Title: Detailed source term estimation of the atmospheric release for the Fukushima Daiichi Nuclear Power Station accident by coupling simulations of atmospheric dispersion model with improved deposition scheme and oceanic dispersion model
Authors: G. Katata et al.
MS No.: acp-2014-330

Author response to reviewer comments

Response to Ref.4

This paper presents a very detailed analysis of the complex release and dispersion situation for the Fukushima reactor accident performed mainly by the Japan Atomic Energy Agency JAEA. It provides both a better methodological background and an update of the preliminary source term of Chino et al. (2011) which has found widespread use. Therefore, this work is a useful addition to the body of literature on the meteorological aspects of the accident.

➔ We appreciate your crucial comments and suggestions on our paper. We tried to respond to suggestions in detail as much as possible, but we sometimes could not bring the whole sentences from the manuscript to this letter because they are too long to cite in the letter. We showed the chapter, section, or subsection number, so please also see the revised manuscript.

MAJOR COMMENTS

1. The paper is very long and not easy to read, even though certain important aspects are still not sufficiently covered. I would recommend that the paper is revised in a way that it would be more systematic, focus on the most important aspects, and would defer minor aspects to the Supplement. For example, the discussion of the single phases on different days of the accident could be trimmed down, moving a systematic description to the Supplement. However, this description should then be really systematic, best in the form of a table with standard information for each phase, possibly also related figures.

On the basis of all reviewer comments, we made the following revisions to reduce the paper length summarized as:

- In Introduction, the sentence after ‘First, …’ of ln.5/ P.14730 was shorten.
- Chapters 3 and 4 are reconstructed. 3.1.1-3.1.9 which describes the source term and the motion of plume during March was revised and only the periods which were different from the previous paper (Terada et al. 2012) were remained. Related this revision, Figs. 5 and 6 were deleted. As for chapter 4 (Discussion), section 3.2 (Verification of source term) was moved to section 4.1. Section 4.1 (Comparison of source terms) was moved to section 4.2. Finally, section 4.2 (Role in deposition process) was moved to Supplement.
- The section for “Validation using several models” was shortened. Related this revision, (old) Figs. 18, 20, and 22 and Tables 7, 8, and 10 were extracted.
- Conclusion was also shortened based on reviewer’s comments.

At the same time, we needed to include more sentences to reflect all reviewers' comments to the manuscript (e.g., Meanwhile, Fig. S3 was moved to the main text as (new) Fig. 12). However, the page numbers and numbers of figures and tables were still reduced compared with our previous paper.

The authors give now more insight into their method of source determination. This is really important as previous publications have not been very explicit on this topic. However, the presentation should still be improved, and be better placed in the context other similar work, as is detailed below.

2. The authors use both the terms "reverse estimation" and "inverse estimation". It seems that the latter is reserved for the part of the source reconstruction using concentration data in ocean water. The authors should explain what they mean by these terms and consider established technical terms. In applied mathematics and related sciences, the term "inverse problem" and derived from it "inverse method / modelling / . . . " is the standard language (however, as will be explained below, the method does not correspond to a formal inversion).

➔ According to the reviewer comment, we defined the reverse and inverse methods with merit and demerit and also described why we choose the reverse method for the Fukushima case at the first part of chapter 2 as "A reverse method evaluates the release rates of radionuclides by comparing measurements of air concentration of a radionuclide or dose rate in the environment with calculated one by atmospheric transport, dispersion and deposition models (ATDM) for a unit release of a radionuclide. The release rate is estimated by the ratio of the measurement to calculation result. The merit of the reverse method is that the comparison can be made with one or more independent data points. For example, the minimum number of data points needed is only one and the measured data used for the estimation can change with time from air concentration to air dose rate and vice versa. The demerit is that this simple comparison without consideration of the uncertainty of the ATDM results may cause the large errors, and, consequently, expert judgment is essential to correct the discrepancy between the measurement and calculation." Also, we changed the description for the method used to assess the source term over the ocean from "inverse" to "reverse".

3. The authors explain that for the land data, they proceed as follows, with my questions in brackets:

(a) Divide the time axis into intervals [Obviously, this is largely subjective. What is the role in this division of steadiness of meteorological conditions, steadiness of the plant state, steadiness and/or availability of measurements?]

➔ In this study, we divided the release interval and looked for when and where the specific segment of plume increased air dose rate, and then found the appropriate observation data which can be used for the source term estimation for the certain plume. Thus, the timing of releases should be clear. The ground-shine was used for the source estimation on 12 and 15-16 March. For the case of 12 March, the release period for wet venting of Unit 1 was determined from the decrease of pressure of drywell and the release by hydrogen explosion is assumed almost instantaneous. For the case of 15-16 March, the release period was basically divided with every hour and the source term for each plume was estimated by the method as mentioned above. The descriptions how to find the appropriate observation and how to determine the release period of the segment of plume are available in (new) subsections 2.1.1 and 2.3.3.

(b) Only a single measurement available for the respective release period: pair maximum of both measurement and model to determine source. [Does it really occur often that only a single measurement is affected? This seems unlikely. Deposition data are always available, even though they were not used here. And how can one ascertain that only a single release phase impacts this measurement? What if this is not the case? What is the time and space window considered for matching the maxima? Even though this method should increase the robustness of the result, it could also be a major source of error if something is paired which does not match.]

➔ Yes, the dust sampling in Fukushima Prefecture was carried out by limited number of monitoring cars and consequently, the data are not temporally successive and the number of data was small per day. Exceptionally, at JAEA-Tokai, temporal variation of air concentrations was observed continuously (subsection 2.2.1). Thus, the number of measurement points for air concentration which are the most suitable for the source estimation was small, and furthermore the data which successfully measured the concentration of plume was almost one or two in the same time. The number of available data can be found by looking for when and where the specific plume increases air dose rate by the WSPEEDI simulation (subsection 2.1.1). Unless we get the temporal variation of deposition, it cannot be used for the reverse estimation. We do not intend to estimate the source term for each event in each reactor. As you know, the source term estimated by reverse estimation is always the total release rate from three reactors. However, when the peak position in both measurement and calculation which assumed the release of specific event is the almost same, it is natural to think that the measurement values is mainly due to the specific event. The window of errors between calculation and measurement are up to three hours at JAEA-Tokai.

(c) If multiple measurements are affected, average both measurements and model. [Will again only the peaks be considered, or will whole time series over some interval be averaged? What is the justification of this approach – should one not rather average the resulting ratios?]

➔ The comparison of peak values is carried out for JAEA-Tokai data, because time series data are only available at this place. The air concentration data from JAEA-Tokai is one hour averaged one. When we use the reverse estimation method, "Since the uncertainty of model simulation is the primary cause of the discrepancy in the spatiotemporal distribution of plume between the measurements and simulation results, the above procedures cannot be applied systematically, and the correction of this discrepancy by 'expert judgment' is necessary to reduce the impact of model uncertainty on the source estimation. The process is to check all available measurements to see if the plume is reproduced appropriately or not for comparison with the measurements, and to determine if the discrepancy is caused mainly by errors in the calculated wind direction. If the plume flow direction is clearly different from the measured wind direction, the calculated plume is rotated to match the measured wind direction and Eq. (1) is applied. The use of peak values corrects any discrepancy in the timing of the arrival of the peak air concentrations between the measurement (JAEA-Tokai) and simulation. We assume that the peak values of the measurement and simulation are comparable even though the timing or temporal pattern of the arrival of the peak is different because the central plume axis passes across the sampling point differently between the measurement and simulation" (new subsection 2.1.1).

4. Usage of dose rates: is it true that only 5 nuclides are considered? Which fraction of the total dose rate can they explain? How does this fraction depend on time and distance?

➔ We estimated the importance of other nuclides for air dose estimation, and recognized it was needed to consider I-133. Thus, we re-estimated the source term considering the release of I-133 in the period using air dose rate (ground-shine) data. These points are added to subsection 2.3.2.

5. The paper is quite obscure concerning how decay is considered, which is relevant for iodine and other short-lived nuclides. This pertains for example to Figure 2 and specifically 2c as well as the adopted I-131/Cs-137 ratio. It must be clearly stated, for all the data (measured, modelled, release) whether they are decay-corrected (and to which time) or not. If not, how can the ratio be assumed constant over days? Is the ratio only used when the source estimate is based on dose rates, or throughout? I tried to analyse this nuclide ratio from the release data in the supplement and found that after ca. 8 days, it suddenly drops from 10 to 5. It may be just coincidental, but this looks like decay has been ignored for 8 days, then some I-131 measurements come in and the ratio drops corresponding to the decay. Soon after, the ratio jumps by a factor of almost 100 within very short time. Unless a convincing explanation is given for that, I don't think that these ratios and thus the iodine releases can be considered reliable, even though I admit that the comparison with the airborne survey is a good support for the source term. However, as also the deposition parameterization has been strongly modified, and as the reconstructed source term has such a large temporal variability, this evidence is not totally conclusive. I am also surprised by the large scatter of the ratios presented in Figure 2, they need a proper explanation, including uncertainties of the measurements.

➔ Our idea for radioactive decay was described in (new) subsection 2.3.3 as "The ratio of ^{137}Cs to ^{131}I at the released time should be different from that at the measurement time because of radioactive decay during the transport of the plume and the difference of deposition processes of both nuclides in the environment. However, the transport time period between the FNPS1 and the monitoring points used to determine ^{137}Cs/^{131}I ratio (Fig. 4c) are within about 10 hours and sufficiently small compared with decay constants of both nuclides (Table 5). Thus, we only considered the latter effect to adjust the ratio obtained at the measurement points to that at the release point." Concerning the ^{137}Cs/^{131}I ratio, we can also estimate the ratio of inventory in the reactors with time considering the decay constants of each nuclide. However, to estimate the ratio when discharged into the atmosphere, it is the only way to believe the ratio measured in the environment even though the ratio highly varies with time for transport and deposition processes. We are also surprised at such large variations in observational data (old Fig. 2), but we do not have further information to clarify the reason, such as the error of measurement and reactor conditions related to the volatility of each radionuclide.

6. The releases tabulated in the supplement contain three physical species of iodine. It should be explained in the paper how they were derived, which measurements were available, and what the uncertainties are for this subdivision.

➔ The ratio of gaseous and particulate iodine is determined from the measurement at JAEA-Tokai. Regarding iodines, "because there are no observed data on the ratio of elemental iodine (I_2) and organic iodine (CH_3I), the ratio of gaseous CH_3I to I_2 is assumed to be constant of 0.6 throughout the simulation period according to the method of RASCAL 4.0 (US NRC, 2012)" (subsection 2.3.3). As written in (new) Conclusion, the ratio "can cause the errors in estimated results affected by wet deposition (Fig. S5b)". The estimation of source term for iodine is also affected by this uncertainty.

7. The explanation in Section 2.3 is difficult to follow. It should be more formalized, but with proper notation. It is not helpful how variable names Ci, Co, Mo are used. It would be better to use a standard notation such as c for concentration, with a subscript such as, e.g., m and o for model and observation, and not to use the letter C for something which isn't concentration – better call the correction factor r or f or similar. Also, if the index j is occurring everywhere, it can just be dropped. Why is a log function used for averaging? Also, some variables are written upright, others in italics, etc.

➔ We modified the equations and these explanations in subsections 2.1.1 and 2.1.2.

8. As visible in the supplement, various assumptions for the source geometry are made for different release periods. Some explanation on that is needed in the methods sections.

➔ We added the detailed explanation to (new) subsection 2.3.2 (Source assumption).

9. A discussion of the method for source estimation is required. It should include reference to less subjective, formal methods for solving inverse problems related to atmospheric dispersion which are available and partly have already been applied to the Fukushima accident. It should then be explain why the authors believe that their own method, which requires a lot of subjective decisions and manual intervention, is preferable. I do understand that in this situation, very complex both with respect to the source processes and the meteorological phenomena, with data of different type and degrees of quality, a selection guided by all the available knowledge can be very useful in obtaining a robust result, and agencies tasked with this emergency have certainly accumulated a lot of such knowledge. On the other hand, there is also the risk of mistakes in the process and of underusing available information through the lack of a comprehensive method. Personally, I would believe that it should also be possible to use a more comprehensive, formal inverse method together with such knowledge. In any case, the impression should be avoided that the approach used here would be the best or only one for dealing with unknown source terms in a nuclear emergency. It would be good if the authors would frankly explain their experiences and rationale for their approach, and to document their knowledge so that others can also use it (see my suggestion above, for a systematic coverage of the release in the Supplement).

➔ At the first part of chapter 2, we defined the reverse and inverse methods with merit and demerit and also described why we choose the reverse method for the Fukushima case as "A reverse method evaluates the release rates of radionuclides by comparing measurements of air concentration of a radionuclide or dose rate in the environment with calculated one by atmospheric transport, dispersion and deposition models (ATDM) for a unit release of a radionuclide. The release rate is estimated by the ratio of the measurement to calculation result. The merit of the reverse method is that the comparison can be made with one or more independent data points. For example, the minimum number of data points needed is only one and the measured data used for the estimation can change with time from air concentration to air dose rate and vice versa. The demerit is that this simple comparison without consideration of the uncertainty of the ATDM results may cause the large errors, and, consequently, expert judgment is essential to correct the discrepancy between the measurement and calculation.” Also, we changed the description for the method used to assess the source term over the ocean from “inverse” to “reverse”.

10. The comparisons based on the regional-scale WMO calculations show, all in all, lower scores when the new source term is used compared to Terada 2012. Interestingly, NAME with ECMWF data performed best, even though ECMWF is said to have precipitation not well reproduced in this case. There is no discussion of these findings, however, they may indicate a kind of overfitting – or were WMO models tuned to the Terada 2012 source term?

➔ Our statistics table using WMO models confused the reviewer. There was no tuning to WMO models for Terada et al. (2012) as “To evaluate the new source term independently of the one dispersion model used to develop the source term, numerical simulations from three atmospheric dispersion models were compared to observations using our new source term estimates” in subsection 4.1.2. Thus, it was not surprising that the performance of each model in (new) subsection 4.1.2 widely varied because these used the source term derived from the different model (WSPEEDI). We should focus on how the new source term works on atmospheric dispersion simulations to some extent. Thus, we extracted all of tables appeared in (new) subsection 4.1.2 and focused on the overall agreement between calculations and observations when using the new source term as “Then, the new source term is further tested using different atmospheric dispersion and meteorological models over regional- and global-scales to evaluate its reliability for general atmospheric dispersion model studies during the FNPS1 accident.” in section 4.1. By the way, the discussion of NAME with different meteorological fields was extracted from the maintext to reduce the manuscript length.

11. Global HYSPLIT simulations and comparison with measurements: I consider this part a candidate for removal from this paper. Global simulations and comparisons with distant sites are not that useful for verifying the fine details of the source with which the present paper is mainly concerned with. There are some aspects of the results that would need to be addressed in more detail, such as the underprediction of particulate iodine and the lack of correlation for caesium. Note that the authors speculate that the correlation for iodine is caused only by the co-factor of age and associated decay – see my comments above on decay correction! I was also wondering why the time series plots are almost all clipped so that they don’t show the arrival of the plume properly. Better work on a separate study than reporting preliminary results without proper explanation.

➔We believe that showing the global HYSPLIT calculation result is valuable to verify the new source term in particular when the plume flowed directory toward the Pacific Ocean estimated in this study. However, since the discussion should be limited for reducing the manuscript length, we deleted several figures and focused on only the first arrival time of the plume and the general time series (old Figs. 5, 6, and 22). Further evaluation is needed in future.

For your other questions: radioactive decay is considered in the WSPEEDI model as mentioned before. The time series plots in all figures related to global HYSPLIT calculations have included all available data from CTBTO from the beginning of FNPS1 accident.

12. Discussion of source terms (Section 4.1): I don't feel so much convinced with these arguments. The increased wet scavenging might be too high, and estimates of dose rates at nearby monitoring stations would depend on effective source heights, which are not so well known and which appear to have been set to some predetermined values based on expert judgement only (the paper is not explicit on this). The emissions are partly attributed, for example, to wet venting of unit 3, "wet venting" or "dry venting" is also referred to in other places. Even though the emergency staff tried to conduct such operations, often it is also reported that "success is not clear" or that operations failed. Furthermore, it is still not well known which release paths to the environment were created during the course of the accident. Thus, one should not be certain that venting operations reported represent properly the release paths. All in all, results therefore may be not as robust as they are appear in the paper.
➔ As suggested by the reviewer, our description in section 4.1 was too conclusive despite many uncertainties. Thus, we addressed the possible reasons which may cause the uncertainty of new source term in section 4.1 and Conclusion.

13. Supplement with release rates: when I open the xlsx file, I am told that a link to an external file exists (which of course is not present). In view of long-term archival, I don't think xlsx is a good format to include the data – a simple text file could be more useful (including both might be an attractive option). Also, give the units for all the values (it is not stated that releases are given as Bq/h). Do not mix text and data, or source geometry that cannot be read automatically, but provide a format allowing data to be easily ingested by programmes.
➔ We changed the format from .xlsx to .csv. The revised file includes the unit of Bq h^{-1}.

14. Measurement data: The paper refers to various sources of measurements, usually web resources. However, many are in Japanese, and usually in the form of PDF files. This kind of sources is not very suitable for use in further studies. I would like to encourage the authors to do their best to contribute to a collection of the relevant environmental data in an accessible, machine-readable format, and provide the most useful sources in their references.
➔ We realize the importance to make the database for monitoring data in English which can be widely used to public. However, at the current stage to respond to all reviewer comments, we have no time to do so and are also afraid that we take mistakes in alternation.

Other comments

1. I think Eq. 2 is trivial enough to be skipped (and writing whole phrases as subscripts is not very appropriate).
➔ The equation was removed.

2. Reference CTBTO (2011) is missing.
➔ The reference is now available in References section as "CTBTO (Comprehensive Nuclear-Test-Ban Treaty

Organization, Preparatory Commission): Fukushima-related Measurements by CTBTO, available at: http://www.ctbto.org/press-centre/highlights/2011/fukushima-related-measurements-by-the-ctbto (last access: 25 April 2014), 2011.”

3. Figures 12 and 15 should match better (same order of variables shown, domain)

➔The figures are unified to the same figure (new Fig. 11) in the revised version with the same domains and the same order of variables.

4. Figure 25: It would be better to separate this important overview into a Cs-137 and I-131 part.

➔ The figure was separated to (new) Figs. 22 and 23 for Cs-137 and I-131, respectively.

5. Figure 17: It is much too small. Showing the Cs-137 results is sufficient, the others are very similar.

➔ As suggested by the reviewer, we showed the figure for only Cs-137 concentration in Fig. 17, resulting in producing larger graphs.

6. I found Figure S3 quite interesting and was wondering why this Figure was moved to the supplement. Concerning the way it is referenced in the text, this type of figure is not suitable to show that for the majority of monitoring sites the model was within a factor of 2 of the observation. There is a number of sites with nice agreement, but also sites which don’t agree and where the new version did not improve much. It would also be of interest to know whether the sites are a subset, or include available ones.

➔ According to your comment, (old) Fig. S3 was moved to subsection 4.1.1 and reffered as (new) Fig. 12 to demonstrate the impact of the new source term from 15-16 March.

7. Figure 18: Not needed and unreadable. Just give a score such as FA5 as a small table.

➔To make the paper shortened, we deleted this figure in the revised version.

8. Figure 19: Why is the color scale totally different than in the observation plot (Fig. 12)? Also it is very difficult to find geographical features in both figures for comparison. Remove the province border lines and show a well-readable geographic grid, same in both figures. Restrict the area in this Figure to the one in Figs. 12 and 15.

➔ According to the suggestion, we unified the color scale for all figures of spatial distribution so that the reader can compare each other (new Figs. 11, 13, 14, 15, 18, S3, andS9).

9. Tables 6–8: This presentation is not sufficiently clear. All comparisons should be done with the same metrics and be shown in the same layout (a clear one!). My preference is FA2/FA5/FA10 (maybe not all of them are needed), correlation coefficient and bias. FAx is the best metric for these quasi-lognormal data. Bias should not be normalized by mean of obs and model (FB) – gives better score to overpredicting model. As all comparisons in one class have the same obs data, no need for normalisation. Or if you want, normalise with the mean observation only. Add units for all columns which contain dimensional data.

➔ To make the paper shortened, we deleted Tables 7-8 for WMO model calculations in the revised version. In addition, Table 6 has been fully revised with the statistics of CC, NMSE, FB, and FA2/FA5/FA10 to demonstrate the difference between four WSPEEDI simulation cases (combinations of the original/modified WSPEEDI and Terada/new source terms) in subsection 4.1.1.

10. Many of the figures (also some tables) are to small to be legible when printed, only on the screen with 200-400% magnification they can we well read (but some are blurred at high resolution, even worse). Please make sure in the final version that all figures included are readable in print format.
➔We adjusted the figure sizes so that the readers can see figures readily.

11. Several captions of figures with model–observation comparisons lack information about the observation data set used.
➔ We added the references to figures from Figs. 11-16.

12. I don't think having an appendix and a supplement is needed. I would suggest to move the description of the deposition schemes to the supplement.
➔ The description for deposition schemes is necessary for readers to know the changes from the previous version of WSPEEDI-II (particularly in subsection 4.1.1). Therefore, we remained the part in Appendix section with the short explanation of the below-cloud scavenging to answer other reviewers.

13. Language proofreading would be a plus, especially as native English speakers are on the list of authors. (Just a single observation: don't use the word 'trend' for general temporal variation.)
➔ One of our co-authors corrected the linguistic problem throughout the manuscript.