We would like to thank the reviewers for taking the time to read the manuscript and for offering suggestions for improvements.

Below, we provide a list of responses to the reviewer's points.

Referee #1 (P. Siebert)

MAJOR COMMENTS

1-The abstract claims that none of the existing approaches uses dose rate measurements, and in the introduction it is stated that Astrup et al. (2004) used synthetic gamma dose rate data to improve the dose rate field. This is not correct. Astrup et al. used real dose rate data, as do Drews et al. (2004) and Duranova et al. (1999). Furthermore, there are several more works which used synthetic dose rate data (see our web site http://www.univie.ac.at/theoret-met/prepare/).

We added the references. Astrup et al. and Duranova et al. validated their approach by using synthetic measurements whereas Drews et al. (2004) validated their method by using dose rate data from an experiment carried out in Mol in Belgium. During the Mol experiment, ^{41}Ar was released from a reactor stack and the resulting gamma radiation field was monitored by 4 NaI detectors. The observed dose rate came from only one specie contributing only to the plume component of the dose rate because ^{41}Ar is a noble gas and does not deposit. Therefore, the inverse problem Drews et al. had to solve by using dose rate measurements was very similar to the same problem solved by using air concentration measurements (i.e. only one specie, contribution of the plume component only). The case studied by Drews et al. is highly simplified compared to a real accident situation. For instance, for the Fukushima accident, the inverse problem to solve by using dose rate measurements is much more complex and is not any more similar to the same problem solved by using air concentration measurements because the release was composed of several species and ST isotopic composition is unknown. We tried to highlight this idea in the text.

2-The authors use the term "reverse method" or "reverse approach" to designate the simple source estimation method of Chino et al. (2011) and others, whereas more elaborate methods are qualified as "inverse method". I think it is not useful to introduce this kind of terminology. "Reverse" and "inverse" don't really have a different meaning, the difference just being that "inverse" is a well-established scientific term while "reverse" may not have a specific meaning. I really ask the authors to refrain from further popularising this not very fortunate expression. It would be better to speak about "simple methods for source estimation".

We agree with the referee and used, as suggested, the expression "simple methods for source estimation" instead of "reverse method".

The authors claim that, in contrast to others, their method does not make use of an a priori source term, both in the introduction and in Section 2.2. However, their assumption σb = 0 is an a priori, namely assuming that the source term is zero. Whereas other authors use the deviation from their nonzero a priori to stabilize the solution, Saunier et al. use the deviation from zero (i.e., the variance of the source vector). Of course, we can argue that

the point here is that the method works without having to construct a specific a priori source term. However, in principle, the other methods could as well use such a "trivial a priori" and giving the regularisation conditions sufficient weight, a solution will also be attainable. In my opinion, the point here is rather that the dose rate network is sufficiently dense and well-distributed to obtain a good solution without specific a priori. Also the statement "When no prior knowledge of the source term is used, it is still possible to solve the inverse problem but the number of measurements must be much larger than the number of unknown parameters" is not generally true – it depends on how strong the regularisation is. Regularisation has the same effect as decreasing the number of unknowns, because implicitly it reduces the resolution of the vector of unknowns.

The referee is right; we used the deviation from zero to stabilize the solution and our argument to make no use of an a priori source term is not strictly correct. We should have said that we do not need to use a specific a priori source term because the dose rate network is sufficiently dense and well distributed to reconstruct most of the release events. The text has been changed accordingly to the suggestions.

It is true that the other methods could as well use the a priori $\sigma_b = 0$. For instance, Winiarek et al. (2012) used σb = 0 to assess the Fukushima source term by using activity concentration. They showed that some important release events were missing in their solution because the Japanese activity concentration network was too sparse, the stations were distributed mainly south of the FD-NPP and the temporal resolution of the devices distributed away from Japan was too low. When the plume and the deposit induced by a release are not detected by the measurement network, the release rates of the retrieved source term are set to the release rates of the a priori.

When using a specific a priori, inverse modeling methods are used to improve the a priori release rates for the observed events. The a priori release rates remain unchanged for the unobserved events (when the initial source term is equal to σb). When using a trivial a priori (e.g. σb = 0), it is supposed that the number of measurements is sufficient to reconstruct most of the release events.

The regularization term weights the relative impact of the first guess and the measurements. A strong regularization minimizes the measurements impact and the inverted source term remains very similar to the first guess. In our approach, the positive release rates values are assessed on the basis of measurements only and the regularization term is used to avoid inconsistent values when the method fails to reconstruct observed events, for example, when the weather forecast is wrong. The regularization term is not chosen to compensate for the lack of measurements.

Similarly, the statement following Eq. 8, that the number of constraints (for the nuclide ratios) has to be at least equal to the number of radionuclides, is not sufficiently founded.

The main purpose of these constraints is to force the method to release simultaneously the different radionuclides. The compliance with the criteria implies to use a number of constraints (for the nuclide ratios) equal to the number of radionuclides minus one. In practice, the inverse problem solved without all the constraints would lead to unrealistic solutions: for example peaks of cesium release may appear without simultaneous iodine release which is physically unrealistic.

The flexibility of the constraints i.e. the width of the range of permitted values determines the size of the solution space.

4-An important part of the method is the division into a step 2 which identifies potential release periods and a step 3 which then quantifies the releases during those periods. It is not clear why this step 2 must be introduced, or whether it is really beneficial. One might expect that the inversion would anyway only attribute nonzero source strength to the episodes that are identified in step 2. There is no directly visible reason why Eq. 5 should lead to different periods than applying Eq. 7 directly without previous restriction of the source time. The only obvious advantage is the reduction of the size of the problem, but this is not given as a motivation for the division into two steps.

The advantage of dividing the method into 3 steps is, indeed, the reduction of the size of the problem. Thanks to step 2, the number of unknown parameters to solve is divided by 3 in the Fukushima application. We changed the text in order to clarify this point.

We initially tried to solve the inverse problem directly without going through step 2. The obtained results showed that the algorithm increased the number of release events and forecasted erroneous release events to compensate for previous model errors. For example, when the model underestimated the measured dose rate, the algorithm made an erroneous release in order to increase the simulated dose rate and minimize the differences between model and observations. The lack of detection of a plume in the measurements was not sufficient to prevent the algorithm to create such a release.
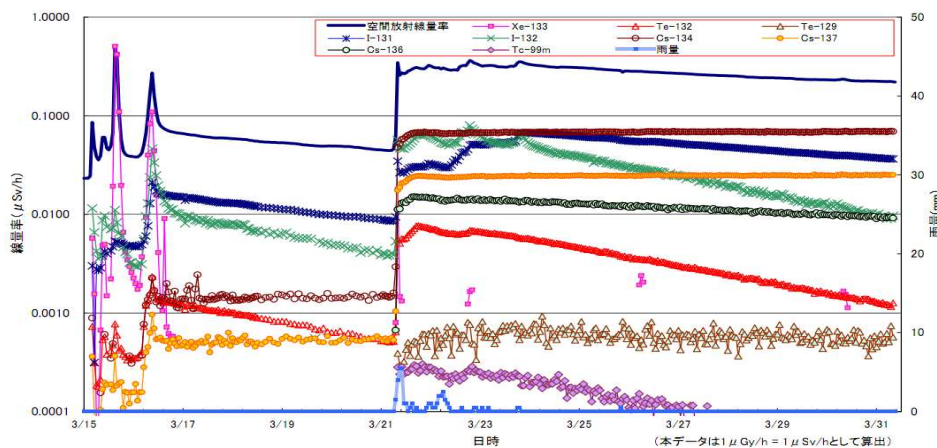
From the point of view of the CPU time, step 2 does not significantly increase the computation time. It runs fast since one considers that the release include only one radionuclide.

5-What is the meaning of the statement "All of the noble gases emitted during the accident are grouped and are estimated as 133Xe emissions." (Section 4.2.2, abstract and conclusions repeat this argument, should in case also be adapted.) If it is true, as said before, that Xe-133 is the only noble gas among the nuclides that contributes significantly to the dose rate, then one should not consider this as grouping and as attributing other nuclides to Xe-133. I am also wondering which fraction of the total dose rate is caused by the selected list of 8 nuclides. Our own calculations (not yet published) showed that for achieving 98% of the total dose committed to people (external and inhalation), we would need about 20 nuclides, including Xe-135 and two Kr isotopes. One would assume that external dose only would show an even higher fraction to be attributed to noble gases and thus probably to isotopes other than Xe-133. If, however, this is not true, than the fact that other noble gas nuclides were excluded has to be introduced before the nuclide list, and it should be justified. Personally, I don't see why one would not want to include them, as their reactor inventory is known and they would thus not increase the effective number of unknowns.

The isotopic composition of the releases is not supposed a priori and is not estimated from the core inventory. The quantities are assessed in such a way that the dose rate of the simulated deposit and, in particular, its slope, fit with the observed ones. The separate half-lives of the radionuclides allow differentiating the contribution of each radionuclide through the correlation between the radioactive decrease and the slope of the dose rate signal due to the deposition. The amount of noble gases is assessed using the fact that they do not deposit so they contribute to the plume dose rate and not to the deposit dose rate. From this point of view, two noble gases will

behave the same. They cannot be distinguished from each other, even if they have very different half-lives. In our application, we considered only the noble gas Xe-133 which means that we used the Xe-133 dose coefficient to compute the dose rate. It is likely that the contributions of the other noble gases emitted during the Fukushima accident are included in the estimated quantities of Xe-133. That should explain the large amount of Xe-133 assessed in the retrieved source term.

The 8 selected radionuclides have been identified mainly from the measurements in Chiba (see Fig. A). The analysis of the Chiba data shows that the 8 radionuclides explain at least 80 % of the dose rate signal. For some events like the plume detection on March 15, the 8 radionuclides explain about 90% of the signal and a large part is due to the contribution of Xe-133. Other radionuclides are present in the release. For example, the analysis of the activity concentration measurements from Wako and Tokai shows the Te-129m activity concentration can be about 30% of the I-131 one. We did not take Te-129m into account because its dose coefficient is lower than the I-131 one. The contribution of the Te-129m to the dose rate remains secondary but it could have been include in the inversion process. Adding radionuclides in the inversion process leads to an increase of the size of the problem and one has to find a balance between the number of unknown parameters we want to realistically estimate and the size of the problem.
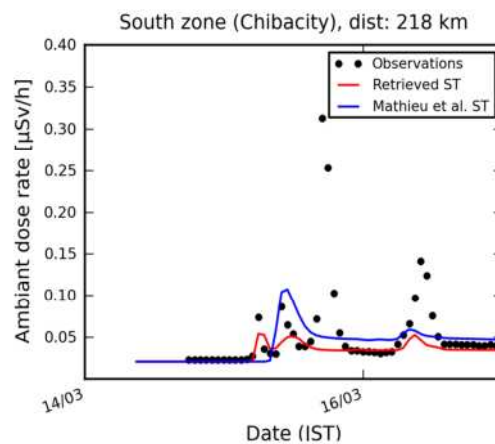

Fig. A: Dose rate measurements in Chiba.

6-The authors state on several occasions that they think that they overestimate the Xe release. I think that the arguments that lead to this conclusion are not sufficiently clear, and also I doubt that there is really an overestimation. It seems that the main argument why the authors believe in an overestimation is provided in Section 4.2.2, p. 15589/90. I don't understand why Te-132/I-132 should behave like a noble gas – after all, the decay of deposits of Te-132 (half-life ca. 3 d) should be clearly visible and the decay of I-132 (ca. 2 h) at least roughly visible from hourly dose rate data. Furthermore, effects of washout (which does not affect the noble gases) would also provide a means of distinguishing. I would rather think that the attribution of all noble gas isotopes as Xe-133 (see discussion above) would lead to an overestimation. Finally, the result of 12 EBq roughly agrees with the inventory of the damaged cores, and certainly isn't too high when the in-growth from I-133 is considered. The conversion between aerosol-borne and gaseous nuclides is, by the way, a potentially relevant influence factor not discussed in the paper.

We believe that the Xe releases are overestimated for several reasons:

- The emissions that went directly toward the Pacific ocean have not been reconstructed by our method because no dose rate measurements were available. Some of them, especially the releases from the Unit 3 probably included a significant amount of Xe-133. However, our retrieved source term contain almost the entire core inventory of Xe-133.

- As discussed before, the method cannot distinguish between noble gases. Therefore, all of the noble gases emitted during the accident are grouped and estimated as Xe-133.

- Depending on the situation, the radionuclides that have half-lives shorter than 2-3 days cannot always be fully identified by the method. In that case, a portion of their contribution to the dose rate may be interpreted as a noble gas contribution. For example, several plumes have been detected between March 15 and March 17 (see, as an example, Fig. B). Between the detection of two plumes, the dose rate signal due only to the deposition does not last long enough to allow the inversion process to interpret the speed of the radioactive decay of the deposit. The dose rate signal due to the deposit influences the inversion process if it lasts, at least, 0.5-1 day without new plume detection. The relatively short life radionuclides that have been deposited on March 15 had partially disappeared on March 17. For example, I-133 (half-life = 20.8h) had partially disappeared. The inversion process cannot distinguish the missing portion of I-133 from a noble gas which would have been emitted on March 15. Similarly, a portion of the Te-132 (half-life 3.3 days) emitted on March 15 may be interpreted as being a noble gas emission. It is the reason why we believe that the retrieved emissions under-estimate the amount of the pair Te-132/I-132 and over-estimate the Xe-133 quantities.

- Since the emissions of I-133 are not estimated, the Xe-133 in-growth from I-133 is assimilated by the method as an emission of Xe-133.

Fig. B: Dose rate observed and simulated in Chiba.

7-Does the inversion method applied include the calculation of a posteriori uncertainties of the retrieved source term? If so, what is their magnitude and how variable is it?

We did no specific calculation of a posteriori uncertainties as proposed by Winiarek et al, 2012 or Stohl et al., 2012. However, Table 4 gives a general overview of the uncertainty of the retrieved ST for each release events. The sources at the origin of the uncertainties are also discussed.

Model-to-data comparisons done with unused measurements (air activity concentration and deposition measurements) also allow assessing the realism of the retrieved ST.

8-At the end of the Conclusions, the authors discuss planned further improvements, among them using a nested approach for better resolving the steep gradients encountered. While this is quite appropriate, it would probably be good to touch upon this issue in the introduction section, as most other inversions rely on Lagrangian particle models which don't have this resolution problem. Furthermore, one could ask whether the smoothing effect of the Eulerian model (note that we are looking at point-to-point source-receptor relationships!) also has a regularizing effect on the inversion, by broadening the sensitivity function in time.

The current inversion methods solve the inverse problem by minimizing a cost function based on the absolute value of the differences between model and measurements (we assume the covariance errors of matrixes B and R are homoscedastic) Therefore, in the cost function estimation, the differences on the high values have more weight than the ones on the low values. Distant measurements (i.e.low values) have therefore little influence on the source term estimation compared to measurements close to the source. This behavior is due to the inversion method and not to the type of model that we used, whether it is a Lagrangian or an Eulerian model. If one wants to reconstruct the source term by using together the measurements located in the US and the measurements located few kilometers from the source and ensure that the distant measurements have an influence on the source term assessment, we need to evolve the method in order to take into account the relative impact of the model-to-measurement differences. At the local scale, the problem is the same: in the area where there are steep gradients, the range of values of the measurements is very extensive and the method is designed mainly to obtain a good agreement in the high values areas and, if necessary, to the detriment of the less contaminated areas.

In our study, the problem is less obvious because the measurements that we used are mainly of the same order of magnitude. Moreover, the lower values (i.e. mostly due to the deposition) are more numerous than the higher values (due to plumes detection). Even if the individual weight on the cost function of the model-to-measurement differences due to the low values is weak, the sum of all their contributions impacts significantly the source assessment.

9-It would be very useful if a supplement with the retrieved source term would be provided, similar to Stohl et al. (2012).

The retrieved source term is now available on line

MINOR COMMENTS

1-Language: While in general the paper is well written, there are numerous instances where grammar, spelling, or vocabulary is not proper English (probably influenced by the authors' mother tongue). It would be very desirable to have a language check done on the final manuscript. A few examples (not exhaustive!):

- *Page 15568, l. 24/25: The result is that the model-measurement agreement for all of the monitoring locations is correct for 80 % of simulated dose rates that are within a factor of 2 of the observed values. – This is not a proper sentence. A possible wording could be: It was found that for 80 % of the monitoring sites, simulated and observed dose rates agreed within a factor of 2.*
- *Page 15569, l. 12: For example, source term related . . . – "the" is missing.*
- *Page 15573, l. 21: vector source term – should be "source term vector" or "vector of source term elements".*
- *ibid., l. 28: humid deposition – should be "wet deposition"*
- *Page 15577, l. 19: real accidental situation – should be "real accident situation".*
- *Page 15580, l. 8: noble gases were rejected – should be "noble gases were emitted" (or "released").*

The text has been corrected.

2-Page 15572, l. 12: The ambient dose rate is not the linear combination of the contributions of all gamma-emitting radionuclides, but their sum. (OK, a sum is also a linear combination, but a trivial one and it is thus misleading to use the term linear combination.)

The text has been changed.

3-Figure 1 and respective paragraph in Section 2.1: The figure caption is insufficient. The numbers 1, 2, 3 must be explained at least briefly in the caption or it has to be said that they are explained in the text. Reference belongs to caption and not to text body. The abbreviation ex. used in the text is not clear. It should be said that this is an idealised situation, for example assuming no wet deposition and that the plume passes over the receptor only once.

The figure caption and the text have been improved. We replaced the abbreviation "ex" by the abbreviation "e.g" in the text.

4-Page 15574, l. 5: Explain what is E. Explicitly state that off-diagonal elements are not considered (and why).

E is the mathematical expectation.
In an accidental context, it is relevant to set off-diagonal elements of the B matrix to zero. Indeed, the source term is often composed of uncorrelated events (minutes to hours). Moreover, the number of available observations is substantial. Thus, the quality of the modelized errors related to the first guess has a limited impact on the solution. To limit the CPU time, we therefore chose a simple parameterization for the B matrix.

The parameterization for the R matrix is chosen as it is usually done in the current inverse modeling approach (Winiarek et al., 2012; Stohl et al., 2012). Off-diagonal elements are set to zero.

The chapter 2.2 describes the general method where no assumption is made on the shape of the matrix. Hypotheses are made chapters 3.3 and 3.4 and the parameterizations used for R and B are given.

5-ibid., l. 18: Quantify which fraction of the dose rate is explained by 10 radionuclides, and also under which conditions.

The referee is right to bring up that point. The number of radionuclides to consider may change depending of the conditions. We changed the sentence.

6-ibid., l. 20: (The relevant radionuclides) can be identified by . . . or by using the core inventory of the damaged facility. – Core inventory plus release fractions for each nuclide group! Same issue after Eq. 4.

We do not use the release fractions for each nuclide group. So far, this information is not known for the Fukushima releases and we do not need to suppose it to resolve our problem.

7-Eq. 6: An erroneous apostrophe appears after the last line.

The correction has been done.

8-Eq. 7: I think it is necessary to include error covariances also in the cost function contribution due to the nuclide ratio, at least in the form of a single weighting factor. Otherwise, you are assigning implicitly a certain weight to this term in comparison to the other terms (probably, due to the shape of the function r, results aren't very sensitive to this, but that is not immediately visible, and also formally it is unsatisfactory).

The barrier-protection function has been added in order to force the optimization process to comply with the constraints. When a release does not satisfy a constraint, it should be rejected and not considered as an option by the optimization process. The constraints must cover a wide range of values containing all the measurements to ensure flexibility of the inversion process.

Ideally, a boolean function should have been used but a discontinuous function is not compatible with the L-BFGS-B algorithm. Instead we used an exponential function to properly penalize the cost function when the constraints are no longer respected and to ensure the rejection of the release. Using the exponential function, the isotopic ratio of a retrieved source term cannot strongly differ from the range of permitted values.

Even with an unrealistic source term (for example a zero release on the Fukushima accident) the barrier-protection function remains higher than the first term of the cost function (e.g. the sum of the differences between model and observations). This result shows that the barrier-function fulfills its function without having to include an error covariance. In addition, the method used during an emergency situation requires to limit the number of weighting factors, always tricky to elaborate. We chose to keep unchanged the cost function.

9-Page 15578, l. 25: Is the output instantaneous or averaged over 1 h?

The model outputs give instantaneous concentrations.

10-ibid.: Is a model top of 3400 m sufficient? Did you verify that?

The simulations are carried out at the Japan scale during a period where no deep convective events were observed. We did one simulation with a model top of 10000 m to check whether the 3400 m top height was sufficient. The outputs showed no significant differences. The choice of a model top of 3400 m instead of a higher top was used to narrow the CPU time of the H matrix computation.

11-Page 15579, l. 4: Give references and/or details for the ldX, C3X and ConsX models. How much time does it take to carry out these 381 simulations, and on which computer platform?

A reference has been added.
It took 6 hours to carry out the simulations by parallelizing the 381 simulations on the 24 processors of an Intel Xeon X5690 @ 3.47GHz. For operational use, the impulses (i.e. the 381 simulations) can be computed in advance.

12-Page 15580, l. 6: I would not say that noble gases are "highly volatile", they are just gaseous (would condense only close to absolute zero). In the reactor or the environment, they never occur in solid or liquid form, so it is not so appropriate to discuss them in terms of volatility.

The text has been changed.

13-ibid., l. 9: I would think that I-132 and Te-132 are not exactly in secular equilibrium, at least not in the first hours after the shut-down. Maybe it should be made more clear how well this assumption is fulfilled at the time for which the first release is anticipated.

The referee is right, I-132 and Te-132 are not exactly in secular equilibrium and the text has been changed.. Following the shutdown, after a time period of about 1 day (~ ten times the I-132 half-life), I-132 and Te-132 are found in equilibrium. Indeed, one day after the shutdown, I-132 present initially in the core inventory had disappeared. Therefore all the I-132 that was observed in the environment resulted from the Te-132 decreasing. Our hypothesis to take a constant I-132/Te-132 ratio then is relevant.

14-ibid., l. 15: It is not correct to say that Cs-137 has no impact if it is decaying too slowly to observe this process. Nevertheless, it is a major contributor to the dose.

The misunderstanding of the referee is probably due to a lack of clarity. We tried to improve the text. Cs-137 is indeed a major contributor to the dose and is considered in the inversion process. We wanted to explain that the half-lives of Cs-134 (2 years) and Cs-137 (30 years) do not allow distinguishing one from the other in our

simulations. The simulations last 15 days which is not long enough to see the deposit concentration of the two radionuclides significantly decrease. Therefore, the algorithm cannot distinguish the Cs-137 contribution from the Cs-134 one.

15-ibid., l. 18: Who showed that the ratio of Cs-134/Cs-137 is constant over Japan (reference)? Constant within which bounds? Shouldn't there be an upper and lower threshold for the ratio in Eq. 10?

We analyzed the air concentration measurements in Tokyo, Tokai, Chiba, Tsukuba, Wako. The measurements show that the ratio of Cs-134/Cs-137 remains roughly constant. For example in Tokyo, more than 95 % of the observations show a Cs-137/Cs-134 ratio varying between 0.9 and 1. In addition, Cs-137 and Cs-134 have the same behavior and it is expected that their ratio remains constant.

16-Eq. 11: It would be of interest to include a table which translates these nuclide ratios used as bounds into ratios of the release fractions of these nuclides, under the assumption of a known and homogeneous inventory.

For the Fukushima accident, the total amount of each radionuclide released in the atmosphere is still unknown and as far as we know, the core inventory has not been disclosed by TEPCO. Therefore we cannot respond to the referee request.

17-Page 15581, l. 19: I would be more careful with respect to stating that _1 can be rigorously determined. That depends very much on the condition of the matrix, and even if a precise condition is formulated, this does not necessarily mean much in practice. Only if errors for all the terms are well known, the solution is truly rigorous.

The correction has been done.

18-Page 15582, l. 1: Please explain briefly how the L-BFGS-B algorithm works, whether it is publicly available in coded form, and whether additional constraints such as positive definiteness are applied.

L-BFGS-B is a limited-memory quasi-Newton algorithm for bound-constrained optimization, i.e. for problems where the only constraints are of the form l<= sigma <= u where the vectors l and u represent lower and upper bounds on the variables. L-BFGS-B can be also efficient for solving unconstrained problems. The user must supply the gradient G1 of the cost function J1. In our inversion method, all the variables have lower bounds (we impose the positivity of the source vector sigma) but there is no upper bound. Open source implementation of L-BFGS in Fortran exists.

19-Section 3.3 in general: Is there a threshold value for _1 being applied? How continuous/intermittent is the resulting possible release time?

The second step of the method identifies 140 potential release events when no threshold is applied and only 120 when the following threshold $\sigma_1(t) > 10^8$ is used. Yet both resulting source terms are very similar. The third step of the method systematically sets to zero the release rates of the 20 additional potential release events

identified when using no threshold. A threshold then can be applied to reduce the CPU time of step 3. In our application, we used $\sigma_1(t) > 10^8$ for that purpose.

20-Page 15583, l. 10: It would interesting to know which percentage of the emissions calculated by Stohl et al. (2012) occurred during periods which did not pass over Japanese dose rate monitoring sites and thus cannot be reconstructed in the framework of the present manuscript. Doing this would allow a more meaningful comparison between these two emission estimates.

Events 3 (Unit 3 ventings) and 4 (Unit 3 ventings and explosion) of Table 4 did not pass over Japanese dose rate monitoring sites and thus are reconstructed to a very limited extend by our method.
To respond to the referee request, we need to assess which percentage of the emissions calculated by all the authors and not only by Stohl et al., 2012. The problem is that the various ST do not contain the same list of radionuclides and the percentage of the emissions is highly dependent on the radionuclide considered. For example, for Stohl et al., 2012, events 3 and 4 are 44 % of Xe-133 total release and only 6 % of Cs-137 total release which cannot be compared of the iodine assessment of Winiarek et al..
We did not change the text.

21-Section 4.2: Maybe this section would better be called "Comparison with observations", as an agreement of simulated and observed dosed rates and/or activity concentration is a necessary but not a sufficient condition for the agreement of the calculated with the true source term.

The correction has been done.

22-Page 15585, Eq. 15: Note that this formulation favours overprediction as compared to underprediction.
23-Table 4: I am wondering how it comes that the biases are always positive. It is quite surprising and needs to be discussed.

There was a mistake in the Eq 15. The absolute value of the difference between dm_i and do_i was missing. The text has been changed.

24-It seems that no distinction has been made between gaseous and aerosol-bound iodine in the dispersion calculations. This should be mentioned explicitly. Could this be responsible (at least partly) for the lower performance with respect to iodine as reported in Table 4?

Indeed, no distinction between gaseous and aerosol-bound iodine have been made in the simulations. To not take account the complex behavior of iodine into the atmosphere could be responsible for the lower performance with respect to iodine. A comment has been added in the text.

25-Page 15589, l. 4: The simulations accurately reproduce the signal's temporal evolution – The agreement is good, but "accurately" is exaggerated.

"Accurately" has been replaced by "properly".

26-Page 1550, l. 6: The comparison is more difficult in the case of the second episode – it is not more difficult, rather it results in more substantial deviations.

The text has been changed into "No comparison can be done for the second episode".

27-Various places: fac2, fac5, fac10 used like words. Please don't use code variable names like words, better write that out, or introduce a proper symbols such as, for example, F2, which should then be introduced at the place where it first appears.

The text has been changed.

28-Page 15591, l. 10: The surface activity increased – did you mean that the deposition rate increased? That is not the same. Also below, deposition activities – does that refer to deposition rates?

"surface activity" has been changed into "deposition amount".

29-ibid., l. 25: These measurements, some of which were taken after the accident,– I thought that all of them were taken after the accident?

The first airborne campaign was carried out on March 17-19.

30-Page 15592, l. 1: I have the impression that while the pattern of the heavy contamination towards NW is fairly reproduced, but its intensity still is too low (how much?)

The north part of the most contaminated pattern is indeed slightly under-estimated by approximately a factor of 2. The under-estimation is higher near the source which is mainly due to the use of an Eulerian model (dilution near the source). The text has been changed.

31-ibid., l. 13. How can it be seen that differences are due to inaccurate meteorological data?

The simulated deposition amounts located 100-150 km west of the FD-NPP are under-estimated as the simulated dose rate in the west area. By analyzing the simulations, a time lag between the wind fields and precipitations is identified. To check if there is no other reason that may explain the model-to-data discrepancies, the inversion process was carried out to reconstruct this single event only. The process failed and the event proved to be impossible to reconstruct by using the ECMWF meteorological fields.

32-Table 3: Does Unit 2 – Unit 3 mean unit 2 and unit 3, or unit 2 or unit 3 (uncertain from where)?

"Unit 2 – Unit 3" is for "Unit 2 or Unit 3"

33-Figures 1 and 2: It is not mentioned whether time is UTC or JST.

We used JST. It has been added where it was missing.

34-Fig. 2: Caption is too short. For example, say where the measurement is taken. Use smaller symbols for the measurements. Consider using a log scale.

We completed the caption and used smaller symbols for the measurements. We prefer to keep the linear scale to be consistent with the Fig. 5 and Fig. 6 where dose rate measurements are plotted.

35-Fig. 3. Include a demarcation of the different zones in the figure, and make the NPP site more visible.

We added three white dashed lines to show the demarcation of the 4 different zones and we increased the size of the purple triangle designating the NPP location.

36-Figure 4: In order to increase the readability, don't use dashed lines (maybe the vertical lines could be plotted in gray). Ticks to the outside on the time axis, and adding minor ticks for subdivision would also be useful. Indicate (somewhere – not necessarily in the Fig. 4 caption) the time shift between UTC and JST.

The Fig. 4 has been changed. The ticks are now located outside on the time axis. We added minor ticks for subdivision on the time axis. Dashed lines have been replaced by full lines to represent source terms. Vertical lines are plotted in gray.

37-Figures 6–8: The dots are not very well readable, and it is not clear whether the width of the dots corresponds to the time interval to which it refers, or otherwise whether dots are centred or at the end of the respective time intervals. Probably horizontal lines (dashes of appropriate length) could be more suitable.

The readability was not improved by replacing the dots by horizontal lines. Instead, the dots are still used and the caption was completed to make clear the meaning of the dots.

38-Figures 2, 4–8: All these figures use linear scales of the ordinate. There are of course pros and cons of linear and logarithmic scales, but I am wondering why all figures are with linear scales. I think at least Fig. 7 would gain in readability if a log scale were used.

The linear scale is usually preferred because it allows one to better distinguish discrepancies between model and observations. The logarithmic scale tends to grant a more significant weight to weak values. Nevertheless, for reasons of readability, a logarithmic scale is used for Fig. 7.

39-Figure 9: The km-annotation on the radii and the legend to the colour bar are too small and hard to decipher. It is not clear to me whether all subfigures use the same scale, and it seems that they don't show the same area. Also, the observation map shows the outline of the provinces while this is missing in the other two

maps. All of the maps lack a geographical grid. It would be best to add that and to re-plot the observation map with the same software (as far as I know, data are public). A reference for the observation map is missing.

The figures have been changed (simulation maps on the same spatial domain than the observations map and same radii for the drawing circles). However, we do not have the observations of deposition, so we cannot re-plot the observation map.

40-It would be useful to add in a prominent location (e.g. abstract and introduction) the scale on which the method works, pointing out that it is not targeted at on-site dose rate measurements. Please introduce the state of the art in a more complete fashion.

The method works on all scales as long as a suitable atmospheric dispersion model is used. Theoretically, the method can use on-site dose rate measurements but external irradiation needs to be taken into account in the dose rate interpretation.

Referee #2

1-Page 15574, Line 5-6 and Page 15581, Line 12: Definitions of parameters in equations should be provided clearly. For example, E, m1, k1,and l.

E is the mathematical expectation.
The parameters m_1 and k_1 are the elements of the diagonal matrixes B and R. They are constant and are defined by the user. They can be also automatically assessed using hyper-parameters techniques (Winiraek et al. 2012 and Davoine and Bocquet 2007).
$\mathbf{I}$ denotes the identity matrix. We replaced $\mathbf{I}$ by $\mathbf{I}_d$.
There was a mistake in Line 6. The B matrix is: $\mathbf{B} = E\left[\left(\boldsymbol{\sigma} - \boldsymbol{\sigma}_b \right)\left(\boldsymbol{\sigma} - \boldsymbol{\sigma}_b \right)^T \right]$.
The text has been changed.

2-Page 15587, Line 15-17 and 25-27: Line 15-17 "From 15 March onwards, the simulations significantly underestimate the dose rates of the North zone monitoring stations close to the Pacific coast and overestimate those of the stations further inland" and Line 25-27, "The fact that the results for the stations located on both sides of the mountain area (see Fig. 6) are incompatible is due to the distance and time lag affecting the precipitation and wind data" – It is difficult to understand that which graph in Fig. 6 the above descriptions explain.

The Yamamoto station is representative of the monitoring stations located in the north zone, near the Pacific Ocean. Fig. 6a shows that the simulations underestimate the dose rate from March 15 onwards. Results for the Yamagata station are presented as an example for the stations further inland. Fig. 6b shows the overestimation of the observed dose rate signal.

The Shirakawa and Minamiaizu stations are located on both sides of the mountain area. Fig.6c and Fig.6d show that the simulations overestimate the dose rate signal in Minamiaizu and underestimate the observations in Shirakawa from March 15 onwards. By analysing the simulations, a time lag affecting the precipitation and wind data has been identified. The inaccurate meteorological fields are probably due to the too low spatial resolution of the meteorological data (12.5 km) enable to fully resolve the impact of the Japan's complex orography.
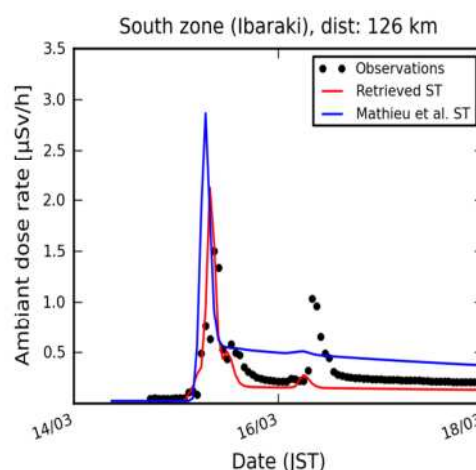The text has been modified.

3-Page 15589, Line 16-19: "On average, the 136Cs activity concentrations are slightly overestimated by a factor of 2 to 3, and the simulation results are most consistent with the measurements in the case of 137Cs. The 131I activity concentrations are overestimated but remain within a factor of 2 of the observations." – This description dose not correspond to Fig. 7. If you explain overall feature, other comparison results such as scatter plot will help readers to understand it. The same can be said for comparison of surface deposition in Section 4.2.3.

The text was not clear and has been improved. Observations in Tokai show an increase of the activity concentration beginning on March 15 at 03:00 JST and a maximum is observed between March 15 at 06:00 and March 15 at 09:00.
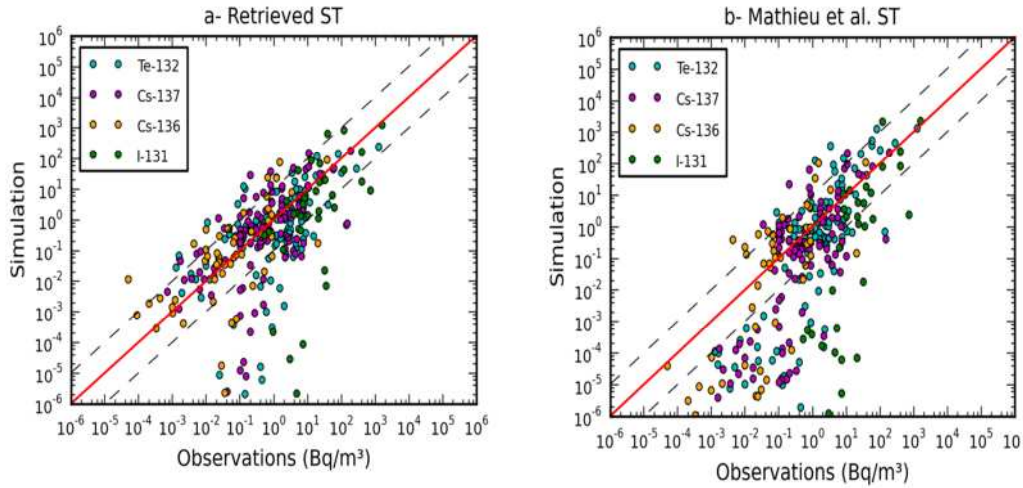Model-to-data comparisons for activity concentration as well as for dose rate (e.g. dose rate signal at the Ibaraki station see Fig. C) show that the simulated arrival time of the peak is one or two hours in advance. Model and observations are in good agreement at the time of the peak but the simulation overestimates the dose rate and the air activity concentration for the rest of the event.
For reasons of readability, we re-plot the fig 7 using a logarithmic scale. Scatter plots are plotted on Fig. D, and Fig. E below but not incorporated into the text.
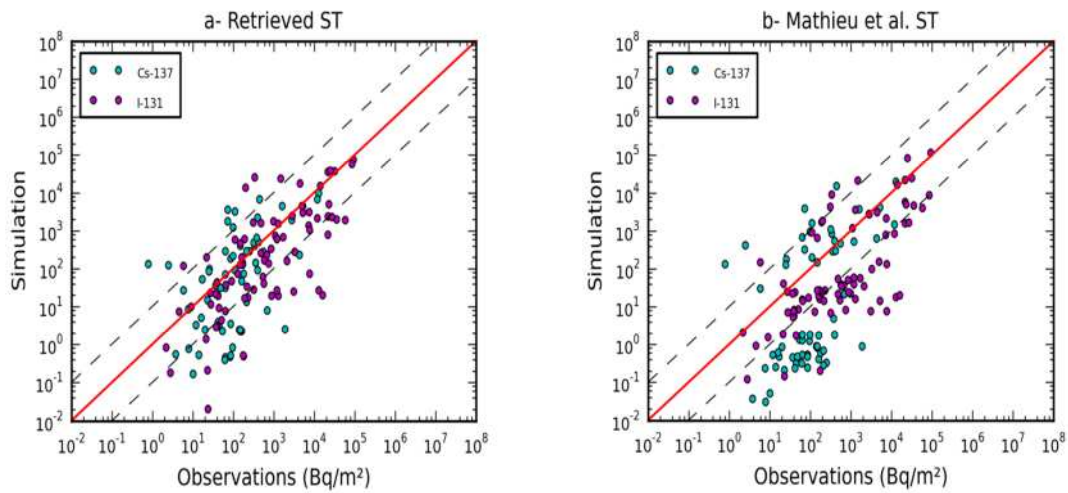

Fig. C: Ambient dose rate observed at Ibaraki station, located near Tokai station.

Fig. D: Scatter plot of air concentration values. Red line: perfect agreement. Dashed lines: factor 10

Fig. E: Scatter plot of daily deposition values. Red line: perfect agreement. Dashed lines: factor 10.

4-Table 2: Total emissions estimated by Terada et al. (2012) may be wrong. For example, the release amount until 31 March of 131I and 137Cs are 1.21017 Bq and 8.61015 Bq, respectively in Hirao et al. (2013).

The referee is right the total emissions are for the Chino et al. estimations. Total emissions estimated by Terada et al. until May 1 are 8.831 x 10^{15} Bq for ^{137}Cs and 1.239 x 10^{17} Bq for ^{131}I. Table 2 has been changed.

5-Table 3: Explanation of "Quality index" should be provided in the caption as well as in the text for better readability

We added an explanation of "quality index" in the caption of Table 3.