We thank reviewer 1 for their feedback and comments. In this document, we provide answers for their questions.

Reviewer 1

The paper is interesting. It extends an approach developed by Krouma et al. (2024) to wind and precipitation forecasts.

Major comments

The reason why precipitation and wind speed are treated differently (with different versions of the SWG) is unclear and should be explained. This seems to add an unecessary complication to the manuscript.

⇒ **We thank the reviewer for this comment. Precipitation and wind speed are treated with different SWGs because the HC-SWG was specifically developed and validated for precipitation and then extreme precipitation, using Z500 as a predictor whose relevance has been extensively demonstrated in previous work (Krouma et al., 2022, 2024).**

In contrast, extreme wind speed is driven by different physical mechanisms and requires different large-scale predictors. We tested forecasts of wind speed using one single atmospheric variable and they did not work. We now specify this in the manuscript. Extending the HC-SWG framework to jointly handle multiple predictors and ensemble reforecasts for wind would significantly increase the model complexity without clear added value. For this reason, we developed a dedicated SWG for wind speed, better tailored to its dynamics. This distinction and its rationale are clarified in the manuscript (L.54-58, p.2-3).

A greater care should be used in the figures. Figure 2 mentions "Santander" on the right column, while it should be "Linkoping". The colors for Fig. 3 and 4 (among others) should be consistent across stations. Figure D1 is unreadable. What are the colors for? Figure E2 is very hard to read.

⇒ **Thank you for mentioning that, we corrected the legend for Figure 2. We verified the consistency of the colors among all figures (Fig 3, 4, 8 and 9). We changed the colors for Figure D1 for better readability, and now explain that they provide a visual illustration of the value of the Brier scores.**

Who is likely to use such systems, which seem to require a complex layer of simulations on top of already complex datasets of ensemble forecasts?

⇒ **This is an interesting question. Such systems are primarily intended for operational forecasting centres and risk management agencies. Several of these have the technical competence needed to implement approaches such**

as the ones we propose, and would then be able to provide the results to local stakeholders involved in impact-based decision making (e.g., emergency managers, infrastructure operators), who in many cases may lack the ability to implement the approaches themselves. By providing calibrated probabilistic information on extremes and compound events at specific locations, the framework can support early warning systems and local adaptation planning, complementing existing ensemble forecast products. We added this in the introduction (l. 45-48 p.2)

Specific comments

Why are the procedures described in Secs. 3.1.1 and 3.1.2 relevant to extremes?

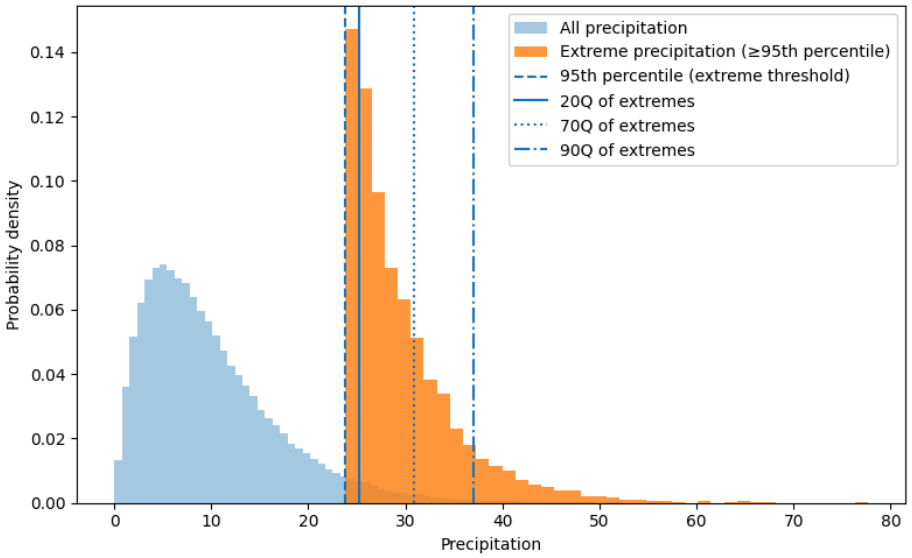
⇒ **The described procedures in Secs 3.1.1 and 3.1.2 explain how each SWG works and is used to forecast each variable. For example for the HC-SWG, we use the whole ensemble reforecast of the ECMWF to find good analogs, which was specifically useful for the extreme precipitation also using different lead time of the reforecast. Similarly, for the wind extremes we explain the inclusion of different atmospheric variables in the SWG instead of one. We included these descriptions to ensure reproducibility of our analysis and to make the study self-contained.**

Please provide a map of the 9 stations (the reader might not be so familiar with European geography).

⇒ **A map showing the 9 stations has been added, in Appendix F.**

What do 20Q, 70Q or 90Q mean? This is not stated in the text. If 20Q is the 20th quantile of precipitation, this cannot be considered as extreme.

⇒ **Apologies for the confusion, but what we do is that we consider the quantiles of the extremes (and that's what 20Q, 70Q and 90Q refer to) and not the quantile of the full distribution as shown in this example below.**



So the quantiles are computed conditionally on extreme precipitation, rather than from the full precipitation distribution. Extreme precipitation is first defined as values exceeding the 95th percentile of daily precipitation at each station. The resulting subset therefore represents only the upper tail of the precipitation distribution. The 20Q, 70Q, and 90Q are then calculated *within this conditional distribution of extreme precipitation* with lower quantiles corresponding to relatively moderate extremes and higher quantiles highlighting the most intense precipitation events. Importantly, these quantiles should be interpreted as percentiles of the extreme precipitation distribution and not as percentiles of all precipitation values. We explained this better in the text (l.184-190, p.8).

The "outperform" verb looks like an exageration. Since the SWGs are based on the forecast products, the least that can be achieved is an improved performance. The improvement for extremes is not clearly outlined.

⇒ We agree with the reviewer that the term “outperform” may be too strong in particular for the HC-SWG as it is built upon ensemble reforecast products. The HC-SWG role is to improve forecast and representation of extremes rather than to fundamentally outperform the underlying forecasts. We therefore revised the wording to “shows improvement against” or “improving forecast skill compared to” to clarify the added value of the SWGs.