Review for egusphere-2024-1336

Indications of authors: answers to each comment are in blue. Text in quotes shows portions of the modified text. Underlined text shows the modifications.

The authors would like to thank the reviewer for his/her thorough analysis of the paper that will definitely increase its quality. We hope that readers will find the answers useful.

The study aims to analyse the performance of convolutional neural networks (CNN) in discriminating between severe (> 2cm) and non-severe (< 2cm) hail cases. Different data sources are used as hail/non-hail observations (ESWD reports, ANELFA hailpads, user reports from MeteoFrance mobile app). A comprehensive pre-selection and quality control is performed on these data to construct training, validation and test dataset. Three CNN architectures are trained using radar-derived input features on images of different input sizes. The input features comprise (polarimetric) radar data as well as radar-based hail proxies. The performance of the trained CNNs in distinguishing between severe hail and rain/small hail events is compared against hail proxies. In addition, feature selection and feature importance are discussed comprehensively. It turns out, that Maximum Estimated Size of Hail (MESH) is the most important input feature of the CNN. The CNN is able to outperform all reference proxies for different verification metrics. However, the study also shows that the discussed hail proxies are able to achieve a similar performance compared to that of the CNN, if they are adjusted/tuned regarding value threshold and area threshold.

The study is very comprehensive and contains interesting results. The results are clearly and comprehensible presented. I appreciate the wise selection and filtering of reference data for severe and non-severe hail events. The approach is well explained and discussed. Also, the CNN model architecture selection and the analysis on feature importance sounds very reasonable. I strongly recommend the publication of the study. Some minor revisions that are proposed below could further improve the paper.

General comments

The introduction gives a detailed overview on hail detection using remote sensing data, but the introduction on hail detection by in-situ measurements or eye-observations and the related issues (representativity, sensitivity on e.g. population density or time of day, …) is somehow missing. That's unfortunate, as these aspects are well discussed in Section 2. In addition, the specific formulation of research questions could be beneficial to outline the story of the paper. For the hail detection techniques in the introduction, the authors wanted to limit themselves to radar-based hail detection techniques for the sake of simplicity. Concerning the formulation of specific scientific questions, the following paragraph has been implemented at the end of the introduction:

"[...] knowledge, none have attempted to use radar polarimetric variables for severe hail detection with CNNs. <u>How do CNNs perform on the task of severe hail detection when applied to polarimetric radar data? Can CNNs outperform existing hail proxies? Can CNNs be used to extract information relevant to the detection of severe hail?</u> To answer these questions, <u>[...]"</u>

Section 2 is quite extensive. Potentially, the discussion on storm mode (sect. 2.5) can be moved to the appendix since it does not contribute to the main storyline of the paper.

The length argument being shared by other reviewers, the following parts have been moved to Appendix to improve the readability of the article: the description of the “second” cell identification algorithm and the storm-mode assessment.

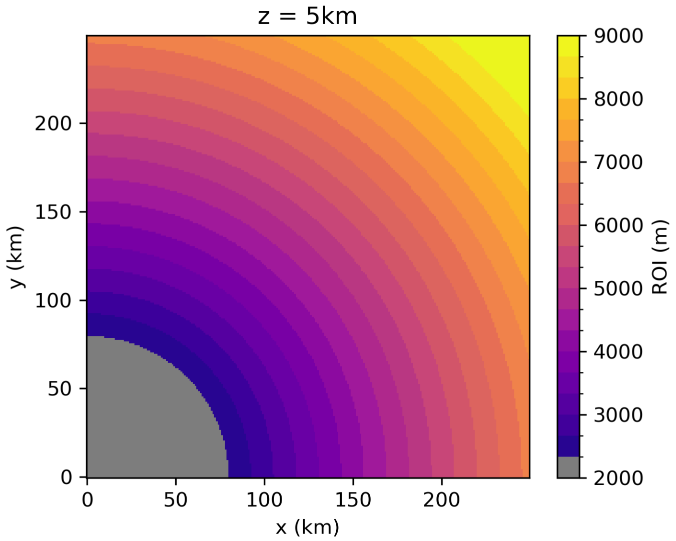
In the conclusions (sect. 5), I would appreciate a more critical discussion on the (operational) applicability of the new CNN approach also with respect to its complexity compared to the much simpler hail proxies. This is a general discussion on the costs and benefits of AI systems in Nowcasting that occurs frequently. Details have been added in the conclusion as follows:

“[...] is recommended to examine cells that have produced reflectivities of at least 45 dBZ. The cell-identification algorithm and the production of input features for the CNN may require a greater investment of computational time and resources than existing hail proxies. The necessary 3D interpolation can be particularly costly. However, this additional computational time can be offset in real-time by the cell-identification algorithm. The input features can be generated [...]”

Minor comments

- ZDR-columns (Snyder et al., 2015) shall be introduced in Section 1 as a well-known precursor of hail. Section 1 (Introduction) introduces known radar-based hail detection techniques in real time. Despite being a known precursor for hail, the Zdr column, particularly the relation between their height and width with hail occurrence, is still being studied and no systematic algorithm nor extensive validation has been made so far for severe hail, to the authors’ knowledge. The authors would like to limit their introduction to established hail detection techniques to remain shorter and clearer. Zdr columns and their predictive skill are anyway discussed later in the paper as they are used as an input feature.
- The NWP perspective (i.e. environmental conditions) on hail/hail size forecasting could be shortly addressed in the introduction (e.g. Battaglioli et al., 2023). The following text has been added in the introduction:

  “Other studies have employed deep learning and machine learning techniques, applied exclusively to environmental variables derived from numerical weather prediction models (NWP), for the purpose of analysing or forecasting hailstorm environments (Gagne et al. 2017; Gagne et al. 2019, Battaglioli et al. 2023).”
- The details on the interpolation of polarimetric radar data on a 3D regular grid, particularly the discussion on the ROI, remains unclear.
  The Figure below is derived from the PyArt’s formulation of the ROI. The example given is computed at an altitude of z=5km. The radar’s location is at (0, 0). The minimum radius is 2000m.



Formulas to obtain the figure are available in the Py-ART's documentation:
https://arm-doe.github.io/pyart/_modules/pyart/map/grid_mapper.html#example_roi_func_dist_beam

- The notation "polarimetric grid" is very confusing (it's a regular grid 3D grid, not in polar coordinates?). Polarimetric grid is a 3D cartesian grid with polarimetric variables (and Zh). It has been renamed as follows:

  "The Z_DR column height was calculated using the 3D Cartesian polarimetric grid"

- Line 234: double "the". It has been corrected.

References

Snyder, J. C., A. V. Ryzhkov, M. R. Kumjian, A. P. Khain, and J. Picca, 2015: A ZDR Column Detection Algorithm to Examine Convective Storm Updrafts. Wea. Forecasting, 30, 1819–1844, https://doi.org/10.1175/WAF-D-15-0068.1.

Battaglioli, Francesco & Groenemeijer, Pieter & Tsonevsky, Ivan & Púčik, Tomáš. (2023). Forecasting large hail and lightning using additive logistic regression models and the ECMWF reforecasts. Natural Hazards and Earth System Sciences. 23. 3651-3669. 10.5194/nhess-23-3651-2023.