Response to referee comments.

Referee comments in black and author responses are in blue.

Interactive comment on "Post processing rainfall forecasts from numerical weather prediction models for short term streamflow forecasting" by D. E. Robertson et al.

J.S. Verkade (Referee)

jan.verkade@deltares.nl

Received and published: 22 July 2013

The paper is generally well written, although I found the sections on the actual postprocessing technique tough to read. Although I have done some postprocessing of precipitation forecasts myself, I find this a difficult topic and I have concentrated on how the resulting forecasts have been verified. Although the ensemble forecasts have been verified to some extent, I think a more extensive verification could improve the manuscript. If, however, the authors decide otherwise, I would like to see a note added that explains the limitations of the 'as is' verification. This pertains to both the (absence of) conditional verification in terms of CRPSS and bias, as well as to the verification of added value of the Schaake shuffle. I don't mind revealing identity (Jan Verkade, Deltares, Delft, the Netherlands). I don't need to see a revised version of the manuscript; I trust that the authors will deal with the comments appropriately.

More specific comments are uploaded in a separate pdf file.

General comments

The paper is generally well written, although I found the sections on the actual postprocessing technique tough to read. Although I have done some postprocessing of precipitation forecasts myself, I find this a difficult topic and I have concentrated on how the resulting forecasts have been verified.

Although the ensemble forecasts have been verified to some extent, I think a more extensive verification could improve the manuscript. If, however, the authors decide otherwise, I would like to see a note added that explains the limitations of the 'as is' verification. This pertains to both the (absence of) conditional verification in terms of CRPSS and bias, as well as to the verification of added value of the Schaake shuffle.

I don't mind revealing identity (Jan Verkade, Deltares, Delft, the Netherlands). I don't need to see a revised version of the manuscript; I trust that the authors will deal with the comments appropriately.

We thank the referee for his detailed review and have made many changes to the manuscript to address his comments.

Specific comments

Section 1: Introduction

(1) P6767, l12 onwards: why single out initial conditions and weather forecasts as sources of uncertainty, and not also mention the models, parameters also?

Sentence added to paragraph to cover concept of hydrological models introducing uncertainty.

"Hydrological models used to transform observed and forecast rainfall to streamflow simulations also introduce uncertainties streamflow forecasts through their simplified representations of the true hydrological processes (Pokhrel et al., 2013; Gupta et al., 2006)."

Section 2: Study catchment and data

(2) I think it would be worthwhile to say something about response times of the catchment and of subcatchments within Ovens. This would further clarify the rationale for having ensemble precipitation forecasts at a sub-daily time step.

Sentence added to discuss catchment response times

"The time of concentration to the catchment outlet is of the order of four to five days; however the time of concentration to some flood sensitive areas within the catchment can be less than 24 hours and therefore models are hydrological models are run at sub-daily time steps (Pagano et al., 2011)."

Section 3: Methods

(3) Throughout the paper, the terms "forecast period" and "lead time" are used. Do they indicate the same thing? If so, maybe good to mention that these are used synonymously, or use a single term only. P6774, l4 onwards is an example of this.

In general we have replaced the term 'forecast period' with 'lead time', however there are a few instances, mainly in the introduction where we believe the term 'forecast period' is more appropriate.

(4) Section 3.1, p6774, l11 onwards: "The post processed probabilistic forecasts of three hour rainfall accumulations (for lead times of 0–60 h) do not contain appropriate spatial and temporal correlation structures.". I believe this may indeed be true, but we really need to see some evidence of this, either by supplying an argument or through some quantitative measure. Alternatively, you could cite evidence gathered elsewhere.

The reasoning for the post processed probabilistic forecasts not having the appropriate spatial and temporal correlations structures has now been added here. Sentence now reads:

"The post-processed probabilistic forecasts of three hour rainfall accumulations (for lead times of 0 – 60 hours) are random samples from independent probability distributions and hence ensemble members created by linking these samples in a simple manner will not contain appropriate spatial and temporal correlation structures.."

A figure (Figure 13) has also been added in the results section to demonstrate this.

(5) You may want to consider moving most of the contents of section 3.3 to an appendix, as these are standard metrics.

The reviewer is correct that these are standard metrics and hence only a brief summary description is given.

(6) Section 3.3.3: note that here (i.e. in the case of ROC), 'unskilled' has a different meaning than in the section on CRPS.

A sentence has been added here to clarify the point.

"Note that unskilled forecasts from a forecast discrimination sense are not the same as climatology forecasts used as a reference for the CRPS skill score, rather they imply that the forecast event probabilities are random."

(7) Section 3.3.3: "Post processing does not influence forecast discrimination". While this claim has been substantiated with a reference, a brief argument would help the reader understand this.

The sentence has been removed as we believe it is unnecessary.

(8) Section 3.3.4: I don't think reliability plots for temporally aggregated forecasts say anything about the Schaake shuffle's ability to restore spatial correlations. Please change accordingly.

The reviewer is correct. Comment about spatial correlations is removed.

Section 4: Results

(9) This section contains both results and conclusions. I recommend renaming it accordingly.

Section renamed Results and discussion

(10) CRPSS and Bias Score are presented for the full available sample only. While for the full sample, the metrics show an improvement in forecast quality, I would be interested in seeing the results for higher quantiles of the climatological distribution. For example, if one would single out observed precipitation events of 5mm/d and higher only, what would CRPSS and bias be then? I would highly recommend augmenting the analysis this way. Please refer to Brown and Seo (2013) for a good example of how this can be done. If, however, authors decide otherwise, then I would like to see a note about interpretation of the scores in results and/or discussion section, explaining the limitations of a full sample analysis only.

The results analysis has been augmented following the reviewers recommendation. Figures for the CRPSS, Bias score and AUC have been generated for a spectrum of thresholds values. See figures 5, 8 and 10.

(11) After reading paragraph 4.2.3, the reader could be forgiven for thinking that the diagonal line in the ROC plot corresponds to the ROC curve of the climatological forecast. This is not the case. I would suggest sharpening the text to distinguish between the ROC curve of a climatological forecast and the diagonal line, which corresponds to a situation in which there is no correlation between a forecast and an observation.

The sentence has been rewritten as follows

"This suggests that as lead time increases the probability of high rainfall in the post processed forecasts becomes less informative and less strongly correlated observed high rainfall events"

(12) Similar to my comment above, I would be interested in learning about ROC curves for more than two events only. As it would be impractical to plot many more figures, you could consider not showing ROC curves, but showing values of the AUC instead. Note that while I think this would constitute an improvement, it is not strictly necessary.

The results section has been augmented to include AUC for a spectrum of threshold values following the reviewers recommendation. See figure 10.

(13) The reliability plots in figures 10 and 11 each contain three points only. Why is that? The plots would be more informative if they would show, for example, observed relative frequencies for forecast probabilities {0, 0.1, 0.2, … 1.0}. Conceivably, this would more evenly distribute samples between the then 11 points on the curve near the now fullest bins.

The number of bins in figures 10 and 11 (now Figures 11 and 12) has been increased to 5. Increasing the number of bins further results in bins containing no forecasts and therefore a less robust assessment of forecast reliability.

(14) In section 4.2.4, it is stated that "For day 2 the forecast probability of a rainfall event of greater than 5mm appears to be unreliable." I agree that it appears to be less reliable than the other forecasts, but when is a forecast reliable and when is it not? To facilitate comparison, you may want to consider looking at the decomposition of Brier's probability score, and then specifically at the reliability component. Again, while I think this would constitute an improvement, it is not strictly necessary.

The wording of this paragraph has been moderated and the sentence now reads:

'For day 2 the forecast probability of a rainfall event of greater than 5 mm appears to be less reliable.'

(15) While I agree that the Schaake Shuffle aims to restore space – time correlations, I don't think the spatial correlation is tested here. As a matter of fact, I'm not so sure about the temporal correlation, either – I would be interested in seeing verification results of aggregated precipitation before and after application of the Schaake Shuffle. I am guessing that this will show that indeed, temporal correlations have been restored, but until then we simply don't know.

The restoration of the temporal correlations is assessed through assessing the reliability of forecasts of cumulative rainfall totals. These cumulative totals can only be reliable if the forecast for individual periods are reliable and the appropriate temporal correlations exist between the individual forecast distributions.

However, to demonstrate that the forecasts do have the appropriate correlations we have introduced figure 13 which explicitly demonstrates this point.

(16) The Results section does not contain any information on the uncertainty in the verification metrics, except for some comments related to small bin sizes for some of the points on the

reliability diagrams. I would like to see a statement on this, preferably in quantitative terms but in any case in qualitative terms.

Uncertainty information has now been added to the reliability diagrams and the new plots of conditional performance contain quantitative estimates of uncertainty in the verification metrics.

(17) I would recommend merging Figures 4 & 5 and 6 & 8 respectively, after connecting the points with lines.

Done.

Section 5: Discussion

The Discussion section is quite good in how it describes merits and limitations of the postprocessing technique. I have no comments to add to that.

(18) I think the real proof of this method is when the postprocessed precip forecasts are used to force the hydrologic model. This will put both the BJP and the Schaake shuffle to the test. I am much looking forward to seeing the results of that!

I agree.

Section 6: Conclusions

(19) This section contains a summary of the manuscript only. I would recommend renaming the title to "Summary". I would then also rename section 4 to "Results and Conclusions". You may even consider removing Section 6 altogether, as it contributes little to what's already stated in the abstract.

Section 4 renamed Results and discussion; Section 5 renamed Further discussion; Section 6 renamed Summary and conclusions

Technical corrections

(20) P6768, l15: "To generalise the approach requires" is grammatically incorrect. Maybe "Generalising the approach requires" would be better.

Corrected

(21) The prefix 'post' in conjunction with 'processing' can be written either with or without a hyphen, but if the latter option is chosen, the words should be joined: postprocessing, not separated.

Corrected

(22) I find the use of the term 'forecast period' somewhat confusing; why not use 'lead time' instead (similar to in Section 2, p6770, l25).

See response to comment (3)

(23) P6771, l7: consider omitting local time. It doesn't contribute much.

Noted, but we will retain to provide context for Australian readers.

(24) P6773, l13: “is used to generated a value”

Corrected

(25) P6777, l18: “by a single point”

Corrected

P6777, l18: “Here, ROC plots…”

Corrected

(26) P6780, l21: “than in the raw forecasts”

Corrected

(27) P6781, l4: “than for the events where rainfall is less than the event of rainfall less than 0.2mm.”

Corrected

(28) P6784, l5: overfitting

Corrected

(29) Fig1 uses a colorscale that I think is somewhat unusual. It is more common to associate blue with lower altitudes, and yellow/red with higher altitudes.

Figure changed to use a more common color scheme.

(30) Figures 2, 9, 10 and 11 use a large margin between plots. I suggest reducing, or removing the margin and labelling top row and upperleft columns only. This will increase legibility of the figures.

Margins have been reduced in all cases

(31) The figures lack legends. Instead, the authors have described the meaning of the lines/dots in the caption. I think the figures would benefit from a proper legend, though.

Some figures do not lend themselves to a simple legend within the figure and in these circumstances we have placed the legend in the caption.

Bibliography

Brown, J. D. and Seo, D.-J.: Evaluation of a nonparametric post-processor for bias correction and uncertainty estimation of hydrologic predictions, Hydrol. Process., 27(1), 83–105, doi:10.1002/hyp.9263, 2013.

References

Gupta, H. V., Beven, K. J., and Wagener, T.: Model Calibration and Uncertainty Estimation, in: Encyclopedia of Hydrological Sciences, John Wiley & Sons, Ltd, 2006.

Pagano, T. C., Ward, P., Wang, X. N., Hapuarachchi, H. A. P., Shrestha, D. L., Anticev, J., and Wang, Q. J.: The SWIFT calibration cookbook: experience from the Ovens, CSIRO, Melbourne, 76, 2011.

Pokhrel, P., Robertson, D. E., and Wang, Q. J.: A Bayesian joint probability post-processor for reducing errors and quantifying uncertainty in monthly streamflow predictions, Hydrol. Earth Syst. Sci., 17, 795-804, 10.5194/hess-17-795-2013, 2013.