On behalf of all authors, we thank you for reviewing and providing feedback on our manuscript. Below you can found our responses to the specific comments.

1. *Line 21. "Specifically, the release of radionuclides – or more precisely: its accompanying ionizing radiation – can potentially pose …"*

   I think this sentence reads more clearly without the hyphenated expression ("- or more precisely … ionizing radiation –") because the same sentence later states that the danger is from exposure to radiation.

   We have removed the hyphenated expression, as we are in agreement with this comment.

2. *Line 119. "The retro-plume dispersion was calculated from the start of each measurement".*

   What does this mean? I understand that particles are released at a constant rate over the whole measurement window. If each particle starts being advected as soon as it is released, should this be "from the end of each measurement"?

   Thank you for pointing this out. This should indeed read "from the end of each measurement".

3. *§2.4. Comments / questions on the adjoint SRS calculations:*
   a. *Fig. 2 is nice, providing an intuitive explanation of the nature of wet / dry deposition observations, and their information content for the purposes of adjoint modelling. It complements the description of the method given in (Eckhardt et al., 2017) (if I am correct that it is the same method?). Since the Eckhardt paper contains many important technical details (e.g. height distribution and masses of model particles), please can you re-reference it in §2.4.*

      Fig. 2 is indeed based on the methods from Seibert and Frank (2004) and Eckhardt et al. (2017). We have now referenced these publications in Sect. 2.4 and the figure caption.
   b. *For completeness, it would be nice for Fig. 2 to include the (adjoint) source functions for the SRS fields, $h(x, y, z, t)$, corresponding to each measurement type (these are proportional to $\hbar$ defined by: $y = \int (h \cdot c) d^3x\, dt$, where c is the activity air concentration field, as in §3 of (Yee et al., 2008)). This essentially rewrites the integrals in a common inner-product framework. Then, I think Fig. 2 would be of even greater use for understanding the SRS method for deposition measurements; in particular, how it relates to the method for air concentration measurements.*

      Thank you for this suggestion. We have reformulated Sect. 2.4 and Fig. 2 to include the (adjoint) source functions.

4. *§2.5. Comments / questions on the performance metrics:*
   a. *How is the 'excluded' part of the domain defined in the FDE metric? Is there a threshold for the cost function / Bayesian posterior value? Does the numerical solver / sampler simply fail in these cases?*

The FDE is indeed based on threshold values. A value for 2 is used for the cost function method, and a value of 0 for the Bayesian algorithm. We have now mentioned this in the text.

b. *The domain enclosed by the grey rectangle in Fig 1 should be described (e.g. using lat-lon coordinates. Perhaps in the caption to Fig 1?). Then, if others repeat this work but with different computational domains, they will be able to calculate an FDE score which is comparable with the present work.*

The grey rectangle has lower-left corner [40° E, 45° N] and upper-right corner [80° E, 70° N]. We have added this description to both the caption of Fig. 1 and to Sect. 2.3.

c. *"[the CDS] can be defined relative to the full domain, or relative to the subdomain defined by the coverage of the location probability".*
Would these different definitions produce different values for the CDS metric? My understanding is that they would not. If this is the case, can it be made clear that the different in definition represents only a different way of calculating the same metric.

The different definitions do produce different values for the CDS metric. The CDS for the full domain is related to the subdomain variant by $\mathrm{CDS}_{\text{full domain}} = \mathrm{FDE} + (1 - \mathrm{FDE})\mathrm{CDS}_{\text{subdomain}}$.

5. *§3.1. Comments / questions on the experiment with meteorological models of different resolutions:*

a. *Line 250. "During the analysis, we noticed that these measurements were highly sensitive to the spatiotemporal resolution of the meteorological data".*

This sentence makes it seem as though the resolution of the meteorological data was of passing interest, rather than being a considered factor in the setup of the dispersion model runs. As it is currently presented, I think that the discussion of different meteorological models would fit better in an appendix. To stay in the main body of the paper, I think it requires rephrasing to present the two meteorological models as candidates which are then chosen between based on the comparison with measurements.

We have changed the wording in both Sect. 2.3 and 3.1 to clarify that we evaluate the meteorological data as a part of the model setup evaluation.

b. *Statistical errors in data analysis can cause meteorological models which are structurally identical (same physics / resolutions) to produce difference analysis data. Please can you discuss / suggest why this is not the primary cause of the differences observed in Fig. 6, and why the difference is produced by structural model differences? Following the suggestion of referee #2, looking at the position of the measurements relative to the plume could help; if the measurements were on the plume edge, they would be sensitive to small meteorological errors (errors in the analysis and short forecasts), so the intrinsic difference between the models might not be significant. Also, reporting the duration of* *the measurements would help to assess how sensitive they are to short-term changes in the plume.*

We observe a significant difference in the plume structure itself, where the bifurcated part of the plume that hits Scandinavia is much wider and contains higher concentrations in the high resolution versus the low resolution met data. We have added this information, and an accompanying figure of the resultant deposition fields to Sect. 3.1.

c. *Could another researcher get the same met data from MARS if they wanted to repeat this study? What information would they need to do this? (I am not familiar with the MARS archive or FlexExtract, so the answer may be obvious).*

The details to access the MARS archive can be found on https://www.ecmwf.int/en/forecasts/access-forecasts/access-archive-datasets. Then one can use FlexExtract on the met data to extract it for use in Flexpart, in the desired resolution.

6. *Line 349. "the distance metric for the rain water data is rather large, at 1500km".Figure 16 suggests this should say fallout data instead (or Fig 16 is incorrectly labelled).*

   Thank you for pointing out this error, it should indeed read "fallout data".

7. *Line 408. "source localisation and reconstruction with deposition measurements… can yield useful results in the context of radiological emergency preparedness".*

   I think this part of the sentence needs further justification (I do not question the rest; the work has undoubtedly proved the use of deposition measurements for CTBT-relevant events). What is meant exactly by 'emergency preparedness' and how would the results support it? If 'emergency preparedness' includes 'emergency response' (as mentioned in line 24), how do the results translate into that context, where source term estimates might be required on shorter timescales than used here (i.e. in this application, the first usable deposition measurements were made days(?) after the event, but source term information might be desired within hours in an emergency).

   The context we refer to is both radiological emergency preparedness and response. We have now made this consistent in the manuscript. Emergency preparedness encapsulates the preparation one requires to properly respond to an unforeseen radiological emergency situation. The results of our study are useful for both emergency preparedness and response since we show that having the ability to perform deposition measurements can aid in source reconstruction, and hence inform the further response.

   In the Ru106 case, the deposition measurements were made several days after the event since that is the time it took the plume to reach these locations. However, it is also true that in general there will be a larger time delay to obtain deposition measurements compared to air concentration measurements, the latter of which are routinely measured in detector networks while deposition measurements are not. Nonetheless, deposition measurements can be made in an emergency response scenario. As part of emergency preparedness there (can) exist mobile units which are equipped with deposition measurement apparatus. Depending on detection limits, measurements can be made within the order of hours to aid real-time source reconstruction. Furthermore, it would also be theoretically possible to build automatic deposition detectors in detector networks.

Deposition measurements can also be used to retrospectively estimate the source term, as has been done after the Fukushima accident. This information can be used to implement emergency measures based on the estimated impact on the food chain and total dose exposure.