We thank the reviewer for the useful comments and suggestions and believe that now this manuscript has been significantly improved. Please, find hereafter our replies to all your comments highlighted in red, whereas all modifications in the main text are in yellow. Thank you for having considered this manuscript.

The manuscript presents the results of a modelling study aiming at quantifying tephra deposition from lava fountain events on the Etna volcano road network. Authors use two models (Tephra2 and FALL3D) to model the transport and deposition of the tephra released from the volcano and find that both models give similar results when taking into account the sensitivity tied to volcanic quantities (here a factor of 5).

I find this to be an interesting study looking at an important problem. However, I feel that despite its novelty, at the current stage the study feels a bit too simple. I do believe though that some additional simulations and analysis can lead to a much more impactful publication. To this end I have written down some recommendation and a number of concerns about the methodology employed:

1. As the authors' note, there is no analysis of the impact of the ESP sensitivity. Considering the relative novelty of the work presented, I would emphatically argue that this is a perfect opportunity to look into the sensitivity of the results to the different combinations of ESP values. I feel that including a robust error analysis (see comment 3) and a quantification of the impact, the study will have a much stronger central point.
>> In paragraph 2 and in the discussion we have added the reference to paper Scollo et al., 2008b, where it is described the sensitivity analysis of ESP on results of Tephra2 and Fall3D. Following your suggestion we have added an error analysis on results derived from both models, from the radar and field data.

2. I am not sure if Tephra2 is the right model to use here as the study focuses on relatively proximal dispersal of unsteady plumes over very complex topography. From the Scollo et al 2019 study, I understand that this is probably due to the fact that Tephra2 is part of the forecast system employed. Is this correct? I think that objectively, FALL3D is a more appropriate model to use, so I would frame this as cross-examination of the Tehpra2 results (necessitated due to the computational constraints) against a more sophisticated model.
>> The reviewer is right. .Tephra2 is the model for the real-time forecast of ash plume dispersion at INGV-OE surveillance system since 2009 but it also includes topography effects. Differently, the Fall3D model is more complex with respect to Tephra2 but due to the computational constraints is less suited for surveillance issues. In this paper, we don't want to explore what is the best model with respect to another but we wish to estimate the uncertainty due to the different physical assumptions and computational constraints. However, comparisons among the results of the two models and field data of 28 February 2021 were also added to estimate the goodness of both models.

3. Even though the eruption studied here are described in Scollo et al 2019, I feel that a section discussing their main characteristics is warranted.
>> We added a section to describe the main features of these Etna events.

4. Despite the presentation of ground observations along the radar-derived values, there is no proper error-based evaluation. This is particularly important for the comparison of the two model results. There's a number of error metrics commonly used (RMSE, MAPE, bias etc) along with correlation coefficients such as Pearson or Kendall tau. Considering the nature of concentrations and depositions, the use of the logarithm error might offer a better tool as it penalises both over and underestimation in the same way. A proper error analysis can help put the conflicting model results into better context.
>> We stressed this point and we added a description of the statistical error analysis on the retrievals of tephra mass such as derived from models and from the XWR, and we have compared them with field data (Pardini et al., 2021); further, we have added a new paragraph 4.5 with some considerations.

5. I'm not sure if I missed information regarding plumes, but how is tephra introduced in the model? Is there a plume model employed, or is it a standard profile or single point release? I know that the representation of tephra concentration along unsteady volcanic plumes is very much an open issue with no proper answer and considering the fact that forecasts do need to be carried out we must accept the use of simplifications, but I think that important information are missing.
>> We added in the paragraph the information about the plume model considered and different assumptions (Lines 206-209 and 212-217).

6. Expanding comment 4, in general there's a lot of important model configuration information that seems to be missing from the manuscript.
>> We added other references which refer to detailed information about employed models (Lines 216-217 and 223-225).

7. The estimation of the road width also needs more information. I have added a relevant comment in the pdf version, but in short, as the results directly scale with the road width chosen further information would help make a more convincing case.
>> We specified in the main text that the roads width have been obtained from both (i) randomly sampling some roads and measuring their width, thus obtaining on average a width of 6 m, and (ii) from the legislation as reported in the text. We also added a width variability of +/- 0.5 m to examine in depth this aspect.

8. This is a nit-pick, but the interpolation method seems computationally inefficient. Is the intermediate step of interpolating everything at 5m really necessary? I feel that directly interpolating over the road network would be more efficient.
>>The reviewer is right, but in this work we opted to interpolate the tephra load over the whole map, as it was still computationally easily feasible. We added a paragraph with a description of uncertainties and results variation.

9. Second nit-pick – the word "resolution" is used throughout the manuscript instead of "grid spacing". The two are not the same, as resolution refers to the scale of phenomena models are able to explicitly resolve. In the case of transport modelling, resolution is mostly tied to the grid spacing of the meteorological data.

>>The reviewer is right, we corrected it and replaced it with grid spacing.

Finally, there are some minor points, language errors, typos etc. I've highlighted some in the pdf, but the manuscript merits another careful read-through by the authors.

Overall, I think that this is an interesting and novel study that requires some additional simulations to truly reach its potential. My overall recommendation would be publication after major revisions as discussed above. I hope that the authors will find the comments constructive.

Kind regards and best of luck with the revisions.

>>We thank the reviewer for the several corrections and suggestions pointed out in the pdf, that have greatly improved the manuscript quality. We have changed the text according to each recommendation, and the various corrections are shown in the main text highlighted in yellow.