Reply to Anonymous Referee #1

Dear Reviewer 1,

We would like to thank you for your review. We believe that your comments and suggestions will help us to improve our manuscript. Please find below a step-by-step reply to your comments and suggestions.

Yours faithfully,
The authors

> *The proposed manuscript presents important question regarding uncertainties associated with inverse modelling problems for unknown atmospheric releases. It is used ensemble approach which is applayed to the ruthenium 106 case from 2017. The study is interesting and clear to read, but more questions remain in conection with the selected case and its suitability for the cause. Although the goal of the paper is not to find the origin of the ruthenium in 2017, the estimated probability regions of the release far from the actual release site need more discussions. It is difficult to draw and to follow conclusions based on results inconsistent with state-of-the-art knowledge about the ruthenium case. Perhaps, wider ruthenium dataset or another case with known location and known release profile could be more appropriate for this type of study.*

Specific comments:

> *Although there is dataset with hudreds positive measurements regarding Ru-106 case in 2017 available (Masson et al., 2019), the authors choose data from 5 locations with12 positive measurements. This is quite suprising and authors should comment this. Moreover, the choice of CTBT stations seems problematic in this case since the main activities have been observed around Mayak and then south-west-wind in Ukraine, Romania etc. (Masson et al., 2019). Hence, my opinion is that the used data can contain rather fractions of information about the release and the results are dominated by the fact that the release period is preselected in the algorithm. This is probably closely related also to the fact that the probable location, Mayak, is not estimated within the probability region in any case (in fact, Dimitrovgrad is much more probable in all cases). This should be discussed in the paper.*

Reply:

There are several reasons for using only a small subset of all the available data:

1. Many observations have redundant information so that including these additional observations would not significantly affect the results.

2. If we would have used observations taken nearby the Mayak facility, indeed the source location could be much more confined. However, the meteorological data and the EDA ensemble that were used in this study are not appropriate for use on local scales, as the ensemble does not capture well local scale uncertainties.

3. In the CTBTO verification context, we typically have available sparse observations around the world at receptor locations from the International Monitoring System that are about O(1,000 km)

apart from each other. Hence, we are dealing with long-range inverse atmospheric transport problems. As this work was carried out in the context of the CTBT, we have chosen to use (a subset of) observations from the IMS.

4. Furthermore, our results are in agreement with other inverse long-range atmospheric transport modelling studies as it shows a broad region in the Southern Urals as possible source location (see for instance Fig 1 of Saunier et al 2019).

5. Another point to take into account is the computational effort associated to ensemble atmospheric transport modelling. If we would use 1,000 observations, this would require 51,000 Flexpart simulations if we assume that the source location is unknown.

We thank you for your comment since indeed we did not discuss these reasons in our original manuscript. We will add the above discussion in the revised manuscript.

> *p. 6, l. 131: The authors claimed that "the release rate is assumed constant during the release period". This assumption seems to be quite strong since the release rate may vary and, in this particular case of Ru-106 release, did vary during the time as estimated by e.g. (Saunier et al., 2019). Is this assumption necessary and what is the impact of it?*

Reply:

It is important to distinguish between different geotemporal scales. While time-varying emissions can have a huge impact nearby the source, these effects are less significant further away from the source due to the atmospheric transport and dispersion processes (and the atmospheric transport model, which filters such information out). Hence, we expect a constant release within release parameters t_start and t_stop to be appropriate to describe the Ru-106 source. In the version of the FREARtool that is described here, the source is parameterised as a fixed release during a certain period. A more recent version of the FREARtool can also deal with time-varying emissions.

> *p. 6, l. 141: The authors assume that "the release is assumed to have occurred between 25 September 2017 0000 UTC and 28 September 2017 0000 UTC", however,the release was estimated before e.g. in (Saunier et al., 2019; Western et al., 2020).Could you, please, comment this choice?*

Reply:

According to Figure 2 of Saunier et al 2019, the release that occurred between 25 and 28 September 2017 is about three orders of magnitude larger than the release outside that time period. Since our source parameterisation deals with a fixed release rate between a certain period, and we do not consider local observations which might be affected by smaller releases that potentially took place before 25 September, we believe it is appropriate to neglect smaller emissions that potentially occurred earlier (or later).

> *p. 6, l. 154: The Currie critical threshold,LC, is used extensively in the paper. Could you please briefly explain basics about this value?*

Reply:
We will add to the revised manuscript:

“Currie (1968) defined a critical level L_C above which a net signal (which is the detected signal from which the effects of background radiation are subtracted) should be in order to declare the net signal to be a detection.”

Note that, without statistical fluctuations, one could simply use L_C = 0, since any positive net signal would be a detection. However, as this is not the case, we set L_C > 0 in order to reduce the number of false positives.

> *p. 9, l. 203: σobs seems to be fixed in your scenario. Do you have uncertainties associated with measurements? How are these uncertainties related to this σobs value?*

Reply:

σ_obs is specific for each observation and is the measurement uncertainty reported by the International Monitoring System. However, there was an inaccuracy in L 203 of the original manuscript, where we wrote:

> "The detected activity concentration c_det is assumed to be Gaussian distributed around the true activity concentration c_true , with a standard deviation σ_obs = L_C /k_α as in Eq. 14."

This is only valid in case of a true non-detection or a false alarm. We have rewritten that sentence as follows:

> "The detected activity concentration c_det is assumed to be Gaussian distributed around the true activity concentration c_true , with a standard deviation σ_obs equal to the reported observation error."

The latter makes it also more clear that σ_obs is not a fixed value.

Technical corrections:

> *It is not necessary to have new paragraph after each equation.*

Reply:

Thank you for pointing this out.

> *The abraviation FREAR(in title) is not used and define in the manuscript.*

Reply:

Thank you for pointing this out. We will add to the revised manuscript:
“FREAR stands for Forensic Radionuclide Event Analysis and Reconstruction”

> *p. 8, l. 189: consider to remove"used".*

Reply:

Thank you for your suggestion.

p. 10, l. 219: there is no si in Eq. (13), please, clarify.

Reply:

Thank you for pointing this out. Eq. (13) is simply Eq. (5) but with c_true replaced by c_det,i. We will also add a reference to Eq. (5) in the manuscript for clarity.

Sec. 4.1: you should specify that this is related to the Fig. 3, LEFT

Reply:

Indeed, thank you for this remark. We will specify this in the revised manuscript.

Reply to Referee #2

Dear Reviewer 2,

We would like to thank you for your in-depth review and interesting thoughts. We believe that your comments and suggestions will help us to improve our manuscript. Please find below a step-by-step reply to your comments and suggestions.

Yours faithfully,
The authors

L 21 – What do the the authors call "anomalous radionuclide detections"? That is something I am perfectly aware of, but it is perhaps not the case of all readers.

Reply:

Thank you for this suggestion. We will add the following definition to the manuscript:

“Anomalous radionuclide detections are detections of anthropogenic radionuclides originating from upwind nuclear facilities, where the detected concentration of (a) specific radionuclide(s) and/or the combination of several detected radionuclides are anomalous with respect to the station’s detection history and/or with respect to what can be expected from these upwind nuclear facilities operating under normal conditions.”

L 25 – According to the authors, atmospheric transport and dispersion modelling is "one of the methods" to relate detections and the source of emission. I do not see other methods. Which other methods do the authors have in mind?

Reply:

In theory, ratios of specific radionuclides (if these are all detected in a certain sample, and assuming no contamination from other sources) could help to discriminate between different sources, without using an atmospheric transport model.

L 30 – In backward modelling, the source-receptor relationships are calculated from fixed receptors to potential sources (not the opposite as written in the sentence in L 30).

Reply:

We will correct this ambiguity in the revised manuscript.

L 32 – The concept of "non-detection" should be explained (or ignored as it is not used in the paper).

Reply:

We will add to the revised manuscript (in green):

“Statistical methods can then be employed to combine the information from all these detections (and possibly non-detections - observations where the activity concentration is below a minimum detectable concentration) in a meaningful way in order to infer relevant information on the source.”

L 44 – In this paper, the model error is considered as a whole. Thus, it does not originate only from the numerical weather predictions, but also from the atmospheric transport and dispersion model. The word "mainly" ("because of the underlying weather prediction data") is questionable. The authors should consider rephrasing the sentence.

Reply:

In our experience, the NWP data results in the largest uncertainty in atmospheric transport modelling using the Flexpart model. In Flexpart, the NWP data determines the transport (by the wind) and dispersion (through parameterisation using atmospheric stability) of particles. Source uncertainties are not applicable here, since we work backward in time. In our experience, Flexpart is fairly robust against perturbations of the Flexpart model parameters.

There is also literature that supports our claim in L 44. We will add these references in the revised manuscript:

Engström, A., & Magnusson, L. (2009). Estimating trajectory uncertainties due to flow dependent errors in the atmospheric analysis. *Atmospheric Chemistry & Physics*, *9*(22).

Harris, J. M., Draxler, R. R., & Oltmans, S. J. (2005). Trajectory model sensitivity to differences in input data and vertical transport method. *Journal of Geophysical Research: Atmospheres*, *110*(D14).

Hegarty, J., Draxler, R. R., Stein, A. F., Brioude, J., Mountain, M., Eluszkiewicz, J., ... & Andrews, A. (2013). Evaluation of Lagrangian particle dispersion models with measurements from controlled tracer releases. *Journal of Applied Meteorology and Climatology*, *52*(12), 2623-2637.

However, we welcome findings or literature from the Reviewer that would contradict or complement the above and remain open to adapt that part of the manuscript accordingly.

L 54 – As for me, it is difficult to create and use a relevant ensemble. The reason is not only (and perhaps not mainly) the computational cost of the ensemble, but the way to constitute it with enough variety, limited redundancy, etc. This complex task should be mentioned in the paper.

Reply:

We agree with that and propose to add the following to the revised manuscript:

“Creating an ensemble with a meaningful spread between its different members (that is, spread which represents the model uncertainty) is a very complex task which requires expert knowledge of all data, processes and their associated uncertainties at each level of the modeling process.”

L 57 – Ditto. It is complicated and not guaranteed that an ensemble captures "most of the possible outcomes". This should be indicated in the paper.

Reply:

We propose the following rewording in the revised manuscript:

“Therefore, ensembles used operationally at major weather institutes around the world are designed in a way that, even with a limited number of members (between 14 and 50, Leutbecher, 2019), the ensemble ~~can capture most~~ tries to capture all (and not more) of the possible outcomes.”

L 59 – What is a "measurement model"?

Reply:

We will refer to Eq. 19 and Eq. 20 in the revised manuscript, and will write that “a measurement model relates the model variable with the observation.”

L 88 – The description of the detections should be gathered in a table with the collection start and stop times (even if I guess that the authors do not wish to develop this aspect of the data).

Reply:

You are correct and this is a helpful suggestion. The text from the paragraph has been reworked to present the data in tabular format.

L 96 – The beginning of the sentence is "the above observation times". I do not see any observation times above?

Reply:

In that paragraph, we mentioned when the observations were made (L 89 – 92 in the original manuscript). However, we will clarify this in the revised manuscript.

L 101 – It is written that FLEXPART is run in backward mode. I wonder how long the simulations go back in time. Could the authors give information about this?

Reply:

We will add to the revised manuscript:
“All simulations ended on 20 September 2017.”

L 110 – It is not obvious that adding and substracting perturbations from an ensemble mean are a legitimate process. Could the authors comment on this?

Reply:

This was motivated by the idea that the unperturbed member could perform slightly better than the perturbed members, so that better results could be obtained by centering the perturbations around the unperturbed member rather than around the ensemble mean.

L 113 – The authors assert that "the spread between the different members represent

the uncertainty". This is undoubtedly a way to account for uncertainty in weather predictions, but are the authors sure that the ensemble perfectly encompasses the uncertainty on the meteorological data? The authors should consider being more cautious and rephrasing this sentence.

Reply:

With that, we rather meant the general principle of an ensemble: viz. the spread between the members represents the uncertainty. Of course, a bad ensemble will result in a bad uncertainty estimate. We propose to rewrite it as follows:

"The perturbations are created in such a way that each ensemble member represents a possible scenario for the true (unknown) atmospheric state, and the spread between the different members ~~represents the uncertainty~~ is simply the model uncertainty as estimated by the ensemble."

L 130 – What are the values of t1 and tm, the first and last time for which source-receptor-sensitivities are available for the source reconstruction?

Reply:

This is discussed in Subsection 3.2 "prior distribution": t_1 is 25 September 2017 0000 UTC and t_m is 28 September 2017 0000 UTC. (Flexpart output files were available for other times too.)

L 131 – The authors assume that the release rate is constant during the release period. I would like to point out that this is a strong assumption as in principle, the release is not known at all. Could the authors comment on this?

Reply:

We repeat our answer to Reviewer 1, who made a similar comment:
"It is important to distinguish between different geotemporal scales. While time-varying emissions can have a huge impact nearby the source, these effects are less significant further away from the source due to the atmospheric transport and dispersion processes (and the atmospheric transport model, which filters such information out). Hence, we expect a constant release within release parameters t_start and t_stop to be appropriate to describe the Ru-106 source."

See also:

De Meutter, P., Camps, J., Delcloo, A., Deconninck, B., & Termonia, P. (2018). Time resolution requirements for civilian radioxenon emission data for the CTBT verification regime. *Journal of environmental radioactivity*, *182*, 117-127.

*L 138 – The total release is assumed to be between 10**10 and 10**16 Bq. This seems to me somewhat arbitrary as it excludes potential releases respectively further downwind and further upwind. Once more, how to proceed when no preconceived solution is available? Could the authors consider commenting on this point?*

Reply:

From the available number of measurements, and the scale at which detections were made, these bounds are not unrealistic. Smaller sources would not have been seen over such a broad geographic area, while larger sources would have been seen at more monitoring locations. The selected bounds represent a conservative, but realistic bound for the source. Furthermore, we have already applied inverse modelling using a cost function approach for this case, which allowed us to make our prior distributions sharper than what can be done without knowledge on this case; please see:

De Meutter, P., Camps, J., Delcloo, A., and Termonia, P.: Source Localization of Ruthenium-106 Detections in Autumn 2017 Using Inverse Modelling, in: Mensink C., Gong W., Hakami A. (eds) Air Pollution Modeling and its Application XXVI. ITM 2018. Springer Proceedings in Complexity., Springer, Cham, https://doi.org/10.1007/978-3-030-22055-6_15, 2020.

L 141 – Ditto. How did the authors choose the time interval of the release (all the more that this time interval is quite short)?

Reply:

(Please also see our reply to your previous comment.) From earlier studies, we knew that the bulk release of Ru-106 likely took place between that period. Since a detailed analysis of the Ru-106 case was not our intention, we have chosen to focus on this time period. An additional benefit of reducing the allowed time interval of the release (when fixing the spatial domain) is that it reduces the memory requirements, which is beneficial when running the case on a personal computer. (Note, however, that the tool can also be run on a server or cluster where more memory is available.)

L 148 – This is another strong hypothesis that the observations are independent while there is likely a space and time dependency between them. Could the authors comment on this?

Reply:

We acknowledged in the manuscript that this is a simplification. Given the large distance (~ 1000 km) between different IMS stations, we believe this approximation is not too incorrect. Furthermore, the authors are not aware of similar studies that take into account geotemporal dependencies between observations, and we would be grateful if the Reviewer could provide some references.

L 160 – Does the index "i" in formula (5) indicate that there are as many applications of this formula (with possibly different values of the s, alpha bar and beta bar parameters) as the number of observations?

Reply:

Eq. 5 is indeed for a single observation. The values for s, alpha bar and beta bar can be made observation-specific (which is also done further in the paper).

L 189 – I wonder if the general-purpose Markov Chain Monte Carlo algorithm MT-DREAM(ZS) is freely available? Who developed this MCMC method?

Reply:

It was developed by Laloy and Vrugt and described in their paper Laloy and Vrugt (2012). Some implementations of DREAM can be found in open source packages on the internet.

Figure 2 – I suppose that "MDC" stands for "Minimum Detectable Concentration" and that we have LC # MDC / 2. In the formulae, it seems that only LC is used. Could the authors confirm this point?

Reply:

In the formulae, L_C is used. With the observations, typically the MDC is reported and not L_C. For the observations in the IMS network of CTBTO, we can assume that L_C = MDC/2.

L 192 – While popular, MCMC methods have well-known drawbacks like the burn-in period or convergence problems. Could the authors consider commenting on this with respect to the MT-DREAM(ZS) algorithm?

Reply:

This depends on the case, but from the authors' experience over the past year, we typically run the tool using ~ 10,000 iterations and convergence occurs after ~ 2,500 iterations (where we discard these first 2,500 iterations). In our previous study however, (De Meutter and Hoffman, 2020) where we studied the Se-75 release, we used 150,000 iterations.
The required number of iterations is also affected by the choice of the uncertainty "s": lower uncertainties require more iterations before convergence takes place.

L 197 – I have the feeling that all technical details in the last part of this paragraph (and notably the "snooker step") would need some more explanations as this part of the text is too concise (and a bit obscure).

Reply:

Regarding the snooker step, we were informed by one of the developers of MT-DREAM(ZS) that the snooker step is theoretically not compatible with the multiple-try part of the algorithm, so that we no longer use the snooker step. The difference in the posterior after using and not using the snooker step is not noticeable in our simulations.

To prove the latter, please find the results below for two simulations for the Ru-106 case, with and without the snooker step:

1/ simulation with the snooker step for the unperturbed member and s_i = 0.5

```
 Running MT-DREAMzs, iteration 7800 of 50001 . Current logp -37.44259 -41.24711 -39.54531
Converged after 7800 iterations
 Running MT-DREAMzs, iteration 50001 of 50001 . Current logp -36.48064 -44.17845 -41.44014
 MT-DREAMzs terminated after 1206.558 seconds
Acceptance rate for chain 1 is 22.24%
Acceptance rate for chain 2 is 22.61%
Acceptance rate for chain 3 is 22.93%
           lon      lat  log10_Q              rstart               rstop
0.025 50.11799 55.50922 14.96976 2017-09-25 00:22:32 2017-09-26 23:34:40
0.5   51.09007 55.91466 15.27527 2017-09-25 07:59:14 2017-09-27 18:17:46
0.975 57.88037 60.75305 15.64360 2017-09-25 22:55:13 2017-09-27 23:34:05
mean  51.98106 56.39979 15.28396 2017-09-25 08:46:06 2017-09-27 16:41:15
```

2/ simulation without the snooker step for the unperturbed member and s_i = 0.5

```
 Running MT-DREAMzs, iteration 12300 of 50001 . Current logp -44.62895 -44.45257 -41.44722
Converged after 12300 iterations
 Running MT-DREAMzs, iteration 50001 of 50001 . Current logp -43.13082 -39.71556 -37.64025
 MT-DREAMzs terminated after 1271.046 seconds
Acceptance rate for chain 1 is 25.26%
Acceptance rate for chain 2 is 24.75%
Acceptance rate for chain 3 is 23.59%
           lon      lat  log10_Q              rstart               rstop
0.025 50.09579 55.51231 14.97063 2017-09-25 00:30:47 2017-09-26 23:01:53
0.5   51.03933 55.88155 15.26998 2017-09-25 08:00:29 2017-09-27 18:35:54
0.975 57.74679 60.13044 15.65440 2017-09-25 22:38:03 2017-09-27 23:37:18
mean  51.85432 56.30864 15.27988 2017-09-25 08:42:32 2017-09-27 16:59:44
```

The following information will be added to the revised manuscript:

~~"The algorithm is designed so that a snooker step occurs with a probability of 20 % to allow jumps between different posterior modes (ter Braak and Vrugt, 2008).~~ To enhance efficiency and to obtain more accurate results, randomized subspace sampling is used (Vrugt et al., 2009). This simply means that not necessarily all source parameters are updated at a time, but instead a randomized subset of the source parameters. Furthermore, MT-DREAM (ZS) makes use of multiple try Metropolis sampling (Liu et al., 2000) to enhance the mixing of the chains. This means in practice that, to advance to Markov chain, several proposals are drawn instead of one proposal in traditional Metropolis sampling. Furthermore, the Metropolis acceptance is calculated in a different way (Liu et al., 2000 , Laloy and Vrugt, 2012)."

L 209 – It is written here that "s" is an estimate of "sigma", but "sigma" is not defined, nor introduced before. Should the reader understand that sigma stands for sigma_mod?

Reply:

Thank you for pointing this out. In L 209, "sigma" should have been "sigma_mod". Note however that in L 222, "sigma" stands for sqrt(sigma_mod^2 + sigma_obs^2). We will add that to the revised manuscript.

L 215 – In formula (17), "sigma_srs" and "srs" are not defined. What do these notations stand for? Moreover, what is the reason for the multiplicative value of 16 (and not another value) in the same formula? Could the authors comment on this?

Reply:

Thank you for pointing that out. "srs" stands for source-receptor-sensitivities (the model output when Flexpart is run in backward mode), and "sigma_srs" is its (unknown) uncertainty. We will add that to the revised manuscript.

The value of 16 is an empirical number that was found to give a good balance between information obtained from detections versus information from non-detections from an earlier case study described in De Meutter and Hoffman (2020). We will add this information in the revised manuscript.

L 216 – The sentence: "as a consequence, the model uncertainty does not depend on the source parameters" is especially unclear or unprecise. What do the authors call "the model"? Is it the weather prediction or the transport and dispersion simulation or

both? As the source parameters are not considered as uncertain, I do not see why and how they should take part in the model uncertainty. Please, consider rephrase this sentence.

Reply:

We can calculate the modeled activity concentrations c_mod as a linear relation between the source-receptor-sensitivities (srs) and the release amount Q:

c_mod = srs(release period, release location) * Q

One way of calculating the uncertainty on c_mod (sigma_c_mod) would then be to use sigma_srs, which could be obtained from the ensemble:

sigma_c_mod = sigma_srs(release period, release location) * Q

However, in that case, sigma_c_mod will depend on the source parameters (the release period, the release location and Q). This resulted in undesired effects in the very beginning of the development of the FREARtool (such as: the model selecting very high Q so that the uncertainty became very large, thereby allowing values of c_mod that did not agree at all with c_obs), so that it was decided to make the model uncertainty independent of the source parameters.

To avoid confusion, we propose to omit this sentence.

L 218 – I wonder how "a part of the plume" can be "subject to more atmospheric transport and dispersion processes". All parts of the plume are subject to atmospheric transport and dispersion processes. Small detections may be obtained at the "edge" of the plume or just far from the source of the release. What does a "small" detection mean? It is just a matter of detection method and device. While I globally agree with the ideas contained in this paragraph, I feel that they should be formulated in a different way.

Reply:

We propose the following revision:
L 218: "This is desirable since small detections are caused by a part of the plume of radionuclides that was subject to more atmospheric ~~transport and dispersion processes~~ dilution..."

L 226 – The whole section 4 uses the ECMWF unperturbed weather prediction. This should be mentioned at the beginning of the section.

Reply:

Indeed, thank you for pointing this out.

L 229 – As I understand "s_i" includes the model error and the observation error. I wonder what the respective parts of each kind of errors are. Could the authors comment on this? The authors present the source location probability map for three values of "s_i". Of course, it is difficult to choose this parameter and it is the central question which the paper deals with. Is it possible for the authors to motivate the choice of the three "s_i" values? Finally, it is written that "the same value s_i is used for all observations". I wonder why different values of s_i should be associated to the observations as the

observation error is by assumption the same for each observation and the model error should depend intrinsically on the model and not on the observation.

Reply:

The interpretation of the different s_i values is straightforward from Eq. 17: it represents a relative error of 30 %, 50 % and 300 % with respect to max(c_det, 16 * L_C). 50 % was our initial "default value". 300 % seemed a good value to go above that (we also tested other values, such as 100 % and 1 000 %). The choice for the lower value is limited by the observation error (lower values for sigma_total would imply an imaginary model error). Furthermore, some members had troubles with convergence when very small s_i values were chosen (10 %).

The observation error is different for each observation as it depends on the background radiation, the sampled volume of air etc… We believe that the model error should also be observation-specific, please see our reply further below to a comment regarding L 347.

Figure 3 – The figure 3 as the following figures seem to me a bit small.

Reply:

We will increase the figure size in the revised manuscript.

L 237 – I do not see what is an "unknown error"? There are observation errors, representativeness errors or model errors including among others the atmospheric processes not resolved by the model. What is "unknown" is not the type of error, but the value to be attributed to the error.

Reply:

In the revised manuscript, we will make the following change:

"~~Besides being an alternative model error, multipliers could also be used to take into account unknown errors (such as errors due to local atmospheric features not resolved by the model).~~"

"Besides being an alternative model error, multipliers could also be used to take into account errors that were not fully captured by the model (such as errors due to local atmospheric features not resolved by the model, measurement errors due to sample inhomogeneity, etc.)"

L 270 – Increasing the value of the parameter s_i results in a shift and an enlargement of the posterior distribution. I wonder why introducing multiplier only results in a shift of the posterior. I suppose that it acts as another way to adjust the posterior without any increase in the level of model uncertainty. Could the authors comment on this?

Reply:

That sounds certainly plausible. The model uncertainty is indeed not affected by the multipliers. The multipliers allow a better match between "m * c_mod" and the observations "c_obs". This better agreement can in theory be obtained with the same source parameters when no multipliers are used (thus, no shift will be seen), or it can be obtained with different source parameters (so that a shift will be seen if the source location is affected).

L 272 – I presume that forcing the model uncertainty with a high value of the parameter s_i predominates against the influence of the multipliers. Do the authors have the same explanation?

Reply:

If the model uncertainties (determined by s_i) are larger than "|c_mod – c_obs|", then indeed the multipliers will have less impact on the posterior.

L 281 – As for me, it is not so obvious that the errors arising from the meteorological input data have the "largest contribution" to the total model error. Would the atmospheric transport and dispersion model be a "bad model" (what is probably not the case of FLEXPART), the dispersion model error would not be negligible. The authors should perhaps moderate their assessment in L 281.

Reply:

Please see also our reply to a related comment concerning L 44. We propose the following (minor) moderation but remain open to consider further moderation if the Reviewer could share findings or literature that shows its necessity.

"While this type of error ~~arguably~~ likely adds the largest contribution to the total model error, other sources of model error are not included."

L 285 – How the data of all grid boxes is aggregated should be more explained. For me, it is not an obvious process.

Reply:

We will add the following in the revised manuscript (in green):

"In order to obtain the error structure, the data of all spatial grid boxes is aggregated into an uncertainty distribution."

Furthermore, we will add to the list:

> "4. The remaining data points are used to make an uncertainty distribution (as in Figure 4)."

L 298 – The probability density function of the SRS members should be presented not only for "an arbitrary observation and an arbitrary time" as in Figure 4, but for other observations and times or all distributions should be considered and their moments computed.

Reply:

It is not feasible to plot the distribution for each time and each observation (288 in total) in one figure, but we will add this as supplementary information.

Figure 4 – There is a typo in the caption: "distributed" versus "distribution".

Reply:

Thank you for noticing this. We will correct this in the revised manuscript.

L 321 – I wonder about the generality of the method presented by the authors, especially in case 4 when the parameters are fitted for each observation and time. As a matter of fact, it means that just adding or removing a detection will not only influence the source term estimate, but also the uncertainty on this estimate (and this with the same meteorological fields). Could the authors comment on this?

Reply:

It is not only the meteorological fields that determine the uncertainty, but also the trajectories that particles follow along these meteorological fields. As a result, the model uncertainty is observation-specific, and indeed, adding or removing observations can alter the uncertainty on the inferred parameters. See also our next reply.

L 347 – Considering "observation-specific" uncertainty parameters is an ad hoc (and interesting) way to fit the model (and observation) error, but it should not be forgotten that the model error should be an intrinsic feature of the model and not depend on the set of observations which is taken into account. I suggest that the authors argue on this.

Reply:

We do not agree that the model error should be an intrinsic feature of the model and does not depend on the observation: the model uncertainty depends on the trajectory of the retro-plume (= the plume that goes from the sampling station backward in time). Observations of a plume that are made three weeks after the release should have higher model uncertainty than observations made two days after the release. Also, depending on the weather conditions along the trajectory, the model error can be observation specific (consider transport associated with a frontal system versus transport associated with the calm conditions found in an anticyclone).

To clarify this, we will add (see text in green) to the revised manuscript:
L341: "In this subsection, it is assessed how the fitted uncertainty parameters vary among different observations and different times. The motivation for this is as follows: first, and somewhat trivial, we can expect the model uncertainty to increase as a function of simulation time. Second, uncertainties are expected to be observation-dependent, since observations are made on different times and at different distances from the source; uncertainties on the trajectories between the receptor and the source will also be affected by the atmospheric conditions along the trajectory, which are expected to be observation-specific."

L 350 – That the model uncertainty grows when going backwards in time is somewhat trivial. At least, the contrary would be surprising.

Reply:

We agree, but it is always good to confirm that our ensemble of atmospheric transport simulations replicates evident features.

L 353 – It is worth noticing that the oscillations have a circadian period. Is it possible to relate them with the day and night alternation of the boundary layer?

Reply:

We believe that L 351 made that notice:
"Also interesting to note is that there is an oscillatory behaviour with a period of eight time steps, corresponding to the diurnal cycle (since SRS fields were produced every three hours). The oscillations are likely associated with boundary layer processes, which often follow the diurnal cycle."

L 365 – It is quite optimistic to assert that both maps in Figure 7 roughly agree. There are many differences. Would the location of the release be the aim of the study, the authors would be certainly quite embarrassed to designate it using one map or the other.

Reply:

Indeed, but we assume the output of the inference will be interpreted by an expert, who is aware that models have uncertainties, and that even the uncertainties are uncertain. Also take into account that we zoom in into the area of interest. If we would plot the full domain, the differences will appear smaller.

L 390 – I would like to point out that there is an interesting result in L 390. As a matter of fact, using the ensemble only to fit the uncertainty parameters or running all members of the ensemble to figure out the uncertainty seems to be equivalent.

Reply:

We believe L 389 in the original manuscript mentions this, but we will try to make it more explicit by adding (in green):

"It seems that overall, a similar picture is obtained when running the Bayesian inference for each ensemble member separately, compared to the procedure explained in Section 5. This suggests that if we use the ensemble only (i) to fit the uncertainty parameters and (ii) to calculate the ensemble median SRS for running the inference as was done in order to obtain Fig. 7, no crucial information from the ensemble is lost with respect to the source location. As a consequence, it is equivalent to running the inference with all members of the ensemble separately to determine the uncertainty."

L 410 – As a conclusion, I would suggest to the authors to apply the different approaches and methods presented in their paper to situations in which the source characteristics (especially the location) is known unambiguously (because in the Ru-106 case the source location was not really recognized). In a situation with a clearly identified location of the emission, it would be interesting to see what results (good or less good) are obtained using the inference in different ways, and also what is the most efficient approach.

Reply:

Thank you for this suggestion, which is in line with the comments made by Reviewer 1. We will add to the conclusions:
“In a future study, we will apply the different approaches and methods presented in this paper to situations in which the source characteristics are known unambiguously. This will help to better evaluate the different approaches proposed in this paper.”

L 435 – As argued by the authors, it seems that using the members of an ensemble in the source term estimate gives more robust results with regard to the choice of the uncertainty parameter as opposed to not using any ensemble. It seems to me quite logical as the ensemble introduces a kind of uncertainty (which is certainly not all the uncertainty, but a “rigorously built” uncertainty). This uncertainty may predominate against the uncertainty arbitrarily fixed by choosing the uncertainty parameter.

Reply:

We agree with that, and propose to add that in the revised manuscript (in green):

“A scenario-based approach (where each ensemble member is used as input for the Bayesian source reconstruction, instead of using the ensemble to fit the uncertainty parameters) gives results which are more robust against the choice of the uncertainty parameters but is more costly compared to directly fitting the uncertainty parameters. This is because the ensemble introduces model uncertainty that may predominate against the uncertainty prescribed by arbitrarily choosing the uncertainty parameter.”

Reply to Anonymous Referee #3

Dear Reviewer 3,

We would like to thank you for your review. We believe that your comments and suggestions will help us to improve our manuscript. Please find below a step-by-step reply to your comments and suggestions.

Yours faithfully,
The authors

General comments:

While the paper relies on a set of ensemble simulations to quantify the model uncertainties for the emission inverse modeling study, it is helpful to include ensemble in the title.

Reply:

Thank you for this suggestion, we propose to revise the title as follows:

"On the model uncertainties in Bayesian source reconstruction using an ensemble of weather predictions, the emission inverse modelling system FREARtool v1.0 and the Lagrangian transport and dispersion model Flexpart v9.0.2"

This paper emphasizes on the spatial patterns of the reconstructed sources. Since the sources also possess the temporal patterns, it is better to describe briefly what the reconstructed sources appear in time. How do the release start time and end time vary with the different approaches in this paper?

Reply:

It is true that we emphasized on the source location; the reason for doing so is that the release location is of primary interest in the context of CTBT verification. Once a location is found (for instance, based on the location of known nuclear facilities within the posterior source region, or based on a seismic signal associated to a nuclear explosion), a new inference could be performed fixing the release location as was done in De Meutter and Hoffman (2020).

We will add this reasoning to the revised manuscript.

Specifics:

Title: FREARtool in the title is never mentioned in text. In the Code and data availability part, it is stated that the "Bayesian inference tool will be made available upon request". If this tool is not mature enough to be available publicly, it is better not to appear in the title.

Reply:

Thank you for noticing this, we will mention ‘FREARtool’ in the text. We will make the tool publicly available after a formal announcement during an in-person CTBT-related event, which was unfortunately not possible so far due to the COVID crisis.

Line 5: It is not clear what the authors mean by “credible intervals”. Is “interval” used to represent the range of emission rates in magnitude? Please clarify this.

Reply:

“Credible interval” is a term used in Bayesian inference and represents the range in which an unobservable parameter falls with a particular probability. A credible interval is thus available for each of the five source parameters that are inferred: the source longitude, source latitude, total emission, release start and release end.

Line 103: It is not accurate to say “model output frequency was three hours”. In addition, the output can be instantaneous or time-averaged quantities. This needs to be clarified.

Reply:

Indeed, we will revise this as follows:

“The ~~model output frequency was~~ time-averaged source-receptor-sensitivities were output every three hours, so that the maximum possible residence time in a geotemporal grid box is 10 800 s.”

Line 105: The emission grid and the concentration grid can be different. Please specify which “grid box” is referred here.

Reply:

Since Flexpart is a Lagrangian particle model, there is no “emission grid”: particles are released from point sources, line sources, area sources and / or volume sources, independent of any grid. With “grid box”, the Flexpart output grid box is meant. If the simulation goes forward in time, this could be interpreted as a concentration grid. If the simulation goes backward in time, one could call this the emission grid instead.

Lines 105-6: Again, it is not accurate to refer the averaging time period as “the output frequency” here.

Reply:

We will correct this.

Lines 107-110: Please specify the resolutions of the meteorological data inputs for FLEXPART.

Reply:

We will add the following in the revised manuscript:

“The EDA system uses a Gaussian grid with 640 latitude lines between pole and equator, but the data was converted to a lon-lat grid having grid spacings of 0.5°.”

Lines 138-9, “Since this spans many orders of magnitude, we take log 10(Q) as source parameter in our implementation and simply impose a uniform prior between 10 and 16”: Does that mean the accumulated release Q is assumed as 10 13 Bq?

Reply:

No, the prior distribution from which initial samples are drawn is a uniform distribution between 10 and 16. The accumulated release is thus assumed to be between 10^10 and 10^16 Bq.

Lines 197-200: These steps are quite important. Brief descriptions of them are suggested here.

Reply:

Regarding the snooker step, we were informed by one of the developers of MT-DREAM(ZS) that the snooker step is theoretically not compatible with the multiple-try part of the algorithm, so that we no longer use the snooker step. The difference in the posterior after using and not using the snooker step is not noticeable in our simulations.

To proof the latter, please find the results below for two simulations for the Ru-106 case, with and without the snooker step:

1/ simulation with the snooker step for the unperturbed member and s_i = 0.5

```
 Running MT-DREAMzs, iteration 7800 of 50001 . Current logp -37.44259 -41.24711 -39.54531
Converged after 7800 iterations
 Running MT-DREAMzs, iteration 50001 of 50001 . Current logp -36.48064 -44.17845 -41.44014
 MT-DREAMzs terminated after 1206.558 seconds
Acceptance rate for chain 1 is 22.24%
Acceptance rate for chain 2 is 22.61%
Acceptance rate for chain 3 is 22.93%
           lon      lat  log10_Q              rstart               rstop
0.025 50.11799 55.50922 14.96976 2017-09-25 00:22:32 2017-09-26 23:34:40
0.5   51.09007 55.91466 15.27527 2017-09-25 07:59:14 2017-09-27 18:17:46
0.975 57.88037 60.75305 15.64360 2017-09-25 22:55:13 2017-09-27 23:34:05
mean  51.98106 56.39979 15.28396 2017-09-25 08:46:06 2017-09-27 16:41:15
```

2/ simulation without the snooker step for the unperturbed member and s_i = 0.5

```
 Running MT-DREAMzs, iteration 12300 of 50001 . Current logp -44.62895 -44.45257 -41.44722
Converged after 12300 iterations
 Running MT-DREAMzs, iteration 50001 of 50001 . Current logp -43.13082 -39.71556 -37.64025
 MT-DREAMzs terminated after 1271.046 seconds
Acceptance rate for chain 1 is 25.26%
Acceptance rate for chain 2 is 24.75%
Acceptance rate for chain 3 is 23.59%
           lon      lat  log10_Q              rstart               rstop
0.025 50.09579 55.51231 14.97063 2017-09-25 00:30:47 2017-09-26 23:01:53
0.5   51.03933 55.88155 15.26998 2017-09-25 08:00:29 2017-09-27 18:35:54
0.975 57.74679 60.13044 15.65440 2017-09-25 22:38:03 2017-09-27 23:37:18
mean  51.85432 56.30864 15.27988 2017-09-25 08:42:32 2017-09-27 16:59:44
```

The following information will be added to the revised manuscript:

~~“The algorithm is designed so that a snooker step occurs with a probability of 20 % to allow jumps between different posterior modes (ter Braak and Vrugt, 2008).~~ To enhance efficiency and to obtain more accurate results, randomized subspace sampling is used (Vrugt et al., 2009). This simply means that not necessarily all source parameters are updated at a time, but instead a randomized

subset of the source parameters. Furthermore, MT-DREAM (ZS) makes use of multiple try Metropolis sampling (Liu et al., 2000) to enhance the mixing of the chains. This means in practice that, to advance to Markov chain, several proposals are drawn instead of one proposal in traditional Metropolis sampling. Furthermore, the Metropolis acceptance is calculated in a different way (Liu et al., 2000 , Laloy and Vrugt, 2012)."

Section 5.2: In this section, the use of "time" (e.g. lines 318, 320, and 321) is confusing. I believe it is used to refer the chosen 3-hr release time intervals. Please clarify.

Reply:

Thank you for noticing this, we will replace "times" by "3 h release time intervals".

On the model uncertainties in Bayesian source reconstruction using an ensemble of weather predictions, the emission inverse modelling system FREARtool v1.0 and the Lagrangian transport and dispersion model Flexpart v9.0.2

Pieter De Meutter[1,2,3], Ian Hoffman[1], and Kurt Ungar[1]

[1]Radiation Protection Bureau, Health Canada, 775 Brookfield Road, Ottawa, Canada
[2]Belgian Nuclear Research Institute, Boeretang 200, Mol, Belgium
[3]Royal Meteorological Institute of Belgium, Ringlaan 3, Brussels, Belgium

Correspondence: Pieter De Meutter (pieter.de.meutter@sckcen.be)

Abstract. Bayesian source reconstruction is a powerful tool for determining atmospheric releases. It can be used, amongst other applications, to identify a point source releasing radioactive particles into the atmosphere. This is relevant for applications such as emergency response in case of a nuclear accident, or Comprehensive Nuclear-Test-Ban treaty verification. The method involves solving an inverse problem using environmental radioactivity observations and atmospheric transport models. The Bayesian approach has the advantage of providing credible intervals on the inferred source parameters in a natural way. However, it requires the specification of the inference input errors, such as the observation error and model error. The latter is particularly hard to provide as there is no straightforward way to determine the atmospheric transport and dispersion model error. Here, the importance of model error is illustrated for Bayesian source reconstruction using a recent and unique case where radionuclides were detected on several continents. A numerical weather prediction ensemble is used to create an ensemble of atmospheric transport and dispersion simulations, and a method is proposed to determine the model error.

1 Introduction

Nuclear facilities release a certain amount of anthropogenic radioactive particulates or gases into the atmosphere, which are transported and dispersed by the wind. These releases can either be routine or accidental. Several countries run a network of stations to monitor airborne levels of environmental radioactivity (Steinhauser, 2018). These monitoring networks allow to verify compliance with regulatory release limits, but also to detect reported and unreported nuclear events. Recent examples include the detections of ^{131}I in Europe (Masson et al., 2018) or ^{106}Ru detections on the northern hemisphere (Masson et al., 2019).

On the international scale, the radionuclide component of the International Monitoring System will consist of 80 stations measuring radioactive particulates (of which at least 40 will be equipped with radioactive noble gas detectors). This network is being set up to verify compliance with the Comprehensive Nuclear-Test-Ban Treaty once it enters into force. In the past, anomalous radionuclide detections were made that are likely linked to a nuclear explosion (Ringbom et al., 2014; De Meutter et al., 2018).

If an anomalous[1] detection occurred (either from a nuclear accident or a clandestine nuclear weapon test), methods are needed that relate the detection with its source - or potential sources if the source is unknown. One of these methods is atmospheric transport and dispersion modelling. An atmospheric transport and dispersion model typically simulates the transport, dispersion, dry and wet deposition and radioactive decay of radionuclides released in the atmosphere. These processes establish a linear relationship between the concentrations at receptors and the release amount at the source. One can calculate such source–receptor-relationships (Seibert and Frank, 2004) between a fixed source and several receptors or stations (when modelling forward in time) or between ~~several potential sources and~~ a fixed receptor and several potential sources (when modelling backwards in time).

A significant event of interest will often be accompanied by multiple detections taken at multiple stations. Statistical methods can then be employed to combine the information from all these detections (and possibly non-detections – observations where the activity concentration is below a minimum detectable concentration) in a meaningful way in order to infer relevant information on the source. In case of an unknown source, the objective is often to find the source location, release time and release amount. In case of a known source, the source location and perhaps also the release times are known. In that case, the release amount and release height can be inferred to refine a previous release estimate obtained through other ways (for instance, an estimation could be made based on accident scenarios and the known or estimated inventory of a reactor). The process of inferring information on the source based on observations is called inverse modelling. Several methods exist, ranging from simply calculating correlations between observations and source-receptor-sensitivities to locate the source (Becker et al., 2007) to more elaborate methods such as optimization methods using cost functions (e.g., Stohl et al., 2012), data assimilation (e.g., Bocquet, 2007) or Bayesian inference (e.g., Yee, 2012).

Of these methods, the Bayesian inference has the advantage of readily providing an uncertainty quantification on the outcome. However, the quality of the inference and the uncertainty quantification depends on the quality of input uncertainties. Typically, one specifies the observation error and model error. Here, the model error relates to errors in the atmospheric transport and dispersion model. These errors are very hard to readily quantify, mainly because of the underlying numerical weather prediction data which is used to calculate the transport and dispersion (Harris et al., 2005; Engström and Magnusson, 2009; Hegarty et al., 201 The errors associated with numerical weather prediction depend on the atmospheric state and thus vary between locations and from day to day.

[1] Anomalous radionuclide detections are detections of anthropogenic radionuclides originating from upwind nuclear facilities, where the detected concentration of (a) specific radionuclide(s) and/or the combination of several detected radionuclides are anomalous with respect to the station's detection history and/or with respect to what can be expected from these upwind nuclear facilities operating under normal conditions.

A well-established method to quantify uncertainties in numerical weather predictions is the ensemble method (Leutbecher and Palmer, 2008). For simplicity, let us focus on single-model ensembles. Such ensembles are created using a single model and consist of a set of one unperturbed and several perturbed scenarios or model predictions. These scenarios are designed in such a way that the spread among the different scenarios represent the uncertainty of an individual model prediction. Each scenario (also called an ensemble member) is created by perturbing certain input data and/or using different parameterisation schemes (except for the unperturbed member, which is created by running the model with the best available input and parameterisation schemes). The key to create a good ensemble lies in providing realistic perturbations. The latter is a very complex task which requires expert knowledge of all data, processes and their associated uncertainties at each level of the modelling process. Ideally, one has a large number of distinct ensemble members (so that many different but realistic scenarios can be obtained). However, the huge computational cost of running an ensemble and storing its vast amount of data limits the number of ensemble members. Therefore, ensembles used operationally at major weather institutes around the world are designed in a way that, even with a limited number of members (between 14 and 50, Leutbecher, 2019), the ensemble ~~can capture most~~ tries to capture all (and not more) of the possible outcomes.

Yee et al. (2014) demonstrated the significant impact of model error on the outcome of Bayesian source reconstruction by employing two different measurement models for the incorporation of the model error (a measurement model relates the model variable with the observation). Nevertheless, given the difficulty to quantify the model error, many studies instead rely on assumptions regarding the model error structure and scale. In this paper, an ensemble of atmospheric transport model simulations is used to determine the model error used in the Bayesian inference. The effect of different model error formulations on the source localisation is studied for a recent case where ^{106}Ru was observed throughout the northern hemisphere in autumn 2017. In this study, the Bayesian source reconstruction tool FREARtool (FREAR stands for Forensic Radionuclide Event Analysis and Reconstruction) described in De Meutter and Hoffman (2020) will be used.

In Section 2, the case that is used to illustrate how the ensemble can be used to determine the model error, is described. The observational and model data are described. Section 3 describes the Bayesian inference system. Section 4 describes the effect of model error on the source location probability maps. The uncertainty parameters are fitted using the ensemble in Section 5. The ensemble is used in a scenario-based way in Section 6. This allows to test whether information is lost when using the ensemble solely to fit uncertainty parameters as performed in Section 5. Finally, conclusions are given in Section 7.

2 Description of the case

In autumn 2017, several national and international monitoring networks reported the detection of ^{106}Ru and, to a lesser extent ^{103}Ru (Masson et al., 2019). However, to date, no country or facility has claimed responsibility for the release. ^{106}Ru (half-life: 373.6 d) and ^{103}Ru (half-life: 39.26 d) are radioactive particulates that do not have natural sources. Since no other fission products such as iodine and cesium were measured, a nuclear reactor accident can be excluded. Several studies using direct and inverse atmospheric transport modelling showed that a release in the region of the southern Ural mountains in Russia can best explain the observations (Sørensen, 2018; Saunier et al., 2019; Bossew et al., 2019; De Meutter et al., 2020). Two major

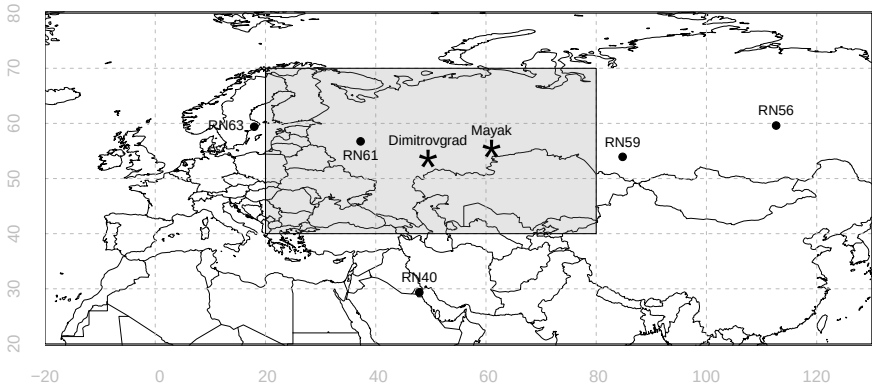

Figure 1. Locations of the five stations from which Ru-106 detections have been used. The shaded box shows the search domain of the Bayesian inference. For reference, the locations of two nuclear facilities are also shown: the Research Institute of Atomic Reactors in Dimitrovgrad, labelled "Dimitrovgrad" (54.19° lat and 49.48° lon) and the Mayak Production Association in Ozyorsk, labelled "Mayak" (55.70° lat and 60.80° lon).

nuclear facilities are located in that area: the Research Institute of Atomic Reactors and the Mayak Production Association (Fig. 1).

Here, we revisit the modelling data used in De Meutter et al. (2020) but perform a Bayesian analysis instead of a cost function analysis. We apply our methods to this case as it offers a very interesting data set since (i) there were detections on multiple continents, (ii) the single-source assumption is likely valid as there is no measurable global background from anthropogenic sources and (iii) it is a recent case so that state-of-the-art numerical weather data is readily available. However, we stress that the case study is used here to examine the model error structure for a Bayesian source reconstruction, and how an ensemble can provide insight in the model error structure and scale – it is not our goal as such to find the origin of the reported ^{106}Ru and ^{103}Ru.

2.1 Activity concentration observations

Twelve ^{106}Ru detections have been used from five different stations of the International Monitoring System which is being commissioned to verify compliance with the Comprehensive Nuclear-Test-Ban Treaty once it enters into force~~: three consecutive detections at station RN40 (Kuwait City, Kuwait) taken between 4 and 6 October 2017; one detection at station RN56 (Peleduy, Russian Federation) taken on 6 October 2017 ; four consecutive detections at station RN59 (Zalesovo, Russian Federation) taken between 3 and 6 October 2017; one detection at station RN61 (Dubna, Russian Federation) taken on 5 October 2017; and three consecutive detections at station RN63 (Stockholm, Sweden) taken between 1 and 3 October 2017.~~ . The observations, with their associated sampling times, are listed in Table 1. The location of these stations are shown in Fig. 1. ~~Although a few stations also measured ^{103}Ru, these detections have not been used for the inference.~~ The stations sample

Table 1. List of twelve observations with their corresponding sampling times and location.

Station	Collection Start (UTC)	Collection Stop (UTC)	Location
RN59	2017-10-02 02:01	2017-10-03 02:03	Zalesovo, Russian Federation
RN59	2017-10-03 02:03	2017-10-04 01:53	Zalesovo, Russian Federation
RN59	2017-10-04 01:53	2017-10-05 01:58	Zalesovo, Russian Federation
RN59	2017-10-05 01:58	2017-10-06 01:49	Zalesovo, Russian Federation
RN40	2017-10-03 06:51	2017-10-04 06:51	Kuwait City, Kuwait
RN40	2017-10-04 06:51	2017-10-05 06:51	Kuwait City, Kuwait
RN40	2017-10-05 06:51	2017-10-06 06:51	Kuwait City, Kuwait
RN61	2017-10-04 09:28	2017-10-05 08:04	Dubna, Russian Federation
RN63	2017-09-30 08:41	2017-10-01 08:41	Stockholm, Sweden
RN63	2017-10-01 08:41	2017-10-02 08:41	Stockholm, Sweden
RN63	2017-10-02 08:41	2017-10-03 08:41	Stockholm, Sweden
RN56	2017-10-04 23:46	2017-10-05 23:46	Peleduy, Russian Federation

approximately 20~~,~~ 000 m^3 of air during a period of 24 h, during which radioactive particulates - if present - are captured on a filter. The filter is then analyzed using gamma spectroscopy to detect the presence of any radionuclides. ~~The above observation times refer to the collection stop of the sample.~~ A few stations also measured ^{103}Ru, but these detections will not be used for the inference. Although the observations were selected to realistically represent this case, we do not attempt to use all (or some optimal selection) of the available detections and non-detections in the inference as it is not our goal to determine the true source location as such. Furthermore, many additional observations contain redundant information. Also, since this work is focussed on Comprehensive Nuclear-Test-Ban Treaty verification applications, we only use a few observations taken at stations from the International Monitoring System. Finally, local observations in Russia could heavily influence the inferred source location, but we do not use these as our numerical weather ensemble is not well suited to capture uncertainties at these smaller spatial and shorter temporal scales.

2.2 NWP and atmospheric transport modelling

We have used the source–receptor-sensitivities associated with the twelve observations from De Meutter et al. (2020). These were obtained by running the Lagrangian particle model Flexpart (Stohl et al., 2005) in backward mode (Seibert and Frank, 2004). All simulations ended on 20 September 2017. The source–receptor-sensitivities represent the residence time of modelled particles in a geotemporal grid box. The ~~model output frequency was~~ time-averaged source-receptor-sensitivities were output every three hours, so that the maximum possible residence time in a geotemporal grid box is 10~~,~~ 800 s. The source–receptor-sensitivities have horizontal grid spacings of 0.5°. An activity concentration (in Bq/m^3) can be related to a release (in Bq) by multiplying the latter with the source–receptor-sensitivity (in s), and dividing by the grid box volume and the ~~output~~

~~frequency~~source-receptor-sensitivities averaging period (which was 10 800 s). An ensemble of numerical weather predictions was used to create an ensemble of atmospheric transport and dispersion simulations. The transport and dispersion processes themselves were not perturbed. The Ensemble of Data Assimilations from the European Centre for Medium-Range Weather Forecasts (ECMWF) was used to run Flexpart. It consists of 26 independent lower-resolution 4D-Var assimilations: one using unperturbed observations and physics, and 25 using perturbed observations, sea-surface temperatures and model physics (Bonavita et al., 2016). The EDA system uses a Gaussian grid with 640 latitude lines between pole and equator, but the data were converted to a lon-lat grid having grid spacings of 0.5°. By adding and subtracting perturbations from the ensemble mean, the number of perturbed members was doubled, so that 50 perturbed and one unperturbed member were obtained. The perturbations are created in such a way that each ensemble member represents a possible scenario for the true (unknown) atmospheric state, and the spread between the different members ~~represents the uncertainty.~~ is simply the model uncertainty as estimated by the ensemble. Then, for each weather ensemble member, Flexpart was run so that an atmospheric transport ensemble of 51 source-receptor-sensitivities was obtained.

3 Bayesian source reconstruction

De Meutter and Hoffman (2020) have developed and applied a Bayesian inference system to find the source parameters of an anomalous ^{75}Se release based on airborne ^{75}Se activity concentrations. The main components of the inference system are summarized here, but details are given in De Meutter and Hoffman (2020) and references therein.

3.1 Source parameters

The unknown source is described by five source parameters:

- the longitude of the source (x_s)
- the latitude of the source (y_s)
- the accumulated release (Q)
- the release start time (t_{start})
- the release end time (t_{stop})

In practice, the release period is parameterized by fractions $rstart$ and $rstop$ to ensure that t_{start} occurs before t_{stop}. The corresponding release period is calculated as follows:

$$t_{start} = t_1 + rstart\,(t_m - t_1) \tag{1}$$

$$t_{stop} = t_{start} + rstop\,(t_m - t_{start}) \tag{2}$$

with t_1 and t_m the first and last time for which source–receptor-sensitivities are available for the source reconstruction. The release rate is assumed constant during the release period. The vertical position of the source is assumed to stretch between the surface and the top of the lowest model layer (at 100 m). Thus, the unknown source is parameterized as follows:

$$\Xi(x,y,z,t;x_s,y_s,z_{top},Q,t_{start},t_{stop}) = \frac{Q}{(t_{stop}-t_{start})\,z_{top}}\,\delta(x-x_s)\,\delta(y-y_s)\,[\mathcal{H}(z)-\mathcal{H}(z-z_{top})] \\ [\mathcal{H}(t-t_{start})-\mathcal{H}(t-t_{stop})] \quad (3)$$

with δ the Dirac delta function and $\mathcal{H}$ the Heaviside step function.

3.2 Prior distribution

Uninformative bounded uniform priors are used for the source parameters. The source longitude is assumed to be between 20° and 80° and the source latitude is assumed to be between 40° and 70° (see Fig. 1 for a map showing the search domain). The accumulated release is assumed to be between 10^{10} Bq and 10^{16} Bq. Since this spans many orders of magnitude, we take $\log 10(Q)$ as source parameter in our implementation and simply impose a uniform prior between 10 and 16. Recall that the first observation that we consider for the inference was taken on 1 October 2017. Therefore, the release is assumed to have occurred between 25 September 2017 0000 UTC and 28 September 2017 0000 UTC. Generally, the upper limit on the release time will exclude solutions further downwind, while the lower limit on the release time will exclude solutions further upwind. Uniform bounded priors between 0 and 1 are used for $rstart$ and $rstop$.

3.3 Likelihood

De Meutter and Hoffman (2020) proposed likelihood equations that can take into account detections, instrumental non-detections, misses and false alarms using Currie detection limits (Currie, 1968). Since non-detections will not be used in this study, only the likelihood of detections will be used here. The possibility of a false alarm, where the detector wrongly identifies a detection, is also considered. For simplicity, the observations $\mathbf{c_{det}}$ are assumed to be independent, thereby neglecting possible geotemporal correlations. As a result, the total likelihood is simply the product of the likelihood associated with individual observations:

$$p(\mathbf{c_{det}}|\mathbf{c_{mod}}) = \prod_{i=1}^{n} p(c_{det,i}|c_{mod,i}) \quad (4)$$

First, let us define c_{true}, which is the true activity concentration that will never be known. Next, we define c_{det}, which is the activity concentration as seen by the detector and which can differ from c_{true} (the observed net signal can even be negative due to the statistical nature of spectroscopic analysis). ~~We use the~~ Currie (1968) defined a critical level L_C above which a net signal (which is the detected signal from which the effects of background radiation are subtracted) should be in order to declare the net signal to be a detection. We use this Currie critical threshold ~~,~~ L_C ~~, (Currie, 1968)~~ to decide, given a net signal, whether a

Table 2. Definition of a true non-detection, a miss, a false alarm and a true detection based on the Currie critical level L_C, the detected activity concentration c_{det} and the "true" activity concentration c_{true} which will never be known.

	$\bar{d}_{obs}$ ($c_{det} < L_C$)	d_{obs} ($c_{det} > L_C$)
$\bar{d}_{true}$ ($c_{true} < L_C$)	true non-detection	false alarm
d_{true} ($c_{true} > L_C$)	miss	true detection

real detection took place (d_{obs} if $c_{det} > L_C$) or not ($\bar{d}_{obs}$ if $c_{det} < L_C$). One can then define a true non-detection, a miss, a true detection and a false alarm as in Table 2. In this application, the risk of a missed detection and a false alarm is set equal to 5 %. c_{true} will never be known, but we assume that the modelled activity concentration c_{mod} corresponding to a source hypothesis Ξ and its associated uncertainty results in a distribution of true activity concentrations $d_{c_{true}}$ given by the following formula:

$$d_{c_{true}}(c_{true}|c_{mod,i}) = \frac{\bar{\alpha}^{\bar{\beta}}\,\Gamma(\bar{\beta}+0.5)}{\sqrt{2\pi}\,s_i\,\Gamma(\bar{\beta})}\,\frac{1}{[\bar{\alpha}+(c_{true}-c_{mod,i})^2/(2\,s_i^2)]^{\bar{\beta}+0.5}} \tag{5}$$

with the index i denoting "corresponding to the i$^{\text{th}}$ observation" (with $i = 1...12$), Γ the gamma function and s, $\bar{\alpha}$ and $\bar{\beta}$ parameters of an inverse gamma distribution. The above equation was used by Yee (2012) as likelihood function for detections, and was obtained by starting from a Gaussian function, replacing the standard deviation σ by an inverse gamma distribution and integrating over all possible values of σ. However, in order to take into account the possibility of a false alarm, instead we propose the following likelihood for the ith detection $c_{det,i}$ given its corresponding modelled activity concentration $c_{mod,i}$ associated with a source hypothesis Ξ:

$$p(c_{det,i}|c_{mod,i}) = p(c_{det,i}|c_{mod,i}, d_{true})\,p(d_{true}|c_{mod,i}) + p(c_{det,i}|\bar{d}_{true})\,p(\bar{d}_{true}|c_{mod,i}). \tag{6}$$

In the equation above, $p(d_{true}|c_{mod})$ is the probability of a true detection, given by:

$$p(d_{true}|c_{mod}) = \int_0^\infty p(d_{true}|c_{true})\,p(c_{true}|c_{mod})\,\mathrm{d}c_{true} \tag{7}$$

$$= \int_{L_C}^\infty p(c_{true}|c_{mod})\,\mathrm{d}c_{true}, \tag{8}$$

and $p(\bar{d}_{true}|c_{mod})$ the probability of a true non-detection, given by:

$$p(\bar{d}_{true}|c_{mod}) = \int_0^{\infty} p(\bar{d}_{true}|c_{true})\, p(c_{true}|c_{mod})\, \mathrm{d}c_{true} \tag{9}$$

$$= \int_0^{L_C} p(c_{true}|c_{mod})\, \mathrm{d}c_{true} \tag{10}$$

$$= 1 - p(d_{true}|c_{mod}). \tag{11}$$

In these equations, $p(c_{true}|c_{mod})$ is simply equal to Eq. 5, but normalized (below, c'_{true} is a dummy variable for integration):

$$p(c_{true}|c_{mod}) = \frac{d_{c_{true}}(c_{true}|c_{mod})}{\int_0^{\infty} d_{c_{true}}(c'_{true}|c_{mod})\, \mathrm{d}c'_{true}} \tag{12}$$

$p(c_{det,i}|c_{mod,i}, d_{true})$ gives the likelihood of detecting $c_{det,i}$ given $c_{mod,i}$ and assuming that the detection was not a false alarm. We assume it is given by Eq. 5 as follows:

$$p(c_{det,i}|c_{mod,i}, d_{true}) = d_{c_{true}}(c_{true} = c_{det,i}|c_{mod,i}) \tag{13}$$

$p(c_{det,i}|\bar{d}_{true})$ is the likelihood of detecting $c_{det,i}$ assuming that the detection is a false alarm:

$$p(c_{det,i}|\bar{d}_{true}) = \int_0^{L_C} \frac{1}{\sqrt{2\pi} L_C/k_\alpha} \exp\left(-\frac{(c_{det} - c_{true})^2}{(L_C/k_\alpha)^2}\right) \mathrm{d}c_{true} \tag{14}$$

with $k_\alpha = 1.645$ for a false alarm risk of 5 %. The likelihood and its different components are plotted for two hypothetical detections in Fig. 2. For the hypothetical detection c_{det} close to its L_C (Fig. 2, left), the likelihood is significantly increased for small c_{mod} because the possibility of a false alarm is considered. For a hypothetical detection c_{det} significantly larger than its L_C (Fig. 2, right), the consideration of a false alarm does not significantly add to the total likelihood. Indeed, for $c_{det,i} \gg L_{C,i}$ and $c_{mod,i} \gg L_{C,i}$, Eq. 6 simplifies to:

$$p(c_{det,i}|c_{mod,i}) \approx d_{c_{true}}(c_{true} = c_{det,i}|c_{mod,i}) \tag{15}$$

3.4 Sampling from the posterior

The posterior distribution was sampled using the general-purpose Markov chain Monte Carlo algorithm MT-DREAM$_{(ZS)}$ (Laloy and Vrugt, 2012)~~is used to sample the posterior distribution~~. A chain of source hypotheses is obtained by repeatedly creating a proposal source hypothesis based on the current source hypothesis, and then accepting the proposal or - if

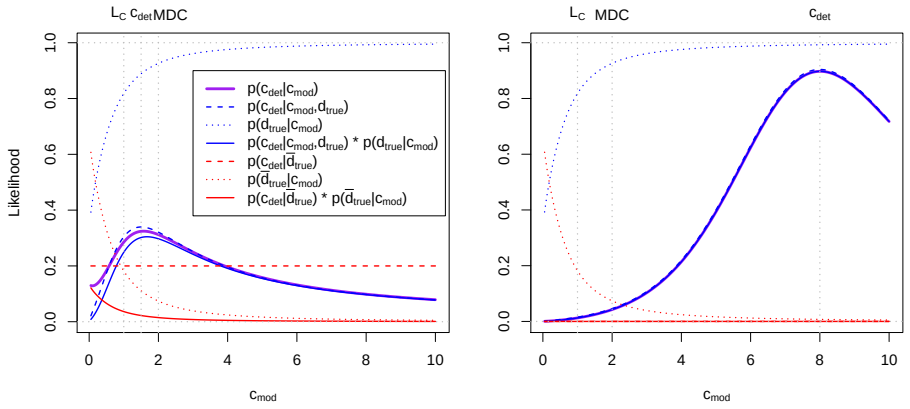

Figure 2. Likelihood function for two detections: c_{det} = 1.5 (left) and c_{det} = 8.0 (right). The gray dotted vertical lines represent $L_C = 1.0$, MDC $= 2.0$ and the detected activity concentration c_{det}. The units are arbitrary in this example.

rejected - retaining the current source hypothesis. A key feature of a Markov chain Monte Carlo algorithm is its ability to construct proposals in such a way that the posterior distribution is efficiently explored and sampled. MT-DREAM$_{(ZS)}$ creates a new proposal by adding a perturbation to the current source hypothesis. Such perturbations are created by taking the difference of two randomly drawn states out of an archive of past states, so that the size and direction of the perturbation are adaptive and optimal to the problem - without the necessity of prior tuning. Every 10th iteration, the current source hypothesis is added to the archive. Three chains are run simultaneously (sharing the same archive of past states) to diagnose convergence more easily. ~~The algorithm is designed so that a snooker step occurs with a probability of 20 % to allow jumps between different posterior modes (ter Braak and Vrugt, 2008).~~ To enhance efficiency and to obtain more accurate results, randomized subspace sampling is used (Vrugt et al., 2009). This simply means that not necessarily all source parameters are updated at a time, but instead a randomized subset of the source parameters. Furthermore, MT-DREAM$_{(ZS)}$ makes use of multiple try Metropolis sampling (Liu et al., 2000) to enhance the mixing of the chains. This means in practice that, to advance to Markov chain, several proposals are drawn instead of one proposal as in traditional Metropolis sampling. Furthermore, the Metropolis acceptance is calculated in a modified way (Liu et al., 2000; Laloy and Vrugt, 2012).

3.5 Observation and model errors

The detected activity concentration c_{det} is assumed to be Gaussian distributed around the true activity concentration c_{true}, with a standard deviation ~~$\sigma_{obs} = L_C/k_\alpha$ as in Eq. 14.~~ σ_{obs} equal to the reported observation error. The model error originates from errors in the source-receptor-sensitivities calculated by the atmospheric transport model. The main sources of error are simplifications in the turbulence parameterization of the atmospheric transport model and errors in the numerical weather prediction data used to run the atmospheric transport model. As it is very hard if not impossible to specify the model error, we

assume Gaussian errors, but replace a fixed σ_{mod} by an inverse gamma distribution, as mentioned earlier in this section and following Yee (2012):

$$\psi(\sigma_{mod,i}|s_i,\bar{\alpha}_i,\bar{\beta}_i) = 2\frac{\bar{\alpha}_i^{\bar{\beta}_i}}{\Gamma(\bar{\beta}_i)}\left(\frac{s_i}{\sigma_{mod,i}}\right)^{2\bar{\beta}_i}\exp\left(-\bar{\alpha}_i\frac{s_i^2}{\sigma_{mod,i}^2}\right)\frac{1}{\sigma_{mod,i}} \tag{16}$$

with the subscript i denoting that these values can be observation-specific. The parameters s, $\bar{\alpha}$ and $\bar{\beta}$ of the inverse gamma distribution are respectively an estimate of ~~σ~~σ_{mod}, a scale parameter and a shape parameter. Yee (2012) proposed to use $\bar{\alpha} = \pi^{-1}$ and $\bar{\beta} = 1$, in which case $\langle\sigma_{mod,i}\rangle = s_i$ and the variance of $\sigma_{mod,i}$ becomes infinite (Eq. 16 will then be a very heavy-tailed distribution). We now have to come up with a value for s_i. Since the source-receptor-sensitivities or srs (calculated by the Flexpart model) typically span many orders of magnitude, it makes more sense to define a relative error rather than an absolute error. Furthermore, because the source-receptor-sensitivities are linearly proportional to the activity concentration, we propose the following for s_i:

$$s_i = \frac{\sigma_{srs,i}}{srs_i}\max(c_{det,i}, 16 * L_{C,i}) \tag{17}$$

~~As a consequence, the model uncertainty does not depend on the source parameters.~~ σ_{srs} is the unknown srs error in the equation above. The value of 16 is an empirical number that was found to give a good balance between information obtained from detections versus information from non-detections (De Meutter and Hoffman, 2020). Note that the above proposal results in a larger relative model error if $c_{det,i} < 16 * L_{C,i}$. This is desirable since small detections are caused by a part of the plume of radionuclides that was subject to more atmospheric ~~transport and dispersion processes~~ dilution and thus should have a larger relative uncertainty than large detections (although the latter have higher absolute uncertainty). In Eq. 13 (which is simply Eq. 5 but with c_{true} replaced by $c_{det,i}$), s_i is replaced by the following in order to take into account both observation and model error:

$$s_i \rightarrow \sqrt{s_i^2 + \sigma_{obs,i}^2} \tag{18}$$

From now on, we will express any parameters s_i and σ (the combined observation and model error) as relative errors, thus as fractions of $\max(c_{det,i}, 16 * L_{C,i})$ (see Eq. 17).

4 Model error and Bayesian source reconstruction

In this section, only the unperturbed member is used. The model uncertainty thus needs to be specified manually, and the impact of different choices on the posterior is discussed here. We focus on the inferred source location, and not on the release period or release amount. This is because the release location is of primary interest in the context of Treaty verification. Once a location is found (for instance, based on the location of known nuclear facilities within the posterior source region, or based

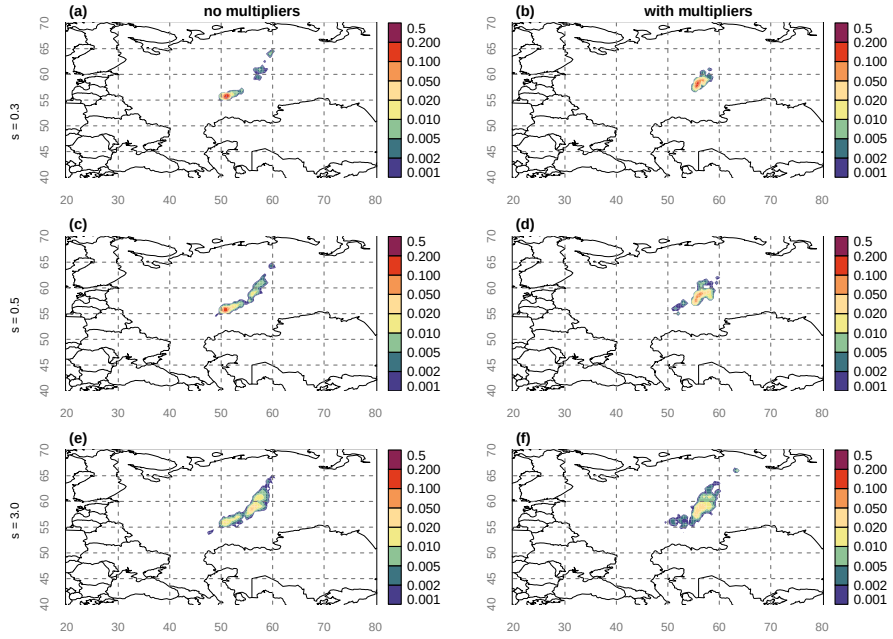

Figure 3. Source location probability maps obtained from the Bayesian source reconstruction. The model uncertainty parameters of the inverse gamma distribution were fixed *a priori*, and different values for s were used in the panels: $s = 0.3$ in (a, b), $s = 0.5$ in (c, d) and $s = 3.0$ in (e, f). No multipliers were used in (a, c, e), while multipliers were used in panels (b, d, f) (see Subsection 4.3 for details). The longitude and latitude are shown by gray numbers and dashed lines.

on a seismic signal associated to a nuclear explosion), a new inference could be performed fixing the release location as was done in De Meutter and Hoffman (2020).

4.1 Posterior effects

In this section, the effect of model error on the posterior is illustrated by varying the parameter s_i of the inverse gamma function (Eq. 16), which roughly fixes the scale of the model uncertainty. The Bayesian inference was independently repeated multiple times, once for each of the chosen parameter values. The resulting source location probability maps are shown in Fig. 3 (a, c, e) for three values of s_i: 0.3, 0.5 and 3. Here, the same value s_i is used for all observations. The source location probability maps illustrate that the model error has a profound impact on the resulting posterior. The source location probability ranges from a narrow spot to a fairly elongated area. Perhaps surprisingly, an increase in the uncertainty parameter s does not simply enlarge the region of possible sources: instead, the enlargement is mostly in one direction (northeast); furthermore, the area with the highest probability shifts (also northeast). Therefore, one cannot simply predict beforehand how the posterior will change when the model error is changed. The resulting source location is in line with what previous studies found (Sørensen, 2018; Saunier et al., 2019; Bossew et al., 2019; De Meutter et al., 2020).

4.2 Introduction to multipliers

In this and the next subsection, an alternative model error structure will be discussed involving multipliers or scale factors. Besides being an alternative model error, multipliers could also be used to take into account ~~unknown errors~~ errors that were not fully captured by the model (such as errors due to local atmospheric features not resolved by the model~~).~~, measurement errors due to sample inhomogeneity, etc.)

Yee et al. (2014) used a small number of activity concentration measurements to retrieve the known source parameters of a major medical isotope production facility. The resulting source longitude, latitude and the release term were compared with the true source parameters in order to evaluate the performance of the source reconstruction. The method they used is similar to the method employed here: a Lagrangian stochastic particle model was used in backward mode to generate the source-receptor-sensitivities, and the source parameters were obtained using Bayesian inference and the MT-DREAM$_{(ZS)}$ algorithm. Yee et al. (2014) showed that the main challenge in source reconstruction lies in the correct specification of the model uncertainty (both scale and structure) by using two different measurement models:

$$c_{obs,i} = c_{mod,i} + \epsilon_i \tag{19}$$

$$c_{obs,i} = m_i\, c_{mod,i} + \epsilon_i' \tag{20}$$

In Eq. 19, ϵ_i is the combined model and measurement error, which is assumed to be drawn from a Gaussian distribution with unknown standard deviation. Again, the unknown standard deviation is replaced by a distribution rather than a single fixed value. In Eq. 20, ϵ_i' only represents the measurement error. The model error is now taken into account by so-called scale factors or multipliers m_i. These multipliers are parameters of the Bayesian inference and they are allowed to vary between 0.1 and 10. The range is chosen so that the multipliers can compensate for model underpredictions or overpredictions of up to a factor 10.

Yee et al. (2014) found that using Eq. 19 resulted in a source reconstruction that did not include the correct source parameters. The multipliers (Eq. 20) on the other hand resulted in a huge shift in the posterior source location, which significantly improved the source reconstruction.

4.3 Multipliers as unknown model error

Here, multipliers (m_i) are introduced to account for unknown model uncertainties that are not yet taken into account by the likelihood formulation. For instance, local atmospheric features which are not resolved by the model might result in incorrect source-receptor-sensitivities. Such errors are very hard to quantify, because the computational power to resolve such features is prohibitively high (at least for continental-scale problems).

The multipliers are additional parameters that need to be estimated during the Bayesian inference. For the sampling of the parameter, we work with $\log_{10}(m_i)$ and assign a uniform prior between $[-1, 1]$. While the multipliers increase the run time of the Bayesian inference, the increase is small - especially if one considers the huge increase in the number of unknown parameters (one additional parameter per observation).

We apply the multipliers using three different values for the model uncertainty parameter s_i (0.3, 0.5 and 3), and show the results in Fig. 3 (b, d, f). Somewhat similar to the results in Yee et al. (2014), the multipliers cause in a shift in the source location probability (though not as dramatic). Furthermore, the multipliers do not cause a widening in the posterior source location. It is interesting to contrast the effect of introducing the multipliers versus increasing the value of the parameter s_i: while the latter result in a shift and a widening of the posterior distribution, the former only result in a shift. While the results with and without multipliers are significantly different for small values of s_i (Fig. 3, a and b), there is substantial agreement in the results when using $s_i = 3.0$ (Fig. 3, e and f). This suggests that the effect of introducing multipliers and the effect of increasing the model uncertainty converges when specifying large model uncertainty.

The results in this section lead to the question: which of the source location maps shown in Fig. 3 is correct. Or more fundamentally: what is the true structure and scale of the model uncertainty. This will be assessed in the next section using the ensemble of source-receptor-sensitivities described in Subsection 2.2.

5 Fitting uncertainty parameters

5.1 The SRS ensemble distribution

In this subsection, it is assessed whether the atmospheric transport model error structure can be obtained from the ensemble of source–receptor-sensitivities (SRS). Note that the ensemble is set up to deal with errors arising from the meteorological input data only. While this type of error ~~arguably~~ likely adds the largest contribution to the total model error, other sources of model error are not included.

As our ensemble contains 51 members (one unperturbed member and 50 perturbed members), there are 51 SRS values available for each spatiotemporal grid box and each observation. In order to obtain the error structure, the data of all spatial grid boxes is aggregated into an uncertainty distribution. This does not necessarily destroy the spatial error correlations in the numerical weather prediction data, since the SRS are the result of an integrated trajectory through the atmosphere associated to a specific observation. The following procedure is applied in order to find the error structure:

1. For each SRS file (associated with a certain observation), and for each spatiotemporal grid box, the ensemble median SRS is calculated; each of the 51 SRS values is scaled by its ensemble median.

2. A Lagrangian particle model can only track a finite number of particles due to computational constraints, and this causes stochastic uncertainty when there are very few particles passing through a geotemporal grid box. However, the SRS variations between ensemble members should represent meteorological uncertainty and should not be impacted by stochastic uncertainty. Therefore, a threshold[2] of $\exp(-20)$ [s] is applied on the median: if the median is smaller than the threshold, all its 51 SRS are omitted from the analysis.

[2]Since the SRS files are output every three hours, the maximum value for the SRS is 10800 s.

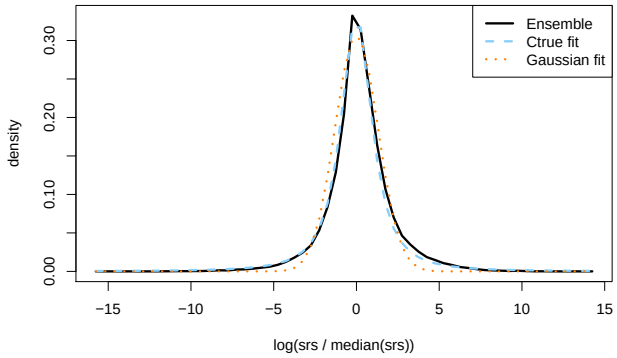

Figure 4. Probability density function showing how the atmospheric transport model ensemble members are ~~distribution~~ distributed around the ensemble median (black solid line). Also shown are two fits, one using Eq. 5 (blue dashed line) and one using a Gaussian distribution (orange dotted line). Note that the SRS are scaled by its ensemble median, and that the natural logarithm is applied.

3. The natural logarithm is applied to all SRS since these span many orders of magnitude. If any ensemble member has an SRS equal to 0 for a specific grid box, all its 51 SRS are omitted from the analysis.

4. The remaining data points are used to make an uncertainty distribution (as in Fig. 4).

As an example, the probability density function of ensemble SRS members around its ensemble median is shown in Fig. 4 for an arbitrary observation and an arbitrary time. It can be seen that most SRS values fall within a factor $\exp(3) \approx 20$ of the ensemble median. Of particular interest is how this density can be approximated by a statistical distribution. In Fig. 4, two distributions are fitted to the density: a Gaussian distribution, and a fit using Eq. 5. Interestingly, while the probability density function can roughly be represented by a Gaussian distribution (effectively being a lognormal distribution with respect to the SRS, since the natural logarithm was applied), its tails are heavier than a Gaussian distribution. On the other hand, a much better fit is obtained using Eq. 5. This distribution describes how the (unknown) true activity concentration is distributed on average around the modelled activity concentration. Similarly, the ensemble gives a distribution of the true SRS around the model predicted SRS. Note that the activity concentrations are simply linear combinations of SRS. This result shows that our choice for the model error structure is supported by the ensemble. This initial result will be elaborated in the next subsection.

5.2 Fitting model uncertainty

In this subsection, the SRS ensemble is used to determine the parameters s_i, $\bar{\alpha}_i$ and $\bar{\beta}_i$ in Eq. 5, which describes how the true activity concentration c_{true} is believed to be distributed given the modelled activity concentration $c_{mod,i}$. The fitting can be performed in several ways, and four different cases are considered. A goodness-of-fit is calculated for each case. As a reference, the fitting procedure is repeated using a Gaussian distribution instead of Eq. 5. The mean of the Gaussian is fixed to be 0 (corresponding to the median of the SRS ensemble), so that only the standard deviation is allowed to vary. Since

our Gaussian fitting thus has only one degree of freedom (and given the suggestion of long tails in Fig. 4), it should not be surprising that the Gaussian fit will be outperformed by the fit using Eq. 5.

The following cases are considered to obtain the uncertainty parameters:

- case 1: the parameters are fixed by *a priori* chosen values; for the fit using Eq. 5, we choose $s_i = 0.5$, $\bar{\alpha}_i = 1/\pi$, $\bar{\beta}_i = 1$ for all observations and ~~times~~all 3 h release time intervals; for the Gaussian distribution, we choose $\sigma = 0.5$
- case 2: the parameters are fitted once for all ~~times~~ release time intervals and all observations (data is aggregated for all release times and all observations)
- case 3: the parameters are fitted for each observation (data is aggregated for all ~~times~~release time intervals)
- case 4: the parameters are fitted for each observation and each ~~time~~ release time interval

The following goodness-of-fit is chosen to quantify how well the fitted probability density function p_{fit} resembles the ensemble probability density function p_{ens}; it involves the integration of the absolute difference of both densities and is simply the fraction of overlap in density:

$$\text{overlap in density} = 1 - \frac{1}{2} \int_{-\infty}^{\infty} |p_{ens}(x) - p_{fit}(x)| \, \mathrm{d}x \tag{21}$$

If the overlap in density equals 1, p_{fit} equals p_{ens}. On the other hand, if it equals 0, there is no overlap between p_{fit} and p_{ens}.

The uncertainty parameters are obtained using the procedures outlined in case 1 to 4. Then, for each case, the overlap in density is calculated by comparing p_{fit} with p_{ens} for each observation and each ~~time~~ release time interval separately, resulting in 288 data points since there are 24 different ~~times~~ release time intervals and 12 observations. The set of overlap in densities is used to construct box plots in Fig. 5. The overlap in density is poor when using the *a priori* values for the uncertainty parameters (labels "InvG" and "Gaus", case 1 in Fig 5). While the outcome would have been entirely different if better *a priori* values would have been chosen, its significance should not be underestimated: (i) although seemingly realistic *a priori* uncertainty parameters have been chosen, the agreement with the ensemble density is poor; (ii) for this case and this particular choice of uncertainty parameters, a Gaussian distribution resulted in a better agreement; (iii) in the absence of an ensemble, using *a priori* chosen uncertainty parameters might be the only available option. When fitting the uncertainty parameters once for all observations and all ~~times~~release time intervals, the overlap in density is significantly improved (labels "InvG" and "Gaus", case 2 in Fig 5). The fit using Eq. 5 performs significantly better than the Gaussian fit. Additional yet smaller improvements are obtained when fitting for each observation separately (labels "InvG" and "Gaus", case 3 in Fig 5) and when fitting for each observation and ~~time~~ release time intervals separately (labels "InvG" and "Gaus", case 4 in Fig 5).

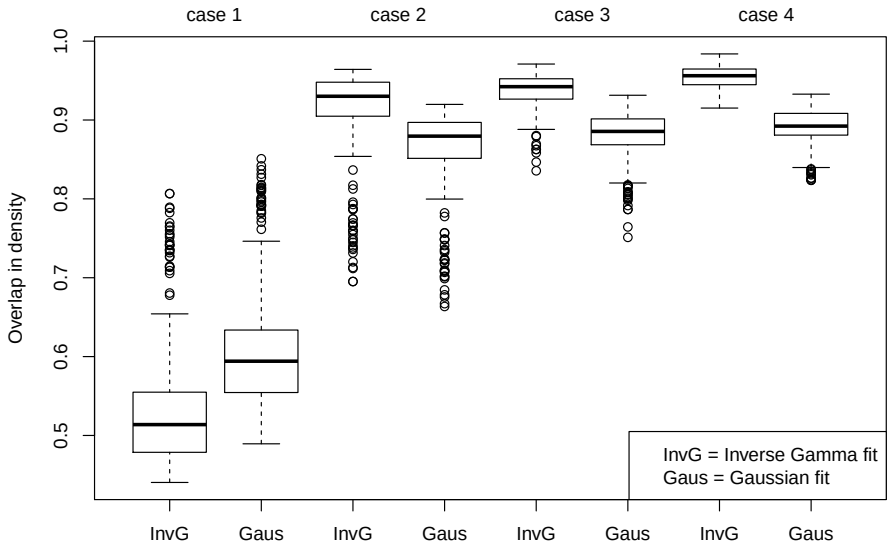

Figure 5. Box plots of the overlap between the ensemble densities and the fitted densities using uncertainty parameters obtained in four ways (case 1 to 4, see Subsection 5.2 for details) and using the distribution in Eq. 5 ("InvG") and the Gaussian distribution ("Gaus") for fitting.

5.3 Error dependency on the release time interval and the observation

In this subsection, it is assessed how the fitted uncertainty parameters vary among different observations and different ~~times~~release time intervals. The motivation for this is as follows: first, and somewhat trivial, we can expect the model uncertainty to increase as a function of simulation time. Second, uncertainties are expected to be observation-dependent, since observations are made on different times and at different distances from the source; uncertainties on the trajectories between the receptor and the source will also be affected by the atmospheric conditions along the trajectory, which are expected to be observation-specific. The interplay of the three uncertainty parameters s, $\bar{\alpha}$ and $\bar{\beta}$ of the inverse gamma function make it less feasible to directly estimate any effect. Therefore, only the fitted standard deviations of the Gaussian distribution are considered here.

The fitted standard deviation for each observation (averaged over all ~~times~~release time intervals) is shown in Fig. 6 (left). The standard deviation varies significantly between the different observations - up to a factor of 2 between the third and the eleventh observation. This can alter the posterior since the uncertainty scales how deviations between the simulated and observed activity concentration are penalized. While it appears difficult to assign meaningful observation-specific *a priori* uncertainty parameters, the ensemble can readily provide such information.

The fitted standard deviation for each ~~time~~ release time interval (averaged over all observations) is shown in Fig. 6 (right). It can be seen that the model uncertainty grows when going backwards in time. The growth is about 30 % during a three day period (note that such growth does not have to be constant over time, and an in-depth assessment is out of the scope of this paper). Also interesting to note is that there is an oscillatory behaviour with a period of eight time steps, corresponding to the

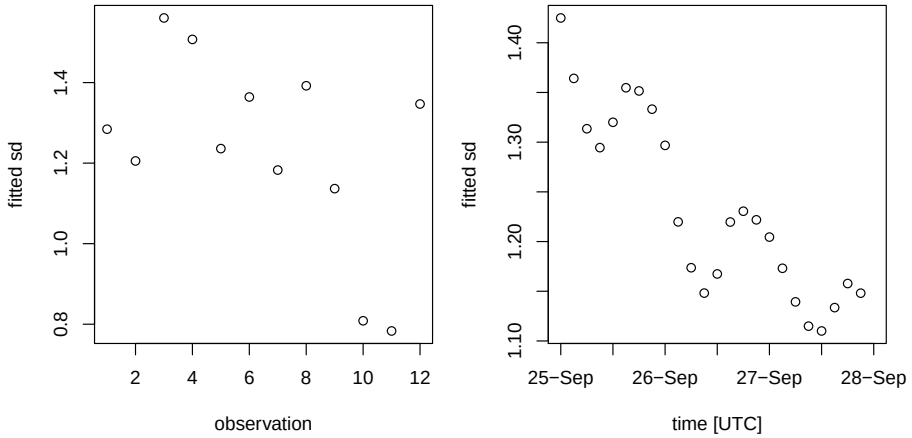

Figure 6. The standard deviation obtained after fitting a Gaussian distribution to the SRS ensemble. Left: standard deviation fitted for each of the twelve observations (all ~~times~~ release time intervals are aggregated); the indices of the observations on the x-axis are arbitrary. Right: standard deviation fitted for each of the twenty-four ~~times~~ release time intervals (all observations are aggregated); ~~the~~ each release time ~~step~~ interval is 3 h long.

diurnal cycle (since SRS fields were produced every three hours). The oscillations are likely associated with boundary layer processes, which often follow the diurnal cycle.

5.4 Resulting source location probability map

In this subsection, the Bayesian source reconstruction is applied using fitted observation-specific parameters s_i, $\bar{\alpha}_i$ and $\bar{\beta}_i$ (this corresponds to case 3, "invG" in Fig. 5). Furthermore, the inference is run using the SRS ensemble median for each geotemporal grid box - before, the unperturbed ensemble member was used. The resulting probability map is shown in Fig. 7 (a). While the source location probability map is distinct from (most of) the individual panels shown in Fig. 3, it is not too different from what Fig. 3 as a whole suggests. It best resembles case (f) of Fig. 3, which corresponds to the combination of multipliers with large model uncertainty. Note that simply changing the *a priori* uncertainty parameters might never yield a near-perfect correspondence with Fig. 7 (a), since the fitting resulted in different uncertainty parameters for each observation, which seems not possible to obtain without an ensemble.

For comparison, the average probability of the six panels in Fig. 3 is shown in Fig. 7 (b). The latter is perhaps the best approach one can take in the absence of an ensemble. While both maps roughly agree at first glance, there are still important differences. Foremost, the most south-west mode out of two modes in Fig. 7 (b) is absent in Fig. 7 (a). The other mode in Fig. 7 (b) is slightly shifted to the west in Fig. 7 (a), and extends further north-east.

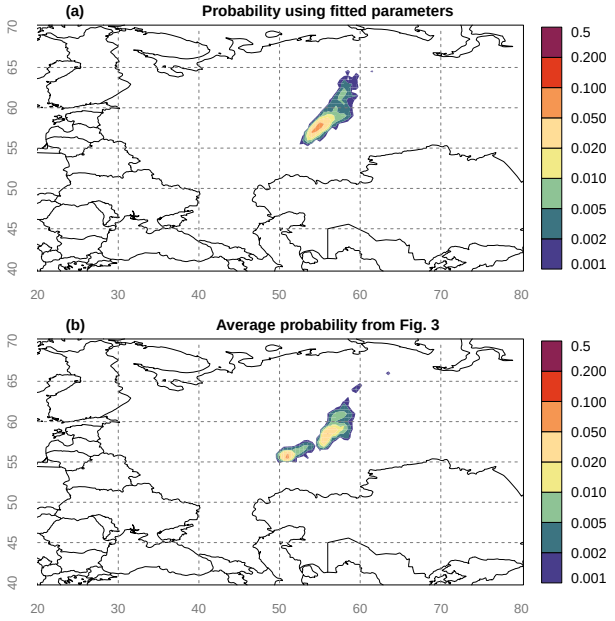

Figure 7. Source location probability map obtained from the Bayesian source reconstruction. Top (a): the parameters of the inverse gamma distribution were obtained using the ensemble SRS for each observation separately; bottom (b): the average source location probability is taken from all six panels shown in Fig. 3. The longitude and latitude are shown by gray numbers and dashed lines.

6 Ensembles as a set of scenarios

In this section, it is assessed whether additional information can be acquired by considering each ensemble member as an independent scenario, thus performing the Bayesian source reconstruction for each ensemble member separately. No fitting of uncertainty parameters is applied here, so that these need to be set *a priori*. The experiment is performed twice, once using $s_i = 0.5$ and once using $s_i = 3.0$. The other two uncertainty parameters remain fixed ($\bar{\alpha} = 1/\pi$ and $\bar{\beta} = 1$).

6.1 Source location probability maps

The Bayesian inference is repeated using the SRS of each ensemble member, so that 51 different posteriors are obtained. These posteriors then need to be aggregated in some way. As before, we focus on the source location probability. While several aggregation methods are possible, here the grid box-wise mean and maximum probability is taken (normalisation is required to ensure that the probabilities sum up to 1). Equal weights were assigned to each ensemble member, since our ensemble is constructed to yield equally likely scenarios or ensemble members. Note that in the case of multi-model ensembles, the latter might not be true, so that a weighting should be applied based on the skill of each model.

Fig. 8 (a, b) shows the results using the unperturbed member only, and using $s_i = 0.5$ and $s_i = 3.0$ – hence, these are identical to panels (c) and (e) in Fig. 3. As discussed earlier in Section 4, the source location probability map differs significantly when

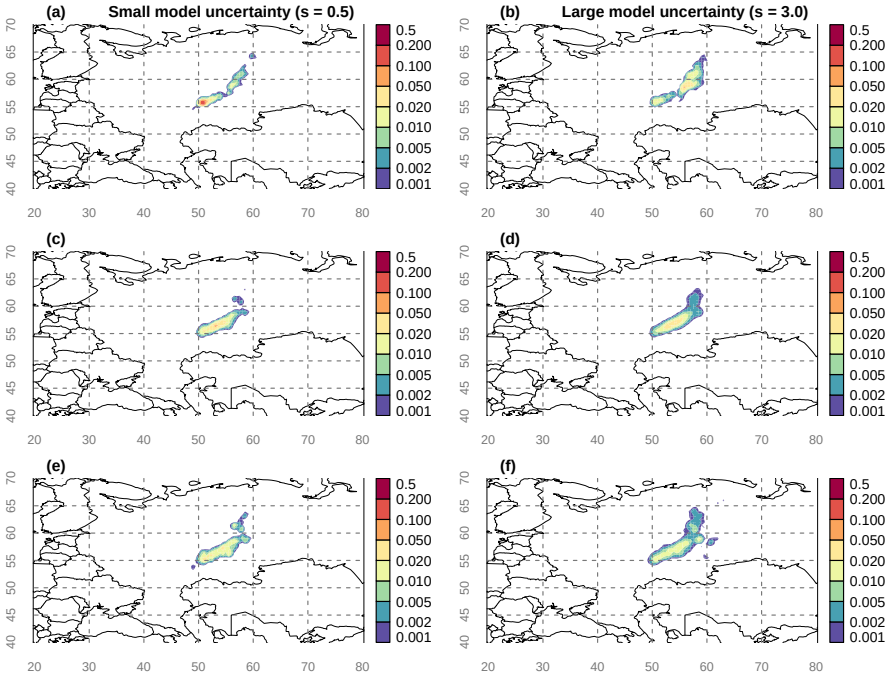

Figure 8. Source location probability maps for (a, b): the unperturbed member; (c, d): the grid box-wise ensemble mean; (e, f): the grid box-wise ensemble maximum. Panels (a, c, e) were obtained using $s = 0.5$; panels (b, d, f) were obtained using $s = 3.0$. The longitude and latitude are shown by gray numbers and dashed lines.

changing the parameter s. Fig. 8 (c, d) show the results for the grid box-wise ensemble mean using $s_i = 0.5$ and $s_i = 3.0$. The results are slightly broader and much smoother, not showing significant bimodal behaviour. Two features are noteworthy: first, the results roughly approximate those in Fig. 7, though important differences are present, as there were between Fig. 7 (a) and (b). Second, the results are generally insensitive to the choice of the uncertainty parameter s, since Fig. 8 (c) and (d) are very similar. The same is true for Fig. 8 (e, f), which shows the grid box-wise ensemble maximum. As one can expect, the resulting source location probability is slightly broader than the one obtained by taking the grid box mean probability; however, the differences are minor.

It seems that overall, a similar picture is obtained when running the Bayesian inference for each ensemble member separately, compared to the procedure explained in Section 5. This suggests that if we use the ensemble only (i) to fit the uncertainty parameters and (ii) to calculate the ensemble median SRS for running the inference as was done in order to obtain Fig. 7, no crucial information from the ensemble is lost with respect to the source location. As a consequence, it is equivalent to running the inference with all members of the ensemble separately to determine the uncertainty.

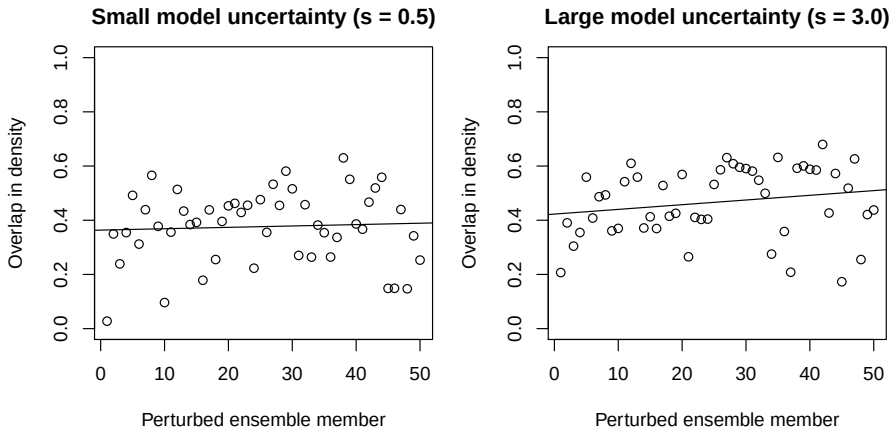

Figure 9. Overlap in source location probability as proxy for how much new information each ensemble member is adding to the ensemble mean (see text). Left: results obtained using an uncertainty parameter $s = 0.5$; right: results obtained using an uncertainty parameter $s = 3.0$. A linear fit is also plotted (solid black line).

6.2 Ensemble convergence

We finally perform a brief assessment on whether or not each ensemble member is adding new information to the ensemble mean source location probability. For each perturbed member $m \in [1, 50]$, the overlap in source location probability is calculated between that member and the mean of all previous members $[0, m-1]$ (with member 0 the unperturbed member). We thus start with the first perturbed member (which is compared with simply the unperturbed member) and end with the last perturbed member (which is compared with the mean of all other ensemble members). The results are shown in Fig. 9 for an uncertainty parameter $s = 0.5$ and an uncertainty parameter $s = 3.0$. Note that the overlap in density is calculated using Eq. 21 and has the same meaning as before, with 0 denoting no overlap (thus being fully informative) and 1 denoting full overlap (providing no new information).

Fig. 9 shows that most ensemble members have an overlap in density less than 0.6. There is significant variance in how much new information is added by each ensemble member. A linear fit suggests that the added information from additional ensemble members is slowly decreasing as expected. Note that the effect is more pronounced for the case with higher model uncertainty (Fig. 9, right). The reason for this is that the source location probability is more spread out due to the higher model uncertainty, making an overlap more likely. We conclude that all ensemble members are adding new information. This is desirable, as it shows that the ensemble is well-constructed (if members are generally not adding new information, they could be a waste of computational resources).

7 Conclusions

Model error has a huge impact on the posterior obtained through Bayesian source reconstruction, a conclusion in agreement with other studies (e.g., Yee et al., 2014). Specifically for the source location, an increase in the scale of the model error resulted in a non-uniform broadening and a shift in the source location probability.

Both the non-uniformity of the broadening and the shift in the source location probability imply that one cannot simply predict beforehand what the result will be when the Bayesian source reconstruction is repeated with different model error.

In the absence of a way to determine the model error, one could perform multiple Bayesian source reconstructions using different model error formulations as shown in Fig. 3. The results could be aggregated by taking the average (as in Fig 7, b) or by using more elaborate procedures (including a weighted mean).

Multipliers can be used to represent model error (as in Yee et al., 2014), or to represent the unknown part of model error as was done in Subsection 4.3. The multipliers result in a shift in the source location probability, but not in a broadening.

We found that the ensemble members of source-receptor-sensitivities are distributed around their ensemble median (Fig. 4) in a way that can be well-described by Eq. 5; the latter describes how the true activity concentration c_{true} are assumed to be distributed around the modelled activity concentration c_{mod}. Therefore, an ensemble of atmospheric transport and dispersion simulations can be employed to determine the parameters associated with the inverse gamma function in Eq. 16. Of course, the effectiveness of such approach largely depends on whether the ensemble is capable to represent model error.

The ensemble showed that model error varies among different observations (up to a factor 2 in the standard deviation when fitting a Gaussian distribution). Therefore, it is expected that having available model error information which is observation-specific can improve the quality of the Bayesian source reconstruction. The model error is also shown to increase when going further backward in time (for this specific case, there was an increase of 30 % during a three day period in the standard deviation when fitting a Gaussian distribution).

The source location probability using the fitted model error obtained from the ensemble (Fig. 7a) is distinct from the source location probability obtained using fixed uncertainty parameters (individual panels in Fig. 3); however, it is not too different from what Fig. 3 as a whole suggests, which demonstrates the usefulness of performing multiple Bayesian source reconstructions using different model error formulations as a sensitivity analysis in the absence of an ensemble.

A scenario-based approach (where each ensemble member is used as input for the Bayesian source reconstruction, instead of using the ensemble to fit the uncertainty parameters) gives results which are more robust against the choice of the uncertainty parameters but is more costly compared to directly fitting the uncertainty parameters. This is because the ensemble introduces model uncertainty that may predominate against the uncertainty prescribed by arbitrarily choosing the uncertainty parameter. No new information is obtained for the source location probability (or stated differently: one does not loose information when using the ensemble only to fit the uncertainty parameters and to calculate the ensemble median for use in the Bayesian inference). The scenario-based approach might be best in case of a small multi-model ensemble, since the fitting of uncertainty parameters might be difficult due to the difference in skill of each ensemble member.

In a future study, we will apply the different approaches and methods presented in this paper to situations in which the source characteristics are known unambiguously. This will help to better evaluate the different approaches proposed in this paper.

Code and data availability. The Flexpart model that was used to generate the SRS data is open source and is available for download (Flexpart, 2020). The SRS data from the Flexpart model is available on Zenodo (De Meutter and Delcloo, 2020). The Bayesian inference tool is available under the GPL3 and can be downloaded from https://gitlab.com/trDMt2er/frear.

Author contributions. All authors contributed to the conceptualization of the study. PDM conducted the simulations and performed the analysis. IH and KU supervised the research. All authors contributed to the manuscript. IH took care of the project administration.

Competing interests. The authors declare no competing interests.

Acknowledgements. The authors acknowledge funding from the Defense Research and Development Canada's Canadian Safety and Security Program through project number CSSP-2018-TI-2393. We would like to thank the reviewers for their constructive comments.

References

Becker, A., Wotawa, G., De Geer, L.-E., Seibert, P., Draxler, R. R., Sloan, C., D'Amours, R., Hort, M., Glaab, H., Heinrich, P., et al.: Global backtracking of anthropogenic radionuclides by means of a receptor oriented ensemble dispersion modelling system in support of Nuclear-Test-Ban Treaty verification, Atmospheric Environment, 41, 4520–4534, https://doi.org/10.1016/j.atmosenv.2006.12.048, 2007.

Bocquet, M.: High-resolution reconstruction of a tracer dispersion event: application to ETEX, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 133, 1013–1026, https://doi.org/10.1002/qj.64, 2007.

Bonavita, M., Hólm, E., Isaksen, L., and Fisher, M.: The evolution of the ECMWF hybrid data assimilation system, Quarterly Journal of the Royal Meteorological Society, 142, 287–303, https://doi.org/10.1002/qj.2652, 2016.

Bossew, P., Gering, F., Petermann, E., Hamburger, T., Katzlberger, C., Hernandez-Ceballos, M., De Cort, M., Gorzkiewicz, K., Kierepko, R., and Mietelski, J.: An episode of Ru-106 in air over Europe, September–October 2017–Geographical distribution of inhalation dose over Europe, Journal of environmental radioactivity, 205, 79–92, https://doi.org/10.1016/j.jenvrad.2019.05.004, 2019.

Currie, L. A.: Limits for qualitative detection and quantitative determination. Application to radiochemistry, Analytical chemistry, 40, 586–593, https://doi.org/10.1021/ac60259a007, 1968.

De Meutter, P. and Delcloo, A.: SRS data, data available for download at Zenodo via https://doi.org/10.5281/zenodo.4003640, 2020.

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

De Meutter, P., Camps, J., Delcloo, A., and Termonia, P.: Source localisation and its uncertainty quantification after the third DPRK nuclear test, Scientific reports, 8, 10 155, https://doi.org/10.1038/s41598-018-28403-z, 2018.

De Meutter, P., Camps, J., Delcloo, A., and Termonia, P.: Source Localization of Ruthenium-106 Detections in Autumn 2017 Using Inverse Modelling, in: Mensink C., Gong W., Hakami A. (eds) Air Pollution Modeling and its Application XXVI. ITM 2018. Springer Proceedings in Complexity., Springer, Cham, https://doi.org/10.1007/978-3-030-22055-6_15, 2020.

Engström, A. and Magnusson, L.: Estimating trajectory uncertainties due to flow dependent errors in the atmospheric analysis., Atmospheric Chemistry & Physics, 9, https://doi.org/10.5194/acp-9-8857-2009, 2009.

Flexpart: Available for download at https://www.flexpart.eu/ (accessed on 1 September 2020), 2020.

Harris, J. M., Draxler, R. R., and Oltmans, S. J.: Trajectory model sensitivity to differences in input data and vertical transport method, Journal of Geophysical Research: Atmospheres, 110, https://doi.org/10.1029/2004JD005750, 2005.

Hegarty, J., Draxler, R. R., Stein, A. F., Brioude, J., Mountain, M., Eluszkiewicz, J., Nehrkorn, T., Ngan, F., and Andrews, A.: Evaluation of Lagrangian particle dispersion models with measurements from controlled tracer releases, Journal of Applied Meteorology and Climatology, 52, 2623–2637, https://doi.org/10.1175/JAMC-D-13-0125.1, 2013.

Laloy, E. and Vrugt, J. A.: High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing, Water Resources Research, 48, 2012.

Leutbecher, M.: Ensemble size: How suboptimal is less than infinity?, Quarterly Journal of the Royal Meteorological Society, 145, 107–128, https://doi.org/10.1002/qj.3387, 2019.

Leutbecher, M. and Palmer, T. N.: Ensemble forecasting, Journal of Computational Physics, 227, 3515–3539, https://doi.org/10.1016/j.jcp.2007.02.014, 2008.

Liu, J. S., Liang, F., and Wong, W. H.: The multiple-try method and local optimization in Metropolis sampling, Journal of the American Statistical Association, 95, 121–134, 2000.

Masson, O., Steinhauser, G., Wershofen, H., Mietelski, J. W., Fischer, H. W., Pourcelot, L., Saunier, O., Bieringer, J., Steinkopff, T., Hyza, M., et al.: Potential source apportionment and meteorological conditions involved in airborne 131I detections in January/February 2017 in Europe, Environmental science & technology, 52, 8488–8500, https://doi.org/10.1021/acs.est.8b01810, 2018.

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, Proceedings of the National Academy of Sciences, 116, 16 750–16 759, https://doi.org/10.1073/pnas.1907571116, 2019.

Ringbom, A., Axelsson, A., Aldener, M., Auer, M., Bowyer, T. W., Fritioff, T., Hoffman, I., Khrustalev, K., Nikkinen, M., Popov, V., et al.: Radioxenon detections in the CTBT international monitoring system likely related to the announced nuclear test in North Korea on February 12, 2013, Journal of environmental radioactivity, 128, 47–63, https://doi.org/10.1016/j.jenvrad.2013.10.027, 2014.

Saunier, O., Didier, D., Mathieu, A., Masson, O., and Le Brazidec, J. D.: Atmospheric modeling and source reconstruction of radioactive ruthenium from an undeclared major release in 2017, Proceedings of the National Academy of Sciences, 116, 24 991–25 000, https://doi.org/10.1073/pnas.1907823116, 2019.

Seibert, P. and Frank, A.: Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode, Atmospheric Chemistry and Physics, 4, 51–63, https://doi.org/10.5194/acp-4-51-2004, 2004.

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

Steinhauser, G.: Anthropogenic radioactive particles in the environment, Journal of Radioanalytical and Nuclear Chemistry, 318, 1629–1639, https://doi.org/10.1007/s10967-018-6268-4, 2018.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmospheric Chemistry and Physics, 5, 2461–2474, https://doi.org/10.5194/acp-5-2461-2005, 2005.

Stohl, A., Seibert, P., Wotawa, G., Arnold, D., Burkhart, J. F., Eckhardt, S., Tapia, C., Vargas, A., and Yasunari, T. J.: Xenon-133 and caesium-137 releases into the atmosphere from the Fukushima Dai-ichi nuclear power plant: determination of the source term, atmospheric dispersion, and deposition, Atmospheric Chemistry and Physics, 12, 2313–2343, https://doi.org/10.5194/acp-12-2313-2012, https://www.atmos-chem-phys.net/12/2313/2012/, 2012.

ter Braak, C. J. and Vrugt, J. A.: Differential evolution Markov chain with snooker updater and fewer chains, Statistics and Computing, 18, 435–446, https://doi.org/10.1007/s11222-008-9104-9, 2008.

Vrugt, J. A., Ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., and Higdon, D.: Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, International Journal of Nonlinear Sciences and Numerical Simulation, 10, 273–290, https://doi.org/10.1515/IJNSNS.2009.10.3.273, 2009.

Yee, E.: Inverse dispersion for an unknown number of sources: model selection and uncertainty analysis, ISRN Applied Mathematics, 2012, https://doi.org/10.5402/2012/465320, 2012.

Yee, E., Hoffman, I., and Ungar, K.: Bayesian inference for source reconstruction: A real-world application, International scholarly research notices, 2014, https://doi.org/10.1155/2014/507634, 2014.