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".