Response to Reviewer 1 comments

Interactive comments on "Volcanic cloud detection using Sentinel-3 satellite data by means of neural networks: the Raikoke 2019 eruption test case" by Petracca et al.

We would like to thank the Reviewer for her/his constructive comments and suggestions, which have improved the manuscript.
Please find our replies to each comment below. Referee comments are reported in black. Our replies are given in red.

This manuscript presents a neural network model in order to detect volcanic ash clouds using Sentinel-3 SLSTR (Sea and Land Surface Temperature Radiometer) daytime products. The neural network is trained with MODIS daytime imagery from the Eyjafjallajökull eruption in May 2010. Then it is applied to the Raikoke eruption in June 2019. The results show that the neural network model can accurately detect volcanic ash from Raikoke compared with RGB visual inspection and BTD (Brightness Temperature Difference) procedure. Moreover, the plumes identified by neural network model agree well with the plume manually identified for the specific SLSTR images.

The manuscript is very well structured and written. It addresses an important issue in detection of the volcanic ash clouds and presents a solution which is beneficial for remote sensing and modeling volcanic ash dispersion. The methods and assumptions are scientifically sound and the results are well elaborated. Thus, I recommend the manuscript for publication after adressing the following points:

1- The authors should use/cite the published data instead of relying on private communication (L92). Specifically, there are several papers in this special issue that present ash and SO2 mass (e.g. Muser et al 2020, ACP). I strongly suggest that the authors review the published papers related to Raikoke and use them in the introduction and discussions.

We changed from:

It is estimated from the AHI data that June 2019 Raikoke eruption produced approximately 0.4–1.8 Tg of ash and 1–2 Tg of SO_2 (Prata, private communication).

To:

It is estimated from the AHI data that June 2019 Raikoke eruption produced approximately 0.4–1.8 Tg of ash (Bruckert et al., 2022; Muser et al., 2020; Prata et al., 2022) and 1–2 Tg of SO_2 (Gorkkavyi et al., 2021; Bruckert et al., 2022).

New references:

Bruckert, J., Hoshyaripour, G. A., Horváth, Á., Muser, L. O., Prata, F. J., Hoose, C., and Vogel, B.: Online treatment of eruption dynamics improves the volcanic ash and SO2 dispersion forecast: case of the 2019 Raikoke eruption, Atmos. Chem. Phys., 22, 3535–3552, https://doi.org/10.5194/acp-22-3535-2022, 2022.

Gorkavyi, N., Krotkov, N., Li, C., Lait, L., Colarco, P., Carn, S., DeLand, M., Newman, P., Schoeberl, M., Taha, G., Torres, O., Vasilkov, A., and Joiner, J.: Tracking aerosols and SO2 clouds from the Raikoke eruption: 3D view from satellite observations, Atmos. Meas. Tech., 14, 7545–7563, https://doi.org/10.5194/amt-14-7545-2021, 2021.

Muser, L. O., Hoshyaripour, G. A., Bruckert, J., Horváth, Á., Malinina, E., Wallis, S., Prata, F. J., Rozanov, A., von Savigny, C., Vogel, H., and Vogel, B.: Particle aging and aerosol–radiation interaction affect volcanic plume dispersion: evidence from the Raikoke 2019 eruption, Atmos. Chem. Phys., 20, 15015–15036, https://doi.org/10.5194/acp-20-15015-2020, 2020.

Prata, A. T., Grainger, R. G., Taylor, I. A., Povey, A. C., Proud, S. R., and Poulsen, C. A.: Uncertainty-bounded estimates of ash cloud properties using the ORAC algorithm: Application to the 2019 Raikoke eruption, Atmos. Meas. Tech. Discuss. [preprint], https://doi.org/10.5194/amt-2022-166, in review, 2022.

2- Raikoke and Eyjafjallajökull are both high-latitude volcanoes. How would the model perform on tropical eruptions like la Soufrière 2021? Is the model transferable to tropical conditions or different ash compositions? It will be interesting to see the application to la Soufrière.

Overall, the main purpose of the paper was to develop a methodology based on a neural network model able to classify SLSTR products for the Raikoke 2019 eruption, investigating the feasibility of training the model with MODIS data at comparable latitudes given the lack of SLSTR products for eruptions at such latitudes.

The complexity of the application suggests that the generalization of the methodology to all types of eruptions is not straightforward, and this was confirmed by some preliminary analysis (also including la Soufrière 2021). For example, the change of latitude has an impact on the characteristics of the atmosphere. At the same time different volcanoes emit different types of ash affecting the variability of the radiance values detected by the sensors. A possible solution to overcome the model transferability issue could be the training of different NN models for specific latitude belts which can be defined to cover the whole globe.

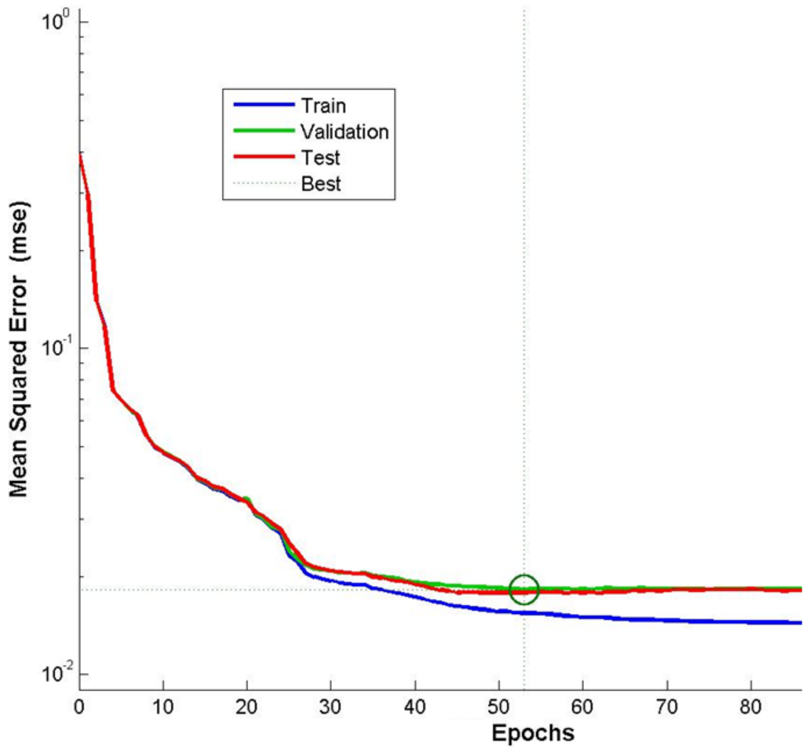
However, we inserted some comments in the discussion/conclusions dedicated to the uncertainties and limitations of the proposed model, as requested in point 4 also.

3- I would like to see the R^2 and RMSE of the neural networks during training, validation and test. The topology of the neural network model (large number of neurons in the hidden layer) and split of the training/validation/test might lead to overfitting. Besides, please add info about the training method.

In the text we added information about activation function, hardware and time needed for training the proposed model.

In Figure 4 we report the confusion matrix during for the validation set which is indicative of the model generalization capability of classification on data which have been not used for training and test the model. Training neural networks for classification problems the accuracy of the confusion matrix (90.9%) on the validation set can be considered as meaningful metric instead of the R^2, which is usually used mostly for regression problems.

Moreover, as the graph below shows, we avoided overfitting through the early stopping technique. The model used for classifying Raikoke SLSTR granules have been trained until epoch 53 where the minimum error on validation have been obtained (MSE = 0.0182).



4- There are no discussions on the uncertainty and the limitations of the presented model.

Some comments on the uncertainty and the limitations of the presented model have been added in sections 5 and 6. In summary:

1. the model transferability is significantly related to the spatial-temporal data availability for the generation of a training dataset which is statistically representative of all the possible scenarios;

2. lack of standard ground truth data for training and validation phases requires the BTD threshold selection by an operator which prevents the method from being fully objective.

Specific comments:

L32-34: this part is not precise. Ash is a part of tephra with D<2 mm. Then we have fine and very fine ash. Please revise.

We revised and changed as below:

"In general, from the start of the eruption, volcanic emissions are composed of a broad distribution of ash particles, ranging from very fine ash (particle diameters, d < 30 µm) increasing in size to tephra (airborne pyroclastic material) with diameters from 2 mm up to 64 mm. Larger fragments and are also generated which fall out quickly; these and ash with d> 30 µm are not considered in this paper. The gaseous part …"

L49: you mean $\Delta T_{11\mu m - 12\mu m}$?

Yes

L70-72: NNs are good for what they are trained for. Their transferability to other eruption at different altitudes and with different ash composition (optics) might be challenging. Please comment on this.

Yes, the model transferability might be challenging in case of different conditions. For example, the change of latitude has an impact on the characteristics of the atmosphere. At the same time different volcanoes emit different types of ash affecting the variability of the radiance values detected by the sensors. Therefore, the generation of an appropriate number of examples, which must be statistically representative of all the possible scenarios, to be included in the training dataset may represent a very difficult task. However, a possible approach could be the design of different neural networks, each associated with a specific scenario.

L159: What is the measure of accuracy? R^2?

The accuracy of the trained model on the MODIS validation dataset was 90.9% as reported in the confusion matrix in Figure 4. Using the proposed vicarious validation to evaluate the performance of the model on SLSTR data some metrics have been added to Table 4 and 5 (see also our reply to the final comment).

Regarding the R^2, please see the reply to the comment n.3.

L205: for consistency, use "meteorological clouds" in the whole manuscript.

Now we use always “meteorological clouds” in the whole text.

L226: this argument is too strong. See my previous comments.

According to the previous comments and replies, we rephrased that statement in the paper.

Tables 4 and 5: It is very difficult to make any quantitative conclusion from these tables. Use other quantitative measures like SAL.

We derived the following metrics to improve quantitative conclusions (added to Tables 4 and 5):

- Precision;
- Recall;
- F-measure;
- Accuracy.

Ref:

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010