

Research Article

Source Reconstruction of Atmospheric Releases by Bayesian Inference and the Backward Atmospheric Dispersion Model: An Application to ETEX-I Data

Yungang Zhao ^{1,2}, Yuanyuan Liu ¹, Li Wang,³ Jianping Cheng,¹ Shilian Wang,² and Qi Li²

¹Key Laboratory of Beam Technology of Ministry of Education, College of Nuclear Science and Technology, Beijing Normal University, Beijing 100875, China

²CTBT Beijing National Data Centre and Beijing Radionuclide Laboratory, Beijing 100085, China

³Department of Physics, Beijing Normal University, Beijing 100875, China

Correspondence should be addressed to Yuanyuan Liu; yyluu@bnu.edu.cn

Received 19 February 2021; Revised 2 April 2021; Accepted 20 April 2021; Published 28 April 2021

Academic Editor: Sheng Fang

Copyright © 2021 Yungang Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Source term reconstruction methods attempt to calculate the most likely source parameters of an atmospheric release given measurements, including both location and release amount. However, source term reconstruction is vulnerable to uncertainties. In this paper, a method combining Bayesian inference with the backward atmospheric dispersion model is developed for robust source term reconstruction. The backward model is used to quantify the relationship between the source and measurements and to reduce the search range of the Bayesian inference. A Markov chain Monte Carlo method is used to sample from the multi-dimensional parameter space of the source term. The source location and release rate are estimated simultaneously, and the posterior probability distribution is produced by applying Bayes' theorem. The proposed method is applied to a set of real concentration data from the ETEX-I experiment. The results demonstrate that the source location is estimated to be $-2.86^\circ \pm 1.01^\circ\text{E}$, $48.25^\circ \pm 0.33^\circ\text{N}$, and the release rate is estimated to be $20.16 \pm 3.56\text{ kg/h}$. The true source location is correctly estimated to be within a one standard deviation interval, and the release rate is correctly determined to be within a three standard deviation interval.

1. Introduction

The accurate and timely evaluation of pollutant source (biological, chemical, and radioactive materials) parameters plays an important role in the emergency response. In nuclear accidents, such as the nuclear accident in Fukushima in 2011 and the Ruthenium-106 plume over Europe accident in 2017, the source locations or release rates are unknown. A major concern in these events is how to locate the pollution source and estimate the release rate. An effective way is to reconstruct the source by using environmental monitoring data and a dispersion model. Such methods have been applied to wind tunnel experiments [1, 2], tracer experiments [3, 4], and real accidents such as nuclear accidents

[5, 6] and radionuclide emissions under normal conditions [7].

Among these studies, two notable methods include Bayesian inference and the backward atmospheric model. Compared with deterministic optimization methods which provide a single optimal solution, Bayesian inference can provide a probability distribution of the source parameters based on the prior information of unknown parameters and monitoring data, but this method is time consuming for enormous sampling [8]. The backward atmospheric model can rapidly provide the possible source region for each measurement site, but it relies on further analysis to quantitatively determine the position and release rate of the source [9].

However, there are few reports on the combination of Bayesian inference and a backward atmospheric model.

For robust source term reconstruction, a new method combining Bayesian inference with the backward run model FLEXPART is proposed in this study. The source-receptor relationship and the possible source region are obtained through backward simulations. A Bayesian method combined with MCMC sampling is used to estimate the source term position and release rate simultaneously, together with their posteriori distributions. The proposed method is validated using the ETEX-I experiment by comparing both the source parameters and the concentration simulations.

2. Methodology

The source term reconstruction method based on Bayesian inference is a method that combines the prior information of unknown parameters with measurements to transform them into posterior information, and it has been applied to different scenarios [10–12]. The following sections briefly summarize Bayesian inference for source reconstruction and the derivation of the source-receptor relationship.

2.1. Bayesian Inference for Source Reconstruction. The aim of Bayesian inference for source reconstruction is to determine the a posteriori probability distribution of source parameters. The governing equation is Bayes' theorem, which can be expressed as follows:

$$P(S|M) = \frac{P(M|S)P(S)}{P(M)}, \quad (1)$$

where S is the state vector formed by the source parameter and M is a vector formed by the measurements.

The terms in equation (1) can be identified as follows. First, $P(S)$ is the prior probability for the source parameters prior to any knowledge. It mainly comes from previous data, historical experience, and subjective judgment. The prior probability distribution includes uniform and normal distribution. Second, $P(M|S)$ is the likelihood function and is the probability given the source parameter S , which indicates the fitness between the modeled and measured concentrations. Here, we define the likelihood function as the normal distribution function.

$$P(M|S) = \frac{1}{\prod_{i=1}^N \sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2} \sum_{i=1}^N \left(\frac{M_i - C_i}{\sigma_i}\right)^2\right), \quad (2)$$

where N is the number of monitoring data, M_i is the i th monitoring data, C_i is the i th data predicted by the atmospheric transport model for a known source S , which will be described in Section 2.2, and σ_i is the error between M_i and C_i . Third, $P(M)$ is the integration of $P(M|S)P(S)$ at all source parameters and is referred to as the evidence. It can be simply a normalization constant. Finally, $P(S|M)$ is the posterior distribution and the answer for the source reconstruction.

2.2. Source-Receptor Relationship (SRR). The determination of the likelihood function requires the computation of C_i , i.e., the simulated concentrations. Because Bayesian inference requires a repeated computation of C_i , a direct simulation using an atmospheric dispersion model can be highly computationally demanding. An alternative method is to calculate a matrix representing the relationship between the release source and the concentration at the stations and to use matrix-vector multiplication instead. This matrix is called a source-receptor relationship. The advantage of calculating an SRR is that the calculation of the simulated concentration can avoid the numerical modeling process, which usually takes several minutes or hours. There are two ways to obtain the SRR: one is based on the source, using forward simulation. A forward simulation can establish the relationship between one source (x_0, y_0, q_0) and a series of monitoring data $C(C_1, C_2, \dots, C_N)$; the other is based on the receptor, using backward simulation [13]. By reversing the sign of the advection, a large number of particles are released at the receptor and transported backward in time. The residence time in the grid is proportional to the contribution of a source in the grid cell. One backward simulation can establish the relationship between one monitoring dataset C_i and a series of source data through the residence time t_r . When we fix the source parameters, the concentration C_i at a station can be computed as follows:

$$C_i = \frac{qt_r}{Vdt}. \quad (3)$$

V is the grid volume, which is calculated based on the grid setting parameters, including the grid resolution and level heights, and dt is the time resolution in the ATM. A more detailed description of the backward simulation can be found in [13].

In the case of unknown source term locations, the possible number of source term locations is generally much larger than the number of monitoring data. More simulations are needed in the forward mode than in the backward mode. Under this condition, the backward simulation is more efficient than the forward simulation in computational cost.

3. Sampling Method and Convergence

Markov chain Monte Carlo (MCMC) algorithms [14] are commonly used to generate random samples to obtain the posterior probability density function of source parameters. Many MCMC algorithms have been developed, including the Metropolis–Hastings (M-H) algorithm and Gibbs algorithm. Currently, the M-H algorithm is commonly used in source term reconstruction research; during one M-H sampling, parameters are disturbed simultaneously.

The process of the M-H algorithm is as follows:

- (1) Initiation: generate the initial point $I = (x_0, y_0, q_0)$ depending on the prior probability.
- (2) Perturbation: the new source $I' = (x_0 + dx, y_0 + dy, q_0 + dq)$ is obtained by unbiased perturbation.

- (3) Accept or reject: the likelihood functions $P(M|I)$ and $P(M|I')$ are calculated and compared, and the ratio $\Delta p = P(M|I')/P(M|I)$ is obtained. At the same time, a random number r between $[0, 1]$ is generated. When $R \leq \Delta p$, I' is accepted; otherwise, I' is rejected.
- (4) Iteration: repeat Steps 2 and 3 until the convergence condition is reached.

In practice, the convergence condition is reached when the posterior distribution is stable. For MCMC sampling with multiple chains, the between-chain variance B and within-chain variance W are generally used to evaluate the convergence of sampling results [15]. For m Markov chains of length n , the variance between chains B is

$$B = \frac{n}{m-1} \sum_{j=1}^m (\bar{S}_j - \bar{S}), \quad (4)$$

where

$$\begin{aligned} \bar{S}_j &= \frac{1}{n} \sum_{i=1}^n s_{ij}, \\ \bar{S} &= \frac{1}{m} \sum_{j=1}^m \bar{S}_j. \end{aligned} \quad (5)$$

The within-chain variance W is

$$W = \frac{1}{m} \sum_{j=1}^m w_j^2, \quad (6)$$

where

$$w_j^2 = \frac{1}{n-1} \sum_{i=1}^n (s_{ij} - \bar{S}_j)^2. \quad (7)$$

The convergence value is

$$\text{Cov} = \frac{\text{var}(S)}{W}, \quad (8)$$

where

$$\text{var}(S) = \frac{n-1}{n} W + \frac{1}{n} B. \quad (9)$$

The posterior distribution of the source parameters can be obtained by statistical analysis of the stable results after the chains reach the convergence condition.

Therefore, the implementation process of source term reconstruction can be shown in Figure 1.

4. Application

4.1. Benchmark. In this section, the proposed methods are used to evaluate the source term location and release rate of ETEX-I.

ETEX-I is an atmospheric tracer experiment organized by many European countries [16]. It was carried out from

October to November 1994. The tracer used was PMCH. The source location was Monterfil, Brittany, France (-2.0083°E , 48.058°N). A total of 168 monitoring stations were set up in 17 countries, and 3104 available monitoring data points were obtained. The monitoring stations and the release point are shown in Figure 2.

To investigate the ability of source term reconstruction with less monitoring data, only 11 stations with 58 monitoring data points were selected, including detection and no-detection data points. The distribution of selected stations is also shown in Figure 2.

4.2. Atmospheric Transport Model and Meteorological Data. The atmospheric transport model used in this work is FLEXPART V8.2 [17]. The meteorological data are NCEP Climate Forecast System Reanalysis (CFRS) 6-hour products [18], a global gridded dataset with a spatial resolution of $0.5^\circ \times 0.5^\circ$ and temporal resolution of 6 h.

The FLEXPART model is a Lagrangian particle diffusion model developed by Norwegian Atmospheric Research Institute. It can run in both forward and backward modes. This work mainly uses the backward mode. Based on the site locations and sampling information of the measurements, the backward simulation of each monitoring data point is carried out, and parts of the results are shown in Figure 3. The backward simulations show that the possible source region moves from the stations to the true source point.

4.3. Prior Distribution of Source Term Parameters. According to Section 4.2, the backward simulation results, the possible sources are mainly distributed in Central and Western Europe, Southern England, and the surrounding maritime areas. Based on these results, the sampling space of source term parameters is set as follows: longitude ($-10^\circ \sim 10^\circ\text{E}$), latitude ($45^\circ \sim 55^\circ\text{N}$), and release rate ($0 \sim 100 \text{ kg/h}$). In this paper, assume the prior distribution of parameters is uniform. According to the prior distribution density function, 1000 source terms are generated as the initial points of the MCMC chain.

5. Results

5.1. Convergence. Source terms are iterated according to the sampling method in Section 3, and the iterative result of location parameters of one chain is shown in Figure 4. The initial point is located at (7.5°E , 53.25°N). During the iterative calculation of Figure 1, the refined source term gradually approaches the true source location and finally stabilizes near the true source location.

The evolution of the mean value and standard deviation of the estimated source parameters is calculated and shown in Figure 5. After 3000 iterations, the mean values of all three parameters become stable. The standard deviations also show obvious reductions for all three estimated parameters. For latitude, the standard deviation is 0.33° at the steady state, which is the smallest among the three parameters. The standard deviation of the longitude is slightly larger, which is 1.01° at the steady state. This phenomenon demonstrates that

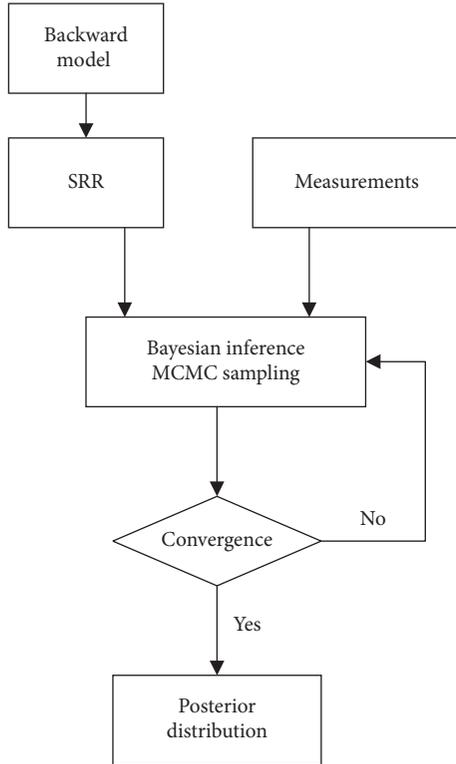


FIGURE 1: Flowchart of source term reconstruction via Bayesian inference.

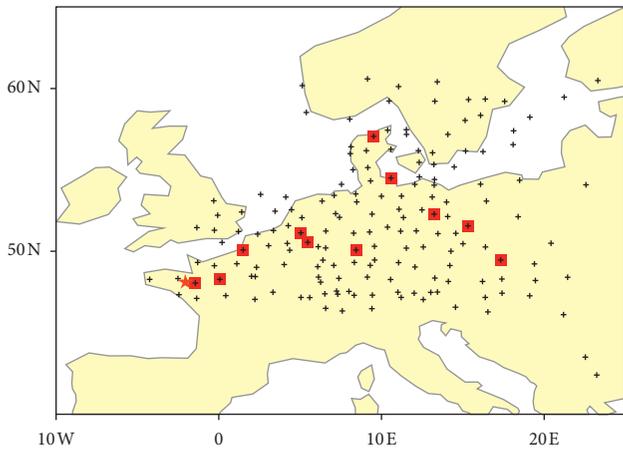


FIGURE 2: ETEX-I monitoring stations and the release point. The red star indicates the source release point, the black cross indicates 168 monitoring stations, and the red square indicates the station measurements selected in this work.

the uncertainties in the estimation of the two location parameters can be different. For the release rate, the standard deviation is 3.56 kg/h.

Based on the convergence method in Section 3, the convergence values of source term parameters are calculated. Figure 6 shows that, among the three parameters of longitude, latitude, and release rate, latitude converges the fastest, and the release rate converges the slowest. The

relatively better convergences of the location parameters may be attributed to the better initial values provided by the backward simulations in Section 4.2.

5.2. Estimated Results. Figure 7 displays the posterior probability distributions of the parameters. The sampling results and true source parameters based on the prior distribution of source terms are also shown in Figure 7. The posterior probability distribution reveals that the mean value, the standard deviation, and the most likely value of the source parameters are $-2.86 \pm 1.01^\circ\text{E}$ (-2.25°E), $48.24 \pm 0.33^\circ\text{N}$ (48.25°N), and $20.16 \pm 3.56 \text{ kg/h}$ (21.0 kg/h). Figure 7(a) shows the longitude values' posterior distribution. The distribution has two peaks to the west of the true point, with deviations of 0.25° and 1.5° . The two peaks are highly influenced by the wind field, which is mainly from west to east. Figure 7(b) shows the posterior distribution of latitude values. The majority of the values are within 0.5° of the true value. Figure 7(c) shows the release rate values of the accepted states. The release rate was correctly determined to be within a three standard deviation interval. The difference between the estimated release rate and the true value mainly comes from the model error. The statistics of the a posteriori distribution are listed in Table 1. Generally, the most likely values are closer to the true values than the mean values. These results indicate that the most likely estimate may be more reliable for source reconstruction in a real nuclear accident. Compared with the location, the release rate shows relatively larger errors and standard deviations.

5.3. Comparison of Simulations and Measurements. The observed concentrations and concentrations simulated by using different source terms were compared to validate the proposed source reconstruction methods.

- (i) FWD_t: forward simulation with the true source term
- (ii) FWD_m: forward simulation with the mean source term in Table 1
- (iii) FWD_p: forward simulation with the most probability source term in Table 1

Figure 8 shows the plumes modeled by different source terms at different times. It is noticeable that the plume shapes are similar to each other for the same times, indicating that the meteorological input plays a dominant role in the simulation. Although the two estimates have relatively large differences in their longitude (Table 1), the concentration distributions are generally much the same, especially for the latter times (the middle and right columns of Figure 8). This phenomenon implies that the simulation is not that sensitive to the displacement of the source positions. However, the high-concentration area of the simulation using the true source term is quite different from that using the most likely and mean estimate, indicating that the release rate differences have a larger influence on the simulation. These differences in sensitivities to the source coordinate and

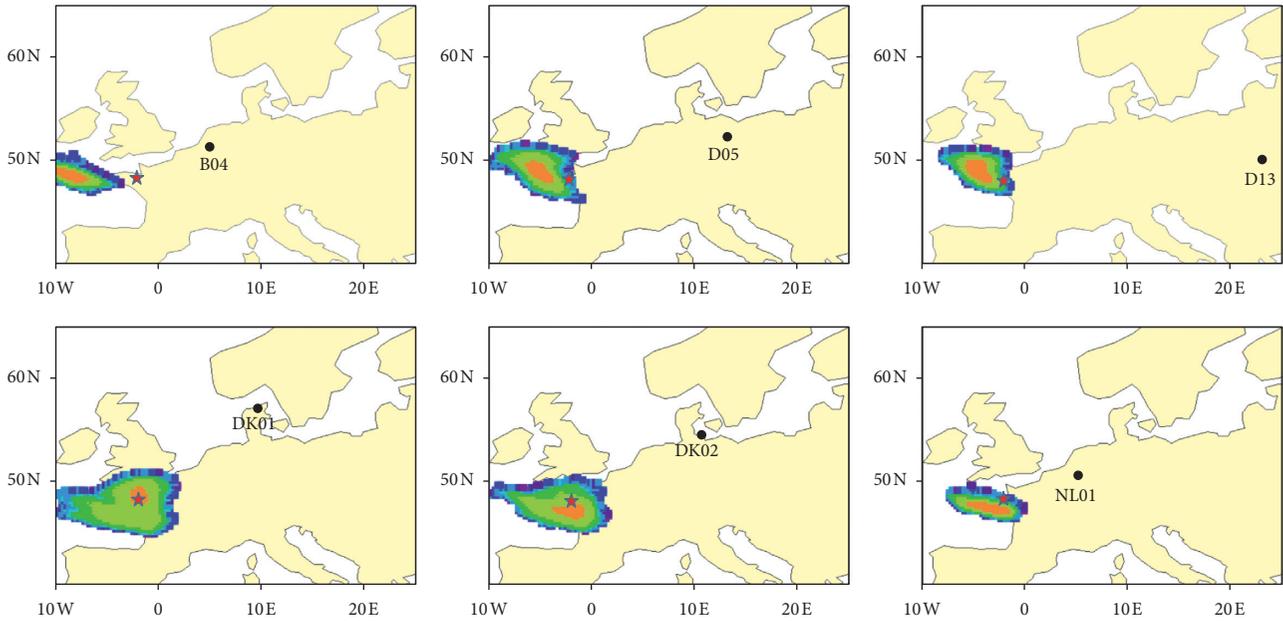


FIGURE 3: Backward simulation results. The red star indicates the source release point, and the black circle indicates the monitoring station.

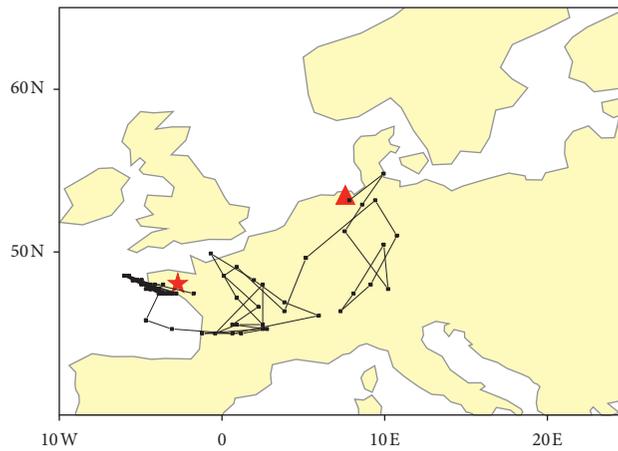


FIGURE 4: A Markov chain Monte Carlo sampling result for the source term location. The red triangle indicates the initial point, and the red star indicates the true source location.

the release rate may partially explain different deviations of the source reconstruction for these two variables (Table 1). At some measurements in the center of the plume, the simulation with the true release rate exhibits even larger deviations than the simulations using the other two source term estimates (the first column in Figure 8). This phenomenon indicates that the air dispersion model may have its own deviations, and such deviations are cancelled out by the deviations in the most likely estimate and the mean estimate to some degree. For those measurements outside the plume, the uncertainties in the meteorological input can be the major source of these deviations.

Figure 9 compares the temporal profiles of the observed and simulated concentrations at representative sites. For all 4 sites, the simulated timings of the peaks generally show deviations within a 3 h interval compared with the monitoring results, whereas the deviations of the simulated peak

values are within 50% at these four sites. For more than 70% of stations, the deviations in the estimation of the peak values are within a factor of 5. There is also a difference in the number of peaks between the simulations and measurements. Taking station D06 as an example, a small peak is shown behind the main peak in the measurement data, while it is not shown in the simulated data. Again, it is noticeable that the simulated profiles using the true release rate show the most obvious deviations at D05 and D13, which significantly overestimate the number and magnitude of the peaks. For the other two source terms, the simulated peak timing is basically the same, but the peak values are different. Thus, in addition to the release rate, it is also necessary to reduce the uncertainties in the input meteorological fields and the physical processes of the dispersion model to improve the simulation results. This improvement shall be our future work.

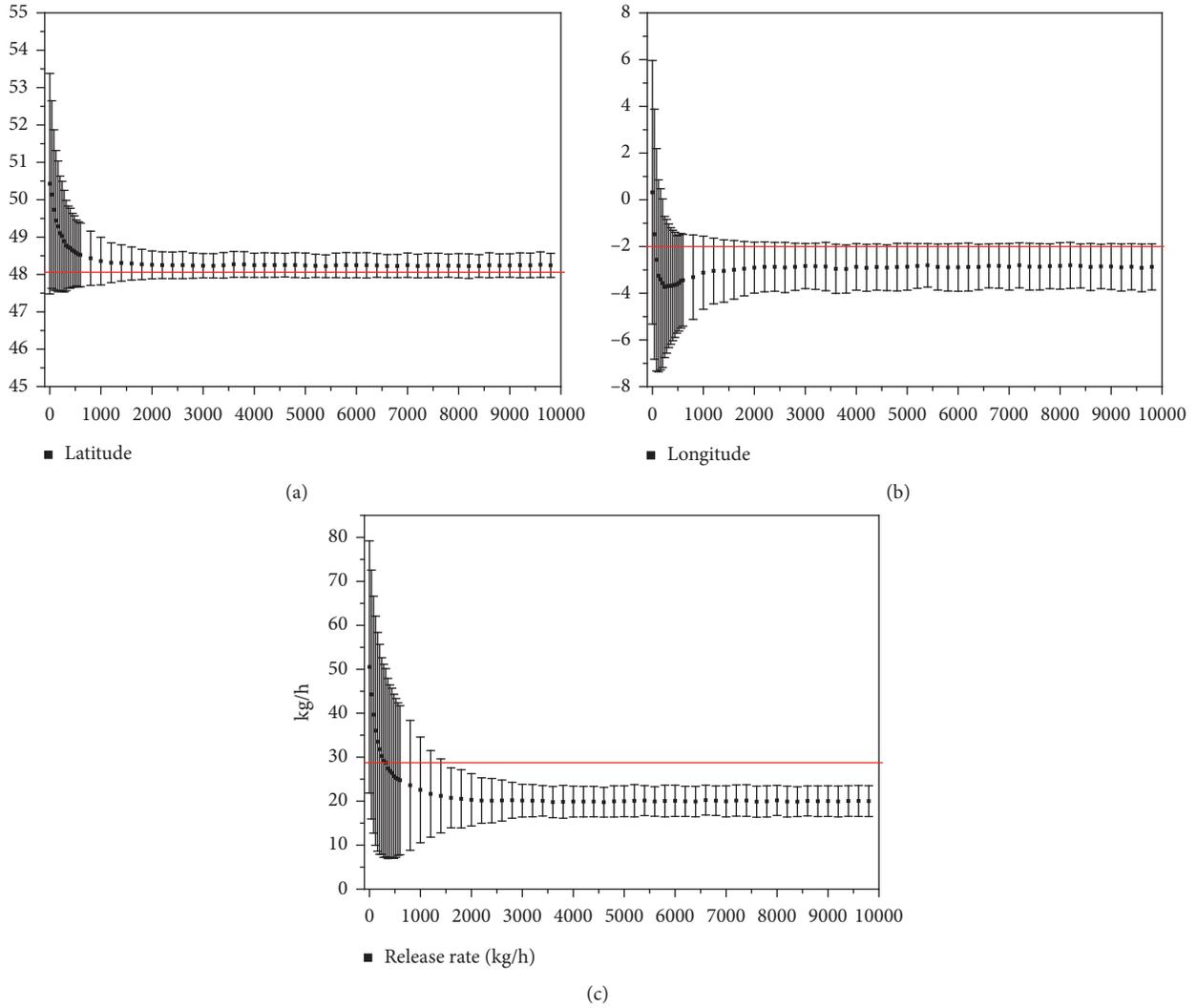


FIGURE 5: The mean values and standard deviations of source parameters after iterating: (a) latitude; (b) longitude; (c) release rate.

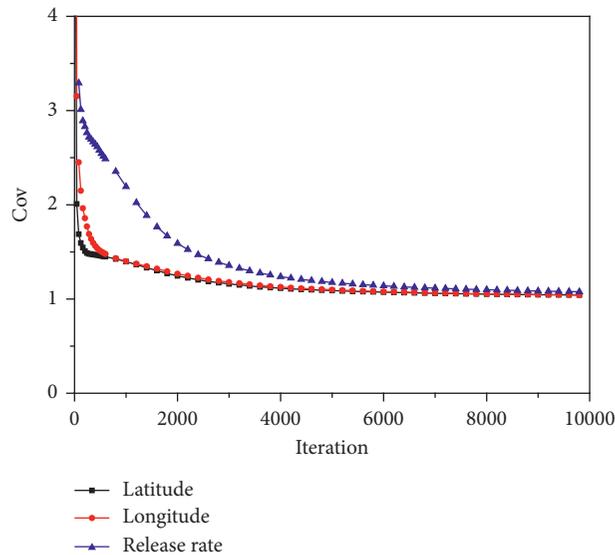


FIGURE 6: Convergence values vs. the number of iterations.

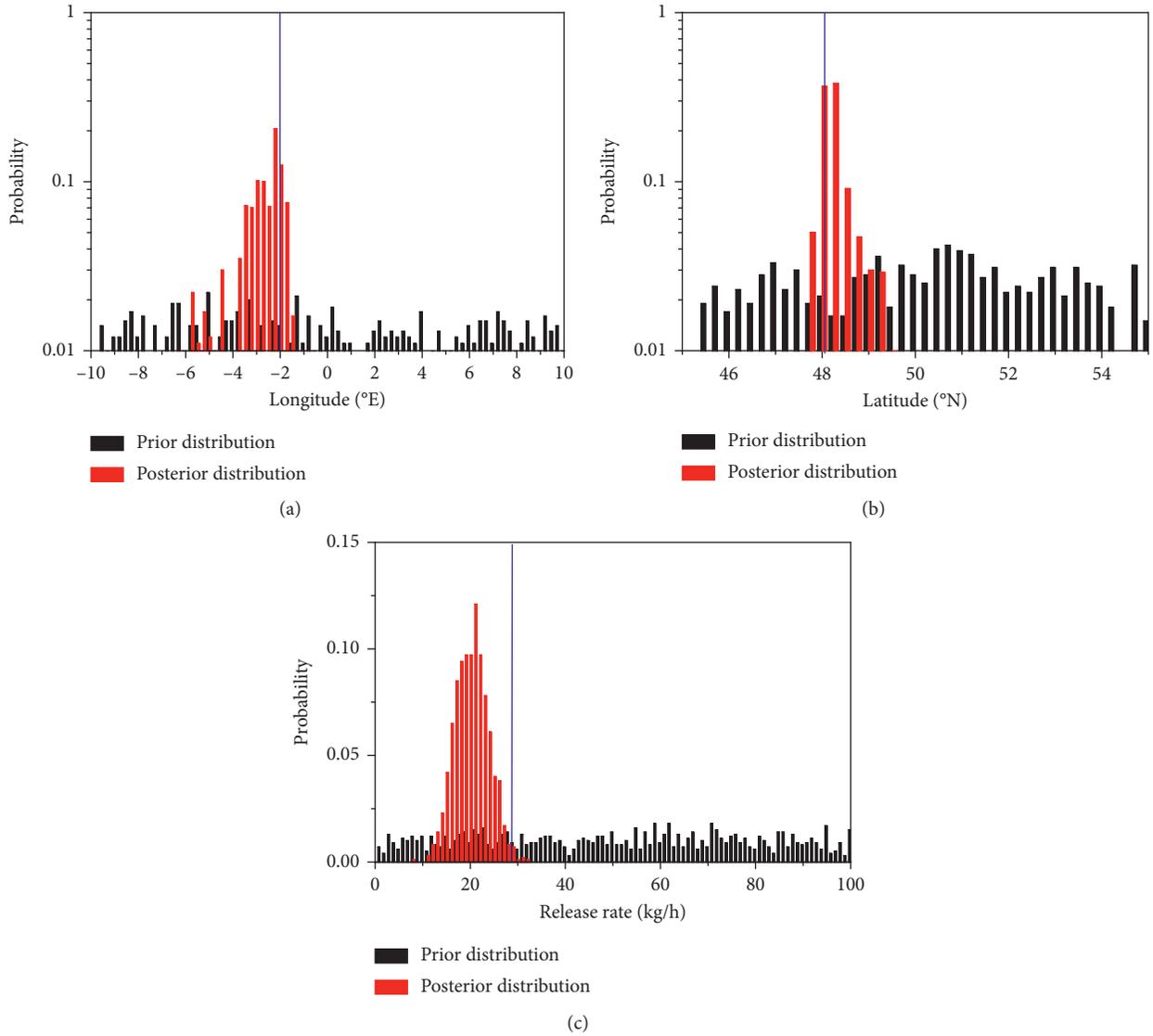


FIGURE 7: Prior distribution (black bar), posterior distribution (red bar), and true value (blue bar) of source term parameters.

TABLE 1: Summary statistics of the MCMC samples.

| | Lon (°E) | Lat (°N) | Rate (kg/h) |
|--------------------|----------|----------|-------------|
| Most prob | -2.25 | 48.25 | 21.0 |
| Mean | -2.86 | 48.24 | 20.16 |
| Standard deviation | 1.01 | 0.33 | 3.56 |
| True | -2.0083 | 48.058 | 28.73 |

Table 2 compares the quantitative metrics of the simulations using three different source terms. The simulations using the true release rate exhibit the worst metrics except

for the correlation R , implying that the model uncertainties may be further reduced. For the other two source terms, their performances are quite close to each other, except FB.

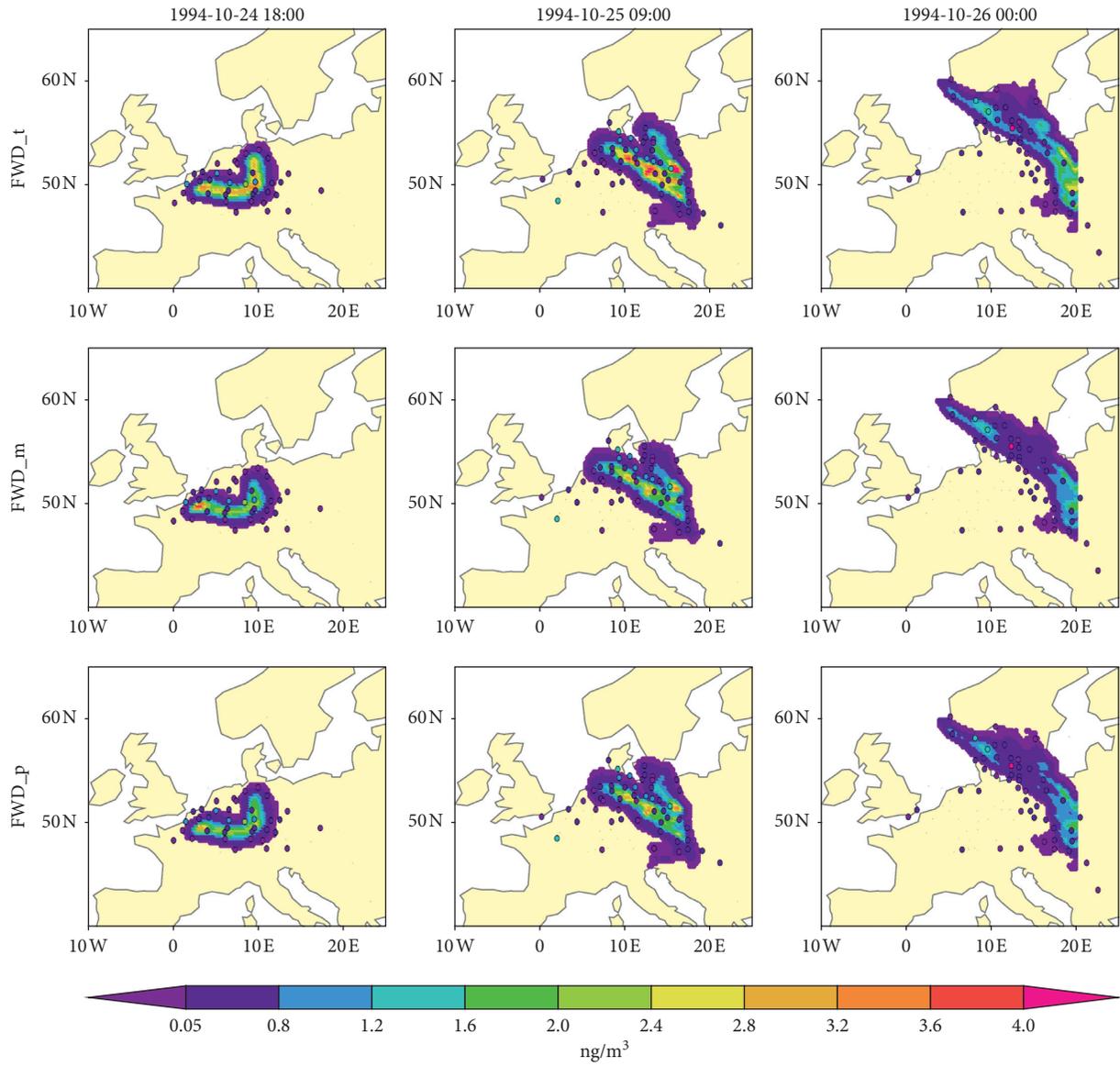


FIGURE 8: Comparison of simulated plumes by different methods for (top) FWD_t, (middle) FWD_m, and (bottom) FWD_p and measurements for different times: (left) 1994-10-24 18:00, (middle) 1994-10-25 09:00, and (right) 1994-10-26 00:00. The circles are the monitoring data.

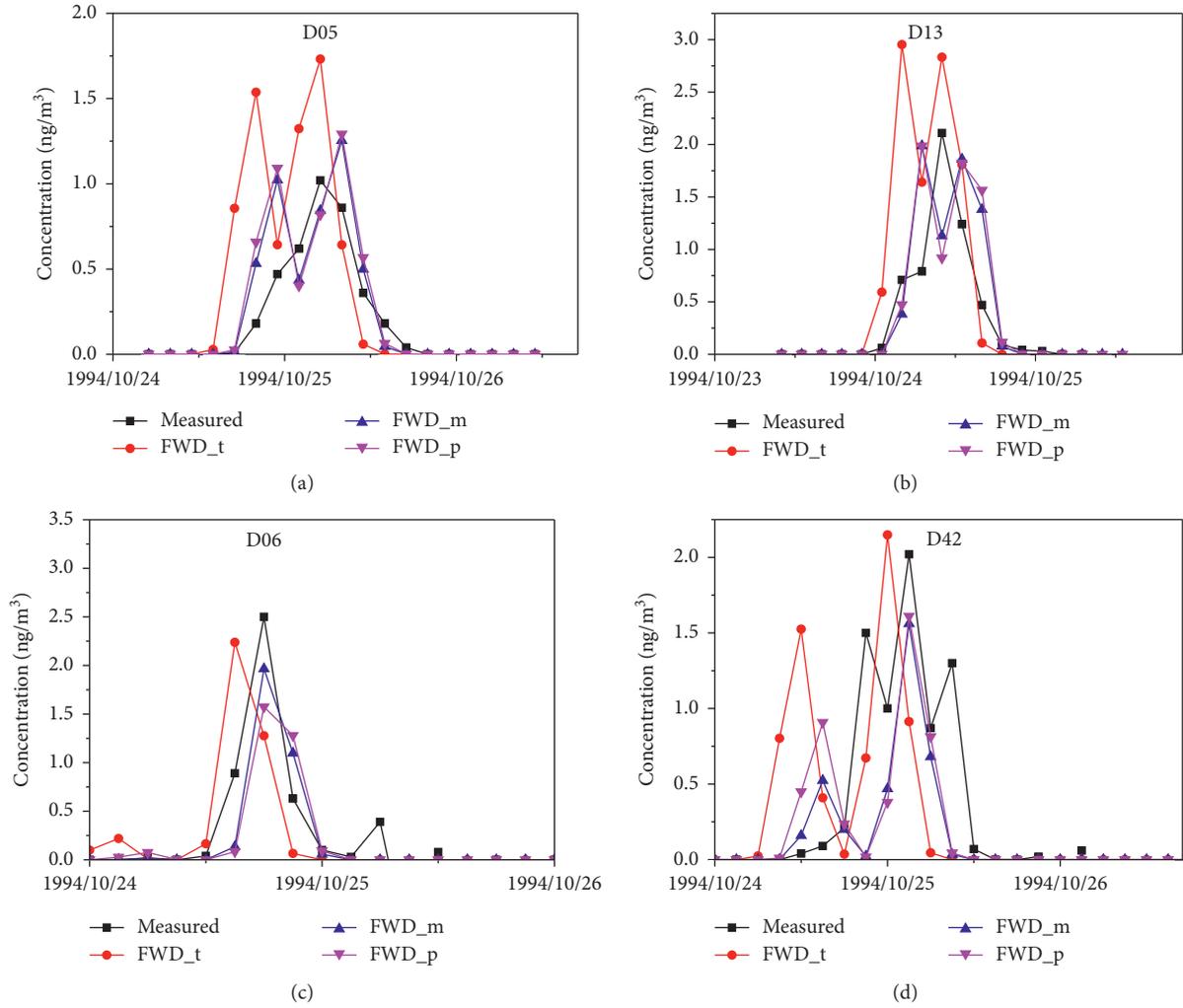


FIGURE 9: Comparison of simulations and measurements at different stations. (a) D04 and (b) D13 are stations used in the source reconstruction. (c) D06 and (d) D42 are stations not used in the source reconstruction.

TABLE 2: Statistical performance indicators for the comparison modeled-measured tracer concentrations. Data at all the 168 stations are used.

| Method | R | FB | FA2 | FA5 | Source |
|--------|------|------|------|------|-----------|
| FWD_t | 0.54 | 0.71 | 0.21 | 0.48 | True |
| FWD_m | 0.28 | 0.12 | 0.27 | 0.56 | Mean |
| FWD_p | 0.30 | 0.33 | 0.26 | 0.56 | Most prob |

6. Conclusion and Discussion

This paper introduces a method to reconstruct the source parameters given a set of measurements. The method combines Bayesian inference (MCMC sampling) and ATM backward simulation and provides the a posteriori probability distributions of the source term. The method has been applied to a tracer experiment (ETEX-I) that occurred in 1994 in Europe. The relationship between the source and receptor is obtained by using the backward run mode of FLEXPART, which reduces the calculation requirement of the likelihood function in a large number of iterative

samplings and improves the calculation efficiency. The reconstructed results are in good agreement with the true parameters of the source term. The position uncertainty was within 1.0°, and the deviation between the release rate and the real rate was within 3 times the standard deviation. After 3000 iterations, the method converges to a steady state, providing not only a single solution but also a posteriori probability distribution. Meanwhile, we find that the validation between the simulation using different source terms and the observation demonstrates that the major uncertainties come from the air dispersion model, including both the meteorological input and the physical process models.

The reduction of these uncertainties can improve the source reconstruction accuracy, which will be addressed in our future work.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] X. Li, H. Li, Y. Liu, W. Xiong, and S. Fang, "Joint release rate estimation and measurement-by-measurement model correction for atmospheric radionuclide emission in nuclear accidents: an application to wind tunnel experiments," *Journal of Hazardous Materials*, vol. 345, pp. 48–62, 2018.
- [2] S. Fang, S. Zhuang, and X. Li, "Automated release rate inversion and plume bias correction for atmospheric radionuclide leaks: a robust and general remediation to imperfect radionuclide transport modeling," *Science of The Total Environment*, vol. 754, Article ID 142140, 2019.
- [3] L. Robertson and J. Langner, "Source function estimate by means of variational data assimilation applied to the ETEX-I tracer experiment," *Atmospheric Environment*, vol. 32, no. 24, pp. 4219–4225, 1998.
- [4] M. Bocquet, "High-resolution reconstruction of a tracer dispersion event: application to ETEX," *Quarterly Journal of the Royal Meteorological Society*, vol. 133, no. 625, pp. 1013–1026, 2007.
- [5] S. Sun, H. Li, and S. Fang, "A forward-backward coupled source term estimation for nuclear power plant accident: a case study of loss of coolant accident scenario," *Annals of Nuclear Energy*, vol. 104, pp. 67–74, 2017.
- [6] X. Li, S. Sun, and X. Hu, "Source inversion of both long- and short-lived radionuclide releases from the Fukushima Daiichi nuclear accident using on-site gamma dose rates," *Journal of Hazardous Materials*, vol. 379, Article ID 120770, 2019.
- [7] E. Yee, I. Hoffman, and K. Ungar, "Bayesian inference for source reconstruction: a real-world application," *International Scholarly Research Notices*, vol. 2014, Article ID 507634, 2014.
- [8] G. Kastner and F. Huber, "Sparse Bayesian vector autoregressions in huge dimensions," *Journal of Forecasting*, vol. 39, no. 7, pp. 1142–1165, 2008.
- [9] J. Sørensen, H. Klein, and M. Ulmoen, "Source localization by inverse methods," 2020, <http://www.nks.org/scripts/getdocument.php?file=111010214696504>.
- [10] L. Monache, J. Lundquist, and B. Kosović, "Bayesian inference and Markov chain Monte Carlo sampling to reconstruct a contaminant source on a continental scale," *Journal of Applied Meteorology and Climatology*, vol. 47, no. 10, pp. 2600–2613, 2008.
- [11] E. Yee, "Automated computational inference engine for Bayesian source reconstruction: application to some detections/non-detections made in the CTBT international monitoring system," *Applied Mathematical Sciences*, vol. 11, no. 32, pp. 1581–1618, 2017.
- [12] D. L. Brazidec, J. Bocquet, and M. Saunier, "Quantification of the modelling uncertainties in atmospheric release source assessment and application to the reconstruction of the autumn 2017 Ruthenium 106 source," *Atmospheric Chemistry and Physics*, vol. 4, 2021.
- [13] P. Seibert and A. Frank, "Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode," *Atmospheric Chemistry and Physics*, vol. 4, no. 1, pp. 51–63, 2004.
- [14] W. R. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC Press, London, UK, 1996.
- [15] A. Gelman, J. B. Carlin, and H. S. Stern, *Bayesian Data Analysis*, CRC Press, Boca Raton, FL, USA, 3rd edition, 2013.
- [16] H. Dop, G. Graziani, and W. Klug, "ETEX: a european tracer experiment," *Large Scale Computations in Air Pollution Modelling*, vol. 57, pp. 137–150, 1999.
- [17] S. Eckhardt, A. Stohl, and H. Sodemann, "The Lagrangian particle dispersion model FLEXPART version 8.0," 2020, <https://folk.nilu.no/%7Eandreas/flexpart/flexpart8.pdf>.
- [18] S. Saha, S. Moorthi, and X. Wu, "NCEP climate forecast system reanalysis (CFSR) 6-hour products," *Computational and Information Systems Laboratory*, vol. 93, 2021.