Interactive comment on "The challenge of forecasting high streamflows in medium sized catchments 1–3 months in advance" *by* J. C. Bennett et al.

J. C. Bennett et al.

james.bennett@csiro.au

Response to Anonymous Reviewer 3

Review comments are in double quotation marks. Line numbers (e.g. lines xxx-yyy) refer to the manuscript attached as the supplement.

Comment: "Overall recommendation: Return to author for some revisions Overall this is a good paper on a very important topic. The paper is well written, well organised and the aims of the study are outlined well in the introduction. The conclusions and discussion are also sound and backed up by the results that were obtained.

The result that antecedent conditions are an important predictor is good...and consistent with previous studies (e.g. Chiew, Verdon etc)... However, I have some concern that the results and main conclusion (i.e. little skill associated with inclusion of large-scale climate drivers as predictors) are more to do with methodology rather than there actually being no skill associated with inclusion of climate indices as predictors."

Response: Thanks for the positive feedback – we've tried to address your suggestions to improve the paper, in particular modifying our major conclusion to state that the addition of climate indices adds little information to that already contained in catchment wetness, noting that catchment wetness implicitly includes a considerable amount of information about the state of the current climate.

Specifically:

1) Comment: "More information is needed on the indices and how exactly they were used? Are you basing your predictions on correlations between climate indices and flow 1mth or3mths later? Or are you using stratification/phase type approaches as per some of the studies you mention (e.g. Verdon, Chiew etc)?"

Response: Thanks for pointing out this wasn't clear. We use lagged climate indices (i.e. relationships between climate indices and high streamflows occurring in the next 1-month period or in the next 3-month period). To clarify this, we have changed the title of the paper to "The challenge of forecasting high streamflows 1-3 months in advance with lagged climate indices in southeast Australia" to emphasise the use of lagged climate indices, define what we mean by 'lagged climate indices' (lines 210-216) and use the term consistently throughout the paper.

2) Comment: "What about SAM, STR-intensity, STR-position, Interdecadal Pacific Oscillation (or similar PDO)?"

Response: We chose our indices from the work of Schepen et al. (2012). This paper did not test the STR specifically, but showed that using lagged climate indices based

on atmospheric variables – e.g. SAM – show very weak relationships with rainfall in Australia. (We have reviewed the predictive power of lagged STR position/intensity for seasonal forecasts but this work is not published: we found that lagged STR adds negligible information to forecasts of seasonal rainfall totals under cross-validation.) We discussed this issue using the Blocking index (B140) as an example, but we now note that the same reasoning could be applied to SAM and the STR (lines 516-518). While it is known that some atmospheric indices have strong concurrent correlations with Australian rainfall, they have much lower autocorrelation than SST variables, which tend to be more persistent, and therefore are of little value for forecasting. IPO-PDO are not practically useful predictors for real-time forecasting systems, because it is not possible to know with certainty whether we are in a PDO +ve or –ve phase at any given moment (see response to comment #13, below). We have now discussed the difficulty of using the IPO index explicitly (lines 543-552).

3) Comment: "What about potential interaction between different climate drivers (see for example Gallant et al Understanding hydroclimate processes in the Murray-Darling Basin for natural resources management. Hydrology and Earth System Sciences, 16, 2049-2068, doi:10.5194/hess-16-2049-2012)....such that two or more climate drivers acting at the same time result in different conditions than if they were acting individually?"

Response: We agree that these interactions may add to the predictability of rainfall in some cases. We now note that characterising such interactions may improve forecasts in future (lines 544-547, 558-560). We also note, however, that identifying and characterising these interactions for the purposes of generating real-time forecasts is extremely difficult. For example, the Gallant et al. paper cited above argue that interaction between SAM and ENSO are an important indicator of rainfall in the Murray-Darling Basin over the period 1904-2005. The finding is difficult to apply to real-time forecasting because: 1. The calculations assume SAM can be calculated concurrently with rainfall. SAM is an atmospheric phenomenon, and therefore less persistent than SST-based

lagged climate indices for the purposes of forecasting rainfall (see response to comment 2, above) 2. The interaction is defined by thresholds (in their case, a positive or negative phase was defined as being outside ± 0.5 standard deviations). Approaches like this (which essentially stratify indices) are very difficult to use in statistical forecasts, because they greatly reduce the data available to train statistical models. 3. Gallant's findings report the ability of correlations to describe historical data, rather than to forecast rainfall. In general, predictive skill of indices (measured, e.g., by cross-validation) are lower (sometimes much lower) than fitting skill. We have found in previous work that including more than one climate variable as joint predictors of seasonal total rainfall/streamflow in the BJP model add negligible skill to forecasts (Robertson and Wang, 2012; Wang et al., 2012), and we now state this in the paper (lines 198-204).

Finally, these interactions should, in theory, be accounted for in dynamical coupled ocean-atmosphere climate models, yet these models do not necessarily generate more accurate rainfall forecasts than merged single-predictor statistical models, in particular at longer lag times (Schepen et al., 2012b), as we now note (lines 560-562).

4) Comment: "For the SST related indices, basing your predictor on only 1 month may not give a proper indication of the true climate state (e.g. for the climate state to be considered a "true" La Niña the SSTs needed to be persistently warmer than average to the north of Australia for several months).....because your method only considers the one mth prior to the period you want to forecast for there is the possibility that whatever happened in that one month prior may not be indicative of the overlying "climate state"....."

Response: We agree that including information from indices at longer lag times can add information to rainfall forecasts, but (Schepen et al., 2012a) show that the strength of the relationship between lagged climate indices and Australian rainfall declines rapidly with lag time, and hence we decided to concentrate on climate indices at the shortest possible lag time, as we now note in the paper (lines 211-214). Perhaps more importantly, much of the information of the climate 'state' is already contained in the estimate

of catchment wetness. This means that the addition of lagged climate indices adds little information, even in regions where teleconnections are strong. We now note this explicitly in the discussion (lines 535-538).

5) Comment: "As you state in the paper for atmospheric related indices it is the opposite to the point above.....1mth prior is too long to pick up things like cut-off lows, east coast lows etc (i.e. the main weather events associated with high flows)......."

Response: We agree. Similar issues apply to other atmospheric indices, including SAM and position/intensity of the STF, as described above. We now mention TSI and SAM in this discussion (with B140) (lines 516-518).

6) Comment: "I think it is too easy to conclude that "including climate indices as predictors adds little skill to the forecasts" when in reality what you have actually shown is that "including climate indices (as chosen and utilised by you in the method chosen and developed by you) as predictors adds little skill to the forecasts (based on skill assessments chosen by you)". There are several assumptions and sensitivities in there and I think caveats should be made in the paper along those lines rather than just concluding catchment wetness is pretty much all there is too it (intuitively and anecdotally this doesn't make sense - as alluded to page 3148, lines 15-25). For example, other work (e.g. SEACI, Verdon, Kiem, Pook, Timbal etc) has shown that the frequency and intensity of synoptic events typically associated with high flows (e.g. east coast lows, cut-off lows etc) is dependent on the overarching climate state. So it isn't so much that climate indices do not add skill as it is that climate indices will only add skill if they are used in such a way as to capture the variable frequency of sub-monthly weather events that are associated with high flow events."

Response: Thanks for this feedback, however, we largely disagree with this sentiment. We agree that we could communicate our method and results better, and we have attempted to address this comment with several additions and qualifications, which we describe below. We believe our methods/analyses are robust and generalisable for

the following reasons: 1. The studies cited above did not attempt to produce pseudo real-time forecasts (i.e. with robust cross-validation and lagged climate indices), but rather examined concurrent (not lagged) relationships between indicators of the climate state and rainfall. When cross-validation methods are employed to relationships between lagged climate indices and rainfall these relationships are often weaker (as demonstrated by Schepen et al. 2012 and Robertson and Wang 2012) than might be supposed from an analysis of historical correlations based on concurrent relationships between climate indices and rainfalls. 2. We show in this (and several other) papers that RMSEP agrees broadly with other commonly used measures of probabilistic forecast performance (in this paper, the Brier Score and the likelihood ratio). For the analyses of the marginal predictive power of climate indices, we show an example catchment (MUR) in Figure 1 below, which compares RMSEP to the well-known Continuous Ranked Probability Score (CRPS). There is very close agreement on both the sign and the magnitude of the skill scores (indeed, RMSEP often shows slightly greater marginal skill) and this holds true for other catchments. We have not included all these analyses in the paper for the sake of brevity. To assist the reader in understanding the RMSEP results we have added additional description of RMSEP to explain its utility (lines 304-309). 3. We have deliberately chosen streams with long and largely complete flow records (see response to comment #12 about streamflow records, below) and employed stringent cross-validation to ensure that our findings are robust. 4. As the reviewer notes, the relative importance of antecedent conditions for forecasting seasonal streamflows with respect to forecasts of rainfall is well established by many studies (e.g. Mahanama et al., 2012; Shukla and Lettenmaier, 2011; Li et al., 2009; Koster et al., 2010; Robertson and Wang, 2013) and this includes statistical as well as dynamical forecasts of rainfall. In short, we have taken a robust method, accounted for all relevant uncertainties and found that by far the largest source of skill in forecasts of high flows is antecedent catchment wetness. In addition, our findings are in strong agreement with many other studies. Therefore we do not agree that we have drawn the conclusion that climate indices add little skill to forecasts of high streamflows easily. We

have gone to some lengths to demonstrate this finding through careful analyses and reasoning. To the degree that large scale climate indices are correlated with precipitation, it is highly likely that they will be correlated with antecedent catchment wetness. In other words, it is not that correlations between lagged climate indices and precipitation are non-existent, it is that lagged indices do not add more (climate) information to forecasts than is already available in antecedent catchment moisture. This seems intuitive (at least to us).

Finally, we agree that we can improve the explanation of both our methods and results to clarify these issues, which we have attempted to do as follows: 1) We have added 'lagged climate indices' to the title of our paper to reflect the scope of the paper 2) We have expanded our discussion of results to note that catchment wetness imparts considerable information about the current state of the climate, which may explain the negligible contribution of lagged climate indices to forecast skill (lines 535-538) 3) As noted above, we have expanded our explanation of RMSEP to explain its utility (Lines 304-309)

7) Comment: "'Normalising and stabilising variance' in predictands and predictors (as the BJP does, as described section 2.2.4) I don't think is a good way to do this given the inherent variability, non-linearity, and non-stationarity associated with flow (especially high flows) in Australia."

Response: It is precisely because the relationships we are attempting to describe are often non-linear and show high variability that we transform the data. Streamflows in general, and high flows in particular, have highly non-normal distributions (accompanied by inhomogeneous variance), and their relationships to the various indices cannot be described without accounting for these properties. The transform parameters are an integral part of the BJP model: all parameters (the transform parameters and the parameters of the multivariate normal distribution) are estimated jointly (lines 237-242). Therefore, the BJP modelling approach can be thought of as a sophisticated regression technique that accounts for non-linear relationships between predictors and pre-

dictands as well as parameter uncertainty. The success of the BJP is then gauged on how well it can use these relationships to forecast independent events.

8) Comment: "On top of this, it seems there is another "smoothing" step introduced via the BMA......given all the "normalising" and "averaging" is it really a surprise that abnormal, infrequent, and highly variable timeseries such as high flows are not well predicted?"

Response: We do not agree that the BMA is necessarily another significant source of smoothing – particularly where the evidence for a given teleconnection is strong. Bayesian model averaging weights the forecasts (given as probability densities) of different models based on their predictive ability (under cross-validation). That is, it weights individual models highly where evidence of a given teleconnection is strong. This allows the predictive densities of different models to be combined to produce the most skilful forecasts possible. If one teleconnection is particularly strong (e.g. the NINO3.4 index in a given season), the BMA will weight that model strongly, and will ignore other models. So the BMA will allow the forecasts to reflect the accuracy (and uncertainty) of a forecast where the evidence for a particular climate index is strong (i.e., it will not get "smoothed"). Where evidence for teleconnections is more equivocal, the BMA will adjust the weights (and accuracy and uncertainty) accordingly. Wang et al. (2012) showed that this improves forecast skill.

Minor comments:

9) Comment: "Table 2. Anomalies are used for some indices. What periods were the anomalies calculated based on? Why were anomalies used for some indices but not all?"

Response: Thanks for pointing this omission out. In fact, all the SST indices we use are anomalies. The anomalies are calculated in relation to 1971-2000 (this is specified in the reference for the SST temperature series, Smith et al. 2008, which we have now cited in response to the comment #10, below). Whether the indices are anomalies or

not makes no material difference to the BJP models.

10) Comment: "Table 2. For the indices, which SST and SLP data sets were the indices calculate from?"

Response: Apologies for this oversight. We have added the following text to address this comment: "Sea surface temperature climate indices were derived from the National Center for Atmospheric Research (NCAR) Extended Reconstruction of Sea Surface Temperature version 3 (Smith et al., 2008). B140 was derived from the National Centers for Environmental Prediction (NCEP)–NCAR reanalysis data (Kalnay et al., 1996). SOI was sourced from the Australian Bureau of Meteorology (BOM)." (Lines 205-209)

11) Comment: "Table 1. Different analysis period was used for different stations. Are the results sensitive to this? Why not use a consistent analysis period?"

Response: We used the maximum amount of data available (always constrained by the availability of streamflow data) to train our models, as we would if generating real-time forecasts. This gives the models the best chance of success. The cross-validation means that the results are less sensitive to the choice of period.

12) Comment: "Where was the daily flow data obtained from? Was it complete over the periods indicated in Table 1? Was any infilling conducted? How were losses/gains due to non- natural influences accounted for (e.g. irrigation, farm-dams, reservoir spills etc)??"

Response: Again, apologies for this oversight. The data were taken from the CWYET dataset (Vaze et al., 2011), a quality-controlled streamflow dataset of catchments with minimal human impact (i.e., no major storages or diversions). The sites were chosen in part because of their long and largely complete records, however there were (of course) some missing values. We have added a column to Table 1 listing the proportion of streamflow data missing. No infilling was carried out, and as the catchments are not

affected by large dams or diversions, no account was made of farm dams or other extractions (which in any case tend to have less effect on high streamflows). We have now noted the source of these data in the paper (lines 114-116).

13) Comment: "I don't think enough (or any) comment is made about the role of multi-decadal climate variability. The paper by Kiem et al 2003 is cited which showed that the analysis period used in this study is split into two epochs (predominance of high flows pre 1978; lack of high flow post-1978). This is associated with the IPO which Scott Power in his 1999 paper also showed influences predictability (i.e. when IPO is negative (i.e. 1948-78) predictive skill, associated with SOI in the Power paper, is enhanced and when IPO is positive (as it has been since late-1970s until 2010) predictive skill is reduced. Given the majority of your study period is dominated by IPO +ve conditions (i.e. low predictability) I think some discussion and caveats along these lines is warranted."

Response: Thanks for pointing this out – we have added discussion of the role of interdecadal variability as follows: "Some studies (e.g. Kiem et al., 2003) use an index describing the Interdecadal Pacific Oscillation (IPO) to relate rainfall/streamflow to climate indices. If we limit our assessment of forecasts only to periods where IPO was in the negative phase, it is possible that ENSO SST indices may add more skill to the forecasts (as suggested by Kiem et al., 2003). However, we sought to assess forecast skill in the context of generating forecasts in real-time. Describing the IPO is not particularly useful for real-time forecasting because it is only possible to define an IPO phase with certainty in retrospect (although informed speculation about the present IPO phase is possible; see, e.g., Cai and van Rensch, 2012). That is, it is often not possible to know with certainty which IPO phase we are in at the present time, so it cannot be used to inform real-time forecasts." (Lines 543-552)

References

Cai, W., and van Rensch, P.: The 2011 southeast Queensland extreme summer rainfall:

a confirmation of a negative Pacific Decadal Oscillation phase?, Geophysical Research Letters, 39, L08702, 10.1029/2011GL050820, 2012.

Kalnay, E., M. Kanamitsua, R. Kistlera, W. Collinsa, D. Deavena, L. Gandina, M. Iredella, S. Sahaa, G. Whitea, J. Woollena, Y. Zhua, A. Leetmaaa, R. Reynoldsa, M. Chelliahb, W. Ebisuzakib, W. Higginsb, J. Janowiakb, K.C. Mob, C. Ropelewskib, J. Wangb, R. Jennec, and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, Bull. Amer. Meteor. Soc., 77, 34, 1996.

Kiem, A. S., Franks, S. W., and Kuczera, G.: Multi-decadal variability of flood risk, Geophysical Research Letters, 30, 1035, 10.1029/2002GL015992, 2003.

Koster, R. D., P., S. P., Mahanama, Livneh, B., Lettenmaier, D. P., and Reichle, R. H.: Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow, Nature Geoscience, 3, 613-616, 10.1038/NGEO944, 2010.

Li, H., Luo, L., Wood, E. F., and Schaake, J.: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, Journal of Geophysical Research: Atmospheres, 114, D04114, 10.1029/2008jd010969, 2009.

Mahanama, S., Livneh, B., Koster, R., Lettenmaier, D., and Reichle, R.: Soil moisture, snow, and seasonal streamflow forecasts in the United States, Journal of Hydrometeorology, 13, 189-203, 10.1175/jhm-d-11-046.1, 2012.

Robertson, D. E., and Wang, Q. J.: A Bayesian approach to predictor selection for seasonal streamflow forecasting, Journal of Hydrometeorology, 13, 155-171, 10.1175/JHM-D-10-05009.1, 2012.

Robertson, D. E., and Wang, Q. J.: Seasonal forecasts of unregulated inflows into the Murray River, Australia, Water Resour Manage, 27, 2747-2769, 10.1007/s11269-013-0313-4, 2013.

Schepen, A., Wang, Q. J., and Robertson, D.: Evidence for using lagged climate indices to forecast Australian seasonal rainfall, Journal of Climate, 25, 1230–1246,

10.1175/JCLI-D-11-00156.1, 2012a.

Schepen, A., Wang, Q. J., and Robertson, D. E.: Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall, Journal of Geophysical Research, 117, D20107, 10.1029/2012JD018011, 2012b.

Shukla, S., and Lettenmaier, D. P.: Seasonal hydrologic prediction in the United States: understanding the role of initial hydrologic conditions and seasonal climate forecast skill, Hydrology and Earth System Sciences, 15, 3529–3538, 10.5194/hess-15-3529-2011, 2011.

Smith, T. M., Reynolds, R. W., T.C.Peterson, and Lawrimore, J.: Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006), Journal of Climate, 21, 2283–2296, 2008.

Vaze, J., Perraud, J., Teng, J., Chiew, F., Wang, B., and Yang, Z.: Catchment Water Yield Estimation Tools (CWYET), 34th World Congress of the International Association for Hydro-environment Research and Engineering and 33rd Hydrology and Water Resources Symposium and the 10th Conference on Hydraulics in Water Engineering, Brisbane, 2011, 1554-1561, 2011.

Wang, Q. J., Schepen, A., and Robertson, D. E.: Merging seasonal rainfall forecasts from multiple statistical models through Bayesian model averaging, Journal of Climate, 25, 5524-5537, 10.1175/JCLI-D-11-00386.1, 2012.

Figures Captions:

Figure 1: Skill (%) of BJP-BMA models calculated with respect to the BJP model that uses only catchment wetness as a predictor for the MUR catchment.

Please also note the supplement to this comment:
http://www.nat-hazards-earth-syst-sci-discuss.net/1/C2022/2013/nhessd-1-C2022-2013-supplement.pdf

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., 1, 3129, 2013.

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

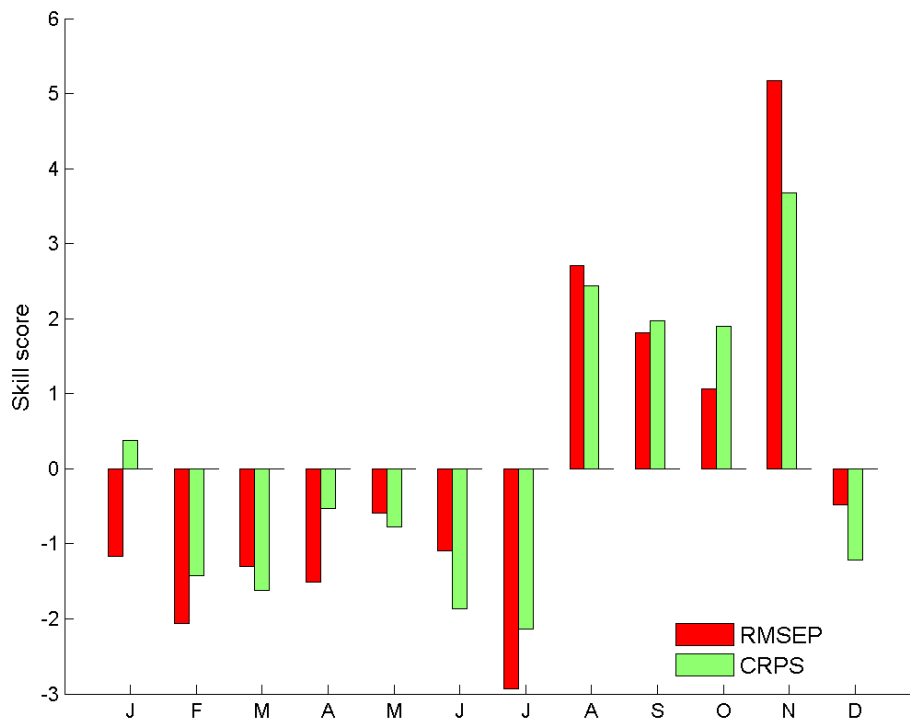Interactive Discussion

Discussion Paper

Fig. 1. Skill (%) of BJP-BMA models calculated with respect to the BJP model that uses only catchment wetness as a predictor for the MUR catchment.