

Response to reviewer 1

Title: Explainable AI shows that a neural network learns extratropical cyclones as predictors of heavy precipitation

Authors: Robin-Guillaume-Castel, Camille Li, Stefan Sobolowski

Recommendation: revision

Summary: In the current version the manuscript is a method paper focusing on the introduction of a novel explainable AI (XAI) method that facilitates the application of XAI for scientific discovery. To this end the authors combined a post-hoc attribution-based XAI method (LRP) with a large scale spatial field analysis to identify impact of atmospheric objects. Building on a clearly defined case study of cyclone impact on Norwegian strong precipitation events, the authors construct a CNN architecture that predicts a heavy precipitation event label for the following seven days (0 or 1 for each day) based on several atmospheric variable fields of cyclone-specific inputs. Across several experiments the author demonstrate that the network relies on the established cyclone patterns and driver.

Quality:

The manuscript is overall well-written and the work highly interesting. The authors backen their claims with extensive results and relevant, straight forward figures. While I find this work very relevant, I am confused by the overall storyline. In my opinion in the current framing the study remains weak, as it attempts to simultaneously introduce a new method and deliver a general proof for the usability of AI, performing only a single case study, which uses an outdated network architecture that ultimately proves the same point already established in several publications since 2020. This seriously hampers the novelty and weakens the value of the experiments. Thus, I strongly recommend reframing the paper into one of two clear directions: Option A, frame it strictly as a scientific discovery paper on cyclone-caused precipitation using AI and XAI methods, this might not need much more work and underlines more strongly the extensive work already done in this publication. In this case, it would suffice to discuss the potential other applications of their method in the outlook. Option B is to frame it as a dedicated method paper, which requires adding at least one more case study, as a single regional study is not sufficiently representative to validate a novel pipeline. For option b, the experiments demonstrate that the proposed explanation technique is useful, it would be essential, however, to show how generalizable this success actually is across other applications. I recommend using one of the trained foundation models as these are operational and make for a stronger case.

We thank the reviewer for taking the time to provide this useful review.

We appreciate the positive feedback and the comments, which will help us clarify the framing of the current manuscript. We agree that the framing was not clear. The paper was always intended to be about the potential of XAI for scientific discovery (along the lines of option A), but we see that we need to do a better job of emphasizing that the main novelty of our paper is the use of weather features to understand and interpret explainability results. Mentions about improving the general trust of AI with an object-based XAI framework will be limited to the discussion in the revised version.

Major Comments:

1. Introduction: While the introduction is well-written, includes timely and relevant citations and guides the reader into the topic. I already struggle with the outlined storyline, which in my opinion, fails to focus on the novelty of this work. The authors aim on the one hand to deliver further proof on the general concept of deep learning and XAI for scientific discovery while also aiming to introduce a new methodology. Such a general proof would require more than one case study and a more timely network architecture and in addition lacks novelty since it has been discussed in many publications (I suggest checking E. A Barnes papers and overview publications such as <https://www.nature.com/articles/s41467-025-56573-8>) since the mentioned work from 2020. Similarly, I would argue that to introduce and validate a new XAI methodology a second case study would be helpful. Therefore, I suggest reframing the paper in terms of a pipeline that enables scientific discovery with physical prediction validation as a crucial step.

As mentioned above, we should indeed clarify the purpose of this study. A number of changes will be implemented to ensure that the focus of the paper is on scientific discovery rather than trust in neural networks. We highlight that our results show how to interpret relevance maps when the drivers of the predictions are transient and mobile objects (i.e. cyclones in our case). The new introduction will be rewritten to emphasize the following points. Currently, some studies have highlighted that transient/mobile drivers such as weather features can be found in XAI *relevance maps*, but these have been manual and not done at scale. There is no current approach that we know of to systematically interpret relevance maps when predictions are driven by moving objects. In our paper, we explore how external tracked weather features can help us understand the neural network internal decision process. We highlight how western Norway is an ideal case to develop this approach.

To focus on how our approach gives more insight than simple composites, we compare our results to the mean relevance map for positive events (i.e. days when there is a heavy rainfall event in western Norway) (see draft figure D1). This comparison shows that, while simple composites can highlight that the neural network puts some importance on the North Atlantic storm track, it fails to show that it focuses on specific cyclones.

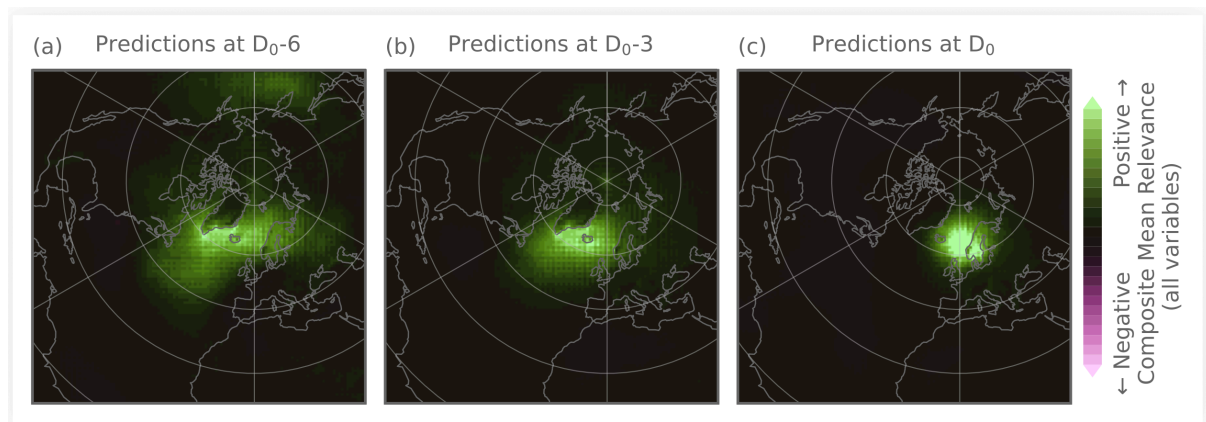


Figure D1: Composite (simple averages) of all our relevance maps (only for positive events).

- Line 26-29: I agree with authors general statement but I believe that this should be a given that forecast validation should be first. I don't understand why this discussion is blown out of proportioner deep learning, since analysis based on numerical model data requires the same validation step, i.e identifying if a certain finding is part of the model bias or due to underlying physics.2. 3. 4.

We agree with the reviewer here, and this part will be removed, especially since it does not fit well within the scientific discovery framework.

- Line 43 -44 (Methodology): I think there are more paper that came after Toms, therefore I suggest to write "works like Toms et al. 2020". Also I am not sure if that is really the line of argument I would make or a narrative I would pursue. I see the novelty of this paper more in the findings and the way the authors leverage XAI which unfortunately I feel is not established clearly, underselling the work's novelty.

We agree with the reviewer. The claim that we directly build from this paper was not needed and contributed to the unclear aims of the original manuscript. It will be removed.

Methods & Data: I congratulate the author's to a thorough well described methodology section and well-designed dataset and model. However, under the current framing, this design feels untimely and extremely limited, both the network choice and the applied XAI methodology are very simplistic and similar studies could have been performed on more elaborate architectures such as Aurora or GraphCast which would have better supported the claimed purpose of the work and would have delivered better generalisability of the results across comparable approaches. In the current state you only learn about the skill of the designed architecture and the applicability of your XAI approach to the given architecture. I believe, however that this study, in the current state does carry novelty and purpose which is why I do not strongly recommend adding more experiments.

We thank the reviewer for the positive feedback. The choices described in this section (e.g., the very simple network choice) will make more sense now when focusing our framing on scientific discovery.

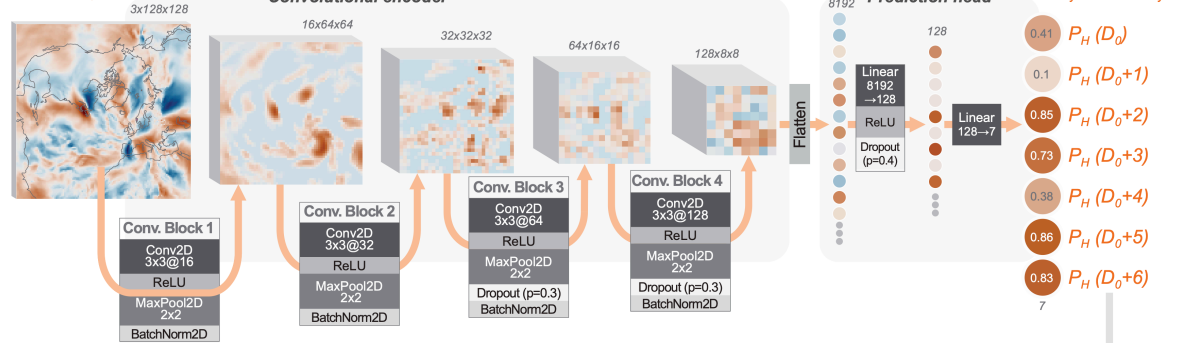
- **Section 2.3.1:** I suggest improving the description of the network and the corresponding figure, integrating visualisations of concrete input and output examples (Figure 2) this would better indicate what each output node corresponds to.

We will make an updated figure of the neural network with concrete input and output values for better understanding the actual processing of the neural network. See below for a draft of how this will look (Draft Figure D2).

(a) Neural network architecture and inference example

Input features

Anomalies of U850, V850
and SLP at D_0



(b) Example of obtention of relevance maps using Layer-wise Relevance Propagation

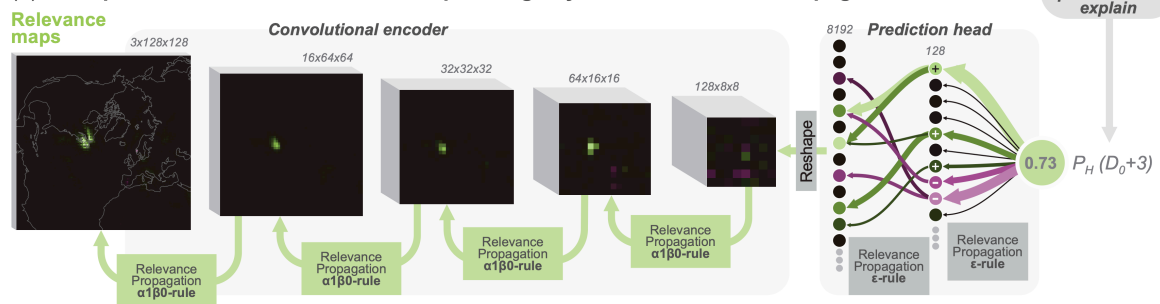


Figure D2: a) Updated figure showing the neural network architecture, along with inputs/outputs concrete example. b) new subfigure showing how relevance maps are obtained from the prediction and the model architecture.

- **Section 3:** Similar to the network section the XAI framework would profit from an illustrative visualisation of the methodology especially since this is one of the main novelties and contributions of this paper.

The updated figure on the neural network will include a panel showing how relevance maps are obtained from the predictions and the network (draft Figure D2.b). We will also provide an updated figure detailing how the neural network outputs are associated with cyclones through the relevance maps (see draft Figure D3).

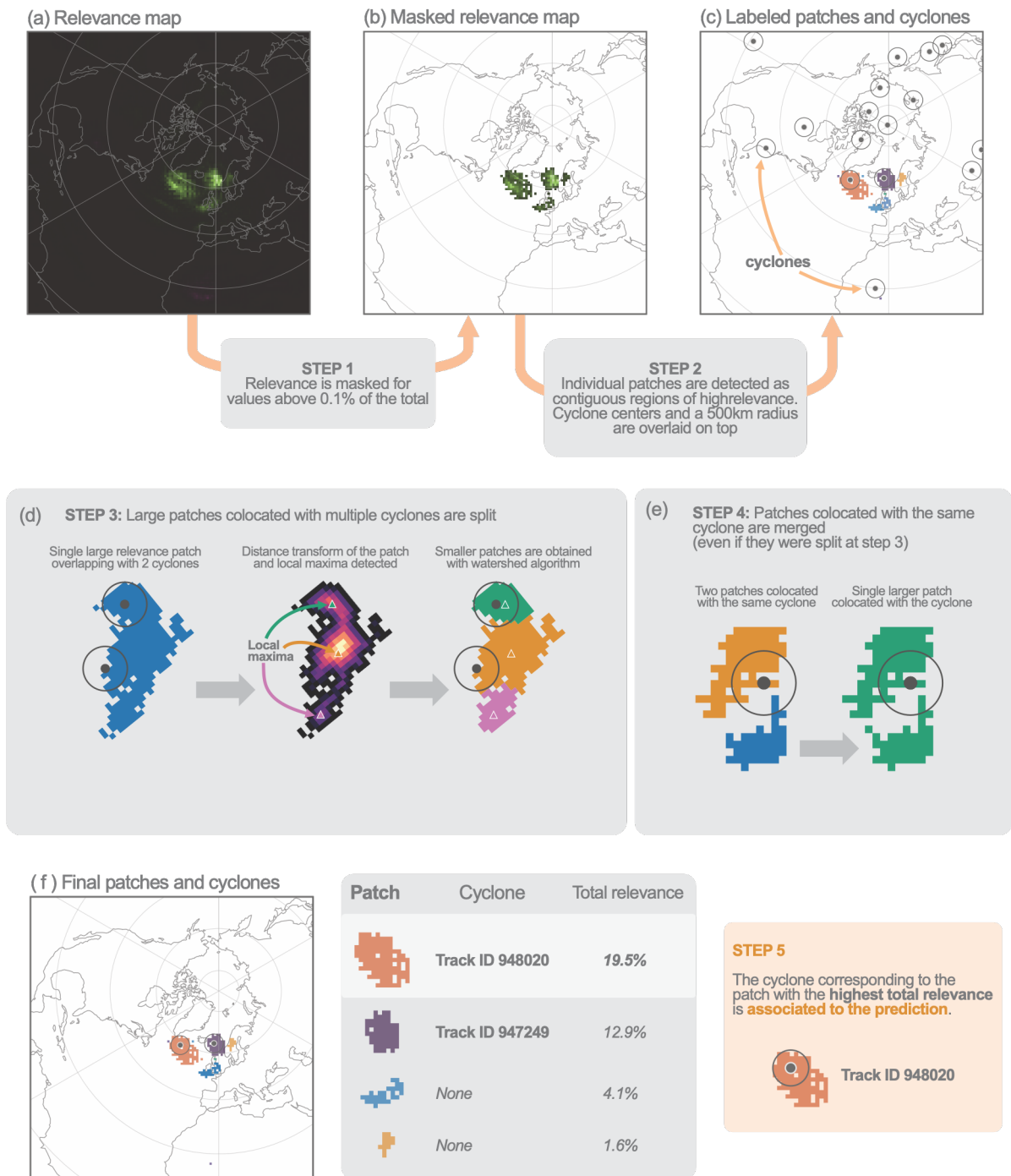


Figure D3: Updated methodology figure detailing our workflow. From one relevance map, we extract contiguous areas of high relevance and associate them with cyclones.

Experiments & Results: The authors already present extensive results and this only requires reshaping according to the reframing.

We thank the reviewer for the feedback. The results section will be reshaped accordingly. In addition, there will be some changes to what material appears in methods versus results based on revisions made in response to reviewer 2's comments (see response to their main comment #2). Notably, the neural network predictive skill assessment will now be included in the results.

Discussion: This section would of course need the same re-writing to change the papers

focus. In addition, I recommend to make it one concise section with paragraphs rather than subsections (but I leave this up to the authors). I recommend reducing the last subsection specifically or integrating it into limitations.=

Following our choice to focus on scientific discovery, we will reshape the last section (discussion and concluding remarks) accordingly. We will highlight that our main contributions consist of interpreting and analysing XAI results with weather features and we will discuss the limitations and potential future uses of this approach.

- **LRP**: there are multiple extensions of LRP for almost any novel architecture type. While it might require some engineering most foundation models in weather and climate should be in fact covered (see G. Montavon Publications)

We thank the reviewer for this comment, and we will mention in the discussion that LRP could be used even on the more novel architectures for a potential extension of this work. We have actually started to work with these state-of-the-art models for future work, but we could not find a ready-to-use python library including LRP some layers of these models, which would explain our misconception.

Corrections:

Line 19-20: I would add more citations here.

We will add citations to the following works:

- Dewitte et al., 2021: Artificial Intelligence Revolutionises Weather Forecast, Climate Monitoring and Decadal Prediction, Remote Sensing, 13, 3209, <https://doi.org/10.3390/rs13163209>,
- Bi et al., 2022: Pangu-Weather: A 3D High-Resolution Model for Fast and Accurate Global Weather Forecast, <https://doi.org/10.48550/arXiv.2211.02556>, 2022.
- Lang et al., 2024: AIFS – ECMWF's data-driven forecasting system, <https://doi.org/10.48550/arXiv.2406.01465>, 2024.
- Bracco et al., 2025: Machine learning for the physics of climate, Nature Reviews Physics, 7, 6–20, 2025.
- Kent et al., 2025: Skilful global seasonal
- predictions from a machine learning weather model trained on reanalysis data, npj Climate and Atmospheric Science, 8, 314, <https://doi.org/10.1038/s41612-025-01198-3>, 2025
- Haji-Aghajany et al., 2025: Beyond the horizon: A comprehensive analysis of artificial intelligence-based weather forecasting models, Engineering Applications of Artificial Intelligence, 162, 112 335, <https://doi.org/10.1016/j.engappai.2025.112335>, 2025.

Line 24-25: I recommend ordering the citations by year.

Thanks, this will be done

Line 90: “is” to “are”

This will be corrected