Bayesian WLS/GLS Regression for Regional Skewness Analysis for Regions with Large Crest Stage Gage Networks

A.G. Veilleux[1], J.R. Stedinger[2], and D.A. Eash[3]

[1]Hydrologist, USGS Water, U.S. Geological Survey, 6000 J Street, Placer Hall, Sacramento, CA 95819-2605 USA; Email: aveilleux@usgs.gov
[2]Professor, School of Civil & Environ. Engineering, Cornell Univ., Hollister Hall, Ithaca, NY 14853-3501 USA; Email: jrs5@cornell.edu.
[3]Hydrologist, USGS Iowa Water Science Center, U.S. Geological Survey, 400 S. Clinton St., Iowa City, Iowa 52244 USA; Email: daeash@usgs.gov

ABSTRACT:

This paper summarizes methodological advances in regional log-space skewness analyses that support flood-frequency analysis with the log Pearson Type III (LP3) distribution. A Bayesian Weighted Least Squares/Generalized Least Squares (B-WLS/B-GLS) methodology that relates observed skewness coefficient estimators to basin characteristics in conjunction with diagnostic statistics represents an extension of the previously developed B-GLS methodology. B-WLS/B-GLS has been shown to be effective in two California studies. B-WLS/B-GLS uses B-WLS to generate stable estimators of model parameters and B-GLS to estimate the precision of those B-WLS regression parameters, as well as the precision of the model. The study described here employs this methodology to develop a regional skewness model for the State of Iowa. To provide cost effective peak-flow data for smaller drainage basins in Iowa, the U.S. Geological Survey operates a large network of crest stage gages (CSGs) that only record flow values above an identified recording threshold (thus producing a censored data record). CSGs are different from continuous-record gages, which record almost all flow values and have been used in previous B-GLS and B-WLS/B-GLS regional skewness studies. The complexity of analyzing a large CSG network is addressed by using the B-WLS/B-GLS framework along with the Expected Moments Algorithm (EMA). Because EMA allows for the censoring of low outliers, as well as the use of estimated interval discharges for missing, censored, and historic data, it complicates the calculations of effective record length (and effective concurrent record length) used to describe the precision of sample estimators because the peak discharges are no longer solely represented by single values. Thus new record length calculations were developed. The regional skewness analysis for the State of Iowa illustrates the value of the new B-WLS/B-GLS methodology with these new extensions.

INTRODUCTION:

For the log-transformation of the flood flows, Bulletin 17B [IACWD, 1982] recommends using a weighted average of the at-site skewness coefficient and a regional skewness coefficient to help improve flood quantile estimators. The Bulletin supplies a national map, but also encourages hydrologists to develop more specific local relations. Since the first map was published in 1976, some 35 years of additional information has accumulated, and better spatial estimation procedures have been developed [Stedinger and Griffis, 2008].

Tasker and Stedinger [1986] developed a weighted least squares (WLS) procedure for estimating regional skewness coefficients based on sample skewness coefficients for the logarithms of annual peak-discharge data. Their method of regional analysis of skewness estimators accounts for the precision of the skewness estimator for each station, which depends on the length of record for each station and the accuracy of an Ordinary Least Squares (OLS) regional mean skewness. More recently, Reis and others [2005], Gruber and others [2007], and Gruber and Stedinger [2008] developed a Bayesian generalized least squares (GLS) regression model for regional skewness analyses. The Bayesian methodology allows for the computation of a posterior distribution of both the regression parameters and the model error variance. As shown in Reis and others [2005], for cases in which the model error variance is small compared to the sampling error of the at-site estimates, the Bayesian posterior distribution provides a more reasonable description of the model error variance than both the GLS method-of-moments and maximum likelihood point estimates [Veilleux, 2011]. While WLS regression accounts for the precision of the regional model and the effect of the record length on the variance of skewness coefficient estimators, GLS regression also considers the cross-correlations among the skewness coefficient estimators. In some studies the cross-correlations have had a large impact on the precision attributed to different parameter estimates [Gotvald, 2009; Parrett and others, 2011].

Due to complications introduced by the use of the Expected Moments Algorithm (EMA) (see Cohn and others [1997]) and large cross-correlations between annual peak discharges at pairs of gages sites, an alternate regression procedure was developed to provide both stable and defensible results for regional skewness coefficient models [Veilleux, 2011]. This alternate procedure is referred to as the B-WLS/B-GLS regression framework [Veilleux, 2011; Veilleux and others, 2011]. It uses an OLS analysis to fit an initial regional skewness model; that OLS model is then used to generate a stable regional skewness coefficient estimate for each site. That stable regional estimate is the basis for computing the variance of each at-site skewness coefficient estimator employed in the WLS analysis. Then, B-WLS is used to generate estimators of the regional skewness coefficient model parameters. Finally, B-GLS is used to estimate the precision of those WLS parameter estimators, to estimate the model error variance and the precision of that variance estimator, and to compute various diagnostic statistics.

To provide cost effective peak-flow data for smaller drainage basins in Iowa, the U.S. Geological Survey (USGS) operates a large network of crest stage gages (CSGs) that only record flow values above an identified recording threshold (thus producing a censored data record). CSGs are different from continuous-record gages, which record almost all flow values and have been used in previous B-GLS and B-WLS/B-GLS regional skewness studies. Thus, while the Iowa regional skewness study described here did not exhibit large cross-correlations between annual peak discharges, it did make extensive use of EMA to estimate the at-site skewness coefficients and its mean square error. Because EMA allows for the censoring of low outliers, as well as the use of estimated interval discharges for missing, censored, and historic data, it complicates the calculations of effective record length (and effective concurrent record length) used to describe the precision of sample estimators because the peak discharges are no longer solely represented by single values. To properly represent these complications, modifications were made to the B-WLS/B-GLS procedure. The steps in this analysis are described below.

METHODOLOGY FOR REGIONAL SKEWNESS MODEL

This section provides a brief description of the B-WLS/B-GLS methodology. Veilleux and others [2011] and Veilleux [2011] provide a more detailed description.

OLS Analysis

The first step in the B-WLS/B-GLS regional skewness analysis is the estimation of a regional skewness model using Ordinary Least Squares (OLS). The OLS regional regression yields parameters $\hat{\boldsymbol{\beta}}_{OLS}$ and a model that can be used to generate unbiased and relatively stable regional estimates of the skewness for all gage sites:

$$\tilde{\mathbf{y}}_{OLS} = \mathbf{X}\hat{\boldsymbol{\beta}}_{OLS} \quad (1)$$

Here **X** is an (*n* x *k*) matrix of basin characteristics, $\tilde{\mathbf{y}}_{OLS}$ are the estimated regional skewness values, *n* is the number of gage sites, and *k* is the number of basin parameters including a column of ones to estimate the constant. These estimated regional skewness values $\tilde{\mathbf{y}}_{OLS}$ are then used to calculate unbiased at-site regional skewness variances using the equations reported in Griffis and Stedinger [2009]. These at-site regional skewness variances are based on the regional OLS estimator of the skewness coefficient instead of the at-site skewness estimator, thus making the weights in the subsequent steps relatively independent of the at-site skewness estimates.

WLS Analysis

A Bayesian Weighted Least Squares (B-WLS) analysis is used to develop estimators of the regression coefficients for each regional skewness model [Veilleux, 2011; Veilleux and others, 2011]. The WLS analysis explicitly reflects variations in record length, but intentionally neglects cross correlations thereby avoiding the problems experienced with GLS parameter estimators [Veilleux, 2011; Veilleux and others, 2011].

GLS Analysis

After the regression model coefficients, $\hat{\boldsymbol{\beta}}_{WLS}$, are determined with a WLS analysis, the precision of the fitted model and the precision of the regression coefficients are estimated using a Bayesian Generalized Least Squares (B-GLS) analysis [Veilleux, 2011; Veilleux and others, 2011]. Precision metrics include the standard error of the regression parameters, $SE(\hat{\boldsymbol{\beta}}_{WLS})$, and the model error variance, $\sigma^2_{\delta,B-GLS}$, pseudo R^2_δ as well as the average variance of prediction at a gage site not used the regional model, AVP_{new}.

DATA ANALYSIS:

Data for Iowa Regional Skewness Study

This study is based on annual peak flow data from 273 stream flow gage sites in Iowa and the surrounding states. The annual peak flow data were downloaded from the USGS National Water Information System: Web Interface (NWISWeb). In addition to the peak flow data, over 65 basin characteristics for each of the 273 sites were available as explanatory variables in the regional study. The basin characteristics available include percent of basin contained within different hydrologic regions, as well as the more standard morphometric parameters such as location of the basin centroid, drainage area, main channel slope, and basin shape among others.

At-Site Skewness Estimators

In order to estimate the at-site log10 skewness, G, and its mean square error, MSE_G, the analysis used the expected moments algorithm (EMA) [Cohn and others, 1997; Griffis and others, 2004]. EMA provides a straightforward and efficient method for the incorporation of historical information and censored data, such as those from a crest stage gage, contained in the record of annual peak flows for a gage site. PeakfqSA, an EMA software program developed by Cohn [2011], is used to generate the at-site log10 estimates of G and its MSE_G, assuming an LP3 distribution and employing a Multiple Grubbs-Beck test for low outlier screening.

Pseudo Record Length

Because the data set includes censored data and historic information, the effective record length used to compute the precision of the skewness estimators is no longer simply the number of peak flows at a gage site. Instead, a more complex calculation should be used to take into account the availability of historic information and censored values. While historic information and censored peaks provide valuable information, they often provide less information than an equal number of years with systematically recorded peaks [Stedinger and Cohn, 1986]. The following calculations provide a pseudo record length, P_{RL}, which appropriately accounts for all peak flow data types available for a site. P_{RL} equals the systematic record length if such a complete record is all that is available for a site.

The first step is to run EMA with all available information, including historic information and censored peaks (denoted EMA_C, for EMA complete). From the EMA run, the at-site skewness without regional information $\hat{G}_C$ and the MSE of that skewness estimator $MSE(\hat{G}_C)$ are extracted, as well as the year the historical period begins, YB_C, the year the historical period ends YE_C and the length of the historical period H_C. (YB_C, YE_C, and H_C are used in Equation 9.)

The second step is to run EMA with only the systematic peaks (denoted EMA_S, for EMA systematic). From the EMA_S analysis, the at-site skewness without regional information $\hat{G}_S$ and the MSE of that skewness estimator, $MSE(\hat{G}_S)$ are extracted, as well as the number of peaks P_S. (P_S is used in Equation 4.)

The third step is to represent, from both EMA_C and EMA_S, the precision of the skewness estimators as two record lengths, RL_C and RL_S, based upon the estimated skew and MSE. The corresponding record lengths are calculated using equation (2) below from Griffis and others [2004] and Griffis and Stedinger [2009]: where RL_C uses $\hat{G}_C$ and $MSE(\hat{G}_C)$, and RL_S uses $\hat{G}_S$ and $MSE(\hat{G}_S)$.

$$MSE(\hat{G}) = \left[\frac{6}{\mathrm{RL}} + a(\mathrm{RL})\right] + \left[1 + \left(\frac{9}{6} + b(\mathrm{RL})\right)\hat{G}^2 + \left(\frac{15}{48} + c(\mathrm{RL})\right)\hat{G}^4\right] \quad (2)$$

$$a(\mathrm{RL}) = -\frac{17.75}{RL^2} + \frac{50.06}{RL^3}$$

$$b(\mathrm{RL}) = \frac{3.93}{RL^{0.3}} - \frac{30.97}{RL^{0.6}} + \frac{37.1}{RL^{0.9}}$$

$$c(\mathrm{RL}) = -\frac{6.16}{RL^{0.56}} + \frac{36.83}{RL^{1.12}} - \frac{66.9}{RL^{1.68}}$$

Next, the difference between RL_C and RL_S is employed as a measure of the extra information provided by the historic and/or censored information that was included in the EMA_c analysis, but not in the EMA_s analysis.

$$RL_{diff} = RL_C - RL_S \quad (3)$$

The pseudo record length for the entire record at the gage site, P_{RL}, is calculated using RL_{diff} from equation (3) and the number of systematic peaks P_S,

$$P_{RL} = RL_{diff} + P_S \quad (4)$$

P_{RL} must be non-negative. If P_{RL} is greater than H_C, then P_{RL} should be set to equal H_C. Also if P_{RL} is less than P_S, then P_{RL} is set to P_S. This ensures that the pseudo record length will not be larger than the complete historical period or less than the number of systematic peaks.

Unbiasing the At-Site Estimators

The at-site skewness estimates are unbiased by using the correction factor developed by Tasker and Stedinger [1986] and employed in Reis and others [2005]. The unbiased at-site skewness estimator using the pseudo record length is

$$\hat{\gamma}_i = \left[1 + \frac{6}{P_{RL,i}}\right] G_i \quad (5)$$

Here $\hat{\gamma}_i$ is the unbiased at-site sample skewness estimate for site i, $P_{RL,i}$ is the pseudo record length for site i as calculated in Equation 4, and G_i is the traditional biased at-site skewness estimator for site i from EMA.

The variance of the unbiased at-site skewness includes the correction factor developed by Tasker and Stedinger [1986]:

$$Var[\hat{\gamma}_i] = \left[1 + \frac{6}{P_{RL,i}}\right]^2 Var[G_i] \quad (6)$$

where $Var[G_i]$ is calculated using [Griffis and Stedinger, 2009]

$$Var(\hat{G}) = \left[\frac{6}{P_{RL}} + a(P_{RL})\right] + \left[1 + \left(\frac{9}{6} + b(P_{RL})\right)\hat{G}^2 + \left(\frac{15}{48} + c(P_{RL})\right)\hat{G}^4\right]$$

$$a(P_{RL}) = -\frac{17.75}{{P_{RL}}^2} + \frac{50.06}{{P_{RL}}^3}$$

$$b(P_{RL}) = \frac{3.92}{{P_{RL}}^{0.3}} - \frac{31.10}{{P_{RL}}^{0.6}} + \frac{34.86}{{P_{RL}}^{0.9}}$$

$$c(P_{RL}) = -\frac{7.31}{{P_{RL}}^{0.59}} + \frac{45.90}{{P_{RL}}^{1.18}} - \frac{86.50}{{P_{RL}}^{1.77}}$$

Estimating the Mean Square Error of the Skewness Estimator

There are several possible ways to estimate MSE_G. The approach used by EMA (taken from Cohn and others [2001, eqn 55]) generates a first order estimate of the MSE_G, which should perform well when interval data are present. Another option is to use the Griffis and Stedinger [2009] formula in Equation 6 (the variance is equated to the MSE), employing either the systematic record length or the length of the whole historical period. However, this method does not account for censored data, and thus can lead to inaccurate and underestimated MSE_G. This issue has been addressed by using the pseudo record length instead of the length of the historical period; the pseudo record length reflects the impact of the censored data and the number of recorded systematic peaks. Figure 1 compares the unbiased MSE_G estimates from the Griffis and Stedinger [2009] approach based upon pseudo record lengths and regional skewness estimates, and the unbiased EMA_C MSE_G estimates based on the estimated at-site skewness.

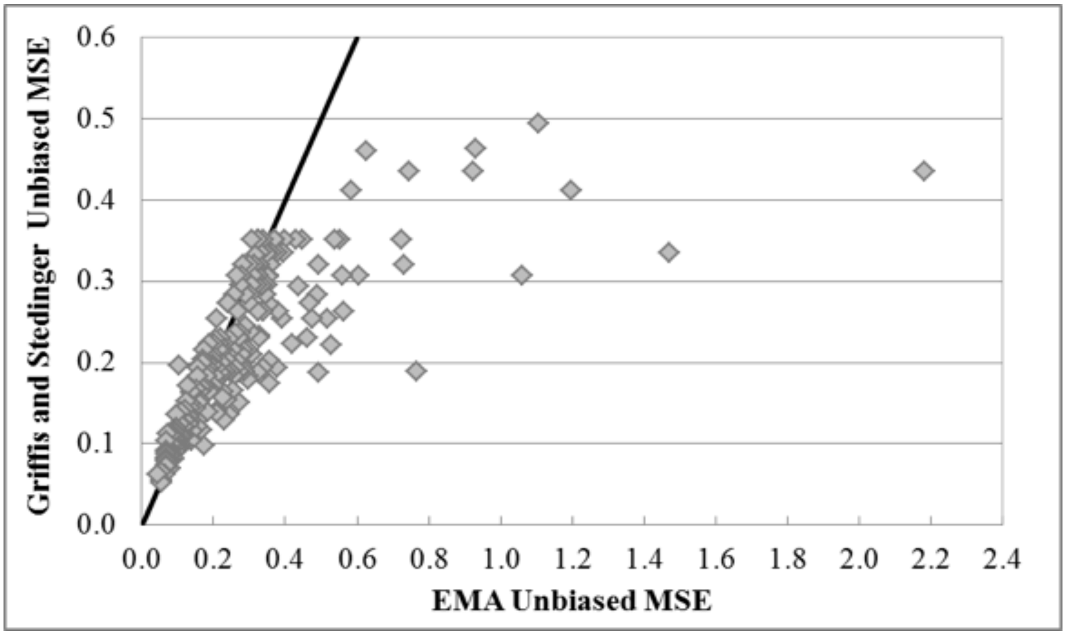

Figure 1: Comparison of EMA and Griffis and Stedinger [2009] MSE_G estimates of at-site skewness estimators for each of the 273 gage sites in the State of Iowa regional skewness study.

As shown in Figure 1, for those gage sites with MSE_G less than about 0.4 the two methods generate similar MSE_G. However, for 33 gage sites, EMA generates unreasonably large MSE_G with values greater than about 0.4. For these sites, the Griffis and Stedinger [2009] formula does not generate a MSE_G greater than 0.5. It appears that for these 33 gage sites EMA is having trouble estimating the parameters due at least in part to the number of censored observations. Of these 33 sites with EMA unbiased $MSE_G > 0.4$, 45% of the sites had 50% or more of their record comprised of censored observations, whereas 81% of the sites had 20% or more of their record comprised of censored observations. Also the average P_{RL} for all 273 sites in the Iowa study is 49 years. However, the longest record of the 33 sites with EMA unbiased $MSE_G > 0.4$ is 43 years, with 85% of the 33 sites having $P_{RL} \leq 35$ years and 42% of the 33 sites have $P_{RL} \leq 25$years. Thus, it appears that for those sites with shorter record lengths and a large percentage of their record comprised of censored observations, EMA has trouble estimating the moments. For this reason, these 33 sites with EMA unbiased $MSE_G > 0.4$ were removed from the analysis.

Thus, there are 240 gage sites remaining from which to build a regional skewness model for the State of Iowa. The unbiased Griffis and Stedinger [2009] MSE_G is used in the regional skewness model because it is more stable and relatively independent of the at-site skewness estimator.

Cross-Correlation Models

A critical step for a GLS analysis is estimation of the cross-correlation of the skewness coefficient estimators. Martins and Stedinger [2002] used Monte Carlo experiments to derive a relation between the cross-correlation of the skewness estimators at two stations *i* and *j* as a function of the cross-correlation of concurrent annual maximum flows, ρ_{ij}:

$$\hat{\rho}\left(\hat{\gamma}_i, \hat{\gamma}_j\right) = Sign\left(\hat{\rho}_{ij}\right) cf_{ij} \left|\hat{\rho}_{ij}\right|^k \tag{7}$$

where $\hat{\rho}_{ij}$ is the cross-correlation of concurrent annual peak discharge for two gaged stations, κ is a constant between 2.8 and 3.3, and cf_{ij}, a factor that accounts for the sample size difference between stations and their concurrent record length, is defined as follows:

$$cf_{ij} = CY_{ij} \Big/ \sqrt{\left(P_{RL,i}\right)\left(P_{RL,j}\right)} \tag{8}$$

CY_{ij}= pseudo record length of the period of concurrent record, and
$P_{RL,i}$, $P_{RL,j}$= the pseudo record length corresponding to sites *i* and *j*, respectively (see equation 4)

Pseudo Concurrent Record Length

After calculating the P_{RL} for each gage site in the study, the pseudo concurrent record length between pairs of sites can be calculated. Due to the use of censored data and historic data, the effective concurrent record length calculation is more complex than determining in which years the two gage sites both have recorded systematic peaks.

The years of historical record in common between the two gage sites is first determined. For the years in common, with beginning year YB_{ij} and ending year YE_{ij}, the following equation is used to calculate the concurrent years of record between site *i* and site *j*.

$$CY_{ij} = \left(YE_{ij} - YB_{ij} + 1\right)\left(\frac{P_{RL,i}}{H_{C,i}}\right)\left(\frac{P_{RL,j}}{H_{C,j}}\right) \tag{9}$$

The computed pseudo concurrent record length depends upon the years of historical record in common between the two gage sites, as well as the ratios of the pseudo record length to the historical record length for each of the two gage sites.

IOWA REGIONAL SKEWNESS RESULTS

This section describes the Iowa regional skewness regression analysis using the B-WLS/B-GLS regression methodology [see Veilleux, 2011; Veilleux and others, 2011] described above. All of the available basin characteristics were considered as explanatory variables in the regional skewness analysis. Available basin characteristics include: precipitation (mean annual, mean monthly, maximum 24 hours over a number of years), soil (hydrologic soil types, percent clay and sand, soil permeability), stream characteristics (main channel slope, stream density, ruggedness, number of first order streams, total stream length), basin measures (drainage area, slope, relief, length, perimeter, shape factor), hydrologic parameters (streamflow variability index, base flow index, base flow recession), and hydrologic regions. A few basin characteristics were statistically significant in explaining the site-to-site variability in skewness, including slope, drainage area, basin length, and the total length of mapped streams in the basin. The best model, as classified by having the smallest model error variance, σ_δ^2 , and largest pseudo R_δ^2, which included a constant and a parameter (or combination of parameters) was the model which included drainage area. Table 1 provides the final results for the constant skewness model denoted "Constant," and the model that uses a linear relation between skewness and log10[Drainage Area], denoted "DA."

Table 1: Regional skewness models for Iowa. [Standard deviations are in parentheses. σ_δ^2 is the model error variance. ASEV is the average sampling error variance. AVP_{new} is the average variance of prediction for a new site. Pseudo R_δ^2 (%) describes the fraction of the variability in the true skewness explained by each model (Gruber and others, 2007)]

Model	egression Param		σ_δ^2	ASEV	AVP_{new}	R_δ^2
	b_1	b_2				
Constant: $\hat{y} = b_1$	-0.40 (0.09) 0%	-	0.15 (0.03)	0.01	0.16	0%
DA: $\hat{y} = b_1 + b_2[log_{10}(DA)]$	-0.78 (0.16) 0%	0.20 (0.05) 0%	0.12 (0.02)	0.01	0.13	19%

Table 1 includes the pseudo R_δ^2 value for both models; pseudo R_δ^2 describes the estimated fraction of the variability in the true skewness from site-to-site explained by each model [Gruber and others, 2007; Parrett and others, 2011]. A constant model does not explain any variability, so the pseudo R_δ^2 equals 0. The "DA" model has a pseudo R_δ^2 of 19 %. The posterior mean of the model error variance, σ_δ^2, for the DA model is 0.12, which is smaller than that for the Constant model for which $\sigma_\delta^2 = 0.15$. This indicates that the inclusion of drainage area as an explanatory variable in the regression helps explain some of the variability in the true skewness. However, this small gain in precision does not warrant the increased model complexity. Thus, the Constant model is chosen as the best regional model for Iowa skewness. The average sampling error variance (ASEV) in Table 1 is the average error in the regional skewness estimator at the sites in the data set. The average variance of prediction at a new site (AVP_{new}) corresponds to the mean square error (MSE) used in Bulletin 17B to describe the precision of the generalized skewness. The Constant model has an AVP_{new}, equal to 0.16, which corresponds to an effective record length of 50 years. An AVP_{new} of 0.16 is a marked improvement over the Bulletin 17B skewness map, whose reported MSE is 0.302 [Interagency Advisory Committee on Water Data, 1982] for a corresponding effective record length of only 17 years. Thus the new regional model has three times the information content (as measured by effective record length) of that claimed by the Bulletin 17B map.

Figure 2 shows the relation between the unbiased at-site skewness and drainage area; the marker selected for each gage site represents the at-site pseudo record length. The sites with the largest drainage area generally have the longest pseudo record lengths. It is not apparent from the data that the upward trend, suggested by the DA model, exists between the unbiased at-site skewness and drainage area. Thus, for this study, the simpler model is chosen, *i.e.* the Constant model.

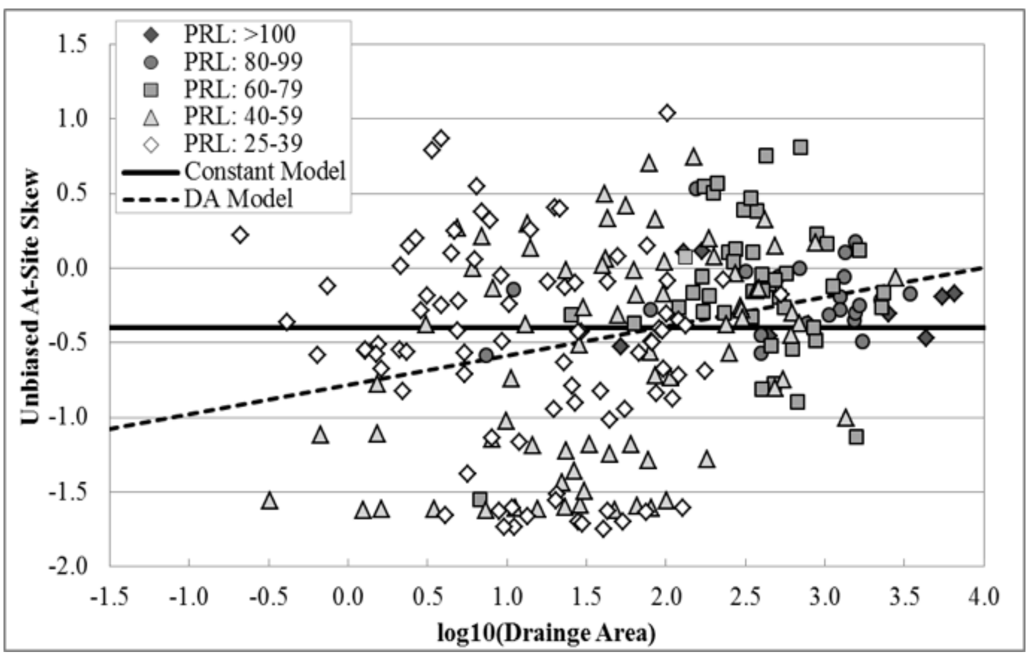

Figure 2: Relation between the unbiased at-site skewness and drainage area for the 240 sites in the State of Iowa regional skewness study. All 240 sites are represented with shapes signifiying five different groupings of at-site pseudo record length (P_{RL}): solid diamonds, record length greater than 100 years; circles, record length between 80-99 years; squares, record length between 60-79 years; triangles, record length between 40-59 years; open diamonds, record length between 25-39 years. The solid black line represents the Constant Model from Table 1, while the dashed black line represents the DA model from Table 1.

CONCLUSIONS

This paper continues efforts to develop a regional statistical methodology for the estimation of skewness parameters. Regional log-space skewness studies to support frequency analysis with the LP3 distribution have extended the Bayesian-Generalized Least Squares methodology presented by Reis and others [2005] [Parrett and others, 2011; Veilleux and others, 2011]. The inclusion of censored data from crest stage gages and historic information in Iowa required significant adaptations of the B-WLS/B-GLS regression procedures. This paper describes those extensions of the B-WLS/GLS algorithm to account for the censored data in record length and concurrent record length calculations required for the GLS covariance matrix. The Bayesian WLS/GLS methodology is used successfully to develop regional skewness models for the log-skewness of Iowa peak flows. The nominal effective record length (ERL) of the regional skewness estimators is 50 years (MSE = 0.16). This ERL is dramatically better than the ERL of 17 years (MSE = 0.302) reported for Plate 1 in Bulletin 17B, the current flood frequency guidelines used by Federal agencies in the United States.

REFERENCES:

Cohn, T.A., (2011), PeakfqSA Version 0.960 (software), Tim Cohn, USGS, http://www.timcohn.com/TAC_Software/PeakfqSA/

Cohn, T.A., W.L. Lane, and W.G. Baier (1997), An algorithm, for computing moments based flood quantile estimates when historical flood information is available, Water Resour. Res., 33(9), 2089-2096.

Cohn, T.A., W.L. Lane, and J.R. Stedinger (2001), Confidence intervals for Expected Moments Algorithm flood quantile estimates, Water Resour. Res., 37(6), 1695-1706.

Gotvald, A.J., Feaster, T.D., and Weaver, J.C., (2009), Magnitude and frequency of rural floods in the southeastern United States, 2006—Volume 1, Georgia: U.S. Geological Survey Scientific Investigations Report 2009–5043, 120 p.

Griffis, V.W., and J. R. Stedinger, (2009), The Log-Pearson Type 3 Distribution and its Application in Flood Frequency Analysis, 3. Sample Skew and Weighted Skew Estimators, J. of Hydrol. Engineering 14(2), 121-130.

Griffis, V. W. , J.R. Stedinger, and T. A. Cohn , (2004), LP3 Quantile Estimators with Regional Skew Information and Low Outlier Adjustments, Water Resources Research, 40, W07503, doi:1029/2003WR002697.

Gruber, A.M., Dirceu S. R. Jr., and J.R. Stedinger, (2007), Models of Regional Skew Based on Bayesian GLS Regression, Paper 40927-3285, World Environmental & Water Resources Conference - Restoring our Natural Habitat, K.C. Kabbes editor, Tampa, Florida, May 15-18.

Gruber, A.M. and J.R. Stedinger, (2008), Models of LP3 Regional Skew, Data Selection and Bayesian GLS Regression, Paper 596, World Environmental and Water Resources Congress – Ahupua'a, Babcock, R.W. and R. Watson editors, Honolulu, Hawai'I, May 12-16.

Interagency Committee on Water Data (IACWD). (1982). *Guidelines for determining flood flow frequency: Bulletin 17B (revised and corrected)*, Hydrol. Subcomm., Washington, D.C., 28.

Martins, E. S., and J.R. Stedinger [2002], Cross correlations among estimators of shape, Water Resour. Res., 38(11), 1252, doi:10.1029/2002WR001589.

Parrett, C., Veilleux, A.G., Stedinger, J.R., Barth, N.A., Knifong, D.L., and Ferris, J.C., (2011), Regional skew for California, and flood frequency for selected sites in the Sacramento–San Joaquin River Basin, based on data through water year 2006: U.S. Geological Survey Scientific Investigations Report 2010–5260, 94 p.

Reis, D. S., Jr., J. R. Stedinger, and E. S. Martins, (2005), Bayesian generalized least squares regression with application to log Pearson type 3 regional skew estimation, Water Resour. *Res.*, *41*, W10419, doi:10.1029/2004WR003445.

Stedinger, J.R., and T.A. Cohn, (1986), Flood Frequency Analysis With Historical and Paleoflood Information, Water Resources Research, 22(5), 785-793.

Stedinger, J.R. and V.W. Griffis, (2008), Flood Frequency Analysis in the United States: Time to Update. (editorial) *J. of Hydrol. Engineering*, April, pp. 199-204.

Tasker. G.D. and J.R. Stedinger, (1986), Regional Skew with Weighted LS Regression, *Journal of Water Resources Planning and Management*, 112(2), 225-237.

Veilleux, A.G., (2011), Bayesian GLS Regression, Leverage and Influence for Regionalization of Hydrologic Statistics, Ph.D. Dissertation, Cornell University, August.

Veilleux, A.G., J.R. Stedinger and J.R. Lamontagne,(2011), Bayesian WLS/GLS Regression for Regional Skewness Analysis for Regions with Large Cross-Correlations among Flood Flows, paper 3103, World Environmental and Water Resources Congress 2011: Bearing Knowledge for Sustainability, ASCE, Palm Springs California, May 22-26.

Veilleux, A.G., J. R. Stedinger, and D. A. Eash, Bayesian WLS/GLS Regression for Regional Skewness Analysis for Crest Stage Gage Networks, Paper 227, Crossing Boundaries, Proceedings World Environmental and Water Resources Congress, ed. Eric D. Loucks, Amer. Society of Civil Engineering, Albuquerque, New Mexico, pp. 2253-63, May 20-24, 2012.