Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-84-AC2, 2016

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "A statistically based seasonal precipitation forecast model with automatic predictor selection and its application to Central and South Asian headwater catchments" *by* Lars Gerlitz et al.

Lars Gerlitz et al.

lars.gerlitz@gfz-potsdam.de

Received and published: 3 June 2016

Dear Mr. Barlow,

thank you very much for your comments on our manuscript. Especially your remarks on the quality of gridded precipitation data sets and the model evaluation are highly appreciated.

Please find enclosed our response as well as some suggestions for improvement.

##

Major comments: 1. Data quality. If the methodology is applied in an automated way to a number of different regions, how can data quality issues, which can vary considerably from location to location, be dealt with? And specifically for the case of Central Asia, the number of reporting stations varies dramatically over the 1948-2014 period considered here. I think the authors need to comment on both the general issue and provide some more information for the specific cases of Central and South Asia (e.g., plot the number of reporting stations as a function of time and assess the sensitivity of their results to the number of stations).

We totally agree on the fact, that the quality of gridded precipitation often does not suit the demands of statistical analyses. Especially for Central Asia a distinct decrease of meteorological stations after 1990 has frequently been reported. In the revised manuscript we will better comment on the general data quality issues and provide some more information for the selected target areas. Further we suggest to include a homogeneity analysis in the model results, which can serve as a basis for the data quality assessment and the model interpretation.

2. False positives if automated. Additionally, if the method is run for a large number of locations, some regions will get high prediction skill purely by chance. (If, say, a 95% significance criterion is applied for the validation period for each location, approximately 5% of the locations will appear significant by chance.) How would this issue be dealt with?

In order to avoid the use of potential predictors, which are only correlated by chance, we splitted the predictor selection and the model calibration into two (independent) parts. "In order to avoid overfitting and to develop robust regression relationship, the model calibration is based on the second random sample and thus is independent from the predictor selection procedure" (p7l33). The potential predictors are based on the correlation analysis and will certainly include false predictors. E.g. not all of the predictor clusters shown in Fig. 2 will have a real forecast potential. The final regression tree algorithm however only selects those variables, which also explain some part of the variance of the independent second sample. Thus the two step predictor selection procedure should deal with that problem.

3. Forecast correlation magnitude. I'm somewhat confused by Table 2. Are the correlations for the training period or for the evaluation period? And is the seasonal cycle included when calculating the correlation or is it removed first? If not removed, then numbers for when it has been removed should also be shown. If I'm reading the table correctly, there are several forecast correlations between 0.7 and 0.86 – I'm not aware of any forecast correlations for precipitation (with seasonal cycle removed) that are anywhere near that high for any region using any forecast method. As an example, it appears that the forecast correlation for Naryn is 0.86 for JFM forecast from Dec. As far as I know, that's also considerably higher than any potential predictor for the region (SSTs, lagged precipitation, etc.). If I've read that correctly, that's a rather extraordinary result that will require extra evidence to be considered plausible – perhaps by identifying a few individual high-correlation predictors and showing that they are linearly independent. It would also be useful to put those numbers into the context of other reported forecast skill for the regions, especially from the usual seasonal forecast centers, and of the skill of a pure persistence forecast.

The assessment of the forecast skill is an important task, which we already intensively discussed in our working group. The results which are shown in Tab. 2 are indeed calculated after removing the seasonal cycle. The correlation for DJF was e.g. calculated under consideration of those winter values only. However, as stated in the manuscript, the results should be interpreted with care, since the validation is only conducted for the independent time series, which includes 10 years of observations. "Although, the analysis based on 10 years only might be insufficient for the precise quantification of statistical model skill, we assume that a general assessment of the model quality is feasible. We abstained from the implementation of a cross-validation procedure due to the high computational demands of the predictor selection routine" (p9l17). Particularly for Naryn, the precipitation variability during the evaluation decade is strongly

linked to ENSO. Thus, the correlations might be distinctly lower in other decades, when the variability of ENSO is less pronounced.

We suggest to better comment on those issues in the manuscript. Further, we could conduct a split sample test, i.e. apply the predictor selection, model calibration and evaluation to shifted time periods. E.g. a second model evaluation based on the period from 1995 to 2004 would be feasible.

A disadvantage might be that this procedure would lead to a second large table and a duplication of Fig. 6, which might be slightly confusing for the reader. So we would be grateful for a final recommendation concerning the model validation strategy,

4. SST relationship for Central Asia. For the March SST correlations shown in Fig. 2, I don't understand why there is no signal at the equator in the central Pacific – I was expecting an ENSO pattern (and that is also what I get if I do a quick correlation based on GPCP data).

The attached figure shows the complete map of correlations between March precipitation time series in Naryn and the SSTs in previous January. Indeed, the entire ENSO region is positively correlated. However, highest levels of statistical significance are reached in those clusters which are highlighted by the polygons. Likewise the Warmpool region around Indonesia certainly reflects the ENSO signal. In order to avoid confusion, we will clearly comment on that point in the revised manuscript.

Minor comments: 1. I found the use of "exemplarily" to be somewhat distracting. I would suggest something more like "the model was applied to two test cases" or "two example cases." If the two regions really are exemplars, what makes them particularly useful or representative of the approach? Were other regions considered and, if so, why were they not included?

We will rather use the terms case studies or test cases in the revised manuscript. Although the study has been conducted in the framework of a research project dealing with the forecast of hydro-climatological conditions in Central Asia, we aim at the development of a statistical model which can be easily transferred to other regions.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-84, 2016.

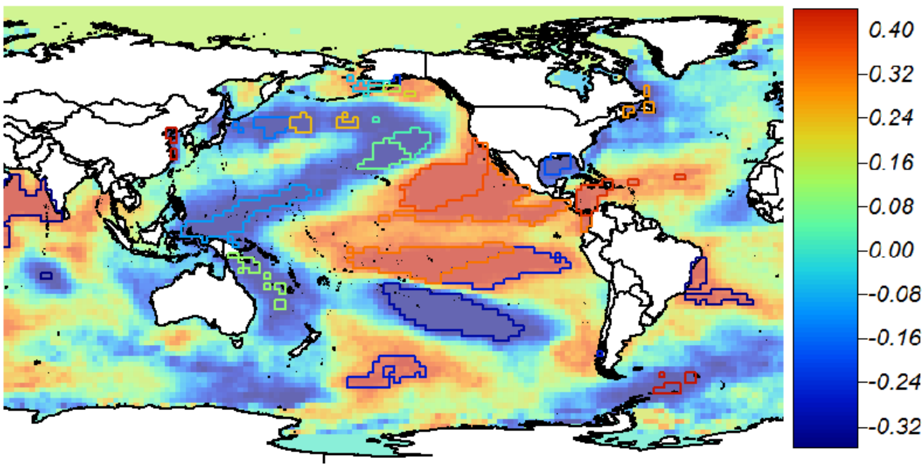

Fig. 1.