1.5.7 Non-nested tests (advanced)

It is commonly the case that there are alternative ways of specifying a SPARROW model. For example, suppose that two alternative SPARROW models, of identical functional form, use different data sets to characterize a given source variable; perhaps one model uses fertilizer sales to characterize the fertilizer nitrogen source, whereas the other model uses an estimate that corrects the sales information for exports associated with the sale of agricultural products. Another example would be the case in which there are two potential specifications for the in-stream decay process: one process uses the discrete streamflow classes approach whereas the other process uses the continuous function of streamflow approach (see section 1.4.4). The question arises: which is the better specification?

The nature of the question just raised poses a challenge for hypothesis testing. The classical hypothesis test investigates the effect on model fit caused by a restriction on the model; that is, the model conforming to the null hypothesis is nested within a more general model specification. As such, the hypothesis is tested by determining the effect on model fit from a restriction imposed on the model. The null hypothesis of a non-nested test, conversely, cannot be structured in this way; first the model is restricted, to negate the alternative specification, and then the model is appended to include the specification corresponding to the null hypothesis. Indeed, the entire classification of specifications into nulls and alternatives is somewhat arbitrary. Clearly, a different testing framework is needed to address this question.

The following describes two approaches to testing non-nested hypotheses. The two approaches, called the *J*-test and the *P*-test, are similar in concept but differ in form: the *J*-test is evaluated using a generalization of the two models that requires estimation by nonlinear methods; the *P*-test evaluates a generalization of two models based on a linear analysis (see Davidson and MacKinnon, 1993). Superficially, the *P*-test would appear to be the easier test to implement, but this is not necessarily the case. Neither test is implemented directly in SPARROW; however, as explained below, output from the SPARROW model can be used to implement either test as a follow-on exercise to a SPARROW analysis.

To describe the two tests, suppose there are two competing specifications of a model explaining some dependent variable *y*. Both specifications are assumed to be nonlinear with respect to their parameters, although they may also be linear. [In fact, if the alternatives are both linear models, the *J*-test and the *P*-test are equivalent.] Let the alternative specifications be defined as

$$y_i = x_i(\boldsymbol{\beta}) + e_i \tag{1.113}$$

$$y_i = z_i(\boldsymbol{\gamma}) + u_i, \tag{1.114}$$

where y_i is the dependent variable for observation *i*, $x_i(\boldsymbol{\beta})$ is the first model, which depends on the parameter vector $\boldsymbol{\beta}$, and $z_i(\boldsymbol{\gamma})$ is the second model, which depends on the parameter vector $\boldsymbol{\gamma}$. The error terms e_i and u_i, having respective variances σ_e^2 and σ_u^2, are each assumed to be identically and independently distributed across observations, but may be correlated to each other for the same observation.

The *J*-test is implemented by first estimating equation (1.114) using nonlinear least squares. The coefficients obtained from that estimation are used to form the prediction $\hat{z}_i = z_i(\hat{\boldsymbol{\gamma}})$. The estimated $\hat{z}_i$ are then inserted into the augmented model

$$y_i = (1-\alpha)x_i(\boldsymbol{\beta}) + \alpha\hat{z}_i + e_i', \tag{1.115}$$

which is estimated for the unknown parameters α and $\boldsymbol{\beta}$ using nonlinear least squares. The null hypothesis under this specification is that equation (1.113) is the correct model. This hypothesis is evaluated by testing the ancillary hypothesis that $\alpha = 0$ using a simple *t*-test. The procedure is then reversed and the predictions from estimating equation (1.113), $\hat{x}_i = x_i(\hat{\boldsymbol{\beta}})$, are entered into the model

(1.116) $$y_i = (1-\theta) z_i(\boldsymbol{\gamma}) + \theta \hat{x}_i + u_i',$$

to obtain nonlinear least squares estimates of θ and $\boldsymbol{\gamma}$, with subsequent testing of the null hypothesis that $\theta = 0$.

There are four possible outcomes for this approach. If the hypothesis $\alpha = 0$ is rejected and the hypothesis $\theta = 0$ is not, we then conclude that the correct model is given by equation (1.114). If the hypothesis $\theta = 0$ is rejected and the hypothesis $\alpha = 0$ is not, then the conclusion is that the model given by equation (1.113) is correct. If both hypotheses are rejected, then we say that both models fit the data well and neither model shows the other to be misspecified. If we fail to reject both hypotheses, then the conclusion is that neither model is satisfactory.

The *P*-test is implemented by first estimating equations (1.113) and (1.114) separately using nonlinear least squares and forming the predictions $\hat{x}_i = x_i\left(\hat{\boldsymbol{\beta}}\right)$ and $\hat{z}_i = z_i(\hat{\boldsymbol{\gamma}})$. Additionally, output from the SPARROW model will include the gradients from each model $x_{i\boldsymbol{\beta}'}\left(\hat{\boldsymbol{\beta}}\right)$ and $z_{i\boldsymbol{\gamma}'}(\hat{\boldsymbol{\gamma}})$. To test the null hypothesis that equation (1.113) is correct, ordinary least squares is used to estimate the model

(1.117) $$y_i - \hat{x}_i = x_{i\boldsymbol{\beta}'}\left(\hat{\boldsymbol{\beta}}\right) b + \alpha(\hat{z}_i - \hat{x}_i) + e_i'',$$

and the hypothesis that $\alpha = 0$ is tested using an exact *t*-test. To test the null hypothesis that equation (1.114) is correct, use ordinary least squares to estimate the complementary model

(1.118) $$y_i - \hat{z}_i = z_{i\boldsymbol{\gamma}'}(\hat{\boldsymbol{\gamma}}) c + \theta(\hat{x}_i - \hat{z}_i) + u_i'',$$

and test the hypothesis that $\theta = 0$, again using an exact *t*-test. As with the J-test, there are four possible outcomes. The model in equation (1.113) is accepted in favor of equation (1.114) if we reject $\theta = 0$ and fail to reject $\alpha = 0$. The model in equation (1.114) is accepted in favor of equation (1.113) if we reject $\alpha = 0$ and fail to reject $\theta = 0$. Both models are acceptable if both hypotheses are rejected, and if we fail to reject either hypothesis then neither model is satisfactory.

1.6 Model Predictions

SPARROW output contains prediction results paired with measures of accuracy. A number of technical issues can arise in the derivation of these statistics—most of these are caused by the nonlinear nature of the SPARROW model. The following sections provide a detailed description of the methods used to obtain prediction results using SPARROW.

1.6.1 Prediction equation

The prediction equation is similar to the calibration equation (1.27) given in section 1.4.1. Because predictions are generated for specific sources, however, it is necessary to decompose flux into source-specific components. Let $F_{i,n}^{*}$ denote the reach *i* model-estimated flux associated with source *n*, and let $F_{j,n}'$ be the source-*n* flux from upstream reach *j*. If reach *j* is monitored (that is, $j \in I$, where *I* is the set of monitored reaches), and predictions are conditioned on measured flux, then $F_{j,n}'$ is apportioned from the measured flux F_j^M according to

$$F'_{j,n} = \frac{F^*_{j,n}}{\sum_{m=1}^{N_y} F^*_{j,m}} F^M_j \,. \tag{1.119}$$

Otherwise, if reach *j* is not a monitored reach (that is, $j \notin I$) or predictions are not conditioned on measurements, then $F'_{j,n}$ is set equal to $F^*_{j,n}$.

The equation defining the model-estimated flux for source *n*, $F^*_{i,n}$, is given by

$$F^*_{i,n} = \left(\sum_{j \in J(i)} F'_{j,n} \right) \delta_i A\left(\mathbf{Z}^S_i, \mathbf{Z}^R_i; \boldsymbol{\theta}_S, \boldsymbol{\theta}_R\right) + S_{n,i} \alpha_n D_n \left(\mathbf{Z}^D_i; \boldsymbol{\theta}_D\right) A'\left(\mathbf{Z}^S_i, \mathbf{Z}^R_i; \boldsymbol{\theta}_S, \boldsymbol{\theta}_R\right). \tag{1.120}$$

The actual flux leaving reach *i* attributed to source *n*, denoted $F_{i,n}$, is based on the modeled flux $F^*_{i,n}$ times a multiplicative error term. The predicted flux leaving reach *i*, denoted $\overline{F}_{i,n}$ is based either on the modeled flux $F^*_{i,n}$ times the expectation of the exponentiated error or, if the prediction is conditioned on measurements, on the apportioned measured flux F^M_i using the apportioning method described in equation (1.119) (see section 1.6.6 for additional discussion concerning the conditioning of predictions on flux measurements).

An assumption of model estimation is that the model error, e_i, for total flux is additive in logarithm space and independent across monitored nested basins $i \in I$ [recall from section 1.4.1 that a nested basin comprises all reaches upstream from monitored reach *i*, exclusive of reaches at or above all upstream monitoring stations]. In deriving predictions, particularly for disaggregated components of monitored flux, such as flux for non-monitored reaches or flux by source, it is necessary to add structure to the error properties of the model. The additional structure is needed to provide a basis for estimating prediction uncertainty. There is only one way this can be done while retaining the simple error structure assumed for estimation: to assume that every additive term in equation (1.120), for every reach i' within the nested basin of monitored reach *i*, is multiplied by the same multiplicative error term $\exp(e_i)$ associated with the monitored flux at reach *i*. Monitored reaches above the nested basin of reach *i* express a separate, independent, multiplicative error. If it is necessary to attribute a source to this error, the only model factor that appears with every additive term in equation (1.120) is the stream attenuation factor $A(\cdot)$ (or $A'(\cdot)$). Thus, the implicit assumption is that all model error is associated with unobserved in-stream attenuation processes.

The error structure just described is certainly unrealistic. It implies the error covariance between reaches is determined arbitrarily by the location of monitoring stations: reaches within a nested basin are perfectly correlated; reaches on either side of a monitoring station are perfectly uncorrelated. It also implies a perfect correlation for errors associated with flux distributed to the individual topical sources (e.g., fertilizer, point sources, and atmospheric deposition). The assumed error structure requires estimates of incremental flux to be multiplied by the same error factor as total, accumulated flux. This can be rationalized by the assumption that error is scale invariant. Indeed, a number of SPARROW applications in the literature exhibit error that is approximately scale invariant (that is, the error has constant variance in logarithm space over the full range of observations; see Preston and Brakebill, 1999 and Moore and others, 2004). The national applications of SPARROW, however, show residuals that are not scale invariant (see, for example, Smith and others, 1997). Despite these shortcomings, the assumed error structure is the best we can do within the current SPARROW modeling framework. Future refinements to the model are planned that will allow for a much more general specification of error.

Because the error terms in the SPARROW model are permitted to be heteroscedastic, it would make sense to assign the variance of the logarithm-transformed error term for nested basin *i* (that is, the error associated with the monitoring station *i*) to equal the variance that was used for purposes of weighting the

observation for model estimation. As shown in section 1.3.1, this variance has relevance for the expectation of the error term transformed back into real space. However, this is not what is done. For reasons that will be made clear later in the next section, a robust estimator of the expectation of the multiplicative error term, that is, the retransformation factor, is one that treats errors as interchangeable, implying they have a common distribution. In that case, the appropriate error term to apply in prediction is one that has a variance equal to the variance of the weighted residual, given by $\sqrt{w_i}e_i$.

Given these considerations, the predicted flux attributed to source *n* leaving reach *i* is

$$\bar{F}_{i,n} = \begin{cases} \dfrac{F^*_{i,n}}{\sum_{m=1}^{N_S} F^*_{i,m}} F_i^M, & i \in I \\ F^*_{i,n}\overline{e^{\sqrt{w}\varepsilon}}, & i \notin I, \end{cases} \tag{1.121}$$

where the retransformation factor $\overline{e^{\sqrt{w}\varepsilon}}$ is the expectation of the exponentiated weighted residual.

1.6.2 Parametric predictions

The discussion of prediction in this and remaining sections adopts a simplified notation—one that focuses on the role of coefficients and errors. To this end, we simplify the notation that associates a prediction to a specific source *n*, and simply note that the expressions given below pertain to the prediction of any reach-level flux quantity, including total flux, flux from the incremental watershed, accumulated flux by source, and incremental watershed flux by source. In section 1.6.6 we include additional comments regarding the special considerations required to make predictions of non-flux quantities.

The relation between actual flux expressed in real space, F_i, modeled flux, F_i^*, and error for reach *i* is given by

$$F_i = F_i^*(\boldsymbol{\beta})e^{\varepsilon_i}, \tag{1.122}$$

where F_i^* is defined by equation (1.120) in section 1.6.1, $\boldsymbol{\beta} = \left[\boldsymbol{\alpha}', \boldsymbol{\theta}_D', \boldsymbol{\theta}_S', \boldsymbol{\theta}_R'\right]'$ is a collection of all model parameters, and $\varepsilon_i = e_i\sqrt{w_i}$ is the weighted error term. In accordance with the efficient notation, we have expressed F_i^* as a function of its parameters only, suppressing the dependence of F_i^* on upstream monitored fluxes, F^M. The expectation of F_i conditioned on model information (and upstream monitored fluxes), denoted $\bar{F}_i$, is given by

$$\bar{F}_i = F_i^*(\boldsymbol{\beta})\overline{e^{\varepsilon_i}}, \tag{1.123}$$

where $\overline{e^{\varepsilon_i}}$ is the expectation of the exponential of the error term. Because the exponential function is concave, by Jensen's inequality, $\overline{e^{\varepsilon_i}} > e^{\bar{\varepsilon}_i} = 1$. As explained in section 1.6.1, the SPARROW model used for prediction assumes all errors have the same distribution, implying $\overline{e^{\varepsilon_i}}$ can be replaced by $\overline{e^{\varepsilon}}$, which is not reach specific. This value is referred to as the "retransformation" factor.

The evaluation of equation (1.123) requires estimates of $\boldsymbol{\beta}$ and $\overline{e^{\varepsilon}}$. The estimation of the SPARROW model provides estimates of both the coefficient vector, $\hat{\boldsymbol{\beta}}$, and residual terms, $\hat{\varepsilon}_j$, $j = 1,\ldots,N$, where *N* is the number of monitored reaches. If the underlying residuals were normally distributed, the retransformation factor

would be $e^{\sigma_\varepsilon^2/2}$; if the estimated residuals were also derived using ordinary least squares, the expectation of an exponentiated estimated residual would be $E\left(\exp\left(\hat{\varepsilon}_j\right)\right) = \exp\left(\sigma_\varepsilon^2\left(1-h_j\right)/2\right)$, where $h_j = \mathbf{X}_j\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}_j$ is the residual's leverage statistic. A more robust estimate of $\overline{e^{\varepsilon}}$, one that does not depend on the residuals being normally distributed, is made by taking a simple average of the exponentiated, weighted estimated residuals, where the assigned weight for residual *j* is $1/\sqrt{1-h_j}$ (see equation (1.62) in section 1.5.1.3 for the determinants of the nonlinear analog to the leverage statistic), resulting in

(1.124) $$\overset{\wedge}{\overline{e^{\varepsilon}}} = \frac{1}{N}\sum_{j=1}^{N} e^{\hat{\varepsilon}_j/\sqrt{1-h_j}}.$$

Such an estimate, called a Smearing estimator (Duan, 1983), is unbiased in small samples under the assumption of normality, and consistent (that is, unbiased in large samples) regardless of the underlying error distribution. For a nonlinear least-squares analysis such as SPARROW, the values of the $\mathbf{X}_j$ vector and $\mathbf{X}$ matrix in the observation-*j* leverage statistic are the gradients evaluated at the optimal parameter estimates (see section 1.5.1 above). In this case, because the coefficients are unknown, the Smearing estimator is not unbiased in small samples, even if the residuals are normally distributed. However, the estimator remains consistent in large samples and the effect of weighting generally reduces bias in small samples.

The values of the estimated parameters $\hat{\boldsymbol{\beta}}$ and the Smearing estimator $\overset{\wedge}{\overline{e^{\varepsilon}}}$ are substituted for the unknown quantities $\boldsymbol{\beta}$ and $\overline{e^{\varepsilon}}$ in equation (1.123) to obtain what is called a *parametric* estimate of F_i

(1.125) $$\hat{F}_i = F_i^*\left(\hat{\boldsymbol{\beta}}\right)\overset{\wedge}{\overline{e^{\varepsilon}}}.$$

Because of the consistency of $\hat{\boldsymbol{\beta}}$ and $\overset{\wedge}{\overline{e^{\varepsilon}}}$ (see section 1.5.2), the parametric estimate $\hat{F}_i$ is also consistent. In small samples, however, due principally to the nonlinearity of F_i^*, the expectation of $\hat{F}_i$ is not equal to $\overline{F}_i$. Moreover, the complex nature of the nonlinearity of F_i^* also implies there is no analytic function of $\hat{\boldsymbol{\beta}}$ that is unbiased.

1.6.3 Bias-corrected predictions based on bootstrapping (advanced)

A widely used approach to bias correction is the bootstrap method. A typical bias-correction bootstrap analysis assumes bias is linear, so that bias is simply the difference between the expected value of the estimator and the true mean. This characterization of bias is inappropriate, however, if the desired estimate must always be positive, as it must for contaminant flux. To show this, consider the case in which you wish to estimate e^{μ} and you have an unbiased estimate of μ, $\overline{x}$, that is distributed normal with mean μ and variance σ^2. As explained in section 1.3.1.2, the naïve estimator $e^{\overline{x}}$ is biased with bias given by

(1.126) $$B = E\left(e^{\overline{x}}\right) - e^{\mu} = e^{\mu}\left(e^{\sigma^2/2} - 1\right).$$

Consider the correction of this bias using the linear parametric bootstrap procedure (see section 1.5.3.1), in which *R* normally distributed random variables $w^{(r)}$, with mean zero and variance σ^2, are generated to evaluate

$$B^R = R^{-1}\sum_{r=1}^{R}\exp\left(\overline{x} + w^{(r)}\right) - \exp\left(\overline{x}\right). \tag{1.127}$$

The expectation of (1.127) is

$$E\left(B^R\right) = e^{\mu+\sigma^2/2}\left(e^{\sigma^2/2} - 1\right) = e^{\sigma^2/2}B. \tag{1.128}$$

So the estimate of bias is itself biased and the "bias-corrected" estimate $e^{\overline{x}} - B^R$ is not an unbiased estimate of e^{μ}, even for large *R*. In fact, the linear bootstrap bias correction procedure results in a bias-corrected estimate of e^{μ} that has expectation

$$E\left(e^{\overline{x}} - B^R\right) = e^{\mu+\sigma^2/2}\left(2 - e^{\sigma^2/2}\right), \tag{1.129}$$

which can be negative for sufficiently large values of σ^2.

An alternative approach is to express the bias as a ratio

$$b = \frac{E\left(e^{\overline{x}}\right)}{e^{\mu}}, \tag{1.130}$$

with an associated bootstrap estimate given by

$$b^R = \frac{R^{-1}\sum_{r=1}^{R}\exp\left(\overline{x} + w^{(r)}\right)}{\exp\left(\overline{x}\right)}. \tag{1.131}$$

The bias-corrected estimate of e^{μ} becomes

$$\left(e^{\mu}\right)^R = \frac{e^{\overline{x}}}{b^R} = \frac{\exp\left(2\overline{x}\right)}{R^{-1}\sum_{r=1}^{R}\exp\left(\overline{x} + w^{(r)}\right)} = \frac{e^{\overline{x}}}{R^{-1}\sum_{r=1}^{R}\exp\left(w^{(r)}\right)}. \tag{1.132}$$

For large *R*, the denominator term goes to $\exp\left(\sigma^2/2\right)$, implying the limit of the expectation of equation (1.132) is

$$\lim_{R\to\infty} E\left(\left(e^{\mu}\right)^R\right) = e^{\mu}. \tag{1.133}$$

In this case, therefore, the bootstrap ratio bias correction is unbiased.

It is suspected that the ratio bias correction will work best in those cases, as above, in which the empirical distribution of the predictions in real space is skewed. A skewed distribution makes it likely that the average of the bootstrap predictions will have large variance, implying it is likely that the bootstrap average will exceed the parametric prediction. For applications in which the predictions from each bootstrap iteration are restricted to be positive, as occurs in cases where the parametric prediction for a strictly positive quantity—such as mass—is forced to be admissible, a skewed empirical distribution is virtually guaranteed. The high variability in the bootstrap average prediction in these applications makes it likely that an additive bias correction, which involves taking the negative of the average bootstrap prediction, will result in a negative estimate—a result that is inadmissible. The ratio form of bias correction represents a sort of weighting in which an unusually large

average bootstrap prediction has a proportionately smaller effect on modifying the parametric estimate, explaining its superior performance in the above example.

All predictions generated by SPARROW are restricted to be positive values, implying the ratio form of bootstrap bias correction should perform better than the additive form. Accordingly, the proportional bias in the parametric estimate of F_i is given by

$$b_i = \frac{E\left(\hat{F}_i\right)}{\bar{F}_i}. \tag{1.134}$$

To compute the proportional bias coefficient, one needs the expectation of $\hat{F}_i$, which is a complex function of the sample estimates $\hat{\boldsymbol{\beta}}$ and $\overline{\hat{e}^{\varepsilon}}$ the joint distribution of which is unknown. The bootstrap method provides a paradigm for making the necessary inference. The basic idea is that Monte Carlo methods can be used to define an empirical joint distribution for $\hat{\boldsymbol{\beta}}$ and $\overline{\hat{e}^{\varepsilon}}$, which bears a relation to the estimated parameter vector $\hat{\boldsymbol{\beta}}$ and retransformation factor $\overline{\hat{e}^{\varepsilon}}$ that is analogous to the relation between the distribution of the estimates $\hat{\boldsymbol{\beta}}$ and $\overline{\hat{e}^{\varepsilon}}$ and the true values $\boldsymbol{\beta}$ and $\overline{e^{\varepsilon}}$. Thus, under this paradigm, the estimation of the bias coefficient can be achieved by using the empirical distribution for $\hat{\boldsymbol{\beta}}$ and $\overline{\hat{e}^{\varepsilon}}$ to evaluate the expectation in the numerator, and using the parametric estimate $\hat{F}_i$ to evaluate the denominator.

SPARROW implements two bootstrap approaches for generating the empirical distribution. The simplest and fastest approach, known as a parametric bootstrap, uses a random number generator to create alternative realizations of the $\hat{\boldsymbol{\beta}}$ vector. Under the assumptions of nonlinear least squares (see section 1.5.2), the coefficient estimates are asymptotically normally distributed. Therefore, the parametric bootstrap generates *R* replications of the $\hat{\boldsymbol{\beta}}$ vector, denoted by $\hat{\boldsymbol{\beta}}^{(r)}$, $r = 1, \ldots, R$, from a multivariate normal distribution with mean and covariance equal to the nonlinear least squares estimates. Each set of randomly generated coefficients is used to compute a corresponding set of *N* residuals from which a retransformation factor $\overline{\hat{e}^{\varepsilon}}^{(r)}$ is derived using equation (1.124).

A more robust bootstrap approach uses random sampling with replacement to generate the empirical distribution. Under this approach, random positive integer weights, summing to the number of reaches with monitored fluxes, are applied to the observations (see section 1.5.3.1 for a description of resampling). The weighted observations are used in the context of nonlinear weighted least squares to obtain bootstrap iteration *r* estimates of the coefficients, $\hat{\boldsymbol{\beta}}^{(r)}$, and associated retransformation factors, $\overline{\hat{e}^{\varepsilon}}^{(r)}$. Repeating this process *R* times generates *R* realizations of the $\hat{\boldsymbol{\beta}}^{(r)}$ coefficient vector and retransformation factor $\overline{\hat{e}^{\varepsilon}}^{(r)}$, which serves as the basis of the requisite empirical distribution.

The *R* realizations of $\hat{\boldsymbol{\beta}}^{(r)}$ and $\overline{\hat{e}^{\varepsilon}}^{(r)}$ can be used to generate *R* realizations of $\hat{F}_i$. Define $\hat{F}_i^{(r)}$ to be the r^{th} bootstrap replication of $\hat{F}_i$, where

$$\hat{F}_i^{(r)} = F_i^*\left(\hat{\boldsymbol{\beta}}^{(r)}\right)\overline{\hat{e}^{\varepsilon}}^{(r)}, \tag{1.135}$$

and let the average bootstrap estimate be

$$\hat{F}_i^R = \frac{1}{R}\sum_{r=1}^{R} \hat{F}_i^{(r)} . \quad (1.136)$$

Using the bootstrap paradigm analogy, the bootstrap estimate of proportional bias is

$$\hat{b}_i^R = \frac{\hat{F}_i^R}{\hat{F}_i} . \quad (1.137)$$

Therefore, the bootstrap proportional bias-corrected estimate of F_i, denoted $\hat{F}_i^c$, is

$$\hat{F}_i^c = \frac{\hat{F}_i}{\hat{b}_i^R} = \frac{\hat{F}_i^2}{\hat{F}_i^R} . \quad (1.138)$$

Because all terms in this equation are positive, this estimate must also be positive.

An important point to note is that although each bootstrap iteration estimate of $\hat{F}_i^{(r)}$ retains mass balance, the bias-corrected estimates do not. That is, the $\hat{F}_i^{(r)}$ have source shares that sum to one, and the sum of flux entering a confluence exactly equals the flux leaving the confluence. The effect of bias correction causes a breakdown in these restrictions because each estimate is corrected individually, without regard for global restrictions on mass balance. The absence of strict mass balance in the bias corrected predictions does not mean that SPARROW is not a mass balance model. In the limit, as sample size becomes large, the bias correction goes to zero and mass balance is retained; moreover, strict mass balance remains a property of the parametric estimates (see section 1.6.2). In order to obtain better individual predictions, however, it is necessary to loosen restrictions that pertain to the comparison of predictions. There are alternative approaches that retain mass balance but they come at a cost. One approach is to use linear bias correction, the same correction applied to the coefficient estimates (see above and section 1.5.3.1). Linear bias correction retains all the restrictions of mass balance; however, it also has the capacity to produce inadmissable negative flux estimates. An alternative estimation approach called Bayesian analysis could also resolve the mass balance problem. Bayesian analysis derives a posterior distribution for predicted quantities based on the posterior distribution of the coefficients. Because no bias correction of these predictions is necessary, the Bayesian predictions perfectly retain the mass balance restrictions. The estimation of Bayesian models, however, requires a significantly different estimation methodology that requires the evaluation of highly dimensioned integrals. The numerical accuracy of estimating a national SPARROW model in this way has yet to be evaluated (however, see the recent paper by Qian and others, 2005, for an example of a regional scale SPARROW model estimated using Bayesian methods). Given these tradeoffs, it was decided to sacrifice small sample strict mass balance in favor of estimates that are individually admissible.

1.6.4 Prediction standard errors based on bootstrapping (advanced)

The variance of the prediction error can be divided into two parts: the sample variance arising from sample error in the coefficient estimates, and the model variance arising from model error. Sample variance is a consequence of using a finite sample to estimate the model and will go to zero as the sample size becomes large. Model variance is fixed by the number of relevant predictors included in the model and can be lowered only by extending the model to include additional explanatory variables. Because the prediction formula is a nonlinear function of the parameters, bootstrap methods are required to estimate sample variance. Model variance is estimated with a robust method that does not require an assumed probability distribution.

Using the abstract model described in sections 1.6.2 and 1.6.3, prediction error can be expressed as

$$u_i = F_i - \hat{F}_i^c = \left(F_i - \overline{F}_i\right) - \left(\hat{F}_i^c - \overline{F}_i\right). \tag{1.139}$$

Prediction error is decomposed into a term depending solely on model error and a second term that depends only on sampling error. Because the model error for each reach is assumed to be independent of the model errors for all other reaches, the model error term and the sampling error terms are also independent. Because $\hat{F}_i^c$ is an unbiased predictor of F_i (at least at first order), the expectation of prediction error is zero.

The variance of prediction error is given by the sum of the variance caused by model error and the variance caused by sampling error. The variance caused by model error is given by

$$E\left(\left(F_i - \overline{F}_i\right)^2\right) = E\left(\left(F_i^*(\boldsymbol{\beta})\left(e^{\varepsilon_i} - \overline{e^{\varepsilon_i}}\right)\right)^2\right) = F_i^*(\boldsymbol{\beta})^2 V\left(e^{\varepsilon_i}\right). \tag{1.140}$$

A consistent estimator of this variance is given by

$$\hat{E}\left(\left(F_i - \overline{F}_i\right)^2\right) = \hat{F}_i^{c2} \hat{V}\left(e^{\hat{\varepsilon}/\sqrt{1-h}}\right) \Big/ \overline{\hat{e^{\varepsilon}}}^{\,2}, \tag{1.141}$$

where $\hat{F}_i^c$ is given by equation (1.138), $\overline{\hat{e^{\varepsilon}}}$ is given by equation (1.124), and

$$\hat{V}\left(e^{\hat{\varepsilon}/\sqrt{1-h}}\right) = N^{-1} \sum_{i=1}^{N} \left(e^{\hat{\varepsilon}_i/\sqrt{1-h_i}} - \overline{\hat{e^{\varepsilon}}}\right)^2. \tag{1.142}$$

The variance of the sampling error is predicated on the assumption that the bootstrap bias ratio $\hat{b}_i^R$ is a constant. An attempt to estimate a distribution for $\hat{b}_i^R$ would require a double bootstrap, a method that is computationally intractable given the complexity of the SPARROW model. Using bootstrap methods, the sampling error variance is

$$\hat{V}^R\left(\hat{F}_i^c\right) = \hat{V}\left(\hat{F}_i^{(r)}\right) \Big/ \hat{b}_i^{R2}, \tag{1.143}$$

where $\hat{V}\left(\hat{F}_i^{(r)}\right) = R^{-1} \sum_{r=1}^{R} \left(\hat{F}_i^{(r)} - \hat{F}_i^R\right)^2$. Equation (1.143) is derived by noting that, with $\hat{b}_i^R$ a constant, the variance of $\hat{F}_i^c$ equals $V\left(\hat{F}_i\right) \Big/ \hat{b}_i^{R\,2}$, and the bootstrap paradigm is used to justify substituting $\hat{V}\left(\hat{F}_i^{(r)}\right)$ for the unknown variance $V\left(\hat{F}_i\right)$. Combining the variances due to model and sampling error in equations (1.141) and (1.143), and taking note that $\hat{b}_i^R = \hat{F}_i \big/ \hat{F}_i^c$ (equation (1.138)), the standard error of the prediction $\hat{F}_i^c$ is given by

$$SE\left(\hat{F}_i^c\right) = \hat{F}_i^c \sqrt{\frac{\hat{V}\left(e^{\hat{\varepsilon}\sqrt{1-h}}\right)}{\overline{\hat{e^{\varepsilon}}}^{\,2}} + \frac{\hat{V}\left(\hat{F}_i^{(r)}\right)}{\hat{F}_i^2}}. \tag{1.144}$$

1.6.5 Prediction intervals based on bootstrapping (advanced)

A prediction interval represents a stochastic interval that has a probability, P_c, called the coverage probability or confidence level, of including some unknown value of predictive interest. The derivation of this interval must account for two sources of uncertainty, the sampling error of the coefficients and the model error associated with the inability of the model to fully account for all factors determining the true value of the predicted variable. Because of the complex nonlinear relation between predictions and underlying coefficient estimates in SPARROW, it is not possible to base intervals on known parametric distributions—such as the normal distribution. The bootstrap method represents an alternative approach capable of overcoming this problem.

A number of different bootstrap intervals have been proposed in the literature (Efron and Tibshirani, 1993). These include the percentile interval, the bootstrap-t, a hybrid method, and an "accelerated" bootstrap method. Shao and Tu (1995) analyze these interval concepts, in the context of a confidence interval, to determine the relative rates at which the different interval's coverage probability converges on the true confidence level. The results of that analysis show that all two-sided, equal-tailed intervals have equal accuracy in terms of the order in which the coverage probabilities converge. It is not clear, however, how relevant this discussion is with respect to prediction intervals. The reason is that prediction intervals, unlike confidence intervals, must incorporate uncertainty associated with a single model error—uncertainty that is not amenable to asymptotic convergence theorems (see section 1.5.2). Consequently, it is difficult to assess the relative accuracies of different prediction interval approaches.

A second complication to be dealt with concerns the admissibility of the prediction interval. Predictions from the SPARROW model are always non-negative quantities, implying that in order for a prediction interval to be admissible, the interval must reside entirely in the space bounded from below by zero. As we saw in the discussion of bias in section 1.6.3, standard bootstrap methods do not guarantee admissible estimates. Some of the bootstrap interval methods listed above are admissible and some are not. Those that are not must be further modified to achieve admissibility.

Finally, the prediction interval we derive must account for the fact that, because of the nonlinearity of the SPARROW model, standard predictions are biased. As shown in section 1.6.3, bootstrap methods provide an estimate of proportional bias that can be used to obtain an admissible, corrected estimate. As we show below, the presence of bias affects different intervals in different ways, with differing levels of complexity required to address the problem.

We shall argue the merits of the various bootstrap prediction intervals in terms of their ability to accommodate the complications described above. To simplify and focus the discussion, the methods are compared in terms of their roles as confidence intervals: the extension of the methodology to obtain prediction intervals is discussed once a suitable method is decided. The percentile method is conceptually the simplest and was used to derive prediction intervals in earlier SPARROW papers (Smith and others, 1997, and Alexander and others, 2000). The interval is computed directly from the quantiles of the distribution of bootstrap replication estimates of the predictions, where the quantiles correspond to the percentiles representing equal probability in each tail and the share of the replications within the interval equals the coverage probability. Because each bootstrap replication is admissible, the quantiles forming the interval must also be admissible. The replications are not centered with respect to the parametric prediction. Therefore, if the replications are biased with respect to the parametric estimate, the bias would need to be removed in order to obtain an unbiased estimate of the interval. In fact, the bias would need to be removed twice: first to remove the bias between the bootstrap replications and the parametric estimate and second to remove the bias between the parametric estimate and the true value.

The accelerated bootstrap method is like the percentile method in that it is derived directly from the quantiles of the empirical distribution, without centering. As such, it must be admissible. Like the percentile method, in order to obtain an unbiased interval the interval must be shifted by twice the bias factor. The specific quantiles forming the interval are selected in a complex manner that requires additional bootstrap replications, the added computations resulting in a more accurate interval. Given the high computational costs of obtaining a bootstrap replication in SPARROW, however, the use of this method presents significant computational challenges.

The remaining bootstrap methods, the hybrid and bootstrap-t, are centered; that is, the empirical distribution is derived from bootstrap replication estimates that are expressed as differences from the parametric

estimate. The hybrid method, described in detail in section 1.5.3.2, uses bootstrap methods to determine the equal-tailed quantiles of the empirical distribution of the difference between the bootstrap replication estimates, $\hat{F}_i^{(r)}$ (see equation 1.135), and the parametric estimate, $\hat{F}_i$ (see equation (1.125). The bootstrap-t approach uses bootstrap methods to determine the quantiles of the empirical distribution of $\hat{F}_i^{(r)} - \hat{F}_i$, standardized (that is, divided) by the standard error of $\hat{F}_i$. If the standard error of $\hat{F}_i$ were known, the standardization would have no effect on the determination of quantiles as compared to the hybrid method except to divide each quantile by the standard error of $\hat{F}_i$. As such, the bootstrap-t and hybrid approaches would be equivalent. The two methods are not equivalent if the standard error of $\hat{F}_i$ is not known. In this case, the bootstrap-t approach requires a nested bootstrap; for each bootstrap iteration, a separate set of bootstrap replications must be generated to compute an independent, iteration-specific, estimate of the standard error. The empirical distribution then pertains to a random variable that has two sources of variation: deviation around the parametric estimate and the standard error used to normalize the difference.

The intervals derived from the hybrid and bootstrap-t methods are immutable to the presence of additive bias in the distribution. To see why this is so, consider a simple example in which $\hat{x}$ is a biased prediction of $\overline{x}$. Let the distribution of the estimate $\hat{x}$ be biased by an additive amount *B* with respect to $\hat{x}^*$, an unbiased estimator of *x*. Thus, $\hat{x} = \hat{x}^* + B$ and $E(\hat{x}) = E\left(\hat{x}^*\right) + B = \overline{x} + B$. For quantile *q*, define $H_{\hat{x}-\overline{x}}(q) = \Pr(\hat{x} - \overline{x} \le q)$ to be the distribution of $\hat{x} - \overline{x}$, and define $H^*_{\hat{x}^*-\overline{x}}(q) = \Pr\left(\hat{x}^* - \overline{x} \le q\right)$ to be the distribution of $\hat{x}^* - \overline{x} = \hat{x} - \overline{x} - B$. Let $q(p)$ and $q^*(p)$ represent the quantiles of $H_{\hat{x}-\overline{x}}$ and $H^*_{\hat{x}^*-\overline{x}}$ corresponding to probability *p*; that is, $p = H_{\hat{x}-\overline{x}}\left(q(p)\right) = H^*_{\hat{x}^*-\overline{x}}\left(q^*(p)\right)$. Because bias is additive, we have $H_{\hat{x}-\overline{x}}(q) = H^*_{\hat{x}^*-\overline{x}}(q - B)$; the distribution $H^*_{\hat{x}^*-\overline{x}}$ is just the distribution $H_{\hat{x}-\overline{x}}$ shifted by an amount –*B*, implying $q^*(p) = q(p) - B$. The equal-tailed prediction interval having coverage probability P_c derived from $H_{\hat{x}-\overline{x}}$ satisfies the relation $\Pr\left(\hat{x} - q\left((1-P_c)/2\right) \ge \overline{x} \le \hat{x} - q\left((1+P_c)/2\right)\right) \ge P_c$, implying the upper and lower bounds of the confidence interval are $\hat{x} - q\left((1-P_c)/2\right)$ and $\hat{x} - q\left((1+P_c)/2\right)$. But, because bias is additive, we have $\hat{x} - q(p) = \hat{x}^* - q^*(p)$; effectively, the bias causing the shift in the distribution of the statistic, which causes a shift in the quantile, is nullified by the bias arising in the statistic used to form the interval. Therefore, the prediction interval derived from additively biased centered statistics is equivalent to the confidence interval derived from bias adjusted statistics, which is the point we wished to establish.

The centering of the distributions raises the possibility that the interval will be inadmissible. This happens for the same reason the additive bias correction leads to inadmissible estimates (see section 1.6.3): the distributions of variables restricted to be positive tend to be skewed towards the high end. In fact, inadmissibility is more likely in this case because it is the potentially highly skewed right tale quantile that is subtracted from the parametric estimate to form the interval's lower bound.

One way to ensure admissibility using centered intervals is to transform the analysis into logarithmic space. The complication of bias is avoided if it is also assumed that bias is additive in logarithm space, in which case, as shown above, no further adjustments are necessary. Thus, all complications in forming an interval can be handled by adopting a centering method. In comparing the hybrid method with the bootstrap-t, the hybrid method is clearly easier to compute. The bootstrap-t requires the difference between the bootstrap replication estimate and the parametric estimate to be normalized by the standard error of the replication estimate prior to forming the empirical distribution. Because the bootstrap replication estimate is not a simple function of the data, such as a mean, the estimation of the standard error requires a separate bootstrap analysis *for each bootstrap iteration* (note, if only a single standard error were used for all bootstrap replications the bootstrap-t

method would default to the hybrid method). The computational burden of doing a double bootstrap analysis in the context of a SPARROW analysis makes the bootstrap-t method impractical.

We are left, then, with weighing the advantages and disadvantages of the hybrid and percentile methods. Both methods are approximately equal with respect to computational intensity, and both provide admissible intervals. Whereas the percentile method requires an estimate of bias in order to unbias the interval, the hybrid method accommodates bias without the need to estimate it. The hybrid method is also conceptually stronger. The standard premise underlying a confidence interval is that the shape of the distribution is known (or assumed) but the location with respect to the variable the interval is designed to cover is not known. This same premise underlies the empirical distribution used with the hybrid method. Thus, the hybrid method exemplifies the bootstrap paradigm by simply substituting the empirical distribution for the parametric distribution; the rest of the method is the same as the standard approach. Conversely, the percentile method contains the implicit assumption that not only is the shape of the distribution known, but also its location with respect to the true value. If this were in fact the case, there would be no need for a random interval to localize the true, unknown value; a simple mean of the distribution would reveal the true value with certainty. Given these considerations, therefore, we adopt the hybrid method to derive prediction intervals in the SPARROW model.

As explained above, the prediction interval is derived in logarithm space. Accordingly, the distribution underlying the prediction interval is for the random variable

$$\ln F_i^*\left(\hat{\boldsymbol{\beta}}\right) - \ln F_i = \ln F_i^*\left(\hat{\boldsymbol{\beta}}\right) - \ln F_i^*\left(\boldsymbol{\beta}\right) - e_i . \tag{1.145}$$

The source of uncertainty in equation (1.145) is given by the term $\ln F_i^*\left(\hat{\boldsymbol{\beta}}\right) - e_i$, where $\hat{\boldsymbol{\beta}}$ is the nonlinear weighted least squares estimate of the coefficient vector. Note that an estimate of the retransformation bias factor, or its logarithm transform, is not present in equation (1.145). Although the retransformation bias factor is a determinant of predictions, it is not required for the development of the prediction interval.

Let the cumulative distribution for equation (1.145) be given by $H_i\left(x\right)$. The standard approach to the determination of prediction intervals assumes this distribution is known, in which case the quantiles $H_i^{-1}\left(\left(1-P_c\right)/2\right)$ and $H_i^{-1}\left(\left(1+P_c\right)/2\right)$ satisfy the relation

$$\Pr\left(H_i^{-1}\left(\left(1-P_c\right)/2\right) \le \ln F_i^*\left(\hat{\boldsymbol{\beta}}\right) - e_i - \ln F_i^*\left(\boldsymbol{\beta}\right) \le H_i^{-1}\left(\left(1+P_c\right)/2\right)\right) \ge P_c , \tag{1.146}$$

where P_c is the prediction interval coverage probability (that is, the confidence level). The corresponding bounds on the prediction interval are

$$\left[\ln F_i^*\left(\hat{\boldsymbol{\beta}}\right) - H_i^{-1}\left(\left(1+P_c\right)/2\right),\ \ln F_i^*\left(\hat{\boldsymbol{\beta}}\right) - H_i^{-1}\left(\left(1-P_c\right)/2\right)\right]. \tag{1.147}$$

The bootstrap analog of $H_i\left(x\right)$ is given by the empirical distribution $H_{i,R}\left(x\right)$ corresponding to the *R* replications of the random variable

$$\ln F_i^*\left(\hat{\boldsymbol{\beta}}^{(r)}\right) - \left(\hat{e}/\sqrt{1-h}\right)_i^{(r)} - \ln F_i^*\left(\hat{\boldsymbol{\beta}}\right), \tag{1.148}$$

where $\hat{\boldsymbol{\beta}}^{(r)}$ is the estimate of $\boldsymbol{\beta}$ using the r^{th} bootstrap replication of the resampled observations (see section 1.6.3), and $\left(\hat{e}/\sqrt{1-h}\right)_i^{(r)}$ is the r^{th} random draw from the *N* leverage-adjusted error estimates obtained from the SPARROW model estimated with the original (non-resampled) data. Note that the empirical distribution is fully

non-parametric; neither the coefficient estimates nor the underlying errors are required to be normally distributed.

The appropriate quantiles of the empirical distribution are obtained by ordering the *R* estimates of $\ln F_i^*\left(\hat{\boldsymbol{\beta}}^{(r)}\right)-\left(\hat{e}/\sqrt{1-h}\right)_i^{(r)}$ in ascending order. Let $\tilde{q}_i(s)$ correspond to the s^{th} ordered value in this list. The quantiles forming the bounds of the prediction interval are given by (see appendix A for an explanation of the appropriate ranks)

$$
\begin{aligned}
H_{i,R}^{-1}\left((1-P_c)/2\right) &= \tilde{q}_i\left(\lfloor R(1-P_c)/2\rfloor+1\right)-\ln F_i^*\left(\hat{\boldsymbol{\beta}}\right), \text{ and} \\
H_{i,R}^{-1}\left((1+P_c)/2\right) &= \tilde{q}_i\left(\lceil P_c R\rceil+\lfloor R(1-P_c)/2\rfloor\right)-\ln F_i^*\left(\hat{\boldsymbol{\beta}}\right).
\end{aligned}
\tag{1.149}
$$

The term $\ln F_i^*\left(\hat{\boldsymbol{\beta}}\right)$ does not vary across the bootstrap replications so it is simply subtracted from the ranked bootstrap replications.

Substituting these quantiles for the parametric ones in equation (1.147) gives the bootstrap hybrid prediction interval in logarithm space; the prediction interval in real space is obtained by transforming the resulting bounds by the exponential function. In order to focus on the defined interval in real space, define $q_i\left((1-P_c)/2\right)$ and $q_i\left((1+P_c)/2\right)$ to be the $\lfloor R(1-P_c)/2\rfloor+1$ and $\lceil P_c R\rceil+\lfloor R(1-P_c)/2\rfloor$ ranked values of $F_i^*\left(\hat{\boldsymbol{\beta}}^{(r)}\right)\exp\left(\left(\hat{e}/\sqrt{1-h}\right)_i^{(r)}\right)$. The prediction interval in real space is given by

$$
\left[\frac{F_i^*\left(\hat{\boldsymbol{\beta}}\right)^2}{q_i\left((1+P_c)/2\right)},\frac{F_i^*\left(\hat{\boldsymbol{\beta}}\right)^2}{q_i\left((1-P_c)/2\right)}\right].
\tag{1.150}
$$

Note that the prediction interval in equation (1.150) collapses to zero if $F_i^*\left(\hat{\boldsymbol{\beta}}\right)$ is zero. Such a result is appropriate if there are no sources of a contaminant for reach *i*; it is less defensible if the zero value results from a constrained estimate of the coefficients (for example, if there is only one non-zero source for a reach but the source coefficient for that source is set at its constrained value of zero).

The prediction interval is potentially undefined if either the lower or upper quantiles $q_i\left((1-P_c)/2\right)$ or $q_i\left((1+P_c)/2\right)$ are zero. This could happen for the same reasons that $F_i^*\left(\hat{\boldsymbol{\beta}}\right)$ becomes zero. As explained above, if the zeros result from no sources, then the prediction interval is properly defined as a point at zero. If a zero results from a constrained coefficient, then the interval is effectively unbounded. At present, SPARROW does not distinguish between these two potential outcomes and reports a bound of zero for all cases in which either the numerator or denominator is zero. Note, however, that a bound of zero may be appropriate in the case of a source coefficient going to zero because the numerator, being squared, goes to zero faster than the denominator.

1.6.6 Discussion

The prediction methods described above are applied to every prediction generated by SPARROW having the same units as the dependent variable (units of mass). That is, the same methodology used to derive prediction results (bias-corrected predictions, standard errors and prediction intervals) for total flux is used to derive prediction results of flux by source, incremental flux (total and by source), flux in reservoirs, flux without in-stream decay (total and by source), etc. The predictions of flux by individual source each receive the same retransformation-bias correction factor under the assumption that the error for the reach is due to a common multiplicative error shared by each of the sources—an assumption that causes the sum of fluxes by individual source to equal total flux. Accordingly, to compute the variance and prediction intervals, a common multiplicative error term, equal to the reach's total flux multiplicative error, is assumed to apply to each source term. This ensures these predictions display model error that approximately adds up to the model error of the reach's total flux. It is, however, a strong assumption that overstates the true correlation between these predictions.

Additional predictions that can be derived directly from the flux predictions are the contaminant yield and concentration. The prediction of yield is obtained by simply dividing the flux prediction by the area of the region that generated the flux. Because area is known, the estimate of uncertainty for yield is simply the estimate of uncertainty for flux divided by area. A prediction of mean flow-weighted concentration is obtained in similar fashion by dividing the prediction for flux by the mean streamflow (see the discussion in section 1.3.1). If streamflow is precisely known, the uncertainty of this estimate is simply the uncertainty of flux divided by mean streamflow.

Given the rather coarse assumptions regarding the error structure in SPARROW (see section 1.6.1), it is best to treat each prediction (total, incremental, and flux by source) as an independent estimate. The estimate of uncertainty associated with disaggregated quantities (for example, flux by source) is meant to be suggestive pending a more rigorous analysis of error structure. One should be particularly wary of trying to infer uncertainty of aggregate quantities that represent averages of individual reach results. The simplifying assumptions regarding error in SPARROW produces a correlation structure that is not sufficient to make a reliable estimate of aggregate uncertainty; the estimates of error at an individual reach are more defensible.

A different kind of simplifying assumption is adopted for predictions that are not evaluated in the units of the dependent variable. Such predictions are typically for unitless quantities—for example, the share of flux delivered to a target reach. Because these variables are generally ratios of masses, the retransformation factors in the numerator and denominator should effectively cancel. Therefore, the prediction of these unitless variables does not receive a retransformation-bias correction factor. Accordingly, their standard errors do not include a model error component and the empirical distributions used to evaluate their confidence intervals do not account for variations due to model error. It is certainly presumptive to assert these unitless predictions exhibit no idiosyncratic variation; but in the absence of any estimate of the magnitude of this variation, it would be hard to justify a less conservative assumption. As remarked above, this problem can be resolved by adopting a richer error structure, one that allows for the estimation of the magnitude of idiosyncratic variation associated with individual processes.

As remarked in section 1.6.1, it is possible in SPARROW to condition predictions on measured flux at monitored reaches. This capability implies that predictions for all reaches at or below the monitored reach will be adjusted to be consistent with the measured flux. It is important to understand, however, that this capability does not necessarily imply that predicted flux immediately downstream of a monitored reach will be close in value to the measured flux. This is because of the assumed error structure for the model in which the error at the downstream reach is assumed to be independent of the error for the upstream reach. Recall from section 1.6.1 that a prediction is composed of two parts: a modeled component that depends only on the source, land-to-water, and in-stream processes specified in the SPARROW model, and an error component that is manifested in the prediction through the application of a retransformation factor. Because the measured flux at the monitored reach incorporates the model error directly, it is not appropriate to apply a retransformation factor to this flux; such an application would effectively double count the error process. The prediction for the reach immediately downstream of the monitored reach, however, does receive the retransformation factor. Thus, although the modeled flux at the downstream reach may be very close in value to the measured flux, the application of the retransformation factor necessary for obtaining the complete prediction, which could represent a large adjustment, may result in a predicted flux that differs from the measured flux by a considerable amount.

It should also be noted that if the SPARROW model is run in prediction mode for purposes of simulating water-quality conditions under alternative management scenarios, then it is not appropriate to condition the predictions on measured flux at monitored reaches. The measured flux is consistent with management actions actually taken. As such, the conditioning of predictions on measured flux would not incorporate the changes in water quality associated with the alternative management scenario. For purposes of comparison, therefore, it is reasonable to obtain predictions for both existing conditions and alternative management scenarios that are not conditioned on measured flux.

1.6.6.1 The effect of error in measured flux (advanced)

Section 1.5.3.5 discusses the potentially biased coefficient estimators that result from estimation with dependent variable data that are measured with error. The bias arises because the dependent variable is a conditional estimate of flux; although the difference between unobserved true flux and estimated flux is orthogonal to estimated flux, it is not necessarily orthogonal to the SPARROW model predictors. The absence of this orthogonal variation in the dependent variable, which is a real variation in true flux, implies the model coefficients are biased if the explanatory variables are capable of explaining part of it. It is shown in section 1.5.3.5 that information on the error of the monitored flux estimate and coefficient and model error in the SPARROW model can be used to bound the bias.

The problem of incomplete variation in the dependent variable causes a similar problem for the assessment of prediction uncertainty (see Mullaney and others, 2002, for a discussion of this problem in the context of a time series model of flux). In standard assessments of prediction error, the model mean squared error incorporates all residual variation not accounted for by the model. If the dependent variable is a conditional estimate of the true dependent variable, however, the error in measured flux potentially represents additional variation not accounted for by the model. A complication in the assessment of total model error is that the measurement error could be correlated with the model error associated with the measured component. Unfortunately, it is not possible to precisely estimate this correlation, making it impossible to fully assess the error of the model; however, as with the case of estimation, it is possible to place bounds on the extent of this "missing" error component.

To solidify concepts, let y_i be the true value of the logarithm of flux at reach *i* and let $\tilde{y}_i$ represent the measured component. As in section 1.5.3.5, let u_i be the difference between true and measured logarithm of flux, so that

$$y_i = \tilde{y}_i + u_i , \tag{1.151}$$

where u_i is assumed to be orthogonal to $\tilde{y}_i$, having variance σ_u^2. Let $\tilde{y}_i$ be divided into two components, a component that is "explained" by the model, x_i, and a model error component, e_i, assumed to be orthogonal to x_i. The relation between true flux and model-explained flux is

$$y_i = x_i + u_i + e_i . \tag{1.152}$$

The measured mean squared error of the model is given by σ_e^2; however, the full model error is not necessarily equal to $\sigma_e^2 + \sigma_u^2$. This is because the terms u_i and e_i could be correlated. To see how this might arise, note that orthogonality between u_i and $\tilde{y}_i$ implies $E(u_i \tilde{y}_i) = E(u_i x_i) + E(u_i e_i) = 0$. If the model predictors are correlated with measurement error term u_i, as was the case discussed in section 1.5.3.5, then $E(u_i x_i)$ is not equal to zero. Consequently, $E(u_i e_i)$ must be different from zero also. The variance of the full model error takes the form

(1.153) $$V\left(u_i + e_i\right) = \sigma_u^2 + \sigma_e^2 + 2\rho_{ue}\sigma_u\sigma_e,$$

where ρ_{ue} is the correlation coefficient between u_i and e_i.

Without data on true flux, it is impossible to know the value of ρ_{ue}, and so it is impossible to know the exact value of full model error. However, by recognizing that the correlation coefficient must be between –1 and 1, it is possible to place bounds on this error. Setting the correlation coefficient to its extreme values, we see that the full model variance $V\left(u_i + e_i\right)$ must have the bounds

(1.154) $$\left(\sigma_e - \sigma_u\right)^2 \le V\left(u_i + e_i\right) \le \left(\sigma_e + \sigma_u\right)^2.$$

As was the case in section 1.5.3.5, the bounds can be evaluated using results from flux estimation and SPARROW model output. The root mean squared error from the SPARROW model is σ_e, and σ_u represents the average across all stations of the ratio of the standard error of measured flux to the estimate of measured flux.

There is no simple way of using these bounds to modify the standard errors of predictions or prediction confidence intervals generated by SPARROW. This is because these estimates are derived using robust non-parametric methods (a combination of the smearing estimator and bootstrapping). One approach that is likely to give reasonable approximate estimates is to rescale the estimated model error terms prior to prediction by the factor $\left|\sigma_e - \sigma_u\right| / \sigma_e$ to obtain the lower bound on model error and by $\left(\sigma_e + \sigma_u\right) / \sigma_e$ to obtain the upper bound. However, this feature is not currently implemented in SPARROW.

1.6.7 Share of flux delivered to a target reach

An additional prediction of general interest included in standard SPARROW output is the fraction of flux leaving a reach that is delivered to some downstream target reach. In many applications, the target is a waterbody of some interest to water-quality managers, examples being estuaries or reservoirs. In the case in which there is only one such target downstream from any given reach, the fraction delivered represents the amount of flux leaving the reach that makes its way through the reach network to the target. If there are multiple target reaches downstream from a given reach, the fraction delivered is interpreted as the fraction of flux leaving the reach that is delivered to the nearest downstream target reach.

The method used to compute the fraction of flux delivered to a target reach accounts for all in-stream attenuation processes defined in the model. If reservoirs represent large sinks for the contaminant being studied and the target is an estuary, then the fraction delivered should be small for all reaches upstream of a reservoir. If there is larger in-stream decay in small streams as compared to large streams, as has been found in SPARROW models of total nitrogen, then a map (figure 1.28) of fraction delivered to estuaries shows the fraction delivered becomes smaller the higher in the network the reach is located. Conversely, a map of the fraction delivered for a contaminant that is very conservative in streams and reservoirs shows little variation across the network.

The method used in SPARROW to compute the fraction of flux delivered to a target accounts for diversions in the reach network. To understand how this works, it helps to understand something about the algorithm used to compute the share of flux delivered. The calculation of the fraction delivered is a reach accumulation process that works in reverse. That is, in normal applications of the SPARROW model, prior to the analysis, reaches are ordered in hydrologic order from upstream to downstream, and the model is processed by accumulating flux in the downstream direction. As long as the ordering is accurate, no reach is processed prior to processing all upstream reaches. Where a diversion is encountered, the flux going to each reach immediately downstream from the diversion is determined by the “fraction-diverted” variable. The sum of the “fraction-diverted” variable is always one across all diverting reaches (see section 1.3.2), so no flux is lost in the process of diversion unless one of the diverted streams terminates—that is, fails to reconnect with the stream network.

For the computation of fraction delivered, the ordering of reaches is exactly reversed. The analysis begins at the target and the fraction delivered to the target is modified at a reach by multiplying by that reach’s

incremental delivery fraction. Moving in the upstream direction, if two reaches are joined at a diversion, the cumulative fraction delivered from each downstream reach is multiplied by that reach's "fraction-diverted" variable prior to being aggregated. In this way, the fraction delivered represents the average rate of delivery across the multiple possible pathways between the reach and the downstream target. Note also that if a diverted reach terminates before reaching a target, the fraction delivered at the point of the diversion is effectively reduced by the "fraction-diverted" amount.

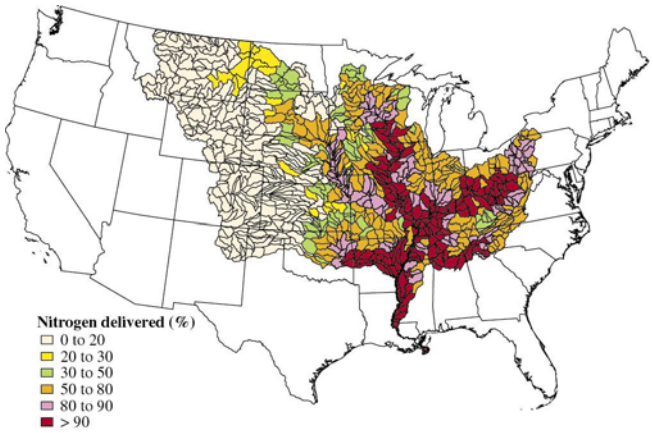

Figure 1.28. SPARROW model estimates of in-stream nitrogen delivered to the Gulf of Mexico. [From Alexander and others (2000).]

1.7 Summary

This completes the discussion of the concepts and theory underlying the SPARROW surface-water model. The SPARROW model is shown to derive from a few simple concepts—mass balance, spatial referencing, and mean annual flux. A rigorous statistical methodology exists for estimating the model and for deriving predictions from it. The results of a SPARROW analysis can be used to test hypotheses regarding hydrologic processes, characterize mean conditions, perform a mass balance accounting of the sources of flux, or simulate alternative management scenarios. Perhaps the most important feature of the SPARROW model, however, is the close connection forged between the results of an application and the basic hydrologic data from which they are derived—a connection that adds considerable credibility to the findings.

Although substantial progress has been made in applying the SPARROW methodology to a number of water-quality constituents across diverse hydrologic settings, there is still much work that needs to be done. There are many contaminants that have yet to be included in a SPARROW analysis—the analysis of biological data represents one particularly fertile area for future research—and additional insights will be gained by the development of multi-year and seasonal mean flux models. Many of these prospective analyses can be performed within the existing SPARROW framework; however, there are also many analyses that will require modifications to the methodology. For example, an extension of the basic SPARROW model to facilitate the simultaneous analysis of multiple contaminants will permit the analyses of nitrogen species and suites of pesticides. The development of a more flexible error structure will provide more defensible model predictions—particularly predictions of aggregated water conditions.

The basic SPARROW methodology described in Part 1 is fully implemented in a model written for the SAS computer software. A detailed description of the model in terms of model specification, required data input, and model output is contained in Part 2 of this documentation. Many of the hydrologic processes introduced in Part 1 of the documentation are further discussed in Part 2 as examples of feasible SPARROW model specifications.

1.8 References

Alexander, R.B., Brakebill, J.W., Brew, R.E., and Smith, R.A., 1999, ERF1—Enhanced river reach file 1.2: U.S. Geological Survey Open-File Report 99-457.

Alexander R.B., Smith, R.A., and Schwarz, G.E., 2000, Effect of stream channel size on the delivery of nitrogen to the Gulf of Mexico: Nature, v. 403, p. 758-761.

Alexander, R.B., Smith, R.A., Schwarz, G.E., Preston, S.D., Brakebill, J.W., Srinivasan, R., and Pacheco, P.A., 2001, Atmospheric nitrogen flux from the watersheds of major estuaries of the United States—An application of the SPARROW watershed model, *in* Valigura, R.A., Alexander, R.B., Castro, M.S., Meyers, T.P., Paerl, H.W., Stacey, P.E., and Turner, R.E. (eds.), Nitrogen loading in coastal water bodies—An atmospheric perspective: Washington, D.C., American Geophysical Union Monograph 57, p. 119-170.

Alexander R.B., Elliott A.H., Shankar U., and McBride G.B., 2002, Estimating the sources and transport of nutrients in the Waikato River basin, New Zealand: Water Resources Research, v. 38, p. 1268-1290.

Alexander, R.B., Johnes, P.J., Boyer, E.A., and Smith, R.A., 2002, A comparison of methods for estimating the riverine export of nitrogen from large watersheds: Biogeochemistry, v. 57/58, p. 295-339.

Alexander, R.B., Smith, R.A., and Schwarz, G.E., 2004, Estimates of diffuse phosphorus sources in surface waters of the United States using a spatially referenced watershed model: Water Science and Technology, v. 49, p. 1-10.

Amemiya, T., 1985, Advanced Econometrics: Cambridge, Mass., Harvard University Press, 521 p.

Arnold, J.G., Williams, J.R., Nicks, A.D., and Sammons, N.B., 1990, SWRRB—A basin scale simulation model for soil and water resources management: College Station, TX, Texas A&M University Press.

Bachman, L.J., Lindsey, B.D., Brakebill, J.W., and Powars, D.S., 1998, Ground-water discharge and base-flow nitrate loads of nontidal streams, and their relation to a hydrogeomorphic classification of the Chesapeake Bay Watershed, Middle Atlantic Coast: U.S. Geological Survey Water-Resources Investigations Report 98-4059, 71 p.

Beaulac, M.N., and Reckhow, K.H., 1982, An examination of land use—nutrient export relationships: Water Resources Bulletin, v. 18, p. 1013-1024.

Beven, K., 2002, Towards an alternative blueprint for a physically based digitally simulated hydrologic response modeling system: Hydrological Processes, v. 16, p. 189-206.

Beven, K., and Kirkby, M., 1979, A physically-based variable contributing area model of basin hydrology: Hydrologic Science Bulletin, v. 24, p. 43-69.

Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Jr., Jobes, T.H., and Donigian, A.S., Jr., 2001, Hydrological simulation program-Fortran, HSPF version 12 user's manual: Mountain View, CA, AQUA TERRA Consultants, 873 p.

Bohlke, J.K., and Denver, J.M., 1995, Combined use of groundwater dating, chemical, and isotopic analyses to resolve the history and fate of nitrate contamination in two agricultural watersheds, Atlantic coastal plain, Maryland: Water Resources Research, v. 31, p. 2319-2339.

Bohlke, J.K., Harvey, J.W., and Voytek, M.A., 2004, Reach-scale isotope tracer experiment to quantify denitrification and related processes in a nitrate-rich stream, mid-continent USA: Limnology and Oceanography, v. 49, p. 821–838.

Boyer E.W., Goodale, C.L., Jaworski, N.A., and Howarth, R.W., 2002, Anthropogenic nitrogen sources and relationships to riverine nitrogen export in the northeastern USA: Biogeochemistry, v. 57, p. 137-169.

Boyer, E.W., Alexander, R.B., Parton, W.J., Li, C., Butterbach-Bahl, K., Donner, S.D., and Skaggs, R.W., and Del Grosso, S.J., in press, Modeling denitrification in terrestrial and aquatic ecosystems at regional scales—current approaches and needs: Ecological Applications.

Brakebill, J.W., and Preston, S.E., 1999, Digital data used to relate nutrient inputs to water quality in the Chesapeake Bay watershed, version 1.0: U.S. Geological Survey Water Open-File Report 99-60, *http://md.usgs.gov/publications/ofr-99-60/*.

Breiman, L., 1968, Probability: Reading, Mass., Addison-Wesley, 421 p.

Bricker, S.B., Clement, C.G., Pirhalla, D.E., Orlando, S.P., and Farrow, D.R.G., 1999, National estuarine eutrophication assessment—effects of nutrient enrichment in the nation's estuaries: U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, p. 20-36, 155.

Caraco, N.F., Cole, J.J., Likens, G.E., Lovett, G.M., and Weathers, K.C., 2003, Variation in NO3 export from flowing waters of vastly different sizes—Does one model fit all?: Ecosystems, v. 6, p. 344-352.

Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N., and Smith, V.H., 1998, Nonpoint pollution of surface waters with phosphorus and nitrogen: Ecological Applications, v. 8, p. 559-568.

Chapra, S.C., 1997, Surface water-quality modeling: New York, McGraw-Hill, 844 p.

Cohn, T.A., Caulder, D.L., Gilroy, E.J., Zynzuk, L.D., and Summers, R.M., 1992, The validity of a simple statistical model for estimating fluvial constituent loads: an empirical study involving nutrient loads entering Chesapeake Bay: Water Resources Research, v. 28, no. 9, p. 2353-63.

Cohn, T.A., Gilroy, E.J., and Baier, W.G., 1992, Estimating fluvial transport of trace constituents using a regression model with data subject to censoring: Proceedings of the Section on Statistics and the Environment of the American Statistical Association, p. 1-9.

Cooper, A.B. and Bottcher, A.B., 1993, Basin scale modelling as a tool for water resource planning: Journal of Water Resources Planning and Management, v. 119, p. 303-323.

Cramer, J.S., 1986, Econometric applications of maximum likelihood methods: Cambridge, Cambridge University Press, 208 p.

Cunnane, C., 1978, Unbiased plotting positions—a review: Journal of Hydrology, v. 37, p. 205-222.

Davidson, J., 1994, Stochastic Limit Theory: Oxford, Oxford University Press, 539 p.

Davidson, R., and MacKinnon, J.G., 1993, Estimation and inference in econometrics: Oxford, Oxford University Press, 874 p.

Davison, A.C., and Hinkley, D.V., 1997, Bootstrap methods and their application: Cambridge, Cambridge U. Press, 582 p.

De Wit, M.J.M., 2000, Modeling nutrient fluxes from source to river load—a macroscopic analysis applied to the Rhine and Elbe basin: Hydrobiologia, v. 410, p. 123-130.

De Wit, M.J.M., 2001, Nutrient fluxes at the river basin scale 1—the PolFlow model: Hydrological Processes, v. 15, p. 743-759.

Doherty, J., 2004, PEST model-independent parameter estimation—user manual (5th ed.): Watermark Numerical Computing, 333 p.

Dole, R.B., 1909, The quality of surface waters in the United States, part I—analyses of waters east of the one hundredth meridian: U.S. Geological Survey Water Supply Paper 236, 123 p.

Duan, N., 1983, Smearing estimate—a nonparametric retransformation method: Journal of the American Statistical Association, v. 78, p. 605-610.

Duan, Q., Gupta, H.V., Sorooshian, S, Rousseau, A.N., Turcotte, R. (editors), 2003, Calibration of watershed models: Washington, D.C., American Geophysical Union Water Science and Application 6, 346 p.

Efron, B., and Tishirani, Robert J., 1993, An introduction to the bootstrap: New York, Chapman and Hall, 436 p.

Elliott, A.H., Alexander, R.B., Schwarz, G.E., Shankar, U., Sukias, J.P.S., and McBride, G.B., 2005, Estimation of nutrient sources and transport for New Zealand using the hybrid physical-statistical model SPARROW: Journal of Hydrology (New Zealand), v. 44, p. 1-27.

Filoso, S., Vallino, J., Hopkinson, C., Rastetter, E., and Claessens, L., 2004, Modeling nitrogen transport in the Ipswich River Basin, Massachusetts, using a Hydrological Simulation Program in Fortran (HSPF): Journal of the American Water Resources Association, v. 40, p. 1365-1384.

Finney, D.J., 1941, On the distribution of a variate whose logarithm is normally distributed: Journal of the Royal Statistical Society Supplement, v. 7, p. 155-161.

Focazio, M.J., Plummer, L.N., Bohlke, J.K., Busenberg, E., Bachman, L.J., and Powars, D.S., 1997, Preliminary estimates of residence times and apparent ages of ground water in the Chesapeake Bay watershed, and water-quality data from a survey of springs: U.S. Geological Survey Water-Resources Investigations Report 97-4225, 75 p.

Gallant, A.R., 1987, Nonlinear statistical models: New York, Wiley, 610 p.

Gallant, A.R., and White, H., 1988, A unified theory of estimation and inference for nonlinear dynamic models: New York, Basil Blackwell, 155 p.

Gilroy, E.J., Hirsch, R.M., and Cohn, T.A., 1990, Mean square error of regression-based constituent transport estimates: Water Resources Research, v. 26, no. 9, p. 2069-77.

Gupta, V.K., and Sorooshian, S., 1985, The relationship between data and the precision of parameter estimates of hydrologic models: Journal of Hydrology, v. 81, p. 57-77.

Hamilton, J.D., 1994, Time series analysis: Princeton, NJ, Princeton University Press, 799 p.

Harvey, J.W., and Wagner, B.J., 2000, Quantifying hydrologic interactions between streams and their subsurface hyporheic zones, *in* Jones, J.A. and Mulholland, P.J., eds., Streams and groundwaters: San Diego, CA, Academic Press, p. 3–44.

Helsel, D.R., and Hirsch, R.M., 1992, Studies in environmental science 49—statistical methods in water resources: Amsterdam, Elsevier, 522 p.

Higgins, J.M. and Kim, B.R., 1981, Phosphorus retention models for Tennessee Valley Authority reservoirs: Water Resources Research, v. 17, p. 571-576.

Hill, M.C, 1998, Methods and guidelines for effective model calibration: U.S. Geological Survey Water-Resources Investigations Report 98-4005, 90 p.

Hill, M.C., 1992, A computer program (MODFLOWP) for estimating parameters of a transient, three-dimensional, ground-water flow model using nonlinear regression: U.S. Geological Survey Open-File Report 91-484, 358 p.

Hoos, A.B., 2005, Evaluation of the SPARROW model for estimating transport of nitrogen and phosphorus in streams of the Interior Low Plateau Ecoregion, Tennessee, Kentucky, and Alabama, during 1992-2002, in Tennessee Water Resources Symposium, 15th, Burns, Tenn., 2005, Proceedings: Tennessee Section of the American Water Resources Association, p. 2A-47.

Howarth, R.W., Billen, G., Swaney, D., Townsend, A., Jaworski, N., Lajtha, K., Downing, J.A., Elmgren, R., Caraco, N., Jordan, T., Berendse, F., Freney, J., Kudeyarov, V., Murdoch, P., and Zhao-Liang, Z., 1996, Regional nitrogen budgets and riverine N and P fluxes for the drainages to the North Atlantic Ocean—natural and human influences: Biogeochemistry, v. 35, p. 75-139.

Howarth, R.W., Boyer, E.W., Pabich, W.J., and Galloway, J.N., 2002, Nitrogen use in the United States from 1961-2000 and potential future trends: Ambio, v. 31, p. 88-96.

ICF Consulting, 2002, User's guide to the Regional Modeling System for Aerosols and Deposition (REMSAD)—version 7: *http://remsad.saintl.com/documents/remsad_users_guide_07-22-02.pdf*

Jakeman, A. and Hornberger, G., 1993, How much complexity is warranted in a rainfall-runoff model?: Water Resources Research, v. 29, p. 2637-2649.

Jobson, H.E., 1996, Prediction of traveltime and longitudinal dispersion in rivers and streams: U.S. Geological Survey Water-Resources Investigations Report 96-4013, 69 p.

Johnes, P.J., Moss, B., and Phillips, G.L., 1996, The determination of water quality by land use, livestock numbers and population data—testing of a model for use in conservation and water quality management: Freshwater Biology, v. 36, p. 951-473.

Jordan, T.E., and Weller, D.E., 1996, Human contributions to terrestrial nitrogen flux—assessing the sources and fates of anthropogenic fixed nitrogen: Bioscience, v. 46, p. 655-664.

Jowett, I.G., 1998, Hydraulic geometry of New Zealand rivers and its use as a preliminary method of habitat assessment: Regulated Rivers—Research and Management, v. 14, p. 451-466.

Judge, G.G., Griffiths, W.E., Hill, R.C., and Lee, T., 1985, The theory and practice of econometrics (2nd ed.): New York, John Wiley and Sons, 1019 p.

Kansas Department of Health and Environment, 2004, Surface water nutrient reduction plan: *http://www.kdhe.state.ks.us/water/download/ks_nutrient_reduction_plan_12_29_final.pdf* (accessed March 1, 2005).

Kelly, C.A., Rudd., J.W.M., Hesslein, R.H., Schindler, D.W., Dillon, P.J., Driscoll, C.T., Gherini, S.A., and Hecky, R.E., 1987, Prediction of biological acid neutralization in acid-sensitive lakes: Biogeochemistry, v. 3, p. 129-140.

Kellogg, R.L., Lander, C.H., Moffitt, D.C., and Gollehon, N., 2000, Manure nutrients relative to the capacity of cropland and pastureland to assimilate nutrients—spatial and temporal trends for the United States: USDA-NRCS-ERS Publication No. nps00-0579, 140 p.

Knopman, D.S., and Smith, R.A., 1993, Twenty years of the clean water act—has U.S. water quality improved?: Environment, v. 35, p. 17-41.

Larson, S.J., and Gilliom, R.J., 2001, Regression models for estimating herbicide concentrations in U.S. streams from watershed characteristics: Journal of the American Water Resources Association, v. 37, no. 5, p. 1349-1367.

Lindsey, B.D., Phillips, S.W., Donnelly, C.A., Speiran, G.K., Plummer, L.N., Bohlke, J.K., Focazio, M.J., Burton, W.C., and Busenberg, E., 2003, Residence times and nitrate transport in ground water discharging to streams in the Chesapeake Bay watershed: U.S. Geological Survey Water-Resources Investigations Report 03-4035, 201 p.

Maddala, G.S., 1983, Limited dependent and qualitative variables in econometrics: Cambridge, Cambridge Univ. Press, 401 p.

Malliaris, A.G., and Brock, W.A., 1982, Stochastic methods in economics and finance: New York, North-Holland, 303 p.

McIsaac, G.F., David, M.B., Gertner, G.Z., and Goolsby, D.A., 2001, Eutrophication—nitrate flux in the Mississippi River: Nature, v. 414, p. 166-167.

McMahon, G., and Roessler, C., 2002, A regression-based approach to understand baseline total nitrogen loading for TMDL planning, *in* Proceedings of Water Environment Federation National TMDL Science and Policy Conference, Phoenix, Arizona, November 13-16, 2002: Phoenix, AZ, Water Environment Federation, 28 p.

McMahon G., Alexander R.B., and Qian S., 2003, Support of TMDL programs using spatially referenced regression models: ASCE Journal of Water Resources Planning and Management, v. 129, p. 315-329.

Michel, R.L., 1992, Residence times in river basins as determined by analysis of long-term tritium records: Journal of Hydrology, v. 130, p. 367-378.

Miller, D.A., and White, R.A., 1998, A conterminous United States multilayer soil characteristics data set for regional climate and hydrology modeling: Earth Interactions, v. 2, n. 2, p. 1-26.

Molot, L.A., and Dillon, P.J., 1993, Nitrogen mass balances and denitrification rates in central Ontario Lakes: Biogeochemistry, v. 20, p. 195-212.

Montgomery, D.C., and Peck, E.A., 1982, Introduction to linear regression analysis: New York, John Wiley and Sons, 504 p.

Moore, R.B., Johnston, C.M., Robinson, K.W., and Deacon, J.R., 2004, Estimation of total nitrogen and phosphorus in New England streams using spatially referenced regression models: U.S. Geological Survey Scientific Investigations Report 2004-5012, 50 p. *http://water.usgs.gov/pubs/sir/2004/5012/*

Mueller, D.K., Ruddy, B.C., and Battaglin, W.A., 1997, Logistic model of nitrate in streams of the upper-midwestern United States: Journal of Environmental Quality, v. 26, p. 1223-1230.

Mullaney, J.R., Schwarz, G.E., and Trench, E.C.T., 2002, Estimation of nitrogen yields and loads from basins draining to Long Island Sound, 1988-98: U.S. Geological Survey Water-Resources Investigations Report 02-4044, 84 p.

NEIWPCC (New England Interstate Water Pollution Control Commission), 2004, New England SPARROW water quality model: Interstate Water Report, v. 1, no. 3, p. 6-7. *http://www.neiwpcc.org/PDF_Docs/iwr_s04.pdf*

Newbold, J.D., Elwood, J.W., O'Neill, R.V., and Van Winkle, W., 1981, Measuring nutrient spiraling in streams: Canadian Journal of Fisheries and Aquatic Sciences, v. 38, p.860–863.

Nolan, J.V., Brakebill, J.W., Alexander, R.B., and Schwarz, G.E., 2002, ERF1_2—Enhanced River Reach File 2.0: U.S. Geological Survey Open-File Report 02-40 [online], *http://water.usgs.gov/GIS/metadata/usgswrd/XML/erfl_2.xml*

National Research Council, 2000, Clean coastal waters—understanding and reducing the effects of nutrient pollution: Washington, D.C., National Academy Press, 405 p.

National Research Council, 2001a, Assessing the TMDL approach to water quality management: Washington, D.C., National Academy Press, 109 p.

National Research Council, 2001b, Ecological indicators for the nation: Washington, D.C., National Academy Press, 180 p.

National Research Council, 2002a, Review of the USGS National Water-Quality Assessment Program: Washington, D.C., National Academy Press, 238 p.

National Research Council, 2002b, Estimating water use in the United States—a new paradigm for the National Water-Use Information Program: Washington, D.C., National Academy Press, 176 p.

National Resources Conservation Service, 2005, National Resources Inventory—a statistical survey of land use and natural resource conditions and trends on U.S. non-Federal lands: U.S. Department of Agriculture. *http://www.nrcs.usda.gov/technical/NRI/*

Peierls, B.L., Caraco, N.F., Pace, M.L., and Cole, J.J., 1991, Human influence on river nitrogen: Nature, v. 350, p. 386-387.

Peterson, B.J., Wollheim, W.M., Mulholland, P.J., Webster, J.R., Meyer, J.L., Tank, J.L., Marti, E., Bowden, W.B., Valett, H.M., Hershey, A.E., McDowell, W.H., Dodds, W.K., Hamilton, S.K., Gregory, S., and Morrall, D.D., 2001, Control of nitrogen export from watersheds by headwater streams: Science, v. 292, p. 86–90.

Poeter, E.P., and Hill, M.C., 1998, A computer code for universal inverse modeling: U.S. Geological Survey Water-Resources Investigations Report 98-4080, 116 p.

Preston, S.D., and Brakebill, J.W., 1999, Application of spatially referenced regression modeling for the evaluation of total nitrogen loading in the Chesapeake Bay watershed: Baltimore, MD, U.S. Geological Survey Water Resources Investigations Report 99-4054, 11 p., *http://md.water.usgs.gov/publications/wrir-99-4054/*.

Puckett, L.J., 1995, Identifying the major sources of nutrient water pollution: Environmental Science and Technology, v. 29, n. 9, p. 408-414.

Qian, Song S., Reckhow, Kenneth H., Zhai, Jun, and McMahon, Gerard, 2005, Nonlinear regression modeling of nutrient loads in streams—a Bayesian approach: Water Resources Research, v. 41, n. 7, W07012 (10 pages).

Rao, C.R., 1973, Linear statistical inference and its applications (2nd ed.): New York, Wiley, 625 p.

Reckhow, K.H., and Chapra, S.C., 1983, Engineering approaches for lakes management, vol. 1—data analysis and empirical modeling: Boston, MA, Butterworth publishers, 340 p.

Royston, J.P., 1982, The W test for normality: Applied Statistics, v. 44, p. 176-180.

Runkle, R.L., Crawford, C.G., and Cohn, T.A., 2004, Load estimator (LOADEST)—A FORTRAN program for estimating constituent loads in streams and rivers: U.S. Geological Survey Techniques and Methods 4-A5, 75 p.

Running, S.W., and Hunt, E.R., Jr., 1993, Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models, *in* Ehleringer, J.R., and Field, C.B. (eds.), Scaling physiological processes—leaf to globe: New York, Academic Press, p. 141-158.

Rutherford, J.C., Williamson, R.B., and Cooper, A.B., 1987, Nitrogen, phosphorus, and oxygen dynamics in rivers, *in* Viner, A.B. (ed.), Inland Waters of New Zealand: Wellington, New Zealand, Science Information Publishing Centre, DSIR, p. 139-165.

SAS, 1993, SAS/ETS user's guide, version 6 (2nd ed.): Cary, NC, SAS Institute Inc., 1022 p.

SAS, 1999, SAS OnlineDoc, version 8, *http://v8doc.sas.com/sashtml/* accessed 4/29/2005, SAS Inc.

Schwarz, G.E., Smith, R.A., Alexander, R.B., and Gray, J.R., 2001, A spatially referenced regression model (SPARROW) for suspended sediment in streams of the conterminous U.S., *in* U.S. Subcommittee on Sedimentation, Proceedings of the seventh Federal Interagency Sedimentation Conference, March 25-29, 2001, Reno, Nevada USA: Washington, D.C., U.S. Subcommittee on Sedimentation, p. VII-80-7.

Seitzinger, S., Styles, R.V., Boyer, E.W., Alexander, R.B., Billen, G., Howarth, R.W., Mayer, B., and vanBreemen, N., 2002, Nitrogen retention in rivers—model development and application to watersheds in the Eastern U.S.: Biogeochemistry, v. 57, p. 199-237.

Shao, J., and Tu, D., 1995, The jackknife and bootstrap: New York, Springer, 516 p.

Sharpley, A.N., and Williams, J.R. (eds.), 1990, EPIC—Erosion/Productivity Impact Calculator—1, model documentation, U.S. Department of Agriculture Technical Bulletin No. 1768, 235 p.

Singh, V.P., 1995, Computer models of watershed hydrology: Highlands Ranch, CO, Water Resources Publications, 1,144 p.

Smith, R.A., and Alexander, R.B., 2000, Sources of nutrients in the nation's watersheds, *in* Managing nutrients and pathogens from animal agriculture, Proceedings from the Natural Resource, Agriculture, and Engineering Service Conference for Nutrient Management Consultants, Extension Educators, and Producer Advisors, March 28-30, 2000, Camp Hill, Pennsylvania: Ithaca, NY, Natural Resource, Agriculture, and Engineering Service, Cooperative Extension, 508 p.

Smith, R.A., Alexander, R.B., and Schwarz, G.E., 2003, Estimating the natural background concentrations of nutrients in streams and rivers of the conterminous United States: Environmental Science and Technology, v. 37, p. 3039-3047.

Smith, R.A., Schwarz, G.E., and Alexander, R.B., 1997, Regional interpretation of water-quality monitoring data: Water Resources Research, v. 33, p. 2781-2798.

Smith, R.A., Alexander, R.B., Tasker, G.D., Price, C.V., Robinson, K.W., and White, D.A., 1994, Statistical modeling of water quality in regional watersheds, *in* U.S. Environmental Protection Agency, Proceedings of watershed '93—a national conference on watershed management, Alexandria, Virginia, March 21-24, 1993: Washington, DC, U.S. Environmental Protection Agency, EPA 840-R-94-002, p. 751-754.

Smith, R.A., Alexander, R.B., and Schwarz, G.E., 2004, Effects of structural changes in U.S. animal agriculture on fecal bacterial contamination of streams—comparison of confined and unconfined livestock operations [abs.]: American Geophysical Union, EOS, Transactions, v. 85, no. 47, Fall Meeting Supplement, Abstract H43B-0371.

Solley, W.B., Pierce, R.K., and Perlman, H.A., 1998, Estimated use of water in the United States in 1995: U.S. Geological Survey Circular 1200, 71 p.

Srinivasan, R., Arnold, J.G., Muttiah, R.S., Walker, D., and Dyke, P.T., 1993, Hydrologic unit modeling of the United States (HUMUS), *in* Yang, S. (ed.), Advances in hydro-science and engineering, vol I, part A: University, MS, Center for Computational Hydroscience and Engineering, School of Engineering, University of Mississippi, p. 451-456.

Stalnacke, P., Grimvall, A., Libiseller, C., Laznik, M., and Kokorite, I., 2003, Trends in nutrient concentrations in Latvian rivers and the response to the dramatic change in agriculture: Journal of Hydrology, v. 283, p. 184-205.

Stream Solute Workshop, 1990, Concepts and methods for assessing solute dynamics in stream ecosystems: Journal of the North American Benthological Society, v. 9, p. 95-119.

Tasker, G.D., Hodge, S.A., and Barks, C.S., 1996, Region of influence regression for estimating the 50-year flood at ungaged sites: Journal of the American Water Resources Association, v. 32, no. 1, p. 163-170.

Thomas, S.A., Valett, H.M., Mulholland, P.J., Fellows, C.S., Webster, J.R., Dahm, C.N., and Peterson, C.G., 2001, Nitrogen retention in headwater streams—the influence of groundwater-surface water exchange: The Scientific World Journal, v. 1 (S2), p. 623-631.

Tobin, J., 1958, Estimation of relationships for limited dependent variables: Econometrica, v. 26, p. 24-36.

Trimble, S.W. and Crosson, P., 2000, U.S. soil erosion rates – myth and reality: Science, v. 289, p. 248-250.

U.S. Geological Survey, 1999, The national hydrography dataset: USGS Fact Sheet 106-99, *http://erg.usgs.gov/isb/pubs/factsheets/fs10699.html*

U.S. Environmental Protection Agency, 1996, The USEPA Reach File Version 1.0 (RF1) for the Conterminous United States (CONUS): *http://www.epa.gov/waterscience/BASINS/metadata/rf1.htm*

U.S. Environmental Protection Agency, 2001, Better Assessment Science Integrating Point and Nonpoint Sources, EPA- 823-B-01-001, U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Vecchia, A.V., 2000, Water-quality trend analysis and sampling design for the Souris River, Saskatchewan, North Dakota, and Manitoba: U.S. Geological Survey Water- Resources Investigations Report 00-4019, 77 p.

Vitousek, P.M., Aber, J.D., Howarth, R.W., Likens, G.E., Matson, P.A., Schinlder, D.W., Schlesinger, W.H., and Tilman, D.G., 1997, Human alteration of the global nitrogen cycle—sources and consequences: Ecological Applications, v. 7, p. 737-750.

Vogel, R.M., 1986, The probability plot correlation coefficient test for the normal, lognormal, and gumbel distributional hypotheses: Water Resources Research, v. 22, no. 4, p. 587-590.

Vogel, R.M., Wilson, I., and Daly, C., 1999, Regional regression models of annual streamflow for the United States: Journal of Irrigation and Drainage Engineering, v. 125, no. 3, p. 148-157.

Vogelmann, J.E., Howard S.M., Yang L., Larson C.R., Wylie, B.K., and Van Driel, N., 2001, Completion of the 1990s national land cover data set for the conterminous United States from landsat thematic mapper data and ancillary data sources: Photogrammetric Engineering and Remote Sensing, v. 67, p. 650-662.

Vollenweider, R.A., 1976, Advances in defining critical loading levels for phosphorus lake eutrophication: Memorie dell'Instituto Italiano di Idrobiologia, v. 33, p. 53-83.

Wade, A.J., Durand, P., Beaujouan, V., Wessel, W.W., Raat, K.J., Whitehead, P.G., Butterfield, D., Rankinen, K., and Lepisto, A., 2002, A nitrogen model for European catchments—INCA, new model structure and equations: Hydrology and Earth System Sciences, v. 6, p. 559-582.

White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity: Econometrica, v. 48, p. 817-838.

Whitehead, P.G., Wilson, E.J., and Butterfield D., 1998, A semi-distributed nitrogen model for multiple source assessments in catchments (INCA), part I—Model structure and process equations: Sci. Total Environ, v. 210/211, p. 547-558.

Wolock, D.M., 1993, Simulating the variable-source-area concept of streamflow generation with the watershed model TOPMODEL: U.S. Geological Survey Water-Resources Investigations Report 93-4124, 33 p.

Yapo, P.O., Gupta, H.V., and Sorooshian, S., 1996, Calibration of conceptual rainfall-runoff models—sensitivity to calibration data: Journal of Hydrology, v. 181, p. 23-48.

Yoder, C.O., 1997, Important concepts and elements of an adequate state watershed monitoring and assessment program, ASIWPCA Standards and Monitoring Task Force: *prepared for* USEPA, Office of Water (Cooperative agreement CX 825484-01-0), 38 p.

Young, R.A., Onstad, C.A., Bosch, D.D., and Anderson, W.P., 1995, AGNPS user's guide, version 5.00, 1995: Morris, MN, Agricultural Research Service, U.S. Department of Agriculture.