 143–154, https://doi.org/10.1016/j.atmosenv.2018.11.001, 2019a.

Li, X., Sun, S., Hu, X., Huang, H., Li, H., Morino, Y., Wang, S., Yang, X., Shi, J., and Fang, S.: Source inversion of both long- and short-lived radionuclide releases from the Fukushima Daiichi nuclear accident using on-site gamma dose rates, J. Hazard. Mater., 379, 120770, https://doi.org/10.1016/j.jhazmat.2019.120770, 2019b.

Lucas, D. D., Simpson, M., Cameron-Smith, P., and Baskett, R. L.: Bayesian inverse modeling of the atmospheric transport and emissions of a controlled tracer release from a nuclear power plant, Atmos. Chem. Phys., 17, 13521–13543, https://doi.org/10.5194/acp-17-13521-2017, 2017.

Ma, D., Tan, W., Wang, Q., Zhang, Z., Gao, J., Wang, X., and Xia, F.: Location of contaminant emission source in atmosphere based on optimal correlated matching of concentration distribution, Process Saf. Environ. Prot., 117, 498–510, https://doi.org/10.1016/j.psep.2018.05.028, 2018.

Masson, O., Steinhauser, G., Zok, D., Saunier, O., Angelov, H., Babić, D., Bečková, V., Bieringer, J., Bruggeman, M., Burbidge, C. I., Conil, S., Dalheimer, A., De Geer, L. E., De Vismes Ott, A., Eleftheriadis, K., Estier, S., Fischer, H., Garavaglia, M. G., Gasco Leonarte, C., Gorzkiewicz, K., Hainz, D., Hoffman, I., Hýža, M., Isajenko, K., Karhunen, T., Kastlander, J., Katzlberger, C., Kierepko, R., Knetsch, G. J., Kövendiné Kónyi, J., Lecomte, M., Mietelski, J. W., Min, P., Møller, B., Nielsen, S. P., Nikolic, J., Nikolovska, L., Penev, I., Petrinec, B., Povinec, P. P., Querfeld, R., Raimondi, O., Ransby, D., Ringer, W., Romanenko, O., Rusconi, R., Saey, P. R. J., Samsonov, V., Šilobritiene, B., Simion, E., Söderström, C., Šoštarić, M., Steinkopff, T., Steinmann, P., Sýkora, I., Tabachnyi, L., Todorovic, D., Tomankiewicz, E., Tschiersch, J., Tsibranski, R., Tzortzis, M., Ungar, K., Vidic, A., Weller, A., Wershofen, H., Zagyvai, P., Zalewska, T., Zapata García, D., and Zorko, B.: Airborne concentrations and chemical considerations of radioactive ruthenium from an undeclared major nuclear release in 2017, Proc. Natl. Acad. Sci. U. S. A., 116, 16750–16759, https://doi.org/10.1073/pnas.1907571116, 2019.

Meutter, P. De and Hoffman, I.: Bayesian source reconstruction of an anomalous Selenium-75 release at a nuclear research institute, J. Environ. Radioact., 218, 106225, https://doi.org/10.1016/j.jenvrad.2020.106225, 2020.

De Meutter, P., Hoffman, I., and Ungar, K.: On the model uncertainties in Bayesian source reconstruction using an ensemble of weather predictions, the emission inverse modelling system FREAR v1.0, and the Lagrangian transport and dispersion model Flexpart v9.0.2, Geosci. Model Dev., 14, 1237–1252, https://doi.org/10.5194/gmd-14-1237-2021, 2021.

Monache, L. D., Lundquist, J. K., Kosoví, 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, J. Appl. Meteorol. Climatol., 47, 2600–2613, https://doi.org/10.1175/2008JAMC1766.1, 2008.

Rajaona, H., Septier, F., Armand, P., Delignon, Y., Olry, C., Albergel, A., and Moussafir, J.: An adaptive Bayesian inference algorithm to estimate the parameters of a hazardous atmospheric release, Atmos. Environ., 122, 748–762, https://doi.org/10.1016/j.atmosenv.2015.10.026, 2015.

Rojas-Palma, C., Aage, H. K., Astrup, P., Bargholz, K., Drews, M., Jørgensen, H. E., Korsbech, U., Lauritzen, B., Mikkelsen, T., Thykier-Nielsen, S., and Van Ammel, R.: Experimental evaluation of gamma fluence-rate predictions from argon-41 releases to the atmosphere over a nuclear research reactor site, Radiat. Prot. Dosimetry, 108, 161–168, https://doi.org/10.1093/rpd/nch020, 2004.

Saunier, O., Didier, D., Mathieu, A., Masson, O., and Dumont Le Brazidec, J.: Atmospheric modeling and source reconstruction of radioactive ruthenium from an undeclared major release in 2017, Proc. Natl. Acad. Sci. U. S. A., 116, 24991–25000, https://doi.org/10.1073/pnas.1907823116, 2019.

Senocak, I.: Application of a Bayesian inference method to reconstruct short-range atmospheric dispersion events, AIP Conf. Proc., 1305, 250–257, https://doi.org/10.1063/1.3573624, 2010.

https://doi.org/10.5194/gmd-2023-173
Preprint. Discussion started: 27 November 2023

Senocak, I., Hengartner, N. W., Short, M. B., and Daniel, W. B.: Stochastic event reconstruction of atmospheric contaminant dispersion using Bayesian inference, Atmos. Environ., 42, 7718–7727, https://doi.org/10.1016/j.atmosenv.2008.05.024, 2008.

Thykier-Nielsen, S., Deme, S., and Mikkelsen, T.: Description of the atmospheric dispersion module RIMPUFF, Riso Natl. Lab. PO Box, 49, 1999.

Tomas, J. M., Peereboom, V., Kloosterman, A., and van Dijk, A.: Detection of radioactivity of unknown origin: Protective actions based on inverse modelling, J. Environ. Radioact., 235–236, 106643, https://doi.org/10.1016/j.jenvrad.2021.106643, 2021.

Wang, Y., Huang, H., Huang, L., and Ristic, B.: Evaluation of Bayesian source estimation methods with Prairie Grass observations and Gaussian plume model: A comparison of likelihood functions and distance measures, Atmos. Environ., 152, 519–530, https://doi.org/10.1016/j.atmosenv.2017.01.014, 2017.

Xue, F., Li, X., and Zhang, W.: Bayesian identification of a single tracer source in an urban-like environment using a deterministic approach, Atmos. Environ., 164, 128–138, https://doi.org/10.1016/j.atmosenv.2017.05.046, 2017a.

Xue, F., Li, X., Ooka, R., Kikumoto, H., and Zhang, W.: Turbulent Schmidt number for source term estimation using Bayesian inference, Build. Environ., 125, 414–422, https://doi.org/10.1016/j.buildenv.2017.09.012, 2017b.

Yee, E.: Automated computational inference engine for Bayesian source reconstruction: application to some detections/non-detections made in the CTBT international monitoring system, Appl. Math. Sci., 11, 1581–1618, https://doi.org/10.12988/ams.2017.74149, 2017.

Yee, E., Lien, F. S., Keats, A., and D'Amours, R.: Bayesian inversion of concentration data: Source reconstruction in the adjoint representation of atmospheric diffusion, J. Wind Eng. Ind. Aerodyn., 96, 1805–1816, https://doi.org/10.1016/j.jweia.2008.02.024, 2008.

Zhao, Y., Liu, Y., Wang, L., Cheng, J., Wang, S., and Li, Q.: Source Reconstruction of Atmospheric Releases by Bayesian Inference and the Backward Atmospheric Dispersion Model: An Application to ETEX-I Data, Sci. Technol. Nucl. Install., 2021, https://doi.org/10.1155/2021/5558825, 2021.