Development of Bayesian GLS regional regression analyses for hydrologic models

Jery R. Stedinger[1], Dirceu S. Reis Jr.[2], Andrea G. Veilleux[3] and Jonathan R. Lamontagne[4]

(1) School of Civil and Environmental Engineering, Cornell University,Ithaca, NY, USA
(2) Department of Civil and Environmental Engineering, University of Brasilia, Brasilia, Brazil
(3) U.S. Geological Survey, Water Mission Area, Analysis and Prediction Branch, Reston, VA, USA
(4) Department of Civil and Environmental, Engineering, Tufts University, Medford, MA, USA

Regional regression analyses are used to estimate parameters of models of hydrologic statistics, such as flood and precipitation quantiles, or distribution parameters such as an LP3 coefficient of skewness or GEV-κ. They can provide Bayesian priors for use with other information. Ordinary Least Squares (OLS) regression does not provide an efficient regional estimator of a hydrologic statistic or the precision of estimated regression parameters because OLS fails to recognize that at-site estimates have different precision due to differences in record length and can be cross-correlated if records include concurrent cross-correlated observations or catchments are nested. Weighted Least Squares (WLS) and Generalized Least Squares (GLS) regression models are more appropriate. WLS accounts for differences in record length, whereas GLS should include estimated cross-correlations among estimators.

This paper reports on development of regional regression models: Bayesian-WLS, Bayesian-GLS and Bayesian-WLS/GLS estimators; such Bayesian analyses often provide a more realistic description of the model error variance than do Method-of-Moments (MM) and Maximum Likelihood (MLE) point estimators. A new set of GLS regression diagnostics include Bayesian Plausibility Values, pseudo adjusted R^2, a pseudo Analysis-of-Variance, and Bayesian metrics for Leverage and Influence (Reis et al., 2020). The Bayesian-WLS/GLS estimators (Veilleux and Stedinger, 2013) have now been used across 50% of the U.S. to develop statistically sound estimates of regional skew and its precision required for U.S. Flood Frequency analyses (England et al., 2018).

The paper reports performance of the B-GLS estimator in regional skew studies conducted by the US Geological Survey (USGS) for the states of California, Oregon, and other U.S. regions. In California and Oregon studies, B-GLS weights used to compute the regional skew appeared to be unreliable. Erratic weights were thought to be related to record length differences and poor estimates of cross-correlations. This paper examines that problem and the solution adopted. New metrics evaluate the stability and performance of GLS considering sets of record lengths and correlation matrices that reflect estimation uncertainties. The GLS results are demonstrated to be relatively robust.

References

England, J.F. Jr., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas Jr., W.O., Veilleux, A.G., Kiang, J.E., and Mason, R.R., Jr., 2017, Guidelines for Determining Flood Flow Frequency – **Bulletin 17C: U.S. Geological Survey Techniques and Methods book 4,** chap. B5, 148 p., https://doi.org/10.3133/tm4B5. March 2018.

Reis, Dirceu S., Jr., Andrea G. Veilleux, Jonathan R. Lamontagne, Jery R. Stedinger, and Eduardo Sávio P. R. Martins, Operational Bayesian GLS Regression for Regional Hydrologic Analyses, *Water Resources Research*, **56**(8), doi.org/10.1029/2019WR026940, 2020.

Veilleux, A.G., and J. R. Stedinger, Appendix. Regional Skewness Regression, In Eash, D.A., Barnes, K.K., and Veilleux, A.G., Methods for estimating annual exceedance-probability discharges for streams in Iowa, based on data through water year 2010: **U.S. Geological Survey Scientific Investigations Report 2013-5086,** 63 p. with appendix., http://pubs.usgs.gov/sir/2013/5086/, 2013.