

Author responses to Referee #1 and Editor comments on “Leveraging Machine Learning techniques and SEVIRI data to detect volcanic clouds composed of ash, ice, and SO₂”

Manuscript: egusphere-2026-727

We would like to thank Anonymous Referee #1 for his constructive comments and suggestions, which have improved the manuscript. Please find our replies to each comment below.

This document is structured according to the following color legend:

- Reviewers' comment.
 - Author's response.
 - Author's changes implemented in the manuscript.
- Lines 192-200: The section on the "calibrator model" is, I believe, new and not at all reproducible. How are the P_{cal} values calculated in the labelled dataset? What is this new model? The following paragraph, ending with your choice of 0.8 as a classification threshold, could be moved to your results section, which would get Figure 3 more views.

We thank the reviewer for these comments and appreciate the opportunity to clarify this aspect. The calibrator model itself is not new. However, its implementation and role within the methodology were explained in greater detail in the revised manuscript version 2. The calibrator model is a standard approach used to obtain calibrated probabilities from the NN model, and in this study it was incorporated as part of the training pipeline. In this new revised version of the manuscript, a more detailed explanation of the calculation of the P_{cal} values has been added, together with an appropriate reference.

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q.: On Calibration of Modern Neural Networks, 34th International Conference on Machine Learning, ICML 2017, 3, 2130–2143, <https://doi.org/https://doi.org/10.48550/arXiv.1706.04599>, 2017.

Niculescu-Mizil, A. and Caruana, R.: Predicting good probabilities with supervised learning, ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning, 625–632, <https://doi.org/10.1145/1102351.1102430>, 2005.

The P_{cal} (depicted in Figure 3a) is estimated by fitting an isotonic regression model using the NN raw output probabilities and the corresponding true labels from a testing set. The model learns the relationship between the raw probabilities and the observed probabilities (Guo et al., 2017; Niculescu-Mizil and Caruana, 2005). The calibrated probability for a new prediction is then obtained by applying the learned isotonic mapping to the NN model's raw output probabilities.

- Line 335: Is this an accidental new paragraph?

The referee is right. While this paragraph originally served to illustrate the SHAP figure interpretation, it has been relocated to the Discussion section, as it more appropriately belongs within that context.

- Line 341-349: This paragraph adds no new information and delays getting to the results. The referee is right. The paragraph has been deleted.

- Line 463: Since the difficulty of manual classification in the validation dataset is brought up in the discussion and conclusion, I suggest adding some information in the dataset description about the degree of difficulty, and if any uncertainty was attached for future use in the training & testing dataset.

We completely agree with the referee's suggestion; adding these details helps clarify the overall procedure and its future implications. We have addressed this point by adding a comprehensive description of the classification difficulty and uncertainty handling in Section 2.3 Dataset.

The manual plume mask generation generally provides the most accurate volcanic cloud detection, but it is a time-consuming process that requires expert interpretation. Despite the expertise of a human operator, some challenging events remain difficult to discriminate. This is particularly true when volcanic clouds are surrounded by meteorological clouds, making the identification of plume mask boundaries more uncertain. In such cases, the human operator tends to adopt a conservative approach, generating a smaller plume mask that includes only pixels with strong evidence of volcanic cloud presence. Consequently, the most ambiguous pixels were intentionally excluded from the manual plume mask, with the aim of allowing the NN model to learn how to discriminate these challenging events during the prediction phase.

- Line 494-5 and elsewhere: I concur that "machine learning" and "neural network" are not proper nouns, so capitalization is not correct when spelled out ("ML" and "NN" are fine)."

Done. Thank you.