Hydrol. Earth Syst. Sci. Discuss., 9, C3343–C3349, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C3343/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Long-range hydrometeorological ensemble predictions of drought parameters" *by* F. Fundel et al.

A. Van Loon (Referee)

anne.vanloon@wur.nl

Received and published: 25 July 2012

Dear editor, dear authors,

Herewith, I provide a review of the original draft manuscript entitled "Long-range hydrometeorological ensemble predictions of drought parameters" by Fundel et al.

I found the study interesting and I potentially a valuable contribution to drought research and especially to drought forecasting. I especially like your evaluation of the use of these forecasts for various users. I have a number of comments related to the study in general and some more specific comment related to the way in which the work is presented. Additionally, I include some technical comments to assist the authors in

revising their manuscript.

1 General comments

It is important to be very clear and consistent in terminology. You use the term “low-flow events” for periods in which streamflow is below a seasonally-varying threshold. Deficiencies in the high-flow season (in the studied catchment the snow melt period) should, however, not be called low-flows as streamflow can be quite high in absolute terms even though it is below the threshold. Therefore, I advise you to use a more appropriate term, such as for example “drought”, “anomaly” or “deficiency” instead of “low-flow”. See Tallaksen and Van Lanen (2004). Please change it throughout the entire manuscript.

In this study, the longest drought in the forecast period is taken as the event for further analysis. Why did you use the longest event and not the most severe or most intense or . . .? Choosing the longest event gives problems with truncation of the event by the end of the forecast lead time. This is clearly shown in Fig.2, where you illustrate the threshold level method. The duration of the drought event that is chosen for further analysis (in this case the event on the right-hand side of the figure) is highly influenced by the maximum lead-time of 32 days. The severity is less influenced and the intensity (maximum deviation from the threshold) even less. So taking the most severe or most intense drought, instead of the longest, would give less impact of the used methodology on the results of this study. Can you please indicate what the effect of a different selection criterion for drought would have on the results of your study?

Additionally, the duration of a drought event is, in your study, also highly dependent on whether or not the streamflow signal reaches just above the threshold. The occurrence of small peaks that divide a long drought into two separate droughts strongly determines which drought is chosen and what the characteristics of that event are. You

HESSD
9, C3343–C3349, 2012

already mention that effect as a problem when you argue the use of a larger catchment (p.6861). A normal procedure in drought research to avoid this strong impact of small peaks is pooling. Various methods are possible, e.g. moving average, inter-event time/volume criterion (Fleig et al. 2006). Please consider applying a pooling method prior to selecting the drought event for further analysis.

In the manuscript, I miss some detailed examples of the performance of drought forecasting. Fig. 7 only gives a very general overview and, according to my view, only limited skill. I would like to see the forecasting skill for two or three cases, for example the 2003-drought (mentioned in the manuscript, but not illustrated) and a drought event/period in the 1990's. For these detailed examples then also the forecasting of timing (i.e. onset and termination) of the drought can be evaluated, because these are quite important features of drought and especially termination is very hard to predict due to the high persistence of drought.

Finally, I would like you to explore any seasonal differences in forecast skill. As winter droughts are caused by very different processes than summer droughts (Van Loon and Van Lanen, 2012), it would be very interesting to see whether one of them is easier to predict. That gives an indication which processes should be improved in the modeling framework to improve forecasting skill of droughts.

2 Specific comments

Abstract

Please provide more information on results (including numbers) in the abstract.

Introduction

You mention the work of Wood et al. (2002), Luo and Wood (2007) and Li et al. (2008) as studies that use a coupled atmosphere-hydrological model for the long-range pre-

HESSD

9, C3343–C3349, 2012

diction of drought. They managed to give reasonable predictions up to several months in advance. You do not clarify what your study adds to these results. What is new in your research? Even more because you evaluate forecasts only up to one month, which is less than in the studies mentioned above.

Data and methodology

Section 2.1: Please give more quantitative information about the catchment. What are "relatively cool conditions"? Give yearly average and minimum and maximum monthly temperature.

Section 2.3: Please give more information about the model. Provide a short summary of the papers you mention on PREVAH physics, parameterization and downscaling, and the papers on the calibration and verification against observations. Also mention the dates for the co-called "extended reference period" and the details of the "meteorological surface observations" (What? Where? When?). Also describe somewhere in this Section, or in Section 2.1, the runoff gauge in Andelfingen (should not be introduced at p.6867). Is data of this gauge used for calibration? Are observations of state variables used in calibration? And please provide the Nash-Suthcliffe value of logQ besides the mean error, as this metric is much used for evaluating model performance on low-flows.

Section 2.4: Here, you mention that a seasonally varying threshold was used, but Fig.1 and 2 seem to show a daily varying threshold and in the caption of Fig.7 you mention a monthly threshold. How was the threshold calculated? Did you define seasons by date? In this section you should describe that the quantile used for calculation of the threshold is selected later, based on your results, and what the criteria for selection were.

Furthermore, you state that the "lead-time is no longer a possible source of forecast error". This is not correct, because the limited lead-time causes a truncation of drought events, and therefore influences prediction of the longest drought (Fig.2).

HESSD

9, C3343–C3349, 2012

Section 2.5: Please define above which score you regard a forecast to be skillful/ beneficial. From Section 3.1 I understand that you denote a forecast as skillful above 0.55 or 0.6.

Results

Section 3.2: Why is the threshold quantile chosen based on forecasting results? For a real forecast this cannot be done, because no observations are available to test which quantile gives best results. Furthermore, in this way it is not related to any user requirements. Please discuss this issue in your manuscript. In the final choice of the 15th quantile, you mention that it is a compromise between the number of drought events and the drought forecasting skills. Why is the number of events such an important issue? If it is so important, mention it already in Section 2.4 and include the number of events as an extra column in Fig.5.

Furthermore, you say that the threshold shows a minimum in October/November, when snow accumulation starts. So, in this catchment winter temperatures are just below or around zero in winter, so that occasional melt takes place? Or is this minimum related to seasonality in precipitation and/or evaporation? If temperatures would be far below zero in winter (like in the Scandinavian countries), then the threshold would decrease in winter and show a minimum just before the snow melt peak. In Fig. 1 there is a second minimum in Februari. Please clarify what the causes are for seasonal changes in the threshold.

Section 3.3: “It is striking how well the duration and severity of the observed low-flow is contained within the range of the ensemble”. Be careful! I would not consider the resemblance strikingly well, especially not during the periods without events in observations.

Section 3.4: “For all users, value scores >50

HESSD

9, C3343–C3349, 2012

3 Technical corrections

p.6858, line 17: Do not use the term “indicators” for drought types. Rather use “types” or “processes”.

p.6859, line 28: “although” should read “however”

p.6860, line 1-4: Sentence not clear, consider revising.

p.6861, line 18 20: Please explain the difference between the 954 ensemble forecasts and 5 members in the reforecast. I, as a layman in forecasting, cannot understand this.

p.6863, line 1-3: Provide references in chronological order (also throughout the rest of the manuscript).

p.6863, line 20: “A lead-time dependency of the low-flow threshold was implemented for the forecasts.” What do you mean? Please rephrase.

p.6865, line 11-12: Please explain what you mean with “continuous observations” and “dichotomous observed outcomes”.

p.6865, line 18-19: Move these sentences to Introduction.

p.6866, line 13: “relative absolute error”?

p.6866, line 15: What do you mean with “beneficial”?

p. 6867, line 22, “varying quantiles”: Please rephrase. Quantile stays the same over the year, but threshold itself varies.

p.6870, line 24: Sect. 2.3 > Sect. 2.4

p.6870, line 25: by > be

figures

Fig.1: Emphasize line of Q15 (chosen quantile for threshold)

HESSD

9, C3343–C3349, 2012

Fig.2: Move information in caption on calculation of severity, magnitude and timing to main text.

Fig.3: Give more information in caption. What do we see? Is it the forecasting score using different quantiles as threshold? And what happened to the ensemble? Is this the ensemble mean? Or the best prediction? Or . . .?

Fig.4: Please show also the lines that are now hidden behind a polygon. Resampling of 1000 times was not mentioned in main text (Methods).

Fig.8: What is on the x-axis?

Fig.10: What are the white and grey dots?

4 References

Fleig, A. K., Tallaksen, L. M., Hisdal, H., and Demuth, S.: A global evaluation of streamflow drought characteristics, Hydrol. Earth Syst. Sci., 10, 535–552, doi:10.5194/hess-10-535-2006, 2006.

Tallaksen, L. M. and Van Lanen, H. A. J.: Hydrological drought: processes and estimation methods for streamflow and groundwater, Developments in Water Science 48, Elsevier Science B.V., The Netherlands, 2004.

Van Loon, A. F. and Van Lanen, H. A. J.: A process-based typology of hydrological drought, Hydrol. Earth Syst. Sci., 16, 1915–1946, doi:10.5194/hess-16-1915-2012, 2012.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 6857, 2012.