EGU-2023-3111

Probabilistic short-range forecasts of high precipitation events: optimal decision thresholds and predictability limits

Bouttier and Marchal

Overall:

The paper presents an interesting approach to deriving optimal decision thresholds for probabilistic ensemble forecasts for two users with different risk appetites and requirements. It provides a means of moving the decision-making process upstream, away from the user, back into the meteorological sphere/space, where the expertise, in terms of interpretation is more likely to exist.

Whilst some in the community may be horrified at collapsing a probability forecast down to a deterministic one, it is what the user does, and as the authors point out, having a sense of where the ensemble adds value over the deterministic is important because it is not a given that the ensemble will add value. In the end most users will either have to choose to do something or nothing. This is a binary decision point.

The lack of skill is clear and potentially disheartening. However, the study provides some tangible evidence of the current state-of-the-art capability in our NWP systems and how to maximise the value they can provide.

I recommend the paper is accepted with some minor revisions.

Specific comments:

Please check the figure numbering and text where the figures are referred to. I did think there are a number of places where things are either not referred to or not referring to the correct figure.

L113-115: I rather like how you have worded this and would like to see you use something similar in the abstract. It would strengthen the abstract in my view.

L179: Perhaps more a comment than anything else, but I would be very wary of applying a neighbourhood verification method on top of using a neighbourhood post-processing approach first.

L186 and L191: the 4- and 30-mm thresholds appear somewhat at random here. Is there any rationale for these values?

L219: I think you meant that it gives 4 times more weight to c than to b?

L220: I would replace non-detections with missed events

L260: I feel this “only” is misplaced given that $0.12 < 0.3$

L271: missing bracket

L274: I think you mean Fig 3 here

L295: great illustration of the value of the FDR over POFD

L317: I am not that familiar with the F2 score but we know that the ETS can be optimised by over-forecasting. It looks like the F2 can as well. Over-forecasting is by far the most useful mechanism for reducing missed events (of course at the expense of the false alarms!)

L320: This is probably an important result and one worthy of elevating more in the abstract perhaps? For user H the ensemble probably has more value, even if the scores are lower.

L332: This would be consistent with the notion that whilst spread is necessary, too much spread is not good.

Figs 5, 6 and 7 are very small. The final version should have bigger panels. These figures contain some important findings.

L343: What Fig 5b shows really well is how the use of some sort of neighbourhood improves the score.

L346-349: You'd hope that the ensemble (especially) can capture some of that spatial uncertainty. Not all heavy rainfall events are convective. Perhaps it would be better to say, that if 30 mm/6h are convection-driven, these come with large spatial uncertainties. To say that location errors affect user H more than L, is perhaps a little disingenuous because user H is a subset of user L. I strongly suspect that if user L was hit by the that 30 mm/6h they would still be affected by it!

L365-368: This is the double penalty effect in action and why we use upscaling.

L383: it might be useful to see what N is here for each of the thresholds

L417: suggest longest instead of highest

L435: What it does suggest is that prolonged periods without rain/events can be detrimental to aggregated statistics. It is therefore not the length of time that is important but the number of events in that time window.

Fig 10. I am not entirely sure I like the highest totals to the left. It is somewhat counterintuitive.

L501: I guess what this example tells me is that the basic characteristics of the forecast (ensemble or deterministic) are still very much fundamental to the success of applying a user-relevant decision threshold. The outcome will only be as good as the forecast itself.

L516: This is a crucial point. Many countries are constrained in not issuing warning areas which are too large. Therefore, having specific and tight warning areas is highly desirable for many met services.

L528-531: Perhaps one of the good things about this way of post-processing the forecasts is that it does provide this alternative "views" of the same forecast.