Bayesian generalized least squares regression with application to log Pearson type 3 regional skew estimation

D. S. Reis Jr.
Water Resources Department, Fundação Cearense de Meteorologia e Recursos Hídricos, Fortaleza, Brazil

J. R. Stedinger
School of Civil and Environmental Engineering, Cornell University, Ithaca, New York, USA

E. S. Martins
Water Resources Department, Fundação Cearense de Meteorologia e Recursos Hídricos, Fortaleza, Brazil

Received 25 June 2004; revised 20 May 2005; accepted 22 June 2005; published 26 October 2005.

This paper develops a Bayesian approach to analysis of a generalized least squares (GLS) regression model for regional analyses of hydrologic data. The new approach allows computation of the posterior distributions of the parameters and the model error variance using a quasi-analytic approach. Two regional skew estimation studies illustrate the value of the Bayesian GLS approach for regional statistical analysis of a shape parameter and demonstrate that regional skew models can be relatively precise with effective record lengths in excess of 60 years. With Bayesian GLS the marginal posterior distribution of the model error variance and the corresponding mean and variance of the parameters can be computed directly, thereby providing a simple but important extension of the regional GLS regression procedures popularized by Tasker and Stedinger (1989), which is sensitive to the likely values of the model error variance when it is small relative to the sampling error in the at-site estimator.


1. Introduction

The development of a rigorous and efficient statistical procedure for the estimation of a regional skew coefficient model and its precision for the log Pearson type 3 (LP3) distribution is of great importance in the United States because U.S. federal guidelines for flood frequency analysis (FFA) in Bulletin 17B [Interagency Advisory Committee on Water Data (IACWD), 1982; McCuen, 1979] require the use of a regional skew estimator and its mean square error, and the use of an accurate regional skew with a small model error variance is known to be of a significant value [Griffies et al., 2004]. Traditional regional skew map procedures are entirely lacking in statistical rigor [Tasker and Stedinger, 1986]. The Bayesian generalized least squares (GLS) regional regression model developed here provides the needed statistical tool for FFA in the United States and other countries that employ regional shape parameter models. The corresponding extension of the traditional regional GLS regression procedures popularized by Tasker and Stedinger [1989] is also of value for regionalization problems in hydrology including flood quantiles, rainfall statistics, nutrient loads, and low-flow statistics.

1.1. Background

[Cunnane [1988] and Groupe de recherche en hydrologie statistique (GREHYS) [1996a, 1996b] review methods that use regional information in the estimation of hydrologic statistics. One versatile approach employs regional information to derive a relationship between streamflow statistics and physiographic characteristics using regional regression analysis. Such regional regression methods have been widely used to estimate hydrologic statistics at ungauged sites [Benson and Matalas, 1967; Matalas and Gilroy, 1968; Thomas and Benson, 1970; Moss and Kerlinger, 1974; Jennings et al., 1994] and to increase the precision of the statistic of interest at sites with short record lengths by adding regional information [Shane and Gaver, 1970; Vicens et al., 1975; IACWD, 1982; Kuczera, 1982; Stedinger, 1983; Madsen and Rosbjerg, 1997; Fill and Stedinger, 1998; Martins and Stedinger, 2000; Walker and Krug, 2003; Shu and Burn, 2004; Reis and Stedinger, 2005; Reis, 2005].

Regional regression models aim to explain spatial variability of the hydrologic statistic by relating it to physiographic variables, such as drainage area, slope of the main channel, and percentage of forest cover. For many years, the parameters of regional model were estimated by the ordinary least squares (OLS) procedure that considers the residual errors of the model to be homoscedastic and independently distributed [Riggs, 1973]. These assumptions are often violated in hydrologic problems because estima-
tors of the statistic of interest at different sites have different precision due to variations in record length, and are usually not independent because streamflow records are often spatially correlated.

[5] Tasker [1980] suggested the use of a weighted least squares (WLS) procedure, which employs weights for each site to account for differences in record length used in the computation of flood quantiles. Tasker concluded that WLS provides a smaller mean square error for the parameter estimates than OLS when the model error is small, and performs as well as OLS when the model error is large.

[6] Kuczera [1983] considers how short record lengths and spatial correlation affect parameter estimation when regional and at-site information are combined through an empirical Bayesian procedure. He introduced a generalized regional regression model that considered both time sampling errors and the spatial correlation among the estimators. Stedinger and Tasker [1985, 1986] developed a practical generalized least squares (GLS) regression estimator for flood quantiles that accounts for differences in record lengths and also for correlation among estimators of the statistic of interest at different sites. They resolved the difficult issue raised by Kuczera [1983] and Tasker [1980] of how to estimate the covariance matrix of the residuals errors.

[7] The estimated covariance matrix should be independent of the estimates of the flood quantiles in order to obtain an unbiased estimator of the parameters of the model. This poses a problem because the variance of the residuals is a function of the at-site variance. Thus Stedinger and Tasker [1985] used a smoothed sampling covariance matrix that employs a regional estimate of the at-site variance and a smoothed correlation function. Their Monte Carlo experiments compare the statistical performance of OLS, WLS, and GLS with the smoothed sampling covariance matrix in the estimation of regional flood quantiles, and maximum likelihood (ML) and generalized mean square error model error variance estimators. WLS and GLS parameter estimators were more efficient than OLS, and GLS performed better than WLS in situations where the cross correlation of the flows is large and the model error is relatively modest or small. Kroll and Stedinger [1998] showed that such smoothing resulted in little loss of efficiency.

1.2. Previous Applications

[8] The GLS procedure introduced by Stedinger and Tasker [1985] and Tasker and Stedinger [1989] has been extensively used nationally and internationally [World Meteorological Organization, 1994; Robson and Reed, 1999] to estimate the parameters of regional regression models of flood quantiles [Tasker et al., 1986, 1996; Kjeldsen and Rosbjerg, 2002; Feaster and Tasker, 2002; Law and Tasker, 2003], low-flow statistics [Tasker, 1989; Vogel and Kroll, 1990; Ludwing and Tasker, 1993; Kroll and Stedinger, 1999], extreme rainfall [Madsen et al., 1995, 2002], loads of chemical constituents [Tasker and Driver, 1988], and parameters of probability distributions [Madsen and Rosbjerg, 1997]. Several of the reports provide comparisons showing that GLS does yield smaller model error and prediction error estimates [e.g., Tasker and Driver, 1988; Vogel and Kroll, 1990; Walker and Krug, 2003; Kothun, 2003]. GLS has also been used as the basis of hydrologic network design [Tasker, 1986; Tasker and Stedinger, 1989; Moss and Tasker, 1991], including in the work by Markus et al. [2000], who contrast the use of the Tasker-Stedinger GLS procedure with a maximum entropy analysis.

[9] Tasker and Stedinger [1986] applied WLS regression to derive a regional skewness estimator for the Illinois River basin obtaining an average variance of prediction equal to 0.10. Rasmussen and Perry [2000] employed WLS with stations in Kansas and obtained a MSE of 0.036, and Pope et al. [2001] considered two statistically significant regions in North Carolina and, with WLS, obtained model error variances of 0.038 and 0.062, whereas for Texas, Judd et al. [1996] combined approximate WLS weights with kriging and reported an average error variance of 0.12. Those studies were unable to use GLS because they did not know how to describe the correlations among skewness estimators. Martins and Stedinger [2002a] have recently developed simple equations for the cross correlation among skewness (and shape parameter \( \kappa \) of GEV and GP distributions) estimators as a function of the cross correlation of the flood flows themselves. Martins and Stedinger [2002b] employed those equations to develop a GLS model for regional skew estimation.

1.3. Model Error Estimation

[10] While Stedinger and Tasker [1986] studied the precision of the generalized method of moment and ML estimators of the model error variance, no general procedure has been developed to describe the precision of those model error estimators. Such a measure of precision is needed, particularly for regional skew analyses wherein the value of the model error variance is of significant concern. Moreover, the Stedinger-Tasker generalized method of moments and maximum likelihood estimators of the model error variance may be zero in situations where the model error variance is small compared to the sampling error. This is often the case for regionalization of shape parameters and this can distort the uncertainty attributed to the regional regression estimator.

[11] This paper introduces a Bayesian approach to the analysis of a GLS model for hydrologic statistics. The new approach allows computation of the posterior distributions of the parameters and the model error variance using a quasi-analytic approach. The Bayesian approach provides both a measure of the precision of the model error variance that the traditional GLS lacks, and a more reasonable description of the possible values of the model error variance in cases where the model error variance is small compared to the sampling errors.

2. Regression Problem

[12] This section develops the basic GLS model and discusses classical analysis procedures. Subsequent sections recast the analysis of the GLS model in a Bayesian framework.

[13] The GLS model assumes that the quantity of interest \( y_i \) at a given site \( i \) can be described by a linear function of physiographic characteristics (or a transformation there of) with an additive error. In matrix notation, the model is

\[
y = X\beta + \epsilon
\]
columns, the vector $\beta$ has the $(k + 1)$ parameters of the model that must be estimated, and $\epsilon$ is the vector containing the errors for each of the $n$ sites used in the regression.

[14] In the statistical literature, the errors $\epsilon$ are generally assumed to have zero mean and covariance matrix defined by

$$E(\epsilon\epsilon^T) = \delta^2 \Omega$$  \hspace{1cm} (2)

wherein $\delta^2$ is the model error variance and $\Omega$ is a positive definite symmetric matrix. Different choices of the matrix $\Omega$ allow one to make different assumptions regarding the nature of the model errors. If $\Omega$ is equal to the identity matrix $I$, the errors are homoscedastic, and the GLS model reduces to OLS. Uncorrelated errors with different variances at different sites can be described using a heteroscedasticity and correlation among residuals. Definite symmetric matrix $\Omega$ were known, the minimum variance unbiased estimator of $\beta$ does not depend on $\delta^2$ and is given by [Aitken, 1935; Rao and Toutenburg, 1999]

$$\hat{\beta} = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} y$$  \hspace{1cm} (3)

with sampling covariance matrix

$$\Sigma[\hat{\beta}] = \delta^2 (X^T \Omega^{-1} X)^{-1}$$  \hspace{1cm} (4)

An unbiased model error variance estimator is

$$\hat{\delta}^2 = \frac{(y - X\hat{\beta})^T \Omega^{-1} (y - X\hat{\beta})}{n - k - 1}$$  \hspace{1cm} (5)

In practice the matrix $\Omega$ is rarely known and must be estimated, which usually causes a loss of efficiency in the estimators of $\beta$ and $\delta^2$ [Rao and Toutenburg, 1999].

2.1. Stedinger-Tasker Model

[16] Stedinger and Tasker [1985, 1986] developed a slightly different GLS model for regional hydrologic analysis. The key difference is the partition of the covariance matrix of the errors. Their model assumes that total error results from two sources: model errors $\epsilon_i$ that are assumed to be independently distributed with mean zero and common variance $\delta^2$, and sampling errors that arise due to the fact the actual values of the $y_i$ are unknown and only estimates of the quantities of interest are available. Therefore equation (1) becomes

$$\hat{y} = X\beta + \omega + \epsilon = X\beta + \eta$$  \hspace{1cm} (6)

where $\omega$ is the sampling error in the sample estimators. Thus the regression model errors $\eta_i$ are a combination of (1) time-sampling error in sample estimators $\hat{y}_i$ of $y_i$, and (2) underlying model error $\epsilon_i$. The total error $\eta$ has mean zero and covariance matrix

$$E[\eta\eta^T] = \Lambda(\delta^2) = \delta^2 I + \Sigma$$

where $\Sigma$ is the covariance matrix of the sampling errors in the sample estimators. Time-sampling errors in estimators of the $y_i$ values are usually correlated among sites because flows at nearby sites experience similar meteorology. Estimation of the sampling covariance matrix in the GLS regression is of great concern and will be discussed later.

[17] Stedinger and Tasker employ the GLS estimator of $\beta$ given by

$$\beta = [X^T \Lambda(\delta^2)^{-1} X]^{-1} X^T \Lambda(\delta^2)^{-1} y$$

whosing sampling covariance matrix for given $\delta^2$ equals

$$\Sigma[\hat{\beta}] = [X^T \Lambda(\delta^2)^{-1} X]^{-1}$$

The model error variance $\delta^2$ can be estimated by either the generalized method of moments or the maximum likelihood estimators considered by Stedinger and Tasker [1986].

[18] The method of moments (MM) GLS estimator is obtained by iteratively solving equation (8) along with the generalized residual mean square error equation

$$\frac{1}{2} \ln[\det(\delta^2 I + \Sigma)] + \frac{1}{2} (y - X\beta)^T [\delta^2 I + \Sigma]^{-1} (y - X\beta)$$

for $n$ sites and $k + 1$ parameters, which is a direct generalization of equation (5). In some situations, the sampling covariance matrix explains all the variability observed in the data, which means the left-hand side of equation (10) will be less than $n - (k + 1)$ even if $\delta^2$ is zero. In these circumstances, the $\hat{\delta}^2$ is generally set to zero [Stedinger and Tasker, 1985].

[19] The maximum likelihood estimator is derived by assuming the residuals are normally distributed with mean zero and covariance matrix described in equation (7). Then, both $\beta$ and $\delta^2$ can be estimated jointly, subject to $\delta^2 \geq 0$, by minimizing

$$\frac{1}{2} \ln[\det(\delta^2 I + \Sigma)] + \frac{1}{2} (y - X\beta)^T [\delta^2 I + \Sigma]^{-1} (y - X\beta)$$

which is the same as equation (9).

2.2. Error Variance Estimators

[20] Clearly, the maximum likelihood estimate of $\beta$ is the same as the one given by (8), except that the value of $\delta^2$ would be different. Because $\beta$ and $\delta^2$ are asymptotically independent [Rencher, 2000], the variance of $\beta$ is described by the inverse of the Fisher information matrix [Bickel and Doksum, 1977].

$$\Sigma[\beta] = \left[ \frac{\partial^2 \ln f(y|\beta)}{\partial \beta \partial \beta^T} \right]^{-1}$$

where $\Sigma[\beta]$ is the sampling covariance matrix of the residuals.
Tasker, 1991; Kroll and Stedinger, 1998] show that the generalized least squares (GLS) procedure with the estimated sampling covariance matrix generally is more efficient than WLS and OLS for hydrologic regression.

[22] On the basis of Monte Carlo simulations, Stedinger and Tasker [1986] concluded that the method of moments model error variance procedure is faster, more robust because no assumptions about the distribution of the residuals are made, and less biased when the true model error variance is moderate to large. The Stedinger and Tasker [1986] data from their experiment 1 are displayed in Figures 1 and 2. The cross correlation among concurrent annual maximum flows are 0, 0.3, 0.6. For small $d^2$, MLEs were much more accurate, but this was not thought to be important because the focus at that time was the estimation of quantiles which generally have larger errors. Actually, the ML estimator for $d^2$ always had a smaller MSE than the method of moments estimator.

[23] These results suggest that if the regional regression analysis exhibits a small model error variance, as one sees

**Figure 1.** Bias in the estimation of the model error variance with cross correlation among concurrent annual maximum flows equal to 0, 0.3, and 0.6 using generalized method of moments (MM) and MLE. Results are from Stedinger and Tasker [1986].

**Figure 2.** Root mean square error of the estimated model error variance with cross correlation among concurrent annual maximum flows equal to 0, 0.3, and 0.6 using generalized MM and MLE. Results are from Stedinger and Tasker [1986].

4 of 14
with the regionalization of skewness estimators, the maximum likelihood procedure should be preferred to the generalized method of moment estimator. Bayesian analysis, which is based on the likelihood function, is also a good candidate for these situations, and would address the bias concern because on average over the prior, Bayesian estimators are unbiased [Stedinger, 1983].

2.3. Predictive Variance

The variance of prediction for a statistic of interest \( Y \) around its estimated value \( \hat{x}_o \beta \) at a new site not used to derive the model with physiographic characteristics \( x_o \) is given by

\[
E \left[ (Y - \hat{x}_o \beta)^2 \right] = \delta^2 + x_o (X^T A^{-1}X)^{-1} x_o^T
\]

(13)

where the two terms in the right-hand side are the regional model error variance and the sampling variance associated with \( x_o \beta \). The variance of prediction as defined in equation (13) could be used as a criterion for model selection if one is interested in prediction at a specific site. However, regression equations are often developed to make predictions over an entire region and not only for a specific site [e.g., Feaster and Tasker, 2002]. In order to evaluate the precision of a regression model for a region, Stedinger and Tasker [1986] used the “average variance of prediction” for new sites (AVP\(_{\text{new}}\)), which reflects how well the model would perform on average at sites like those at which one has data

\[
\text{AVP}_{\text{new}} = \delta^2 + \frac{1}{n} \sum_{i=1}^{n} x_i (X^T A^{-1}X)^{-1} x_i^T = \delta^2 + \text{ASV}
\]

(14)

where ASV is the average sampling variance.

Sometimes, regression models are derived to provide a regional value of the statistic of interest that will be used with at-site information to improve the precision of an estimator [e.g., Kuczera, 1983]. In these cases, it is common to include the data for the site in question in the regression study. In these instances, the formula for the AVP that accounts for the correlation between \( x_o \beta \) and the model error is

\[
\text{AVP}_{\text{old}} = \delta^2 + \frac{1}{n} \sum_{i=1}^{n} x_i (X^T A^{-1}X)^{-1} x_i^T - 2\delta^2 x_i (X^T A^{-1}X)^{-1} X^T A^{-1} e_i
\]

(15)

where \( e_i \) is a column vector with one at the \( i \)th row and zero elsewhere [Reis, 2005].

3. Bayesian Approach

3.1. Background

The current GLS model analysis methodology based on the work by Tasker and Stedinger [1989] does not provide an estimate of the uncertainty in the estimated model error variance, nor does it include that uncertainty in the estimator of the variance of \( \beta \). Moreover, in situations where the model error variance is small, such as in the regionalization of the skewness estimator, both MM and MLE-GLS may yield a regional model error variance estimator equal to zero, which implies the model is perfect. This result causes an overestimation of the precision of the regional estimate leading eventually to an underestimation of the uncertainty in the flood quantile and too much relevance on the regional model. Bayesian analysis is a natural methodology to solve these problems because it can provide the full posterior distribution of both the parameters and the error variance of the regional model, and describes in a more reasonable fashion the possible values of the model error even when the MLE is zero. Bayesian analyses of spatial models are reported by Kitanidis [1986], Guadard et al. [1999], Best et al. [2000], Schmidt and Gelfand [2003], Wikle and Anderson [2003], Wikle [2003], and Richardson and Best [2003]. For instance, Holland et al. [2000] used Bayesian analysis to estimate regional trends in sulfur dioxide over the eastern United States. Like our GLS model, they also had a covariance matrix with two additive terms, one that accounts for the actual model error and another for measurement error with a specified covariance matrix. They use a Markov Chain Monte Carlo algorithm to obtain the regional trends and their standard errors, as well as the posterior distribution of the model error variance.

3.2. Bayesian Inference

Bayesian inference is an alternative to the classical statistical point of view. In a Bayesian framework, the knowledge about the parameters of the model is described by a probability distribution. The Bayesian approach combines the available data with prior knowledge about the parameters, which may come from other data sets or a modeler’s experience and intuition, using Bayes theorem [Zellner, 1971]:

\[
p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}
\]

(16)

Here \( p(\theta|y) \) is the posterior distribution of the parameter vector \( \theta \) given the information \( y \) in the available data set, \( p(\theta|y) \) is the likelihood function for the data, and \( p(\theta) \) is the prior distribution of \( \theta \). The denominator is a normalizing constant that ensures that the area under the posterior pdf equals one.

Providing a full posterior distribution of the parameters is an advantage of the Bayesian approach over classical methods, which usually report a point estimate of the parameters and make use of asymptotic normality assumptions to evaluate uncertainties [Bickel and Doksum, 1977]. In the Bayesian framework, one does not have to use such approximations to evaluate the uncertainties because the full posterior distribution of the parameters is available. Moreover, credible intervals on the parameters, or any function of the parameters, are more easily interpreted than the concept of confidence intervals of classical statistics [Congdon, 2001].

3.3. Bayesian GLS Model

3.3.1. Prior Distribution

Here prior information concerning the \( \beta \) parameters is modeled by a multivariate normal distribution with mean vector \( \beta_0 \) and precision matrix \( P \), which is the inverse of the covariance matrix \( \Sigma_\beta \). If no prior information on any of
the β parameters is available, an almost noninformative prior distribution with a large variance can be used (thus P ≈ 0). A multivariate normal prior is

\[ \xi(\beta) = \frac{|P|^{1/2}}{(2\pi)^{n/2}} \exp\left[-0.5(\beta - \beta_0)^T P (\beta - \beta_0)\right] \] (17)

wherein β has dimension k + 1.

[30] When no previous information for the model error variance is available, the prior can be represented by the reciprocal of the variance [Zellner, 1971]. This prior is improper but results in a proper posterior if it is combined with the likelihood function for the traditional regression model [Zellner, 1971, p. 45]. Another way to represent prior information for the error variance is to use a gamma distribution so that

\[ \exp \left( -0.5 (\beta - \beta_0)^T P (\beta - \beta_0) \right) \]

is the joint posterior of the parameters of the model precision [Zellner, 1971; Congdon, 2001]. This is equivalent to modeling the error variance with an inverse gamma distribution. When no previous information is available, one should choose parameters α and β such that the pdf is relatively flat or “gentle” on the region of interest.

### 3.3.2. Likelihood Function

[31] The likelihood function for the data is a multivariate normal distribution so that

\[ L(\hat{y}|\beta, \sigma^2) = (2\pi)^{-n/2} |\Lambda|^{-1/2} \exp\left[-0.5(\hat{y} - \hat{y}|X|^2)^T \Lambda^{-1}(\hat{y} - X\beta)\right] \] (18)

where the covariance matrix is defined in equation (7), n is the number of gauges, \( \hat{y} \) is the vector with the sample values of the hydrologic characteristic of interest, and X is the matrix of covariates.

### 3.4. Quasi-analytic Solution

[32] Computing the normalizing constant in (16) can be a computationally intense task depending on the dimension of the problem. One possible solution is to use a Markov chain Monte Carlo (MCMC) algorithm [Gilks et al., 1996] that generates a correlated sample of the parameters of the model from the joint posterior distribution in (16) without knowing the normalizing constant. This sample would then be used to obtain the posterior distribution of the model error variance and β parameters. While we initially implemented an MCMC analysis, we discovered that this problem could be solved more easily using the quasi-analytical approximation of the marginal posterior of the model error variance presented here.

[33] This section derives a quasi-analytical approximation of the posterior distribution of the model error variance using the same factorization employed by Kitanidis [1986] and Zellner [1971, equation 8.14]. Numerical integration of a one-dimensional integral provides the mean and variance of the regional model error variance. Knowing the posterior distribution of the model error variance allows the computation of the posterior moments of the parameters β by another one-dimensional numerical integration.

[34] Given the likelihood function in (18) and the prior for β in (17), in simple cases, one can integrate the joint posterior distribution of \( \sigma^2 \) and β over the possible values of β to obtain numerically the marginal posterior for \( \sigma^2 \), except for the normalizing constant. Thus:

\[ f(\sigma^2|\hat{y}) = \int f(\beta|\sigma^2, \hat{y}) f(\sigma^2|\beta) d\beta \] (19)

where \( f(\beta|\sigma^2, \hat{y}) \) is the joint posterior of the parameters, \( f(\sigma^2|\beta) \) is the likelihood function, and \( \xi(\beta, \sigma^2) \) is the joint prior for \( \sigma^2 \) and β. Using the prior distribution for β in (17), the joint prior distribution is

\[ \xi(\beta, \sigma^2) \propto \xi(\sigma^2) \exp\left[-0.5 (\beta - \beta_0)^T P (\beta - \beta_0)\right] \] (20)

where \( \xi(\sigma^2) \) is the prior for the model error variance. The integral in (19) yields a marginal posterior distribution for \( \sigma^2 \) given by

\[ f(\sigma^2|\hat{y}) \propto \xi(\sigma^2) \int \exp\left[-0.5 (\hat{y} - X\beta)^T \Lambda^{-1}(\hat{y} - X\beta)\right. \\
- 0.5 (\beta - \beta_0)^T P (\beta - \beta_0)\big] d\beta \] (21)

With the substitution

\[ (\hat{y} - X\beta)^T \Lambda^{-1}(\hat{y} - X\beta) = (\hat{y} - X\beta)^T \Lambda^{-1}(\hat{y} - X\beta) \\
+ (\beta - \beta_0)^T X^T \Lambda^{-1}X(\beta - \beta_0) \] (22)

the integral in equation (21) can be rewritten

\[ \exp\left\{-0.5 \left[ (\hat{y} - X\beta)^T \Lambda^{-1}(\hat{y} - X\beta) + \beta_0^T P \beta_0 + \beta_0^T X^T \Lambda^{-1}X \beta_0 - \beta_0^T C \beta_0 \right] \right\} \\
\cdot \int \exp\left\{-0.5 (\beta - \beta_0)^T C (\beta - \beta_0)\right\} d\beta \] (23)

wherein \( \beta_0 = C^{-1} [P \beta_0 + X^T \Lambda^{-1} X \beta_0] \) and \( C = P + X^T \Lambda^{-1} X \). Thus the marginal posterior distribution of \( \sigma^2 \) is given by

\[ f(\sigma^2|\hat{y}) \propto |A|^{-1/2} \exp\left\{-0.5 (\hat{y} - X\beta)^T \Lambda^{-1}(\hat{y} - X\beta) \\
+ \beta_0^T P \beta_0 + \beta_0^T X^T \Lambda^{-1}X \beta_0 - \beta_0^T C \beta_0 \right\} \xi(\sigma^2) \] (24)

For an almost noninformative prior on the parameters β, so that \( P \rightarrow 0 \), then \( C \rightarrow X^T \Lambda^{-1} X \) and \( \beta_0 \rightarrow \beta \) in equation (8), so the posterior distribution for \( \sigma^2 \) reduces to a simpler expression published earlier by Reis et al. [2003]:

\[ f(\sigma^2|\hat{y}) \propto |A|^{-1/2} \exp\left\{-0.5 (\hat{y} - X\beta)^T \Lambda^{-1}(\hat{y} - X\beta) \right\} \xi(\sigma^2) \] (25)

With either (24) or (25) one can compute numerically the posterior pdf of the model error, and its mean and variance. The pdf of the model error variance can then be used to obtain the posterior distribution of the β parameters by numerically computing

\[ f(\beta|\sigma^2, \hat{y}) = \int f(\beta|\sigma^2, \hat{y}) f(\sigma^2|\beta) d\sigma^2 \] (26)
where \( f(\beta|\delta^2, \hat{\gamma}) \) is actually a multivariate normal distribution for each value of \( \delta^2 \) corresponding to \((\beta|\delta^2, \hat{\gamma}) \sim N(\beta_0, C^{-1})\).

Thus posterior moments of \( \beta \) are easily evaluated numerically using

\[
\mu_j = \int E(\beta|\hat{\gamma}) d\beta = \int E(\beta|\hat{\gamma}) f(\delta^2|\hat{\gamma}) d\delta^2
\]

\[
Var(\beta|\hat{\gamma}) = \int \{ (\beta - \mu_0)^2 f(\beta|\delta^2, \hat{\gamma}) / C^{-1} \} f(\delta^2|\hat{\gamma}) d\delta^2
\]

This result turns out to be a simple extension to the GLS procedure developed by Stedinger and Tasker [1985]. With use of efficient numerical integration procedures, the integrals in (27) and (28), as well as the mean and variance of \( \delta^2 \), are easily computed.

4. Examples for Skew Estimation

The examples presented below address regional models of the log space skewness coefficient. The current methodology for flood frequency analysis in the U.S. consists of fitting a log Pearson type III (LP3) distribution to the gauged data by estimating the mean, standard deviation, and skew of the logarithms of the flows. A problem is that the at-site skewness estimator is highly variable with typical record lengths. In order to improve the precision of the estimator, Bulletin 17B recommends combining the at-site estimator with a regional estimate of the skew coefficient [IACWD, 1982; see also McCuen, 1979; Tung and Mays, 1981a, 1981b; McCuen and Hromadka, 1988; Griffis et al., 2004].

This section illustrates the use of the new Bayesian GLS procedure by applying it to regional skew estimation for two river basins, the Muskingum basin (Ohio, USA) and Tibagi basin (Parana, Brazil). The section begins with a description of the prior distributions for the \( \beta \) parameters and for the model error variance \( \delta^2 \), and the estimation of the sampling covariance matrix for the regional skew problem. The section concludes with a discussion of the results for both basins.

4.1. Prior on the Parameters of the Model

The standard family of distributions for the prior for the model error variance is the inverse gamma. The reason is that the inverse gamma distribution is a conjugate prior for normal regression problems when the covariance matrix is known except for the scale \( \delta^2 \), as in model (2). However, the inverse gamma distribution is not a conjugate prior for our GLS model described in (6) and (7), so its use is not as appealing. Besides, the inverse gamma distribution has a heavy right-hand tail, meaning it provides a relatively large probability for big variances when compared to other distributions such as for the exponential. Therefore a distribution with such a heavy tail does not seem appropriate as an informative prior for the error variance of regional skew models. Moreover, the pdf of the inverse gamma distribution goes to zero at the origin so that it does not allow the model error variance to be equal to zero even if the likelihood function is strictly positive at zero and in a neighborhood of zero.

To avoid these problems, an exponential distribution is employed for the prior. The exponential distribution has a thinner right-hand tail, which is more consistent with what we believe to be the likely values of the error variance for regional skew models. It also has a nonzero pdf at zero, which would allow the data, represented by the likelihood function, to provide information about the error variance near zero. The exponential pdf is

\[
\xi(\delta^2) = \lambda e^{-\lambda \delta^2}, \quad \delta^2 > 0
\]

Were skews uniformly distributed between \( \pm 1 \), a seemingly extreme case given reported values [Hardison, 1974; Landwehr and Matalas, 1978; IACWD, 1982; Griffis and Stedinger, 2005], then the variance would be 0.33. So, the model error variance should be less than 0.33. However, in order to be conservative, the prior distribution should give a modest probability of \( \delta^2 \) being greater than 0.33 in a region is particularly abnormal. Thus the parameter \( \lambda \) was set equal to 6, which represents a model error variance with mean equal to 1/6 and a probability of just 0.14 that \( \delta^2 \) is greater than 0.33. In order to understand the influence the parameter \( \lambda \) has on the estimated parameters of the regional model and on the model error variance, a sensitivity analysis was carried out and the results are discussed in section 5.

If there is no previous information on the parameters \( \beta \), an almost noninformative prior can be used. In our case, the elements of the mean vector \( \beta_0 \), and precision matrix \( P \) in equation (17) are set equal to zero. Then, the posterior distribution of the model error variance is described by equation (25) with equation (29) describing the prior on \( \delta^2 \), which was employed in these examples.

4.2. Estimation of Sampling Covariance Matrix \( \Sigma \)

The estimation of the matrix \( \Sigma \) was based upon the recent results of Griffis [2003] and Martins and Stedinger [2002b]. Griffis [2003] provides the following accurate approximation of the variance of the traditional skewness
The factor \(\frac{1}{1 + \frac{6}{N}}\) of the skewness coefficient estimators determined the relationships between the cross correlation and Stedinger [2002a]. Their Monte Carlo experiments \(W_{10419}\) and therefore is in the range \(2.8 \leq \kappa \leq 3.3\), \(n_j\) is the common sample size and \(\gamma\) is the true value of skew. The factor \(\frac{1}{1 + \frac{6}{N}}\) in equation (30) should be employed only if the bias correction factor proposed by Tasker and Stedinger [1986] is adopted in the estimation of the at-site skew, which is computed by

\[
\gamma = \left[ 1 + \frac{6}{N} \right] \left\{ \frac{N \sum_{i=1}^{N} (z_i - \bar{z})^3}{(N-1)(N-2)s^3} \right\}
\]

(31)

Here \(z_i\) is the logarithm of the annual peak flow in year \(t\), and \(s\) is the standard deviation of the \(z_i\). Because the true values of skews at each site are unknown, the regional average of the skews was used in (30).

[44] The cross correlations among skewness estimators \(\rho(\gamma_i, \gamma_j)\) can be obtained using the expressions of Martins and Stedinger [2002a]. Their Monte Carlo experiments determined the relationships between the cross correlation of the skewness coefficient estimators \(\gamma\) of Pearson type 3 series as a function of the distance between gauges as shown in Figure 3. The cross correlation between skewness estimators was described by \(\rho(\gamma_i, \gamma_j) = cf_{ij} (1 - 0.00291d_{ij})^{-0.85}\), where \(d_{ij}\) is the distance between the gauges in kilometers for \(d < 300\) km.

4.3 Results

[46] A Bayesian GLS analysis was used to develop regional skew models for the Muskingum Basin (Ohio, USA) with 44 stations and record lengths varying from 23 to 93 years, and the Tibagi Basin (Parana, Brazil) with 17 stations and record lengths ranging from 15 to 65 years. The slope and the logarithms of both drainage area and length were used as explanatory variables in the regression analysis for the Muskingum, while the logarithms of slope and drainage area were used for Tibagi. All explanatory variables were centered so that the constant of the model represents the regional mean of the skews.

[47] The skewness coefficient was regressed against every possible combination of the explanatory variables. Tables 1 and 2 present the results of the Bayesian GLS for the two areas. Tables 1 and 2 include the average variance of prediction among the models analyzed, and zero

![Table 1. Results of the Bayesian GLS Regression for the Muskingum Basin (44 Sites)](image)

<table>
<thead>
<tr>
<th>Model</th>
<th>Constant</th>
<th>ln(A)</th>
<th>ln(L)</th>
<th>Slope</th>
<th>Model Error Variance</th>
<th>Average Sampling Variance</th>
<th>AVP_new</th>
<th>AVP_old</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.221a</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.124</td>
<td>0.009</td>
<td>0.133</td>
<td>0.128</td>
</tr>
<tr>
<td>1</td>
<td>0.214b</td>
<td>0.022</td>
<td>-</td>
<td>-</td>
<td>0.129</td>
<td>0.017</td>
<td>0.146</td>
<td>0.134</td>
</tr>
<tr>
<td>2</td>
<td>0.214b</td>
<td>0.274</td>
<td>-0.471</td>
<td>-</td>
<td>0.124</td>
<td>0.023</td>
<td>0.147</td>
<td>0.130</td>
</tr>
<tr>
<td>3</td>
<td>0.224a</td>
<td>0.237</td>
<td>-0.695b</td>
<td>-0.034b</td>
<td>0.087</td>
<td>0.027</td>
<td>0.115</td>
<td>0.099</td>
</tr>
<tr>
<td>4</td>
<td>0.216c</td>
<td>-</td>
<td>0.024</td>
<td>-</td>
<td>0.130</td>
<td>0.017</td>
<td>0.147</td>
<td>0.135</td>
</tr>
<tr>
<td>5</td>
<td>0.202a</td>
<td>-</td>
<td>-</td>
<td>-0.014b</td>
<td>0.113</td>
<td>0.016</td>
<td>0.129</td>
<td>0.119</td>
</tr>
<tr>
<td>6</td>
<td>0.222a</td>
<td>-0.120b</td>
<td>-</td>
<td>-0.031b</td>
<td>0.103</td>
<td>0.023</td>
<td>0.125</td>
<td>0.111</td>
</tr>
<tr>
<td>7</td>
<td>0.227a</td>
<td>-</td>
<td>-0.276b</td>
<td>-0.035b</td>
<td>0.092</td>
<td>0.022</td>
<td>0.114</td>
<td>0.101</td>
</tr>
</tbody>
</table>

*aZero is not contained in the 95% credible region for the parameter.
*bZero is not contained in the 90% credible region for the parameter.

![Table 2. Results of the Bayesian GLS Regression for the Tibagi Basin (17 Sites)](image)

<table>
<thead>
<tr>
<th>Model</th>
<th>Constant</th>
<th>ln(Slope)</th>
<th>ln(A)</th>
<th>Model Error Variance</th>
<th>Average Sampling Variance</th>
<th>AVP_new</th>
<th>AVP_old</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-0.195</td>
<td>-</td>
<td>-</td>
<td>0.065</td>
<td>0.046</td>
<td>0.110</td>
<td>0.103</td>
</tr>
<tr>
<td>1</td>
<td>-0.152</td>
<td>-0.235</td>
<td>-</td>
<td>0.046</td>
<td>0.054</td>
<td>0.100</td>
<td>0.090</td>
</tr>
<tr>
<td>2</td>
<td>-0.207</td>
<td>0.118</td>
<td>0.055</td>
<td>0.055</td>
<td>0.055</td>
<td>0.113</td>
<td>0.100</td>
</tr>
<tr>
<td>3</td>
<td>-0.145</td>
<td>-0.261</td>
<td>-0.026</td>
<td>0.051</td>
<td>0.075</td>
<td>0.126</td>
<td>0.108</td>
</tr>
</tbody>
</table>

*aZero is contained in the 90% credible region for all parameters.
Table 3. Results of the Skew Regression for the Muskingum Basin (Model 7)\(^a\)

<table>
<thead>
<tr>
<th>Method</th>
<th>Constant</th>
<th>ln(Len)</th>
<th>Slope</th>
<th>Model Error Variance</th>
<th>Average Sampling Variance</th>
<th>AVP(_{new})</th>
<th>AVP(_{old})</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.234 (0.072)</td>
<td>–0.254 (0.110)</td>
<td>–0.034 (0.011)</td>
<td>0.228</td>
<td>0.016</td>
<td>0.234</td>
<td>0.212</td>
</tr>
<tr>
<td>MM-GLS</td>
<td>0.226 (0.090)</td>
<td>–0.277 (0.119)</td>
<td>–0.035 (0.011)</td>
<td>0.072</td>
<td>0.020</td>
<td>0.092</td>
<td>0.082</td>
</tr>
<tr>
<td>MLE-GLS</td>
<td>0.226 (0.090)</td>
<td>–0.277 (0.119)</td>
<td>–0.035 (0.012)</td>
<td>0.072</td>
<td>0.021</td>
<td>0.093</td>
<td>0.082</td>
</tr>
<tr>
<td>Bayesian GLS</td>
<td>0.227 (0.093)</td>
<td>–0.276 (0.124)</td>
<td>–0.035 (0.012)</td>
<td>0.092 (0.051)</td>
<td>0.022</td>
<td>0.114</td>
<td>0.101</td>
</tr>
</tbody>
</table>

\(^a\)Standard errors are presented in parentheses (44 sites).

\(^b\)Statistically significant at 95% level except for the Bayesian GLS when it means that zero is not contained in the 95% credible region for the parameter.

is not contained in the 95% credible regions for any of the parameters. The AVP\(_{new}\) is equal to 0.114, which means that the predicted skew is roughly equivalent to an at-site skewness estimator based on 63 years of data (using equation (30) with \(\gamma = 0.227\)). A simple constant, model 0, has an AVP\(_{new}\) of 0.133 which is equivalent to 55 years of record, and is almost as good.

For the Tibagi basin, model 0, which is simply a regional mean skew, is most attractive because the other three models have coefficients whose 90% credible regions include zero. Model 0 would predict a regional skew that would be roughly equivalent to an at-site skewness estimator based on 64 years of data. One could argue that the constant in model 0 should be set to zero because the 90% credible interval contains zero. This reasoning is based on the hypothesis the nominal value for the regional skew is zero. However, there is little reason to believe that zero is the right mean value: there is no theory that suggests the distribution of the logarithms of the maximum floods should be exactly lognormal, corresponding to a regional skew of zero.

Table 3 and 4 compare the results of the Bayesian GLS with those of the OLS, and MM-GLS, and MLE-GLS when applied to the recommended models: model 7 for Muskingum and model 0 for Tibagi. The results of model 1 for Tibagi were also included in Table 4 because both MM-GLS and MLE-GLS results indicate that this model should be selected.

One can see that OLS yields a larger model error variance and average variance of prediction when compared to the GLS model with MM, MLE, and Bayesian model error variance estimators. This happens because OLS does not distinguish between the variance due to the model error and the variance due to time sampling error in \(\hat{y}\).

Tables 3 and 4 compare the results of the Bayesian GLS with those of the OLS, and MM-GLS, and MLE-GLS when applied to the recommended models: model 7 for Muskingum and model 0 for Tibagi. The results of model 1 for Tibagi were also included in Table 4 because both MM-GLS and MLE-GLS results indicate that this model should be selected.

One can see that OLS yields a larger model error variance and average variance of prediction when compared to the GLS model with MM, MLE, and Bayesian model error variance estimators. This happens because OLS does not distinguish between the variance due to the model error and the variance due to time sampling error in \(\hat{y}\). The

parameters of the regional model determined by MM, MLE, and the Bayesian procedure are very close for these two cases. For the Muskingum basin, the Bayesian analysis generated a posterior model error variance equal to 0.092, which is larger than those obtained by both the MM and MLE. This difference results in an increase in both the standard errors of the parameters of the model and the average variance of prediction.

For the Tibagi basin with model 1, the GLS results obtained with MM and MLE model error variance estimators do not agree with those obtained with the Bayesian approach. Both the MM and MLE generate a model error variance equal to zero for model 1, suggesting that this model should be selected because it has the smallest average variance of prediction among the models considered. However, the Bayesian approach obtained an average model error variance equal to 0.046 for model 1, which increases the standard error of the parameters, and the 90% credible region for \(\hat{\beta}_{\text{slope}}\) then includes zero, indicating model 0 should be selected. The Bayesian result makes more sense in this case because, unlike MM and MLE, it considers the possibility the regional model is not perfect, as is almost surely the case.

The problem of zero estimates for variance components is a widely recognized issue in the statistical literature. For example, Berger [1985, pp. 81, 117–119] encountered this situation in a mathematically similar problem wherein he attempted to estimate the variance for a prior distribution for the unknown means \(\mu_i\) for several populations \(N[\mu_i, \sigma^2]\), wherein observations were available from each population, and \(\sigma^2\) is known. More recently, Crainiceanu and Ruppert [2004] address inference for a complex normal linear mixed model with several variance components; they found that MLEs of some variance components were often zero and as

Table 4. Results of the Skew Regression for the Tibagi Basin (Models 1 and 0)\(^b\)

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>Constant</th>
<th>ln(Slope)</th>
<th>Model Error Variance</th>
<th>Average Sampling Variance</th>
<th>AVP(_{new})</th>
<th>AVP(_{old})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>OLS</td>
<td>–0.129 (0.107)</td>
<td>-</td>
<td>0.196</td>
<td>0.012</td>
<td>0.207</td>
<td>0.184</td>
</tr>
<tr>
<td>0</td>
<td>MM-GLS</td>
<td>–0.191 (0.199)</td>
<td>-</td>
<td>0.014</td>
<td>0.040</td>
<td>0.054</td>
<td>0.052</td>
</tr>
<tr>
<td>0</td>
<td>MLE-GLS</td>
<td>–0.194 (0.204)</td>
<td>-</td>
<td>0.029</td>
<td>0.042</td>
<td>0.071</td>
<td>0.067</td>
</tr>
<tr>
<td>0</td>
<td>Bayesian GLS</td>
<td>–0.195 (0.213)</td>
<td>-</td>
<td>0.065 (0.056)</td>
<td>0.046</td>
<td>0.110</td>
<td>0.103</td>
</tr>
<tr>
<td>1</td>
<td>OLS</td>
<td>–0.129 (0.110)</td>
<td>–0.073 (0.161)</td>
<td>0.206</td>
<td>0.024</td>
<td>0.230</td>
<td>0.182</td>
</tr>
<tr>
<td>1</td>
<td>MM-GLS</td>
<td>–0.135 (0.196)</td>
<td>–0.264 (0.114)</td>
<td>0.000</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
</tr>
<tr>
<td>1</td>
<td>MLE-GLS</td>
<td>–0.135 (0.196)</td>
<td>–0.264 (0.115)</td>
<td>0.000</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
</tr>
<tr>
<td>1</td>
<td>Bayesian GLS</td>
<td>–0.152 (0.210)</td>
<td>–0.235 (0.148)</td>
<td>0.046 (0.049)</td>
<td>0.054</td>
<td>0.100</td>
<td>0.090</td>
</tr>
</tbody>
</table>

\(^a\)Standard errors are presented in parentheses (17 sites).

\(^b\)Statistically significant at 95% level except for the Bayesian GLS when it means that zero is not contained in the 95% credible region for the parameter.
a result a likelihood ratio test was very ineffective. Use of a Bayesian framework with a reasonable prior on $\delta^2$ addresses the problems encountered in the analysis of the GLS model.

4.4. Posterior Distribution of $\delta^2$

[54] Looking at the graphs of the marginal posterior and prior distributions of $\delta^2$, as well as of the likelihood function for the data helps one understand the differences between the maximum likelihood and the Bayesian estimators of the model error variance for the two basins studied. For this purpose the marginal and profile likelihood functions are useful one-dimensional projections on $\delta^2$ of the joint likelihood function.

[55] The profile likelihood function $L(\delta^2|\beta = \hat{\beta}, \hat{y})$ represents the maximum value of the likelihood function $L(\beta, \delta^2|\hat{y})$ for a given value of $\delta^2$, and is computed using equation (18) with $\beta$ given by equation (8) [Sprott, 2000,

Figure 4. Profile and marginal likelihood functions and both prior (dashed) and posterior (solid) pdfs for model error variance $\delta^2$ for Muskingum basin (model 7).

Figure 5. Profile and marginal likelihood functions and both prior (dashed) and posterior (solid) pdfs for model error variance $\delta^2$ for Tibagi basin, model 1.
The marginal likelihood $L(\theta^2 | y)$ is the function that when multiplied by the prior distribution $\xi(\theta^2)$ results in the marginal posterior distribution $f(\theta^2 | y)$ for $\theta^2$. The marginal likelihood function is obtained by integrating the product of the likelihood function $L(\beta, \theta^2 | y)$ and the prior for $\beta$ parameters $\xi(\beta)$ over the possible values of $\beta$, thus:

$$L(\theta^2 | y) = \int L(\beta, \theta^2 | y) \xi(\beta) d\beta \quad (33)$$

As a result the marginal posterior distribution for $\theta^2$ equal to $f(\theta^2 | y)$ is proportional to $L(\theta^2 | y)\xi(\theta^2)$.

Figure 4 and 5 show the marginal $L(\theta^2 | y)$ and the profile $L(\theta^2 | \beta = \hat{\beta}, y)$ likelihood functions for $\theta^2$, and both its prior and the posterior distributions for the Muskingum (model 7) and Tibagi (model 1) basins, respectively. The same pattern shown in Figure 4 is also observed for Tibagi (model 0) basin.

One can see that the exponential prior used in the Bayesian analysis has some influence on the posterior distribution in both basins. In all three cases, including Tibagi (model 0), the posterior density function for the model error variance is nonzero at the origin, as will always be the case if $|\Sigma| \neq 0$.

From the profile likelihood function for model 1 for Tibagi in Figure 5 it is clear why the MLE of the model error variance is zero in this case: the mode of the profile likelihood function is located at $\theta^2 = 0$. Figures 4 and 5 also help to explain why MLEs for the model error variance behave as they do. In all the cases presented here, both the profile and marginal likelihood functions are skewed to the right, so the MLE will be to the left of the center of mass of the likelihood function, thus negative bias in the estimator is to be expected as observed by Stedinger and Tasker [1986] (see Figure 1). The Bayesian estimator, on the other hand, is based on the center of mass of the marginal posterior distribution, and should better represent the possible values of the model error variance given the data available and the prior information. An advantage of Bayesian estimators is that on average (over the prior distribution), the posterior mean of the model error variance is an unbiased estimate of the true average model error variance. [Stedinger, 1983].

### 5. Sensitivity Analysis for the $\theta^2$ Prior Distribution

The results in section 4 are based on an exponential prior distribution for the model error variance whose parameter $\lambda$ represents the reciprocal of the mean of the prior distribution. A value of $\lambda = 6$ was used in section 4. This section considers the impact of the choice of $\lambda$ on the posterior mean and variance of $\theta^2$, on the posterior distribution of the parameters $\beta$, and on the average variance of prediction.

The regression analysis was repeated for both the Muskingum and Tibagi basins using values of $\lambda$ between 3 and 15. The smaller the value of $\lambda$, the larger the mean of the prior, and less information the prior provides because the prior pdf is flatter and less precise with a variance equal to the mean squared.

For both basins, the models with the smallest average variance of prediction are always the same as those presented in Tables 3 and 4, regardless of the adopted $\lambda$ value over this range. See Tables 5 and 6. The differences in the values of the $\beta$ parameters are also unimportant, less than 0.5% in both basins for $3 < \lambda < 15$ when compared to the base case with $\lambda$ equal to 6. Differences in the standard errors of the $\beta$ parameters are less than 4.0% for the Muskingum and are around 2.0% for the Tibagi. As expected, the major differences from use of different $\lambda$ values are found in the posterior distribution of the model error variance itself, and in the average variance of prediction.

Figure 6 shows how the posterior mean of the model error variance and average variance of prediction change with $\lambda$. One can see the influence of the parameter $\lambda$ on the final regression estimation is relatively small. A 50% decrease in $\lambda$ from 6 to 3 (100% increase in the prior mean) increases the posterior model error variance by less than 10% in the Muskingum basin and about 18% in the Tibagi basin.

### Table 5. Sensitivity Analysis on the Prior for the Model Error Variance Using Model 0 for Tibagi Basin

| $\lambda$ | Constant | Model Error Variance | Average Sampling Variance | $\xi(\theta^2 = 0) \xi(\theta^2 = 0.3 | \lambda)$ |
|-----------|-----------|---------------------|--------------------------|-------------------------------------------------|
| 3         | $-0.195 (0.216)$ | 0.076 (0.068) | 0.047 | 0.123 | 0.114 | 2.5 |
| 6         | $-0.195 (0.214)$ | 0.065 (0.056) | 0.046 | 0.110 | 0.103 | 6.0 |
| 9         | $-0.194 (0.212)$ | 0.057 (0.048) | 0.045 | 0.101 | 0.095 | 14.9 |
| 12        | $-0.194 (0.210)$ | 0.050 (0.043) | 0.044 | 0.094 | 0.088 | 36.6 |
| 15        | $-0.194 (0.209)$ | 0.045 (0.038) | 0.043 | 0.089 | 0.084 | 90.0 |

### Table 6. Sensitivity Analysis on the Prior for the Model Error Variance for Muskingum Basin (Model 7)

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>Constant</th>
<th>ln(Length)</th>
<th>Slope</th>
<th>Model Error Variance</th>
<th>Average Sampling Variance</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.227 (0.095)</td>
<td>$-0.275 (0.126)$</td>
<td>$-0.035 (0.012)$</td>
<td>0.100 (0.055)</td>
<td>0.023</td>
<td>0.123</td>
</tr>
<tr>
<td>6</td>
<td>0.227 (0.094)</td>
<td>$-0.276 (0.124)$</td>
<td>$-0.035 (0.012)$</td>
<td>0.092 (0.051)</td>
<td>0.022</td>
<td>0.114</td>
</tr>
<tr>
<td>9</td>
<td>0.227 (0.093)</td>
<td>$-0.276 (0.122)$</td>
<td>$-0.035 (0.012)$</td>
<td>0.085 (0.048)</td>
<td>0.021</td>
<td>0.106</td>
</tr>
<tr>
<td>12</td>
<td>0.227 (0.092)</td>
<td>$-0.276 (0.120)$</td>
<td>$-0.035 (0.012)$</td>
<td>0.078 (0.045)</td>
<td>0.021</td>
<td>0.099</td>
</tr>
<tr>
<td>15</td>
<td>0.227 (0.091)</td>
<td>$-0.277 (0.119)$</td>
<td>$-0.035 (0.011)$</td>
<td>0.073 (0.042)</td>
<td>0.021</td>
<td>0.093</td>
</tr>
</tbody>
</table>
the other hand, a 150% increase in \( \lambda \) from 6 to 15 (60% decrease in the prior mean) decreases the model errors variance by about 22% in the Muskingum and 30% in the Tibagi. The same degree of influence is observed in the standard error of the model error variance (Tables 5 and 6). It is clear that decreasing the mean of the prior has a greater effect in the estimation of the model error variance than increasing it. That happens because as the mean of the exponential prior gets smaller, the variance, equal to the mean squared for the exponential distribution, also gets smaller. The two effects combine to result in a very informative posterior distribution for the model error variance. The last column in Table 5, which shows the ratio between the values of the prior pdf for \( \delta^2 = 0 \) and 0.3, respectively, illustrates the point. The values in the last column of Table 5 apply to both basins. It is surprising how little the effect was from choice of very different values of \( \lambda \) for the exponential prior.

6. Conclusions

[64] This paper develops a Bayesian analysis of the GLS model for regionalization of hydrologic data. The posterior distribution of the model error variance is derived, except for a normalization constant, allowing numerical computation of the pdf, mean, and standard error of the model error variance and the \( \beta \) parameters using simple one-dimensional numerical integrations. The quasi-analytic Bayesian approach provides the full posterior distribution of regional model parameters, as well as of the model error variance. This fills the gap left by the traditional GLS procedure advanced by Tasker and Stedinger [1989] that lacks a description of the precision of the computed model error variance. In some cases, where the sampling errors Var[\( \gamma \)] are larger than the regional model error \( \delta^2 \), the earlier GLS procedure obtains a model error variance equal to zero, which causes an overestimation of the precision of the regional estimate leading eventually to an underestimation of the uncertainty in the flood quantile and too much reliance on the regional model. The Bayesian approach with an appropriate prior provides a more reasonable description of the possible values of the model error variance in these cases. The quasi-analytic solution provides a simple and needed extension of the GLS procedure described by Stedinger and Tasker [1985] and Tasker and Stedinger [1989] and the WLS regional skew model of Tasker and Stedinger [1986]. It can equally as well be used for estimation of skews, regional \( L \) skews, regional \( \kappa \) for GEV models, or other statistics of interest.

[65] Two examples considered regional models of the coefficient of skewness for the log Pearson type 3 distribution. They showed that the new Bayesian GLS procedure with its more careful interpretation of the data resulted in selection of different models and different estimates of the model error variance than obtained with ordinary least squares or the traditional method of moments and maximum likelihood model error variance estimators. Both regional skew estimators were as precise as 60 or more years of records at site.

[66] Acknowledgments. We thank the Brazilian Government agency CNPq (Conselho Nacional de Desenvolvimento Cientifico e Tecnologico), who provided support for much of this research. We also thank the Associate Editor and six reviewers whose constructive criticism improved the manuscript.

References

Congdon, P. (2001), Bayesian Statistical Modelling, John Wiley, Hoboken, N. J.


Interagency Advisory Committee on Water Data (IACWD) (1982), Guidelines for determining flood flow frequency, Bull. 17B, 28 pp., Hydrol. Subcomm., Washington, D. C.


Reis, D. S., Jr. (2005), Flood frequency analysis employing Bayesian regional regression and imperfect historical information, Ph.D. thesis, Sch. of Civ. and Environ. Eng., Cornell Univ., Ithaca, N. Y.

Reis, D. S., Jr., and J. R. Stedinger (2005), Bayesian MCMC flood frequency analysis with historical information, J. Hydrol., in press.


Richardson, S., and N. G. Best (2003), Bayesian hierarchical models in ecological studies of health-environment effects, Environmetrics, 14(2), 129–147.


Tung, Y., and L. Mays (1981b), Reducing hydrologic parameter uncer-

Vicens, G. J., I. Rodríguez-Iturbe, and J. C. Schaake, Jr. (1975), A Bayesian

Vogel, R. M., and C. N. Kroll (1990), Generalized low-flow frequency

Walker, J. F., and W. R. Krug (2003), Flood-frequency characteristics of

Wikle, C. K. (2003), Hierarchical Bayesian models for predicting the spread

Wikle, C. K., and J. Anderson (2003), Climatological analysis of tornado

World Meteorological Organization (1994), *Guide to Hydrological Prac-
tices. Data Acquisition and Processing, Analysis. Forecasting and Other

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*,
John Wiley, Hoboken, N. J.

E. S. Martins and D. S. Reis Jr., Water Resources Department,
FUNCEME, Avenida Rui Barbosa, 1246 Fortaleza, Ce, Brazil. (esm9@secrel.com.br; direceu@funceme.br)

J. R. Stedinger, School of Civil and Environmental Engineering, Cornell
University, Hollister Hall, Ithaca, NY 14853-3501, USA. (jrs5@cornell. edu)