Response to interactive comment by anonymous referee #1

General comment

I found the paper very well written and it is well-arranged. In my opinion, the manuscript fits pretty well into this special issue and its content is relevant for publication in HESS. The extensive experiment of the continent-wide forecasting system is described in depth regarding the underlying methodology (including the statistical and rainfall-runoff models), the verification procedure as well as the numerous results. This work is a consistent step to continue previous work reported e.g. by Bennett et al. (2016) and Turner et al. (2017).

Response: Thanks very much for the careful review and the positive feedback.

Major comment

I suggest trying to condense section 2.2 (especially its subsections) in order to strengthen the role of the Bayesian prior, which is assessed in experiment 3 and which seems to be most promising / sensitive to improve FoGSS output. The error model approach is very interesting, but as a reader, I slightly lost the focus on the relevant aspect (the prior) for your study.

Response: We will shorten this section, as suggested. We will also reorganise it to emphasise the prior, by separating the hydrological model from the error model, and moving the description of parameter estimation after the description of the error model. This means the reader arrives much more directly at the description of the bias-correction, and we will also more directly flag its use in the experiments with the prior.

Minor / Technical comments

Page 2, line 18: I suggest adding a short explanation, why BJP is not (or even may be cannot be) suitable for those long-range forecasts in order to assist readers, who aren't familiar with the BJP approach.

Response: We will note that the BJP has no mechanism for generating hydrographs, which are required for streamflow forecasts for long time series.

Page 2, line 25: It might be beyond the scope of this paper, but did you experience that the preference of some water agencies to use stochastic scenarios (instead of seasonal forecasts) might be based on the fact that they still have to gain confidence to this "new" source of information (they are used to the scenarios, they comprehend it, ..)? So the "practical" advantages you mentioned might also contain such more psychological aspects instead of purely technical ones?

Response: This is a thought-provoking question, and one we can only speculate on as the FoGSS forecasting system is not yet operational. (At present, stochastic scenarios or climatology are really the only option for many water agencies looking for 12-month panning scenarios.) We based our comments on the use of stochastic scenarios partly on responses of water agencies to a formal (but unpublished) survey conducted by the Bureau of Meteorology, and partly on our own (more anecdotal) interactions with water agencies. Neither source of information gives clear indication of possible barriers to adopting a new forecasting service. In some cases water agencies show a great deal of faith in the existing 3-month ensemble forecasts (generated with the BJP), and they simply wanted longer range forecasts: in these cases we feel that they would probably be quite happy to use long-range monthly forecasts if they were available (notwithstanding computational barriers to using large ensembles, which is another discussion). In other cases, however, we agree that the

resistance to using forecasts is not purely technical, and it may be due to 'institutional inertia' (which is possibly the result of psychological preferences of key staff). As the reviewer surmises, however, we feel that this discussion is outside the scope of the paper. It is probably worthy of a more detailed discussion in its own right in a paper directed at barriers to adoption of new forecasting services.

Page 3, line 8: I recommend mentioning how the three variants of the POAMA model are generated (variation of model parameter)?

Response: We will note that these variants are generated by changing model physics.

Page 3, line 29: I suggest to split section 2.2 and to add a separate chapter "Hydrological model" (or something similar). Even this chapter might be relatively short, I suggest to have a separate section for each of the main FoGSS components, which correspond to the three experiments described in section 3 of this paper.

Response: We will follow this suggestion, as noted above.

Page 3, line 35: Have you thought about / tested using data assimilation techniques to reduce hydrological model errors?

Response: To date, we have focussed our research on improving the hydrological error model. This has a similar effect to state updating, but we feel it is a more approach direct (streamflow errors are directly measurable, while states are not). Importantly, this allows us to correct reliability more directly. Additional updating of states may possibly improve our forecasts. However, existing methods of state updating would be difficult to implement alongside our error model (at least, as it is currently formulated). However, we may consider state updating in future (if we get the time!).

Page 5, line 8: As you state that the upper limit of d is arbitrary, it would be interesting to know, if you have tested other thresholds before (and you ended up with 2)?

Response: We also considered upper limits of 1 and 1.5. We found these too restrictive in some catchments (some of the biases are very strongly non-linear), but we felt that values greater than 2 could cause unrealistically large bias corrections under cross-validation (indeed, our use of the prior showed this to be true even for the upper limit of 2). The use of the prior makes the choice of the upper limit less important.

Page 11, line 1: I guess that "perennial" and "ephemeral" have to be switched?

Response: These are correct as they are, but we can see that this is confusingly phrased. We will rephrase to avoid the confusion.

Page 12, line 4-6: Could you please explain, why you are planning to improve Wapaba instead of using the GR2M model, as the latter one seems to perform better, especially in drier catchments? I think you give some kind of explanation on page 13 (line 13-18), but I suggest to add a link or to explain your decision to the reader already at the end of section 4.3.

Response: We agree that this is not a strong justification for future research, and we will remove this statement.

Appendix A: As the state is a relevant information in your list, I suggest to explain the acronyms used, as several reader might not be familiar with the different Australian states.

Response: We will add explanations of these acronyms.

Figure 1: Why does the arrow linking "rainfall-runoff model" and "Climatology PE" point in the direction of the climatology? Isn't climatology potential evaporation an input to the rainfall-runoff model?

Response: Thanks for picking this up – we will change the direction of the arrow.

Figure 3 + 13: I suggest to explicitly mentioning the CRPSS as "skill measure". Figure 5, 6, 7, 8, 9, 10, 11, and 12: I suggest adding the total number of catchments in each panel (e.g. in brackets behind the title).

Response: We will add the information, as suggested.

Typographical corrections

Page 4, line 5: Shouldn't it be "homoscedastic" in this context?

Page 4, line 24: I suggest to insert a comma before ": : : a and b are parameters."

Page 5, line 1: I think "takes" should be deleted.

Page 5, line 3: The word-wrap slipped (comma at the beginning of line 4 should be in line 3).

Response: Thanks for reading our manuscript to closely – these are errors that we will correct.