Response to comments from Reviewer 1

1.1 I found this manuscript to be of high quality, interesting and very well written. In particular, I find the figures very clear and of high quality. I have a few minor comments that I would like to see addressed, but this should not require a lot of time.

We thank Reviewer 1 for their encouraging feedback, and for the valuable suggestions that will improve our paper.

1.2 When I first started reading the manuscript, including the title, I was very confused by what you refer to when you talk about a "model". To me, and I think for most hydrologists, our first reflex is to think about the hydrological model (GR4J in your case). It later became clear that you were most of the time referring to the post-processing method, but I was still bothered throughout the paper by the fact that it could be clearer. For one thing, I would suggest changing the title to indicate that you are interested by comparing two post-processing methods.

We agree that our use of “model”, when referring to “post-processing model”, was too vague. In particular, the phrase “streamflow model” in the title of the paper was a poor choice. We apologise for the confusion this caused.

There are several models used in the work, including the deterministic model, the residual model, etc. The usage of "model" in the original manuscript was a bit loose and we can see how this caused confusion.

For the revised paper we will

i. Clarify specifically which models we refer to in specific sentences, to avoid ambiguity/ confusion
ii. Change the title to clearly indicate the study focuses on streamflow forecasting and compares two post-processing methods

1.3 Second, I would avoid the use of the QPP acronym and instead refer to the "post-processing method" or "post-processing model". I recognize that it would be a bit longer, but in my opinion it would be clearer. I'm one of those persons who don't like acronyms too much...

We agree that the use of acronyms can often make it difficult to follow the text. On the other hand it can also lead to more compact wording. We will review the usage of acronyms across the paper and reduce them where possible.

1.4 Section 2 should be named "Post-processing methods" or "Post-processing models", and maybe the titles of the subsections could be adjusted accordingly.

We do agree that the loose usage of word "model" across the paper created considerable confusion.

The current title "Forecasting Models" is intended to make it clear we are talking about forecasting methods. A title "Post-processing" on its own does not communicate this.

In addition, the section content is not limited to post-processing methods – it also includes the deterministic hydrological model itself. It would be strange and potentially confusing in its own right to introduce the deterministic model in a section titled "post-processing" methods.

We do need some clear term for the complete system, which includes the deterministic model plus the post-processing model. We currently believe "Forecasting model" is an appropriate term, once it is clearly defined in the revised manuscript. However this will be reviewed as part of the comprehensive tightening of terminology across the paper

1.5 Finally, I would suggest mentioning GR4J explicitly in section 2.1 and referring to section 3.2 for greater details. In fact, I was initially a bit bothered by the fact that GR4J was described in the case study and not in the "Forecasting models" section, because to me, GR4J is a model and it can be used jointly with meteorological forecasts to produce streamflow forecasts. I think I understand what you wanted to do here: you wanted to avoid drawing too much attention to GR4J because it is not the central element of your study. So in the end, I don't mind too much GR4J being described in section 3.2, but I really think it would help clarify/distinguish things if you could mention it in section 2.1 too.

We do not focus on GR4J in Section 2.1 because the streamflow post-processing models can be used with any hydrological model, not just GR4J. As such, we simply referred to a "deterministic hydrological model h" in Section 2.1 and introduce GR4J in the case study in Section 3.2.

However, we do see value in providing some context regarding the term "deterministic hydrological model", and will refer to GR4J in Section 2.1 as an example of such deterministic model used in this study. We will revise the text accordingly.

1.6 This is linked to point 1. When I read the abstract and intro, I had a bit of difficulty understanding what you meant by "designed" for a single time step", because I had, again, a hydrological model in mind. So I was think that, for instance, if GR4J runs at a daily time step, there is nothing stopping you from aggregating it at the monthly time step. So I wasn't completely sure were you were going with that, until I read section 2. It then became very clear that the monthly post-processing method is completely inapplicable at the daily time step and I understood what you meant in the abstract and intro. I think it would be better it could be clearer right from the start, and I think that making it more explicit that you refer to post-processing methods would really help.

Thanks for highlighting this problem. We see that this would have been confusing.

In the revised paper we will explicitly refer to "post-processing models" when we state that forecasts are commonly designed for a single time step (i.e. in the first paragraphs of the abstract and the introduction).

See also response to comment 1.2 on the use of term "model" across the paper.

1.7 Page 1 line 20: Would it be more accurate to refer to ACCESS-S as an atmospheric model? When I think of a climate model, I think about the ones used for climate change study, that really model the climate over long periods and for which the fluctuations at smaller time steps have no real significance. I know they are basically very similar in their structures, but if I understand correctly, they are not run in the same way at all

The reviewer is correct that "climate model" is not the appropriate term. We will refer to ACCESS-S as a "numerical weather prediction model" in the revised paper.

1.8 I was really thankful for Figure 1. Section 2 was a bit heavy, but this figure is really helpful to understand how everything is put together. Great work.

Thank you.

1.9 Very small point: is it possible that the acronym REM is not defined in the text? Maybe I missed it? I understand it means Residual Error Model but can you please make sure it is defined in the text when first mentioned? Just for clarity.

Good catch. We will make sure this term is defined.

1.10 Very very small point: Line 111, why chose the subscript "foc"? Why not "fcst" if it is forecasted? I think it would be clearer?

"foc" is an abbreviation of "forecast" and is shorter than "fcst". We will review the usage of acronyms across the paper.

1.11 Page 5 line 141: do you mean each individual member or is it really each forecast? I understand you reduce (or "collapse") the forecasts to deterministic ones, so I guess it is really "each forecast", but I'm not sure. Can you maybe clarify? I was a bit lost.

We should have stated that the error model is applied to each individual member (i.e. each ensemble member). We will clarify this in the text.

Note for the MuTHRE model we do not collapse the ensemble of forecasts to a deterministic forecast. We see why the previous wording may have suggested otherwise, and believe the revised text will overcome this confusion.

1.12 Page 6 line 167: I like that you use the word "collapse" when referring to the transition from ensemble to deterministic. I think I will adopt this terminology myself in the future.

Nice.

1.13 Section 2.4 is very helpful. Very clear. Thank you.

Thank you!

1.14 Page 10 line 258: I would suggest referring to Schepen et al (2018) as a pre-processing method instead of a post-processing method, especially in the context of your study, since you use the result from this method as an input.

Good point. We will now say that “the rainfall forecasts are pre-processed using the method of Schepen et al (2018)” and use similar wording elsewhere in the paper.

1.15 I like section 3.4.1. The choice of performance indicators is very well justified and they are clearly appropriate.

Thanks.

1.16 I was surprised that Figure 5 b was not discussed more, especially in relations to the findings in the next page (statistical significance for the different metrics. When looking at Figure 5b, I can't help but think that the monthly post-processing method is not doing a very good job, and this is a bit surprising given that it is trained for monthly forecasts. First, there is an underestimation of streamflow between 2000-2001, and then (2006-2008) a pretty large overestimation. Why is that? Any idea? Then later we learn that, according to the different performance metrics, the differences in performance are actually not often statistically significant. I was surprised by that, because of that figure. I would really appreciate a bit more discussion/explanation of what is happening in 2000-2001 and 2006-2008.

Reviewer 1 is correct that the time series presented in Figure 5b for the monthly QPP model are surprising, and do not appear consistent with the metrics in Figure 6.

It turns out that there was a bug in the code used to produce this time series plots for the monthly QPP model.

The correct time series in Figure 1 below (which we will use in the revised paper) shows very little difference between monthly forecasts from the MuTHRE and monthly QPP models. This is much more consistent with metrics presented in Figure 6 of the paper.

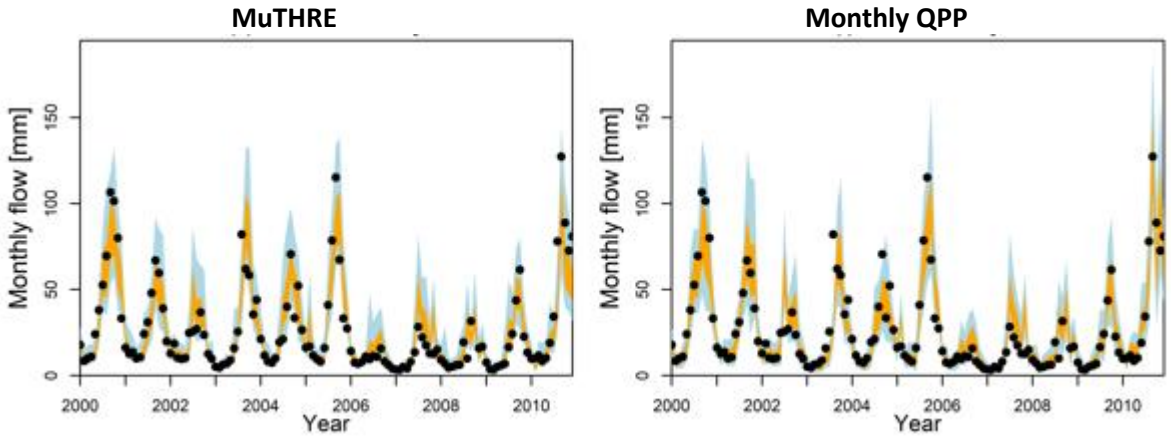

Figure 1: ***Revised Figure 5*** *for the manuscript, showing time series of monthly probabilistic forecasts for Murray River at Biggara.*

We are very thankful to Reviewer 1 for their attention to detail, which has allowed us to identify and resolve this problem.

1.17 Line 388: I suggest removing the word "actually" to avoid repetition with the previous sentence

Will do.