Hydrol. Earth Syst. Sci. Discuss., 7, C3082–C3086, 2010
www.hydrol-earth-syst-sci-discuss.net/7/C3082/2010/

Hydrology and Earth System Sciences Discussions

Interactive comment on "On the sampling distribution of the coefficient of L-variation for hydrological applications" *by* A. Viglione

A. Viglione

viglione@hydro.tuwien.ac.at

Received and published: 22 October 2010

I would like to thank Björn Guse and the anonymous referee for their positive reviews and their useful comments. While preparing the responses to their observations and discussing the paper with a friend (José Luis Salinas), I realised that my paper has a fundamental problem, which I think should be corrected. The problem does not regard the methods and the results, but the wording/notation used. I try to explain it briefly. The title of the HESSD paper is "On the sampling distribution of the coefficient of L-variation for hydrological applications", i.e., on the distribution of t, the sample L-CV. In reality I am not so much interested in the distribution of t itself, but in the estimation of τ, the population L-CV. In particular, I use the distribution of t to evaluate the confidence

intervals for τ. This is what I check by using the uniformity plots and the Anderson-Darling test: i.e., if the assumed distribution of t provides good confidence intervals for τ. Result: it does not, because in many cases t is biased respect to τ (and also the variance estimated with Elamir and Seheult, 2004, is biased). Therefore I correct these biases to obtain a distribution which gives better confidence intervals for τ, but which is no more the sampling distribution of the L-CV!

Therefore I have changed the title to "Confidence intervals for the coefficient of L-variation in hydrological applications". Also the abstract and part of the text have been changed in order to avoid the misunderstanding between the sampling distribution of t and the estimation of the confidence intervals for τ. The abstract now reads:

The coefficient of L-variation (L-CV) is commonly used in statistical hydrology, in particular in regional frequency analysis, as a measure of steepness of the frequency curve of the hydrological variable of interest. As opposed to the single-value estimation of the L-CV, in this work we are interested in the estimation of the interval of values (confidence interval) in which the L-CV is included at a given level of probability (confidence level). Several candidate distributions are compared in terms of their validity as estimators of confidence intervals for the population L-CV, i.e., by checking if the nominal coverage probability (confidence level) holds. Monte-Carlo simulations of synthetic samples from distributions frequently used in hydrology are used as a basis for the comparison. The best estimator proves to be the log-Student t distribution whose parameters are estimated without any assumption on the underlying parent distribution of the hydrological variable of interest. This distribution-free parametric estimator is shown to outperform also the non-parametric bias-corrected and accelerated bootstrap method. An illustrative example of how this result could be used in hydrology is presented, namely in the comparison of methods for regional flood frequency analysis. In particular, it is shown that the confidence intervals for the L-CV can be used to assess the amount of spatial heterogeneity of flood data not explained by regionalization models.

Moreover, the title of Section 3, "Goodness of fit method", has been changed to "Validity of the confidence interval estimators" and a figure has been added to show the validity of the method on the L-CV–L-CA space of Figure 1.

Responses to the comments of the reviewers follow:

1. *p. 5471, line 13: The superscript 'T' is not defined.*

   The sentence "and C is the $n \times n$ lower triangular matrix..." has been changed to "and C^T is the transponse of the $n \times n$ lower triangular matrix..."

2. *p. 5471, line 16: Consider a more detailed explanation of equation 4. Maybe by including an intermediate step.*

   I agree that the sentence appears complicated. The reason, to me, is that eq. (4) is introduced before defining the variance matrix of the L-moments. I have rewritten the paragraph by defining the estimator of Eq. (5) and the variance matrix Θ before equation (4).

3. *p. 5472, line 4: To improve the understanding of section 2, consider a subdivision of section 2 into two sub-sections. The sub-section 2.1 could end at this line with the equation 8, since this equation is an intermediate result which is an input for equations 9, 10, 12, 13 and 14, i.e. the five distributions. This could help the reader to understand the procedure and the results in a better way. For example, you could name the subsection 2.1 "variance of L-moment ratio" and 2.2 "selection of distributions" or something similar.*

   This is indeed a very good suggestion. I have used the following titles: 2.1 Variance of the sample L-CV; 2.2 Candidate distributions for the sample L-CV.

4. *page 5472, line 23: maybe it should be pointed out that the variance of the distribution of Eq. (10) is not var(t) because the variance of the Student distribution is not 1. The same consideration applies to Eq. (14).*

   The following lines have been added: "Note that the variance of the distribution defined by Eq. (10) is greater than $\widehat{\mathrm{var}(t)}$, since the variance of T_{n-1} is greater than 1."

5. *page 5474, line 9: change "N=10000 samples are" to "N=10000 samples of length n are".*

   Change made.

6. *page 5474, line 12: "should be uniformly distributed between 0 and 1". Please add a Reference (e.g., Laio et Tamea, 2007).*

   The sentence has been modified to: "The candidate distribution P of t can be consistently used to estimate the confidence intervals of τ if the probability of non-exceedance $P(\tau)$ is uniformly distributed between 0 and 1 (which is a consequence of the fact that $P(\tau)$ is the probability integral transform of τ; see e.g., Kottegoda and Rosso, 1997, Section 8.2.1)."

7. *p. 5474, line 19 and figure 2: In the text, the empirical cumulative distribution function is named Ri/N. In figure 2, it is denoted as F. Consider to use one variable name consistently.*

   I replotted Figures 2, 4 and 5 using Ri/N on the y-axis.

8. *Figure 6: Could you explain why some sites which are close to the regional value are outside of the significance level and other which are far from it are in the significance level.*

   Because each sample has its own confidence bounds. Therefore one (short) sample far from the regional value includes it because it has wide confidence bounds and, on the other side, a (long) sample close to the regional value does not include it because it has tiny confidence bounds.

9. *page 5476, line 25: in how many cases the estimate of var(t) by Elamir Seheult (2004) is negative?*

   In few of the cases for n=10. I have added the sentence "For very short samples ($n = 10$), in some few cases (1% of the cases for GEV and 3% for P3), the estimate $\widehat{\mathrm{var}(t)}$ of Elamir and Seheult (2004) is negative. In those cases, the Monte-Carlo simulations have been discarded when producing the curves in Fig. 2c and d. Therefore, the results for very short samples are even worse than what shown in the figure".

10. *p. 5477, line 18: The article is partly not concise in the use of gray (AE) and grey (BE). Please use BE or AE consistently.*

    Amended.

11. *p. 5482, line 15: A bracket is missing after (17).*

    Amended.

12. *p. 5486, line 4: Please change "HandBook" into "Handbook".*

    Amended.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 7, 5467, 2010.