Dear editor,

We would like to thank the reviewer for his additional comments that helped us to further improve the clarity of the manuscript, although we do not think that they needed neither a major revision to be addressed nor more than three months to be delivered. As we explain below, a few comments were relevant to clarify some points, while we found other comments/suggestions not fully relevant for the goal of the paper, that is a gas hazard analysis and not a fluid dynamic study on a specific test case, nor a validation of the simulation tools (DISGAS and TWODEE-2) used by VIGIL.

In the following we list the reviewer's comments that needed to be addressed in italics and our responses. When lines in the manuscript are reported, these refer to the track-changes version of the file.

It's unclear what the domain is being described as "small" in relation to. The significance of a regime change is determined by the dilution rate, and no domain is inherently "small": its size depends on the specific variability scale of the phenomena being modeled. In my view, the possible change of regime as the gas flow dilutes downstream, along with the model's approximations, should be explicitly addressed in the description of the methodology to provide clarity on how the domain size relates to the physical processes being studied.

(…)

I agree this point and, at my advice, a sentence like this one should be put in the text.

We added a few lines addressing this point (lines 195-201). We agree that the transport regime can change as the gas flow dilutes downstream, but as we explain in the answer to the next point, the use of one model (e.g., DISGAS) or the other (e.g., TWODEE-2) does not result in a significant discrepancy and, from a hazard assessment point of view, does not imply an underestimation of the hazard.

It is not correct to conclude that an effect is insignificant by comparing RMSE values with maximum values. Instead, RMSE should be compared with the mean value above the background concentration across both the entire domain and the reduced domain. This mean value should be calculated as:

$$MV = \frac{\sum_{i=1}^{N} C_{i,DISGAS}}{N} - C_{background}$$

These values should be included in Table 5 of the revised manuscript as a reference point for the RMSE values provided. Furthermore, since the hazard is presented as a map rather than a spatial average, a map showing the difference between the two models in either the worst-case or a representative scenario is necessary. This will help illustrate the potential epistemic error in the hazard maps as a function of location. This difference map should be generated at the same height at which the hazard maps were extrapolated. Additionally, the manuscript should specify the height at which the MV and RMSE were calculated.

(…)

The same comment applies here as previously mentioned: the mean value (MV) and a representative difference map should be included and discussed in the paper. These additions will help provide a clearer understanding of the model's accuracy and potential epistemic errors across the domain.

As suggested by the reviewer, we calculated the Mean Values (MV) for both tests in the following ways:

- For the Richardson number-dependency test (Table 1), for each tested Richardson number scenario, MV was calculated as suggested by the reviewer but, since the RMSE was calculated comparing the concentrations predicted with DISGAS and TWODEE2, MV was calculated by averaging the MVs obtained by each model.
- For the resolution-dependency tests (Table 2 and 3), for each test (3m vs 1.5 m and 6 m vs 1.5 m), MV was calculated using the concentration obtained with a resolution of 1.5 m.

In both cases, the background concentration was set to 400 ppm. Furthermore, both RMSE and MV were calculated across all the domain and taking all the output heights into account.

As it can be seen from Table 1, the ratios between RMSE and MV in all tested cases range from 0.02% to 5.11%, therefore we can conclude that both the error introduced by the scenario selection (dilute vs. dense gas) and the error introduced by the resolution selection is not significant. For the same reason, we do not deem necessary to introduce new maps in the manuscript that can make the reading heavy without providing useful information. However, we have now explained this point at lines 446-453 and modified Table 5 in the manuscript. Table 2 and 3 presented here, are now added in Appendix 1, with an explanation on the test and its results now included at lines 463-467.

Table 1. Results of RMSE calculations for the Richardson-dependency test.

Ri	RMSE All domain [ppm]	MV All domain (ppm)	RMSE/MV [%]	RMSE Reduced domain [ppm]	MV Reduced domain [ppm]	RMSE/MV [%]
0.438	54.71	1246.25	4.39	102.38	4360.19	2.35
0.625	60.84	1223.66	4.97	159.56	8051.21	1.98
0.812	77.82	2326.30	3.35	155.10	9576.85	1.62

Table 2. Results of RMSE calculations for the resolution-dependency test for the DISGAS case

resolution	RMSE All domain [ppm]	MV All domain [ppm]	RMSE/MV [%]	RMSE Reduced domain [ppm]	MV Reduced domain [ppm]	RMSE/MV [%]
3 m vs 1.5 m	1.07	1223.77	0.09	3.15	15691.99	0.02
6 m vs 1.5 m	4.73	1230.35	0.38	11.91	9232.30	0.13

Table 3. Results of RMSE calculations for the resolution-dependency test for the TWODEE2 case

resolution	RMSE All domain [ppm]	MV All domain [ppm]	RMSE/MV [%]	RMSE Reduced domain [ppm]	MV Reduced domain [ppm]	RMSE/MV [%]
3 m vs 1.5 m	44.53	1915.58	2.32	187.22	26863.64	0.70
6 m vs 1.5 m	98.18	1920.91	5.11	286.52	15037.50	1.91

Thank you for your clarification regarding the model's parameters. I understand that some of these parameters may vary significantly, and while this variability is important, I believe it would still be highly beneficial to present them in a table. For parameters with a wide range, simply reporting the range of variability should suffice. However, for parameters that are kept fixed, such as the numerical diffusion mentioned, it is crucial to include these explicitly in the manuscript. Providing a clear list of all parameters used in the simulation ensemble would greatly enhance the reader's ability to quickly understand the key assumptions and approximations underlying the model. This would also improve the transparency and reproducibility of the work.

We do not agree there is need for additional tables, as the few parameters that are fixed in DISGAS and TWODEE-2 simulations are clearly specified in the manuscript while the other that are varied are fully discussed or the proper literature reference provided (e.g., the turbulent diffusion coefficients for DISGAS, calculated as explained in lines 123-126). The main variables that were varied to capture their statistical variability were the wind field in the whole domain, based on the DIAGNO outputs initialized from ERA-5 reanalysis (for the details on how DIAGNO works we properly cited the DWM User Guide that obviously cannot be summarized in a table or even in the article) and the emission rate, for which we explicitly said that the values were varied sampling from a normal distribution (lines 225-226), whose statistical parameters are reported clearly in the manuscript. Moreover, all the data, codes, input files, instructions to re-run the PHA are clearly explained in the manuscript and provided in the Zenodo repository. For these reasons we consider the reviewer comments that constantly question the transparency and reproducibility of the work unfounded.

OK. I suggest explicitly stating that the regime is selected by calculating the Richardson number using the wind field at 10 m above the ground.

Done in line 250.

Additionally, details regarding the data used to evaluate the Monin-Obukhov length scale should be included in the manuscript to enhance transparency.

Done in lines 132-136.

Regarding the vertical wind profiles, I understand that Figure 7 represents data at a fixed height and that the available meteorological data have relatively low vertical resolution. However, for this very reason, it is crucial to comment on the variability of the original dataset. For instance, if the dataset includes only two vertical points, it would be useful to illustrate this variability using Rose diagrams. This would help explain how the limited data *are translated into the much more detailed wind field produced by the DIAGNO model. To improve clarity, one option could be to present Rose diagrams side by side: one for the heights where meteorological data are available and another for the wind field at 10 m above ground level, as generated by DIAGNO. Additionally, I believe it would be helpful, if possible, to show an example of a vertical wind profile from the DIAGNO model, overlaid with the two original points from the ERA5 dataset. This would provide a clearer view of how DIAGNO refines the coarser input data into a more resolved wind profile.*

We disagree with the reviewer on this comment and on the need to add further figures to our manuscript in addition to figure 7. As we explained above and in the text, we used ECMWF ERA5 reanalysis meteorological data, which resolution is too coarse to capture local variability and topographic effects, to initialize the mass consistent wind module DIAGNO, which provides the local high resolution wind field near the ground that is then used to run the gas transport model. The goal here is not to compare the wind field from large to local scale, for which the reader can refer to the DWM user guide, but to statistically capture its variability in order to build a probability map. From a statistical point view, it was demonstrated that, for hazard analysis scopes, the used dataset is not critical for capturing the aleatoric uncertainty (see for example Macedonio et al., 2016 analysed the effects of different dataset for the case dispersion of volcanic ash on a larger scale: Macedonio G., Costa A., Scollo S., Neri A. (2016) Effects of eruption source parameter variation and meteorological dataset on tephra fallout hazard assessment: an example from Vesuvius (Italy), J. Appl. Volcanol., 5, 5, 1-19, doi:10.1186/s13617-016-0045-2).

However, we added few lines to stress this point at 132-136.

OK. The fact that the gas emission rate is varied by randomly sampling it across the 1,000 simulations should be explicitly stated also in lines 240-250 to prevent any confusion. In its current form, Section 4.1 gives the impression that 1,000 days with different meteorological conditions were selected independently of the source conditions. However, as I now understand, this is not the case. Each day could potentially have a different source condition, randomly sampled as briefly described later in the conclusion. This raises a question about the stability of the results: if the same workflow were run again, different outcomes could emerge due to the random sampling. While I recognize the constraints imposed by computational resources and am comfortable assuming this effect is likely negligible, I believe this point should still be explicitly addressed in the manuscript. Clearly describing this aspect will help readers understand the inherent variability in the results and the model's robustness.

The emission rate was sampled randomly by a normal distribution as already stated in lines 225-226. We added a new sentence in lines 250-251 which we hope will make clearer that the emission rate was varied, within the normal distribution, for each of the sampled day.

OK. Please incorporate this reasoning into the manuscript for clarity.

The lack of a recent measurement campaign was already pointed out in lines 491-492 (surveys in the area are also quite dangerous so not very frequent). However, we included further lines in the Conclusions section (lines 494-497).

I believe Figure 7 could be highly useful for readers attempting to understand how many scenarios fall into one regime or another. To enhance clarity, I suggest adding a reference *to Figure 7 in support of the discussion around line 225. This would help readers better visualize the distribution of scenarios across different regimes.*

We are sorry but in our opinion presenting Figure 7 in the position where the reviewer suggests is not useful. Figure 7 shows the domain-averaged wind strength at 10 m above the ground for each season and was originally introduced to explain the seasonal control on the probabilistic outputs of CO_2 concentration. In our opinion, Figure 7 does not help understanding how many scenario fall into one regime or another, an information provided clearly in lines 254-256.

Additionally, I would like to point out that the reasoning around line 225 is based on the mean gas emission rate. Please ensure this is made explicit in the text to avoid any ambiguity.

Done in line 229.