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)

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.

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.