*Line 1. "Satellite retrievals of column mass loading of volcanic ash are **incorporated** into the HYSPLIT transport and dispersion modeling system".*

Does that mean that the procedures illustrated in the paper are now available with Hysplit?

Line 7. "the end of life of the ash cloud".

What do you mean with "end of life". Do you refer to the settling of ash particles on the ground, or do you refer to ash concentration in the atmosphere getting lower than a fixed threshold?

Line 10. "small pieces of ash".

What are the small pieces of ash? It is not clear if small refer to the size of tephra, or to small portion of the original ash cloud. If you are referring to a portion of the ash cloud, please change to "parts", also in the rest of the text, "pieces" is confusing.

Line 20. "However, within the next five years".

Please add a reference.

Line 38. "The resulting ash cloud reached a plume height of around 9km"

Whenever a plume height is given, it should be stated if it is above sea level or above the vent.

Fruthermore, according with Horvath et al. 2021 (https://doi.org/10.5194/acp-21-12207-2021), "At 20:50 UTC (Fig. 10f and g), the overshooting top of the eruption column reaches its maximum altitude of ~ 15.3 km according to the side view technique".

Can you comment this difference in the estimation of maximum column height?

Lines 42-44. "The resulting plume forms a complicated three dimensional structure as it is stretched and folded by the wind field over the course of less than a day. As shown later, the exact location and shape of these structures is difficult to forecast."

Throughout the manuscript, the use of the terms "plume" and "cloud" are sometime confusing. For example, in line 38, the term plume seems to be used for the volcanic column. Because of that, here it is not clear if the 3D structure of the plume refers to the column or the cloud. I think that a choice should be done at the beginning and the use of the terms should be clearly stated, and then it should be consistent throughout the paper.

Line 46. "most of the ash is drawn into an area of low pressure, the location of which is fairly easy to forecast. The end fate of these large ash clouds is similar to that of the Bezymianny cloud discussed here".

Here the text seems to suggest that a pressure gradient drives the ash trajectories, while it is the drag exerted by the wind that controls them.

Line 49. "path of ash parcels emerging from a low pressure area".

I don't understand what this means. How do ash particles emerge from low pressure areas?

Line 50. Before this paragraph, I would be useful and interesting to have more details on the event (VEI, mass eruption rate, total grain size distribution, observed concentrations, deposit distance, wind condition at the vent), in order to have an idea of the size of the eruption.

Line 52. "the dissipation of the ash cloud".

Maybe "dispersion" instead of "dissipation". I've never seen this term used for ash cloud by the volcanological community.

Lines 56-61. It is important to give more information about the satellite retrievals. It is difficult to understand how the total mass or the total area are computed without knowing what are the data used. What is the maximum/minimum size of particles detected? Are you using an estimated columnar content? Is there a threshold used for the detection? Is there an uncertainty associated with the detection? To better understand how the area is computed, it would also be useful to have here two satellite images at different times, with a contour delimiting the area.

Line 64. "as well as a mass fraction of fine ash of 0.1".

Please justify this value and add a reference.

Line 64. "this would result in a plume height of about 8.2 km".

Without knowing the wind condition, it is difficult to judge if the use of the Mastin 2009 relationship gives reliable results. Is the plume strong or weak?

Line 68. "the decrease in mass over time is due to physical processes such as dissipation due to dispersion".

Mass is not dissipated, it is always conserved. The local concentration could decrease because of dispersion, but this does not decrease the total mass.

Line 69. "gravitational settling, and wet and dry deposition".

Here it seems that settling and deposition are two different things. Is this correct?

Lines 80-84.

This paragraph needs more details on the way Hysplit was used. From my knowledge, Hysplit can be used with particles or puff and, accordingly with this choice, with different dispersion/diffusion model associated with turbulence. Also, the way ash concentration is computed depends on the choice of particles or puffs. In both the cases, have you performed a convergence analysis on the number of particles/puffs needed to obtain a stable output? Forthermore, being Hysplit a Lagrangian model (i.e. it does not solve directly for mass concentration), it should be explained how concentration is computed from particles/puff position.

Line 110. "plume width was 1 km".

Again, it is not clear here if with "plume" you are referring to the volcanic column or to the ash cloud. Please clarify. If you are referring to the volcanic column, it is not clear where you assume this width/diameter? Is it at the base or at the top? In general, the diameter grows a lot from the base to the neutral buoyancy level. It is also not clear how the puffs/particles are released for this RunA. Is it from a cylinder, from a line, from the lateral surface of a cylinder? Maybe this does not make

any difference, but in any case, if the reader wants to try to replicate the results, he/she needs to have all the information required.

Line 110. "Vertical mass distribution was uniform from the vent at 2.88 km to the plume height".

Is the mass distribution or the release of ash uniform? In both the cases, is this assumption justified? For small ash particles, most of the mass reach the neutral buoyancy level and it is released in the atmosphere at that height.

Line 111. "and a mass fraction of fine ash"

Please specify the value used.

Lines 112-114. "would result in an MER of …".

It is not clear if the MER refers to the eruption rate of fine ash only or if it is the total rate. In general, this term is used to the total rate, so a different use would be confusing. If it is already the total eruption rate, the values reported in these lines seems low for a 10km volcanic column.

Line 116. "The inversion algorithm".

Because Hysplit allows for inversion of trajectories, I think that a reader could get easily confused here. Please give some more details on the inversion procedure from Chai et al.

Line 118. "and 1 km in the vertical and area of above the vent"

Something is missing here.

Line 118. "The modeling system only consider ash passively advected by the wind".

Aren't you considering the gravitational settling?

Line 128. "with the mass distributed"

Please change "mass" to "mass release".

Line 143. "The difference between using 20 and 6 μm particles".

What is the difference in the settling velocity for the two sizes?

Line 148. "lower emissions"

Does "lower" refer to the height or to the rate?

Line 153. "we find that this is in large part due to the dispersion of the ash cloud not being adequately represented by the model"

Is this really a limitation of the model or a limitation or of the retrieval algorithm?

Line 162. "Bias correction with CDF matching"

Starting from this section, I really struggled to understand what has been done, mostly because the algorithms and techniques applied are not described with enough details to make then clear. For example, for the CFD matching, it is written that "model values and observed values" are sorted, but there is no mention to what are the values used. Are they probabilities, concentrations, pixel values, values averaged for all the pixels? I really have no idea. And the figure does not help, because there are no units on the x-axis. You need first to explain clearly for which parameter/variable you compute the CDF.

Line 167. "A linear fit is applied to the difference between the pairs as a function of forecast value. Although, Reichle and Koster (2004) use higher order fits, we find that a linear fit is adequate."

Maybe I looked at the wrong reference, but I can't find any mention of fit in the paper cited here. Please check.

Line 172. "among the number of ash layers present".

What do you mean with layers of ash?

Line 172. "mass loading values".

Usually the volcanological community use this for the deposit (i.e. loading on the ground).

Lines 187-190.

I would move these lines in the previous subsection.

Line 195. "Increasing the spread of the cloud at early times to more closely match observations, caused the spread of the cloud at later times to have a larger mismatch."

The umbrella cloud intrudes as a gravity current at it takes some time to reach the maximum upwind and crosswind spreading. Have you tried to increase the size with time?

Line 198. "The modeled horizontal dispersion of the cloud is too fast".

This could be also done to a release of ash particles from a too large vertical interval, coupled with vertical wind shear. What happens when a larger fraction of particles is released from the neutral buoyancy level? As previously commented, particles of size you are considering here should, in large part, reach the top of the column.

Line 215. "We make the assumption that verification of modeled column mass loading values can be used as a proxy for verification of forecast concentrations."

The distinction between the use of column of mass loading and concentration should have been done at the beginning, because in most of the previous sections it was not clear which of the two forecasts and observations were referring to.

Line 229. "output of ash column mass loadings shown in Figure 8"

The use of a linear color scale makes more difficult to compare the results. Please use a log scale. Also the choice of colors does not help.

Lines 233-234. "By 12 UTC, this line of ash has broken into three small pieces, one just to the east of the volcano, one to the south, and one to the northwest."

Are these parts at different heights?

Line 249. "An example is shown in Figure 9(c) which shows number of model runs exceeding a given mass loading threshold"

I would remove "an example", because the number is not normalized in figure 9(c).

Line 250. "84 %".

I think that here and in the rest of the text you should remove the space between the number and the percentage symbol.

Lines 259-261. "In later sections we will utilize measures such as the precision recall curve, PRC, to evaluate the ATL at various probability thresholds. Keep in mind that these statistics are the same for the APL with the caveat that the point for 5 % probability of exceedance threshold for the ATL becomes the point for the APL of 95 %."

This is difficult to understand, because the PRC has not been introduced so far in the manuscript.

Line 263. "Some sources"

Which sources? Add references.

Line 267. "If velocity is constant then D is not sensitive to spatial averaging that is performed parallel to the flight path."

This is true, but when computing the concentration in the grid cell, averaging is preformed both in directions parallel and normal to the flight path, so computed dosage depends on grid resolution.

Line 292. "Instead predicting the time at which the ash cloud breaks into small enough pieces to be ignored becomes important."

Written in this way it seems that it is common to observe the ash cloud breaking into small parts, but I'm not so sure it is always the case. It also not completely clear to me what "breaks into small pieces" means. Is it because between these "small pieces" ash concentration is very low and so it is not detected? Or is the ash cloud really splitting into different and isolated parts?

Lines 295-298. "The extent of lower mass loading of ash, 0.2 g m^{-2} continues to increase through 12 UTC. In contrast, the extent of the higher mass loadings follows the observations more closely. This mismatch occurs because the spatial gradient in the observations is much steeper than in the simulation. This is again an indication that the modeled turbulent dispersion which controls the spread of the ash is not reproducing what is observed."

This is not clear to me. If the extent of the area exceeding a threshold is too large, that does not necessarily mean that there is problem with turbulent dispersion. In fact, when the bias correction is applied, it seems that the areas are better reproduced.

Lines 309-315.

I'm sorry but the description is the reliability diagram is extremely confusing to me. In the simulations you have the ensemble of simulations, the output at different times, the output at different pixels. When you write "the probability of observing the event" what do you mean? Is it the probability for a single element of the ensemble, considering all the pixels at a single time? Is it the probability associated with the variability in the ensemble elements? Is it the variability associated with the different output times in a time interval? I have the same problem with the variable plotted in the vertical axis of the middle column in Fig.11, what does "Fraction observations" mean? I have the same problem with the refinement distribution. "The second part of the diagram is the refinement distribution which is a histogram showing how often the modeled probability, y_i, occurred". You need to state more clearly what you mean with "modeled probability".

Section 6.5.2

In this section I have problems to understand the discussions on the reliability diagram and the refinement distribution because, as written above, it is not clear what is plotted in the figures.

Lines 379-380. "Observed and modeled fractions are then computed for different neighborhood sizes, n, by convolution of the field with a square kernel of that size".

I suggest to change to:

"Observed and modeled fractions are computed by convolution of the field with a square kernel, for different kernel sizes, n.".

Line 384.

Before equation 3, please define O(n) and m(n). I assume they are the fractions, but it would be better to state it explicitly.

Lines 389-391. "When computing the FSS, it is standard for the reference forecast to be defined as the largest possible MSE that can be obtained from the forecast and observed fractions".

Isn't this number just 1? By looking at Eq.3, you maximize the MSE when you maximize each addend of the sum, and this is obtained when, in the difference between O(n) and m(n), one value is 0 and the other is 1. I also don't understand why in Eq.5, with respect to Eq.3, the square moved inside the parenthesis and there is a difference instead of a sum.

Line 397. "At some value of n <= 2N-1"

What is N?

Line 419. "Number of occasions"

Here I have the same problem I had with the description of the reliability diagram. The definition of the parameters seems to ignore that you are applying the technique to a specific application, and it is given as an abstract definition. What does the term "occasions" mean? I don't understand if you refer to the number of pixel for a simulation of the ensemble, or to the number of ensemble for a pixel.

Line 431. "For instance, a measure such as F is highly sensitive to the domain size "

What is F here?

Line 435. "various probability thresholds".

Is this correct or should it be just "various thresholds", without probability? Is this the same threshold used to define a in Lines 419-420? "both the forecast and observation are above threshold". If it is really a probability threshold, please can you write explicitly to what the probability refers to?

Line 436-437. "However, some care must be taken in the interpretation of the curves as POFD which is plotted on the x axis".

Where is the x-axis you mention? There is no reference to any figure here.

Line 443-end of subsection.

I confess that I was lost here, because I could not understand clearly the probability thresholds, as written above. In the two first columns of Fig.16 there are also a lot of markers, but I can't understand what they represent. Probably there is something I missed.

Figures

Figure 1. Please add letters to the panels. In panel (a), is is e^(-xx), with e being the Euler's number, or 10^(-xx)? In Panel (d), please write if the height is a.s.l.

Figure 6. In panel b and d add a label to the vertical axis.

Figure 8. The color scale does not help to distinguish low concentrations from the background. I think it would also work better a log scale.

Figure 9. I think you should remark that the color scale here is not linear or log, but I think a mix of the two.