Discussion: Quantification of uncertainties in the assessment of atmospheric release source with application to the autumn 2017 ^{106}Ru event

Joffrey Dumont Le Brazidec[1,2], Marc Bocquet[2], Olivier Saunier[1], and Yelva Roustan[2]

[1]IRSN, PSE-SANTE, SESUC, BMCA, Fontenay-aux-Roses, France
[2]CEREA, Joint laboratory École des Ponts ParisTech and EDF R&D, Université Paris-Est, Marne-la-Vallée, France

Correspondence: Joffrey Dumont Le Brazidec (joffrey.dumont@enpc.fr)

Note to the Editor and reviewers

A mistake in the meteorological fields used in this paper has been found by one of the co-authors. More precisely, the ldX Eulerian transport model codes used to compute the operators assimilated the rainfall fields as cumulative fields when they were in fact decumulated fields. The correction of this error and the subsequent reconstruction of the distributions a posteriori slightly modifies the conclusions of this paper:

- no changes in the position of the source reconstructed;
- the new estimated total release is higher (100-350 TBq before to 200-450 TBq now with the enhanced ensemble of observation operators);
- no changes regarding the interest of the observation sorting algorithm;
- with the HRES meteorological data, the use of several likelihoods which led to:
  - a major impact on the distribution of coordinates reconstruction;
  - a minor impact on the distribution of the total release reconstruction;

  now leads to a moderate impact on both variables' pdfs;
- in the case of the use of the ensemble of observation operators, the use of several likelihoods in combination with the strategy of integrating the weights of the operators still leads to:
  - a major impact on the distribution of coordinates reconstruction;
  - a major impact on the distribution of the total release reconstruction.

Report 1

We would like to thank Dr. De Meutter for his numerous and very interesting comments which allowed us to improve the quality of our manuscript.

Major comments

- 1/ I suggest to reformulate parts of Section 1.2: - "An efficient way to use these forecasts to better estimate uncertainties is to combine them": it is not clear to me what the authors mean here. Please be more specific. (If you would combine the members of an ensemble into a best estimate of the true state of the atmosphere, and use only that best estimate, you would loose the uncertainty information.)

  An answer to this question is given below (last answer of the first point).

- "This approach is known as multimodel ensemble forecasting (Zhou and Du, 2010).": It is not clear to me what "This approach" is. If you use the same model with perturbed input parameters and / or perturbed physics, I would not call it multimodel ensemble forecasting.

  We have specified "the approach of combining these forecasts (the forecasts where each realisation is the result of both the initials conditions and the numerical model pertubations) in a single forecast".

- What is "sequential aggregation"? Do you mean sequential in time (= an ensemble of deterministic forecasts run at different starting times)?

  We have deleted this term which does not add anything and can be confusing. Thank you.

- "An aggregated forecast is then formed by the weighted linear combination of the forecasts of the ensemble": I wonder why the authors want to create a best realization, rather than extracting the uncertainty from the ensemble. I suggest to add some discussion to explain the reasoning for this.

  The short state of the art on ensemble methods presented in the beginning of the paper does not introduce the method that we propose. The methodology proposed in this paper related to the estimation of the uncertainties linked to the meteorological and transport models is presented later in section 2.2. This first section acts simply as a short introduction on the topic of ensemble methods. We have reworded this part to make it less confusing.

- 2/ While the discussion on the choice of the likelihood function is valuable and interesting, I'm not sure if all arguments made in the paper are valid. - I'm not sure if I can agree with the discussion in Lines 111 - 117. There, you ignore the fact that the uncertainties will be larger for the higher observation-prediction couple; both observation-prediction couples could have the same penalty if they have the same relative uncertainty, which is not unlikely considering the error from the atmospheric transport and dispersion model. Therefore, I would rather say that the problem that you mention is a result of the oversimplified error covariance and not because of the Gaussian likelihood.

Firstly, the discussion is indeed a consequence of

- using a Gaussian likelihood,
- the simplification $\mathbf{R} = r\mathbf{I}$ (which is a very classic assumption).

We have recalled in the text a second time that $\mathbf{R} = r\mathbf{I}$ to make it clearer, this is indeed important. Thank you.
Secondly, we agree that the fourth criterion results from a "multiplicative" consideration of the error distribution, which is not necessarily true. However, we believe that this simplification is relevant in the case where a limited number of hyperparameters are available to model $\mathbf{R}$. We have specified in the text that this criterion relies on this underlying simplification.

- Similarly, you cannot make a statement about relativity in line 125 without considering the uncertainty.

  A sentence has been added to be more specific. Thank you.

- Note also that some authors use Gaussian likelihood, but work with ln(y) and ln(Hx)

  This is an other solution to solve the relativity problem which might be less elegant however (which should correspond to a log-normal likelihood in disguise).

- In Section 3.3.3, the posterior is shown for the different likelihood functions. It can be seen that the posteriors don't overlap too much for the longitude and latitude parameters, but they do overlap for the Total Retrieved Released Activity. While it is not explicitly stated in the paper, the operator H was calculated on a grid with grid spacings of 1°? That would imply that the likelihood for the location is extremely sharp, thereby hinting to an unphysically small uncertainty in the location for all considered likelihood functions. I would expect that, if the uncertainties would be larger for all likelihood functions, then they would overlap much more, as is already the case for the TRRA. The conclusion would then be that the likelihood has some impact on the posterior shape, but not too much, which is what I would expect a priori.

  In section 3.1.1, physical parametrisation : "The chosen mesh spatial resolution of G is 0.5 degrees for the simulations with ECMWF ERA5 HRES observation operator". In section 3.3, the HRES observation operator is used so the grid spacing is 0.5°.
  Furthermore, pdfs of the coordinates with the new HRES meteorological data have been modified and the overlap on the longitude and latitude pdfs has indeed increased (and the overlap on the TRRA distributions has decreased).
  However, we do agree with your a priori and we have the same intuition. More generally, if all uncertainties were "well" quantified, then the impact of the choice of the likelihood between log-normal, log-Cauchy, and log-Laplace should not matter too much since the overlap should be more important. Thank you for this interesting comment.

- 3/ Threshold values Line 157: "As a consequence, it can be deduced that a "good" threshold for the log-normal distribution in a case involving important quantities released should lie between 0.5 mBq.m^{-3} and 3 mBq.m^{-3} . " Could you give some explanation on how the values were deduced? I have a concern that the thresholds that are mentioned here are

large compared to instrumental detection thresholds, which might explain why many observations in Central Europe are non-informative (r = 0.09) in Fig 3. As an alternative, De Meutter and Hoffman (2020) formulated likelihood functions that explicitly consider detections, non-detections, false alarms and misses.

The values were deduced after computations of the penalties for diverse couples and diverse likelihoods with various thresholds. The computations can be found in table 1. We have however decided to not include this table in the paper in order to not over-complicate the discussion.

In our opinion, the role of the likelihood threshold is to explicit how small couples (i.e., couples with a close-to-zero or zero observation or prediction) should be compared to big couples. Furthermore, the variability of the instrumental detection tresholds of the ensemble of stations is very important. Therefore, in this paper, we solely focus on studying this comparison to choose the likelihood threshold. This "role" of the threshold becomes logical when we consider the cost function as a function that must judge how to guide the inverse problem.

In the configuration of table 1 provided, for a log-normal choice with a threshold of 0.5mBq.m^{-3}, the cost of a (prediction=1, observation=0)mBq.m^{-3} is 0.6 which is comparable to the cost of a (prediction=400, observation=100)mBq.m^{-3} which is 0.96. This shows that small differences are taken in account and will impact the reconstruction of the posterior distribution. A threshold of 3mBq.m^{-3} is an upper limit (and actually, it is the likelihood which yields a surprising pdf in the enhanced ensemble case).

	$(y_i, (\mathbf{H}\boldsymbol{x})_i)$										
y_t	(20,0)	(5,0)	(8,0.5)	(0.5,8)	(6,1)	(1,0)	(400,100)	(170,50)	(25,10)	(12,10)	(120,100)
Log-normal											
0.1	14.06	7.73	3.39	3.39	1.47	2.87	0.96	0.75	0.41	0.02	0.02
0.5	6.9	2.87	2.29	2.29	1.08	0.6	0.96	0.74	0.39	0.02	0.02
1	4.63	1.61	1.61	1.61	0.78	0.24	0.95	0.73	0.37	0.01	0.02
3	2.07	0.48	0.66	0.66	0.33	0.04	0.93	0.70	0.29	0.01	0.02
5	1.30	0.24	0.37	0.37	0.18	0.02	0.91	0.67	0.24	0.01	0.02
Log-Laplace											
0.1	5.30	3.93	2.6	2.6	1.71	2.4	1.39	1.22	0.91	0.18	0.18
0.5	3.71	2.40	2.14	2.14	1.47	1.1	1.38	1.22	0.89	0.17	0.18
1	3.04	1.79	1.79	1.79	1.25	0.69	1.38	1.21	0.86	0.17	0.18
3	2.04	0.98	1.15	1.15	0.81	0.29	1.36	1.18	0.77	0.14	0.18
5	1.61	0.69	0.86	0.86	0.61	0.18	1.35	1.16	0.69	0.13	0.17
Log-Cauchy											
0.1	3.37	2.80	2.05	2.05	1.37	1.91	1.07	0.91	0.6	0.03	0.03
0.5	2.69	1.91	1.72	1.72	1.15	0.79	1.07	0.91	0.58	0.03	0.03
1	2.33	1.44	1.44	1.44	0.94	0.39	1.07	0.9	0.55	0.03	0.03
3	1.64	0.67	0.84	0.84	0.51	0.08	1.05	0.88	0.46	0.02	0.03
5	1.28	0.39	0.55	0.55	0.31	0.03	1.04	0.85	0.39	0.02	0.03
Gaussien	200	12.5	28.12	28.12	12.5	0.5	inf	inf	112.5	2	200

Table 1. Several likelihoods and their corresponding cost for several observation-prediction pairs. The values of the observations y_i, the predictions $y_{S,i}$, and the threshold y_t are expressed in the same unit (e.g., in $\mathrm{mBq.m^{-3}}$). A dimensionless variance equal to 1 is used and normalisation constants are omitted in the calculation of quantities.

- 4/ Line 179: "Indeed, the error is a function of time and space and is obviously not common for every observation-prediction couple." I wonder why the authors do not prescribe the uncertainty on the observation and the prediction, and make it observation-specific? In De Meutter et al. (2021), the observation uncertainties are combined with the prediction uncertainties, which were obtained from an ensemble. As a result, the uncertainty on the input is no longer a parameter that needs to be inferred. Ideally, the distribution of these uncertainties should also be consistent with the likelihood function, which could be mentioned in Section 2.1. The principle of the methodology proposed in this paper is the hierarchical Bayesian methodology. The goal is to include as many parameters as possible in the vector of variables that will be sampled, because all the parameters are linked with each other. For example, let us consider r_i a hyperparameter of the error covariance matrix $\mathbf{R}$ corresponding to a couple observation-prediction $(y_i, (\mathbf{H}\boldsymbol{x})_i)$. If we assume a Gaussian likelihood (for simplicity), then a good estimation of this hyperparameter is directly proportional to $(y_i - (\mathbf{H}\boldsymbol{x})_i)^2$ (this result can be found by simply minimising the corresponding cost function). But $\boldsymbol{x}$ which here corresponds to the variables describing the source (location, vector of release rates) is not known and is subject to uncertainties. So $(\mathbf{H}\boldsymbol{x})_i$ might be a very variable value. Hence, r_i, to be well estimated, should be sampled with the other variables, in order to be chosen according to the sampled values of the coordinates and the release rates. Furthermore, a full hierarchical Bayesian methodology is also very practical because we "let the algorithm do the job". This methodology has of course some drawbacks : the estimation of uncertainties might be challenging. The proposal of this paper is to improve this estimation (this ambition is behind the algorithms 2.2 and 2.3).

  Furthermore, since our desire is to include the hyperparameters of $\mathbf{R}$ as variables, and since MCMC algorithms can only estimate a limited number of parameters, we had to make a compromise.

  Finally, we explain in a later answer why in a hierarchical Bayesian approach, we cannot make the estimation of the uncertainty observation-specific.

- 5/ There is limited discussion on the results using the spatial clustering (Lines 379-383). Could you provide some discussion, for instance whether you would recommend it or not, and why? And what is the effect of changing the threshold (please see also my comment 3)?

  These results have been slightly changed since the correction of the meteorological fields. The impact seems negligible in comparison with the other methods (even with 9 groups.) Therefore, we do not recommend it. Changing the likelihood threshold impacts how small couples (i.e., couples with a close-to-zero or zero observation or prediction) are compared to big couples.

- 6/ Enhanced ensemble It is not surprising that a pointwise comparison will give the result in Figure 6a: a ten-member global ensemble can only represent the uncertainty on large spatio-temporal scales (which is of interest here, since you do a long-range atmospheric transport and dispersion calculation). Also, it seems strange to suggest to compensate underdispersiveness in the weather data by perturbing the atmospheric transport and dispersion model. The latter has its own uncertainties which should ideally be taken into account. In Lines 419 - 428, the discussion is inconsistent with the (incorrect) motivation for perturbing ldX.

We agree that uncertainties exist both in the meteorological data and in the transport model. The motivation to perturbate parameters of the transport model is independent from the motivation to use an ensemble of weather forecast. We believe there might an unclear part in the paper which suggests that we want to compensate underdispersiveness but we are not sure we do understand which part? The discussion in lines 419-428 supposes that weather data member choice does not impact the posterior distribution whereas transport parameters perturbation does.

- 7/ In the conclusions, it is stated: "Moreover, we provided a method to add meteorological and dispersion uncertainties to the reconstruction of the distributions of a source, improving its evaluation." However, no improvement is mentioned or discussed in Section 3.3.4.

  We have added a sentence "In other words, the uncertainty emanating from meteorological data and the transport model is better quantified." and we had used an other sentence "In other words, the integration of weights member interpolation adds uncertainty not only over the magnitude of the release but also over the timing of the release (here, the day)." to explain why this methodology improves the sampling. The general idea is the following : in the previous paper (Dumont Le Brazidec et al., 2020) we had understood that variance of the distributions of the TRRA and the coordinates were underestimated. In this paper, the integration of pertinent uncertainties increases the variance of the TRRA and coordinates distributions. This might be counterintuitive to think that "adding uncertainties" is an improvement; this should rather be seen as : better estimating uncertainties.

Minor comments

- Line 6: Firstly,... Secondly, ..., Finally, ...

  Thank you for this, it has been corrected.

- Line 55: "modelling choices": it is not clear what is meant with this. Is it the atmospheric transport and dispersion model, or does it also include the likelihood and error covariance?

  This means transport model choice, meteorological data definition, likelihood definition (which includes the definition of the error covariance matrix). And in our opinion, it should also include the priors choices ... But this is not analysed in this paper.

- Line 56: (see also the above comment): "The objective of this study is to investigate the various sources of uncertainties compounding the problem of source reconstruction": but the title suggested modelling uncertainties, which I would associate to the atmospheric transport and dispersion modelling.

  Indeed, but we believe that the likelihood choice should also be considered as a modelling choice. However, to avoid confusion, we have changed the title from "Modelling uncertainties" to "Uncertainties".

- Line 58: "The quantification of the uncertainties largely depends on the definition of the likelihood and its components." Could you clarify this?

The sentence has been changed to « The quantification of the uncertainties largely depends on the definition of the likelihood and its components (for example, a corresponding covariance matrix) » in order to state what we meant in this case by « components ». Thank you.

- Section 1.4: the section numbering is confusing. I suggest to use more sections, for instance a new section for "Summary and Conclusions".

  Indeed, this is clearer with a new section for the final conclusion. Thank you.

- Line 108: "... the likelihood part of the cost should be zero and it should increase when the difference between the observation and the prediction values grows.": you mean the cost part of the likelihood.

  This was indeed unclear. The cost function is properly defined as the opposite of the logarithm of the posterior distribution. Therefore, the cost function is proportional to the sum of the cost corresponding to the likelihood and the cost corresponding to the prior. Here, we focus on the first part. We have changed the sentence to make it clearer.

- Page 5, criteria for the likelihood function: there is a contradiction between the first and the fourth criterion. The likelihood should indeed measure the difference between observations and predictions (fourth criterion), so that the positive support requirement becomes invalid (first criterion). If you consider the differences, I would rather suggest that it should be symmetric around its maximum, which should be at 0 (zero difference between observation and prediction).

  First criterion : by positive support, we mean that the likelihood function should be defined for values which are positive by nature : the « input » space should be positive. Fourth criterion : this is the « output space » which is indeed not necessarily positive or negative, it is equal to its maximum when the observation is equal to the prediction and symmetric around (which is the second criterion). We have added some precisions to make it clearer. Thank you.

- Lines 232 - 238: I suggest to omit this.

  These few sentences are for introducing methods used in application section. We have added a reference to the corresponding section.

- Table 1: the spatial resolution, vertical resolution and time resolution: this is for ldX and not for ERA5? Furthermore, ldX was run forward in time? With one simulation for each day and each grid point? And this grid had grid spacings of 1°, while the output grid spacings are 0.28125°? I suggest to make this information more explicit.

  These are ldX resolutions. It has been mentioned in the title of the table. ldXs runs foward in time, this had been added. Thank you. Simulations are indeed ran for each point in G (of spacing 0.5 degrees in HRES case and 1 degree in enhanced ensemble case). We have added more details to make all these pieces of information more explicit.

- Line 299-300: units are missing for the variances.

  Indeed, the unit on the coordinates variance was missing. This has been added. Thank you.

- Line 301: "When the algorithm to discriminate pertinent observations presented in section 2.2 is used, ..." What are "pertinent" observations? Previously, you used the terms "discriminant" and "non-discriminant"?

  The term « pertinent » has been deleted for clarity. Thank you for the remark.

- Line 333-334: units are missing for the error variances.

  The error variances have actually no dimension in the case of the use of the log-normal, log-Laplace, or log-Cauchy likelihoods.

- Line 341-342: same as above.

  (no dimension)

- L 368: Figure 4c should be Figure 4b.

  Indeed, thank you.

- L 379: Figure 4b should be Figure 4c.

  Indeed, thank you.

- Line 405: "... and the standard deviation (std) of the joint multi-model TRRA is therefore far more important than the std of the joint HRES TRRA." What is the meaning of standard deviation here? And what do you mean with "important"?

  We have been more careful with the term "important". By standard deviation we simply mean the square root of the variance.

Thank you for these minor technical comments and the very interesting discussions.

Report 2

We thank the reviewer for his/her constructive comments, revisions and technical comments, which allowed us to clarify some points of the paper.

The manuscript presents an evaluation of the impact on assumptions surrounding statistical model selection on posterior estimates of source properties of (unknown) radiological releases. This manuscript will be informative to researchers and operational users. I have tried to avoid repetition if the previously posted comment. One major drawback of the manuscript is that much of the motivation is based on a straw-man argument, i.e. the original Gaussian set-up is designed in the manuscript such that it will fail. Below are a number of suggestions for revisions to improve the manuscript, followed by technical comments.

- Paragraph starting line 110: This is a straw-man argument. The assumption from the start is that the error is larger for higher measurements. This may not always be valid, e.g. an incorrectly dispersed wide plume with a very high concentration. It may be that the error is much larger for the smaller measurements than the larger measurements. There are other approaches to improve the validity of Gaussian (or any) likelihoods through transformation of variables. For example, using a non-linear forward model. Caveats, justification and 'typical errors' needs explaining, preferably at the start of section 2.

  We agree that the fourth criterion results from a "multiplicative" consideration of the error distribution, which is not necessarily true. However, we believe that this simplification is relevant in the case where a limited number of hyperparameters are available to model $\mathbf{R}$. We have specified in the text that this criterion relies on this underlying simplification.

- Much of the arguments surround having independent and identically distributed (iid) model-measurement error in the covariance. Many of the arguments throughout can be countered by the use of non-iid covariances, e.g. tR, where the diagonal of R is the measurement value and t is a scaling – equivalent to having e.g. a 10% model-measurement error. This needs further discussions and better justification for the arguments proposed (e.g. non-negativity).

  This is true, our argument is indeed based on the hypothesis of modelling $\mathbf{R}$ by some hyperparameters. However, it should be noted that the variance hyperparameters of the matrix $\mathbf{R}$ are estimated through a hierarchical Bayesian approach inside the MCMC procedure. In other words, it is not estimated independently of the MCMC in an empirical Bayesian approach. More precisely, if we note $\boldsymbol{x}_S = (x_1, x_2, \ln \mathbf{q})$ the vector of variables describing the source except for the hyperparameter r, we try to estimate :

$$\begin{aligned} p(\boldsymbol{x}_S, r_1, ..., r_N|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{x}_S, r_1, ..., r_N)p(\boldsymbol{x}_S, r_1, ..., r_N) \\ &\propto p(\boldsymbol{y}|\boldsymbol{x}_S, r_1, ..., r_N)p(\boldsymbol{x}_S|r_1, ..., r_N)p(r_1, ..., r_N) \end{aligned} \tag{1}$$

  and

$$p(\boldsymbol{x}_S, r_1, ..., r_N|y) = p(\boldsymbol{x}_S|y, r_1, ..., r_N)p(r_1, ..., r_N|y) \tag{2}$$

If we model $\mathbf{R}$ with a term t of scaling model-measurement, we suppose that the values $r_1, ..., r_N$ depend on t, i.e., we suppose that these values depend on the difference between the observations $\boldsymbol{y}$ and the predictions $\boldsymbol{y}_S = \mathbf{H}\boldsymbol{x}_S$. However, this difference depends on the definition of $\boldsymbol{x}_S$. In other words, with a definition of $\mathbf{R}$ depending on t then $r_1, ..., r_N$ depend on $\boldsymbol{x}_S$. This would not be rigourously defined in a hierarchical Bayesian formalism.

Furthermore, even if we agree that $\mathbf{R} = r\mathbf{I}$ is a simplification and that better and more rigorous modelling are possible, this is a very classic choice in the literature.

- Section 3.2.3: It would be useful for many readers to provide a brief conceptual introduction to MCMC methods (i.e. asymptotically exact methods not reliant on closed form solutions or conjugacy).

  A few sentences have been added.

- Section 3.4: This section needs expanding considerably. It is a paper with lot of content. At current, the summary provides an overview of the approach of the paper but no summary of the finding. The summary should summarise the results and findings to adequately inform the lazy reader on the paper's content.

  This is true and we have extended the conclusion accordingly. Thank you.

- I understand the following would require a lot of extra work, and so do not mandate it for publication. It would however, much improve the paper. Seeing as an ensemble is used, it would seem sensible to me to use a simulated dataset (a simulation using an ensemble member) to draw conclusions from the various experiment. It is not perfect, but would be useful to have a "truth".

  Indeed, we agree that this may be interesting to use a synthetic case. However, we decided in this paper to focus on the development of methods rather than on the maximum exploitation of the results on the ^{106}Ru case.

- It would be useful in the analysis and plot to also show the original case of a Gaussian likelihood. This is needed to prove the worth of using a non-Gaussian likelihood.

  We believe that the relevance of a non-Gaussian likelihood can be proved in theory but not on a particular case (Liu et al., 2017). Inverse modelling with a Gaussian likelihood might provide pertinent results in some cases : the problem is that these results are based on the use of the few largest observations instead of the whole dataset. The information in some of the data is not exploited, which can lead to discrepancies between the reconstructions in some cases. The longitude and latitude distributions are reconstructed with a Gaussian likelihood in figure 1. We can note that the results differ significatively from the ones obtained using log-normal, log-Laplace, or log-Cauchy likelihoods.

Technical comments:

- Title: "modelling uncertainties" is ambiguous as it can refer to a statistical model or a transport model. The title also is not grammatically correct. A suggested improvement is "Quantification of uncertainties in atmospheric release source assessment applied to the autumn 2017 Ruthenium 106 source".

  We have changed the title following the suggestion of the reviewer. Thank you.

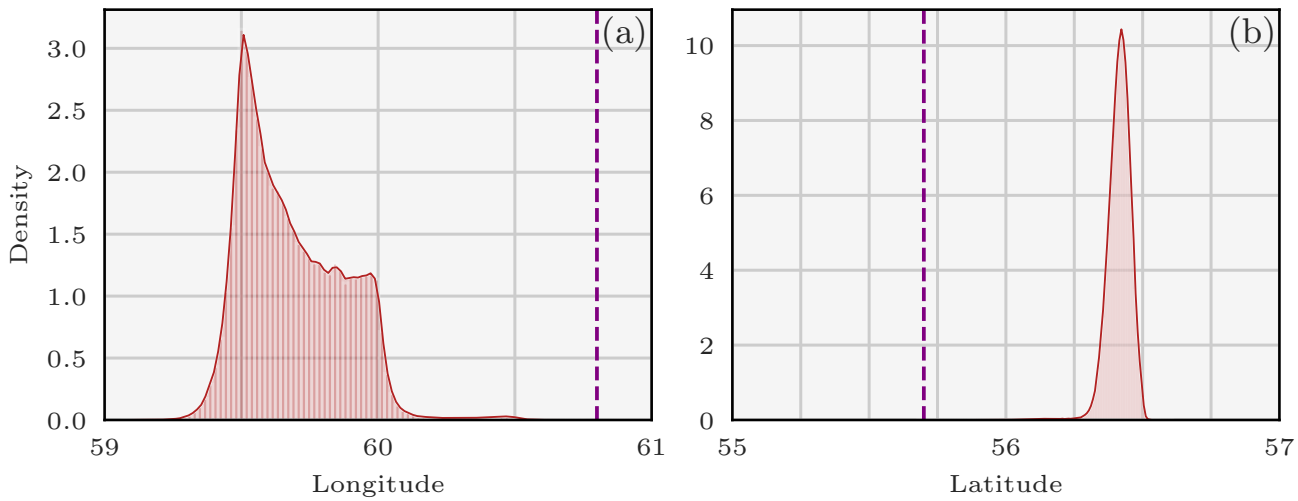

Figure 1. Density of the coordinates describing the ^{106}Ru source for a Gaussian likelihood of the Ruthenium source sampled using the parallel tempering method with the help of the observation sorting algorithm and the HRES meteorology..

- Abstract, line 5: "improve on these distributions" is vague. 'Better quantify' or 'improve estimates of these distributions' would be better.

  We have corrected the sentence following this suggestion.

- Line 7: A space is not needed before a colon in English.

  This has been corrected. Thank you.

- Line 8: Clarify 'model errors' (I assume transport errors?)

  We have added « physical » to clarify.

- Line 10: 'several suited distributions for the errors are advised' the passive voice reads as though you are advising the distributions. Better "we suggest several suitable distributions for the errors" or "are suggested" if sticking with the passive.

  Active voice in now used. Thank you for this advice.

- Line 17: sources or a source; remove 'many'; 'Therefore' doesn't follow – delete.

  This has been corrected, Thank you for these corrections.

- Line 31: 'finds it origin in' to 'originates from'

  This has been corrected. Thank you.

- Line 38: This is an oversimplification of the weighting strategy. See, for example, importance resampling or Ensemble Kalman filter methods. We use a linear weighting strategy in order to be able to include the weights as variables of the ensemble to be sampled.

- Line 46: Why index vector x elements with x1, x2 and then ln(q)? Perhaps x3=ln(q) would be clearer.

  Classic coordinates names are (x,y) but y name is already used for describing the observations vector. $\ln \mathbf{q}$ being a vector, this might be confusing to use x_3.

- Line 47: R is better described as the covariance matrix containing the model-measurement errors.

  This has been corrected.

- Line 52: Introduce for the reader what the prior is (i.e. the probability distribution of prior knowledge before considering data).

  This has been added.

- Line 57: reconstructed posterior distribution

  This has been corrected. Thank you

- Line 60: 'transformation' is better than 'parameterisation'

  This has been corrected.

- Line 61: 'are the results' would be better as 'are the results of a simulation'

  This has been corrected.

- Line 62: 'and are therefore depending' to 'and depend on'

  This has been corrected. Thank you a lot.

- Line 70: This isn't an expansion.

  We mean that the computation time is high.

- Line 103: A cost function is a non-probabilistic concept and so better to refer to as simply the negative log-likelihood.

  The authors do agree but the term of "cost" or "loss" is actually widespread in the statistic, data assimilation, or inverse problem literature. The use of the term "cost" rather than "negative log-likelihood" allows us to save a lot of space considering the number of occurrences of the term. Furthermore, the cost can correspond to the negative of the log-posterior distribution.

- Equation 4: There should be no divide by 2 in the first term.

  This comment has been removed by the reviewer.

- Line 112: A space not full stop is needed between units

  These have been corrected. Thank you.

- Line 115: Capital G on Gaussian.

  This has been corrected.

- Line 120: Unless there has been a transform (e.g. ln(y)). Square bracket is facing the wrong way. We do not understand. Observations can be positive or equal to zero.

- Line 123: Space between units.

  This has been corrected. Thank you.

- Line 156: 'Large multiple'

  This has been corrected.

- Line 178: 'as this paper'

  This has been corrected.

- Line 278: What are the upper/lower bounds of the uniform distribution?

  This has been added. Thank you a lot.

- Line 280: What are the shape parameters of the log-gamma distribution?

  This has been added.

- Line 348: 'Harmful' is an incorrect choice of work here. You can just delete it;

  This has been removed.

- Line 348-350: This sentence isn't correct. Observations don't have a high likelihood. Please rephrase.

  This has been corrected.

- Line 351: Change 'totally legitimate' to 'valid'

  This has been corrected.

- Line 353-355: This sentence does not make sense. I'm unsure of its meaning, please revise.

  This has been corrected. Thank you.

- Line 368: 4c and 5a

  This has been corrected (4b and 5a).

- Equation A1 and A2: Second term is incorrect, not divide by 2 but multiplied by 2.

  This comment has been removed by the reviewer.

Thank you very much for all these technical comments.

Report 3

We thank Dr. Tichý for all his valuable comments, recommendations, and interest in the paper.

The manuscript presents interesting study based on estimation of atmospheric release from ambient concentration measurement coupled with atmospheric model. Few prior models of a release are presented, discussed and evaluated on Ruthenium 106 case from 2017. Here, there is consensus on release location and approximate release time-profile which makes this case very interesting and a playground for model testing. The manuscript is nicely written and clear to understand. What I lack is clarification and verification of some statements. I also recommend to extend conclusion (or discussion) by some suggestions and recommendations for future cases, see specific comments bellow. In sum, I would recommend the paper for publication after these clarifications.

Specific comments:

- p. 5, line 115: Although I understand the importance of lower values in measurements, there might be a good reason for high significance of higher values since they may bring more confident infromation wiht lower uncertainty, especially in case with spatialy and temporaly long transport.

  We agree that the fourth criterion results from a "multiplicative" consideration of the error distribution, which is not necessarily true. However, we believe that this simplification is relevant in the case where a limited number of hyperparameters are available to model $\mathbf{R}$. We have specified in the text that this criterion relies on this underlying simplification.

- Figure 2: I am curious whether similar results are obtained for latitude. Considering the dominant direction of the atmospheric transport is probably in longitude axis, it is maybe different in latitude axis. Please, comment.

  The difference in std for the latitude is indeed lower. We have chosen a special case where the difference is very important. Please see figure 2 for the difference between the latitude distribution with observation sorting algorithm (for two different threshold values) and without observation sorting algorithm. In most scenarios, and for most variables, the difference between the stds of reconstructed distributions with and without the application of the sorting algorithm is about 10 to 30 $\%$.

- p. 14, line 375: Regarding temporal profiles of the estimated release, what is exactly the time-resution of the posterior, is it one day? Is it possible to plot the release profiles somehow, e.g. using medians or similar? Did you estimate some significant activity also in other days except 25th and 26th September?

  The release is estimated as a vector of daily release rates. Each release rate on each day between the 22nd and the 28th is estimated as a 1D variable. We plot the evolution of the release rate (using medians and variances) on figure 3. No significant activity is estimated in other days except for the 25th or 26th as described in our first paper (Dumont Le Brazidec et al., 2020). There is however an exception in the case with the enhanced ensemble of observation operators,

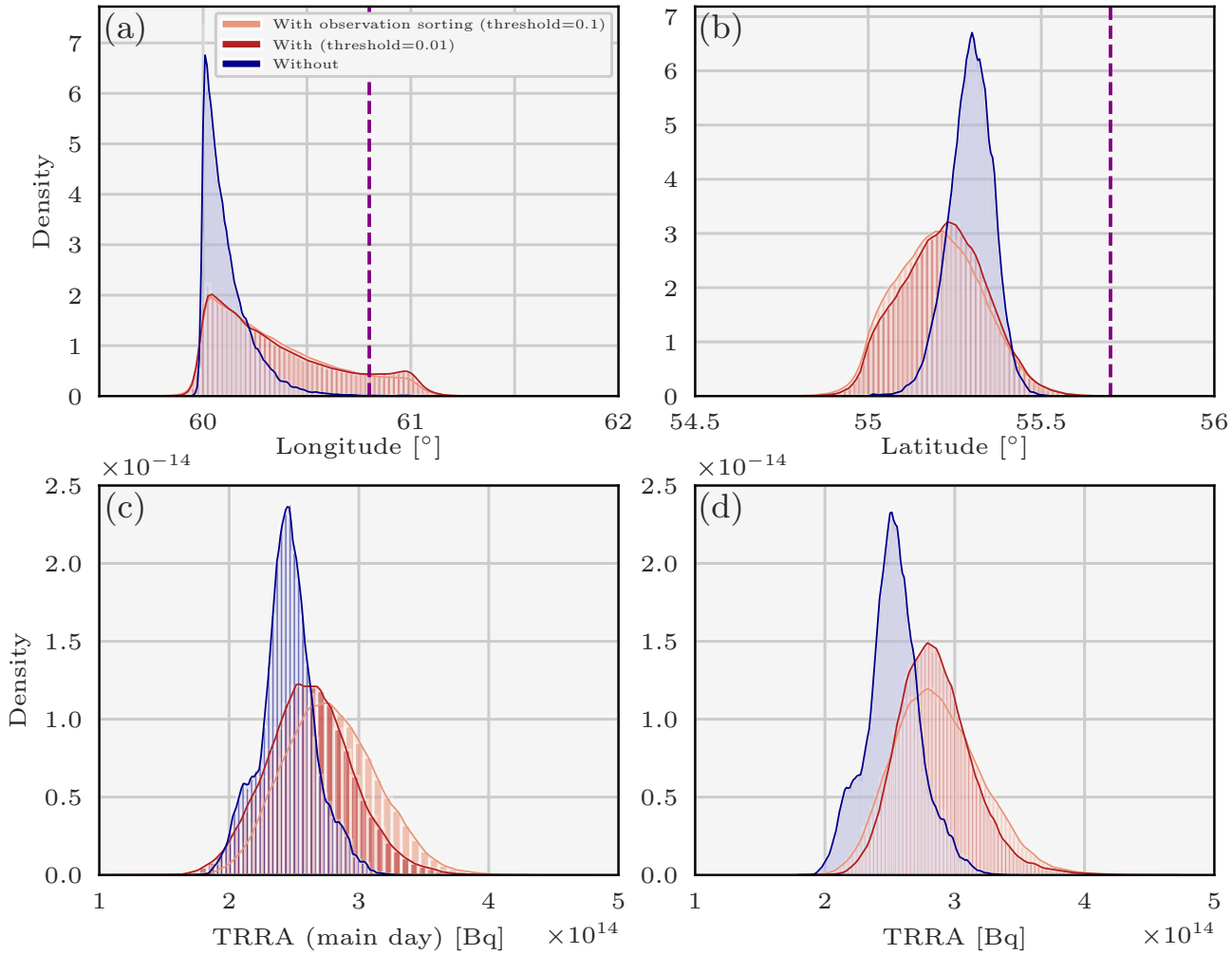

Figure 2. Density of the variables describing the ^{106}Ru source for a log-Laplace likelihood with threshold $y_t = 0.1\,\mathrm{mBq.m^{-3}}$ of the Ruthenium source sampled with the enhanced ensemble of observation operators using the parallel tempering method with or without the help of the observation sorting algorithm. When applied, the threshold of the observation sorting algorithm is $\epsilon_d = 0.1\,\mathrm{mBq.m^{-3}}$ or $0.01\,\mathrm{mBq.m^{-3}}$.

with a log-normal likelihood of threshold 3mBq.m^{-3} (the coordinates of the source distribution is displaced and so is the release.)

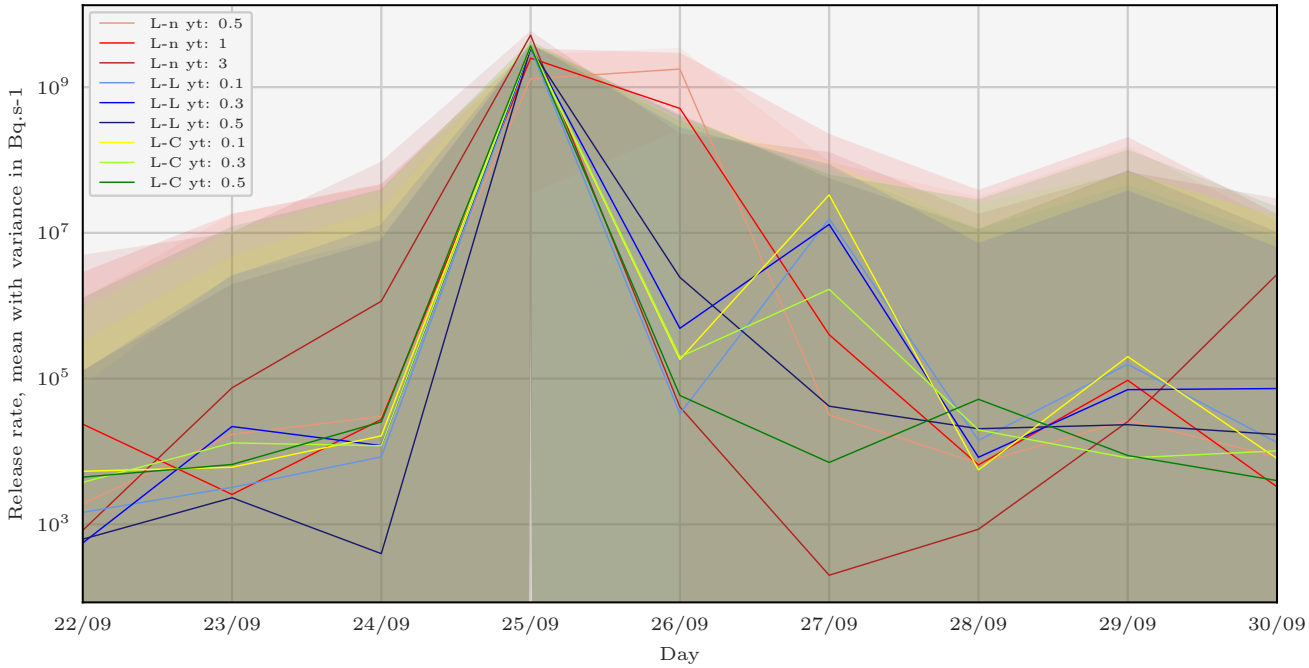

Figure 3. Evolution of the release rate of ^{106}Ru for several likelihoods with several thresholds sampled with the enhanced ensemble of observation operators using the parallel tempering method with the observation sorting algorithm. Line represents the median. Area with less opacity represents the area between the mean added to the standard deviation and the mean minus the standard deviation.

- p. 17, line 398: Could you please clarify the choice of reference source in [60, 55] while Ozyorsk (near where the plant is located) is located at 60°43' E 55°45' N and modelled spatial resolution is 0.5 degree? Shouldnt it be closer point [61.5 55.5]? Or maybe you have different numbering, please, clarify.

  We used the source of the Saunier et al. (2019) retrieved using inverse problem methods instead of the Mayak facility as reference.

- In general, I lack discussion and recommendation what settings should one choose when situation similar to the Ru-106 case occur in the future and, let say, one location and one total of the release need to be reported. Also, are your findings rather general, or case specific?

  We believe the most trustworthy results are to be found in the Figure 7 where all methods are applied at once. If one had to give a single location, it must be at the peak of most longitude distributions (60 degrees) and latitude distributions (55.15 degrees). For what concerns the release magnitude, the total of the release posterior distribution maximum is

around 300-350 Tbq. But the beauty of Bayesian results is that we have access to a full range of possibilities instead of a single optimum!

Furthermore, we believe that the methods that we exploit are rather general. Each of them can be applied on any type of accidents.

Technical corrections:

- p. 3, line 63: S should be defined in term $\mathbf{y}_S$, probably Sth obsevation

  We are not sure to understand. S goes for simulation if this is the question (with $\boldsymbol{y}$ being the observations vector, $\boldsymbol{y}_S$ is the predictions vector or simulated observations vector).

- Eq. (2): the $\mathbf{x}$ is used as the source term previously while $\mathbf{q}$ it is used here. Please, clarify whether they are the same of have different meanings.

  x is the ensemble of variables describing the source that we want to sample and q is the source term : the vector of release rates. q is included in x.

- Eq. (5a-5c): norm with two indexes in subscript should be defined (although clear for many readers, for many may not).

  The definitions have been added. Thank you.

- p. 5, line 141: "Secondly, a location term appears..." I am not sure what you mean by this statement, please, clarify.

  We have clarified this sentence.

- p. 6, line 163: there sould be (5c) instead of (6), probably.

  This is not the original definition of the log-Cauchy cost. We have added a threshold term to the log-Cauchy definition in this equation.

- Figure 8: there are missing labels (a) - (f) in subplots.

  This has been added. Thank you for all your comments.

Thank you a lot for these interesting comments.

Report 4

We thank the reviewer for his/her technical comments and general advices to improve the manuscript.

This paper does exactly what it says it will do: source term estimation for an emission of Ruthenium using observations, an atmospheric model, and Bayesian inversion. It excels at explaining the concepts involved, making it especially accessible to someone who has not done this exact type of problem before. However, many corrections are needed to the wording, particularly in section 2, and the organization, particularly in section 3.

Lines or references were missing in the report, so not every comment could be taken in account in the manuscript.

Comments:

- In the climate modeling community we would refer to an initial conditions ensemble of the same atmospheric model as a “single-model ensemble” rather than a “multi-model ensemble.”

  We have changed the sentence according to this comment. Thank you.

- Missing “a”

  We did not find where the "a" was missing.

- 60-63. Comment: I like this concise explanation of how a model, source, observations, and likelihood fit together.

  Thank you for your kind words.

- Suggested “These three sources of uncertainty are explored in an application of source term estimation for the 106Ru release. . . ”

  We have changed the sentence.

- state of the art of

  We did not understand this comment.

- 89-94. Can you rephrase this so that it flows monotonically, i.e. reference section 2 before section 3?

  Reference to section 2 is simply used as a reminder. But the topic of section 2 is already described before the topic of section 3.

- Missing a word here, which obscures the meaning of the sentence.

  We do not know where to find the missing word.

- It may be helpful for the reader if you reference the section in which the threshold is discussed.

  We removed the passage « and will be discussed later on » as the threshold discussion is two paragraphs after.

- If the observation sorting algorithm is the division into r and rnd, then you should not start a new paragraph for sentence 191.

  Thank you, we have modified this part.

- 268, 274. “22nd”

  Thank you, we have modified this part.

- This summary section should be clarified if possible. For uniformity, I recommend starting each bullet point with a section number, e.g.

  Section 3.3.2 is an application of the observation sorting algorithm. . . ; Section 3.3.3 is an application of the different likelihood functions and spatial clustering . . . ; Section 3.3.4 is an application of the perturbed dispersion parameters and enhanced ensemble. . .

  Indeed, this part gains to be clarified. Thank you.

- Secondly, the section heading “Summary” section seems out of place, especially since you have a summary section later. I would suggest renaming 3.3 “Results” and renaming 3.3.1 “Overview.”

  This seems indeed necessary. Thank you.

- “Probable sources”

  This has been corrected.

- “which is not justifiable.

  This has been corrected.

- ”Explain when and where this accident took place, and maybe ad some thoughts about how this might compare to what you just did.

  We are again very sorry, but we were not able to find what the comment was referring to.

- 436-440. I think more discussion would be helpful for the reader. Remember, many readers skim the paper until they get to the conclusions.

  This has been added. Thank you a lot.

- The math is correct but the wording is not quite right. I think you mean that r is a positive coefficient and R (and rI) is a positive diagonal matrix; r itself is not a “positive diagonal coefficient.”

  This has been corrected.

- 111-112. Suggest something like…choosing Gaussian likelihood penalizes the largest errors to an extent that smaller errors are negligible.

  Thank you, this has been corrected.

- Rephrase. Consecutive sentences starting with “in other words.”

  We have deleted this sentence.

- I think you should delete the sentence starting with “Every “ as the wording is confusing. Your example (100,120) vs (10,12) has already made this point.

  Indeed, we have also deleted this sentence. Thank you

- Bracket typo.

  We did not find where was the typo.

- What do you mean by mitigated here? I think you can say “should be 1” or “should be close to 1.”

  We remove "mitigated".

Thank you for these technical comments.

References

Dumont Le Brazidec, J., Bocquet, M., Saunier, O., and Roustan, Y.: MCMC methods applied to the reconstruction of the autumn 2017 Ruthenium-106 atmospheric contamination source, Atmospheric Environment: X, 6, 100 071, https://doi.org/10.1016/j.aeaoa.2020.100071, 2020.

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