Response to RC1:

This is an important issue - not just for users, but for professionals formulating warning messages. Anchoring can be positive, e.g. if the anchor line is set at an agreed response threshold e.g. reasonable worst case, or for a specific cost/loss ratio, but it can also be negative especially in a low probability high impact situation, if highlighting of the median leads to underestimation of risk. Thanks for your comment.

Line 84: It should be noted that the sample is not representative of typical users. I don't believe this undermines the results in any way, but a comment on the differences between the sample and typical real life users might be worth including. We have noted that these participants are not necessarily representative of all populations

Line 105-6: The meaning of the three probability levels is not clear. I believe it may mean that there were three forecast scenarios presented with data adjusted to give 30%, 50% and 70% probabilities of exceeding 1 metre, and that each was presented in 7 different ways. However, I am still not sure if that is a correct interpretation. There is no figure that shows what the three scenarios look like (maybe the deterministic presentations would clearly display them?). I think it would also be helpful to include a brief description of how the 3 scenarios were created. Your interpretation is correct. We have reworded to make this clearer. For conciseness, we have not included depictions of the 50% and 70% probabilities but have included a description for how these were created (we used the spaghetti plot as a base and had 3, 5, and 7 forecast lines above the 1-meter mark, respectively. Then we converted the spaghetti plot into the other plot types).

Line 131: The authors should not assume that all readers will be familiar with the fit statistics for the ANOVA or Bonferroni tests. I suggest adding a short section 2.3 to introduce the statistical tests used and the meaning of the fit statistics (supplemented by a suitable reference). Thanks for this comment. We will add an explanation of the statistics used in the updated paper.

Line 179: The explanation of the reason that good interpretations were made of the spaghetti plots despite even less attention to the key than for the boxplots may be correct, but it runs counter to the general view that spaghetti plots should be avoided because they are difficult to interpret (due to crossing lines etc). I think this view should be acknowledged and responded to. We have cited a paper that found that climate information including individual model estimates helps user interpretation of data, which is counter to the general view you cite, however, we are happy to include reference to this in the manuscript

Line 189: This paragraph does not tell us what the impact on answer time actually is, merely what the statistics are of the differences. We should not have to look at the figure to work this out. We have updated the manuscript to make sure that the direction of the effects reported is stated explicitly in the text.

Line 198: a comment on the increase in time to estimate the maximum in a spaghetti plot with median is needed. This analysis was conducted combining all forecast representations together, not conducted separately for each forecast representation, like in section 3.2. Therefore, a comment specifically on the spaghetti plot would not be appropriate.

Line 211: This line starts with "However", which seems out of place given that the previous paragraph was also talking about the dangers of anchoring. Thanks for pointing this out. We've amended the manuscript to remove the "however".