Hydrol. Earth Syst. Sci. Discuss., 8, C3237–C3241, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C3237/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Effects of seasonality on the distribution of hydrological extremes" *by* P. Allamano et al.

P. Rasmussen (Referee)

rasmusse@cc.umanitoba.ca

Received and published: 4 August 2011

The objective of this investigation is to demonstrate that ignoring the seasonal nature of daily precipitation processes can lead to erroneous estimates of design events. The authors assume a simple, idealized world where daily precipitation occurs according to a non-homogeneous Poisson process with seasonally varying parameter and where daily precipitation amounts follow an exponential distribution with a mean level that also varies seasonally. Specifically, sinusoidal variations in the Poisson and exponential parameters are assumed. Quantiles estimated from this parent process is compared to quantiles estimated from a model that disregards seasonality and simply fits a single exponential distribution to the pooled information.

The specific case of the peaks-over-threshold (POT) model with seasonal variation of parameters was studied in a paper by Rasmussen and Rosbjerg (1991) who – in a manner very similar to Allamano et al. – assessed the impact of pooling inhomogeneous seasonal exponential distributions into a single population that was then fitted by an exponential distribution. In terms of model formulation, the only real difference between the two studies is that Allamano et al. assume a sinusoidal variation of parameters whereas Rasmussen and Rosbjerg's paper assumed seasons with constant parameters. In either case, when ignoring seasonality a bias is introduced in the estimation of quantiles, because the combination of different exponentials is not an exponential. However, bias is not the only consideration in assessing the quality of an estimator. Rasmussen and Rosbjerg also assessed the standard deviation and the root-mean-square error (RMSE) of the estimator that ignores seasonality. The RMSE, which combines bias and standard deviation, may be a better measure of whether it is worthwhile taking seasonality into account or not. More specifically, pooling seasonal events into a single sample increases the ratio of the number of data to the number of model parameters which generally implies smaller sampling variability. This gain has to be balanced against the increased bias. The present manuscript looks exclusively at bias, as quantified by the RT ratio in Eq. 9. One should not be surprised to find biases when fitting an exponential distribution to a population that is not exponential.

Of course, a fundamental question is if nature is really exponential at the seasonal level. The answer is probably no since nature tends to be much more complex than what can be described by a 1-parameter distribution. The results in the paper are only truly applicable in the assumed simplistic world of seasonality. It is not clear how significant the results are in the real world where it could well happen that daily precipitation events pooled over the seasons are roughly exponentially distributed, despite the presence of seasonality. Indeed, the fundamental problem here is not the seasonality itself, but the fact that an incorrect distribution is being fitted to the non-seasonal sample.

The type of seasonality assumed in Eq. 2 may not fit real data very well and therefore limits the use of the results in practical applications. It is perhaps acceptable as an exploratory model, but one still wonders if a simpler version of seasonality could have been used. Although the sine assumption is relatively simple, it leads to a quite complicated analytical expression for the marginal distribution of daily precipitation amounts. The derivation of Eq. 6 is a nice analytical result, but the complexity and effort to derive it may obscure the real point the authors want to make.

The reparameterization of Eq. 1 is critical for the paper, but its rationale is not well explained. It would be instructive to provide some examples of how well the densities in Eqs. 1 and 3 can be matched since the reparametrization is only approximate. We plotted a few cases and found that for some parameters the agreement is very good while for others it is not. It seems that the larger the amplitude of the sine wave, the higher the discrepancy. A more troubling problem is that the discrepancy depends on α_0, with higher α_0-values leading to poorer agreement, even in relative terms. This implies that the agreement between Eqs. 1 and 3 is scale-dependent, or, in other words, that one would get different results depending on whether precipitation is measured in millimetres or in inches.

Detailed comments

P4790, L18. "Annual time scales" should be "seasonal time scales".

P4790, L25. " . . . could belong to different populations." This is an inaccurate statement and contrary to what the authors argue, this is not a problem for standard frequency analysis. If seasonality of the type considered in the paper is ignored, the population distribution is the one given in Eq. 6. A problem only arises if one tries to fit an incorrect distribution to this population.

P4791, L10-12. The statement about the hypotheses of identically distributed random variables for monthly and annual extremes needs to be explained and justified. It is not clear what the authors mean.

P4791, L12. The statement regarding estimation uncertainty of design values based on monthly maxima is incorrect. The ratio of the number of parameters to number of data should stay the same, so the modeling of separate seasons should not affect the sampling uncertainty.

P4792, L16. Use "cycles" instead of "peaks" in this sentence.

P4792, L24. Bayes' theorem is not used to obtain the marginal distribution. The marginal is obtained by simply integrating the conditional distribution.

P4793, L4. The expression for α'(t) is missing the term 2π/365.

P4793, Eq. 3. The dependence of this integral on x should be specified on the left-hand side.

P4794, Eq. 6. The middle expression should be divided by 365.

P4794, Eq. 7. How was this expression derived?

P4796, L20. Here the authors argue that statistical tests typically fail to recognize that the marginal distribution of exceedances is not exponential. We have not tried to reproduce the test results, but are curious about the results. For example, in Figure 2, with the case of λ=50 and a 20 year record, one would have on average of 1000 observations in a sample. This is a very large sample and goodness-of-fit tests should be able to pick up discrepancies from the assumed distribution. In fact, when samples are very large, statistical tests tend to become "too powerful" and are usually avoided. This is the well-known issue of "statistical significance" versus "practical significance".

P4799, L6. A value λ=20 and a record length of 74 years will give sample of 1500 events. Our guess is that there would be significant benefit in raising the threshold, because the lower the threshold, the more significant the problem of seasonality will be. For estimation of a, say, 100-year event, one is probably better off using only the 100 or 200 largest values.

P4799, L17. Remove expression in parentheses.

P4801, L7-9. This is only true if the additional events in the POT model are consistent with the assumed distribution. Change "sample dimension" to "sample size".

References

Rasmussen, P.F. and D. Rosbjerg (1991) Prediction uncertainty in seasonal partial duration series. Water Resources Research, 27 (11): 2875-2883.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 4789, 2011.