

Article



# Probabilistic Inverse Method for Source Localization Applied to ETEX and the 2017 Case of Ru-106 including Analyses of Sensitivity to Measurement Data

Kasper Skjold Tølløse <sup>1,2,\*</sup>, Eigil Kaas <sup>1,2</sup>, and Jens Havskov Sørensen <sup>1</sup>

- <sup>1</sup> Danish Meteorological Institute, DK-2100 Copenhagen, Denmark; ek@dmi.dk (E.K.); jhs@dmi.dk (J.H.S.)
- <sup>2</sup> Niels Bohr Institute, University of Copenhagen, DK-2100 Copenhagen, Denmark
- \* Correspondence: ktoe@dmi.dk

**Abstract:** In recent years, cases of unexplained, elevated levels of radioactive particles have demonstrated an increasing need for efficient and robust source localization methods. In this study, a Bayesian method for source localization is developed and applied to two cases. First, the method is validated against the European tracer experiment (ETEX) and then applied to the still unaccounted for release of Ru-106 in the fall of 2017. The ETEX dataset, however, differs significantly from the Ru-106 dataset with regard to time resolution and the distance from the release site to the nearest measurements. Therefore, sensitivity analyses are conducted in order to test the method's sensitivity to these parameters. The analyses show that the resulting source localization depends on both the observed temporal resolution and the existence of sampling stations close to the source. However, the method is robust, in the sense that reducing the amount of information in the dataset merely reduces the accuracy, and hence, none of the results are contradictory. When applied to the Ru-106 case, the results indicate that the Southern Ural region is the most plausible release area, and, as hypothesized by other studies, that the Mayak nuclear facility is the most likely release location.

**Keywords:** source localization; atmospheric dispersion modelling; inverse modelling; Bayesian inference; ETEX; Ru-106

# 1. Introduction

In the case of an accidental release of a substance to the atmosphere, the time and location of the release may be unknown, in which case, only indirect information about the source location is available. If a network of sampling stations has detected the substance, this provides an indication of the geographic location of the plume of the substance at times later than the release time. There is no unambiguous way of calculating the source location from this information; instead, inverse problem theory is needed in order to relate the measurements to possible release scenarios. In the case of an unreported release of harmful particles or gasses, source localization is likely to be an operational task carried out by national emergency management agencies. The aim of this study is to develop and illustrate a method suitable for operational usage, and therefore, the focus is on efficiency and robustness.

Examples of previous work on source localization include different applications of adjoint dispersion modelling, i.e., running the model backwards in time, such as the methods by Pudykiewicz [1], Wotawa et al. [2], Seibert et al. [3,4], and Sørensen [5]. Some studies have combined adjoint dispersion modelling with Bayesian inference and Markov chain Monte Carlo methods, such as the method by Keats et al. [6] and Yee et al. [7,8]. One of the main challenges in inverse dispersion modelling is that model predictions are subject to large unknown errors, which complicate direct comparison of model and measurements. This is demonstrated by Yee et al. [8], who suggest altering the cost function to take this into account by representing the standard deviations with probability distributions allowing for



**Citation:** Tølløse, K.S.; Kaas, E.; Sørensen, J.H. Probabilistic Inverse Method for Source Localization Applied to ETEX and the 2017 Case of Ru-106 including Analyses of Sensitivity to Measurement Data. *Atmosphere* **2021**, *12*, 1567. https:// doi.org/10.3390/atmos12121567

Academic Editors: Irène Korsakissok and Susan J. Leadbetter

Received: 14 October 2021 Accepted: 19 November 2021 Published: 26 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). variation around the estimated uncertainties. Other studies suggest using a cost function that is less sensitive to outliers, i.e., model predictions with large errors, than the typical Gaussian distribution, such as the log-normal distribution [9,10]. As opposed to the previously mentioned studies, Saunier et al. [9] and Le Brazidec et al. [10] both estimate the source term using forward-based methods, utilizing a variational approach and a Markov chain Monte Carlo method, respectively. Another approach, which is also less sensitive to outliers, is to use a correlation-based cost function such as the methods by Efthimiou et al. [11], Kovalets et al. [12,13], and Tomas et al. [14]. An additional advantage of this approach is that evaluation of the probability is possible without a need for explicitly specifying the uncertainties.

A recent case, which has demonstrated the need for better source localization methods, is the release of Ru-106 in the fall of 2017 (see Section 2.2). To date, no release has been reported, but several studies have attempted to estimate the source location [5,9,10,13–15]. These studies are in overall agreement and point towards the Southern Ural region as the most likely release area. Within this geographical area, it has been suggested that the release site could be either the Mayak nuclear facility, cf. http://www.po-mayak.ru/ (accessed on 13 October 2021), or the NIIAR nuclear facility, cf. http://www.niiar.ru (accessed on 13 October 2021).

The methodology developed in this study is inspired by the combination of Bayesian inference and adjoint dispersion modelling used by Keats et al. [6] and Yee et al. [7,8] but with a correlation-based measure for the probability, similar to the method by Tomas et al. [14]. First, the method is validated against the European tracer experiment (ETEX) (see Section 2.1) and then applied to the Ru-106 case (Section 2.2). However, there are a few important differences between the ETEX dataset and the dataset of Ru-106 measurements: first, ETEX consists of a large set of three-hour measurements, whereas most measurements in the Ru-106 case are conducted over either approximately 12 h, 24 h, or one week; this is at least the case for the dataset used in this study, where a number of measurements have been discarded, because they do not meet the quality control requirements defined in Section 2.2. Second, in the ETEX case, there are several measurement stations located less than a few hundred kilometers from the release site, one of which is located directly downstream (see Section 3.1 for a discussion of this). In the Ru-106 case, on the other hand, most measurements that fulfill the quality control requirements are more than 2000 km away from the estimated source. There are a few measurements from locations in Russia close to the estimated source. However, these measurements are conducted over a week or more and therefore contain only limited information about when the plume of Ru-106 passed the measurement stations. To study the effects of the differences between the two datasets, the validation against the ETEX case includes analyses of the sensitivity to these parameters. In addition, the importance of including non-detections is examined, i.e., measurements conducted in the relevant geographical area and period, which have not detected concentrations above the detection limit. Further, there are a few differences that could be mentioned: the tracer gas used in the ETEX experiment was non-depositing and non-decaying, whereas Ru-106 is subject to both radioactive decay (although with a quite long half-life of 371.5 days) as well as dry and wet deposition.

Section 2 describes the data and methodology, Sections 2.1 and 2.2 describe the measurement datasets, Sections 2.3 and 2.4 describe the meteorological data and the dispersion model, while Sections 2.5 and 2.6 describe the methodology. Next, the results are presented and discussed in Section 3, Section 3.1 presents the results of the validation on the ETEX dataset, including tests of sensitivity to the data quality, and Section 3.2 shows the results of applying the methodology to the Ru-106 case. Finally, Section 4 presents a summary and the conclusions of the study.

#### 2. Materials and Methods

# 2.1. ETEX Dataset

The European tracer experiment (ETEX) is a tracer gas experiment designed to test and compare the capabilities of atmospheric transport models to predict long-range atmospheric

dispersion [16,17]. ETEX consisted of two separate experiments, ETEX-1 and ETEX-2, of which only the first is considered in this study. The non-decaying and non-depositing gas perfluoromethylcyclohexane (PMCH) was used as tracer, and a total of 340 kg of the gas was released to the atmosphere with a constant release rate during the release period, starting from 16.00 UTC on 23 October 1994 and lasting 11 h and 50 min. The location of the release ( $48^{\circ}03'30''$  N,  $2^{\circ}00'30''$  W) is close to the village of Monterfil in Brittany, France.

The sampling network consisted of 168 ground-level sampling stations distributed in 17 European countries. The sampling was carried out over 30 three-hour intervals, the first starting at 15.00 UTC on 23 October 1994, i.e., one hour before the release started. The different sampling stations were planned to start sampling about six hours before the expected time of tracer arrival. Therefore, each sampling station has not necessarily measured the concentration in all possible sampling intervals. A total of 3046 valid samples are available, of which 935 are measurements of non-zero concentrations, and the remaining are non-detections, i.e., levels below the detection limit.

As mentioned previously, the sensitivity analyses include a deliberate reduction in the temporal resolution of the ETEX dataset. This is performed by computing averages of four measurements at adjacent time intervals, such that a 12 h measurement interval is obtained. However, there are some measurement intervals, where measurements are lacking, which means that it is not possible to combine the measurements to 12 h averages. These intervals occur randomly throughout the measurement campaign. The measurements that only combine to shorter time intervals are discarded. This approach has resulted in a dataset consisting of a total of 665 12 h measurements, of which 281 are measurements of non-zero concentrations. This is repeated for 24 h intervals, resulting in a dataset consisting of 288 24 h measurements, of which 123 are measurements of non-zero concentrations.

# 2.2. Ru-106 Dataset

During September and October 2017, small concentrations of Ru-106 were detected in high-volume air samples in several European countries by different sampling networks. The concentration levels were below those requiring public protective actions, but the large geographical area affected suggested a release of considerable magnitude [18].

The dataset used in this study is adapted from Masson et al. [18]. The full dataset consists of more than 1000 air concentration measurements from 34 different countries, of which some are non-detections, i.e., levels below detection limits. However, for some of these measurements, information about the sampling period is limited to a start and end date, with no information on the time of the day. For measurements conducted over, e.g., 24 h, this incomplete information gives rise to large uncertainties. For longer measurements, on the other hand, imposing starting and ending time is not as problematic. In this study, all measurements conducted over less than five days where no starting and ending time is given have been discarded. For longer measurements, we assume that the measurement interval starts and ends at 12.00 UTC if the times are not specified. Further, a few measurement periods either start or end outside the simulation period (see Section 2.3), and hence, these measurements have been discarded as well. Finally, the dataset includes a number of non-detections with quite high detection limits compared to the numerical values of some of the non-zero measurements. As described in Section 2.6.1, the non-detections are interpreted as zeros, and therefore, non-detections with high detection limit have also been discarded from the dataset. The maximum accepted detection limit is chosen somewhat arbitrarily to the 5th percentile of all included non-zero measurements, which ensures that all included non-detections are, if not zero, at least small in comparison with the majority of the non-zero measurements. The limits defined in this way are  $0.22 \text{ mBg m}^{-3}$  for measurements conducted over up to 36 h and  $0.0030 \text{ mBg m}^{-3}$ for measurements conducted over more 36 h.

The dataset that remains, after discarding the measurements described above, consists of 583 samples, of which 349 are measurements of non-zero concentrations and the remaining are non-detections. However, approximately half of these measurements are concentration averages over a week or more. It is questionable whether a weekly mean value is useful for source localization, since there is no information about when, during this period, the plume actually passed the measurement station. To address this issue, a source localization is first based on all 583 measurements, and next, a localization is based on two separate datasets: the first only including measurements conducted over up to 36 h, similar to approaches by other studies [5,14], and the second only including measurements conducted over more than 36 h. The purpose is to examine whether the weekly averages contain valuable information, but also whether including them in the dataset introduces a risk of reducing the accuracy of the localization.

# 2.3. Meteorological Data

The simulations have been conducted using meteorological data from the global numerical weather prediction model of the European Centre for Medium-Range Weather Forecasts (ECMWF) [19,20]. For the ETEX case, ERA5 reanalysis data [19] of horizontal resolution  $0.25^{\circ} \times 0.25^{\circ}$  are employed; for the Ru-106 case, cycle CY47R1 data [20] of  $0.1^{\circ} \times 0.1^{\circ}$  resolution. The meteorological fields are available hourly in regular lat-lon grids. The model simulation for the ETEX case starts on 27 October 1994 at 09.00 UTC and runs (backwards in time) until 22 October 1994 at 06.00 UTC. The model simulation for the Ru-106 case starts on 10 October 2017 at 00.00 UTC and runs (also backwards) until 21 September 2017 at 00.00 UTC. The domains used for the simulations are limited area domains: for the ETEX case, the domain limits are 20° W to 30° E and 40° N to 65° N, and for the Ru-106 case, the domain limits are 15° W to 80° E and 30° N to 70° N.

# 2.4. Dispersion Modelling

The atmospheric dispersion is modelled using the Lagrangian model DERMA, the Danish Emergency Response Model of the Atmosphere [21,22]. The model can be represented mathematically with the operator:

$$L(\cdot) = \frac{D}{Dt}(\cdot) - \nabla \cdot (\mathbf{K}\nabla(\cdot)) + \lambda(\cdot), \tag{1}$$

where  $\frac{D}{Dt}(\cdot) = \frac{\partial}{\partial t}(\cdot) + \mathbf{u} \cdot \nabla(\cdot)$  is the total derivative and thus contains both the time derivative and the advection terms,  $\mathbf{u}$  is the three-dimensional wind field,  $\nabla \cdot (\mathbf{K}\nabla(\cdot))$  is the turbulent diffusion term, where  $\mathbf{K}$  is the turbulent diffusion tensor, and  $\lambda$  denotes radioactive decay and deposition. Provided a source function Q, the concentration field C can then be obtained by numerically integrating the differential equation L(C) = Q.

Our method for source localization relies on the use of adjoint dispersion modelling, which implies that the sign of the total derivative is changed and the particles, and thus moves opposite the wind direction. The mathematical operator for the adjoint dispersion model is defined as [1,23]:

$$L^*(\cdot) = -\frac{D}{Dt}(\cdot) - \nabla \cdot (\mathbf{K}\nabla(\cdot)) + \lambda(\cdot).$$
(2)

The adjoint concentration field  $C^*$  can be obtained by applying a source function,  $Q^*$ , and numerically integrating the differential equation  $L^*(C^*) = Q^*$ . Note that only the total derivative changes sign, which can be interpreted as if the particles are moved backwards in time, but the physical processes diffusion, decay, and deposition behave as in the original forward-in-time model.

#### 2.5. Source-Receptor Relationship

Although the aim is to estimate the location of the source, it is necessary to consider additional source term characteristics such as start time, duration, and amount of released material, since these characteristics are closely related to the release location. In this study, only stationary point releases with a constant release rate in a finite time interval are considered. Further, the source is assumed to be located inside the planetary boundary layer (PBL), where the concentration is assumed constant with height and, consequently, only two spatial coordinates are necessary. Thus, all possible sources can be described by the following source term model:

$$\mathbf{m} = (\mathbf{x}_{\text{release}}, t_{\text{start}}, \Delta t_{\text{release}}, q), \tag{3}$$

where  $\mathbf{x}_{\text{release}} = (\phi_{\text{release}}, \lambda_{\text{release}})$  are the horizontal coordinates of the source location, latitude, and longitude, respectively.  $t_{\text{start}}$  is the start time of the proposed release,  $\Delta t_{\text{release}}$  is the duration, and q is the constant release rate.

In order to establish the source–receptor relationship, i.e., a function that relates a proposed source term model **m** to a set of expected detections  $\hat{\mathbf{y}}$  (implying that a set of detections **y** exists), a source function  $Q(\mathbf{m}, \mathbf{x}, t)$  must be related to the proposed source term model **m**:

$$Q(\mathbf{m}, \mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{x}_{\text{release}}) q_t(\mathbf{m}, t),$$

$$q_t(\mathbf{m}, t) = \begin{cases} q & t_{\text{start}} \le t \le t_{\text{start}} + \Delta t_{\text{release}} \\ 0 & \text{otherwise} \end{cases}.$$
(4)

Here, *Q* is the release rate corresponding to the proposed source term model **m** as function of location **x** and time *t*, and  $\delta(\mathbf{x})$  is the Dirac delta function. Assuming that the *i*th measurement  $y_i$  is a point measurement at the location  $\mathbf{x}_i$ , initiated at time  $t_i$ , and of duration  $\Delta t_i$ , the expected detection  $\hat{y}_i$  can be computed as the following inner product:

$$\hat{y}_i = \langle C, h_i \rangle \equiv \int_t \int_V Ch_i \, \mathrm{d}V \, \mathrm{d}t = \frac{1}{\Delta t_i} \int_{t_i}^{t_i + \Delta t_i} C(\mathbf{x}_i) \, \mathrm{d}t, \tag{5}$$

where the filter function  $h_i$  extracts the concentration value *C* at the location and time of the *i*th measurement. *C* is related to  $Q(\mathbf{m}, \mathbf{x}, t)$  as described in Section 2.4. To obtain the last equality in Equation (5), the following definition of the filter function is used:

$$h_{i}(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{x}_{i}) \tau_{i}(t),$$

$$\tau_{i}(t) = \begin{cases} \Delta t_{i}^{-1} & t_{i} \leq t \leq t_{i} + \Delta t_{i} \\ 0 & \text{otherwise} \end{cases}.$$
(6)

Using Equation (5) to estimate the expected detections requires a forward run with the dispersion model for each proposed source term model, **m**. If a large number of realizations of **m** is needed, this approach may be computationally inefficient. Instead, the method developed here relies on the adjoint source–receptor relationship [6–8], which is obtained by using the Lagrangian duality relationship. Only the main result is shown here, and for further details, the reader is referred to [1,6,23]. The Lagrangian duality relationship states:

$$\langle C, Q^* \rangle = \langle C^*, Q \rangle,$$
 (7)

where, as described in Section 2.4, *C* is the concentration field obtained by using the source function Q, and  $C^*$  is the adjoint concentration field obtained by using the source function  $Q^*$ . Equation (7) provides a relation between the value of an adjoint concentration field  $C^*$  originating from a given location and the actual concentration *C* evaluated at the same location. By setting  $Q^* = h_i$  and using Equations (5) and (7), it follows that the expected detections  $\hat{y}_i$  can be computed using the adjoint concentration field:

$$\hat{y}_i = \langle C_i^*, Q \rangle = \int_t \int_V C_i^* Q \, \mathrm{d}V \, \mathrm{d}t, \tag{8}$$

where  $C_i^*$  is the adjoint concentration field obtained by using the filter function  $h_i$  as source function. Thus, when using this adjoint source–receptor relationship, it is sufficient to run

the adjoint dispersion model once per measurement instead of once per realization of **m**. Further, using the definition of the source function, Equation (4), it follows that:

$$\hat{\mathbf{y}} = q\mathbf{X}, \text{ where}$$

$$X_i = X_i(\mathbf{x}_{\text{release}}, t_{\text{start}}, \Delta t_{\text{release}}) \equiv \int_{t_{\text{start}}}^{t_{\text{start}} + \Delta t_{\text{release}}} C_i^*(\mathbf{x}_{\text{release}}) \, \mathrm{d}t.$$
(9)

Thus, the expected detection is proportional to the time integrated adjoint concentration field, and the proportionality constant is the proposed release rate *q*.

#### 2.6. Proposed Method for Direct Marginal Posterior Estimation

Given a set of concentration measurements, y, which contains indirect information about the source term, the probability distribution for the elements of m, Equation (3), can be determined by applying Bayes' theorem:

$$P(\mathbf{m}|\mathbf{y}, I) = \frac{P(\mathbf{m}|I) P(\mathbf{y}|\mathbf{m}, I)}{P(\mathbf{y}|I)},$$
(10)

where  $P(\mathbf{m}|I)$  is the prior probability distribution for  $\mathbf{m}$ ,  $P(\mathbf{y}|\mathbf{m}, I)$  is the likelihood, and  $P(\mathbf{y}|I)$  is a statistical model-independent constant called the evidence. *I* is any background information that may be available about, e.g., source location or amount of released material. To evaluate Equation (10), the quantities  $P(\mathbf{m}|I)$  and  $P(\mathbf{y}|\mathbf{m}, I)$  need to be estimated for a selection of realizations of  $\mathbf{m}$ , and the resulting posterior probability distribution  $P(\mathbf{m}|\mathbf{y}, I)$  can then be estimated by normalizing the distribution. To obtain a good estimate of the probability distribution, it is important to make sure that the relevant parts of this parameter space is sampled. One option is to use Markov chain Monte Carlo methods to sample the parameter space [6–8,10]. This approach is especially useful for sampling very large parameter spaces. However, a few simplifications are described below, which further reduce the dimensionality of the parameter space, such that evaluation of the probability for the entire model grid is computationally feasible. Therefore, there is no need for using Markov chain Monte Carlo methods to sample the probability distribution.

Instead of sampling the posterior distribution for **m**, the idea presented here is to obtain a direct estimate of the marginal posterior distribution for  $\mu = (\mathbf{x}_{release}, t_{start}, \Delta t_{release})$ . First, the posterior probability distribution for **m** may be rewritten as:

$$P(q, \boldsymbol{\mu}|\boldsymbol{y}, I) = \frac{P(q|\boldsymbol{\mu}, I) P(\boldsymbol{\mu}|I) P(\boldsymbol{y}|q, \boldsymbol{\mu}, I)}{P(\boldsymbol{y}|I)}.$$
(11)

Further, it is useful to define the conditional posterior distribution for q, where  $\mu$  is assumed known:

$$P(q|\mathbf{y}, \boldsymbol{\mu}, I) = \frac{P(q|\boldsymbol{\mu}, I) P(\mathbf{y}|q, \boldsymbol{\mu}, I)}{P(\mathbf{y}|\boldsymbol{\mu}, I)}.$$
(12)

The marginal distribution for  $\mu$  can be related to Equations (11) and (12), and by reordering the terms, it can be shown that:

$$P(\boldsymbol{\mu}|\mathbf{y}, I) = \frac{P(q, \boldsymbol{\mu}|\mathbf{y}, I)}{P(q|\mathbf{y}, \boldsymbol{\mu}, I)} = \frac{P(\boldsymbol{\mu}|I) P(\mathbf{y}|\boldsymbol{\mu}, I)}{P(\mathbf{y}|I)}.$$

Focusing on the case where no quantifiable prior information about the source is available, both  $P(\mu|I)$  and  $P(q|\mu, I)$  can be assumed uniform. However, for the prior probability to be uniform for the source location, a factor of  $\cos(\phi_{\text{release}})$  is necessary to compensate for the convergence of longitude lines approaching the poles. The marginal distribution for  $\mu$  is proportional to the evidence of the conditional posterior distribution for q,  $P(\mathbf{y}|\mu, I)$ , which, in this case, reduces to:

$$P(\boldsymbol{\mu}|\mathbf{y}, I) \propto \int \cos(\phi_{\text{release}}) P(\mathbf{y}|q, \boldsymbol{\mu}, I) \, \mathrm{d}q.$$
(13)

#### 2.6.1. Likelihood and Uncertainty Quantification

In order to evaluate Equation (13), one first needs to define the likelihood,  $P(\mathbf{y}|q, \boldsymbol{\mu}, I)$ . The traditional approach is to evaluate the probability of observing  $y_i$  given the model prediction  $\hat{y}_i = qX_i$  and some assumed probability distribution for the residual  $y_i - qX_i$ . For example, one could assume that  $y_i - qX_i$  is Gaussian distributed with variance  $\sigma_i^2$ , such as the methods by Keats et al. and Yee et al. [6–8]. The variance can then be related to the uncertainties associated with both the observation and the model prediction  $\sigma_i^2 = \sigma_{o,i}^2 + \sigma_{m,i'}^2$ , where  $\sigma_{o,i}$  is the observation uncertainty, and  $\sigma_{m,i}$  is the modelling uncertainty. The resulting likelihood is the joint Gaussian distribution:

$$P(\mathbf{y}|q,\boldsymbol{\mu},I) = \frac{1}{\sqrt{(2\pi)^{N}|\mathbf{R}|}} \exp\left[-\frac{1}{2}(\mathbf{y}-q\mathbf{X})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-q\mathbf{X})\right],$$
(14)

where N is the number of measurements,  $\mathbf{R} = \mathbf{O} + \mathbf{M}$  is the error covariance matrix, with **O** and **M** being the error covariance matrices for observation and modelling errors, respectively.  $|\mathbf{R}|$  denotes the determinant of **R**. Assuming that measurements are unbiased, the observation errors can be assumed uncorrelated, and hence **O** is a diagonal matrix with diagonal elements  $\sigma_{\alpha i}^2$  related to the accuracy of the measurement equipment. The elements of **M**, on the other hand, are unknown, and the structure is likely to be complex. There are three main contributions to the uncertainty on the time integrated adjoint concentration  $X_i$ : (1) errors in the estimated trajectories due to uncertain meteorological data, (2) errors in the turbulent diffusion and deposition due to inaccurate parameterizations, and (3) numerical errors. The two first contributions, which we assume are dominant, may cause systematic errors, e.g., due to systematic over- or underestimation of wind speeds or of the turbulent diffusion coefficients. Considering this nature of the modelling errors, it is likely that the errors on predictions that are close in time and space are correlated. However, estimating the off-diagonal elements of **M** is highly non-trivial, and therefore, the best solution might still be to assume that **M** is diagonal; previous studies also assume uncorrelated modelling errors [6-8,10]. Based on these assumptions, the likelihood in Equation (14) can be related to the following cost function:

$$J = \sum_{i}^{N} \left( \frac{y_i - qX_i}{\sigma_i} \right)^2, \tag{15}$$

such that  $P(\mathbf{y}|q, \boldsymbol{\mu}, I) \propto \exp(-J)$ . Although the values of  $\sigma_{o,i}$  may be known, we assume that the errors on the model prediction are generally much larger than the observation uncertainties  $\sigma_{m,i} \gg \sigma_{o,i}$ , and hence  $\sigma_i \approx \sigma_{m,i}$ . Thus, minimizing *J* as defined in Equation (15) requires quantitative estimates of  $\sigma_{m,i}$ . In this study, we use an alternative correlation-based cost function, which allows for evaluation of the likelihood without the need for quantifying uncertainties. As described below, the cost function defined in Equation (15) will then only be used to determine the source strength. As mentioned previously, this approach is inspired by previous methods using correlation-based cost functions [11–14]. However, in contrast to the previous studies, we implement the correlation-based cost function in the Bayesian framework described in the previous section. Our method resembles the method by Tomas et al. [14], where the probability of a given source location is assumed to be proportional to the time integrated Pearson correlation coefficient. In our Bayesian framework, this corresponds to marginalizing over the  $t_{\text{start}}$  dimension of  $P(\boldsymbol{\mu}|\mathbf{y}, I)$ . Thus, the main difference is that the method presented here allows for releases of different durations.

The likelihood is assumed proportional to the reflective correlation coefficient defined as:

$$r(y_i, \hat{y}_i) = \frac{\sum_i y_i \hat{y}_i}{\sqrt{\sum_i y_i^2 \sum_i \hat{y}_i^2}}.$$
(16)

It should be stressed that this assumption is not theory-based, and that correlation cannot, in general, be interpreted as probability. Consequently, the resulting probability

8 of 19

distribution should be interpreted as a rough estimate of the true probability distribution. The use of the reflective correlation coefficient is also different from the previous studies, which use the regular Pearson correlation coefficient [11–14]. The reflective correlation coefficient can be thought of as a non-central version of the Pearson correlation coefficient. Both are normalized measures of the co-variation of two variables, but instead of comparing the variation around the mean values of each variable, the reflective correlation coefficient, *r*, as defined in Equation (16), is not invariant under translation, i.e., adding a constant to all instances of one of the variables will change the value of *r*. Both correlation coefficients are equal to either 1 or -1 for a perfect linear relation; however, the Pearson correlation coefficient assumes a relation of the form y = ax + b, whereas the reflective correlation assumes a relation of the form y = ax. Thus, *r*, as defined in Equation (16), is more suited for the problem considered here. Further, since concentration variables are non-negative, *r* varies between 0 and 1. Therefore, probabilities are naturally ensured to be non-negative as well.

Finally, as mentioned previously, one of the challenges in this problem is direct comparison of a measurement and a model prediction with a large, unknown error. However, the correlation is a measure of how well high and low values follow each other in the two variables. Thus, the exact value of the individual model prediction is less important, and hence, the correlation-based probability is potentially a robust alternative to more traditional approaches. Based on these considerations, it is also reasonable to include non-detections by assuming that these are measurements of zero concentration; as long as the detection limit of the filter station considered is small compared to the numerical values of the non-zero measurements.

Another important feature of the correlation coefficient is the independence of the release rate, q, due to the direct proportionality  $\hat{\mathbf{y}} = q\mathbf{X}$ . This means that the likelihood can be taken outside the integral in Equation (13), and thus, the marginal distribution  $P(\boldsymbol{\mu}|\mathbf{y}, I)$  can be written as:

$$P(\boldsymbol{\mu}|\boldsymbol{y}, I) \propto \cos(\phi_{\text{release}}) \frac{\sum_{i} y_{i} X_{i}}{\sqrt{\sum_{i} y_{i}^{2} \sum_{i} X_{i}^{2}}},$$
(17)

where, as in Equation (9),  $X_i$  is the time integrated adjoint concentration at the location of the proposed release and over the proposed release interval.

Although the likelihood is independent of q, it is still possible to estimate the most likely release rate, denoted  $\hat{q}$ , since this is just the slope of the best linear fit. By minimizing the least-squares cost function defined in Equation (15) with respect to q, one obtains the expression:

$$\hat{q}(\mathbf{\mu}, \mathbf{y}) = \frac{\sum_{i} \sigma_{i}^{-2} y_{i} X_{i}}{\sum_{i} \sigma_{i}^{-2} X_{i}^{2}}$$

Thus, in order to estimate  $\hat{q}$ , one needs to quantify  $\sigma_i$ . As described previously, we assume  $\sigma_i \approx \sigma_{m,i}$ . We have then examined the perhaps three simplest assumptions: (1) assuming  $\sigma_i \propto X_i$  as suggested by Keats et al. [6], (2) assuming  $\sigma_i \propto y_i$ , and (3) assuming  $\sigma_i$  constant, i.e., the uncertainty is the same for all model predictions. The idea behind the two first approaches is that the uncertainty should be of the same magnitude as the predicted concentration itself. The first approach, however, leads to the assumption that predictions of zero concentration has no uncertainty, and since these predictions are not rare, this approach does not work well in practice. The idea behind the third approach is that the model might be better at predicting higher concentrations, and accordingly, one should give greater weights to these terms. We found that the third approach gives the overall best results, while the second approach leads to systematic underestimation of the release rates for the ETEX case. It should be noted that this is not necessarily a general

result; it may depend on the dispersion model used as well as on the case considered. By assuming  $\sigma_i$  constant, we obtain the following simple expression for  $\hat{q}$ :

$$\hat{q}(\boldsymbol{\mu}, \boldsymbol{y}) = \frac{\sum_{i} y_{i} X_{i}}{\sum_{i} X_{i}^{2}}.$$
(18)

# 3. Results and Discussions

Since the elements of  $\mu$  are continuous variables, there are infinitely many possible realizations of  $\mu$  within the limited area domain of the simulation. However, it is assumed that the resolution of the output data is sufficient to properly resolve the main features of the probability distribution. The spatial resolution of the output is the same as the input meteorological data, see Section 2.3, and the concentration fields are output every third hour. Thus, only model grid points are considered as possible source locations, only model output times are considered possible start times, and only durations, which are multiples of the time-resolution of model output are considered possible release periods. In addition, an upper limit for the release duration has been chosen, of 36 h.

As described in the previous section,  $P(\mu|\mathbf{y}, I)$  is a four-dimensional probability distribution for the source location, start time, and duration. When nothing is known about the release time, the best estimate of the source location is obtained by marginalization:

$$P(\mathbf{x}_{\text{release}}|\mathbf{y}, I) = \int_{\Delta t_{\text{release}}} \int_{t_{\text{start}}} P(\boldsymbol{\mu}|\mathbf{y}, I) \, \mathrm{d}t_{\text{start}} \, \mathrm{d}\Delta t_{\text{release}}.$$
 (19)

Another useful concept, which was applied to inverse atmospheric dispersion modelling by Tomas et al. [14], is the highest posterior density region (HDR) [24]. Let *M* be the sample space of  $\mathbf{x}_{\text{release}}$ , in this case, the coordinates of all grid points in the simulation domain. Given the probability density function  $P(\mathbf{x}_{\text{release}}|\mathbf{y}, I)$ , the  $100(1 - \alpha)$ % HDR is defined as the subset  $R(P_{\alpha})$  of the sample space of *M*, such that

$$R(P_{\alpha}) = \{ \mathbf{x}_{\text{release}} | P(\mathbf{x}_{\text{release}} | \mathbf{y}, I) \ge P_{\alpha} \}.$$
(20)

Thus, the  $100(1 - \alpha)$ % HDR is the smallest possible region, in which the probability density integrates to  $100(1 - \alpha)$ %. The HDR can be thought of as a generalization of a confidence interval, where  $R(P_{\alpha})$  is the most credible region for the confidence level  $100(1 - \alpha)$ %. However, due to the assumptions described in Section 2.6, the probability distribution, and therefore the HDRs, should be interpreted with caution. Thus, it cannot a priori be assumed that, e.g., the 10% HDR corresponds to a 10% confidence level. Instead, the results based on the ETEX case can give an indication of how to interpret the different HDRs.

# 3.1. Application to the ETEX Case

First, the methodology is applied to the set of three-hour measurements, both with and without non-detections. The two-dimensional marginal distributions for the release location, shown in Figure 1, are obtained by initially applying Equation (17) and then Equation (19). Figure 1a shows the results based on non-zero measurements only, while Figure 1b shows the results when including non-detections. It should be noted that the plots do not use the same colorbars; instead, the minimum and maximum values on each colorbar are defined by the minimum value of the 90% HDR and the maximum value of the probability densities is considered more relevant than comparing the actual values of the probability densities. Further, Table 1 shows the coordinates for the locations of maximum probability as well as the distances from these locations to the true release site. The table also shows results based on the different datasets used as part of the sensitivity analyses described in Section 3.1.1.



**Figure 1.** Marginal posterior probability distribution for release location for ETEX in units of  $km^{-2}$ . The blue triangle shows the location of highest probability density, the red diamond shows the actual release location, and the yellow circles show the locations of the sampling stations. The results are based on the set of three-hour measurements. (**a**) is based on non-zero measurements only, while (**b**) is the result when including non-detections. A zoom of the area near the maximum probability is shown in Figure 2.

Data Set	Coordinates for Location of Maximum Probability	Distance to True Release Site
3 h measurements excluding non-detections	48.0° N, 1.0° W	75 km
3 h measurements all measurements	$48.0^{\circ}$ N, $1.0^{\circ}$ W	75 km
12 h measurements all measurements	48.0° N, 0.75° W	94 km
24 h measurements all measurements	$48.0^{\circ}$ N, $0.75^{\circ}$ W	94 km
24 h measurements excluding data within 200 km	47.75° N, 2.5° W	50 km
24 h measurements excluding data within 400 km	46.5° N, 0.25° W	218 km
24 h measurements excluding data within 800 km	47.0° N, 1.5° W	123 km

**Table 1.** List of locations of maximum probability for the different versions of the ETEX dataset. The table also shows the distance to the true release site.

Figure 2 shows the same results as in Figure 1, only for a smaller geographical area. Here, it is easier to see that the probability is generally high in the area close to the true release location. Further, comparison of the two figures clearly indicates the added effects of the non-detections: the result in Figure 2b is based on more sampling stations (more yellow circles) than the result in Figure 2a. Associated with several of these added sampling stations, there are areas of lower probability upstream from the measurement station, e.g., the stations north east and south east of the release location.

Finally, it is worth noticing that, in both figures, the sharp peaks of the probability distributions are located near the sampling station in Rennes (the yellow circle directly east of the estimated release location). Examining the dataset, it was discovered that this sampling station measured the highest values in the dataset, which emphasizes the importance of testing the sensitivity to measurements close to the source location.



**Figure 2.** Marginal posterior probability distribution for release location for ETEX in units of  $\text{km}^{-2}$ . The blue triangle shows the location of highest probability density, the red diamond shows the actual release location, and the yellow circles show the locations of the sampling stations. The results are based on the set of three-hour measurements. (a) is based on non-zero measurements only, while (b) is the result when including non-detections. The probability distributions are identical to those shown in Figure 1 but for a smaller geographical area.

#### 3.1.1. Sensitivity Analyses

Now, the sampling periods are deliberately extended to 12 and 24 h, as described in Section 2.1. In both cases, non-detections are included. The resulting probability distributions are shown in Figure 3, and to make comparison easier, the same geographical area as in Figure 2 is used. Thus, Figures 2b and 3a,b show the results based on measurements of 3, 12 and 24 h, respectively. As one might expect, the coarser time resolutions result in wider, less accurate, probability distributions. Nonetheless, all three distributions have the area of maximum probability in the same general area, indicating that even 24 h average concentrations are useful for source localization.



**Figure 3.** Marginal posterior probability distribution for release location for ETEX in units of  $km^{-2}$ . The blue triangle shows the location of highest probability density, the red diamond shows the actual release location, and the yellow circles show the locations of the sampling stations. The results are based on both non-zero measurements and non-detections. (**a**) is based on 12-h measurements, while (**b**) is based on 24-h measurements.

Figure 4 shows the results based on the 24-h measurements, where in Figure 4b–d, measurements within a radius of 200, 400, and 800 km of the release location, respectively,

are excluded from the dataset. The resulting probability distribution strongly depends on the distance from the release site to the nearest measurements: the precision of the source localization gradually decreases as measurements are removed from the dataset, i.e., the peak of the distribution becomes less sharp. The location of maximum probability, on the other hand, is closest to the true release site for the dataset, where measurements within a radius of 200 km are excluded, see Table 1. This may be a coincidence, or it may be explained by the fact that the measurement station in Rennes dominates the localization when these measurements are included.



**Figure 4.** Marginal posterior probability distribution for release location for ETEX in units of  $\text{km}^{-2}$ . The blue triangle shows the location of highest probability density, the red diamond shows the actual release location, and the yellow circles show the locations of the sampling stations. The results are based on the 24 h measurements and include both non-zero measurements and non-detections. (a) is based on all measurements, (b) is the result when excluding measurements within a radius of 200 km from the source, (c) is the result when excluding measurements within a radius of 400 km from the source, and (d) is the result when excluding measurements within a radius of 800 km from the source.

For all four probability distributions, the true source location is located well within the 30% HDR, and only in one case, Figure 4d, the true source is located just outside, almost at border, of the 10% HDR. This indicates that, rather than considering the location of maximum probability, one should consider, e.g., the 10% HDR and combine this with independent information about possible release site candidates, if such exists.

#### 3.1.2. Time and Magnitude of the Release

Determining the most likely release period and the amount of released material can be approached in different ways. One option is to consider the two-dimensional marginal posterior distribution  $P(t_{\text{start}}, \Delta t_{\text{release}} | \mathbf{y}, I)$ . However, the start time and duration vary significantly with the release location considered. Therefore, instead, the two-dimensional conditional posterior distribution  $P(t_{\text{start}}, \Delta t_{\text{release}} | \mathbf{y}, \mathbf{x}_{\text{release}}, I)$  is considered for the most likely release sites. In case no information about possible release sites exist, the best estimate of the source location would be the location of maximum probability. However, if we imagine that ETEX was a nuclear accident, it would most likely be possible to obtain a list of nuclear facilities within the 10% HDR, which could be used to suggest one or a few concrete source locations. Thus, for the ETEX case, the time and magnitude of the release will simply be evaluated for the true release site.

The resulting conditional posterior distribution is shown in Figure 5a using the dataset of three-hour measurements, and in Figure 6a, using the dataset of 24 h measurements, where measurements within a radius of 800 km are excluded. In both figures, the 25, 50, and 75% HDRs are plotted together with the probability densities. Further, based on all release periods within the 50% HDR, the distribution of  $\hat{q}\Delta t_{\text{release}}$ , i.e., the time integrated release, is estimated and shown in Figures 5b and 6b, where the values of  $\hat{q}$  are estimated using Equation (18). Since the amount of released material may vary several orders of magnitude, the logarithm of the time integrated release is considered.

According to the result shown in Figure 5a, the most likely release started on 23 October 1994 at 18.00 UTC and lasted six hours, which is two hours after the true release started and half the duration. However, both the 25% and 50% HDR indicate that releases that start earlier but have a longer duration have almost similar probabilities. The same pattern, only more pronounced, is seen in in Figure 6a, where an even larger area of the parameter space have almost the same probability. This is expected, since this result is based on the dataset with the coarser time resolution and without measurements within the first 800 km. In both cases, the true release is located inside the 50% HDR, and hence, this is interpreted as the best possible estimate of the release period.



**Figure 5.** Time and magnitude of the release based on the true release site. The blue triangles show the most likely release, and the red diamond shows the true release. (a) Marginal posterior probability distribution for start time and duration for ETEX in units of  $h^{-2}$ . The result is based on the set of three-hour measurements. (b) Probability distribution for time integrated release based on the start times and durations within the 50% HDR of the result in (a).



**Figure 6.** Time and magnitude of the release based on the true release site. The blue triangles show the most likely release, and the red diamond shows the true release. (a) Marginal posterior probability distribution for start time and duration for ETEX in units of  $h^{-2}$ . The result is based on the set of 24 h measurements, where measurements within a radius of 800 km are excluded. (b) Probability distribution for time integrated release based on the start times and durations within the 50% HDR of the result in (a).

Surprisingly, the probability distributions for the logarithm of the time integrated release rates, Figures 5b and 6b, indicate that the second dataset, where the data quality has been reduced, gives a better prediction of the magnitude of the release. The probability distribution in Figure 5b predicts a release between 30 and 240 kg, with the most likely release being 90 kg. In comparison, the true release is 340 kg, which is of the same order of magnitude as the largest estimated releases but not included in the predicted interval. The probability distribution in Figure 6b, on the other hand, predicts a release between 90 and 840 kg with the most likely release being 310 kg. As mentioned previously, examination of the dataset showed that the measurement station in Rennes had measured the highest values in the dataset and that this measurement station seems to dominate the resulting probability distribution, see Section 3.1. It is, therefore, possible that the estimate of  $\hat{q}$  in Figure 5b is also dominated by a few measurements close to the source, which may make it less robust than the estimate of  $\hat{q}$  in Figure 6b.

## 3.2. Application to the Ru-106 Case

The two-dimensional marginal distributions for the release location, shown in Figures 7 and 8, are obtained by first applying Equation (17) and then Equation (19). Figure 7 shows the result based on all Ru-106 measurements, while Figure 8 shows the results based on two different sub-datasets: Figure 8a shows the result based on measurements of up to 36 h, of which the majority are either 12-h or 24-h averages, and Figure 8b shows the result based on measurements of more than 36 h, of which the majority are weekly averages (see Section 2.2 for details). Further, Table 2 shows the coordinates for the location of maximum probability as well as the distance from this location to the NIIAR and Mayak nuclear facilities, respectively.

The results in both Figures 7 and 8a show areas of high probability near the Mayak nuclear facility, which is located within the 10% HDR. In both cases, the NIIAR nuclear facility is located only just inside the 50% HDR. Assuming that the HDRs can be interpreted as in the ETEX case, Section 3.1, we would expect the true source location to be located inside or at least close to 10% HDR. Therefore, we conclude that Mayak is the most likely release site.

The 10% HDR of the probability distribution based on long-period measurements only, Figure 8b, extends further to the north and does not contain Mayak. However, Mayak is located just outside the 10% HDR, whereas NIIAR is only included in the 50% HDR. Hence, this result also indicates that the Mayak nuclear facility is the most likely release site.



 $2.2\times10^{-8} \quad 8.0\times10^{-8} \quad 1.4\times10^{-7} \quad 2.0\times10^{-7} \quad 2.5\times10^{-7} \quad 3.1\times10^{-7} \quad 3.7\times10^{-7}$ 

**Figure 7.** Marginal posterior probability distribution for the Ru-106 case in units of km<sup>-2</sup>. The blue triangle shows the location of highest probability density, the red diamond shows the location of the Mayak nuclear facility, the blue square shows the location of the NIIAR nuclear facility, and the yellow circles show the locations of the sampling stations. The result is based on all non-zero measurements and non-detections.



**Figure 8.** Marginal posterior probability distribution for the Ru-106 case in units of  $\text{km}^{-2}$ . The blue triangle shows the location of highest probability density, the red diamond shows the location of the Mayak nuclear facility, the blue square shows the location of the NIIAR nuclear facility, and the yellow circles show the locations of the sampling stations. (a) the result is based on all non-zero measurements and non-detections, both up to 36 h. (b) the result is based on all non-zero measurements and non-detections, both up to 36 h. (b) the result is based on all non-zero measurements and non-detections, both up to 36 h. (b) the result is based on all non-zero measurements and non-detections, both up to 36 h.

**Table 2.** List of locations of maximum probability for the different versions of the Ru-106 dataset. The table also shows the distance to the Mayak and NIIAR nuclear facilities.

Data Set	Coordinates for Location of Maximum Probability	Distance to Mayak	Distance to NIIAR
All measurements	57.9° N, 60.0° E	241 km	771 km
Measurements conducted over up to 36 h	57.8° N, 59.4° E	239 km	735 km
Measurements conducted over more than 36 h	58.4° N, 59.9° E	296 km	794 km

Time and Magnitude of the Release

The release period is estimated by considering the conditional posterior distribution  $P(t_{\text{start}}, \Delta t_{\text{release}} | \mathbf{x}_{\text{release}})$  based on the most likely release site. Further, as described in Section 3.1.2, the amount of released material is estimated by considering the distribution of  $\hat{q}\Delta t_{\text{release}}$  based on all release periods within the 50% HDR. Again, the values of  $\hat{q}$  are estimated using Equation (18). Assuming that the Mayak nuclear facility is the release site, we obtain the conditional posterior distribution  $P(t_{\text{start}}, \Delta t_{\text{release}} | \mathbf{x}_{\text{release}})$  shown in Figure 9.

The result in Figure 9a indicates that the release started between 23 September 2017 around 18.00 UTC and 26 September 2017 at 15.00 UTC and lasted between 3 and 36 h; generally, earlier releases (before September 00.00 UTC) would be of longer duration. In fact, the probability distribution has at least three local maxima: one on 24 September around 00.00 UTC, one on 25 September around 08.00 UTC, and one on 26 September around 02.00 UTC. This may indicate that the release possibly consisted of a few different release periods, which has also been hypothesized by Saunier et al. [9]. However, since the source term model only allows for constant releases within a single time interval, the probability of the combination of two release periods is not considered explicitly, and hence this hypothesis cannot be tested with our method. The 50% HDR covers quite a large area of the parameter space, and it is therefore difficult to estimate the release time and duration more accurately than this.

The probability distribution for the logarithm of the time integrated release rate, Figure 9b, predicts a release between approximately 210 and 1600 TBq, with the most likely release being 620 TBq, which is in accordance with the estimates of other studies [5,9,10,14]. The probability distribution also allows for a release that is significantly larger, up to about 8250 TBq, although this is in unlikely scenario. By comparison with the distribution  $P(t_{\text{start}}, \Delta t_{\text{release}} | \mathbf{x}_{\text{release}})$  in Figure 9a, it is likely that this option of a much larger release is related to the possibility of an early release, i.e., 23–24 September.



**Figure 9.** Time and magnitude of the release based on the location of the Mayak nuclear facility. The blue triangles show the most likely release. (a) Marginal posterior probability distribution for start time and duration for the Ru-106 case in units of  $h^{-2}$ . The result is based on all measurements, i.e., also including measurements conducted over more than 36 h. (b) Probability distribution for time integrated release based on the start times and durations within the 50% HDR of the result in (a).

#### 4. Summary and Conclusions

The methodology combines the adjoint source–receptor relationship with a simplified Bayesian approach and a correlation-based probability measure. This provides an efficient way to obtain an estimate of the four-dimensional probability distribution for latitude, longitude, start time, and duration of a release, based on the assumption that the release rate is constant during the release period. The two-dimensional probability distribution for the source location is then obtained by marginalization. The validation against ETEX shows that the methodology yields an accurate source localization when applied to the entire set of three-hour measurements; in this case, the location of maximum probability is 75 km downstream from the true source location, and the latter is located well within the 10% HDR. When excluding measurements close to the source or extending the measurement intervals, the width of the probability distribution is increased, but the area of high probability, e.g., the 10% or 30% HDR, still contains the true release site. The exact location of maximum probability, on the other hand, depends less systematically on these parameters, cf. Table 1, indicating that this measure is not useful alone but should be interpreted in combination with the 10% HDR. Thus, since the location, a more robust interpretation of the result is to consider a larger area of high probability, such as the 10% HDR, and combine this result with information about possible release sites in the area to be able to suggest concrete source locations.

For a suggested release site, the start time and duration of the release is estimated by considering the conditional posterior distribution. When evaluating the conditional posterior distribution for the true source location in the ETEX case, a rough estimate of the start time and duration of the release is obtained. The methodology does not accurately pinpoint the correct release period but rather gives an interval of possible start times and durations. For the ETEX case, this means that releases of 3 to 36 h duration are possible, with the starting times ranging from 22 October 1994 around 10–15 UTC until approximately 32 h later. Thus, this method may not necessarily be ideal for accurately estimating the release period.

An estimate of the magnitude of the release is obtained by considering the distribution of the logarithm of the time-integrated release. For the ETEX case, the estimate is better when excluding the measurements close to the source. This may be related to the fact that the estimated optimal release rate is dominated by fewer data points with high values, when the measurements close to the source are included. As described in Section 2.6.1, the implication of setting  $\sigma_i$  = constant is that measurements of higher values dominate the estimate of  $\hat{q}$ , which is likely to explain why the magnitude of the release is better estimated when excluding these measurements. However, a more thorough analysis of this is needed in order to reach a conclusion. For the two different versions of the ETEX datasets considered, the source was estimated between 30 and 240 kg and between 90 and 840 kg, respectively. In both cases, the interval of estimated releases includes values of the roughly the same order of magnitude as true release, 340 kg.

Further, the non-detections are included in the dataset by assuming that these are measurements of zero concentration, though they can be non-zero concentrations below the detection limit. The results show that the non-detections do add information, e.g., by locally decreasing the probability in areas upstream from measurement stations that only measured concentrations below the detection limit. However, the result is quite similar to that based on the dataset of only non-zero measurements. This indicates that, for the case considered, non-detections are not crucial for source localization, but they should be included if they exist.

The methodology is then applied to the Ru-106 case using three different datasets: the first including all measurements that fulfill the quality control requirements described in Section 2.2, the second only including measurements conducted over up to 36 h, and the third only including measurements conducted over more than 36 h. The result based on the first two datasets are almost indistinguishable by visual comparison. The third share similar features, but the 10% HDR is shifted towards north and covers a larger geographical area. However, despite these differences, the method proves to be robust, since the three datasets lead to the same overall conclusion. The two-dimensional marginal probability distribution for the location indicates that the Southern Ural region is the most plausible release area. By comparison of the 10% HDR and previously suggested release sites, we conclude that the most likely release site is the Mayak nuclear facility. Based on this assumption, the results indicate that the release started between 23 September 2017

at 18.00 UTC and 26 September 2017 at 15.00 UTC and lasted between 3 and 36 h. Finally, the magnitude of the release is estimated between 210 and 1600 TBq, with the most likely release being 620 TBq.

The main objective of this study was to develop a robust and efficient method suitable for operational use. The simplifications made allow for efficient evaluation of the probability, while the correlation-based probability measure provides a robust way of comparing measurements and model results despite the potentially large, unknown modelling errors. Through various sensitivity analyses, it has been demonstrated that the methodology yields useful results, even when the information contained in the dataset is reduced.

**Author Contributions:** Conceptualization, K.S.T., E.K. and J.H.S.; methodology, K.S.T. and J.H.S.; software, K.S.T.; validation, K.S.T. formal analysis, K.S.T.; investigation, K.S.T.; writing—original draft preparation, K.S.T.; writing—review and editing, K.S.T., E.K. and J.H.S.; visualization, K.S.T.; supervision, E.K. and J.H.S.; funding acquisition, K.S.T., E.K. and J.H.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Innovation Fund Denmark grant number 0196-00017B and by the Danish Meteorological Institute.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** The ETEX observation dataset is published by Nodop et al. [17]. The Ru-106 observation dataset is described and made publicly available in the work of Masson et al. [18]. The ERA5 dataset is available via the https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5 (accessed on 13 October 2021).

Conflicts of Interest: The authors declare no conflict of interest.

# Abbreviations

The following abbreviations are used in this manuscript:

DERMA	Danish Emergency Response Model of the Atmophere
PBL	Planetary Boundary Layer
ETEX	European Tracer Experiment
ECMWF	European Centre for Medium-Range Weather Forecasts
PMCH	Perfluoromethylcyclohexane
HDR	High Density Region

#### References

- 1. Pudykiewicz, J.A. Application of adjoint tracer transport equations for evaluating source parameters. *Atmos. Environ.* **1998**, 32, 3039–3050. [CrossRef]
- Wotawa, G.; De Geer, L.E.; Denier, P.; Kalinowski, M.; Toivonen, H.; D'Amours, R.; Desiato, F.; Issartel, J.P.; Langer, M.; Seibert, P.; et al. Atmospheric transport modelling in support of CTBT verification—Overview and basic concepts. *Atmos. Environ.* 2003, 37, 2529–2537. [CrossRef]
- Seibert, P. Methods for source determination in the context of the CTBT radionuclide monitoring system. In Proceedings of the Informal Workshop on Meteorological Modelling in Support of CTBT Verification, Vienna, Austria, 4–6 December 2000; pp. 1–6.
- Seibert, P.; Frank, A.; Kromp-Kolb, H. Inverse modelling of atmospheric trace substances on the regional scale with Lagrangian models. In Proceedings of the EUROTRAC-2 Symposium, Garmisch-Partenkirchen, Germany, 11–15 March 2002; pp. 11–15.
- Sørensen, J.H. Method for source localization proposed and applied to the October 2017 case of atmospheric dispersion of Ru-106. J. Environ. Radioact. 2018, 189, 221–226. [CrossRef] [PubMed]
- 6. Keats, A.; Yee, E.; Lien, F.S. Bayesian inference for source determination with applications to a complex urban environment. *Atmos. Environ.* **2007**, *41*, 465–479. [CrossRef]
- 7. Yee, E.; Lien, F.S.; Keats, A.; D'Amours, R. Bayesian inversion of concentration data: Source reconstruction in the adjoint representation of atmospheric diffusion. *J. Wind Eng. Ind. Aerodyn.* **2008**, *96*, 1805–1816. [CrossRef]
- 8. Yee, E.; Hoffman, I.; Ungar, K. Bayesian Inference for Source Reconstruction: A Real-World Application. *Int. Sch. Res. Not.* 2014, 2014, 12. [CrossRef]
- 9. Saunier, O.; Didier, D.; Mathieu, A.; Masson, O.; Le Brazidec, J.D. Atmospheric modeling and source reconstruction of radioactive ruthenium from an undeclared major release in 2017. *Proc. Natl. Acad. Sci. USA* 2019, *116*, 24991–25000. [CrossRef] [PubMed]

- 10. Le Brazidec, J.D.; Bocquet, M.; Saunier, O.; Roustan, Y. MCMC methods applied to the reconstruction of the autumn 2017 Ruthenium-106 atmospheric contamination source. *Atmos. Environ. X* **2020**, *6*, 100071. [CrossRef]
- 11. Efthimiou, G.C.; Kovalets, I.V.; Venetsanos, A.; Andronopoulos, S.; Argyropoulos, C.D.; 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.* **2017**, *170*, 118–129. [CrossRef]
- 12. Kovalets, I.V.; Efthimiou, G.C.; Andronopoulos, S.; Venetsanos, A.G.; Argyropoulos, C.D.; Kakosimos, K.E. Inverse identification of unknown finite-duration air pollutant release from a point source in urban environment. *Atmos. Environ.* **2018**, *181*, 82–96. [CrossRef]
- 13. Kovalets, I.V.; Romanenko, O.; Synkevych, R. Adaptation of the RODOS system for analysis of possible sources of Ru-106 detected in 2017. *J. Environ. Radioact.* 2020, 220, 106302. [CrossRef] [PubMed]
- 14. Tomas, J.M.; Peereboom, V.; Kloosterman, A.; van Dijk, A. Detection of radioactivity of unknown origin: Protective actions based on inverse modelling. *J. Environ. Radioact.* **2021**, 235, 106643. [CrossRef] [PubMed]
- Sørensen, J.H.; Klein, H.; Ulimoen, M.; Robertson, L.; Pehrsson, J.; Lauritzen, B.; Bohr, D.; Hac-Heimburg, A.; Israelson, C.; Buhr, A.M.B.; et al. NKS-430: Source Localization by Inverse Methods (SLIM); Technical Report; Nordic Nuclear Safety Research (NKS): Roskilde, Denmark, 2020.
- 16. Graziani, G.; Klug, W.; Mosca, S. *Real-Time Long-Range Dispersion Model Evaluation of the ETEX First Release*; Office for Official Publications of the European Communities: Luxembourg, 1998.
- 17. Nodop, K.; Connolly, R.; Girardi, F. The field campaigns of the European Tracer Experiment (ETEX): Overview and results. *Atmos. Environ.* **1998**, *32*, 4095–4108. [CrossRef]
- Masson, O.; Steinhauser, G.; Zok, D.; Saunier, O.; Angelov, H.; Babić, D.; Bečková, V.; Bieringer, J.; Bruggeman, M.; Burbidge, C.; et al. Airborne concentrations and chemical considerations of radioactive ruthenium from an undeclared major nuclear release in 2017. Proc. Natl. Acad. Sci. USA 2019, 116, 16750–16759. [CrossRef] [PubMed]
- 19. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]
- 20. PartIII: Dynamics and Numerical Procedures. In IFS Documentation CY47R1; IFS Documentation, ECMWF: Reading, UK, 2020.
- Sørensen, J.H. Sensitivity of the DERMA long-range Gaussian dispersion model to meteorological input and diffusion parameters. *Atmos. Environ.* 1998, 32, 4195–4206. [CrossRef]
- Sørensen, J.H.; Baklanov, A.; Hoe, S. The Danish emergency response model of the atmosphere (DERMA). *J. Environ. Radioact.* 2007, 96, 122–129. [CrossRef] [PubMed]
- 23. Marchuk, G.; Shutyaev, V.; Bocharov, G. Adjoint equations and analysis of complex systems: Application to virus infection modelling. *J. Comput. Appl. Math.* 2005, 184, 177–204. [CrossRef]
- 24. Hyndman, R.J. Computing and graphing highest density regions. *Am. Stat.* **1996**, *50*, 120–126.