Appendix B. Methodology for Computing Trends in Annual Streamflow, Trends in Load, and Flow- and Non-Flow-Adjusted Trends in Concentration

The non-flow adjusted trend estimates are derived from parameter estimates, and associated covariances, obtained from a model of flow and model of water-quality concentration, both models being estimated in logarithm space. The model of flow consists of an intercept, a linear trend term (decimal time), sine and cosine functions of decimal time (the seasonal component of flow), and a serially correlated error term; the flow residual is assumed to follow a 30-order autoregressive (AR(30)) process. The flow model is estimated using maximum likelihood methods as employed by the SAS AUTOREG procedure (SAS Institute, 1989). The high order autoregressive process is necessary to remove as much serial correlation as possible from the residuals, thereby reducing bias in the estimated coefficient covariance matrix of the flow model.

The water-quality model relates the natural logarithm of contaminant concentration to various predictors. An abstract representation of the model is given by

$$c_t = b_0 + m(q_t)b_q + h(T_t)b_T + x_t b_x + e_t, \qquad \text{(B1)}$$

where:

c_t is the natural logarithm of contaminant concentration in period t;

q_t is natural logarithm of streamflow;

T is decimal time, in years;

x_t is a vector of ancillary predictors such as the sine and cosine functions of decimal time;

e_t is an independent and identically normallly distributed random error;

$m(.)$ and $h(.)$ are multi-element vector functions of q and T; and

b_0, b_q, b_T and b_x are associated coefficients to be estimated.

The multi-element vector function of the natural logarithm of streamflow, $m(.)$, consists of the natural logarithm of flow and the square of the natural logarithm of flow; the multi-element vector function of decimal time, $h(.)$, consists of second-order polynomial terms and step functions of decimal time.

The water-quality model estimates values using either ordinary least squares, if the water-quality data contain no censored observations, or the maximum likelihood method, if censored observations are present. The exact methods differ slightly from those employed by LOADEST (Runkel and others, 2004) and small discrepancies can be expected in the coefficient and covariance estimates. In particular, LOADEST uses exact analytic derivatives to estimate the covariance and maximum likelihood bias adjustment, whereas the method employed in this model uses a numerical central-difference approach. The maximum likelihood bias adjustment requires estimates of the detection level even for uncensored observations. If the detection level is not supplied, a default detection level is used that equals the maximum of the parameter's median detection level for all censored observations across all stations or the reported, uncensored value.

The estimate of non-flow adjusted trend (NFAT) is based on the flow and trend coefficients from the water quality model (b_q and b_T), and the coefficient on decimal time in the flow model, subsequently denoted a. Let t_1 and t_2 define the beginning and ending dates of the trend period (in this case, the dates of the first and last water-quality sample collected during the analysis period October 1, 1992 to September 30, 2004). The trend in the natural logarithm of flow in period t, $\tilde{q}_t$, is given by

$$\tilde{q}_t = \bar{q} + a\left(T_t - \bar{T}\right), \qquad \text{(B2)}$$

where:

$\bar{q}$ and $\bar{T}$ are the averages of the natural logarithm of flow and decimal time over the trend period.

If flow is upward trending, then a will be positive and trend in the natural logarithm of flow will be less than the mean value of the natural logarithm of flow for the first one-half of the analysis period and greater than the mean value thereafter. Note that the average of the natural logarithm of flow, $\bar{q}$, implicitly accounts for the intercept and average of the seasonal terms that are included in the flow model but not otherwise apparent in the formulation of equation (B2).

The full trend in water quality in period t, $\tilde{c}_t$, is defined as

$$\tilde{c}_t = b_0 + m(\tilde{q}_t) b_q + h(T_t) b_T + \bar{x} b_x, \qquad \text{(B3)}$$

where:

$\bar{x}$ is an average, over the trend period t_1 to t_2, of the non-flow/non-trend variables in the water-quality model given in equation 1, and the average of the error term is absent because it is set to its expected value of zero.

Note that in forming this estimate of $\tilde{c}_t$, the trend in the natural logarithm of flow is substituted for the actual natural logarithm of flow in the function $m(.)$. This implies that variations in streamflow not reflected in trend do not determine the proposed measure of full trend in water quality. Because of the nonlinearity of the function $m(.)$, this might lead to a bias in the evaluation of full water-quality trend if flows are becoming more or less variable over time.

The full trend in water-quality concentration, τ_c, over the trend period, t_1 to t_2, is given by

$$\begin{aligned}\tau_c &= \tilde{c}_{t_2} - \tilde{c}_{t_1} \\ &= \left(m(\tilde{q}_{t_2}) - m(\tilde{q}_{t_1})\right) b_q + \left(h(T_{t_2}) - h(T_{t_1})\right) b_T \\ &= \left(m\left(\bar{q} + a\left(T_{t_2} - \bar{T}\right)\right) - m\left(\bar{q} + a\left(T_{t_1} - \bar{T}\right)\right)\right) b_q \\ &\quad + \left(h(T_{t_2}) - h(T_{t_1})\right) b_T .\end{aligned} \qquad \text{(B4)}$$

The full trend in water-quality concentration depends on the trend and flow coefficients from the water-quality model, b_q and b_T, as well as the trend coefficient a from the streamflow model. The trend in load, τ_L, is similarly defined but includes an additional term to reflect the direct effect streamflow has on the determination of load

$$\begin{aligned}\tau_L &= a\left(T_{t_2} - T_{t_1}\right) \\ &\quad + \left(m\left(\bar{q} + a\left(T_{t_2} - \bar{T}\right)\right) - m\left(\bar{q} + a\left(T_{t_1} - \bar{T}\right)\right)\right) b_q \\ &\quad + \left(h(T_{t_2}) - h(T_{t_1})\right) b_T .\end{aligned} \qquad \text{(B5)}$$

The full trend, expressed in percent per year, is given by

$$\text{Percent Trend} = 100 \frac{\left(\exp(\tau) - 1\right)}{T_{t_2} - T_{t_1}}, \qquad \text{(B6)}$$

where:

τ is either τ_c or τ_L.

The estimator of full trend is obtained by evaluating the coefficients a, b_q and b_T in (A4) and (A5) at their estimated values, where the water-quality parameter estimates are obtained using the adjusted maximum likelihood method (Cohn, 2005). Generally, the transformation of the estimated values of τ_c or τ_L into real space, via the exponential function appearing in (A6), induces a small degree of bias. No correction of this bias was applied in this case to ensure agreement between the sign of the percent trend and the sign of the test statistic used to evaluate the statistical significance of trend

A statistical test of the null hypothesis that trend is zero can be undertaken by forming t-statistics from the estimates of either τ_c or τ_L, and dividing by their respective standard error. Standard errors of the estimated values of τ_c and τ_L are complicated to derive owing to the nonlinear manner in which the flow trend coefficient and the water-quality flow coefficients interact in the determination of full trend. An approximation to the standard error, which is suitable for large samples, is obtained by taking a first-order Taylor approximation of the full trend estimate (either equation (B4) or (B5)) with respect to the flow and water-quality model coefficients. Let $b = \{b'_q \; b'_r\}$ represent the vector of combined flow and trend coefficients from the water-quality model and let V_b be the covariance matrix of this vector. Under the plausible assumption that streamflow is exogenous with respect to water quality, meaning that changes in streamflow cause changes in water quality but changes in water quality do not cause changes in streamflow, the covariance between the estimated values of a and b is zero. Consequently, the standard error of τ_c and τ_L, denoted σ_c and σ_L, are given by

$$\sigma_c = \sqrt{V_a \left(\frac{\partial \Delta m}{\partial a} b_q\right)^2 + A V_b A'}, \qquad \text{(B7)}$$

and

$$\sigma_L = \sqrt{V_a \left(\frac{\partial \Delta m}{\partial a} b_q + \left(T_{t_2} - T_{t_1}\right)\right)^2 + A V_b A'}, \qquad \text{(B8)}$$

where

V_a is the variance of the estimated flow trend coefficient, a,

$$\frac{\partial \Delta m}{\partial a} = \frac{\partial m\left(\bar{q} + a\left(T_{t_2} - \bar{T}\right)\right)}{\partial a} - \frac{\partial m\left(\bar{q} + a\left(T_{t_1} - \bar{T}\right)\right)}{\partial a} \qquad \text{(B9)}$$

and

$$\begin{aligned}A &= \left\{ m\left(\bar{q} + a\left(T_{t_2} - \bar{T}\right)\right) \quad h\left(T_{t_2}\right) \right\} \\ &\quad - \left\{ m\left(\bar{q} + a\left(T_{t_1} - \bar{T}\right)\right) \quad h\left(T_{t_1}\right) \right\}\end{aligned} \qquad \text{(B10)}$$

In large samples, the t-statistics τ_c/σ_c and τ_c/σ_c are distributed standard normal. Therefore, the two-sided p-value for significance of trend is given by

$$p = 2\left(1-\Phi\left(\left|\tau/\sigma\right|\right)\right), \tag{B11}$$

where:

τ/σ is either τ_c/σ_c or τ_L/σ_L, and

$\Phi(.)$ is the standard normal cumulative distribution.

The unit trend is the change in concentration or load over a given period, expressed in the units of concentration or load, divided by the length of the period, expressed in years. Both the unit trend in concentration and the unit trend in load are based on trend-denominated variations in the logarithm-transformed water-quality and flow models given in equations (B1) and (B2). The unit trend in concentration, denoted τ_c^u, is given by

$$\tau_c^u = \frac{\exp\left(\tilde{c}_{t_2}\right)-\exp\left(\tilde{c}_{t_1}\right)}{T_{t_2}-T_{t_1}} = \exp\left(\tilde{c}_{t_1}\right)\left(\frac{\exp\left(\tau_c\right)-1}{T_{t_2}-T_{t_1}}\right), \tag{B12}$$

where

$\tilde{c}_t$ is given by equation (A3), and

τ_c is given by equation (A4).

The units of τ_c^u are milligrams per liter per year (mg L^{-1} y^{-1}). A similar expression applies to the determination of the unit trend in load, τ_L^u,

$$\tau_L^u = k\exp\left(\tilde{c}_{t_1}+\tilde{q}_{t_1}\right)\left(\frac{\exp\left(\tau_L\right)-1}{T_{t_2}-T_{t_1}}\right), \tag{B13}$$

where

k is a conversion factor that expresses τ_L^u in units of kilograms per year per year (kg y^{-2}), and

$\tilde{q}_{t_1}$ is given by equation (A2), evaluated at the date of the first water-quality sample in the analysis period.

The estimators of τ_c^u and τ_L^u, given by (B12) and (B13), are based on the sample estimates of τ_c and τ_L described above, and on estimates of $\tilde{c}_{t_1}$ and $\tilde{q}_{t_1}$, given in equations (B3) and (B2), with coefficients evaluated at their sample estimates (see above), and with the period set to t_1, the date of the first water-quality sample in the analysis period. A correction was applied to account for the transformation bias arising from the exponential transformation of $\tilde{c}_{t_1}$ in equation (B12) and $\tilde{c}_{t_1}+\tilde{q}_{t_1}$ in (B13). The correction was applied only to these terms, and not to the terms involving τ_c and τ_L, because correction of the latter could affect the sign of the unit trend estimate as compared to the sign of the test statistic used to statistically evaluate the significance of trend. The transformation bias correction is based on a first-order Taylor approximation that linearizes the expressions for $\tilde{c}_{t_1}$ and $\tilde{c}_{t_1}+\tilde{q}_{t_1}$ with respect to all model coefficients. The estimation of the variance of $\tilde{c}_{t_1}$ and $\tilde{c}_{t_1}+\tilde{q}_{t_1}$ then follows by evaluating the quadratic form implied by this linearization, similar to the quadratic forms appearing inside the square-root radicals in equations (B7) and (B8) above, employing the estimated variance of a and the covariance matrix of the estimated values of b_0, b_q, b_T and b_x. Let the estimated variances of $\tilde{c}_{t_1}$ and $\tilde{c}_{t_1}+\tilde{q}_{t_1}$ be denoted $V\left(\tilde{c}_{t_1}\right)$ and $V\left(\tilde{c}_{t_1}\right)$. Under the assumption that the estimated coefficients are approximately normally distributed, as occurs in large samples, the transformation bias correction is obtained by multiplying the expressions for τ_c^u and τ_L^u by the respective factors $\exp\left(-V\left(\tilde{c}_{t_1}\right)/2\right)$ and $\exp\left(-V\left(\tilde{c}_{t_1}+\tilde{q}_{t_1}\right)/2\right)$.

References Cited

Cohn, T.A., 2005, Estimating contaminant loads in rivers—An application of adjusted maximum likelihood to type 1 censored data: Water Resources Research, v. 41, no. 7, W07003.

SAS Institute, Inc., 1989, SAS version 6, original copyright, 1989: SAS Institute, Inc., Cary, N.C.

Runkel, 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 Book 4, Chap. A5, 69 p.

This page intentionally left blank.