Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2017-183-RC1, 2017

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Regression-based season-ahead drought prediction for southern Peru conditioned on large-scale climate variables" *by* Eric Mortensen et al.

Anonymous Referee #1

Received and published: 18 May 2017

The paper 'Regression-based season-ahead drought prediction for southern Peru conditioned on large-scale climate variables' by Mortensen et al describes a statistics-based system for seasonal prediction of droughts in Peru. The results are interesting and are relevant to society.

Does the paper address relevant scientific questions within the scope of HESS? yes Does the paper present novel concepts, ideas, tools, or data? not really Are substantial conclusions reached? not really Are the scientific methods and assumptions valid and clearly outlined? yes Are the results sufficient to support the interpretations and conclusions? yes Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)? can

be improved. Do the authors give proper credit to related work and clearly indicate their own new/original contribution? Can include more references and cite more previous work on statistical techniques. Does the title clearly reflect the contents of the paper? yes Does the abstract provide a concise and complete summary? yes Is the overall presentation well structured and clear? can be improved with more well-defined data and methods sections with more details. Is the language fluent and precise? yes Are mathematical formulae, symbols, abbreviations, and units correctly defined and used? yes Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated? yes - some repetition. Are the number and quality of references appropriate? could cite some more work with similar techniques or addressing the same type of problems. Is the amount and quality of supplementary material appropriate? NA

Details.

p. 4, L. 1. Monthly precipitation (totals or means?) were derived from presumably daily rain gauge data. It is interesting to look at the number of days per month with precipitation, as the statistical sample of precipitation amounts involves a small (hence greater sampling fluctuations and less well defined mean estimates) in regions with few wet days. To get larger samples, one may use seasonally (3 months) or annually aggregated statistics.

Precipitation may be regarded as having two types of statistical distributions: for dry and for wet days. The dry-day statistics is trivial (zero), whereas the wet-day distribution is often described with the gamma distribution (of exponential for a simple approximation). The classification of the data into 'dry' and 'wet' makes sense because different physical conditions are present when it rains and when it doesn't.

p. 4. L. 2: 'multi-regression' should perhaps be 'multiple regression' or 'multivariate regression'

p. 4. L. 6-11: Please specify if it is version 1 of the reanalysis. Also, the time period

covered and the area selected are important. It is important to have sufficient information so that the analysis can be replicated by others independently. Some of this is discussed further down, but it may be easier for the reader if this is provided in a methods section before the results.

p. 5. L. 1. EOFs often refer to principal components analysis (PCA) of gridded data, weighted by the gridbox area. PCA is mathematically the same thing, but a term used more generally than EOFs. However, this is a matter of taste.

p. 5. L. 8. How much of the variance do the subsequent modes capture, and presumably the second order suggests a bi-pole type pattern? There is no need to show this, but perhaps worth describing its character. It is interesting that the leading PC so closely reproduces the station (not area?) mean precipitation,. What does that suggest? That the precipitation is dominated by large-scale climatic phenomena (at least aggregated over 3 months) and that other modes are essentially regional perturbations from the large-scale precipitation? I think it may be worth commenting these aspects, but perhaps later in the discussion.

Unless the precipitation has been gridded to onto a regular mesh, the term 'area average' should be replaced with 'station average'.

p.6, L. 10. Perhaps the higher modes of the PCA shows the orographic effects.

P. 8. figure 6: the scatter suggests that more than one factor affects the precipitation, but there is also a discernible anti-correlation between Nino3.4 and the precipitation. An ordinary linear regression can quantify the relationship (and associate a p-value), as can the correlation coefficient. This can be repeated with a subset of the data where the three outliers are excluded to estimate how exceptional they were.

p. 11, L11-16. subtracting 9 driest season during El nino Years from the 9 wettest from La Nina years is bound to produce an ENSO signal by design of the analysis. This paragraph seems to be repeated on p. 12, L. 1-5.

p. 13., L. 10.14. Is is evident that the PCA applied to area means according to Table 1., but this should also be stated more clearly in the main text. Also, the text should explain how different units are handled - was the PCA applied to standardised indices?

p. 14, L. 16-17. It is not clear how hindcasts were generated from residuals from cross-validation. Please elaborate.

p. 19, L.6-8. From a sample of test results, one will expect som to score high from pure chance. If there are 100 stations, one would expect 5 to score above the 95% confidence level by chance - this is how the confidence interval is defined.

split discussions and conclusions into two sections. The conclusions should be brief and repeat the main findings.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-183, 2017.