General comments

This paper studies a novel use of deposition measurements: the locating of an undeclared source. Source inversion in this context is usually a data sparse problem; the method presented enables more of the available data to be used a consistent manner, increasing confidence in inversion results. The method is systematically investigated and its presentation in this paper is clear and methodical.

Below, I have a few comments / questions which I think require consideration before the paper is ready for publication. If these are addressed, I encourage the publication of this work.

My thoughts largely overlap with the comments from the first two anonymous referees. Hence, I restrict my comments to those which differ from / extend theirs.

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.

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"?

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.
   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 h defined by: $y = \int_{-\infty}^{\infty} (h \cdot c)\, d^3x\, dt$, 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.

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?
    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.
    c. *"[the CDS] can be defined relative to the full domain, or relative to the sub-domain 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 difference in definition represents only a different way of calculating the same metric.

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.

    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.

    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).

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).

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).