Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2020-205-AC2, 2020

Atmospheric Measurement Techniques Discussions

AMTD

Interactive comment

Interactive comment on "Real-time measurement of radionuclide concentrations and its impact on inverse modeling of ^{106}Ru release in the fall of 2017" *by* Ondřej Tichý et al.

Ondřej Tichý et al.

otichy@utia.cas.cz

Received and published: 1 December 2020

We would like to thank you for providing us a detailed review of our manuscript. We are glad that we can submit a revision of our paper. In the following text, we will respond to all comments.

This paper presents an inverse modelling study of airborne Ru-106 detections made in the Czech Republic in the fall of 2017. An existing inverse modeling algorithm, already successfully applied in earlier studies, is used and the results are compared (and found compatible) with other studies that considered

Printer-friendly version

Discussion paper

the Ru-106 detections. Furthermore, in this study the authors also describe two new detection systems. The motivation for the new detection systems is that they will provide observations with a higher temporal resolution, which is obtained by reducing the sampling time (CEGAM system) or by measuring during sampling (AMARA system). Finally, the authors consider different data sets to perform the inverse modelling. They conclude that inverse modelling using data with shorter sampling times (thus having a higher temporal resolution) performs equally well or better than inverse modelling using data having sampling times of a few days up to one week. This paper is relevant and results are compatible with previous studies, although I think the conclusions related to the added value of the new measurement systems are not well supported by the results. Furthermore, the chosen case study – although being a very important and interesting case – is likely not well-suited to fully demonstrate the added value of such systems given the large geotemporal scales of the Ru-106 release (having source-receptor distances of thousands of kilometers).

Specific comments:

In the abstract, the authors wrote: "Since reasonable temporal resolution of concentration measurements is crucial for proper source term reconstruction, the standard one week sampling interval could be limiting". Although it is sensible that better temporal resolution will lead to better source reconstruction, I'm wondering how important the limiting effect is. The effect is likely case-dependent, and in particular more pronounced for problems with shorter geotemporal scales. In that light, the Ru-106 case might not fully demonstrate the added value of short sampling times. A test with a fictitious source and fictitious measurements would be instructive (one test at scales of a few hunderds of kilometers, and another test at a few thousands of kilometers). The fictitious

experiment could demonstrate and quantify the limiting effect of long sampling times in a more controlled way.

Authors response: We agree with the reviewer that the Ru-106 case is not a perfect match to study the influence of a fast measuring system in details. We prefer to avoid a synthetic study since its results would be sensitive to our simulation setup. The Ru-106 event was the first significant release with a fully operational AMARA system and it is also well studied in the literature which allows discussion of the obtained results. To provide more solid evidence on the added value of the fast measurements, we extended the paper by additional simulation using the FLEXPART atmospheric transport model (see details in other responses below). The new simulation using a completely different simulation tool resulted in the same conclusion, supporting our previous claim that was based on a single model (HYSPLIT) and thus could have been obtained by chance.

Changes made in the paper: We employ the FLEXPART model to the same datasets to demonstrate that the results are not obtained by change but are systematic.

The AMARA and CEGAM measurement system descriptions are not clear to me. Specifically: p 5, line 9: "The achieved MDAC for Ru-106 is at a level of 1 mBq/m3 per one-hour integration time and 12 hours of sampling." Does this mean that an activity concentration measurement is available every hour, and that the filter is renewed every 12 hours? And for the CEGAM system, a measurement is available every 4 hours, and the filter is renewed every 4 hours? What is the philosophy of having two different systems, and will both systems be used and maintained in the coming years? p 5, line 8: the reference to Fig 2 is slightly confusing since no 4-hour averaging is applied for the AMARA system?

Authors response: There are two time intervals affecting the final MDAC - the duration of measurement and the duration of the sampling. Spectra in the AMARA system are measured (sampling duration) every 5 minutes and therefore arbitrary sums (measurement duration) could be computed afterwards. On the other hand, CEGAM system is limited by the time step of carousel mechanism therefore the sampling duration is equal to measurement duration.

Changes made in the paper: We have extended section 2 in order to better describe the measurement/sampling logistics. The difference between both systems and their intended use is also briefly discussed.

p 6, line 5: "Unfortunately, the CEGAM system was not yet operational during the Ru-106 incident but we have simulated its output by integrating the AMARA results in a 4 hours window." Some additional information would be helpful here. If CEGAM pseudo-observations are used based on the AMARA observations, then they would contain the same information? I assume the simulated output is not used for the inversion, but it would be good to confirm this in the text.

Authors response: Indeed, the use of simulated output would be pointless since it would contained the same information. All data used for inversion are based on AMARA system, the CEGAM system was set to operational regime later. We agree that the sentence was not clear and we state clearly this fact in the current version of the manuscript.

Changes made in the paper: We reformulated the sentence to avoid misunderstanding.

Section 2.3: Dataset description: I think it would be good to add a figure or table that summarizes the different datasets (range of observed activity concentrations, number of observations, number of (non-)detections. After consulting the Supplementary Information, I am a bit worried that the differences between data set "RAW" and "FAST" are too small to be significant. Also, why is the integration window set to be between 3 and 13 hours? From Sections 2.2.1 and 2.2.2, I

expected that measurements from the AMARA system would be available every hour (and measurements from the CEGAM system every 4 hours)?

***Authors response*:** Thank you for pointing this out, we agree that graphical representation of measurements would be instructive. Since the main differences can be observed in the case of the Prague station (equipped with the AMARA system), we provide a figure that summarizes measurements from this station.

The integration window was set adaptively to maintain the sufficient response to the Ru-106 activity. Difference between the real-time measurement values and values obtained by the measurement of the whole filter in laboratory was within approx. 15 %. This error margin is also compatible with our previous findings where we compared the laboratory values and real-time values of natural Be-7.

***Changes made in the paper*:** We added a figure with visualization of measurements from Prague as well as related description in the text.

p 9 line 5: it would be instructive to get an estimate of the values used for σ^2_{length} in the calculation of the inverse covariance matrix R.

***Authors response*:** Indeed, we miss out to define the $\boldsymbol{\sigma}_{length}$ coefficient in the text which is now corrected. It is defined as $\boldsymbol{\sigma}_{length} = \frac{measurementhours}{6}$ where the 6 hours window is motivated by the GFS data resolution. Varying the length of this window does not affect the results significantly.

***Changes made in the paper*:** We define this coefficient in Sec. 3.2 in the revised manuscript.

Section 4.1: Atmospheric transport modeling: Numerical weather prediction data, which is used to drive the atmospheric transport model Hysplit, was available every 6 hours. This is likely sufficient for the geotemporal scales of the

problem. However, it might not if one wants to explore the added value of measurements with sub-daily sampling periods.

***Authors response*:** We agree that the 6-hours resolution of the meteorology may be limiting and may somehow blur the results. Therefore, we run a new simulation using the FLEXPART model driven by 3-hour meteorological analyses from the European Centre for Medium-Range Weather Forecasts (ECMWF) and, subsequently, we select 3 hours temporal resolution of the output grid. We conclude from the estimated source terms that better temporal resolution of measurements improves the temporal specificity of the source term. We demonstrate this in the case of source term estimation from the most probable location 2, Mayak, in Sec. 4.4.

***Changes made in the paper*:** We extend the paper by FLEXPART simulation with higher temporal resolution. The FLEXPART configuration is given in Sec. 4.1.2 while the results for the most probable location, Mayak, are given in Sec. 4.4.

Table 3: Can the authors think of any reason why the release length is significantly different for the four considered locations when using the data sets "WEEKS" and "CUT", but not when using the datasets "RAW" and "FAST"? If one does not assume a priori that a short release period is better, Table 3 could be interpreted as if data set "FAST" gives less information regarding the release duration than "WEEKS", as it is less sensitive to the location. Also, I guess that the regularization will have a larger impact on the release duration than the choice of the data set. From these considerations, I am not convinced that the real-time monitoring data results in a better temporal specification of the release, as stated in Conclusions.

***Authors response*:** The LS-APC algorithm was designed to minimize the number of tuning parameters (they are estimated from the data) leaving its result sensitive only to the initial conditions. As demonstrated on the ETEX dataset in the original publication, it is rather insensitive even to the initial conditions. Further confirmation can be found recently in (TichÃ¡, O., Ulrych, L., Šmídl, V., Evangeliou, N., and Stohl, A.: On the tuning of atmospheric inverse methods: comparisons with the European Tracer Experiment (ETEX) and Chernobyl datasets using the atmospheric transport model FLEXPART, Geosci. Model Dev., 13, 5917–5934, https://doi.org/10.5194/gmd-13-5917-2020, 2020.).

In fact, LS-APC assigns a higher prior probability to shorter releases than to longer ones. The preference is rather weak and informative data overrule this prior. However, this is probably the reason for different lengths e.g. in the case of the WEEKS dataset. When the observations could be explained by a shorter release, LS-APC considers it a more likely solution.

***Changes made in the paper*:** We extended discussion of the results in Sec. 4.2 and also add a reference to the sensitivity study of the used LS-APC algorithm to Sec. 3.1.

Figure 7 is important for assessing the quality of the inverse modeling results that were obtained using different data sets, by comparing simulated activity concentrations with the IAEA measurements. However, Figure 7 seems to suggest that the temporal resolution of the observations do not really matter for this case. Perhaps other metrics might reveal an improvement from the use of higher temporal resolution, but I doubt that that will be the case for this specific case study (large geotemporal scales and 6-h meteorological data). In the same Figure 7, data set "RAW" performs slightly worse than data set "WEEKS". Do the authors have an explanation for that? From temporal resolution considerations, I would expect that "FAST" performs equally well or better than "RAW", and "RAW" equally well or better than "WEEKS". Also, from Figure 4 and knowing the true source location, I do not see why the results using the "FAST" data set would be better than the results from the other data sets. Concerning the

table in Figure 9, I wonder whether other metrics would come to different conclusions (the NMSE, although widely used, is not unbiased, see Poli and Cirillo, 1993 - Poli, A. A., & Cirillo, M. C. (1993). On the use of the normalized mean square error in evaluating dispersion model performance. Atmospheric Environment. Part A. General Topics, 27(15), 2427-2434.). These considerations make it hard for me to agree with the statement made on p 17 line 5.

***Authors response*:** We agree that the agreement with the IAEA measurements on former Figure 7 is rather insensitive to the choice of the temporal resolution. Therefore, we perform an additional simulation using the FLEXPART model with 3-hour temporal resolution and present these results in the updated version of the manuscript. The results from the FLEXPART runs are summarized in Sec. 4.4 for location 2, Mayak. We believe that the temporal specificity of the FAST dataset is better demonstrated there. Although the FLEXPART slightly overpredicted some of the IAEA observation, the estimation using the FAST dataset provides the best fit.

We are grateful to the reviewer for pointing out deficiencies of the NMSE coefficient. In the current version of the manuscript, we use four coefficients: the normalized mean square error (NMSE), the normalized mean square error of the distribution of the normalized ratios (NNR) suggested by Poli and Cirillo, and also other coefficients: the figure of merit in space (FMS) and the fractional bias (FB). In all cases, the results by the FAST dataset are the closest to the IAEA result.

***Changes made in the paper*:** First, we extended the manuscript by the FLEXPART simulation with finer temporal resolution and study the results for location 2, Mayak, in Sec. 4.4. Second, we extended Sec. 4.3 by additional coefficients, NNR, FMS, and FB.

p 17 line 1: how are the probabilities of the source location calculated? Is it the

evidence / marginal likelihood, but normalized so that its sum over the whole domain equals 1?

***Authors response*:** The marginal log-likelihood is normalized using the maximum of each domain, hence, the maximum of each normalized domain is equal to 1. This information was missing in the manuscript and we added it to the present version.

***Changes made in the paper*:** The information on the normalization of the displayed marginal likelihood is added to Sec. 3.3.

In Conclusions, the authors wrote: "It is safe to state that the installation of multiple devices such as AMARA and CEGAM over a larger region (on European scale) would certainly yield additional improvements in source location and in source term estimation in the event of a radionuclide atmospheric release." There is a trade-off between detector sensitivity and the sampling length (more observations will have a higher minimum detectable concentration). I suggest to briefly discuss this trade-off also in the conclusions. Also, although I agree that there is potential in using observations with higher temporal resolution, I don't think that its added value is clearly demonstrated in this study.

***Authors response*:** The trade-off mentioned by the reviewer is now discussed in section 2, we acknowledged the reviewers' remark. We also mentioned in conclusion that there is possible limitation of the continental scale scenario, however, we believe that the effect of real-time monitoring system is still observable.

***Changes made in the paper*:** We extended section 2 significantly and we also discuss some issues regarding the scale of the experiment in conclusion.

Minor issues:

p 3, line 1: location –> localisation

***Authors response*:** Thank you, we corrected this typo.

p 10, line 13: “. . . and run for the period . . .” –>“. . . and release particles during the period ...”

***Authors response*:** We reformulated this accordingly.

p 12, line 16: “The estimated source terms are displayed for the RAW dataset using blue lines, for the WEEKS dataset using magenta lines, for the FAST dataset using red lines, and for the CUT dataset using green lines.” –> I suggest to omit this sentence as this is already mentioned in the caption of Figure 6.

***Authors response*:** Indeed, we removed the color code description from here.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2020-205, 2020.