Hydrol. Earth Syst. Sci. Discuss., 11, C3884–C3889, 2014
www.hydrol-earth-syst-sci-discuss.net/11/C3884/2014/

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Seasonal predictions of agro-meteorological drought indicators for the Limpopo basin" *by* F. Wetterhall et al.

Anonymous Referee #2

Received and published: 18 September 2014

Review of: Seasonal predictions of agro-meteorological drought indicators for the Limpopo basin Authors: Wetterhall et al. Recommendation: Accept subject to major revisions

General Comments: The paper presents a relatively simple post-processing approach for ensemble seasonal forecasts that have agricultural significance, and I do not doubt the basic conclusions that the post-processing improves the forecasts. The paper as a whole is relatively brief compared to most HESS papers, and lacks the typical depth of analysis and detail, but the results are practical and likely to be of interest to the forecast applications community. To reach a standard that is acceptable for publication, however, the authors must provide additional detail for various aspects of their approach that are unclear. In particular, the QM must be explained in more detail, and

I strongly suggest that the authors either bolster their rational for applying the thresholds as they do, or evaluate a suggested variation (see comments below). In addition, further diagnosis of the results is warranted, given that some aspects (such as the disparity in performance in Fig. 5) are difficult to understand.

I would strongly urge the authors to bring the paper up to the density of a typical HESS article through the addition of further visualizations of the results (including, perhaps, timeseries comparing each years predictions and observations for the main predictands, among other analyses) – such displays offer different insights than those captured in the summary metrics alone. The more general discussion of the value of forecasts is useful and appropriate, and probably could be left as is.

Specific Comments to the Authors:

865-23: 'for shorter lead times, such as . . .' (clarify)

Table 1: Comparisons might be better presented as %bias, so that the reader doesn't have do the arithmetic in his/her head.

Figure 3: This figure might be improved by an X-axis that shows increments of month?

867-10: The detail on the modeling system is perhaps more than needed, but as long as detail is being given, please include the strategy to create ensemble members. Are their atmos components initialized at various lags from time zero, eg?

868-28: The refs given for the QM bias-correction approach all relate to climate change applications. Perhaps better refs would be Voisin et al (2010), where it was applied to medium range forecasting, or an earlier seasonal forecasting paper, Wood et al JGR (2002). ** Voisin, N., Schaake, J. C., and Lettenmaier, D. P.: Calibration and Downscaling Methods for Quantitative Ensemble Precipitation Forecasts, Weather Forecast., 25, 1603–1627, 2010. ** Wood, A. W., E. P. Maurer, A. Kumar, and D. Lettenmaier, Long-range experimental hydrologic forecasting for the eastern United States, J. Geophys. Res., 107(D20), 4429, doi:10.1029/2001JD000659, 2002.

868-24: ‘a simple mean monthly bias correction would not correct biases’ – I disagree – if all daily rainfall values in a month were scaled in the process of correcting monthly means, the step would alter daily rainfall amt distributions and could indirectly improve such biases. Also, please clarify: Are quantile mappings applied to each grid cell separately or to are the CDFs created for all cells in the domain? Explain whether the forecast and obs rainfall distributions are uniform enough across the domain to justify this choice if so. Could there be spatial variations in bias that would cause one correction applied across the whole domain to have suboptimal performance?

869-4: perhaps you mean quantiles 0.5-99.5? I don’t believe there is a unique 100th quantile in theory, at least if it is taken to mean non-exceedence percentile.

869-6: filter -> ‘rainy day threshold’ or some other more descriptive term; filter is a little unclear.

869-7: describe the problem that this step (bootstrapping followed by averaging) solves – ie, a single empirical CDF mapping can be jagged due to inadequate sampling?

869-7: what is done with SYS4 forecast values that lie beyond the extremes of the SYS4 CDF? Or because the mapping is done only for forecasts that contributed to the SYS4 CDF, this does not occur? In real-time application, this might have to be addressed unless the CDFs were recalculated each forecast time to include the latest.

869-29: it would be helpful to add one sentence that describes why biases in features like dry spells and precip intensities are biased in NWP or climate forecast model predictions.

870-5: grammar – phrase starting with ‘Firstly’ is not a sentence; perhaps remove ‘the fact that’

870-13: evaporative cooling from intercepted canopy moisture and increased humidity do, however, influence the crop – though the roots do not receive water. Minor point.

870-20: If a flow chart of the processing can be given, including both the forecasts &

obs, and the BC and the thresholding, that would be helpful in illustrating the experiment. Also, “This results for each lead time and for each year of available data, in. . .” – perhaps also for each threshold? Are the thresholds applied identically to forecasts and obs pcp? If so, clarify the logic. I would think a more obvious strategy is to apply it only to the forecasts as a ‘calibration’ step for dry-days/spell lengths, to be verified against the obs (EGPCP). If at the same time the obs are adjusted, the dry-day calibration has a moving target, and the skill of the SYS4 may not be optimized. Clearly the EGPCP at large scales will have biases in dry day and other pcp characteristics relative to local observations and crop impacts, but massaging and interpreting that relationship is a different topic and objective.

871-12: ‘poor man’s ensemble’ is a somewhat jargony way of saying a climatological ensemble, composed of members drawn from the historical observations for the forecast calendar period. I have seen the phrase used in other ways in forecasting (ie, to denote an ad-hoc collection of single-value forecasts such as control runs from different NWP models). Perhaps it’s not quite correct here?

872-1: ‘distribution’ might be clarified by the qualifier ‘magnitude’ or ‘amount’, since in the experiments, temporal distribution is also a feature of interest.

3.1: I did not expect the QM alone to affect the dry spells because the text earlier gave the impression that only forecast rainy day amplitudes were corrected (thus zero rainfall forecast days remained zeros, and the precipitation frequency would be unadjusted). The next few comments go back to the earlier section to suggest further corrections, the first of which is organizational.

2.3: this section presents both forecast data description and a method description – it should stick to the former and end on line 14.

868-15: the remainder of section 2.3 should be moved into one devoted to ‘forecast improvements’ or some appropriate title – which would also include the dry day threshold text. It could be a two part section on ‘amplitude correction’ and ‘calibration of dry day threshold'. A possible third or final subsection could be 'experimental design', basically including material after 870-20, describing the various trials that you assessed, and including the schematic of the flow of data & adjustments and end-point assessments (dry day, dry length).

869-5: the statement that only rainy days were corrected appears to be misleading, as it somehow also changed the frequency of rainy days, hence dry spell length. The CDF mapping must have been allowed to translate rainy to dry days or vice versa, to change their intermittency, correct? The typical problem is that the forecast model drizzles, and the frequency of non-zero days is too small. This is simply handled by applying QM to the entire distribution including zero-precip days – in which case some of the model's rainy days map to zeros in the lower quantiles of the obs CDFs. In the alternate case, in which the model is too frequently dry, zeros with a non-unique quantile (eg every quantile below obs (1-prob. of precip)) must be mapped to a quantile range that contains both zero and rainfall amts. This can be done with random estimates of quantiles within that range. In any case, it's not clear from the text how this issue is handled, and the authors must supply more detail. Related to this point, if only rainy days are mapped, why do the CDFS in Figure 3 (top row) not include the full probability range from (0,1)? My guess is that the QM quantiles were established using all days, but only the rainy day portion is shown, which would also explain the altered frequencies.

873-5: there is something missing in this sentence. "after the performance is comparable to the other areas"??? Can the authors give more diagnosis of the spatial variation in results? What is it about the distributions or forecasts in different areas that gives such a variation?

Table 1 – can the authors explain why the frequency of dry spells forecasted becomes worse with the correction for the first 3 lead times, despite the fact that their length forecast has improved? Perhaps plotting the mean of the forecast ensemble for each metric versus the obs as a scatter plot would illustrate some basic features of the impact

of the correction and the thresholding.

Fig 5 – I don't see intuitively why the skill scores for the raw forecasts with a correction fall so dramatically as the threshold rises, and this should be diagnosed and explained. The CDFs suggest that rainfall amounts are not badly biased (most correction factors not that far from 1). Please diagnose this result more completely. In addition, it would be worth comparing to the case in which the threshold is applied only to the forecast (both with and without QM) for verification against the non-thresholded obs. That, I think, would resemble most other applications of post-processing that I've seen.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 11, 861, 2014.