MCMC methods applied to the reconstruction of the autumn 2017 ruthenium 106 atmospheric contamination source

Joffrey Dumont Le Brazidec, Marc Bocquet, Olivier Saunier, Yelva Roustan

ATMOSPHERIC ENVIRONMENT: X 6 (2020) 100071

Contents lists available at ScienceDirect

Atmospheric Environment: X

journal homepage: http://www.journals.elsevier.com/atmospheric-environment-x

MCMC methods applied to the reconstruction of the autumn 2017 Ruthenium-106 atmospheric contamination source

Joffrey Dumont Le Brazidec [a,b,*], Marc Bocquet [b], Olivier Saunier [a], Yelva Roustan [b]

[a] IRSN, PSE-SANTE, SESUC, BMCA, Fonstenay-aux-Roses, France
[b] CEREA, Joint Laboratory École des Ponts ParisTech and EDF R&D, Université Paris-Est, Marne-la-Vallée, France

ARTICLE INFO

Keywords:
Inverse problems
Bayesian inference
MCMC methods
Source term assessment
Ruthenium-106
Parallel tempering

ABSTRACT

In autumn 2017, small amounts of Ruthenium-106 of unknown origin were detected in Europe by several independent monitoring networks. To study the dispersion of this radionuclide, inverse modelling methods are applied to retrieve the location, time, duration and magnitude of the source. The inverse problem is solved within the Bayesian framework to yield a full reconstruction of the uncertainties.

We first develop a classical Markov chain Monte Carlo (MCMC) method - the Metropolis Hastings (MH) algorithm - to reconstruct the source. However, the algorithm fails to converge in acceptable time because of local minima of the cost function, whose existence is due to a poor distribution of the observations.

To overcome this obstacle, the parallel tempering algorithm, an enhanced MCMC method, is assessed and applied to the Ruthenium dispersion event. The convergence of the algorithm is studied and keys to implement and accelerate it are provided. The probability distribution functions of the variables associated with the source and the observation errors are obtained: they point to an area in south Ural and a total release of several hundreds of TBq on the 26th of September 2017. The results are compared and proven to be consistent with other estimates. Moreover, the method converges fast enough to make it suitable for operational use.

1. Introduction

1.1. The source term estimation problem

The assessment of air pollutant emissions, for instance of radiological nature, is a topic of great importance in atmospheric sciences. The estimation of substance emission is directly linked to the problem of source term estimation (STE) which can be tackled through two approaches. The first method, called a bottom-up approach (Tichý et al., 2016), relies mainly on building an inventory of the releases with the help of a physical or statistical model describing the emission process. The second method, known as a top-down approach (Nisbet and Weiss, 2010), based on inverse modelling methods, consists in combining environmental measurements and an atmospheric transport model. In the case of an accidental radionuclide release into the atmosphere, information about the release rate and its time evolution, the release altitude and sometimes the source location can be incomplete. Consequently for these cases, classical approaches lie within the framework of inverse modelling for source term determination.

1.2. Inverse modelling of the source term

Inverse modelling is a challenging exercise subject to significant uncertainties (Abida and Bocquet, 2009) which can been handled in different ways. Variational approaches (Winiarek et al., 2012; Saunier et al., 2013; Bocquet, 2012) consist in estimating the optimal source term by minimising a cost function measuring differences between observed and simulated measurements using gradient-based descent optimisation tools (Saunier et al., 2019) or genetic algorithms (Schmehl et al., 2012). Machine learning algorithms and hypercube sampling are used in the work by Lucas et al. (2017) to retrieve the source term of a trace gas released from a coastal California nuclear power plant. Another class of inverse modelling methods is based on the maximum entropy on the mean principle (Bocquet, 2005), where information prior to the reconstruction can be adjusted. Nowadays, more and more inverse methods are set within the framework of probabilities and especially Bayesian inference, in which uncertainties in the source reconstruction can be quantified and priors for the source can be described. Indeed, due to inherent noises acquired in the process of environmental measures, or the nonlinear nature of the physical mechanisms controlling the

* Corresponding author. IRSN, 31 Avenue de la Division Leclerc, 92260, Fontenay-aux Roses, France.
E-mail address: joffrey.dumontlebrazidec@irsn.fr (J. Dumont Le Brazidec).

https://doi.org/10.1016/j.aeaoa.2020.100071
Received 23 July 2019; Received in revised form 4 March 2020; Accepted 8 March 2020
Available online 20 March 2020

meteorology, the STE problem is subject to many errors and a quantification of the uncertainties is essential.

1.3. Bayesian inverse problem

Several Bayesian methods have been developed within the context of STE problems: Rajaona et al. (2015) develop an adaptive scheme based on importance sampling (IS), enhancing its estimation process by learning from the previous sampling rounds. Their method is tested afterwards on the Fusion Field Trials 2007 experiment. Tichý et al. (2016) propose an iterative variational Bayes method with no need of tuning and, as a result, suitable for operational use. The algorithm is then tested using real data from the European Tracer Experiment (ETEX). However, the most common ways to estimate the posterior distributions of the source's variables is through the use of Markov chain Monte Carlo (MCMC) algorithms. These methods are described by Delle Monache et al. (2008) who apply MCMC techniques for a stochastic estimation of the Algeciras incident source location. In Keats et al. (2007), the source parameters of a complex urban environment are sampled using the Metropolis-Hastings and hybrid Hamiltonian algorithms. Yee et al. (2014) reconstruct the location and the emission rate of a medical isotope production facility using a mutiple-try differential evolution adaptive Metropolis algorithm with an archive of past states. Liu et al. (2017) compare several Bayesian methods, including MCMC methods, on the Chernobyl and Fukushima Daiichi accidents to assess the source term and the associated uncertainties. MCMC techniques are also applied by Senocak et al. (2008) on both real and synthetic problems, and by Hazart et al. (2014) on several simulated cases. Chow et al. (2008) use stochastic sampling to extract source variables in urban environments.

In this paper, to reconstruct the distributions associated with each variable describing a source and the corresponding uncertainties, we explore and develop advanced Bayesian sampling methods based on Monte-Carlo Markov Chains and apply them on a real accidental release case.

1.4. Objectives of this study

The objective of this study is to apply a sophisticated MCMC method to the reconstruction of the source of unknown origin of the Ruthenium-106 detected in Europe in the autumn of 2017. Uncertainties are integrated into the MCMC modelling.

In Section 2, we provide context, describing recent works and findings about the ^{106}Ru detection event studied in this paper. Furthermore, the observation dataset, meteorological fields, and atmospheric dispersion models are presented and discussed. In Section 3, the Bayesian framework is introduced and a full probabilistic description of the problem is given. Notably, discussions on the choice of the distributions are provided. Thereafter, a classic MCMC method - the MH algorithm - is presented and applied to retrieve the longitude, the latitude, the magnitude and the time of the release. The failure of the algorithm to converge within an acceptable time period when the location is assumed to be unknown is studied. In Section 4, we present and use the parallel tempering (PT) algorithm, an advanced MCMC method, to deal with the problems encountered while using the MH technique. Discussions to optimise this method and ensure its convergence are provided. In particular, the choice for a gamma prior on the release rates is explained. The distributions of the coordinates, the time-dependent release of the ^{106}Ru source and the observation error covariance matrix are sampled using the PT algorithm. Subsequently, the pdfs are analysed and the main outcomes are discussed. Finally, a summary and conclusions are given in Section 5.

2. Airborne radioactivity increase in Europe in autumn 2017

In this section, we give a complete description of the event of ^{106}Ru detection in Europe. A review of previous works is provided. The observation dataset, atmospheric dispersion models and meteorological data used in this study are also described.

2.1. Context

During the months of September and October 2017, small amounts of ^{106}Ru were detected in Europe by several networks involved in the monitoring of atmospheric radioactive contamination. Although concentration levels of a few milliBecquerels per cubic meter of air were too low to pose any human health or environmental issue, the widespread detection suggested that the source term must have been quite high. More precisely, concentration levels range from several microBecquerels per cubic meter of air to more than a hundred milliBecquerel per cubic meter of air (Masson et al., 2019).

The location of the source being unknown, investigations based on measurement results and inverse modelling were carried out to retrieve the release zone and to estimate the quantity of Ruthenium released, as well as the time and duration of the release.

2.2. Recent works

Several methods have been used to estimate the ^{106}Ru source. Kovalets and Romanenko (2017) propose to retrieve the location of the source by computing correlations between the observations and a produced corresponding set of predictions. Saunier et al. (2019) apply inverse modelling methods to retrieve the ^{106}Ru source using variational techniques and identify the southern Ural as the most likely geographical location of the accident for a total release of 250 TBq. The source is reconstructed in the work of Shershakov et al. (2019) by simulating the time-reversed transition of tracers from virtual sources located at the measurement sites where the most probable location is assessed in the South or Central Urals. Sørensen (2018) applies inverse modelling techniques to retrieve the source and the time of the release.

2.3. Observation dataset

The full observation dataset ($>$ 1000 air concentration measurements of ^{106}Ru in September and October 2017) is described in the work of Masson et al. (2019). ^{106}Ru was observed at levels between a few microBecquerel per cubic meter of air to more than 170 milliBecquerel per cubic meter of air in Romania, where it achieved its maximum. Only stations located further west in Europe (Portugal, Spain, and Ireland) did not report detections of ^{106}Ru. In total, 31 countries on the European continent have reported concentration levels of ^{106}Ru (Saunier et al., 2019).

Note that the air sampling period is not consistent over the different countries. It ranged between 11 h (in Romania) to 3 months (in Russia). There is also a considerable variability in the detection limit of the measurements; it is very low in France (under 1 μBq/m3) but higher in Germany or in Romania.

Fig. 1 indicates the maximum air concentration values observed in Europe (Masson et al., 2019). Note that besides air concentration observations, measurements of deposition have been collected in several European countries. Nevertheless, given the small number of deposition observations compared to the number of air concentration observations, we will not take the deposition measurements into account in this study.

2.4. Modelling

The atmospheric dispersion of the 106Ru plume is simulated using the Eulerian ldX model, a part of IRSNs C3X operational platform (Tombette et al., 2014), validated on the Algeciras incident, as well as on the Chernobyl accident (Quélo et al., 2007). The main features are set up based on previous work (Saunier et al., 2019) and are recalled in Table 1. To conduct simulations in this study, we will use meteorological

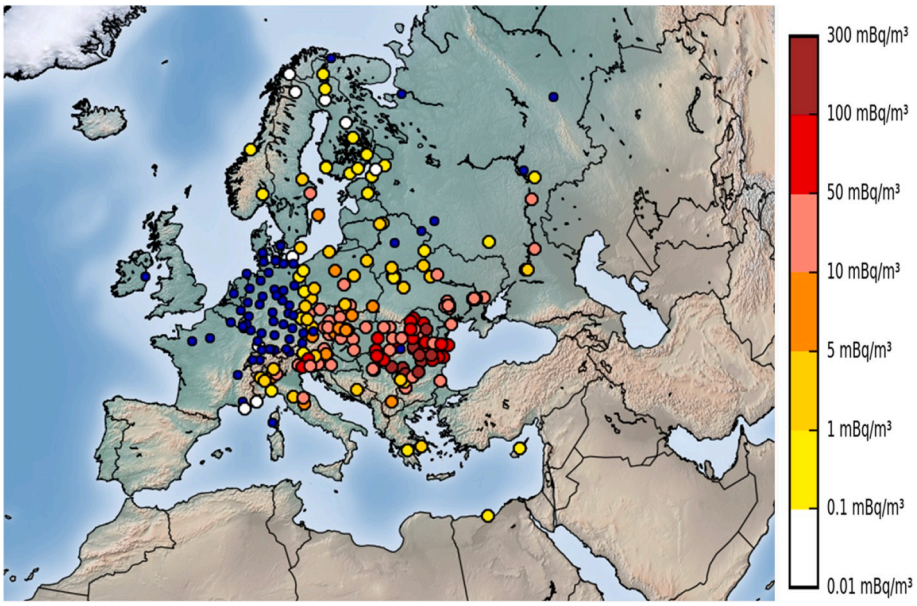

Fig. 1. Maximum air concentrations of ^{106}Ru observed over Europe in mBq.m^{-3}. Blue points measured concentrations below the detection limit. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

data from two sources: ECMWF (European Centre for Medium-Range Weather Forecasts) ERA5 HRES reanalysis of the global weather (Hersbach et al., 2019) and Météo-France global reanalysis from the numerical ARPEGE model (Pailleux et al., 2000).

The model simulations start on the 22nd of September, 2017 at 00.00 UTC and run until the 13th of October, 2017 (time of the last observation). We assume that the release occurred at ground-level. The domain will be restricted to a domain extending from the west of Europe to the east of Kazakhstan. This follows previous work by two of the authors, which excluded the possibility of the source being located anywhere else (Saunier et al., 2019).

Inverse modelling involves the comparison of the observation dataset with a corresponding set of predictions produced by the atmospheric dispersion model given a characterised release. As explained in more detail by, e.g. Winiarek et al. (2011), the observation equation

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \varepsilon \tag{1}$$

expresses this relationship, where **H** is the observation operator, **x** is the vector of variables describing the source of the release (evolution of the release, nature of the radionuclides, location, etc), and ε is the observation error. In this way, **Hx** represents the atmospheric dispersion of a release described by **x**. More precisely, the product **Hx** yields the vector of air concentration predictions for the atmospheric dispersion model, given a source **x** to compare to the available set of observations **y**. In our study, all parameters describing the source (nuclide composition, height of the release, etc) are assumed to be known, except for the source term **q** and the coordinates x_1, x_2 of the source. The computed observation operator **H** scales linearly with the release rate ($\mathbf{H}(x_1, x_2, \mathbf{q}) = \mathbf{H}(x_1, x_2, 1)\mathbf{q}$) and, in this way, can be computed independently from **q**. However, there is no particular relationship between the observation operators of two different locations, which means that the observation operator is dependent on the location of the source. Consequently, to save computation time, the spatial domain is subdivided into regular meshes of resolution 2° × 2°, where a set of observation operators $\mathbf{H}_k$ will be computed at the nodes, which correspond to the black points in Fig. 2. In this way, the observation operator $\mathbf{H}(x_1, x_2, 1)$ of the product $\mathbf{H}(x_1, x_2, 1)\mathbf{q}$ will be determined as a bilinear interpolation of these pre-computed $\mathbf{H}_k$.

Table 1
Main configuration features of the simulations for the ^{106}Ru detection event. Domain extent and parameters relative to the model dispersion are the same for both meteorological datasets.

^{106}Ru detection event (Autumn 2017)		
	ARPEGE Météo-France	ECMWF ERA5
Computational domain extent	[6W; 70E] and [34N; 68N]	
Spatial resolution	0.5° × 0.5°	0.28125° × 0.28125°
Vertical resolution	11 terrain following layers (from 0 to 4400 m)	15 t.f.l. (from 0 to 8000 m)
Time resolution	3 h	1 h
Vertical mixing	K-diffusion following the parametrisation of Louis' closure (Louis, 1979) and (Troen and Mahrt, 1986) in unstable conditions in the PBL.	
Horizontal mixing	Constant horizontal eddy diffusion coefficient	
Wet scavenging	Groëll et al. (2014)	
Dry deposition	constant deposition velocity $v_d = 2.10^{-3}$ m s^{-1}	

3. Bayesian inverse problem

Bayes' rule, with **x** being the set of variables characterising the source and **y** the observations vector, a set of air concentrations, is written as

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \tag{2}$$

The objective is to determine the source variable distribution according to the observations and a prior: for this we will use the Bayesian formula described above and a class of methods for obtaining a sequence of random samples from a probability distribution: the Markov chain Monte Carlo (MCMC) algorithms. Other sampling methods could be used, such as Randomise-Then-Optimise method (Bardsley et al., 2014; Liu et al., 2017), an importance sampling algorithm (Liu et al., 2017) or hypercube sampling (Lucas et al., 2017).

More precisely, we want to portray the joint distribution of the variables describing the source: the coordinates (x_1, x_2), the time-dependent release rate, which will be described as a vector **q** of the ^{106}Ru source, and a term describing the uncertainties **R** described in the following.

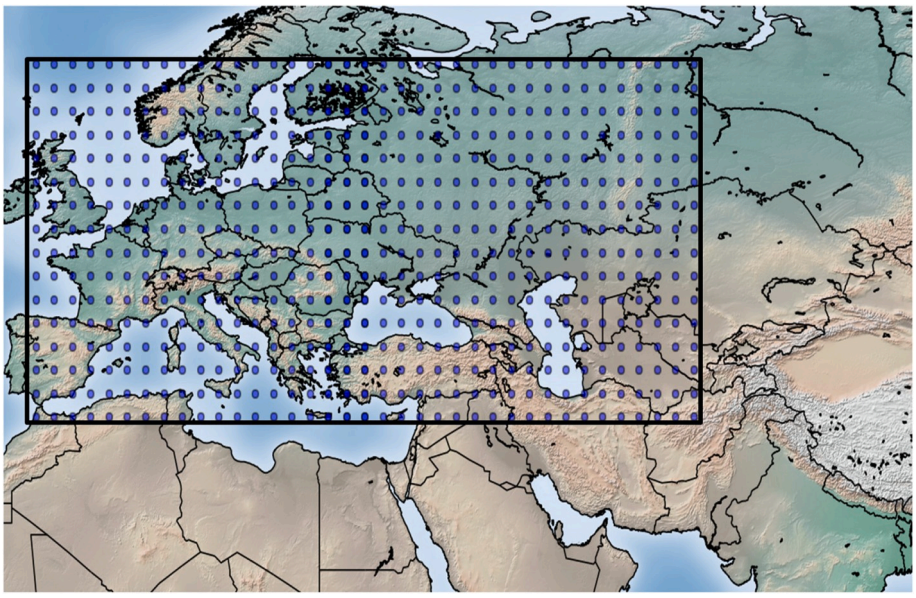

Fig. 2. The black box corresponds to the domain of research of the source of the Ruthenium. For each blue point of coordinates (x_1, x_2), a corresponding observation operator $\mathbf{H}(x_1, x_2, 1)$ is computed. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

In this section, we will first describe the different probability distributions of our problem and the rationale behind theses choices. Then, we will introduce a basic MCMC approach: the Metropolis Hastings algorithm and the associated proposal distributions before applying them to the ^{106}Ru detection event.

3.1. Probabilistic description of the problem

3.1.1. Likelihood

The likelihood quantifies the probability of the observations given a source. More precisely in our case, the likelihood quantifies the difference between the N_{obs} measurement vector $\mathbf{y}$ and a corresponding set of modelled concentrations from the source $\mathbf{x}$ multiplied by the observation operator $\mathbf{H}$.

Most authors rely on a Gaussian description of the likelihood (Tichý et al., 2016; Rajaona et al., 2015; Yee, 2008) or a Gaussian-based description (Yee et al., 2014). However, as in Liu et al. (2017), we choose log-normal statistics to quantify the observations - predictions difference, for the reason that the ability of the log-normal statistics to deal with observations of significantly different magnitude is a major advantage.

In this case, the likelihood is written as a multivariate log-normal distribution of density

$$p(\mathbf{y}|\mathbf{x}) = \frac{e^{-\frac{1}{2}\ln\left(\frac{\mathbf{y}}{\mathbf{Hx}}\right)^{\top}\mathbf{R}^{-1}\ln\left(\frac{\mathbf{y}}{\mathbf{Hx}}\right)}}{\sqrt{(2\pi)^{N_{obs}}|\mathbf{R}|^{1/2}\prod_{i=1}^{N_{obs}} y_i}}, \tag{3}$$

where ln stands for the natural logarithm, $\ln\left(\frac{\mathbf{y}}{c_{ref}}\right) = \left[\ln\left(\frac{y_1}{c_{ref}}\right), \ldots, \ln\left(\frac{y_{N_{obs}}}{c_{ref}}\right)\right]$ with $c_{ref} = 1\text{mBq}\cdot\text{m}^{-3}$ and where $\mathbf{R}$ is the observation error covariance matrix usually defined for simplicity as $\mathbf{R} = r\mathbf{I}$, $r > 0$ as in Winiarek et al. (2011). While unrealistic, such a representation of the observation error covariance matrix gives a rather good approximation of its magnitude. Note that $|\mathbf{R}|$ stands for the determinant of $\mathbf{R}$.

Among the data, there are many null observations (observations under the detection limit). Those are useful data which lie outside the logarithm's domain of definition. To fix this problem, we have chosen to set a threshold y_t and a resulting function ξ_t that is:

$$\xi_t(\mathbf{y}) = \ln\left(\frac{\mathbf{y} + \mathbf{y_t}}{c_{ref}}\right), \tag{4}$$

where $\mathbf{y_t}$ is simply a constant vector of length N_{obs}: $\mathbf{y_t} = y_t \mathbf{1}_{N_{obs}}$. ξ_t is used instead of the logarithm function in the definition of the log-normal distribution (Liu et al., 2017). To set y_t, we need to understand its role. A very small value of y_t would give a great deal of importance to the null observations: more specifically, maximising the likelihood would be mainly equivalent to minimising the errors between the null observations and the corresponding predictions (or the null predictions and the corresponding observations). More generally, the smaller the value of y_t is, the greater the importance is given to the null or very small observations. However, a too high-valued y_t would conversely decrease the importance of the null observations: we need to find a satisfying balance. We set $y_t = 0.5\ \text{mBq.m}^{-3}$. The motivation of this choice can be found in Appendix B.

3.1.2. Variables prior

The ^{106}Ru source $\mathbf{x}$ will be described by a set of variables, the missing information: the longitude x_1, the latitude x_2, the logarithm of the time-dependent release rate and described by its time-independent components $\mathbf{q} = (q_1, \ldots, q_N)$ defined more precisely in the following, and a term representing the uncertainties. In what follows, we will set the priors for these different variables separately, as they are considered independent.

First, the prior for the location of the source is described by a uniform distribution in the domain defined in Table 1, since its origin is assumed to be completely unknown:

$$x_1,\ x_2 \sim \mathscr{U}(-6^{\circ}, 70^{\circ}) \times \mathscr{U}(34^{\circ}, 68^{\circ}), \tag{5}$$

where $x \sim \mathscr{U}$ means that x is drawn from the uniform distribution. Another possibility is to set a multimodal Gaussian distribution corresponding to the known nuclear plants in a given area. Nevertheless, this would be a very faulty prior in case of undeclared (non-referenced) nuclear activity.

With regard to the releases, and as mentioned above, we decided to model the time-dependent true release within the Bayesian framework as a vector $\mathbf{q}$ of N time-independent release rates. As discussed in the work of Tichý et al. (2016), release rates may be dependent on one another but we consider them as independent random variables. Since the sampling period of the air concentration observations of Ruthenium

ranges from 11 h to several days for most of the measurements, and for the sake of quick computation, the vector of release rates components $\mathbf{q}$ is composed of daily release rates; i.e., the release on a day is considered as constant and described by a component of $\mathbf{q}$. Consequently, $\mathbf{q}$ is described as a vector of size N_{imp}, the number of its daily constant release rates.

The prior for the release rate vector acts as a regularisation term that fixes ill-defined STE problems. Different distributions can be considered, such as the truncated Gaussian distribution (Winiarek et al., 2012; Tichý et al., 2016), or the log-normal distribution (Liu et al., 2017), which are used to enforce the positivity of the source term within the Bayesian framework. Indeed, log-defined statistics are very appropriate to deal with the large range of values of pollutant concentrations and their errors.

In our case, and for special considerations further explained in Appendix E, we need to significantly strengthen the constraint on very high releases; i.e., lower the probability of a very high release. For this purpose, the gamma distribution is very appropriate. It is written as:

$$p\left(\frac{\mathbf{q}}{q_{\text{ref}}};k,\theta\right)=\frac{\left\|\frac{\mathbf{q}}{q_{\text{ref}}}\right\|^{k-1}e^{-\left\|\frac{\mathbf{q}}{q_{\text{ref}}}\right\|/\theta}}{\theta^{k}\Gamma(k)}, \tag{6}$$

where $||\cdot||$ is the max norm: $\left\|\frac{\mathbf{q}}{q_{\text{ref}}}\right\|=\max\left(\left|\frac{q_1}{q_{\text{ref}}}\right|,\ldots,\left|\frac{q_{N_{\text{imp}}}}{q_{\text{ref}}}\right|\right)$ and $q_{\text{ref}}=1\text{Bq.s}^{-1}$. In this way, we only make an assumption on the magnitude of the release and not on the timing of the release.

The prior should be set so that the total magnitude of the release does not exceed 10^{18} Bq. By way of comparison, the total ^{137}Cs released during Fukushima accident was about 10^{16} Bq and the total ^{131}I released was around 10^{17} Bq. However since we work with daily releases in Bq.s^{-1}, the magnitude limit on one day for a minor accident should be: $\frac{10^{18}}{86400}\approx e^{30}$ Bq.s^{-1}. To reduce the probability of very high releases, θ is the control parameter of importance. θ corresponds to the spreading term of the gamma distribution and should therefore be set as e^{30}.

As justified in Appendix C, the source term variables of interest are the logarithms of the release rates instead of the release rates. The corresponding prior is therefore:

$$f_{\ln(\mathbf{Q})}\left(\ln\frac{\mathbf{q}}{q_{\text{ref}}};k,\theta\right)=\frac{e^{k\ln\frac{\mathbf{q}}{q_{\text{ref}}}}e^{-e^{\ln\left\|\frac{\mathbf{q}}{q_{\text{ref}}}\right\|}/\theta}}{\theta^{k}\Gamma(k)}=\frac{\left\|\frac{\mathbf{q}}{q_{\text{ref}}}\right\|^{k}e^{-\left\|\frac{\mathbf{q}}{q_{\text{ref}}}\right\|/\theta}}{\theta^{k}\Gamma(k)}. \tag{7}$$

This new prior is sometimes defined as a log-gamma distribution. As regards the shape parameter k, setting $k>0$ would reduce the probability of minor releases, which is not something desirable. We should therefore set k to 0.

Concerning the prior for the hyperparameter r, which represents the variance parameter of our likelihood distribution, we have chosen to sample it within the hierarchical Bayesian framework:

$$\begin{aligned}p(\mathbf{x}|\mathbf{y})&=p(\mathbf{x}_S,r|\mathbf{y})=\frac{p(\mathbf{y}|\mathbf{x}_S,r)p(\mathbf{x}_S,r)}{p(\mathbf{y})}\\&=\frac{p(\mathbf{y}|\mathbf{x}_S,r)p(\mathbf{x}_S|r)p(r)}{p(\mathbf{y})}=\frac{p(\mathbf{y}|\mathbf{x}_S,r)p(\mathbf{x}_S)p(r)}{p(\mathbf{y})},\end{aligned} \tag{8}$$

since we assume the variables describing the source $\mathbf{x}_S=\left(x_1,x_2,\ln\frac{\mathbf{q}}{q_{\text{ref}}}\right)$ to be independent from the uncertainties.

The prior for this variance is simply defined as a uniform distribution on $\mathbb{R}_*^+$. A Jeffrey prior could also be considered and would be defined as $p(r)=\sqrt{\frac{1}{r}}$ (Liu et al., 2017). Note that there are many other viable paths for estimating the hyperparameter r, such as the expectation-maximisation algorithm used in Liu et al. (2017), which estimate r in an independent algorithm.

Note also that the matrix $\mathbf{R}$ accounts for the three usual sources of errors (e.g., Rajaona et al., 2015):

- the observation error, due to the built-in sensor noise and bias;
- the modelling error: the resolvent matrix $\mathbf{H}$ is processed using both modelled meteorological fields and atmospheric dispersion models. These two modelling processes are subject to uncertainty;
- the representativeness error, due to the interpolation process in time used to represent the time-dependent release rate and the bilinear interpolation process in space to construct the observation operators.

The assumption that the matrix $\mathbf{R}$ is diagonal supposes that correlations between observations are negligible, which is an acceptable hypothesis for measurements distant from each other. The validity of this hypothesis is studied in the work of Saunier et al. (2019). The homoscedasticity property (the error variance is the same for all observations) suggests that the measurement instruments have similar precision and that the modelling precision does not vary significantly in space and time. The unique variance hyperparameter r captures an average of the variance parameters of the observations.

3.1.3. Cost function

We have defined the likelihood and the priors that yield the posterior distribution of interest: the joint distribution of $\mathbf{x}=\left(x_1,x_2,\ln\frac{q_1}{q_{\text{ref}}},\ldots,\ln\frac{q_{N_{\text{imp}}}}{q_{\text{ref}}},r\right)$ conditionally on y. These probability distributions can be replaced by cost functions, which are often more practical for computations. The cost function $\mathcal{J}_{\mathbf{y}}$ is defined as

$$\mathcal{J}_{\mathbf{y}}:\mathbf{x}\rightarrow\mathcal{J}_{\mathbf{y}}(\mathbf{x})=-\ln p(\mathbf{x},\mathbf{y})=-\ln p(\mathbf{x}|\mathbf{y})p(\mathbf{y})=-\ln p(\mathbf{y}|\mathbf{x})-\ln p(\mathbf{x}), \tag{9}$$

and is defined for the variable vector of a specific source: we will call $\mathcal{J}_{\mathbf{y}}(\mathbf{x})$ the cost of a source $\mathbf{x}$.

As it will be discussed in more detail later, the cost of a source $\mathbf{x}$ is only used in comparison to the cost of an other source $\mathbf{x}$'. That being said, we can derive the cost function keeping only the relevant parts, i.e., the parts not depending on $\mathbf{x}$.

Assuming that the hyperparameters θ, N_{obs} and the set of observations $\mathbf{y}$ are all defined before any Bayesian reconstruction of the source and thus not included in $\mathbf{x}$, $\mathcal{J}_{\mathbf{y}}$ can therefore be written as:

$$\mathcal{J}_{\mathbf{y}}\left(x_1,x_2,\ln\frac{\mathbf{q}}{q_{\text{ref}}},r\right)=\frac{1}{2r}\ln\left(\frac{\mathbf{y}+\mathbf{y}_t}{\mathbf{H}\mathbf{x}+\mathbf{y}_t}\right)^{\mathrm{T}}\ln\left(\frac{\mathbf{y}+\mathbf{y}_t}{\mathbf{H}\mathbf{x}+\mathbf{y}_t}\right)+\frac{N_{\text{obs}}}{2}\ln(r)+\|\frac{\mathbf{q}}{q_{\text{ref}}}\|/\theta, \tag{10}$$

where the first part corresponds to the cost of the likelihood, while the second part corresponds to the cost of the prior for the logarithms of the release rates. Note that we do not derive the costs of the priors for the coordinates or the covariance matrix term for the sake of readability, since those are just given by indicator functions.

3.2. Markov chain Monte Carlo sampling: the Metropolis-Hastings algorithm

In this subsection, we present the MCMC methods (e.g., Gilks et al., 1995), a class of algorithms that we will use to sample from the probability distribution $p(\mathbf{x}|\mathbf{y})$, which has been defined in Section 3.1. A Markov chain $(Q_t)_{t\in T}$ is a stochastic model describing a sequence of possible events (or random variables) in which the probability of each event depends only on the state attained in the previous event. The principle of MCMC methods is to build a Markov chain that has for

invariant distribution the target distribution. Hence, after sufficient time (burn-in), the Markov chain samples values of this distribution.

A particularly popular MCMC method is the Metropolis-Hastings (MH) algorithm. We describe here the foundations and the principle of the MH algorithm in the case where it is aimed at sampling from the ^{106}Ru source's variables vector conditional distribution $p(\mathbf{x}|\mathbf{y})$.

The Metropolis-Hastings algorithm relies on the theory of the Markov processes, which states that the existence and uniqueness of a stationary distribution is met under the condition of the ergodicity of the Markov process and the detailed balance. The effective algorithm is derived from these two conditions. To initiate the process, an initial set of variables. should be generated, either randomly or according to the priors, and then cycling through the following three steps:

- Generate a new candidate $\mathbf{x}$' at iteration i from previous state $\mathbf{x}_i$ according to a transition probability $g : \mathbf{x} \rightarrow \mathbf{x}$' described in the next section.
- Compute the corresponding acceptance ratio $\alpha_{\mathbf{y}}(\mathbf{x}'|\mathbf{x}_i)$:

$$\alpha_{\mathbf{y}}(\mathbf{x}'|\mathbf{x}_i) = \frac{p(\mathbf{y}|\mathbf{x}')p(\mathbf{x}')g(\mathbf{x}_i|\mathbf{x}')}{p(\mathbf{y}|\mathbf{x}_i)p(\mathbf{x}_i)g(\mathbf{x}'|\mathbf{x}_i)}. \tag{11}$$

- Generate $u \sim \mathscr{U}(0,1)$. If $u \leq \alpha_{\mathbf{y}}(\mathbf{x}'|\mathbf{x}_i)$: set $\mathbf{x}_{i+1} = \mathbf{x}'$. If not, set $\mathbf{x}_{i+1} = \mathbf{x}_i$.

3.3. Definition of the transition probabilities

The transition probabilities that are used in the MH sampling strategies must be defined. Note that the choice of these transition probabilities does not change the nature of the posterior distribution. However, they should be carefully chosen since the efficiency of the MH algorithm relies mainly on the choice of the proposals. The reversibility of the transition probabilities is an advantage for the efficiency and the simplicity of the MH algorithm, which means that we would like the following condition to be observed:

$$g(\mathbf{x}'|\mathbf{x}_i) = g(\mathbf{x}_i|\mathbf{x}'), \tag{12}$$

where g is a transition probability, $\mathbf{x}$' plays the role of the proposal and $\mathbf{x}_i$ of the current state. Many distributions share this property: the transition probabilities that we use are based on the folded normal distribution whose probability density function is given by

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}} + e^{-\frac{(x+\mu)^2}{2\sigma^2}}\right)1_{\mathbb{R}_+^*}(x), \tag{13}$$

where μ is the mean and σ^2 is the variance of the unfolded distribution. The folded normal distribution is folded in 0 but a similar distribution folded on another point can be defined. The rationale behind the use of this distribution is the following: let us assume that we are sampling from a positive variable u, which means that we have set a prior probability for u for which $p(u \leq 0) = 0$. Consequently, a MH step making a proposal u' with u' negative would be systematically rejected and thus would be a waste of time. Setting a folded normal distribution as the transition probability of u ensures that the proposal u' will never be negative. Consequently, this saves time for the algorithm and does not change in any aspect the posterior distribution. Note that the following transition probabilities are defined independently for each component of the source vector $\mathbf{x}$ so that for two variables u and v describing $\mathbf{x}$:

$$g((u',v')|(u,v)) = g_u(u'|u)g_v(v'|v). \tag{14}$$

The transition probability on the coordinates x_1, x_2 can be described as a folded Gaussian distribution with a standard deviation $\sigma_{x_1} = \sigma_{x_2}$ folded in a way that the coordinate variables are restricted to a spatial domain (see Table 1).

Regarding the transition probability of the logarithm of the release, we can use a Gaussian distribution. It is simply written as

$$\ln\left(\frac{\mathbf{q}_{i+1}}{q_{\text{ref}}}\right) \sim \mathscr{N}\left(\ln\left(\frac{\mathbf{q}_i}{q_{\text{ref}}}\right), \boldsymbol{\sigma}_{\mathbf{lnq}}\right) \tag{15}$$

where $\mathscr{N}$ stands for the normal distribution. Taking the exponential of these logarithms yields the release rates. Note nevertheless that the probability distribution of $\mathbf{q}$ cannot be derived immediately from the distribution for the logarithm. The Jacobian must first be derived, since the logarithm function is not linear.

Different transition probabilities can be used for the hyperparameter r. Log-normal statistics again are well suited but, in contrast to the magnitude of the release rates, the magnitude range of r is very small. Indeed, $\mathbf{R}$ is the covariance matrix of a log-normal distribution and, as such, is a dimensionless parameter representing the uncertainties between the observations and the predictions. A positive Gaussian distribution, the folded normal distribution, is thus more appropriate in this case.

Transition probability variances do not have any influence on the final distribution, but a classic way to proceed is to choose them so that the acceptance rate converges to a value in the vicinity of 0.234, which is the optimal acceptance ratio under certain assumptions (Roberts et al., 1997). We have chosen $\sigma_x = 0.3$, $\sigma_{\text{lnq}} = 0.05$ and $\sigma_{\text{r}} = 0.03$. Note that since each variance has a leverage on the final acceptance rate, many combinations of $(\sigma_x, \sigma_q, \sigma_r)$ satisfy this optimal acceptance rate and this specific choice has been made based on numerical experiments to increase the convergence rate.

3.4. Application of the Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm is applied to retrieve the source of the ^{106}Ru. The release rate vector $\mathbf{q}$ is defined as a vector of seven uniform daily release rates from the 22$^{\text{th}}$ to the 28$^{\text{th}}$ of September 2017. Indeed, Saunier et al. (2019) indicate that the Ruthenium was mainly released on the 26$^{\text{th}}$ of September, which motivates this choice. In this way, we reconstruct seven time-independent rate components q_{22}, …, q_{28}, each component corresponding to a uniform emission. Variables are initialised randomly in the state space: more specifically, coordinates are initialised randomly in the domain [6W; 70E] and [34N; 68N], which corresponds to the domain of the meteorological fields. The seven release rate variables are initialised randomly in $[0\ \text{Bq.s}^{-1}, 10^{10}\ \text{Bq.s}^{-1}]$ and r randomly in $[1, 20]$. In this subsection, only the ARPEGE Météo-France meteorological data are used. The MH chain runs for 5.10^5 iterations.

The histograms of the probability distribution of the coordinates are constructed using the MH algorithm for the different initialisations in Fig. 3. Fig. 3.a is the histogram for an initialisation in England, while Fig. 3.b corresponds to a random initialisation. Chain samples are greatly influenced by their initialisation, even after a large number of iterations, and are caught in local minima basins, regardless of the step sizes of the jumps (i.e., the variance of the transition probability). These local minima appear because of a lack of information: in our case, this lack of information is due to a poor distribution of the observations. Since our multimodal joint distribution cannot be sampled correctly in operational time, we must accelerate the convergence of the MH algorithm. For this reason, we will now introduce a method to accelerate and ensure the convergence of the algorithm.

4. Accelerating MCMC algorithm with the parallel tempering algorithm

As previously seen, MCMC convergence issues in real cases studies can be due to systems poorly constrained, resulting in many modes where the probability mass is located in different areas of the state

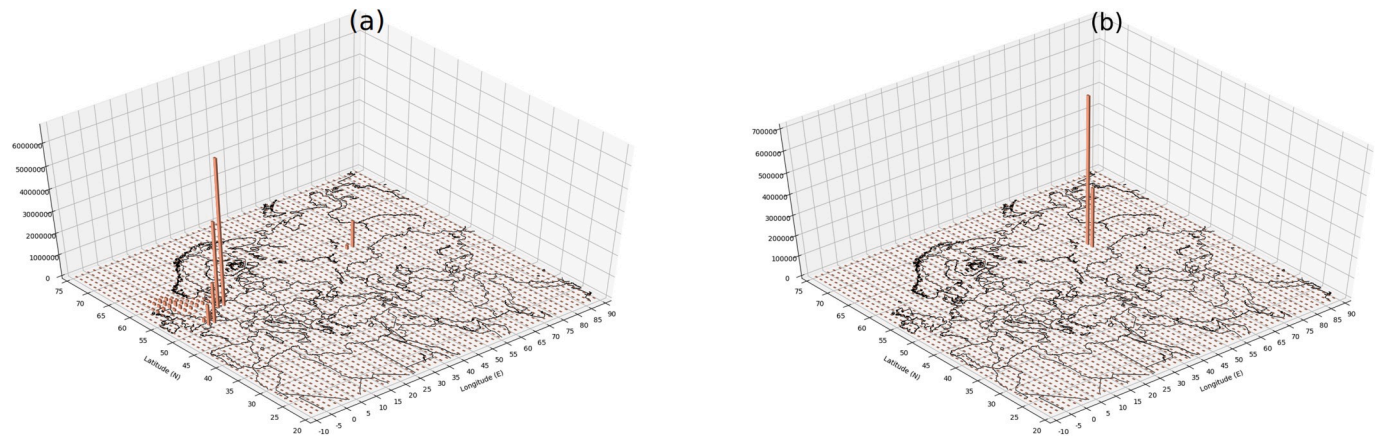

Fig. 3. Joint distribution of the coordinate variables sampled using a MH algorithm with initialisation in England of the coordinates variables (a) and random initialisation of the coordinate variables (b).

space. To address this difficulty, a variety of methods is described in the literature, some of them relying on the idea of tempering. In what follows, we present the parallel tempering algorithm originally introduced by Swendsen and Wang (1986) and developped by Geyer (1991), also called in the literature Metropolis-coupled Markov chain Monte Carlo (MCMCMC) or temperature swapping (Earl and Deem, 2005; Baragatti, 2011; Atchadé et al., 2011). The main idea of parallel tempering is to combine different MCMC chains at different temperatures. High temperatures $T > 1$ flatten out the target distribution, thus allowing the corresponding chains to explore the entire state space and avoid local minima, provided that the temperature is high enough. The method has already been used in the context of phylogenetic inference by Altekar et al. (2004), who describes a particular temperature swapping strategy. Yee (2008) depicts a strategy using, in particular, parallel tempering for inverse modelling in the case of several radioactive sources. In the following, we describe the process in detail.

4.1. Overview of parallel tempering

Let us call the target distribution $\pi : \mathbf{x} \rightarrow \pi(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})$. A chain at temperature T that we will also call a T-tempered chain is a chain which has for target distribution:

$$\pi^{\frac{1}{T}} : \mathbf{x} \rightarrow \pi(\mathbf{x})^{\frac{1}{T}} \left(\int \pi(\mathbf{x})^{\frac{1}{T}} d\mathbf{x} \right)^{-1}, \tag{16}$$

with $T \geq 1$. In particular, note that the non-tempered chain or 1-tempered chain is the one targeting the posterior distribution of interest. With this definition, the tempered chain sample from a flattened posterior distribution: the higher the temperature is, the easier it is for the tempered chain to overcome the issue of local minima and consequently quickly sample from its target distribution.

Let us now define $1 < T_1 < T_2 < \ldots < T_n$ the selected temperatures and the corresponding n+1 Markov chains with invariant distribution π, $\pi^{\frac{1}{T_1}}, \pi^{\frac{1}{T_2}}, \ldots, \pi^{\frac{1}{T_n}}$ respectively. Parallel tempering (PT) proceeds by running each chain in parallel using a Metropolis-Hastings algorithm. Swap moves between chains at adjacent temperature levels are allowed according to an acceptance ratio satisfying the detailed balance and thus respecting the MCMC theory. The principle of the PT approach relies on connecting fast-mixing high-tempered Markov chains with a slow-mixing low-tempered Markov chain (Robert et al., 2018).

More precisely, PT switches between two dynamics:

- *Single-temperature move*: in some iterations we attempt to update the parameters of each n+1 chain with a traditional MCMC acceptance process, such as the Metropolis-Hastings algorithm;
- *Swapping (two chains at adjacent temperatures)*: in some iterations, we attempt to swap the variables of two Markov chains at different temperatures. The probability of such swapping is calculated according to the Metropolis-Hastings theory. Everything happens as if the swapping step was just another MH step. One can note that swapping variables is equivalent to swapping temperatures except for its computation time, since swapping of two temperatures is always equivalent to swapping of two real numbers while the number of variables can be large.

Let us derive the acceptance ratio, which corresponds to swapping of the variables of the chains T_j and T_k. We call $\mathbf{x}_j$ the variables of the T_j -tempered chain before the swap and $\mathbf{x}_k$ the variables of the T_k -tempered chain. For the T_j -tempered chain, the transition probability defined in the acceptance ratio of the MH theory is $\pi^{\frac{1}{T_k}}$. Indeed, the probability that a variables vector $\mathbf{x}$ is proposed to the chain at temperature T_j from the chain at temperature T_k is simply the probability $\pi^{\frac{1}{T_k}}(\mathbf{x})$ of the chain at temperature T_k to sample the variables vector $\mathbf{x}$. The MH acceptance ratio of the chain at temperature T_j to make the transition towards $\mathbf{x}_k$ given $\mathbf{x}_j$ is consequently (Baragatti, 2011; Atchadé et al., 2011):

$$\min\left(1, \frac{\pi^{\frac{1}{T_j}}(\mathbf{x_k})\pi^{\frac{1}{T_k}}(\mathbf{x_j})}{\pi^{\frac{1}{T_j}}(\mathbf{x_j})\pi^{\frac{1}{T_k}}(\mathbf{x_k})}\right). \tag{17}$$

4.2. Optimisation of the parallel tempering algorithm

Before any computation, the way in which the different chains interact with each other and how they are defined must be chosen. In particular, we must define the number of chains, the temperatures of each chain and when two chains swap. We already decided to only allow swaps between chains at adjacent temperatures, as in the work of Robert et al. (2018) or Earl and Deem (2005) but, for instance, Altekar et al. (2004) has chosen to define a more complex process and Baragatti (2011) opts for a random process.

The choice of the number of temperatures and the spacing between them is crucial. If the gap between two temperature levels is too large, the swap acceptance ratio will be too small: the swap will rarely be accepted and this will delay the convergence of the MCMC. Nevertheless, if the gap between two temperature levels is too small, a very large number of intermediate chains and a large computation power might be

needed to really improve the mixing. Furthermore, a very large number of intermediate chains can slow down the process if the information from the hot states requires too much time before having an influence on the cold chains. This choice of the gap between the temperatures relies on the "Goldilocks principle": not too large, not too small (Robert et al., 2018).

Baragatti (2011) suggests to initialise temperatures by setting the highest temperature and the lowest temperature. The lowest should be set to 1 by construction and the highest should be chosen sufficiently high for the corresponding chain to quickly converge to its corresponding target distribution. In other words, the highest temperature T_n should be set so that the corresponding chain rapidly explores the entire state space ensuring that no mode is forgotten. Once the highest temperature is chosen, several strategies to implement the other temperatures exist. As presented in the work of Robert et al. (2018), most of the historical literature suggests the implementation of geometrical spacing between temperatures. In practice, c and n must be derived so that $T_n = c^n$. Each temperature T_i is then equal to c^i. Overall, the parameters c and n should be chosen so that the n swap acceptance ratios converge to a value between 0.1 and 0.3. Indeed, Atchadé et al. (2011) or Roberts and Rosenthal (2014) prove that, under certain conditions, the optimal swap acceptance ratio between two temperature levels should converge precisely to 0.234. Adaptive spacing could be derived from this condition but, in our case, we have decided to set geometrical spacing for simplicity.

Typically, the mass of the probability landscape of the chains at high temperature will be well distributed over the entire state space. In one dimension, if the weight of the probability distribution $p(x)$ of a variable x is fully described in a state space of a certain width, the weight of the probability distribution with tempering $p(x)^{\frac{1}{T}}$ of the same variable might be described in a much broader state space. Since we use MCMC algorithms to sample from these distributions, we use random walks with specific transition probability variances. These variances are set so as to improve the mixing of the cold chain and, consequently, are typically too small to allow the chains at higher temperature to explore the entire state space rapidly. To solve this issue, variances of our transition probabilities are modified: for a transition probability variance of the cold chain noted σ, the transition probability variance of the chain at temperature T is chosen as

$$\sigma_T = \sigma\sqrt{T}. \tag{18}$$

4.3. Application of the PT algorithm

The parallel tempering algorithm is applied to retrieve the source of the ^{106}Ru event. Seven parallel chains at different temperatures are initialised with a geometrical spacing (with a geometrical rate c of 3). The spacing rate and the number of temperatures have been chosen so as to increase the convergence rate according to the MCMC theory; i.e., the swap acceptance ratios converge to values around 0.2. Both sets of meteorological data, ARPEGE Météo-France and ERA5 ECMWF, are used independently.

Each chain, as for the Metropolis-Hastings algorithm, samples from a variable vector composed of the longitude, the latitude, the logarithms of seven release rates from the 22nd to the 28th of September, and a variance parameter r. Transition probability variances are modified as previously described. PT uses 5×10^5 samples. Note that all logarithms are assumed to be natural logarithms.

The sampling process has been done in two steps: at first, the PT algorithm has been applied over the domain of research defined in Table 1.

Fig. 4.a and 4.b describe the reconstruction of the longitude and latitude of the source using Météo-France and ERA5 data. It shows that the regions between Volga and the Southern Ural are the most reliable areas of the potential release. However, the pdf of the longitude reconstructed using ERA5 data is significantly different from that reconstructed using Météo-France data (difference of the means of the pdfs approaches 8°~800 km). To improve the accuracy of the source reconstruction, and to show that this discrepancy is due to interpolation representativeness error, a sub-domain d of dimension [47E; 62E] and [53N; 58N] has been extracted and divided into meshes of $0.5^\circ \times 0.5^\circ$ where a set observations operator $\mathbf{H}_i$ has been computed. The spatial resolution of the dispersion modelled on this sub-domain is thus $0.5^\circ \times 0.5^\circ$. From now on, all of the following presented results have been run on the full domain defined in Table 1 with the difference that the simulation of the dispersion of a source located in d is computed using bilinear interpolation on the $\mathbf{H}_i$ instead of the $\mathbf{H}_k$ introduced in Section 2.4.

Fig. 5a and 5b describe the probability distribution functions of the source's longitude and latitude using both meteorologies. Distribution parameters are described in Table 2. With the ERA5 data, the mean of the distribution of the source's coordinates is [58.614E; 56.325N] while using Météo-France data, the mean is given by [58.968E; 55.72N]. The location of the source estimated in the work of Saunier et al. (2019) also gives similar estimation. The difference in longitude between the two results is 0.354° and in latitude 0.605°, which is a small difference

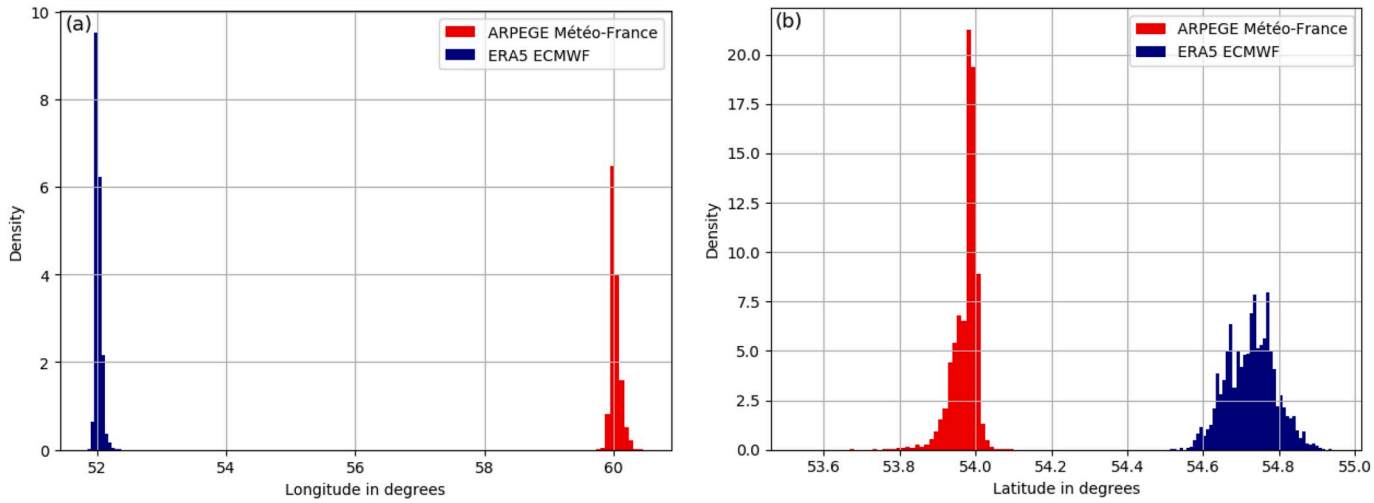

Fig. 4. Distribution of the variables describing the Ruthenium source sampled using the parallel tempering algorithm in the domain defined in Table 1: longitude (a), latitude (b).

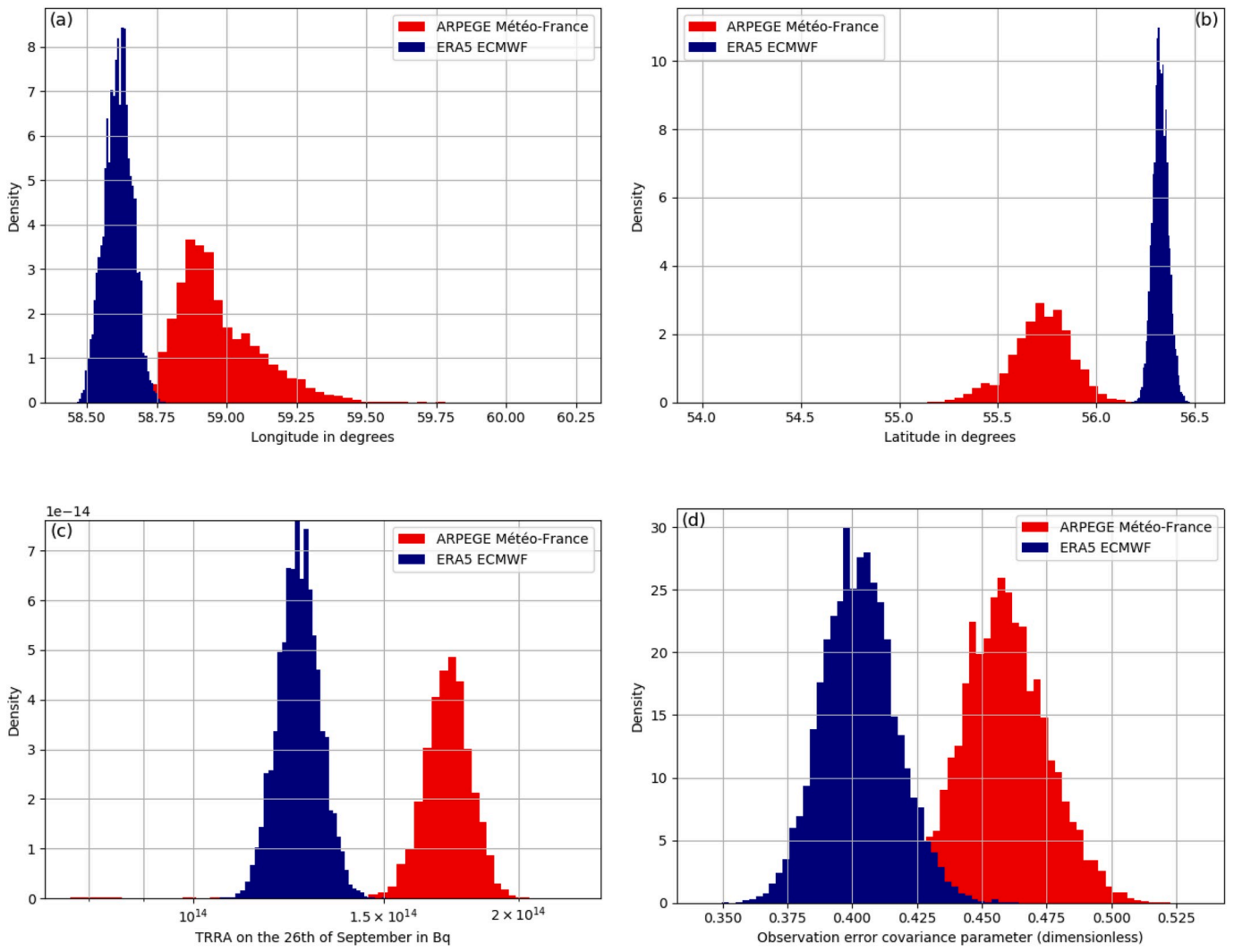

Fig. 5. Distribution of the variables describing the Ruthenium source sampled using the parallel tempering algorithm: longitude (a), latitude (b), TRRA on the 26th of September (c), and the observation error covariance parameter r (d). Dispersion on the domain d is computed using bilinear interpolation on the $\mathbf{H}_i$ instead of the $\mathbf{H}_k$.

compared to the spatial resolution of the meteorological ARPEGE fields ($0.5^{\circ} \times 0.5^{\circ}$). Both distributions indicate that the source must be located in the Southern Ural. Furthermore, the shape of the pdfs states that the probability of the source being located somewhere else in Europe is very low. The standard deviations of the longitude and latitude distributions corresponding to the ARPEGE fields are more than three times greater than those corresponding to the ERA5 fields. The ERA5 reconstructed distributions are almost Gaussian as well as the distribution of the latitude using the ARPEGE fields. However, the distribution of the longitude using the ARPEGE fields has a log-normal shape.

Table 2
Statistical parameters of the distributions of the variables of interest sampled using ARPEGE Météo-France or ECMWF ERA5 data. Dispersion in the domain d is computed using bilinear interpolation on the $\mathbf{H}_i$ instead of the $\mathbf{H}_k$.

	Mean	Median	MAP	Std
Longitude (ARPEGE)	58.968°	58.934°	58.915 °	0.15 °
Longitude (ERA5)	58.614°	58.614°	58.625°	0.05°
Latitude (ARPEGE)	55.72 °	55.728°	55.724°	0.153°
Latitude (ERA5)	56.325°	56.324°	56.292°	0.039°
Log. release rate 26th Sept. (ARPEGE)	21.407	21.414	21.434	0.071
Log. release rate 26th Sept. (ERA5)	21.098	21.097	21.093	0.046
Obs. error parameter (ARPEGE)	0.459	0.458	0.451	0.017
Obs. error parameter (ERA5)	0.403	0.402	0.387	0.014

Fig. 5c shows the pdfs of the logarithm of the release rate on the 26th of September, since most of the Ruthenium was released on this day according to the samples. Both pdfs have Gaussian shape. Using the ERA5 meteorology, the mean of the TRRA (Total Retrieved Released Activity) distribution on the 26th of September is 125 TBq and the TRRA is between 113 TBq and 139 TBq. Using the Météo-France meteorology, the mean of the TRRA distribution on the 26th of September is 171 TBq and the TRRA is between 140 TBq and 207 TBq. The standard deviation of the ARPEGE corresponding distribution is again greater than its counterpart, but the retrieved distribution values are very similar. These results are consistent with estimations performed by the IRSN using variational approaches (Saunier et al., 2019).

Fig. 5d shows the pdfs of the hyperparameter r using both meteorologies. Both distributions have Gaussian shapes and an average of 0.403 and 0.459 for ECMWF ERA5 and Météo-France data, respectively. The ERA5 error term is smaller than the Météo-France error term. This makes sense given that the spatial and time resolutions of the ERA5 HRES fields are finer than the Météo-France ARPEGE fields that are used here. Both standard deviations for the two meteorologies are very small and very similar. Uncertainties are low and show great confidence in our results.

Fig. 6a displays the histograms of the reconstructed logarithms of the release rates using ERA5 data and explains why we focus on the distributions of the release on the 26th of September. Indeed, the release on

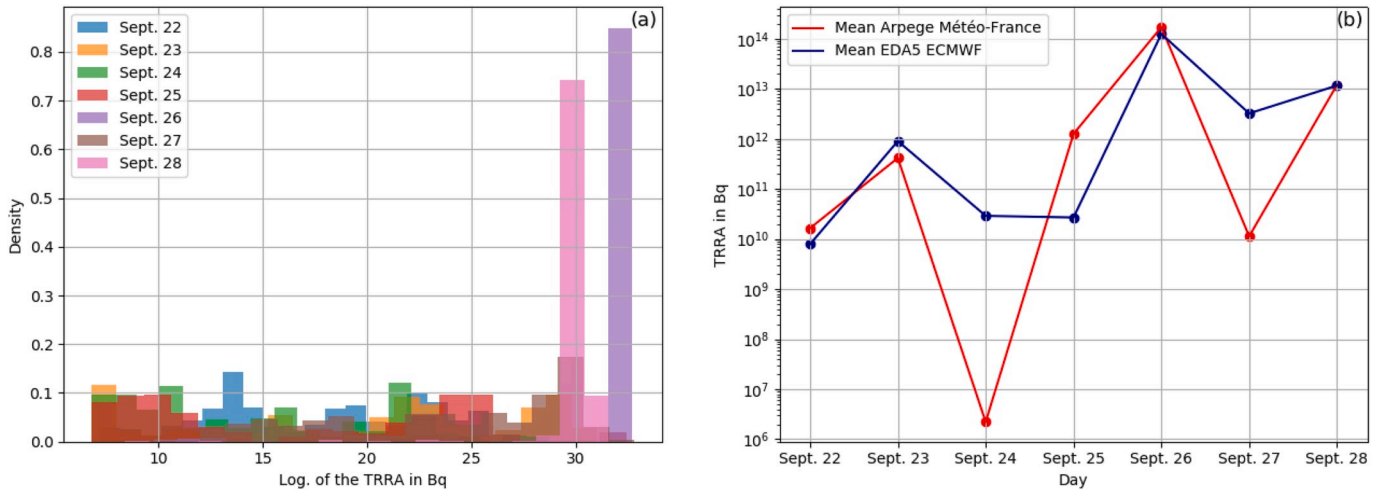

Fig. 6. Figure (a) shows the histograms of the natural logarithms of the TRRA for the different dates sampled using the parallel tempering algorithm and the ERA5 ECMWF meteorological data. Figure (b) displays the evolution of the mean of the reconstructed distribution of the TRRA.

the 26th of September is estimated to be more than 10 times greater than the release on the 28th of September and every other day. Fig. 6b describes the evolution of the mean of the release rate logarithm distribution from the 22nd of September to the 28th using both sets of meteorological data. Note that the use of a regularisation term on average releases such as a Gaussian prior (Tichý et al., 2016) may decrease the daily release rates obtained. Note also that using the Météo-France fields, there was a non-null (but very low) probability that the release primarily occurred on the 25th of September instead of the 26th of September. These samples are not visible in the figures below.

4.4. Summary and conclusions

In this paper, we have explored several Bayesian inverse modelling methods to reconstruct the distributions of the ^{106}Ru source's variables following the event of ^{106}Ru detection in Europe in autumn 2017. The inverse problem was to retrieve the complete time-dependent source term as well as the coordinates of the source since its location is still not known. Estimating the distribution of the observation error covariance matrix was also an objective of the study.

Log-normal statistics have been considered to model the likelihood of the observations. The location prior was set as a uniform distribution over a very large area. The prior for the observation error hyperparameter was set as a uniform distribution over the positive real numbers. We have decided to sample from the logarithms of the release rates.

To sample from the probability distribution of these variables, two MCMC methods have been used. First, a traditional MCMC approach, the MH algorithm, has been tested. Transition probabilities used were based on the folded Gaussian distribution to accelerate the convergence of the chain. However, the algorithm fails to converge within an operational time period unless the coordinates of the source are initialised in the Southern Ural. After exploitation of the data, this failure seems to be due to the sparse distribution of the observations. To solve the convergence issues, the parallel tempering algorithm is used. To ensure convergence of the chains at high temperatures, a log-gamma prior is set on the logarithms of the release rates. Temperatures are initialised following a classical procedure and a geometrical spacing. The transition probabilities are initialised differently for each temperature, as their corresponding state spaces vary. The ability of the parallel tempering algorithm to escape local minima is a major advantage in the sampling of the source's location variable. The algorithm converges in a very short time on a cluster ($<$ 1 h) and provides a full reconstruction of each variable.

Distributions of each variable have been sampled based on two sets of meteorological data: ECMWF ERA5 and ARPEGE Météo-France.

Sampling of the source coordinate distributions describes an area located in the Southern Ural. The maximum of the source coordinate distribution is reached at coordinates [58.6E; 56.3N] for ERA5 data and [58.9E; 55.7N] for Météo-France data. Furthermore, the shape of the pdfs indicates that the probability of the source being located somewhere else in Europe is very low.

The results show that Ruthenium was primarily released on the 26th of September with quantities ranging between 100 and 200 TBq. The magnitude of the error observation parameter term is estimated at 0.40–0.45 using log-normal statistics.

Data availability

The observation dataset used is described in detail and is publicly available in the work of Masson et al. (2019) on supplementary information.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Joffrey Dumont Le Brazidec: Software, Methodology, Conceptualization, Investigation, Writing - original draft, Visualization. **Marc Bocquet:** Methodology, Conceptualization, Writing - review & editing, Visualization, Supervision. **Olivier Saunier:** Resources, Methodology, Conceptualization, Writing - review & editing, Visualization, Supervision. **Yelva Roustan:** Methodology, Software, Writing - review & editing.

Acknowledgements

The authors would like to thank the reviewer for very complete comments and suggestions, and Météo-France and the European Centre for Medium-Range Weather Forecasts (ECMWF) for the meteorological fields used in this study. The authors also acknowledge the support provided by CEREA, the joint laboratory of École des Ponts ParisTech and EdF R&D. CEREA is a member of Institut Pierre Simon Laplace (IPSL).

Appendix A. Acronyms

UQ	Uncertainty quantification
STE	Source term estimation
MAP	Maximum a posteriori
MCMC	Markov chain Monte Carlo
MH	Metropolis-Hastings
PT	Parallel tempering
pdf	Probability density function
TRRA	Total retrieved released activity
IRSN	Institut de radioprotection et de sûreté nucléaire

Appendix B. Choice of a likelihood threshold

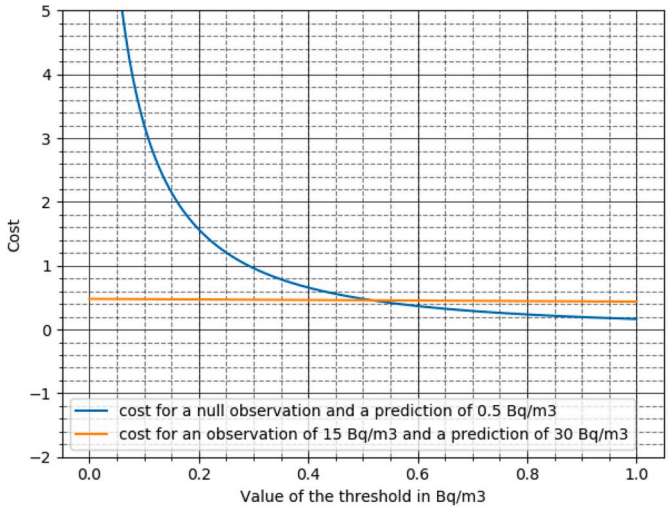

Fig. B.1. Log-normal cost for two sets of corresponding observation and prediction according to the threshold. The blue curve describes the error cost for a null observation and a corresponding predicted concentration of 0.5 mBq.m^{-3}. The orange curve describes the error cost for an observed concentration of 15 mBq.m^{-3} and a corresponding predicted concentration of 30 mBq.m^{-3}.

Applying inverse problem techniques to a large set of observations implies that one needs to assign a cost to differences between observations and predictions. Most of the cost in the Ruthenium case comes from either a false non-detection (where the prediction is zero and the observation is positive) or a false detection (where the prediction is positive and the observation is zero) or either a large difference between a positive observation and a positive prediction. When using a log-normal likelihood, it is necessary to set a threshold to deal with zero observations or predictions and this threshold should be chosen based on how one wants to compare the cost of a false non-detection or detection with the cost of a large difference in a positive couple observation-prediction.

In Fig. B.1, we compare the costs of two differences; i.e., we compare the importance given to the error between an observed concentration and a corresponding predicted concentration $(y_i, (\mathbf{Hx})_i) = (0\ \text{mBq}\cdot\text{m}^{-3}, 0{\cdot}5\ \text{mBq}\cdot\text{m}^{-3})$ and $(y_j, (\mathbf{Hx})_j) = (15\ \text{mBq}\cdot\text{m}^{-3}, 30\ \text{mBq}\cdot\text{m}^{-3})$. The cost function computed here corresponds precisely to $\ln\left(\frac{y_i+y_t}{(\mathbf{Hx})_i+y_t}\right)$. We use these two couples because we have no reason to favour one couple over the other: based on different numerical experimentations, we believe that these two errors should affect the choice of the source approximately in the same way. This amounts to saying that both types of couples (false non-detection or detection and large differences, here a factor of 2) should be taken in account when trying to retrieve the source.

In Fig. B.1, we observe that the orange curve that corresponds to $(y_j, (\mathbf{Hx})_j)$ is constant and equal to 0.5, which is consistent with the definition of ξ_t, which depends on the threshold only for small concentrations, while the blue curve follows an exponentially decreasing trend. We observe that these two costs are equal for a threshold of approximately 0.5 mBq.m^{-3}. For values smaller than 0.1 mBq.m^{-3}, the orange curve cost becomes significantly higher than its counterpart.

In practice, for threshold values from 0.1 mBq.m $^{-3}$ to 1 mBq.m^{-3}, differences in the source assessment are very small and negligible. Saunier et al. (2019) use a threshold of 0·1 mBq·m^{-3} for the Ruthenium case. We use here a threshold of 0.5 mBq.m^{-3}. Another solution would be to include the estimation of the threshold y_t in the Bayesian framework as a hyperparameter.

Appendix C. The logarithms of the release rates as control variables

Let us assume that we use the release rate vector $\mathbf{q}$ as a control variable. Due to the large range of values that release rates may have, log-normal statistics are very convenient. It furthermore has the advantage of ensuring the positivity of the releases. Consequently, a multivariate log-normal distribution can be used. With $\mathbf{q}_{i+1}$ the proposal computed from $\mathbf{q}_i$, we have:

$$g(\mathbf{q}_{i+1}|\mathbf{q}_i) = \frac{e^{-\frac{1}{2}\ln\left(\frac{\mathbf{q}_{i+1}}{\mathbf{q}_i}\right)^\top \sum_q^{-1} \ln\left(\frac{\mathbf{q}_{i+1}}{\mathbf{q}_i}\right)}}{(2\pi)^{\frac{N_{imp}}{2}}\left|\Sigma_q\right|^{1/2}\prod_{j=1}^{N_{imp}} q_{i+1}^j}, \tag{C.1}$$

with $\mathbf{\Sigma_q}$ which can be considered, for the sake of simplicity, as diagonal and equal to $\sigma_q\mathbf{I}$. Consequently, each daily release rate only depends upon its previous value, which yields:

$$\frac{g(\mathbf{q_{i+1}}|\mathbf{q_i})}{g(\mathbf{q_i}|\mathbf{q_{i+1}})} = \frac{\prod_{j=1}^{N_{imp}} {q_i}^j}{\prod_{j=1}^{N_{imp}} {q_{i+1}}^j}. \tag{C.2}$$

This dissymmetry would be included in the acceptance ratio of the Metropolis-Hastings algorithm and may slow down the process of sampling. Therefore, we decided to use the logarithms of the release rates as variables to avoid this dissymmetry.

Appendix D. Details about the derivation of the PT algorithm

In what follows and for the sake of readability, we define the inverse temperature $\beta_j = \frac{1}{T_j}$.

D.1 Single-temperature move step

This step is the classic move of a MH algorithm with the difference that each Markov chain converges to a different invariant distribution. The acceptance ratio $\alpha_{\mathbf{y}}^{T_j}(\mathbf{x'}_j|\mathbf{x}_j) = \alpha_{\mathbf{y}}^{\beta_j}(\mathbf{x'}_j|\mathbf{x}_j)$ corresponding to the acceptance of the proposal $\mathbf{x}'_j$ given $\mathbf{x}_j$ for the chain at inverse temperature β_j is:

$$\alpha_{\mathbf{y}}^{\beta_j}(\mathbf{x'}_j|\mathbf{x}_j) = \frac{\pi^{\beta_j}(\mathbf{x_j'})}{\pi^{\beta_j}(\mathbf{x_j})} = \frac{\pi(\mathbf{x_j'})^{\beta_j}}{\pi(\mathbf{x_j})^{\beta_j}} = \frac{p(\mathbf{y}|\mathbf{x_j'})^{\beta_j} p(\mathbf{x_j'})^{\beta_j}}{p(\mathbf{y}|\mathbf{x_j})^{\beta_j} p(\mathbf{x_j})^{\beta_j}}, \tag{D.1}$$

which does not depend on the normalisation constants of π^{β_j} $(= \pi^{T_j})$.

D.2 Swapping step

In this part, we describe the *temperature swap* step in more detail. We attempt to swap $\mathbf{x}_j$ and $\mathbf{x}_k$ the states of the chains at inverse temperatures β_j and β_k, respectively. The corresponding acceptance ratio $\alpha_{\mathbf{y}}^{T_j,T_k} = \alpha_{\mathbf{y}}^{\beta_j,\beta_k}$ is calculated exactly as was done for the *within temperature swap*:

$$\alpha_{\mathbf{y}}^{\beta_j,\beta_k}(\mathbf{x}_j,\mathbf{x}_k) = \min\left(1, \frac{\pi^{\beta_j}(\mathbf{x_k})\pi^{\beta_k}(\mathbf{x_j})}{\pi^{\beta_j}(\mathbf{x_j})\pi^{\beta_k}(\mathbf{x_k})}\right) = \min\left(1, \frac{\pi(\mathbf{x_k})^{\beta_j}\pi(\mathbf{x_j})^{\beta_k}}{\pi(\mathbf{x_j})^{\beta_j}\pi(\mathbf{x_k})^{\beta_k}}\right). \tag{D.2}$$

We derive the logarithm of the acceptance ratio, after simplification:

$$\begin{aligned} \ln\alpha_{\mathbf{y}}^{\beta_j,\beta_k}(\mathbf{x}_j,\mathbf{x}_k) &= -\beta_j\mathscr{J}_{\mathbf{y}}(\mathbf{x_k}) - \beta_k\mathscr{J}_{\mathbf{y}}(\mathbf{x_j}) + \beta_j\mathscr{J}_{\mathbf{y}}(\mathbf{x_j}) + \beta_k\mathscr{J}_{\mathbf{y}}(\mathbf{x_k}) \\ &= (\beta_k - \beta_j)\left(\mathscr{J}_{\mathbf{y}}(\mathbf{x_k}) - \mathscr{J}_{\mathbf{y}}(\mathbf{x_j})\right). \end{aligned} \tag{D.3}$$

Note that this step requires no computation time, since both $\mathscr{J}_{\mathbf{y}}(\mathbf{x}_k)$ and $\mathscr{J}_{\mathbf{y}}(\mathbf{x}_j)$ are already computed. In this way, the PT algorithm is almost as fast as running $n+1$ independent chains, which can be parallelised. Let us see what this formula implies for the probability of swapping the variables vector $\mathbf{x}_{k-1}$ of a chain at temperature T_{k-1} and the variables vector $\mathbf{x}_k$ of a chain at temperature T_k for $T_{k-1} < T_k$. Following the Metropolis-Hastings procedure, $u \sim \mathscr{U}(0,1)$ is generated and the swap is accepted if

$$\begin{aligned} \ln u &\leq \ln \alpha_{\mathbf{y}}^{T_{k-1},T_k}(\mathbf{x}_{k-1},\mathbf{x}_k) \\ &\leq (\beta_{k-1} - \beta_k)\left(\mathscr{J}_{\mathbf{y}}(\mathbf{x_{k-1}}) - \mathscr{J}_{\mathbf{y}}(\mathbf{x_k})\right). \end{aligned} \tag{D.4}$$

Since $\beta_{k-1} > \beta_k$, the swap is accepted automatically if $\mathscr{J}_{\mathbf{y}}(\mathbf{x_k}) \leq \mathscr{J}_{\mathbf{y}}(\mathbf{x_{k-1}})$. If the variables from the chain at a higher temperature are better (i.e., their cost is lower) then the two chains swap their variables. However, the swap can also be accepted if $\mathscr{J}_{\mathbf{y}}(\mathbf{x_{k-1}}) \leq \mathscr{J}_{\mathbf{y}}(\mathbf{x_k})$ and the difference is of the same order of magnitude as $\frac{\ln \mathscr{U}(0,1)}{\beta_{k-1}-\beta_k}$. Finally, note that the observation error term does not participate in this process, as it would change the temperature of the hot chains.

Appendix E. Issues with parallel tempering

In this section, we present the rationale behind our choice of a log-gamma prior distribution on the logarithms of the release rates. We compare the results of two Metropolis-Hastings algorithms with the same invariant distribution:

$$\pi_{\mathbf{y}}^{\frac{1}{T}} : \mathbf{x} \rightarrow p(\mathbf{x}|\mathbf{y})^{\frac{1}{T}} \left(\int p(\mathbf{x}|\mathbf{y})^{\frac{1}{T}} d\mathbf{x}\right)^{-1}, \tag{E.1}$$

with $T = 3^6 = 729$ (which is a very high temperature and, correspondingly, a very flattened distribution) but a log-gamma prior for the logarithms of

the release rates with two different scale hyperparameters: in the first run, we set $\theta = 30$ and in the second run, $\theta = \infty$. The MH chains run for 10^4 iterations: the convergence of the MH algorithm should be rapidly reached since the target distributions are supposed to be very flat.

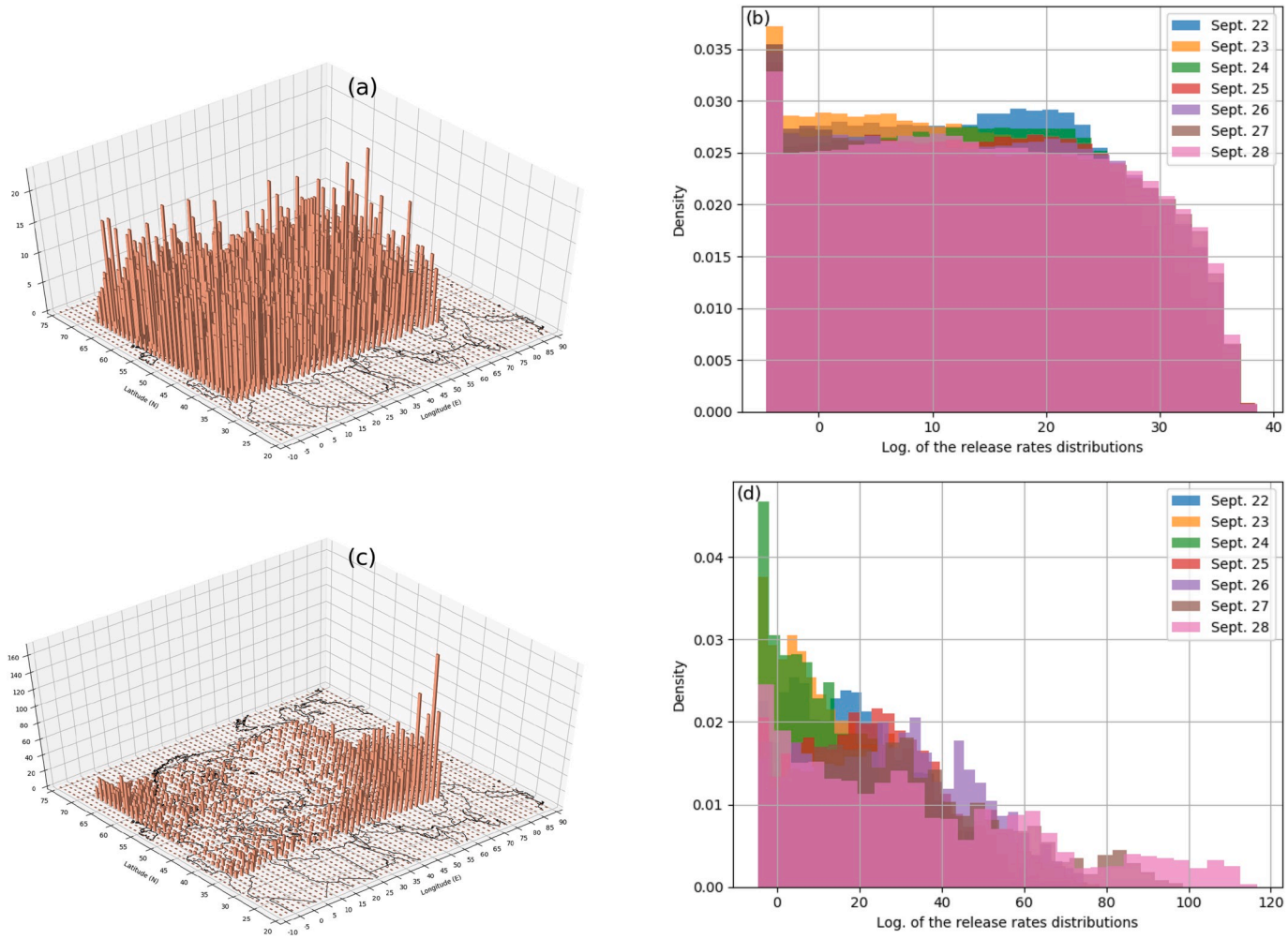

Figure E.1. Figures (a) and (b) represent the joint distribution of the coordinates and the distribution of the logarithms of the release rates respectively using a MH algorithm applied with a log-gamma prior with hyperparameter $\theta = 30$. Figures (c) and (d) show the use of a log-gamma prior with hyperparameter $\theta = \infty$.

The distributions sampled in the case $\theta = 30$ (Figs. E.1a and E.1b) show the expected behaviour: the joint distribution of the coordinates and the logarithms of the release rates are very flat. However, in the second case with $\theta = \infty$, Fig. E.1c indicates that chains are sampling mainly from an area located around [70E, 34N] and Fig. E.1d shows that the logarithms of the release rates reach very high values (up to a source term 10^{34} larger than that of the full magnitude of the Fukushima accident).

In the $\theta = 30$ case (Figs. E.1a and E.1b), the prior for the logarithms of the release (even flattened) rates is very binding: release rates are prevented from growing to excessively high values. In the other case (Figs. E.1c and E.1d), very high release rates are only constrained by the observations. However, the search domain is not entirely covered by observations: there are areas variable vector with high release rates $\mathbf{q}$, it then necessarily converges to certain spatial areas (here a domain extending from Turkey to Pakistan) that will not activate the monitoring networks to avoid producing large errors. As a result, a high temperature chain can remain frozen in local minima of coordinates. Since the convergence of our chain of interest (at s) is conditioned by the convergence of the chains at a high temperature, this is a big issue. Setting a high constraint using the log-gamma distribution with adapted θ solves this problem.

Appendix F. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aeaoa.2020.100071.

References

Abida, R., Bocquet, M., 2009. Targeting of observations for accidental atmospheric release monitoring. Atmos. Environ. 43, 6312–6327. https://doi.org/10.1016/j.atmosenv.2009.09.029.

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20, 407–415. https://doi.org/10.1093/bioinformatics/btg427.

Atchadé, Y.F., Roberts, G.O., Rosenthal, J.S., 2011. Towards optimal scaling of metropolis-coupled Markov chain Monte Carlo. Stat. Comput. 21, 555–568. https://doi.org/10.1007/s11222-010-9192-1.

Baragatti, M., 2011. Sélection bayésienne de variables et méthodes de type Parallel Tempering avec et sans vraisemblance. thesis, vol. 2. Aix-Marseille.

Bardsley, J., Solonen, A., Haario, H., Laine, M., 2014. Randomize-then-optimize: a method for sampling from posterior distributions in nonlinear inverse problems. SIAM J. Sci. Comput. 36, A1895–A1910. https://doi.org/10.1137/140964023.

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

Bocquet, M., 2012. Parameter-field estimation for atmospheric dispersion: application to the Chernobyl accident using 4D-Var. Q. J. R. Meteorol. Soc. 138, 664–681. https://doi.org/10.1002/qj.961.

Chow, F.K., KosoviÃ, B., Chan, S., 2008. Source inversion for contaminant plume dispersion in urban environments using building-resolving simulations. J. Appl. Meteorol. Climatol. 47, 1553–1572. https://doi.org/10.1175/2007JAMC1733.1.

Delle Monache, L., Lundquist, J., Kosović, B., Johannesson, G., Dyer, K., Aines, R.D., Chow, F., Belles, R., Hanley, W., Larsen, S., Loosmore, G., Nitao, J., Sugiyama, G., Vogt, P., 2008. Bayesian inference and Markov chain Monte Carlo sampling to reconstruct a contaminant source on a continental scale. J. Appl. Meteorol. Climatol. 47, 2600–2613. https://doi.org/10.1175/2008JAMC1766.1.

Earl, D.J., Deem, M.W., 2005. Parallel tempering: theory, applications, and new perspectives. Phys. Chem. Chem. Phys. 7, 3910. https://doi.org/10.1039/b509983h.

Geyer, C.J., 1991. Markov Chain Monte Carlo Maximum Likelihood. Interface Foundation of North America, p. 8.

Gilks, W.R., Richardson, S., Spiegelhalter, D., Richardson, S., Spiegelhalter, D., 1995. Markov Chain Monte Carlo in Practice. Chapman and Hall/CRC. https://doi.org/10.1201/b14835.

Groëll, J., Quélo, D., Mathieu, A., 2014. Sensitivity analysis of the modelled deposition of 137 Cs on the Japanese land following the Fukushima accident. Int. J. Environ. Pollut. 55, 67–75. https://doi.org/10.1504/IJEP.2014.065906.

Hazart, A., Giovannelli, J.F., Dubost, S., Chatellier, L., 2014. Inverse transport problem of estimating point-like source using a Bayesian parametric method with MCMC. Signal Process. 96, 346–361. https://doi.org/10.1016/j.sigpro.2013.08.013.

Hersbach, H., Bell, B., Berrisford, P., Horányi, A., Sabater, J.M., Nicolas, J., Radu, R., Schepers, D., Simmons, A., Soci, C., Dee, D., 2019. Global Reanalysis: Goodbye ERA-Interim, Hello ERA5, vol. 159, pp. 17–24.

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

Kovalets, I., Romanenko, A., 2017. Detection of Ruthenium-106 in 2017: Meteorological Analysis of the Potential Sources. https://doi.org/10.13140/RG.2.2.36537.67685.

Liu, Y., Haussaire, J.M., Bocquet, M., Roustan, Y., Saunier, O., Mathieu, A., 2017. Uncertainty quantification of pollutant source retrieval: comparison of Bayesian methods with application to the Chernobyl and Fukushima Daiichi accidental releases of radionuclides. Q. J. R. Meteorol. Soc. 143, 2886–2901. https://doi.org/10.1002/qj.3138.

Louis, J.F., 1979. A parametric model of vertical eddy fluxes in the atmosphere. Boundary-Layer Meteorol. 17, 187–202. https://doi.org/10.1007/BF00117978.

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

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

Nisbet, E., Weiss, R., 2010. Top-down versus bottom-up. Science 328, 1241–1243. https://doi.org/10.1126/science.1189936.

Pailleux, J.P., Geleyn, J.F., Legrand, E., 2000. La prévision numérique du temps avec les modèles ARPEGE et ALADIN.

Quélo, D., Krysta, M., Bocquet, M., Isnard, O., Minier, Y., Sportisse, B., 2007. Validation of the Polyphemus platform on the ETEX, Chernobyl and Algeciras cases. Atmos. Environ. 41, 5300–5315. https://doi.org/10.1016/j.atmosenv.2007.02.035.

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

Robert, C.P., Elvira, V., Tawn, N., Wu, C., 2018. Accelerating MCMC algorithms. Wiley Interdiscipl. Rev.: Comput. Stat. 10, e1435. https://doi.org/10.1002/wics.1435.

Roberts, G.O., Gelman, A., Gilks, W.R., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. Ann. Appl. Probab. 7, 110–120. https://doi.org/10.1214/aoap/1034625254.

Roberts, G.O., Rosenthal, J.S., 2014. Minimising MCMC variance via diffusion limits, with an application to simulated tempering. Ann. Appl. Probab. 24, 131–149. https://doi.org/10.1214/12-AAP918.

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

Saunier, O., Mathieu, A., Didier, D., Tombette, M., Quélo, D., Winiarek, V., Bocquet, M., 2013. An inverse modeling method to assess the source term of the Fukushima Nuclear Power Plant accident using gamma dose rate observations. Atmos. Chem. Phys. 13, 11403–11421. https://doi.org/10.5194/acp-13-11403-2013.

Schmehl, K.J., Haupt, S.E., Pavolonis, M.J., 2012. A genetic algorithm variational approach to data assimilation and application to volcanic emissions. Pure Appl. Geophys. 169, 519–537. https://doi.org/10.1007/s00024-011-0385-0.

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

Shershakov, V.M., Borodin, R.V., Tsaturov, Y.S., 2019. Assessment of possible location Ru-106 source in Russia in Septemberâ€“October 2017. Russ. Meteorol. Hydrol. 44, 196–202. https://doi.org/10.3103/S1068373919030051.

Swendsen, R.H., Wang, J.S., 1986. Replica Monte Carlo simulation of spin glasses. Phys. Rev. Lett. 57, 2607–2609. https://doi.org/10.1103/PhysRevLett.57.2607.

Sørensen, J.H., 2018. Method for source localization proposed and applied to the October 2017 case of atmospheric dispersion of Ru-106. J. Environ. Radioact. 189, 221–226. https://doi.org/10.1016/j.jenvrad.2018.03.010.

Tichý, O., Šmídl, V., Hofman, R., Stohl, A., 2016. LS-APC v1.0: a tuning-free method for the linear inverse problem and its application to source-term determination. Geosci. Model Dev. (GMD) 9, 4297–4311. https://doi.org/10.5194/gmd-9-4297-2016.

Tombette, M., Quentric, E., Quélo, D., Benoit, J., Mathieu, A., Korsakissok, I., Didier, D., 2014. C3X: a software platform for assessing the consequences of an accidental release of radioactivity into the atmosphere. In: Poster Presented at Fourth European IRPA Congress, pp. 23–27.

Troen, I.B., Mahrt, L., 1986. A simple model of the atmospheric boundary layer; sensitivity to surface evaporation. Boundary-Layer Meteorol. 37 (1), 129–148. https://doi.org/10.1007/BF00122760.

Winiarek, V., Bocquet, M., Saunier, O., Mathieu, A., 2012. Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant: application to the reconstruction of the cesium-137 and iodine-131 source terms from the Fukushima Daiichi power plant. J. Geophys. Res.: Atmosphere 117. https://doi.org/10.1029/2011JD016932.

Winiarek, V., Vira, J., Bocquet, M., Sofiev, M., Saunier, O., 2011. Towards the operational estimation of a radiological plume using data assimilation after a radiological accidental atmospheric release. Atmos. Environ. 45, 2944–2955. https://doi.org/10.1016/j.atmosenv.2010.12.025.

Yee, E., 2008. Theory for reconstruction of an unknown number of contaminant sources using probabilistic inference. Boundary-Layer Meteorol. 127, 359–394. https://doi.org/10.1007/s10546-008-9270-5.

Yee, E., Hoffman, I., Ungar, K., 2014. Bayesian inference for source reconstruction: a real-world application. Int. Sch. Res. Not. https://doi.org/10.1155/2014/507634, 2014.