Editor's comments:

1. Please address the comments from reviewer #3.

- ✓ **We are grateful for the valuable comments and suggestions from both yourself and Referee #3. We have taken all of the concerns into consideration and made the necessary revisions to the manuscript (indicated by line numbers in red). Our responses to the feedback are highlighted in bold text, and the updates to the manuscript are shown in bold and italicized text.**

2. You should reference and read the paper of Karion et al. (2019) (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6605086/) in which WRF-Chem and dispersion models coupled to WRF are compared.

- ✓ **It is considered in the lines 244-246: "*These results point out that particle dispersion modeling is primarily influenced by both the meteorological inputs and the dispersion model of choice, which is consistent with the results of Karion et al. (2019).*"**

3. A missing part in the paper is the validation of the met data, and especially the precipitation field. How realistic are they? In its present form, you list differences between FLEXPART coupled with FNL, CFSV2, FNL-WRF and ERA5-WRF. Which member has the lowest bias in wind speed and precipitation anyway? Do you improve the meteorological data by coupling FNL to WRF with a downscaling to 10km?

- ✓ **Thanks for your suggestion. We have added subsection 3.1 to the text. This part includes the comparison of meteorological inputs and observations. All of the above issues are addressed in this subsection. In this part we conclude that lines 256-258: "*From these results, we conclude that the downscaled datasets correlate better with observations, at the expense of an increase in error and bias values, especially for the wind speed.*"**

4. Let's assume that a nuclear accident happens in your region of interest. Which meteorological input would you use with FLEXPART anyway?

- ✓ **As you mentioned, the only option would be the GFS dataset if the goal is prediction. Otherwise, we have accentuated the need to an ensemble of different meteorological inputs for the dispersion modeling of radionuclides. For example, lines 316-317: "*the differences have resulted in TIDI values varying by a factor of 2 to 10 in the south of the area of interest between ensemble members.* lines 517-518:**

We attribute this to the fact that differences, however small, in the meteorological inputs that lead to cumulative deviations in the transport and concentration calculations of atmospheric pollutants.”.

5. It is worth considering to use NCEP GFS at 0.25x0.25 degree, which is one of the easiest weather model forecasts to access and would be a natural choice to predict the dispersion of a plume.

✓ **The NCEP GFS dataset is used instead of the CFSv2 dataset in the revised submission.**

6. Detailed information on your WRF simulations is missing: which PBL scheme do you use, which convection scheme, surface scheme etc.

✓ **Overview of the WRF model configuration is presented in Table 1.**

7. By using the total column, you discard, to some extent, any effects on vertical transport due to the PBL development. Actually, the capability of the model to simulate correctly the PBL is critical when estimating surface concentration or deposition. The work on the total column should be limited to the general analysis of the transport pattern. Any intercomparison between models should be removed and only the first 100 or 150m should be used.

✓ **The total column analysis is thoroughly removed. Lines 156-158: “*Thickness-weighted averages of simulated concentrations, hereafter referred to as near-surface concentrations, are calculated from concentrations within the bottom four model layers between 5 and 100 m agl (with layer thicknesses of 5 m, 5 m, 40 m, and 50 m).*”**

8. You show total deposition while there is interesting information in wet and dry depositions. Wet deposition will vary with the cloud field and precipitation in the met data.

✓ **Thanks for your suggestion. The performance analysis of the ensemble members is considered separately for dry and wet Cs-137 deposition in subsection 3.3 in lines 453-470. Please also see Figure 13.**

9. The dry deposition scheme will be directly affected by the boundary layer scheme. 417 trajectories per hour is not enough to get a good statistic at the surface. You should increase the number of trajectories by a factor of 10. You can reduce the compuation time by limiting the trajectory length (with an ageclass) to 96 hours.

✓ **We performed a test run based on the GFS dataset with a ten-fold increase in the number of particles (100000 particles in the first 24 hours of each 96-hour**

simulation period or ~ 4000 per hour). The results showed that the dry deposition simulations based on 10000 vs. 100000 particles agree well (S1). Therefore, to reduce the computational load, 10000 particles were used for all runs. Please see the lines 197-201*: "To assess if this rate is sufficient, we study the dry deposition process, which is directly influenced by the boundary layer conditions. We performed a preliminary test run with GFS data with a 10-fold increase in particles. As shown in S1, the simulated dry deposition of ^{137}Cs and ^{131}I and the wet deposition of ^{137}Cs do not undergo significant changes with the increase of an order of magnitude."*

7) CFSV2 is definitively an outlier in your ensemble. As referee #3 mentioned, the CFSV2 at a 0.5x0.5 degree resolution met data should not be used within the analysis since obviously, only 3 grid cells separate the source and receptor. You should use the 0.2 x 0.2 degree version. As referee #3 mentioned: "the poorer spatial resolution of CFSv2 inputs caused the faulty separation of land and sea boundary layer process along the coastlines of Qatar. This also may lead to the suboptimal modeling of particle dispersion across the study area, especially along the coastlines"

✓ **The NCEP GFS dataset is used instead of the CFSv2 dataset in the revised submission.**

8) The downscaling exercise is difficult. In your paper, you compare the results of CFV2 to the WRF results, but they are based on completely different models. I recommend that you test your best FLEXPART-WRF set up, and go to a 4km resolution. I understand that it will be computationally hard to reduce the resolution to 4km, but you can probably do it for representative weather situations (a couple of days for each season for instance).

✓ **We have tried to implement your suggestion by downscaling FNL to the 4 and 2 km resolutions. However, in both cases, we have faced technical issues to reconcile the simulation output with the observations. Therefore, we decided not to include the 4km- and 2km sensitivity analysis in this paper, and leave that for a future study.**

✓ **However, based on your suggestion, we have added subsection 3.1 for the evaluation of the meteorological inputs. In this section, the effect of downscaling on the meteorological inputs and, subsequently on the radionuclide simulations, is examined by comparing the FNL and FNL-WRF datasets with**

meteorological observations and with each other (before implementing different dispersion models).

Referee #3's comments:

Major concerns:

1) 157: "Modeled concentrations are vertically integrated…" I see no reason what this should be good for (if not for a lack of particles in the surface layer following from too few particles being released). For any impact analysis, it is only layer (mostly surface up to ~ 150 m) concentrations (or doses) that count. Moreover, this approach makes subsections 3.2 and 3.3 incomparable. Model inter-comparison or evaluation will be biased to better outcomes because the vertical cloud or particle positions are no longer important in a total column comparison and upper layer concentrations are less impacted by tricky boundary layer processes. Total columns are mainly used in air quality studies for contrasting model values to satellite observations and this is the first time that I find total column values in a radionuclide dispersion study. The problem culminates in l. 390: "…is of importance in preparedness programs. Figure 9-A shows the frequency of occurrences (FoO) of 131I column densities…" If the authors seriously refer to column loads in the context of nuclear accident preparedness programs, there seems to be a lack of understanding.

- ✓ **Thank you for your suggestion. In the revised version, the weighted average of surface (0-100 m agl) radionuclide concentrations is used instead of column loads in the revised submission.**
- ✓ **With respect to point in l. 390, we have assumed that the high column loads coincide with the high surface values. However, column loads are not used in any of analyses in the revised version.**

2) Meteorological input data: l. 166/167: "…CFSv2 can be used to provide 6-hourly forecast inputs for FLEXPART at the spatial resolution of 0.5 degrees…" Given the scale of the study the use of this data set should be completely avoided. The emitter in UAE and the eastern boarder of the receptor area are just ~1.5° (i.e., three grid boxes) apart. It comes as no surprise that a lot of spatial gradients are lost in Figures 5 and 6. I fear that a lot of differences the authors are discussing between this member and the others are mere artefacts caused by the poor spatial and temporal resolution of the CFSv2 data. The authors even emphasize the problem of low resolution themselves

several times in the paper. The CFSv2 data should by no means be used for the purpose of forecasting in the case of an emergency.

- ✓ **The NCEP GFS dataset is used instead of the CFSv2 dataset in the revised submission.**
- ✓ **In the previous submission, we had discussed the limitations of CFSv2 for local-scale modeling in the previous submission. However, you have correctly raised the point that the use of this dataset has impacted our analysis and respective conclusions. Therefore, it is replaced by GFS dataset in the revised submission.**

Additionally, the (re-)analysis ensemble members are correlated (especially ERA5-WRF and FNL-WRF) and not well suited to quantify meteorological uncertainty.

- ✓ **In subsection 3.1 (lines 213-240) we discussed that simulations based on the ERA5-WRF and FNL-WRF datasets show only a relatively better agreement. The comparison of these meteorological inputs with observations and with each other (Figure 2), and also the comparison of the resulting simulations (Figures 12 and 13) show significant differences though. Even in the case of the FNL- and GFS-based simulations, the significant similarity between the meteorological inputs (Figure 2) does not prevent significant differences between the resulting simulations (Figures 4, 6, 7, 12, and 13). There are notable differences in the spatio-temporal distribution of radionuclide simulations based on the FNL and GFS datasets. We believe that our results highlight the need to the ensemble of different meteorological inputs for the dispersion modeling of radionuclides. For example, *lines 312-317*: "*The advance of TIDI above 2500 µSv to the southeast of Qatar in simulations based on FNL inputs occurs in both fall and winter, but is observed only in winter in the GFS-based simulations. TIDI values peak in the fall in both downscaled runs, with inputs from the ERA5- and FNL-WRF datasets, but the extent of high TIDI values to the southeast of Qatar in fall is much larger in the ERA5-WRF run. This is also the case when comparing the FNL-WRF and FNL-based simulations. The differences have resulted in TIDI values varying by a factor of 2 to 10 in the south of the area of interest between ensemble members*"**

3) The authors often mix integrated with maximum I-131 concentrations or integrated Cs-137 deposition with completion of C-137 deposition. It is only the maximum concentration and the completion of deposition which can be reasonably contrasted to

particle release times. Any time-integrated value naturally loses its time stamp information. Terms are correctly introduced by the authors in l. 326-328. But the statement in l. 348-349 already starts to confuse the reader. It culminates in the statement in l. 371/372, l. 523 or in the caption of Figures 7 and 8 that the integrated I-131 [muSv] concentration is converted to maximum hourly dose [muSv/h].

- ✓ **All above issues have been addressed. In the corrected version, it is stated that we have only checked the release time and age of particles corresponding to the maximum amount of I-131 concentrations and completion of C-137 deposition.**

4) The comparison with the results of Maurer et al. (2018) (l. 441/442, 447-450 and l. 569-571) to me demonstrates that the authors have a poor knowledge of atmospheric radionuclide dispersion modelling. Not only that Xe-133 – in contrast to I-131 or Cs-137 – is an inert tracer which undergoes no deposition (thus being easier to model) and that Maurer et al. (2018) employed atmospheric dispersion runs based on NWP (re-)analyses only (no forecast was involved), the scale of this study was completely different. Whereas the scale of the present study covers 200 or 300 km the source and the receptors in Maurer et al. (2018) are mostly several 1000 km (up to 17000 km) apart! Finally, IMS sampling times are not one hour but rather 12 to 24 hours. So, I am sorry to say, this is comparing apples with pears.

- ✓ **We agree with the referee's comment. In response, we have removed any comparison between the simulations of I-131 or of Cs-137 and Xe-133. In the revised version, we cite Maurer et al. (2018) only as a reference for the evaluation metrics.**
- ✓ **We only wanted to show that two radionuclides that are more difficult to model than Xe-133 are still within a reasonable range compared to simulations based on the reanalysis dataset (which in our study represent observations). However, as you also pointed out, we now find this comparison far from valid for several reasons.**

The abstract should cover the full paper in a balanced way. E.g., there is no word about the dose calculations.

- ✓ **We have revised the abstract to better represent the scope and significance of the work. The updated abstract provides a clearer and more comprehensive overview of the research aims, methods, results, and conclusions. We believe that the revised**

abstract more accurately reflects the contributions of our study and will be of greater benefit to the readers.

The English is sometimes poor and sentences hard to understand. Tenses are often switched arbitrarily (present tense versus past tense). Some of the minor issues below are mere suggestions, others clearly reflect a lack of care in terms of contents or wording. Sometimes authors are even contradicting themselves within the paper. It is urgently needed to increase coherence and consistency within the paper.

- ✓ **We have thoroughly reviewed the English language of the manuscript to ensure its clarity and coherence. Our efforts have focused on correcting grammar, spelling, punctuation, and syntax errors, as well as improving the readability of the text. To optimize the language of the MS, we will also use the Copernicus proofreading service.**

Minor issues:

- l. 11 ff.: "intensity of radionuclides" -> "(activity) concentrations of radionuclides". No proper wording. Occurs numerous times throughout the paper. Remove all occurrences of "intensity" or "intensities" in the very same context in the paper.
- ✓ **Done**.
- l. 12, l. 50, l. 182, l. 184: "a fictitious accident" -> "fictitious accidents". You investigated in fact 365 scenarios.
- ✓ **Done.**
- l. 23/24: The difference in input PBLH explains well the inter-member variations of simulated radionuclide concentrations. See major concern 2). The PBLH alone will not explain all the differences.
- ✓ **The entire discussion of differences in PBLH is removed.**
- l. 24/25: "Simulated concentrations were found with the same level of consistency as reported for real case studies". See major concern 4).
- ✓ **Corrected.**
- l. 38: "…from the Fukushima nuclear power plant accident…": Somehow a contradiction to what is said above ("…case studies of real accidents of the order of a few days are not suited to examine the impact of seasonal (atmospheric) changes on the radionuclide dispersion."). This was a real accident and the effect of East Asian northeast monsoon on radionuclide transport was evidently studied.
- ✓ **Here, we express the need to study a nuclear event with similar emission intensity but under different weather conditions. In fact, if the Fukushima incident had occurred under different weather conditions, the distribution of radionuclides could be very different from what was observed. This is where it becomes necessary to conduct studies similar to what we have done.**
- ✓ **lines 42-47**: ***In addition, Long et al. (2019) studied the effects of the East Asian northeast monsoon on the transport of radionuclides from the Fukushima nuclear power plant accident to the tropical western Pacific and Southeast Asia. They found***

that in these regions, radioactivity levels are lower than in other regions of the Northern Hemisphere, which is due to the late arrival of the radionuclide plumes carried by the monsoon circulations. That is, the dispersion of radionuclides from this accident could potentially be different under other atmospheric conditions, which are only captured by the hypothetical, iterative simulation of this event at different times of the day and year.

- l. 40: "northern hemisphere" -> "Northern Hemisphere"

✓ **Done.**

- l. 48: "…transport and surface concentration and deposition…" -> "transport, surface concentration and deposition"

✓ **Done.**

- l. 53: "131 I concentration": The main reason for the significance of I-131 are thyroid doses, not just the high activity of I-131.

✓ **Modified. Please see section 3.2 (Figure 6) where I-131 concentrations are converted to the thyroid internal dose from inhalation (TIDI, in units of μSv).**

- l. 55 and l. 86: Please add "Pisso et al., 2019" to the references.

✓ **Done.**

- l. 63: "lack of accuracy". Please specify. With regard to the internal modelling time step?

✓ **It is reworded to clarify. lines 67-69: *In some cases, the particles may not remain well-mixed during simulation (Brioude et al., 2013). This is mainly due to the treatment of the stochastic motion of the particles and/or the mass balance of vertical velocity with the horizontal winds.***

- l. 68: "perturbations". Please specify and/or provide a reference.

✓ **All three approaches are discussed by Galmarini et al. (2004). In the revised version, this reference precedes the approaches.**

- l. 69: "suite of different meteorological models": Difference in practice may be limited due to NWP models being similar to each other and thus an ensemble can easily give an incomplete picture of meteorological uncertainty. See for example your ERA5-WFR versus FNL-WRF inputs.

✓ **Although the GFS- and FNL-based (and also FNL-WRF- and ERA5-WRF-based) simulations have the higher agreement due to the similarities between their meteorological inputs. But significant differences are also reported in the resulting spatio-temporal distribution of simulations. For example, *lines 312-317: The advance of TIDI above 2500 μSv to the southeast of Qatar in simulations based on FNL inputs occurs in both fall and winter, but is observed only in winter in the GFS-based simulations. TIDI values peak in the fall in both downscaled runs, with inputs from the ERA5- and FNL-WRF datasets, but the extent of high TIDI values to the southeast of Qatar in fall is much larger in the ERA5-WRF run. This is also the case when comparing the FNL-WRF and FNL-based simulations. The differences have resulted in TIDI values varying by a factor of 2 to 10 in the south of the area of interest between ensemble members.***

✓ l. 71-73 "…is compared against the (re)analysis members. (Re)analysis-based simulations are expected to be closer to (unavailable in a real-world scenario) actual values than forecast-based ones (Leadbetter et al., 2022)." -> "…is compared against (re)analysis members. (Re)analysis-based simulations (unavailable in a real-world scenario) are expected to be closer to actual values than forecast-based ones (Leadbetter et al., 2022).”

- ✓ **Done.**
- l. 96/97: Please check the suitability of references. Tipka et al. describes the preprocessing of ECMWF fields before being ingested into FLEXPART, the other two papers deal with convection. But likely not removal processes (decay and deposition).
- ✓ **Relevant references are added in lines 101 and 102 “computing various removal processes (Stohl et al., 2005, Grythe et al., 2017)”**
- l. 113: “…by (Hanna, 1982).” -> “…by Hanna (1982).”
- ✓ **Done.**
- l. 114: “…method as Maryon (1998) is followed.” -> “…method as in Maryon (1998) is followed.”
- ✓ **Done.**
- l. 115: “In dispersion modeling…” -> “In dispersion modeling of radionuclides…”
- ✓ **Done.**
- l. 118: “…from already calculated the radionuclide…” -> “…from the radionuclide…”
- ✓ **Done.**
- l. 123: “…the Weather Research and Forecasting (WRF)…” -> “…the Weather Research and Forecasting (WRF) model…”
- ✓ **Done.**
- l. 126 + 127: “cloudy pixels” -> “cloudy grid cells”
- ✓ **Done.**
- l. 130: “…any grid cells beneath these grid cells…” Do you mean beneath a value of 80% or beneath in terms of altitude?
- ✓ **Corrected. please see the lines 132-137 *“In previous versions, including version 9.0.2 used in the development of FLEXPART-WRF 3.2, in-cloud grid cells are defined as those with precipitation and relative humidity above 80%. The grid cells below the in-cloud grid cells up to the surface are defined as below-cloud grid cells (Seibert and Arnold, 2013, Pisso et al., 2019). In recent updates to the FLEXPART's source code, the above threshold has been modified using the 3D cloud water mixing ratio (qc) fields. The threshold of qc > 0 (qc = 0) now identifies grid cells within the cloud (below the cloud) (Pisso et al., 2019).”***
- l. 131: “…cloud water mixing ratio…” This field is now used to distinguish between below- and in-cloud grid cells. Please state this explicitly.
- ✓ **Corrected. please see the above answer.**
- l. 141: “Eq7” –> “Eq. 7”
- ✓ **Done.**
- l. 148: “…estimate the transport…” -> “…estimate the temporal characteristics of transport…”
- ✓ **Done.**
- l. 150: “…is added to the history output grids that have a horizontal resolution of 10 km…” -> “…is added to the output grid that has a horizontal resolution of 10 km…”
- ✓ **Done.**
- l. 152/153 & caption of Fig. 3: “smooth density estimates”. In how far smooth? Were the distributions smoothed?

- ✓ **The intensity of the smoothing is controlled by the kernel bandwidth. In our study, it is determined by using the method called Scott's rule: n**(-1./(d+4)), where n is the number of data points and d is the number of dimensions.**
- l. 153: “normalized difference” -> “maximum normalized difference”
- ✓ **Done.**
- l. 170/171: “…reanalysis data that covers from January 1, 1950, to nearly the present. They are produced at a spatial resolution of about 31 km at hourly time steps.” -> “…reanalysis data that covers January 1, 1950, to nearly the present. They are produced at a spatial resolution of about 0.25° at hourly time steps.
- ✓ **Done.**
- l. 175/176: “A single simulation code is built for each meteorological dataset to be ingested by FLEXPART…” I do not really understand why. In FLEXPART 10.4 there is even only one executable for both, ECMWF and NCEP data. So I wonder why there should be different codes for two NCEP data sets. Finally, authors are contradicting themselves in l. 303 (“same simulation code”).
- ✓ **We could not run FLEXPART with input from the FNL and CFSv2 datasets using the same executable. We had to compile two different executables files. However, as you said, the simulation code is the same. In the revised manuscript, the term “executable” is removed and we use only "the simulation code" throughout the manuscript. The paragraph above is deleted.**
- Table 1 needs to be improved. Use capitals consistently in the header, e.g., “Temporal resolution”. First line with entries: Better remove “x”. Second line with entries: Add “0.25°” and “3-hourly” for FNL. Better remove “x”.
- ✓ **Done.**
- l. 184: “May 2011” - > “March 2011”
- ✓ **Done.**
- l. 193: “This experiment has been performed…” -> “This experiment was performed…”
- ✓ **Removed.**
- l. 194/195: “For the diurnal and seasonal stratification of simulations…” –> “For stating particle ages related to simulations…”
- ✓ **Removed.**
- Caption of Figure. 1: “Figure 1 A is the study area embracing the B-NPP (red square) and the state of Qatar. The base map and overlaying information are taken from Google Earth. B is the schematic illustration of the LPDM simulation cycle.” -> “Figure 1. A: Study area embracing the B-NPP (red square) and the state of Qatar. The base map and overlaying information are taken from Google Earth. B: Schematic illustration of the LPDM simulation cycle.” I suggest a similar style for all the figures. Use full stops and colons accordingly.
- ✓ **Done. This correction is applied to all figures with a similar caption structure.**
- l. 205: “…since it resides mostly in the gaseous phase and has a short half-life of 8 days” This is not true according to my knowledge. The best assumption is a 50:50 partition between gaseous and particulate iodine. Again, the thyroid doses are an important aspect of iodine.
- ✓ **Based on data by the “Ring of Five (Ro5)”, an informal network of European national authorities (which comprises more than 150 sampling systems of high**

volume samplers and some with activated coal traps), the average gaseous/total ratio for 131I is 77.2 ± 13.6 % (Masson et al., 2011). The US Environmental Protection Agency (EPA) RadNet station measurements detected 81 % of the ambient 131I in the gas and 19 % in the particle phase (Ten Hoeve and Jacobson, 2012). The average of 71 ± 11 % reported from the Fukushima site from 22 March to 4 April 2011 (Stoehlker et al., 2011). Therefore, we decided to consider only the gaseous phase of I-131, especially the Cs-137 can represent particulate pollutants.

- l. 208: "…season and time of day in which…" -> …season and time of day in/at which…"
- ✓ **Removed.**
- Figures 2 and S1: "FNL-WRF" -> "Forecast". Are times in Fig. 2B/S1B UTC-times? Y-axis: "all intensity" -> "all loads". Caption of Figure 2: "Figure 2 A: the smooth density estimates of air parcel ages corresponding to all intensities of 131I column densities (top row) and of those above the 66th percentile (bottom row). B: the same as A, but for four times of the day." -> "Figure 2. A: Density estimates of air parcel ages corresponding to all 131I column loads (top row) and of those above the 66th percentile (bottom row). B: The same as A, but for four times of the day."
- ✓ **Done. All day times are converted to local time in the revised version.**
- l. 210-212: "…differences in the transport characteristics of these radionuclides, such as the wet and dry deposition rate and radioactive decay, are not large enough to cause the abundance of cases where 131I and 137Cs particles are not present in a common grid." This is even not to be expected, because removal process have no influence on particle positions in LPDM. Just on particle masses. However, the transport for gases and particulates (undergo gravitational settling) will be different to some extent.
- **Corrected. Lines 243-244: *the high degree of similarity between the age distributions of 131I and 137Cs is due to the fact that the removal process in LPDMs only affects mass concentrations and not particle positions*.**
- Figure S2: "…lines are corresponding kernel density estimates of counts." Completely obscure to me. Needs (sufficient) explanation in the text.
- ✓ **The explanation is added to the caption of S2 (S7 in the revised version) *"The distribution of the ratio of 137Cs wet deposition to total 137Cs deposition for each ensemble member (colors). The counts of the ratios are smoothed with a Gaussian kernel. The counts on the y-axis are on a logarithmic scale."***
- l. 216: "All ensemble members in all seasons simulated an abrupt increase…" I think this is not true for summer.
- **This conclusion holds true in all seasons after column loads are replaced by near-surface concentrations.**
- l. 225: "In addition to having finer spatial resolution than CFSv2, FNL assimilates observations like ERA5." Also CFSv2 needs to assimilate observations at some point, i.e., at the analysis time step based on which the forecast is made.
- ✓ **Due to replacement of the CFSv2 datasets by GFS dataset, the statement is removed.**
- l. 226: "distributed in a wider range" Did you check the significance of this feature?
- ✓ **We explained this feature in more detail in lines 345-355. We believe that further elaboration of this feature is beyond the scope of this paper.**

- l. 226/227: "…in FNL…" -> "in FNL-based FLEXPART simulations" or "FNL simulated…" -> "FLEXPART-FNL simulated…" FNL does not yield the output directly. It needs FLEXPART in addition. Numerous analogous formulations throughout the paper. Please adapt them as outlined.
- **Corrected throughout the text.**
- l. 228 + l. 357: "…air parcels reaching further receptors…" -> "…air parcels reaching receptors further away…"
- ✓ **Done.**
- l. 232: "column (mass) densities" (or density). Please note that this is not a proper term. Occurs numerous times in the paper. It has to be "column load". If you vertically integrate concentrations [Bq/m3] you will end up with column loads [Bq/m2]. However, as stated above, column loads are not used in radionuclide studies.
- ✓ **This correction does not apply to the revised submission because the weighted average of surface (0-100 m agl) radionuclide concentrations replaced the column loads.**
- l. 242/243: "For low intensities of both radionuclides, however, the age distributions are almost the same as that seen in all intensities. The peak of newly arriving air parcels in all intensities occurs earlier in spring than in other seasons." -> "For low column loads of both radionuclides, however, the age distributions are almost the same as that seen for all column loads. The peak of fast arriving air parcels for all column loads occurs earlier in spring than in other seasons." I would rather say spring, winter and to some degree also fall behave very similar in terms of fast arriving particles.
- **This statement is removed because all of the above patterns are not seen in the distribution of particle ages corresponding to different levels of near-surface concentration, which replace column loads in the revised version.**
- l.244: "…in low and all intensity column densities…" -> "…for low and all column loads…"
- ✓ **Removed.**
- l. 245: "…led to the more distant transport of radionuclides to northern parts of the study area…" Can this really be concluded that easily? See the options stated at the end of the above paper paragraph.
- ✓ **This statement is removed. Please see the response to your point about l. 242/243.**
- l. 249: "…the age distribution of the particles that is released…" -> "…the age distribution of the particles that are released…"
- ✓ **Done.**
- l. 250/251: "…the number of long-lived air parcels (including all intensities of concentrations) increased in all members…" -> "…the number of long-lived air parcels (including all levels of column loads) increased for all members…" Anyway, not true for Cs and FLXPART-FNL (Figure S1).
- ✓ **In order to shorten the article (at the same time adding section 3.1 related to the comparison of meteorological inputs with observations) and also to avoid unnecessary discussions, we have removed the discussion of the effect of the release time of particles on their age distribution. However, the effect of particle release time on the intensity of radionuclide concentration and deposition are discussed in Section 3.2 (Figures 6, 7 and 8).**
- l. 253: "longer Lagrangian particle ages" -> "higher Lagrangian particle ages"

- ✓ **Please see the response to your point about l. 250/251"**
- l. 254: "less time to travel by the end of simulation period" -> "less time to travel until the end of simulation period"
- ✓ **Please see the response to your point about l. 250/251.**
- l. 257: "…between 20 and 70 hours after release…" I would rather say "…up to 40 or 50 hours after the release…" for moderate and high column loads."
- ✓ **Please see the response to your point about l. 250/251"**
- l. 257/258: "Consequently, the difference in release time of Lagrangian particles is not significantly affected by their age spectrum." Sentence makes no sense. If, then the other way round.
- ✓ **Please see the response to your point about l. 250/251"**
- l. 258/259: "…FNL have simulated a larger number of shorter-aged air parcels…" -> …FLEXPART-FNL simulated a larger number of shorter-lived air parcels…"
- ✓ **Please see the response to your point about l. 250/251".**
- l. 261/262: "Like seasonal distribution, the diurnal variations of air parcel ages for high concentrations are very similar in all members." -> "Like for seasonal distributions, the diurnal variations of air parcel ages for high column loads are very similar for all members."
- ✓ **Please see the response to your point about l. 250/251".**
- l. 276: This equation is quite obscure to me. If j=1 (first simulation time step) "a" can only be equal to one (j=a=1). There cannot be particles older than one hour at this stage of the simulation. For j >1, e.g. j=50, i can only range between 26 and 50 and cannot adopt a value equal to 1 in the nominator given that particles were released over the first 24 hours of the simulation.
- ✓ **Thanks for your comment. We have made changes and corrections To Eq. 9 to include all possible conditions leading to CS deposition equal to zero. Please see lines 276-284**

"To analyze the relationship between the age composition of air parcels and the amount of ^{137}Cs deposition, the deposition values cumulatively aggregated across time steps (j) and age spectra (i) are normalized to the total amount of ^{137}Cs deposition in each grid cell (k) at the end of each simulation run (l).

$$^{137}\mathrm{Cs}_{klna_{norm\_depso}} = \begin{cases} \dfrac{^{137}\mathrm{Cs}_{kl(n-1)n}}{\sum_{j=1}^{96}\sum_{i=1}^{95}{}^{137}\mathrm{Cs}_{klij}} & if\ n = 2 \\ \dfrac{\sum_{j=1}^{n-1}\sum_{i=1}^{j-1}{}^{137}\mathrm{Cs}_{klij} + \sum_{i=1}^{a}{}^{137}\mathrm{Cs}_{klin}}{\sum_{j=1}^{96}\sum_{i=1}^{95}{}^{137}\mathrm{Cs}_{klij}} & if\ n > 2 \end{cases} \quad (9)$$

where n is the time step with a maximum of 96 (the last time step) and a is the given particle age with a maximum of n-1. $^{137}\mathrm{Cs}_{klij} = 0$ in two conditions (1) $n >= 26$ if $i \in [1, n-25]$ and (2) $i >= j$."

- l. 277: "…in winter when both dry and wet deposition occur in the study area. We found very similar results in other seasons (Fig. S5)." Something got quite confused here. Comparing Figure S5 with Figure 4 it is easy to see that, e.g. results for FLEXPART-FNL, are not identical for season winter.
- ✓ **Regarding "…in winter when both dry and wet deposition occur in the study area", this shows why we chose the simulations in winter to discuss in the main text.**
- ✓ **We did not claim they are identical but similar: please see lines 286-287: "*Similar deposition patterns are obtained for other seasons (S6). As shown in S7, the main reason for the small difference between the seasonal deposition patterns is the lack of precipitation and subsequent wet deposition in the region.*"**
- I guess Figure 4 in fact depicts the full year. I also wonder about similar results in other seasons given that wet deposition will hardly occur in summer in the study area.
- ✓ **Corrected. Sorry for the mistake. It was an old figure left over from the first draft.**
- l. 277/278: "As expected, the values of the 137Cs deposition increase cumulatively with the time after the accident." This indeed should be the case. But this is not what can be seen in Figure 4. There is no steady increase in, e.g., the median. Consequently, the statement in the next line, i.e., "deposition (the median of normalized deposition > 0.8) happens within 80 hours after the assumed accident" is misleading to me. According to my understanding Figure 4 probably says that 80% of the deposition is accomplished by particles up to an age of 80 hours integrated over all possible time steps.
- ✓ **Regarding <u>There is no steady increase in, e.g., the median,</u> this represents deposition variations in the areas far from the source.**
- ✓ **lines 288-290: "*Although the spatial pattern of the deposition varies considerably, as indicated by the range of quartiles, the median of the normalized deposition shows that about 80 percent of the deposition occurs within 80 hours after an accident. The cumulative deposition at the end of the simulation period is mostly in the areas farthest from the source.*"**
- Caption of Figure 4: "This plot (S5) shows simulations in winter (other seasons)." Weird sentence in this context.
- ✓ **Corrected. lines 285-286: "*Figure 5 shows the normalized deposition amounts ($^{137}Cs_{klna_{norm\_depso}}$) in winter, when both dry and wet deposition occur in the study area. Similar deposition patterns are obtained for other seasons (S6).*"**
- l. 292: "…the seasonal and diurnal changes…" -> "…the spatio-seasonal distribution…"
- ✓ **Modified.**
- **Figure 5: Too busy with regard to release time isolines. I think they should be thinned out, local minima and maxima be removed, respectively. Please also state the unit, i.e., muSv.**
- ✓ **Corrected.**
- l. 296: "131I_intg_conc_seas": Needs to be explained when first introduced.
- ✓ **Done**.
- l. 297: "The model inputs…" -> "Dose coefficients…"?

- ✓ **Corrected. we meant "The conversion coefficients". Please see the lines 301-303: "*Since approximately 90% of the population of Qatar is in the adult age group (UNStats, 2020), TIDI is specifically calculated for this age group using the coefficients defined by WHO (2012) (Fig. 6).*"**
- l. 301: "…in most parts of study area than the other members." -> "…in most parts of the study area compared to the other members."
- ✓ **Due to the replacement of column loads by near-surface concentrations, this statement is removed.**
- l. 306: "…over emission point in UAE and, with less intensity, within the study area…" -> "…over the emission point in UAE and, with lower heights, within the study area…" What is the reason for this discrepancy? For which time (noon?) is Figure S9 valid?
- **Due to the replacement of CSFv2 by GFS, this statement is removed.**
- l.307: "dilution" -> "vertical dilution"
- ✓ **Removed.**
- l.308: "process" -> "processes"
- ✓ **Removed.**
- l.311: "cold period" Rather mostly fall instead of winter.
- ✓ **Corrected. See lines 305-308: "*Our ensemble simulations show that the TIDI values above 2500 µSv occur frequently in the cold period of the year, especially in fall, in the simulations of all members. This may be due in part to the lower (higher) PBLH in the cold (warm) seasons and to the synoptic conditions.*"**
- l. 314: "exceptional transport" Why is this feature not present in the both WRF simulations?
- ✓ **This pattern is only seen in FNL-based simulations. Therefore, I think you mean why it is not seen in the simulations based on GFS dataset. If so, we believe that "*This may be caused by a rare atmospheric circulation in summer, discussed later, and it may change as the modeling period is extended*" (lines 308-309).**
- l. 315-319: Comparison to the Fukushima accident is too vague. How do you define "adjacent"? If I am not mistaken the paper shows doses over 96 hours for iodine only. But the paper cited evidently provides integrated doses for three months and three nuclides.
- ✓ **We discuss the thyroid internal dose from inhalation (TIDI) in the revised version. Therefore, the above comparison is removed.** Lines 303-305: "***As reference values for comparison, the total TIDI values, collectively calculated for 15 studied radionuclides for the adult age group, were found to be between about 2,000 and 50,000 µSv in the first year following the Fukushima accident in areas close to the power plant (see Table 4 in WHO (2012))***."
- l. 324/325: "For example, the simulations of FNL in southern Qatar in fall are more than twice that of FNL-WRF (more details in the subsection 3.3)." -> "For example, the simulations based on FNL in southern Qatar in fall result in more than twice the dose compared to that based on FNL-WRF (more details in the subsection 3.3)."
- ✓ **Due to the replacement of column loads by near-surface concentrations, the discussion is modified**.
- l. 328-329: "To identify the highest possible level of pollution at each point, regardless of its frequency, local maxima are calculated only from non-zero intensities." Sentence is completely obscure to me.

- ✓ **Modified. lines 318-321: “*To investigate the influence of the particle release time on the radionuclide dispersion, the seasonal median of the particle release time (hours in local time (LT)) coinciding with the maximum concentration of 131I and the completion of 137Cs deposition is considered (contours in Figures 6 and 7). The results show that the highest 131I concentrations coincide with particles released between 9 a.m. and 2 p.m. LT in most parts of the study area*”**
- l. 330: “is caused” -> “are caused”
- ✓ **Done**.
- Figure 6 and 7B: Needs to be redone with units converted to kBq/m2, critical thresholds (40 and 90 kBq/m2, see text) clearly distinguishable and probably replacing the whitish part of the color scale.
- ✓ **Done**. **Please see Figure 7.**
- l. 335-343: As a result of the deficiencies of Figure 6, the discussion is not really traceable looking at the figure. Specifically, 40 and 90 kBq/m2 cannot be distinguished (both thresholds fall within the withish part of the color scale).
- ✓ **Corrected. Due to the replacement of CSFv2 by GFS, 90 kBq/m2 threshold (found in CFSv2-based simulations) is removed.**
- l. 346: “…the higher CFSv2 137Cstot_depos_seas in winter…” It is not really higher, but rather affects a broader area.
- ✓ **Due to the replacement of CSFv2 by GFS, this statement is not considered in the revised version.**
- l. 362 + caption of Figures 5 & 7: “…particle ages simultaneous…” -> “…particle ages occurring simultaneously…”
- ✓ **Captions are rephrased**.
- Figure 8: Convert Bq/m2 to kBq/m2.
- ✓ **Done**.
- l. 375: “…extremely high levels of inhalation doses (higher than 200 µSv) and of 137Cstot_depos_seas (higher than 100 kBqm−2)…” Please provide a reference for these – according to your opinion - “extremely high” thresholds.
- ✓ **It is just a comparative conclusion about the simulations in the study area. Therefore, we believe it does not need a reference. We have already provided a reference for the threshold (40 kBqm-2) used for identifying contaminated areas.**
- ✓ **Please also note that inhalation doses are replaced by TIDI that changes the critical levels in the discussion above. For example, “…. *the extremely high levels of TIDI (greater than 10,000 µSv)*….”**
- l. 379: “…131Iintg_conc_seas inhalation doses…” -> “…131Iintg_conc_seas related inhalation doses…”
- ✓ **Corrected. we refer to TIDI in the revised version.**
- l. 379/380: “…there are no pronounced seasonal and inter-member differences in 137Cstot_depos_seas at different population levels.” This is hard to believe looking at Figure 6.
- ✓ **Corrected. Lines 369-372: “*Due to the exceptional weather pattern that occurs in the simulations based on the FNL dataset in the eastern part of Qatar (where the most densely populated areas are located) in summer, these simulations cause the highest values of TIDI in densely populated areas. Otherwise, all ensemble members simulate the highest TIDI during the cold seasons. The highest* $^{137}Cs^{tot\_depos\_seas}$ *in the***

same areas are observed based on ERA5 and FNL-WRF in the cold period of the year."

- l. 391/392: "…north of Qatar…" -> "…in northern Qatar…"
- ✓ **Done**.
- l. 394: "…15% of extreme events have reached receptors in Qatar." -> "…15% of extreme events occur in Qatar."
- ✓ **Done**.
- l. 398: "has caused"-> "causes"
- ✓ **Done**.
- l. 400/401: "…occurs mainly in the late winter-early spring period simultaneous with the southward movement of westerlies and the eastward movement of the Saudi Arabian subtropical high pressure." -> "…occurs mainly in the late winter-early spring period simultaneously with the southward movement of westerlies and the eastward movement of the Saudi Arabian subtropical high pressure system."
- ✓ **Done**.
- l. 410: "As shown in Figure 10-A, the median of simulated column densities…" -> As shown in Figure 10-A for FLEXPART-ERA5-WRF, the median of simulated column loads…"
- ✓ **Due to the replacement of the column loads by near-surface concentrations, the discussion is modified**.
- l. 411: "There is no comparable change in other seasons that is reflected in a minute increase in the full-year distribution of…" -> "There is no comparable change in other seasons which is reflected in a minute increase in the full-year distribution of I-131 load for…"
- ✓ **Due to the replacement of column loads by near-surface concentrations, the discussion (lines 404-414) is modified**.
- l. 413-414: "The upper quartile of 137Cs deposition with STM in winter is around 25% higher than those simulations with GTM in the same period." -> "The upper quartile of 137Cs deposition related to STM in winter is around 25% higher than that related to GTM in the same period.
- ✓ **Corrected (lines 404-414)**.
- l. 415: "The implementation of STM in FNL-WRF…" -> "The implementation of STM in FLEXPART coupled with FNL-WRF…"
- ✓ **The discussion (lines 404-414) is modified**.
- l. 417: "mainly in fall": What is the synoptic pattern related to the feature being pronounced in fall?
- ✓ **In this part, we discuss the effect of the turbulence scheme of choice a local scale. Therefore, the discussion of synoptic patterns seems irrelevant. In addition, the discussion (lines 404-414) is modified due to the replacement of column loads by near-surface concentrations.**
- Figure 10: Please change the y-axis units to kBq/m2 (you can probably stay with Bq/m3 for concentrations if you remove the vertical integration) and introduce scientific notation.
- ✓ **Done. Please see Figure 11.**

- l. 434: "…from the six International Monitoring System (IMS) stations…" -> …from six International Monitoring System (IMS) stations…" There are more than 70 IMS radionculide stations in total!
- ✓ **Done**.
- l. 436: "…normalized by the sum of the two means…" -> "…normalized by the sum of the simulation and measurement means…"
- ✓ **Done**.
- Figure 11: I wonder about the spurious beams in the regression analysis. Especially about the blue ones in the regression analysis for the full year (yellow data points). I would expect simple lines. Moreover, the seasonal subplots can be skipped in my opinion. In the present layout they are too busy with the red points for fall covering much of the remaining data points. Additionally, they are hardly discussed at all in the text (just l. 443 and 444)
- **All season-related discussion is removed to address the above issues and to make room for the discussion of wet deposition.**
- l. 446 ff.: "FB5" -> "F5" Occurs numerous times in the subsequent paragraphs.
- ✓ **Done**.
- l.452: "…the normalized RMSE which is less sensitive to extreme values…" Sorry, I do not understand. Large deviations will be even enforced due to the quadratic term involved. Also, the NMSE is not equal to the normalized RMSE. Apart from normalization the square root is not applied for NMSE.
- ✓ **Thanks for the point. In fact, in our analysis in the previous submission, we used the normalized RMSE. But after your comment, we checked the reference and now use the correct one.**
- ✓ **Due to the replacement of the column loads by the near-surface concentration and the replacement of CFSv2 by GFS, this discussion is no longer valid.**
- l. 453: "(2)" -> "(2.6)"
- ✓ **Please see the answer above**.
- l. 454/454: "Feeding downscaled FNL inputs into FLEXPART-WRF (FNL-WRF) increased the correlation of ERA5-WRF and FNL-WRF to 0.7." - > "Feeding downscaled FNL and ERA5 inputs into FLEXPART-WRF (FNL-WRF and ERA5-WRF) increased the correlation to 0.7.
- ✓ **Rephrased**.
- l. 456/457: "…indicating that the downscaling of inputs in FNL-WRF did not have much effect on the association of their simulations." Comparing this statement with Figure 5 (where there is a considerable difference between FNL and FLN-WRF based simulations) the inconsistency introduced by using column loads and surface concentrations (doses) alternately becomes very evident.
- ✓ **We discussed here just the relative association (correlation) of simulations (not their absolute difference). Anyway, the column loads are replaced by the near-surface concentrations.**
- l. 458-460: "This suggests that the downscaling of similar (FNL) and different (FNL and ERA-5) meteorological datasets increased and decreased the absolute differences between resulting simulations, respectively." This sentence makes no sense to me. FNL is similar to what? Overall it is evident from the statistics and does not come as a surprise that simulations based on ERA5-WRF and FNL-WRF are most closely related.

- **Rephrased.** **lines 441-444**: “***On the other hand, the downscaling of the inputs and the application of the same simulation code yield higher agreement of the simulations based on the FNL-WRF and ERA5-WRF datasets than to those based on the FNL and FNL-WRF datasets (r=0.7 vs. 0.67, FB=0.01 vs. -0.93, F5=69.83% vs. 53.3%, RMSE=36012.82 Bq/m3 vs. 113888.80 Bq/m3, and NMSE=2.74 vs. 10.08)***.”
- l. 460/461 and caption of Figure 11: “…relative distribution of simulations…” -> “…relative column load distribution of simulations…”

✓ **Modified**.

- l. 461/462: “All distributions here depicted the higher frequency of low 131I column densities in spring than in other seasons (as also seen in Figure 2-A).” Figure 2-A does not display the frequency of column loads but rather that of particle ages! What is reason for this exceptional feature? Why does it not occur in summer as well?

✓ **Regarding old results, this may be due to synoptic patterns in spring that transport radionuclides to areas further away from the source.**

✓ **As requested, all season-related discussion is removed.**

- l. 463: “While the RMSEs have increased tangibly…” I would suggest checking this result. I would have expected it rather the other way round because simulated I-131 levels will overall be higher compared to Cs-137 levels due to the larger source term. Thus, absolute differences between modelled values are prone to be larger for I-131.

✓ **Corrected. We apologize for this error. The I-131simulations seem to be drawn in the unit of ng/m3 in the previous submission. They are in the correct unit (Bq/m3) in the revised submission.**

- l. 474: “The other statistics…” -> “The statistics…”

✓ **Removed.**

- l. 496: “We quantified meteorological uncertainties by producing an ensemble model…” You can hardly quantify the meteorological uncertainties based on three re-analysis members which are correlated (FNL – FNL-WRF, but above all ERA5-WRF – FNL-WRF). The ECMWF, e.g., uses 50 (!) non-redundant ensemble members for uncertainty quantification. Running FLEXPART based on these (or at least a fraction of these) members would deserve the term “uncertainty quantification”. See major concern 2).

✓ **We discussed that simulations based on the ERA5-WRF and FNL-WRF datasets are only in a relatively better agreement. The comparison of these meteorological inputs with observations and with each other (Figure 2), and also the comparison of the resulting simulations (Figures 12 and 13) show significant differences. Even in the case of the FNL- and GFS-based simulations, the significant similarity between the inputs (Figure 2) does not prevent significant differences between the resulting simulations (Figures 4, 6, 7, 12, and 13). For example, lines 316-317: “*the differences have resulted in TIDI values varying by a factor of 2 to 10 in the south of the area of interest between ensemble members.* Lines 517-518: *We attribute this to the fact that differences, however small, in the meteorological inputs that lead to cumulative deviations in the transport and concentration calculations of atmospheric pollutants.”*.**

- l. 502: “…were also examined concerning the population density…” -> “…were also examined in relation to the population density…”

✓ **Done.**

- l. 513 “in spring” I would say in fall too.
- ✓ **Due to the replacement of column loads by near-surface concentration, this conclusion is no longer valid.**
- l. 527: “stronger” -> “larger”
- ✓ **Done.**
- l. 535: “We found…” -> “We suspect…”
- ✓ **The discussion related to PBLH is removed.**
- l. 541: “in winter” It looks like (Figure 6) this pattern is even more pronounced in fall.
- ✓ **Corrected. Lines 520-523: “*The largest expansion of areas with $^{137}Cs^{tot\_depos\_seas}$ greater than 40 kBqm^{-2}were found in the simulations based on the ERA5-WRF and FNL-WRF datasets in winter. The highest levels of $^{137}Cs^{tot\_depos\_seas}$ (above 300 kBqm^{-2}) were found in the simulations based on the ERA5-WRF dataset in the southeastern corner of Qatar in the fall.*”**
- l. 550: “…in which south/southeast winds transport…” -> “in which south/southeast winds in the cold season transport…”
- ✓ **This is not necessary, as “winter” is mentioned in two lines up (line 531).**
- l. 556/557: “…decreased the median of simulated 131I concentrations…” Only in fall.
- ✓ **After the replacement of total column by near-surface concentrations results changed as follow: Lines 537-542 “*The implementation of the Skewed Turbulence Model (STM) instead of the Gaussian Turbulence Model (GTM) increased the occurrence of high levels of ^{131}I concentrations and ^{137}Cs deposition. For example, the quartiles of simulations of ^{131}I concentration and ^{137}Cs deposition based on the ERA5-WRF dataset increased by 21% and 12%, respectively. According to Pisso et al. (2019), this can be interpreted as the enhancement of concentrations and deposition in the areas around the source under skewed turbulence conditions*”**
- l. 561: “In general, CFSv2 simulations of 137Cs and 131I column density are most highly correlated with FNL…” -> “CFSv2 simulations of 137Cs and 131I column loads are most highly correlated with FNL…” Mind that FNL – FNL-WRF correlation is highest.
- ✓ **Due to the replacement of CFSv2 by GFS, this conclusion is removed.**
- l. 563: “FB = 0.02” -> “FB = -0.02”
- ✓ **Removed.**
- l. 560-575: Please avoid stating numerous statistical scores in the conclusion sections. There can be two or three of them but the conclusion section should contain take-away-messages rather than multiple numbers.
- ✓ **Done.**
- l. 573: “RMSE=14.6 x 10^8 and 2946.3 x 10^7”? Exponents?
- ✓ **Corrected. Please see also answers for l. 561 and l. 463.**
- l. 576: “…simulations from FNL and FNL-WRF with identical meteorological inputs.” FNL and FNL-WRF are not “identical”, but of course correlated to some degree. The authors should decide at some point in the text whether their ensemble input quantifies meteorological uncertainties (see l. 496) or is identical which makes the use of an ensemble obsolete.
- ✓ **In the corrected version, in several parts, we concluded that the simulations of the ensemble members have significant differences from each other. We believe that**

our results highlight the necessity of the ensemble of different meteorological inputs for radionuclide dispersion modeling. Please see the response to l. 496.

BRIOUDE, J., ARNOLD, D., STOHL, A., CASSIANI, M., MORTON, D., SEIBERT, P., ANGEVINE, W., EVAN, S., DINGWELL, A. & FAST, J. D. 2013. The Lagrangian particle dispersion model FLEXPART-WRF version 3.1. *Geoscientific Model Development,* 6**,** 1889-1904.

GALMARINI, S., BIANCONI, R., KLUG, W., MIKKELSEN, T., ADDIS, R., ANDRONOPOULOS, S., ASTRUP, P., BAKLANOV, A., BARTNIKI, J. & BARTZIS, J. 2004. Ensemble dispersion forecasting—Part I: concept, approach and indicators. *Atmospheric Environment,* 38**,** 4607-4617.

GRYTHE, H., KRISTIANSEN, N. I., GROOT ZWAAFTINK, C. D., ECKHARDT, S., STRÖM, J., TUNVED, P., KREJCI, R. & STOHL, A. 2017. A new aerosol wet removal scheme for the Lagrangian particle model FLEXPART v10. *Geoscientific Model Development,* 10**,** 1447-1466.

KARION, A., LAUVAUX, T., LOPEZ COTO, I., SWEENEY, C., MUELLER, K., GOURDJI, S., ANGEVINE, W., BARKLEY, Z., DENG, A. & ANDREWS, A. 2019. Intercomparison of atmospheric trace gas dispersion models: Barnett Shale case study. *Atmospheric Chemistry and Physics,* 19**,** 2561-2576.

MASSON, O., BAEZA, A., BIERINGER, J., BRUDECKI, K., BUCCI, S., CAPPAI, M., CARVALHO, F., CONNAN, O., COSMA, C. & DALHEIMER, A. 2011. Tracking of airborne radionuclides from the damaged Fukushima Dai-ichi nuclear reactors by European networks. *Environmental Science & Technology,* 45**,** 7670-7677.

PISSO, I., SOLLUM, E., GRYTHE, H., KRISTIANSEN, N. I., CASSIANI, M., ECKHARDT, S., ARNOLD, D., MORTON, D., THOMPSON, R. L. & GROOT ZWAAFTINK, C. D. 2019. The Lagrangian particle dispersion model FLEXPART version 10.4. *Geoscientific Model Development,* 12**,** 4955-4997.

STOEHLKER, U., NIKKINEN, M. & GHEDDOU, A. 2011. Detection of radionuclides emitted during the Fukushima nuclear accident with the CTBT radionuclide network. *Monitoring research review: Ground-based nuclear explosion monitoring technologies***,** 715-724.

STOHL, A., FORSTER, C., FRANK, A., SEIBERT, P. & WOTAWA, G. 2005. The Lagrangian particle dispersion model FLEXPART version 6.2. *Atmospheric Chemistry and Physics,* 5**,** 2461-2474.

TEN HOEVE, J. E. & JACOBSON, M. Z. 2012. Worldwide health effects of the Fukushima Daiichi nuclear accident. *Energy & Environmental Science,* 5**,** 8743-8757.