Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-745-RC1, 2018

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "The effect of climate on timescales of drought propagation in an ensemble of global hydrological models" *by* Anouk I. Gevaert et al.

Anonymous Referee #1

Received and published: 15 January 2018

Review of "The effect of climate on timescales of drought propagation in an ensemble of global hydrological models" by Gevaert et al.

In this study a model ensemble is used to characterize the propagation timescales from meteorological drought to soil moisture drought and to hydrological drought. The propagation times are analysed in respect of climate, season and model type and evaluated with observed streamflow. This is an interesting, well-structured study, with some substantial conclusions that will be a useful addition to literature. The supplements, tables and figures (apart from Figure 3, see comments below) are self-explaining and appropriate. However, I have three main issues with the methodological approach, that

Printer-friendly version

Discussion paper

should be addressed before publication.

Main issues: The authors chose for their analyses eight accumulation periods that represent different timescales (1, 2, 3, 6, 9, 12, 24 and 36 months for sub-seasonal, seasonal and annual timescales). These accumulation periods are similar to those that are often used, but still arbitrary. For example, they could have chosen only 1, 3, 6, 12 and 24 months for the same reasons. For the determination of propagation timescales this choice is probably of minor relevance, however, for the applied statistical tests I think it can have quite some impact. The tests used in this study are designed for variables on the interval scale (apart from spearman's rho) but the variables are on ordinal scale. The authors are aware of that problem and state they "assume that the difference between accumulation periods of 12 and 24 months (. . .) to be equivalent to the difference between 1 and 2 months" (p.5, l.5ff). Nevertheless, it is still very relevant to check whether there is an influence of the arbitrary choice of accumulation periods and the related assumption on the results of the statistical tests. Additionally, it needs a strong rationale for using tests designed for interval scaled variables instead of tests appropriate for ordinal scaled variables (e.g. the chi-squared test of independence).

The second main issue is about the way the model ensemble mean is calculated: "The model ensemble mean was calculated as the average of the SIs" (p.6, l.29). A very important reason for using standardised indices is to ensure that all time series have the same distribution and are directly comparable (see e.g. Bloomfield and Marchant, 2013; Kumar et al., 2016). Averaging two or more timeseries, that have a standard normal distribution, will lead to a timeseries which distribution has a smaller standard deviation that might favour certain (higher) SPI-n. Moreover, the comparison with the results from the original model time series as it is carried out in chapter 4.3 is not really "fair" anymore, since time series are not directly comparable. The correct way is to average the raw model outputs first and standardize afterwards all seven time series plus the model ensemble mean.

Finally, the authors use an explanatory analysis to identify relevant model characteristics causing differences in drought propagation timescale (p.15f). They are aware of the difficulties using only seven models for that and the problem of collinearity between the groups. In fact, these limitations inhibit any useful result. For example, the factors GHM/LSM and (no)reservoir are highly correlated. Based on Table 1, it is only the model W3RA that is classified into another group. That means, in a study without this model, the groups would have been identical, similar to what is reported about the snow scheme. Accordingly, the graphs in Figure 7 of the two groups have a very similar shape. The authors wonder about the reason for the high influence of reservoirs on soil moisture (p.16, l.12), but the real problem is, that both factors (GHM/LSM and (no)reservoirs) represent the combined effect of (no) reservoirs, GHM/LSM, snow scheme and probably several other relevant model structures. As it is not possible to relate the differences of the groups to one model structure we cannot learn much from this analysis.

Other points the authors might want to look at: In the introduction the authors acknowledge that an important component of drought propagation is the time lag (p.2, l.13ff). However, time lags are not considered in the analysis but listed to be important for future research (p.19, l.16). Including an analysis of time lags which also might differ for the models would increase the relevance of this study. If lags are not included, there should be at least a rationale for excluding them despite their relevance.

In chapter 4.1 analyses of the “mean SPI-n” are presented (e.g. p.10, l.17; caption of Figure 3). For me it does not become clear, whether this is really the arithmetic mean of the SPI-n or rather the mean of the ranks. For example in Figure 3: If there were the two accumulation periods of 1 and 36 months, is “mean SPI-n” (1+36)/2=18.5 or rather (1+6)/2=3.5? This is quite relevant for the plotted circles. If they are calculated as an arithmetic mean, it might be very hard to read the values from the plot due to the very non-linear y-axis.

Moreover, it is important to report somewhere the ‘sample size’, i.e. the absolute number of cells which are not masked for the different climates and drought types. Otherwise it is for example hard to understand, that the t-test leads to significant different SPI-n means for winter and summer in runoff of TMP (Figure 3).

On page 10, l.16 the authors describe the results of the ANOVA: “The means of SPI-n for winter hydrological droughts in continental and polar climates are not significantly different”. Again, for me it is not clear whether the rank mean or the arithmetic mean is meant here. However, more important is the fact that it sounds like two categories were directly compared to each other. In this case, it would have been a t-test rather than an ANOVA what was used. Please clarify, which variables were used for the ANOVA and in which cases a t-test was used.

The stations used for the evaluation against observations are distributed very uneven (as the authors write on p.18, l.7). In Figure 8 it looks like there were very few to no stations in the climates polar and tropical wet. However, the authors state that “errors between models and observations are not related to climate”. To enable the reader to comprehend this important finding, I think it is necessary to give more information on the number of stations per climate zone, the test used to reach this conclusion as well as the results of the test.

References: Bloomfield, J. P., & Marchant, B. P. (2013). Analysis of groundwater drought building on the standardised precipitation index approach. Hydrology and Earth System Sciences, 17, 4769-4787. Kumar, R., Musuuza, J. L., Van Loon, A. F., Teuling, A. J., Barthel, R., Ten Broek, J., Mai, J., Samaniego, L., and Attinger, S. (2016). Multiscale evaluation of the Standardized Precipitation Index as a groundwater drought indicator Hydrology and Earth System Sciences, 20, 1117-1131.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-745, 2018.