Review - Copernicus

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 focussing 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a new method and deliver a general proof for the usability of AI, performing only a single case study, which uses an outdated network architecture that ultimately proves the same point already established in several publications since 2020.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.

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.
   - *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.

- 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.

2. *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.
    - **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.
    - **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.
3. *Experiments & Results: T*he authors already present extensive results and this only requires reshaping according to the reframing.
4. *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.
    - **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)

Corrections:
Line 19-20: I would add more citations here.
Line 24-25: I recommend ordering the citations by year.
Line 90: “is” to “are”