WATER RESOURCES RESEARCH, VOL. 34, NO. 1, PAGES 121–128, JANUARY 1998

Regional hydrologic analysis: Ordinary and generalized least squares revisited

Charles N. Kroll

Environmental Resources and Forest Engineering, College of Environmental Science and Forestry, State University of New York, Syracuse

Jery R. Stedinger

School of Civil and Environmental Engineering, Cornell University, Ithaca, New York

Abstract. Generalized least squares (GLS) regional regression procedures have been developed for estimating river flow quantiles. A widely used GLS procedure employs a simplified model error structure and average covariances when constructing an approximate residual error covariance matrix. This paper compares that GLS estimator ($\hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}}$), an idealized GLS estimator ($\hat{\beta}^{E}_{\mathrm{GLS}}$) based on the simplifying assumptions of $\hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}}$ with true underlying statistics in a region, the best possible GLS estimator ($\hat{\beta}^{T}_{\mathrm{GLS}}$) obtained using the true residual error covariance matrix, and the ordinary least squares estimator ($\hat{\beta}^{T}_{\mathrm{OLS}}$). Useful analytic expressions are developed for the variance of $\hat{\beta}^{T}_{\mathrm{GLS}}$, $\hat{\beta}^{E}_{\mathrm{GLS}}$, and $\hat{\beta}^{T}_{\mathrm{OLS}}$. For previously examined cases the average sampling mean square error (mse_s) of $\hat{\beta}^{E}_{\mathrm{GLS}}$ was the same as the mse_s of $\hat{\beta}^{T}_{\mathrm{GLS}}$, and the mse_s of $\hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}}$ usually was larger than the mse_s of both $\hat{\beta}^{E}_{\mathrm{GLS}}$ and $\hat{\beta}^{T}_{\mathrm{GLS}}$. The loss in efficiency of $\hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}}$ was mostly due to estimating streamflow statistics employed in the construction of the residual error covariance matrix rather than the simplifying assumptions in presently employed GLS estimators. The new analytic expressions were used to compare the performance of the OLS and GLS estimators for new cases representing greater model variability across sites as well as the effect return period has on the estimators' relative performance. For a more heteroscedastic model error variance and larger return periods, some increase in the mse_s of $\hat{\beta}^{E}_{\mathrm{GLS}}$ relative to the mse_s of $\hat{\beta}^{T}_{\mathrm{GLS}}$ was observed.

1. Introduction

Regional regression models are often used to estimate flow statistics at ungaged river sites. Relationships between flow statistics and geomorphic, geologic, climatic, and topographic parameters have been formulated in many regions for low flows [*Thomas and Benson*, 1970; *Thomas and Cervione*, 1970; *Riggs*, 1972; *Vogel and Kroll*, 1992] and for flood flows [*Benson*, 1962; *Matalas and Gilroy*, 1968; *Thomas and Benson*, 1970; *Jennings et al*., 1994; *Tasker et al*., 1996]. Traditionally, the parameters of these models were estimated using ordinary least squares (OLS) regression procedures employing data for gaged river sites [*Thomas and Benson*, 1970]. For OLS parameter estimators to be efficient the model residuals should be independent and homoscedastic.

Tasker [1980] developed a weighted least squares (WLS) regression technique to account for the varying sampling error in the at-site quantile estimators. *Stedinger and Tasker* [1985, 1986] extended this work by developing generalized least squares (GLS) regression techniques to account for the varying sampling error and the cross correlation among concurrent flows. *Tasker and Stedinger* [1989] discuss an implementation of GLS estimators which also accounts for varying model error variance among sites and variations in the cross correlation of concurrent observations. Using Monte Carlo simulation, *Stedinger and Tasker* [1985] demonstrated that GLS procedures provided more accurate parameter estimators, better estimators of parameter sampling variances, and an almost unbiased estimator of the model error variance. In particular, the average sampling mean square error (mse_s) of the GLS estimators was smaller than the mse_s of the OLS estimators when the model error variance was small or the cross correlations among the annual flows were large. *Moss and Tasker* [1991] showed that GLS procedures describe model accuracy in regional analyses better than OLS procedures do.

This paper addresses a number of unresolved issues regarding GLS and OLS regional regression procedures. The paper (1) clarifies assumptions that have been made when implementing GLS and OLS regional regression procedures, (2) examines the loss of efficiency of OLS estimators and GLS estimators that employ a residual error covariance matrix different than the true residual error covariance matrix, (3) determines whether this loss in efficiency in GLS estimators is due to smoothing of the sampling covariance matrix or to implementing an inadequate model of the model error variance, and (4) determines when OLS estimators are adequate and when a GLS estimator which accounts for varying model error variance is needed to achieve efficient parameter estimates. To examine the efficiency of GLS estimators, the practical GLS estimator developed by *Stedinger and Tasker* [1985] is compared with an idealized GLS estimator that uses the true regional statistics and the simplifying assumptions Stedinger and Tasker used to construct the residual error covariance matrix, and an ideal GLS estimator that uses the true residual error covariance matrix.

Paper number 97WR02685.
0043-1397/98/97WR-02685$09.00

This paper is structured as follows. Section 2 presents the regional regression problem. Section 3 describes *Stedinger and Tasker*'s [1985] Monte Carlo experiments. Section 4 discusses the construction of the residual error covariance matrices. Section 5 compares the results for one of *Stedinger and Tasker*'s [1985] Monte Carlo experiments with analytic expressions. Section 6 uses analytic expressions to compare the efficiency of OLS and GLS estimators for several new cases not considered by *Stedinger and Tasker* [1985, 1986]. Finally, section 7 presents our conclusions.

2. Regional Regression Model

Following the notation by *Stedinger and Tasker* [1985], let θ be a vector of the true flow statistics for river sites in a region and let X be a matrix of drainage basin characteristics associated with the sites augmented by a column of ones. Assume that the relationship between θ and X is described by the linear model

$$\theta = X\beta + \varepsilon \tag{1}$$

where β contains model parameters and ε contains the residual errors. Here var $(\varepsilon_i) = \gamma_i^2$ is the model error variance. In practice the true flow statistic, θ, is not known, and an estimator, $\hat{\theta}$, of the statistic of interest is obtained with available streamflow records. Assume that $\hat{\theta}$ is an unbiased estimator of θ so that

$$E[\hat{\theta}] = \theta \tag{2}$$

and

$$E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] = \Sigma \tag{3}$$

where Σ is the sampling covariance matrix associated with the estimator $\hat{\theta}$. The covariance of $\hat{\theta}$ about the regression mean $X\beta$ defines the residual error covariance matrix,

$$E[(\hat{\theta} - X\beta)(\hat{\theta} - X\beta)^T] = \Lambda_T = \Gamma + \Sigma \tag{4}$$

where Γ = diag $[\gamma_i^2]$, and γ_i^2 is the model error variance associated with site i.

In practice, OLS regression procedures have been used to estimate the parameters of linear models such as (1). If Λ_T is equal to a diagonal matrix with a constant variance γ^2 along the diagonal, then OLS parameter estimators are efficient since they have minimum variance among all unbiased linear estimators [*Johnston*, 1984, p. 173]. In practice the variance, and therefore the diagonal elements of Λ_T, will differ from site to site. However, the assumption of a constant variance (called homoscedasticity) is often adequate for many practical problems [*Draper and Smith*, 1981].

The OLS estimator of the parameter vector is [*Johnston*, 1984, equation 5-26]

$$\hat{\beta}_{\text{OLS}}^T = (X^TX)^{-1}X^T\hat{\theta} \tag{5}$$

Because the residual errors associated with the model in (1) are not independent and identically distributed with equal variances, the standard relationship for the variance of the OLS parameter estimator [*Johnston*, 1984, equation 5-33]

$$\text{var } (\hat{\beta}_{\text{OLS}}^T) = \gamma^2(X^TX)^{-1} \tag{6}$$

does not describe the actual sampling variance of the OLS parameter estimator. Instead the correct expression is [*Johnston*, 1984, equation 8-13]

$$\text{var } (\hat{\beta}_{\text{OLS}}^T) = (X^TX)^{-1}X^T\Lambda_TX(X^TX)^{-1} \tag{7}$$

Equation (7) involves Λ_T, the true residual error covariance matrix.

An extension of OLS parameter estimators are GLS parameter estimators that employ some estimate of Λ_T. These estimators weight each observation to reflect the variance of the residual error associated with that observation and the covariance of the residual error with other residual errors. If Λ_T is known, one can employ the optimal GLS estimator [*Johnston*, 1984, equation 8-20]

$$\hat{\beta}_{\text{GLS}}^T = (X^T\Lambda_T^{-1}X)^{-1}X^T\Lambda_T^{-1}\hat{\theta} \tag{8}$$

whose sampling variance is [*Johnston*, 1984, equation 8-12]

$$\text{var } (\hat{\beta}_{\text{GLS}}^T) = (X^T\Lambda_T^{-1}X)^{-1} \tag{9}$$

Both $\hat{\beta}_{\text{GLS}}^T$ and $\hat{\beta}_{\text{OLS}}^T$ are unbiased estimators of the parameters in the model given by (1); however, $\hat{\beta}_{\text{GLS}}^T$ has a smaller variance than $\hat{\beta}_{\text{OLS}}^T$ because the former correctly weights each of the observations. The parameter estimator $\hat{\beta}_{\text{GLS}}^T$ is the best (minimum variance) linear unbiased estimator (BLUE) of the model parameters [*Greene*, 1990; *Johnston*, 1984].

Stedinger and Tasker [1985] were interested in developing an estimator of the parameters of the model in (1) that had an efficiency approaching that of the GLS estimator $\hat{\beta}_{\text{GLS}}^T$. Employing $\hat{\beta}_{\text{GLS}}^T$ is not practical because Λ_T is unknown; Λ_T requires knowledge of the model error variance associated with each observation, γ_i^2, and the sampling covariance matrix, Σ, both of which need to be estimated. *Stedinger and Tasker* [1985] proposed constructing an approximation of Λ_T, which we denote as Λ_E, by approximating Γ by diag (γ^2) with a constant value of γ^2 that must be estimated, and by approximating Σ using averaged or smoothed estimates of the sampling variances associated with the at-site streamflow statistics of interest. An average value of the cross correlation between concurrent flows in the region was used to calculate the covariance terms which are the off-diagonal elements of Σ. In later work a relationship between the cross correlation and the distance between gaged river sites was employed, and γ^2 was allowed to vary across sites [*Tasker and Stedinger*, 1989].

If smoothed estimates of the variance of individual streamflow observations are available, and the average cross correlation of concurrent streamflows across sites and the average model error variance are known, one can construct Λ_E and compute the corresponding idealized GLS estimator

$$\hat{\beta}_{\text{GLS}}^E = (X^T\Lambda_E^{-1}X)^{-1}X^T\Lambda_E^{-1}\hat{\theta} \tag{10}$$

For particular X and Λ_T matrices the sampling variance of this GLS parameter estimator is

$$\text{var } (\hat{\beta}_{\text{GLS}}^E) = (X^T\Lambda_E^{-1}X)^{-1}X^T\Lambda_E^{-1}\Lambda_T\Lambda_E^{-1}X(X^T\Lambda_E^{-1}X)^{-1} \tag{11}$$

While Stedinger and Tasker wanted to employ Λ_E as an approximation to Λ_T, as a practical matter they had to employ an estimator of Λ_E, which we denote as Λ_{MC}. *Stedinger and Tasker* [1985] encountered problems when at-site sample variances were used to construct an estimator of the sampling covariance matrix Σ, because those weights were then correlated with the at-site quantile estimators, the dependent variable in the regression equation. To avoid such problems, the variance of the individual streamflow observations used to compute the elements of Σ was estimated using a regression

relationship developed between at-site sample variances and physiographic basin characteristics. Λ_{MC} was constructed using the computed estimate of the variance of individual observations from this regression relationship, a computed average of the sample cross-correlation estimators, and a computed generalized mean square error estimator of the average model error variance. In Stedinger and Tasker's Monte Carlo experiment, for every replicate of the experiment a different value of Λ_{MC} was constructed to allow computation of a GLS parameter estimator $\hat{\beta}^{MC}_{GLS}$ for that replicate. They also employed Λ_{MC} to estimate the variance of the GLS parameter estimator as

$$\text{var}\,(\hat{\beta}^{MC}_{GLS}) = (X^T\Lambda^{-1}_{MC}X)^{-1} \tag{12}$$

Thus Stedinger and Tasker had two estimators of the variance of their GLS parameter estimator: the observed empirical variability of the individual $\hat{\beta}^{MC}_{GLS}$ estimators across replicates of the Monte Carlo experiment and the average of (12) across replicates. The observed variability of the individual $\hat{\beta}^{MC}_{GLS}$ estimators across replicates should correspond to an average of equation (11) where Λ_E is replaced by Λ_{MC}:

$$E[\text{var}\,(\hat{\beta}^{MC}_{GLS})] = E[(X^T\Lambda^{-1}_{MC}X)^{-1}X^T\Lambda^{-1}_{MC}\Lambda_T\Lambda^{-1}_{MC}X(X^T\Lambda^{-1}_{MC}X)^{-1}] \tag{13}$$

and the expectation is taken over the joint distribution of Λ_{MC}, Λ_T, and X; for every replicate Stedinger and Tasker randomly generated new drainage areas for the region so that X and the corresponding Λ_T were both random. In the Monte Carlo experiments reported by *Stedinger and Tasker* [1985], the difference between (13) and the average across all Monte Carlo replicates of (12) was relatively small.

In the following sections we compare the variance of Stedinger and Tasker's GLS estimator, $\hat{\beta}^{MC}_{GLS}$, from their Monte Carlo experiment, with the variance of the GLS estimator using the true residual error covariance matrix, $\hat{\beta}^{T}_{GLS}$, and the variance of the GLS estimator, $\hat{\beta}^{E}_{GLS}$, based on a constructed residual error covariance matrix using the average variance of the streamflows at each site, the average cross correlation between concurrent flows, and the average model error variance. The parameter estimators $\hat{\beta}^{T}_{GLS}$ and $\hat{\beta}^{E}_{GLS}$ are based on population values of these statistics, whereas $\hat{\beta}^{MC}_{GLS}$ employed sample estimators of the corresponding parameters. The variance of $\hat{\beta}^{T}_{GLS}$ and $\hat{\beta}^{E}_{GLS}$ will be obtained using the analytic expression in (9) and (11), respectively.

In their Monte Carlo analysis, Stedinger and Tasker randomly generated new values of the drainage areas for every replicate. To account for the random drainage areas, the analytic expressions were averaged over 100 replicates, with drainage areas randomly generated for each replicate. This Monte Carlo analysis was also necessary to generate a true random log space standard deviation of the flows at each site which is needed to construct the true residual error covariance matrix. These issues are discussed in section 4.

3. Stedinger and Tasker's Monte Carlo Experiment

The results from *Stedinger and Tasker*'s [1985] first Monte Carlo experiment will be used to compare the variance of their GLS parameter estimator, $\hat{\beta}^{MC}_{GLS}$, to the variance of $\hat{\beta}^{T}_{GLS}$ and $\hat{\beta}^{E}_{GLS}$. This experiment considered a regional regression model for 50-year floods. The region included 30 sites with drainage area randomly selected from a uniform distribution ranging in logarithmic space from 10 to 20,000 miles^2 (25.9 to 51,800 km^2). It was assumed that the annual maximum flows at each site are lognormally distributed. The log space mean, μ_i, and standard deviation, σ_i, at each site were a function of drainage area at that site, A_i, following the models

$$\mu_i = \alpha_\mu + \beta_\mu \ln(A_i) + \upsilon_i \tag{14}$$

$$\sigma_i = [\alpha_\sigma + \beta_\sigma \ln(A_i)] \exp(\delta_i) \tag{15}$$

where $\alpha_\mu = 0$, $\beta_\mu = 0.75$, $\alpha_\sigma = 1.5$, $\beta_\sigma = -0.14$, and υ_i and δ_i are independent normally distributed random error terms with means 0 and $-0.03125\sigma^2_\upsilon$, and variances σ^2_υ and $\sigma^2_\delta = 0.0625\sigma^2_\upsilon$, respectively. For each site an annual flow record of length n_i was randomly generated from a lognormal distribution with moments given by (14) and (15), and the cross correlation among the logarithms of concurrent flows in a region equal to a constant value, ρ. Using the generated record for a site, the sample moment estimators, $\hat{\mu}_i$ and $\hat{\sigma}_i$, were calculated and used to compute an at-site estimator of a flow quantile

$$\hat{Q}_{p,i} = \hat{\mu}_i + z_p\hat{\sigma}_i \tag{16}$$

where $\hat{Q}_{p,i}$ is an at-site estimate of a flow quantile with a nonexceedance probability of p, and z_p is the pth percentile of a standard normal distribution (for the 50-year flood $p = 0.98$ and $z_p = 2.054$). Combining (14), (15), and (16), the underlying regression model is

$$\hat{Q}_{p,i} = \alpha + \beta \ln(A_i) + \varepsilon_i \tag{17}$$

where $\alpha = \alpha_\mu + z_p\alpha_\sigma$, $\beta = \beta_\mu + z_p\beta_\sigma$, and ε_i is the residual error.

In their first experiment, Stedinger and Tasker considered the cases where $\rho = 0.0$, 0.3, 0.6, and 0.9, and $\sigma_\upsilon = 0.0$, 0.1, 0.3, 0.5, and 0.9. For 10 sites the record length was set to $n_i = 50$, for 10 sites $n_i = 20$, and for the remaining 10 site $n_i = 10$.

4. Construction of the Residual Error Covariance Matrix

The residual error covariance matrix constructed in Stedinger and Tasker's Monte Carlo experiment for their GLS estimator, Λ_{MC}, is an estimator of the covariance matrix, Λ_E, they proposed to employ. Both of these matrices differ from the true residual error covariance matrix, Λ_T. This section examines the assumptions employed to develop the sampling covariance matrix and model error variance for each of the GLS estimators. Table 1 summarizes those assumptions.

4.1. Estimation of Sampling Covariance Matrix

The diagonal elements of the sampling covariance matrix, Σ, correspond to the variance of the at-site quantile estimators, and the off-diagonal elements correspond to the covariance among quantile estimators for different sites. When the individual observations are normally distributed, the correct diagonal elements of the sampling covariance matrix, Σ_{ii}, are

$$\Sigma_{ii} = \text{var}\,(\hat{Q}_{p,i}) = \text{var}\,(\hat{\mu}_i + z_p\hat{\sigma}_i) = \text{var}\,(\hat{\mu}_i) + z_p^2\,\text{var}\,(\hat{\sigma}_i) = \frac{\sigma_i^2}{n_i} + z_p^2\sigma_i^2\left[1 - \left(\frac{2}{n_i - 1}\right)\left(\frac{\Gamma(n_i/2)}{\Gamma((n_i-1)/2)}\right)^2\right] \tag{18}$$

Equation (18) incorporates an exact expression for the vari-

Table 1. Alternative Assumptions and Equations Employed to Construct Residual Error Covariance Matrices, Λ, for the GLS Estimators $\hat{\beta}^{T}_{\mathrm{GLS}}$, $\hat{\beta}^{E}_{\mathrm{GLS}}$, and $\hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}}$

Assumption/Equation	$\hat{\beta}^{T}_{\mathrm{GLS}}$ (Λ^{T})	$\hat{\beta}^{E}_{\mathrm{GLS}}$ (Λ_E)	$\hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}}$ (Λ_{MC})
Model error variance for each parameter estimator	$\Gamma_{ii} = \gamma_i$, true model error variance (equation (32))	$\Gamma_{ii} = \bar{\gamma}$, expected model error variance (equation (33))	$\Gamma_{ii} = \hat{\gamma}$, estimator of average model error variance for each replicate (equation (34))
Estimator of variance of observations, σ_i^2, employed to calculate sampling error associated with at-site quantile estimators, $\Sigma_{ii} = \mathrm{var}\ (\hat{\theta}_i)$	σ_i^2, true value of variance (equation (15))	$E[\sigma_i^2]$, expected value of variance (equation (28))	$(E[\hat{\sigma}_i])^2$, square of estimated expected value of standard deviation (equation (29))
Formula employed to compute the variance of at-site quantile estimator, $\Sigma_{ii} = \mathrm{var}\ (\hat{\theta}_i)$	exact estimator (equation (18))	first-order approximation (equation (20))	first-order approximation (equation (20))
Correlation coefficient used in first-order approximation (equation (21)) to compute covariances among at-site quantile estimators, $\Sigma_{ij} = \mathrm{cov}\ (\hat{\theta}_i, \hat{\theta}_j)$	true value	true value	average regional estimate for each replicate

ance of the at-site estimator of the standard deviation [*David*, 1981] and is employed in Λ_T. Stedinger and Tasker employed the first-order approximation

$$\mathrm{var}\ (\hat{\sigma}_i) = \frac{\sigma_i^2}{2n_i} \tag{19}$$

for the variance of the at-site estimator of the standard deviation when constructing the sampling covariance matrix [*Stedinger*, 1983]. Using this approximation, (18) becomes

$$\mathrm{var}\ (\hat{Q}_{p,i}) = \frac{\sigma_i^2}{n_i}\left[1 + \frac{z_p^2}{2}\right] \tag{20}$$

Stedinger and Tasker used a smoothed estimator of the at-site variance of the observations, σ_i^2, in (20) when constructing their residual covariance, Λ_{MC}. The idealized residual error covariance matrix, Λ_E, implements (20) with the expected value of σ_i^2 for each site i.

Stedinger and Tasker also employed a first-order approximation to the covariance among at-site quantile estimators which are the off-diagonal elements of the sampling covariance matrix,

$$\Sigma_{ij} = \frac{\rho_{ij} m_{ij} \sigma_i \sigma_j}{n_i n_j}\left[1 + \rho_{ij} z_p^2/2\right] \qquad i \neq j \tag{21}$$

where m_{ij} is the number of concurrent years of data at sites i and j, and ρ_{ij} is the lag zero correlation coefficient between the annual flows at sites i and j. The exact expression for the covariance among the at-site quantile estimators is quite complex, so (21) will be employed to compute Λ_T and Λ_E in the calculations reported here. Λ_T and Λ_E will employ the true value of the cross-correlation which was constant across the region in Stedinger and Tasker's Monte Carlo experiment. The matrix associated with the GLS estimator implemented by Stedinger and Tasker, Λ_{MC}, used a regional average of at-site estimators of the correlation coefficient.

In (18) and (20), which describe the sampling variance associated with the at-site quantile estimators, an estimate of the variance of the annual flows, σ_i^2, is required. The expected variance of the log-space annual flows is

$$E[\sigma_i^2] = \mathrm{var}\ [\sigma_i] + (E[\sigma_i])^2 \tag{22}$$

The formula for the at-site standard deviation given by (15) assumes that the standard deviation is lognormally distributed. In terms of the logarithm of (15)

$$\ln\ (\sigma_i) = \ln\ [\alpha_\sigma + \beta_\sigma \ln\ (A_i)] + \delta_i \tag{23}$$

the log space moments of the standard deviation are

$$E[\ln\ (\sigma_i)] = \ln\ [\alpha_\sigma + \beta_\sigma \ln\ (A_i)] + E[\delta_i] \tag{24}$$

and

$$\mathrm{var}\ [\ln\ (\sigma_i)] = \mathrm{var}\ [\delta_i] = \sigma_\delta^2 \tag{25}$$

The real-space moments for the lognormal distribution can be calculated using the log space moments [*Loucks et al.*, 1981], resulting in

$$E[\sigma_i] = \alpha_\sigma + \beta_\sigma \ln\ (A_i) \tag{26}$$

and

$$\mathrm{var}\ [\sigma_i] = [\alpha_\sigma + \beta_\sigma \ln\ (A_i)]^2[\exp\ (\sigma_\delta^2) - 1] \tag{27}$$

Substituting (26) and (27) into (22) yields

$$E[\sigma_i^2] = [\alpha_\sigma + \beta_\sigma \ln\ (A_i)]^2 \exp\ (\sigma_\delta^2) \tag{28}$$

Equation (28) corresponds to the average variance of the annual flows at a site with drainage area A_i and is used in the residual error covariance matrix, Λ_E. Equation (28) is not the same as the variance of the flows in each replicate because δ_i in (15) is replaced by an expectation in (28). In this study a Monte Carlo analysis was employed to account for that difference.

In addition, for every replicate of their Monte Carlo experiment, Stedinger and Tasker randomly generated new drainage areas for the region. To account for the variability due to random drainage areas and the difference between σ_i in (15)

and $E[\sigma_i^2]$ in (28), 100 random sets of different drainage areas and δ_i values were generated. In the construction of the true residual error covariance matrix, Λ_T, the random values of δ_i and A_i were used to calculate the true value of σ_i at each site using (15), and this value was used in (18) for constructing one realization of Σ_{ii}. This allows calculation of the average value over 100 replicates of the sampling variance of $\hat{\beta}^E_{\mathrm{GLS}}$ and $\hat{\beta}^T_{\mathrm{GLS}}$ in (9) and (11), respectively, using generated values of X and associated Λ_T and Λ_E matrices.

Stedinger and Tasker used

$$(E[\sigma_i])^2 = (\hat{\alpha}_\sigma + \hat{\beta}_\sigma \ln (A_i))^2 \tag{29}$$

instead of (28) for an estimator of the variance of the observations. This estimator systematically underestimates the variance of the observations, but the bias was small in Stedinger and Tasker's Monte Carlo experiment because σ_δ^2 was much smaller than 1, and thus var (σ_i) was much smaller than $(E[\sigma_i])^2$.

4.2. Estimation of Model Error Variance

The residual error covariance matrix also depends upon the model error variance, γ_i^2, for each site i. Assuming the annual flows are lognormally distributed, γ_i^2 is

$$\begin{aligned}\gamma_i^2 &= \text{var } [\ln (Q_{p,i})] = \text{var } [\mu_i + z_p\sigma_i] \\ &= \text{var } [\mu_i] + z_p^2 \text{ var } [\sigma_i] + 2z_p \text{ cov } [\mu_i, \sigma_i]\end{aligned} \tag{30}$$

The regional model adopted in the experiment had cov $(\mu_i, \sigma_i) = 0$. From (14)

$$\text{var } [\mu_i] = \text{var } [\upsilon_i] = \sigma_\upsilon^2 \tag{31}$$

The var (σ_i) is given in (27). Substituting (27) and (31) into (30) yields

$$\gamma_i^2 = \sigma_\upsilon^2 + z_p^2[\alpha_\sigma + \beta_\sigma \ln (A_i)]^2[\exp (\sigma_\delta^2) - 1] \tag{32}$$

The model error variance is a function of drainage area and thus varies across sites. *Stedinger and Tasker* [1985] proposed a GLS estimator that uses an average model error variance. This assumption was relatively good in their Monte Carlo experiment because the model error variance varied only slightly among sites. In that study $\sigma_\delta^2 = 0.0625\sigma_\upsilon^2$, so that the variance of the standard deviation of the observations, σ_δ^2, was small compared to the variance of the mean of the observations, σ_υ^2, and γ_i^2 in (32) never varied more than 25% from the average model error variance.

The average value of the model error variance is obtained by taking the expectation over the drainage areas of interest on right hand side of (32) to obtain

$$\begin{aligned}E[\gamma_i^2] = \sigma_\upsilon^2 + z_p^2\{&\alpha_\sigma^2 + 2\alpha_\sigma\beta_\sigma E[\ln (A_i)] \\ &+ \beta_\sigma^2 E[(\ln (A_i))^2]\} [\exp (\sigma_\delta^2) - 1]\end{aligned} \tag{33}$$

Stedinger and Tasker's experiments included six values of σ_υ which correspond to $E[\gamma_i^2] = 0.0$, 0.011, 0.102, 0.284, 0.557, and 0.922. These average values can be used to construct a diagonal model error variance matrix Γ with constant elements to compute the residual error covariance matrix Λ_E.

When implementing the GLS estimator in their Monte Carlo experiment, Stedinger and Tasker computed a generalized mean square error estimator of the average model error variance by solving

$$(\hat{\theta} - X\hat{\beta})^T[\Lambda_{\mathrm{MC}}]^{-1}(\hat{\theta} - X\hat{\beta}) = N - k \tag{34}$$

for $\hat{\gamma}^2$ where $\Lambda_{\mathrm{MC}} = \hat{\gamma}^2 I_N + \hat{\Sigma}(\hat{\theta})$, N is the number of sites, k is the number of degrees of freedom in the model, and $\hat{\theta}$ is the parameter estimator for each replicate. This estimator of the average model error variance is dependent upon the at-site data and thus varies from replicate to replicate because of different sets of flows and drainage areas in each replicate.

5. Comparison of Stedinger and Tasker's Monte Carlo Results With Analytic Expressions

In their Monte Carlo analysis (experiment 1) *Stedinger and Tasker* [1985] computed estimates of the variance of the parameter estimators and the average sampling mean square error (mse_s) of the OLS and GLS quantile estimators. The mse_s of a quantile estimator is the average over sites of the squared difference between the quantile estimator and its true value at such sites. The mse_s of unbiased quantile estimators is

$$\begin{aligned}\mathrm{mse}_s = E_{A,\hat{\alpha},\hat{\beta}}\{[\hat{\theta} - \theta]^2\} &= \text{var } (\hat{\alpha}) + 2E[\ln (A)] \text{ cov } (\hat{\alpha}, \hat{\beta}) \\ &+ E[(\ln (A))^2] \text{ var } (\hat{\beta})\end{aligned} \tag{35}$$

In this study the Monte Carlo results from *Stedinger and Tasker* [1985] for the mse_s of $\hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}}$ are compared to calculated values of the mse_s of $\hat{\beta}^T_{\mathrm{GLS}}$ and $\hat{\beta}^E_{\mathrm{GLS}}$ computed using the average values over 100 replicates of the analytic expressions for the variance of the parameter estimators, (9) and (11), respectively. Record lengths of 10, 20, and 50 were randomly assigned to a third of the sites, as in Stedinger and Tasker's first experiment.

For this comparison the efficiencies of the estimators $\hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}}$ and $\hat{\beta}^E_{\mathrm{GLS}}$ are computed as

$$\begin{aligned}\text{Efficiency } \hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}} &= \frac{\mathrm{mse}_s[\hat{\beta}^T_{\mathrm{GLS}}]}{\mathrm{mse}_s[\hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}}]} \\ \text{Efficiency } \hat{\beta}^E_{GLS} &= \frac{\mathrm{mse}_s[\hat{\beta}^T_{\mathrm{GLS}}]}{\mathrm{mse}_s[\hat{\beta}^E_{\mathrm{GLS}}]}\end{aligned} \tag{36}$$

The efficiency of an estimator approaches unity when the mse_s of an estimator is almost as small as the mse_s of $\hat{\beta}^T_{\mathrm{GLS}}$, which has the smallest possible mse_s for a linear unbiased estimator. Figure 1a is a plot of these efficiencies over the range of cross correlations and average model error variances examined in Stedinger and Tasker's experiment 1. The first set of four columns correspond to cross correlations of 0.0, 0.3, 0.6, and 0.9, respectively, when the average model error variance $\bar{\gamma}^2 = 0.0$. Other sets of four columns correspond to different values of $\bar{\gamma}^2$. Figure 1a also contains the efficiencies of the OLS estimator $\hat{\beta}^{\mathrm{MC}}_{\mathrm{OLS}}$ reported by Stedinger and Tasker.

In general, the efficiency of $\hat{\beta}^E_{\mathrm{GLS}}$ is nearly 100% and the efficiency of $\hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}}$ is greater than 90% for the cases examined. The apparent exception is the efficiency for $\hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}}$ when $\rho = 0$ and $\bar{\gamma}^2 = 0.0$, in which case the computed efficiency is only 80%. The computed mse_s of $\hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}}$ for this case (0.004) is small, and Stedinger and Tasker reported only one significant digit; thus this low efficiency is likely due to rounding error. The efficiency of $\hat{\beta}^{\mathrm{MC}}_{\mathrm{GLS}}$ based on the reported variance of the individual parameter estimators was approximately 90% [*Kroll*, 1996], which further confirms that this low efficiency is due to rounding error.

In the experiment reported in Figure 1a the high efficiencies of $\hat{\beta}^E_{\mathrm{GLS}}$ indicate that the assumptions Stedinger and Tasker made when simplifying the residual error covariance matrix had relatively little effect on the performance of the estimator

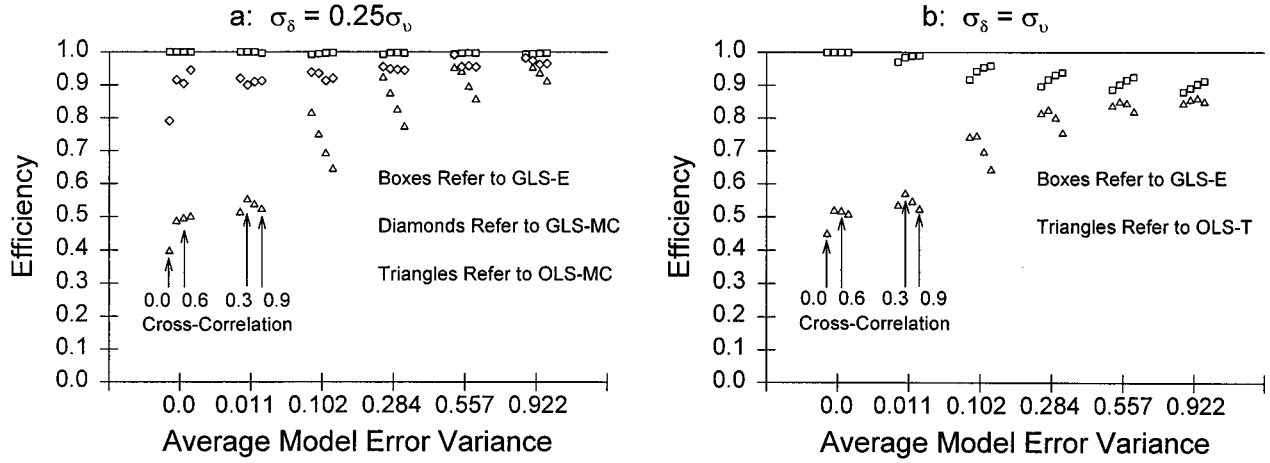

Figure 1. Efficiency of GLS and OLS estimators versus average model error variance for 50-year return period.

compared to an estimator based on knowing the true residual error covariance matrix. The high efficiencies of $\hat{\beta}^{MC}_{GLS}$ indicate the practical implementation of Stedinger and Tasker's proposed GLS estimator $\hat{\beta}^{E}_{GLS}$ in the form of $\hat{\beta}^{MC}_{GLS}$ resulted in little reduction in the performance of the estimator, especially when $\bar{\gamma}^2$ was greater than 0.1.

Figure 1a also includes the efficiency of the OLS estimator $\hat{\beta}^{MC}_{OLS}$ compared to the best GLS estimator $\hat{\beta}^{T}_{GLS}$. For large values of $\bar{\gamma}^2$ the efficiency of the OLS estimator is close to the efficiency of $\hat{\beta}^{MC}_{GLS}$. For small $\bar{\gamma}^2$ the efficiency of the OLS estimator drops considerably. For moderate model error variances, as the cross correlation increases, the relative efficiency of the OLS estimator decreases. The efficiency of the OLS estimator depends on whether the elements along the diagonal of Λ are close to constant and on the relative magnitude of the off-diagonal elements of Λ compared to the diagonal elements. If the off-diagonal elements are relatively small and the diagonal elements are close to constant, the OLS estimator is almost as efficient as the GLS estimator. For small $\bar{\gamma}^2$ the effect of heteroscedasticity due to the sampling error in the at-site quantile estimators is greater than when $\bar{\gamma}^2$ is large, and thus the OLS estimator performs worse when $\bar{\gamma}^2$ is small. *Kroll* [1996] showed that the mse_s and variance of the OLS estimator reported in *Stedinger and Tasker*'s [1985] Monte Carlo analysis, $\hat{\beta}^{MC}_{OLS}$, was nearly identical to the average mse_s and variance of the OLS estimator using (11), $\hat{\beta}^{T}_{OLS}$, as they should be.

6. Comparison of OLS and GLS Estimators Using Analytic Expressions

Obtaining the mse_s of the OLS and GLS quantile estimators requires significantly less effort using the average over 100 replicates of the analytic expressions (equations (7), (9), and (11)), instead of Stedinger and Tasker's complete Monte Carlo simulation using randomly generated streamflows. Using analytic expressions, we examined how the mse_s of the OLS parameter estimator $\hat{\beta}^{T}_{OLS}$ compares to the GLS parameter estimator $\hat{\beta}^{E}_{GLS}$ for a number of different cases. The previous section demonstrated that $\hat{\beta}^{MC}_{GLS}$ performed almost as well as $\hat{\beta}^{E}_{GLS}$ in Stedinger and Tasker's Monte Carlo experiment 1.

In experiment 1 the variance of the residual terms in (14) and (15) were related by $\sigma^2_\delta = 0.0625\sigma^2_\upsilon$. This relationship determines the variability in the model error variance across sites, given by (32). In experiment 1 the maximum variation in the model error variance, γ^2_i, from the average model error variance, $\bar{\gamma}^2$, in the region was 25%. In this case the model error variance was relatively constant across sites, and thus $\hat{\beta}^{E}_{GLS}$, which employs a constant model error variance, performed well compared to $\hat{\beta}^{T}_{GLS}$. Of interest is the relative performance of $\hat{\beta}^{E}_{GLS}$ and $\hat{\beta}^{T}_{GLS}$ when the model error variance has greater variability across sites in a region, as well as the effect return period has on the estimators' performance.

The case where $\sigma^2_\delta = \sigma^2_\upsilon$ was examined, which yields a maximum variation of γ^2_i from $\bar{\gamma}^2$ of 93%. This corresponds to a realistic but perhaps extreme case wherein the relative precision with which the median flood flow and the log space standard deviation of the flood flows can be estimated by their respective models is roughly the same. The same values of $\bar{\gamma}^2$ and ρ examined in Stedinger and Tasker's experiment 1 were examined for a quantile estimator with a 50-year return period. Figure 1b contains the efficiency of $\hat{\beta}^{E}_{GLS}$ and $\hat{\beta}^{T}_{OLS}$ relative to $\hat{\beta}^{T}_{GLS}$ for this case. As σ^2_δ increases relative to σ^2_υ, the model error variance varies more across sites and we observe a drop in the relative efficiency of both $\hat{\beta}^{E}_{GLS}$ and $\hat{\beta}^{T}_{OLS}$. The decrease in efficiency of $\hat{\beta}^{E}_{GLS}$ is larger for cases with larger model error variances, because the heteroscedasticity along the diagonal terms of Λ due to variations in the model error variances is greater for these cases. For the cases presented in Figure 1b the efficiency of $\hat{\beta}^{E}_{GLS}$ is generally greater than 90%. This result indicates that the assumption of a constant model error variance in regions where the model error variance varies as much as 90% from the average model error variance only produces a 10% drop in the efficiency of $\hat{\beta}^{E}_{GLS}$ for quantile estimators with a 50-year return period. As the cross correlation increases, the efficiency of $\hat{\beta}^{E}_{GLS}$ increases, since $\hat{\beta}^{E}_{GLS}$ correctly describes the cross correlation between the quantile estimators. For these cases one would expect that the efficiency of $\hat{\beta}^{MC}_{GLS}$ to track that of $\hat{\beta}^{E}_{GLS}$ and be somewhere between the efficiency of $\hat{\beta}^{E}_{GLS}$ and the OLS estimator $\hat{\beta}^{T}_{OLS}$.

Also of interest is the effect of return period on the performance of $\hat{\beta}^{E}_{GLS}$ and $\hat{\beta}^{T}_{OLS}$. Figures 2a and 2b contain the efficiency of $\hat{\beta}^{E}_{GLS}$ and $\hat{\beta}^{T}_{OLS}$ relative to $\hat{\beta}^{T}_{GLS}$ for a quantile estimator with a 2-year return period when $\sigma^2_\delta = 0.0625\sigma^2_\upsilon$ and $\sigma^2_\delta = \sigma^2_\upsilon$, respectively. With a 2-year return period, $z_p = 0$ in (32), and thus the model error variance is constant across sites

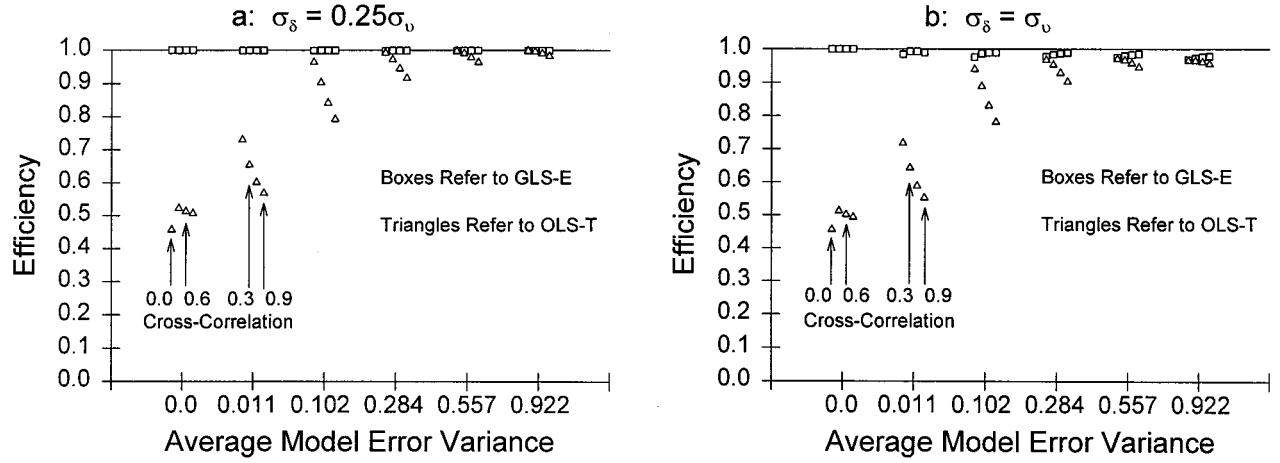

Figure 2. Efficiency of GLS and OLS estimators versus average model error variance for 2-year return period.

in a region regardless of the relationship between σ_δ^2 and σ_υ^2. In Figure 2b, where $\sigma_\delta^2 = \sigma_\upsilon^2$, we see a slight loss in efficiency in $\hat{\beta}^E_{\mathrm{GLS}}$ compared to Figure 2a, for which $\sigma_\delta^2 = 0.0625\sigma_\upsilon^2$. This loss is due to errors incurred when modeling the sampling covariance matrix in $\hat{\beta}^E_{\mathrm{GLS}}$. As σ_δ^2 increases relative to σ_υ^2, the variance of the at-site standard deviation increases, which increases the error in modeling the variance of the flows as $E[\sigma_i^2]$ by (28), as opposed to the true value σ_i^2 in (15). The relatively small loss in efficiency observed in Figure 2b indicates that modeling the variance of the flows as $E[\sigma_i^2]$ produces only minor loss of efficiency in the estimator $\hat{\beta}^E_{\mathrm{GLS}}$.

In Figures 2a and 2b we also observe that the loss in efficiency of the OLS estimator $\hat{\beta}^T_{\mathrm{OLS}}$ for a 2-year return period is much smaller than when the return period was 50-year (Figures 1a and 1b). This is most dramatic for larger model error variances and smaller cross correlations when the efficiency of $\hat{\beta}^T_{\mathrm{OLS}}$ is almost as large as the efficiency of $\hat{\beta}^E_{\mathrm{GLS}}$. This is because with a 2-year return period the model error variance is constant across sites, so for larger average model error variances the diagonal term of the true residual error covariance matrix is nearly homoscedastic. The effect of cross correlation on the efficiency of $\hat{\beta}^T_{\mathrm{OLS}}$ decreases as the average model error variance increases.

Figures 3a and 3b consider 100-year quantile estimators. For a 100-year return period $z_p = 2.326$, which produces more heteroscedasticity in the model error variance across the region than when the return period is 2 or 50 years. Because $\hat{\beta}^E_{\mathrm{GLS}}$ models the model error variance as constant across sites, we observe a greater loss in efficiency in $\hat{\beta}^E_{\mathrm{GLS}}$ as the return period increases and σ_δ^2 increases relative to σ_υ^2, as in Figure 3b, though the efficiency of $\hat{\beta}^E_{\mathrm{GLS}}$ is always greater than 80% for the cases examined. This result indicates that a GLS estimator which accounts for varying model error variance, as proposed by *Tasker and Stedinger* [1989], should be implemented when the return period of the quantiles of interest are 100 years or greater if high efficiency is desired, especially when moderate to large average model error variances are present. In Figure 3b we also observe an increased loss in efficiency of $\hat{\beta}^T_{\mathrm{OLS}}$ because of the increased heteroscedasticity.

7. Conclusions

Stedinger and Tasker [1985] used Monte Carlo simulation to compare the average sampling mean square error (mse_s) of ordinary least squares (OLS) and generalized least squares (GLS) quantile estimators in regional hydrologic regression

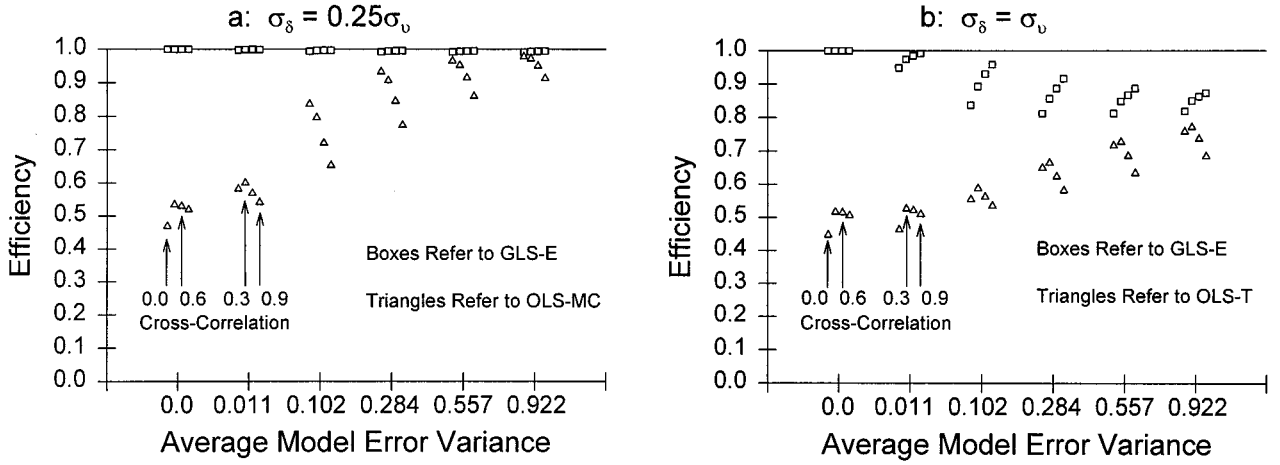

Figure 3. Efficiency of GLS and OLS estimators versus average model error variance for 100-year return period.

analyses. Stedinger and Tasker made a number of simplifying assumptions regarding the structure of the model error variance and the sampling error associated with at-site quantile estimators when constructing their residual error covariance matrix. They used smoothed estimators of the at-site variance of the flows, adopted an average regional cross correlation, and employed a single average generalized mean square model variance estimator in their ideal residual error covariance matrix, Λ_E. In practice, all of these parameters and statistics had to be estimated, so the residual error covariance matrix actually employed, Λ_{MC}, differs from Λ_E. Both of these matrices are different from the true residual error covariance matrix, Λ_T, associated with the underlying model in their Monte Carlo experiment.

The variance of the GLS parameter estimator implemented by Stedinger and Tasker in their Monte Carlo experiment, $\hat{\beta}^{MC}_{GLS}$, was compared to the variance of GLS parameter estimators based on the residual error covariance matrices Λ_E and Λ_T, denoted $\hat{\beta}^{E}_{GLS}$ and $\hat{\beta}^{T}_{GLS}$, respectively. Since the parameter estimator $\hat{\beta}^{T}_{GLS}$ is based on Λ_T, this estimator is unbiased and has minimum variance among all linear unbiased estimators. Using the average over 100 replicates of the new analytic expressions for the mse_s of $\hat{\beta}^{E}_{GLS}$ and $\hat{\beta}^{T}_{GLS}$, it was shown that the mse_s of $\hat{\beta}^{E}_{GLS}$ is almost indistinguishable from the mse_s of $\hat{\beta}^{T}_{GLS}$ for the cases considered by Stedinger and Tasker. For these cases the approximations employed to obtain a smoothed covariance matrix would result in almost no loss of efficiency. In addition, the parameter estimator implemented by Stedinger and Tasker, $\hat{\beta}^{MC}_{GLS}$, had an mse_s almost as small as $\hat{\beta}^{E}_{GLS}$. In most cases the difference was less than 10%. Thus for this case the difference in efficiency between $\hat{\beta}^{T}_{GLS}$, which is the best linear unbiased GLS estimator one could implement, and $\hat{\beta}^{MC}_{GLS}$, which represents the practical GLS estimator employed by Stedinger and Tasker, is relatively small.

The mse_s of the OLS estimator from *Stedinger and Tasker*'s [1985] Monte Carlo experiment, $\hat{\beta}^{MC}_{OLS}$, was also compared to the GLS estimators. For large model error variances, the efficiency of the OLS estimator is close to the efficiency of $\hat{\beta}^{MC}_{GLS}$, but for a small model error variance the efficiency of the OLS estimator drops considerably. For moderate model error variances, the relative efficiency of the OLS estimator decreases as the cross correlation increases.

Using analytic expressions for the mse_s, the performance of $\hat{\beta}^{T}_{OLS}$ and $\hat{\beta}^{E}_{GLS}$ relative to $\hat{\beta}^{T}_{GLS}$ were compared for a number of cases not considered by *Stedinger and Tasker* [1985]. In particular, a model with a more heteroscedastic model error variance was considered for return periods of 2, 50, and 100 years. For models where the model error variance varied considerably across sites, some loss in efficiency in $\hat{\beta}^{E}_{GLS}$ was observed (up to 20%). It was shown that the loss in efficiency of $\hat{\beta}^{E}_{GLS}$ resulted from using an average model error variance and not from smoothing the sampling covariance matrix. The loss of efficiency was particularly apparent for large return periods and moderate to large average model error variances. In this case a GLS estimator which accounts for varying model error variance such as that developed by *Tasker and Stedinger* [1989] should be implemented. They used a three-parameter error model to account for correlation between the log space means and standard deviations of the flood flows.

Overall, for a small return period and moderate to large model error variance, the OLS estimator $\hat{\beta}^{T}_{OLS}$ performed nearly as well as $\hat{\beta}^{E}_{GLS}$, especially when the cross correlation of the flows was small. In this case an OLS estimator, which is much easier to implement than a GLS estimator, could be implemented with little or no loss in efficiency. One should note that the efficiency of the β estimators is only one of the advantages of GLS procedures: *Stedinger and Tasker* [1985] observe that GLS estimators also provide more accurate estimators of model error variances and the precision of estimated parameters than do OLS analyses.

Acknowledgment. This material is based upon work partially supported by the Cooperative State Research Service, USDA, under project 97CRMS06102.

References

Benson, M. A., Factors influencing the occurrence of floods in a humid region of diverse terrain, *U.S. Geol. Surv. Water Supply Pap. 1580-B*, 61 pp., 1962.

David, H. A., *Order Statistics*, John Wiley, New York, 1981.

Draper, N. R., and Smith, H., *Applied Regression Analysis*, John Wiley, New York, 1981.

Greene, W., *Econometric Analysis*, Macmillan, New York, 1990.

Jennings, M. E., W. O. Thomas, and H. C. Riggs, Nationwide summary of U.S. geological survey regional regression equations for estimating magnitude and frequency of floods for ungaged sites, 1993, *U.S. Geol. Surv. Water Resour. Invest. Rep*. *94-4002*, 1994.

Johnston, J., *Econometric Methods*, McGraw-Hill, New York, 1984.

Kroll, C. N., Censored data analyses in water resources, Ph.D. thesis, School of Civ. and Environ. Eng., Cornell Univ., Ithaca, N. Y., 1996.

Loucks, D. P., J. R. Stedinger, and D. A. Haith, *Water Resource Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, N. J., 1981.

Matalas, N. C., and E. J. Gilroy, Some comments on regionalization in hydrologic studies, *Water Resour. Res*., *4*(6), 1361–1369, 1968.

Moss, M. E., and G. D. Tasker, An intercomparison of hydrological network-design technologies, *J. Hydrol. Sci*., *36*(3), 209–221, 1991.

Riggs, H. C., *Low-Flow Investigations: U.S. Geological Survey Techniques of Water Resource Investigations*, book 4, chap. B1, U.S. Geol. Surv., Denver, Colo., 1972.

Stedinger, J. R., Estimating a regional flood frequency distribution, *Water Resour. Res*., *19*(2), 503–510, 1983.

Stedinger, J. R., and G. D. Tasker, Regional hydrologic analysis, 1, Ordinary, weighted, and generalized least squares compared, *Water Resour. Res*., *21*(9), 1421–1432, 1985. (Correction, *Water Resour. Res*., *22*(5), 844, 1996.)

Stedinger, J. R., and G. D. Tasker, Regional hydrologic analysis, 2, Model-error estimators, estimation of sigma and log-Pearson type 3 distribution, *Water Resour. Res*., *22*(10), 1487–99, 1986.

Tasker, G. D., Hydrologic regression and weighted least squares, *Water Resour. Res*., *16*(6), 1107–1113, 1980.

Tasker, G. D., and J. R. Stedinger, An operational GLS model for hydrologic regression, *J. Hydrol*., *111*, 361–375, 1989.

Tasker, G. D., S. A. Hodge, and C. S. Barks, Region of influence regression for estimating the 50-year flood at ungaged sites, *Water Resour. Bull*., *32*(1), 163–170, 1996.

Thomas, D. M., and M. A. Benson, Generalization of streamflow characteristics from drainage-basin characteristics, *U.S. Geol. Surv. Water Supply Pap. 1975*, 1970.

Thomas, M. P., and M. A. Cervione, A proposed streamflow data program for Connecticut, *Conn. Water Resour. Bull*., *23*, 1970.

Vogel, R. M., and C. N. Kroll, Generalized low-flow frequency relationships for ungaged sites in Massachusetts, *Water Resour. Bull*., *26*(2), 241–253, 1990.

C. N. Kroll, Environmental Resources and Forest Engineering, College of Environmental Science and Forestry, State University of New York, Syracuse, NY 13210-2778. (e-mail: cnkroll@mailbox.syr.edu)

J. R. Stedinger, School of Civil and Engineering, Cornell University, Ithaca, NY 14850-3501.

(Received August 20, 1996; accepted September 22, 1997.)