Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC2
https://doi.org/10.5194/nhess-2021-270-RC2, 2021

Comment on nhess-2021-270

Anonymous Referee #2

Referee comment on "VADUGS: a neural network for the remote sensing of volcanic ash with MSG/SEVIRI trained with synthetic thermal satellite observations simulated with a radiative transfer model" by Luca Bugliaro et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2021-270-RC2, 2021

General comments:

In general, I think this is a very interesting paper on an important topic. The potential of neural networks for volcanic ash detection and retrieval is significant given the measurements that are known to provide information about the volcanic ash present, and the complexities of extracting that information given the influence of other variables. Additionally, as new instruments become available it is not unreasonable to expect neural networks to be able take advantage of the extra information. This paper describes a neural network that has been implemented well with the validation showing it is effective. Additionally, the method of using simulated imagery to train and validate the neural network is a sensible approach given the flaws with the different observations or models that have been used as the ‘truth’ to train NN’s in previous studies. Furthermore, this paper and the method described formed the basis for the successor, VACOS, which addresses some of the limitations of the methods described here. As stated by the previous reviewer, there are a number of sections, in particular relating to the production of the simulated observations where a lot of detail is provided. While this detail is critical for others to be able to reproduce the data, where possible detail that is also provided in references could be removed, and where this is not the case, the details could be moved to an appendix. Below I have included a number of minor points where I feel small modifications or additional explanation would be beneficial.

Specific comments:

- Title: Could ‘radiative transfer calculations’ be replaced by ‘simulated radiances’ or similar? To me there are multiple ways that radiative transfer calculations can and are used in volcanic ash retrieval and detection. The key use in this paper is to produce the simulated radiances for the training set. However, it is up to you which you think better summarises the more unique points in the paper.

- Abstract: Lines 8-14: Here the performance of the method of detection and retrieval is detailed. Can these values be compared to values using more conventional non neural network methods? There have been previous studies into this such as (Prata, A. J., and A. T. Prata (2012), Eyjafjallajökull volcanic ash concentrations determined using Spin Enhanced Visible and Infrared Imager measurements, J. Geophys. Res., 117, D00U23, doi:10.1029/2011JD016800.). These are primarily focused on the retrieval part rather than the detection. Most of these studies are with real data that have been validated against aircraft and/or Lidar data, which will be worth noting if comparing to validation completed against simulated imagery. For the detection part, could you provide some comparison with other leading volcanic ash detection algorithms, perhaps from one or both of the intercomparison workshops?
- Section 3.1.3, Lines 225-265: I feel this section is longer than necessary. It is not clear to me if the refractive indices used in this study were taken from work completed by others or if they are similar to those found by others but with small differences in method. If they are the same as those found in other published work, I think it would be better to reference the work with a statement saying followed the method of 'authors et al…'. Equally if the approach is similar to that of a previous study could the above approach be used but with the differences from that method stated. If the differences are significant could these be detailed in an Appendix.
- Lines 257-258: It seems a limitation of the training data to have used the refractive index from one eruption (at least in terms of the applicability to other eruptions). Given the dependency on refractive index and size distribution is strong it seems plausible that using values that are relatively well defined for Eyjafjallajökull may mean the scheme may be less accurate for other eruptions. I believe this limitation has been addressed in the successor version VACOS as mentioned in the conclusion. However, it would be good to state this here or perhaps when VACOS is first mentioned.
- Line 294: "radiative transfer simulations are always run for a set of 41 cosines of the viewing zenith angles from 0.2 (corresponding to ~ 78 viewing zenith angle) to 1.0 (nadir)". I am not a neural network expert, but could this be part of the reason the method performs best at moderate zenith angle? One might expect it to perform best at low (or zero) zenith angle where viewing conditions are most optimum? I wonder if running simulations for 41 cosines of viewing angles might cause the neural network to favour scenarios that perhaps aren't likely, as presumably certain scenarios in terms of the atmospheric conditions will be more likely to occur at particular zenith angles given the fact that higher latitudes will all have relatively high zenith angle.
- Lines 335-336: "To optimally cover different seasonal conditions we select day 15, 12UTC for 12 months from February 2010 to January 2011." What about impacts from different times of day, in particular inaccurate surface temperatures as the surface warms/cools? Again, this potential limitation may have been addressed in the successor and if so could be worth mentioning here.
- Lines 365-366: "Large VAMC > 5 gm^{-2} correspond to brightness temperatures between 260 and 280 K, thus corresponding to medium height levels up to approximately 5 km (see Fig. 4)." Does this mean there were no situations where VAMC > 5gm-2 occurred above 5km in the training set? This doesn't seem ideal given that high column loadings mostly occur close to the source when the plume is highest (often above 5km). If this is the case it would be good to explain in the paper why.
- Lines 372-373: "Of course, other BTDs, like the most used BTD(10.8-12), can be implicitely obtained by the NN through combination of the available ones." Would it have made any difference to have used 10.8-12 or does the fact they are implicitly obtainable to the NN make it irrelevant which specific BTD's are used? In Kox et al 2014 it is suggested that the speed and accuracy of the neural network is dependent on the specific BT's/BTD's used. Could you explain why that may not be the case here or perhaps mention that some improvement could be found through use of different BTD's.
- Lines 384-385: "Surface emissivity is neglected here since its variability is not very strong." Ashpole et al (2012)

https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2011JD016845 suggested that spatial variability in the surface emissivity can cause issues in the detection of dust (and by extension ash). Do you have a reference or could you explain why the variability is not considered sufficient to be significant here?

- Lines 451-452: “With the help of this additional filtering, the overall POD sinks to 0.84 but the FAR also sinks to 0.05.” Is this an improvement on 0.92 and 0.08? Could you explain why reducing FAR appears to have been prioritised over maintaining POD? As mentioned earlier, if this scheme was used in either of the intercomparisons it would be good to discuss its relative performance.
- Lines 553-554: “The distribution of VADUGS-retrieved VATH peaks at 0 km and has a flank reaching up to 19 km, with a notable minor peak at about 9 to 12 km.” Do you have any thoughts why the main peak is around 0km? This kind of ash should be hard to detect. Is this a result of there being little training data very close to 0km or perhaps there being lots of training data between 0 and 5 km? Could you explain why the training data appears to sharply reduce in frequency above 14km? Large eruptions can emit ash higher than this, although if the focus of the scheme is on tropospheric ash then limiting the training data above the tropopause may be wise. Also it may be worth mentioning that the general underestimation of VATH is likely related to (or perhaps the cause of) the typical overestimations in VAMC.
- Figure 7: Could the gaps at certain height levels (i.e every 3km) be explained?

Technical corrections:

- Line 131: “are all points correctly classified as ash and the number of false negatives, NFN, all missed ash clouds (see Tab. 1).” I think either a comma after the first use of ‘ash’ above would help or ‘NFN, *are* all missed ash clouds.’
- Line 136-137: “are all points falsely classified as ash, i.e. the false alarms and the number of true positives, NTN, all points correctly identified as ash free.” As above and also ‘true positives’ should be ‘true negatives’.
- Line 372: ‘implicitely’ should be ‘implicitly’
- Line 375: ‘channels’ should be ‘channel’
- Line 380: ‘induce’ should be ‘deduce’?
- Line 429: ‘For each’ I think a comma after each would improve readability
- Line 430: ‘profiles’ should be ‘profile’
- Line 543: ‘bahaviour’ should be ‘behaviour’
- Lines 621: ‘been’ should be ‘be’