

Reviewer Comment

Thank you for the detailed response and the substantial revisions. I think the manuscript has improved considerably, and the novelty and objective of the study are now much clearer. In particular, the response explains well why the inundation analysis adds value to understanding forecast failure, and why an end-to-end assessment of the NOAA operational forecasting chain for a single high-consequence event is a useful and novel contribution. I have a few remaining comments, mainly related to terminology and figure clarity.

Main point 1 regarding novelty of the study

The authors have generally addressed this point well. In particular, the response now more clearly explains why the inundation analysis materially improves the understanding of forecast failure, and why an end-to-end assessment of the NOAA operational forecasting chain for a single high-consequence event is novel. The additional insights gained from evaluating flood impacts are now explained much more clearly. I particularly appreciate the proposed clarification: “We now state explicitly that two locations may have similar CSI but very different building-impact counts depending on whether the spatial errors fall in populated areas, and that impact-based metrics provide complementary information to extent-based skill scores.” However, I could not identify the corresponding revised text in the manuscript. I recommend either adding this clarification to the manuscript or indicating where it has been included.

Response: Thank you for your comment. Based on your suggestion, we have revised the manuscript and added this sentence in section-4, Line 312.

“From the FIM evaluation, we also found that two locations, such as in Hunt and Kerrville, can have similar CSI scores while experiencing different levels of flood impacts. This indicates the need to combine skill-based and impact-based assessments to better quantify the spatial errors in FIM.”

Main point 2 regarding drivers of impact

The authors have responded well to this comment. I only suggest refining the wording slightly, as “components” is somewhat vague in the current formulation. For example, “components of risk reduction” or “components of effective warning response” may be clearer, depending on the intended meaning.

Response: Thank you for your comment. We have updated this sentence in line number 351.

Figures

Figure 2: The figure has improved, but I think my earlier comment on the colour scheme may have been misinterpreted. By an increasing colour scheme, I meant that each SRF could still have a distinct colour, while the colour intensity would intuitively indicate the development of the forecasts over time. Alternatively, fully distinct colours, as used previously, would also work, but previously the same colours were used for different SRFs. However, the current colour scheme is difficult to interpret. I also suggest showing the x-axis ticks and labels only in panels d and e. Also, currently an empty legend is visible in the top left of every panel.

Response: Thank you for your comment. We have updated Figure-2 in the revised manuscript.

Figure 3: The purple circles now appear to be centred, and the buildings are much more clearly visible. However, the station names seem to be missing. I am also unsure what “Buildings Hit” refers to. If this indicates buildings hit under the forecasted flood extent, then the current blue colour works well because it resembles the forecast flood extent. However, if it refers to buildings hit in the benchmark analysis, using the same blue colour as the forecasted flood extent may be confusing. I leave the choice of colours to the authors, but I recommend making the distinction between the different building-impact values clear in the caption.

Response: Thank you for your comment. We have updated the caption for Figure 3 to clarify the building-impact information. The blue text in the lower-left corner of each panel reports the number of buildings intersected by the benchmark flood extent (Buildings Hit BM) and the forecast flood extent (Buildings Hit), as clarified in the revised caption. We tried to change the color of "Buildings Hit BM" to red to match the color of the benchmark flood maps in the figure, but this made the text much less visible, so we kept the original text color.

Minor point

I appreciate the revised terminology distinguishing skill-based, pixel-based, and impact-based metrics. I assume these terms are clearly defined when they are first introduced in the manuscript.

Response: Thank you for your comment. We are now consistently using Skill-based and Impact based evaluation terminology. We have incorporated this sentence in the revised manuscript. Please refer to line number 245 in the revised manuscript

“We used both skill-based and impact-based assessments for FIM evaluation. Skill-based evaluation primarily assesses FIM performance using categorical metrics, whereas impact-based assessment focuses on affected infrastructure, such as buildings and road networks, within the mapped flood extent.”

Once these final minor points have been addressed, I would be supportive of publication.