1.3.1.5 Guidance on station and record selection

To provide guidance in determining criteria for selecting monitoring stations and specifying an acceptable record length, a simple analysis was performed using data from the USGS's National Stream Quality Accounting Network (NASQAN). In the analysis, long-term mean annual total nitrogen and total phosphorus flux was estimated for approximately 700 monitoring stations across the nation using water-quality and flow data available from Alexander and others (2000). Flux estimates were based on a simple model in which the logarithm of concentration is regressed on the logarithm of contemporaneous flow, a time trend, and a sine and cosine term. The method of adjusted maximum likelihood (Cohn, Gilroy, and others, 1992) was used to accommodate stations with censored data. All available flow and water-quality data between the period 1970 and 2000 were used at each station. As a result of NASQAN program changes, however, the number of observations and exact periods of record varied considerably across stations. An estimate of flux was made only if a station had at least 15 water-quality observations. The mean flux estimates were detrended if the water-quality period of record included the base date June 30, 1992. A standard error was computed for each flux estimate using the methods described above.

Table 1.3 presents the results from a simple analysis in which the logarithm of the standard error of long-term flux is regressed on various features of the basin in which the estimate is made and characteristics of the sample used to estimate the station-specific flux model. The reported coefficient estimates can be interpreted as the change in percent standard error of the mean flux estimate from a one-percent change in the untransformed explanatory variable. The table shows considerable agreement between the determinants of flux precision in the two models. The estimate of mean flux, either total nitrogen (TN) or total phosphorus (TP), is more precise in basins with larger mean flow, smaller variability of flow and, not unexpectedly, smaller unexplained error (RMSE). In terms of sample characteristics, holding features of the basin fixed, the estimate of TN or TP flux is more precise as more days are used to predict mean flux, as more observations are used to estimate the water quality model, the higher the mean flow and the larger the variation in flow on days in which water quality is sampled, and the smaller the number of days between water-quality samples. Samples with more censored observations tend to correspond to reduced precision of the mean flux estimate, but this effect is not statistically significant. The finding that a larger variation in streamflow on water-quality sampling days causes improved precision in flux estimates is due to reduction in the error of the flux model coefficient estimates caused by increased variability in the streamflow predictor.

Table 1.3. Factors affecting the percent standard error of total nitrogen (TN) and total phosphorus (TP) flux estimates at 700 National Stream Quality Accounting Network monitoring stations.

[POR refers to period of record, and RMSE refers to the root mean square error of the station flux model]

	TN	TP
Intercept	-0.068	0.578***
Log mean flow POR	-0.148***	-0.149**
Log std flow POR	0.336***	0.387***
Log No. predict days	-0.042*	-0.044*
Log No. WQ obs	-0.319***	-0.378***
Log mean flow for WQ samples	-0.080	-0.132*
Log std flow for WQ samples	-0.126***	-0.119**
Log median days betwn obs	0.152***	0.073**
No. WQ censored obs	0.130	0.000
Log RMSE	1.143***	1.337***
No. obs	700	693
R^2	0.854	0.836

* Significant at 0.05
** Significant at 0.01
*** Significant at 0.001

1.3.1.6 Guidance for specifying monitoring station flux models

The principal source of statistical inference in SPARROW applications comes from having a large set of water-quality stations covering a wide range of values of the predictor variables. In this context, the need for highly accurate flux estimates is of secondary concern. A simple model of water quality may be preferred to a more complex model, particularly if the complex model has the capability of producing unacceptable estimates of flux. For example, models that include quadratic terms for flow or time may provide extremely poor estimates if the models must extrapolate to conditions that are significantly different from those exhibited by the sample. Unless the researcher has considerable time and patience to examine each of the station estimates in detail, it is suggested that the specified model err towards being too simple. This reasoning argues for models that include only a few variables, such as the logarithm of flow, a time trend, and one or two seasonal harmonics. Quadratic terms should be included only with great care. Even seasonal harmonics can be a problem if the sample does not include observations from all seasons.

The ideal water-quality record is one that has sufficient observations to reliably estimate the coefficients in the model. Models that include both flow and time require a fairly long record, one that includes the base date for detrending, and contains observations covering a wide range of hydrologic conditions. One consideration that places limits on the length of the record is the need for flux estimates to be representative of base-year conditions. By this, it is not meant that the record must be representative of the hydrologic conditions in the base year: the SPARROW model is estimated using mean estimates of water quality with the intention of removing variations due to hydrologic conditions. Rather, what is meant is that the long-term patterns in water quality should be adequately captured by the model specification. Given the argument above for a simple model, this implies the period of record should not be so long as to violate the assumption that water-quality trend is reasonably approximated by a simple linear function of time. It is, however, generally true that longer water-quality records display more complex patterns of trend. Too long a record, therefore, runs the risk of misrepresenting the trend in water quality for any given year, adding noise to the detrended flux estimate. For national analyses, Smith and others (1997) and Alexander and others (2000) have based mean flux on water-quality data that span a 15 to 20 year period. Regional analyses by Preston and Brakebill (1999) and Moore and others (2004) have used water-quality data that span 20 to 25 years. Shorter periods of about six years were used by Alexander, Elliott, and others (2002). Shorter records may certainly be used to obtain an estimate of mean flux, although a shorter record will generally result in a larger standard error in the mean flux estimate. For very short records, however, the specification of trend becomes unreliable. It is recommended, therefore, that trend terms be excluded from the model if the water-quality record spans less than three years.

One problem with including trend in an analysis is the potential for bias in applying the estimated trend to predict periods beyond the beginning and ending dates of the sample record. If the results are to be detrended, the potential for bias arises if the base period is not within the beginning and ending dates of the sample record. Conversely, if the base date is within the sample record, the potential for bias is not a concern: any trend that would be applied to periods outside the sample period is ultimately removed through the process of detrending. Consequently, it is recommended that the trend term be suppressed, and no detrending of the data be undertaken, if the base date is more than a year beyond the upper and lower bounds of the sample period. If the results are not detrended, it is recommended that the trend term be suppressed if the prediction period extends beyond 50 percent of the sample period.

As discussed in section 1.3.1.3, the detrending of water-quality data requires a coincident detrending of streamflow data via the estimation of a streamflow model, and it is possible that only one of the two models, either the flow model or the water-quality model, meets the criteria for inclusion of trend. In this case, the trend in the model that does not meet the criteria is suppressed and the trend in the other model is retained. Thus, it is possible—actually common—for the water-quality flux to be partially detrended; the flow component of the model is detrended because the flow record meets the criteria for trend estimation, whereas the trend term in the water-quality model is removed. In this case, the partially detrended water-quality series will exhibit no trend because all estimated trend terms influencing the series are detrended.

As explained at the end of section 1.3.1, if flux estimates are detrended there is little risk from extrapolating a water-quality model to a prediction period that greatly exceeds the water-quality sample period. In order to adequately account for hydrologic variability, this argues for having a prediction period that is as long as the flow record would allow. Another point made above, however, is that the period should not be so long as to invalidate the specification of the trend function. Therefore, because trends in flow are not generally as large as trends in water quality, a reasonable upper limit for prediction is 30 years. This is, as it happens, a

practical limit as well because the time required to estimate the standard error of mean flux grows with the square of the length of the prediction period. A prediction period greater than 30 years causes significant increases in the computer time required to estimate this standard error.

As was demonstrated in table 1.3 in section 1.3.1.5, one of the principal determinants of accuracy in flux estimation is the number of observations input to the water-quality model. Generally, it is recommended that an estimate of flux depend on no fewer than 15 uncensored water-quality observations. However, it may be necessary to deviate from this threshold in some cases; situations in which a site is subject to high variability in streamflow, the other principal determinant of accuracy in table 1.3, dictate that more water-quality observations are required to obtain acceptable error in water-quality estimation.

Ultimately, the final consideration for inclusion of data for a station in the SPARROW analysis is that the standard error of the mean flux estimate not be too large. SPARROW has the capability of weighting observations according to their standard errors; it would appear, therefore, that wholesale exclusion of high-error flux estimates is unnecessary. If, however, a high-error flux estimate enters the model as an upstream source for some downstream flux observation, the so-called nested basin arrangement, the potential for biasing estimates is increased. High-error flux estimates also present a problem in assessing the error of SPARROW predictions if the predictions are conditioned on observed, upstream flux estimates. In national analyses, we have frequently excluded mean flux estimates that have a standard error greater than about 20 to 30 percent.

1.3.1.7 Related topics

There are applications of the SPARROW model that can be envisioned in which the subject of the analysis is mean flux for a given time interval of the year, say a particular month or season (for example, mean flux for spring). The mass balance assumption for these analyses may be reasonable if the time required for the majority of the contaminants to move through the hydrologic system is considerably less than the time interval being analyzed. The dependent variable in these situations is the long-term mean flux pertaining to the given interval. This variable can be estimated using the same methods described above with the modification that only prediction days corresponding to the specified month or season are included in the determination of average flux. Similarly, a model of flux during certain hydrologic conditions, such as low flow, might also justify mass balance. The dependent variable for these analyses would be the average of daily flux estimates for days in which flow meets the specified condition.

Another potential application of the SPARROW model is to assess the determinants of changes in flux between two periods. In this case, one can contemplate the construction of two SPARROW models, one for each of the two periods, with each model relating time-specific sources to temporal estimates of mean flux. The mean flux concept is retained in this analysis by estimating a single set of station-specific models of flux over all time, but using the trend terms of these models to detrend the estimates to the two separate periods of the analysis. Complications in this kind of analysis could arise if explanatory variables of a fundamentally different character are used in the two periods; for example, if the distribution of land-use category estimates for each period is obtained from different studies employing inconsistent methodologies.

The methods described above for detrending flux could (and should) also be applied to any source variables that depend on hydrologic conditions and have sufficient time series data to support detrending. One example of such a source variable is wet atmospheric deposition, which is variable from year to year due to climatic changes in precipitation. To remove some of the noise in this variable, it is advisable that the data be detrended and a long-term mean estimate normalized to the base year of the SPARROW model be used.

1.3.1.8 Mean flow-weighted concentration (advanced)

The SPARROW model can be used to compute an estimate of mean, detrended, flow-weighted concentration. Let c_t^*, f_t^* and q_t^* be detrended concentration, water-quality flux and streamflow in period *t*. The mean, detrended, flow-weighted concentration, defined over the period $t = 1,\ldots,T$, is given by

$$\overline{c}_q^* = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{q_t^*}{T^{-1}\sum_{s=1}^{T} q_s^*}\right)c_t^* = \frac{T^{-1}\sum_{t=1}^{T} f_t^*}{T^{-1}\sum_{t=1}^{T} q_t^*} = \frac{\overline{f}^*}{\overline{q}^*}, \tag{1.21}$$

which is simply the ratio of mean detrended flux to mean detrended streamflow. The estimation of mean, detrended, flow-weighted concentration can be obtained by developing two SPARROW models, one model for water-quality flux and a second model for streamflow. The mean, detrended, flow-weighted concentration at reach *i* is simply the ratio of the reach *i* SPARROW estimates of mean water-quality flux and streamflow.

1.3.1.9 Mean time-weighted concentration (advanced)

An estimate of mean, detrended, time-weighted concentration can be made by combining mean detrended flux estimates from SPARROW with information on the covariance between the logarithms of concentration and streamflow. The estimate requires an assumption that water-quality concentration and streamflow are log-normally distributed. Under this assumption, the limit of $\overline{c}_q^*$ given in equation (1.21) as T goes to infinity is

$$\begin{aligned}\lim_{T\to\infty}\overline{c}_q^* &= \frac{\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} c_t^* q_t^*}{\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} q_t^*} = \frac{E\left(c_t^* q_t^*\right)}{E\left(q_t^*\right)} \\ &= \frac{\exp\left(E\left(\tilde{c}^*\right)+E\left(\tilde{q}^*\right)+\frac{1}{2}\left(V\left(\tilde{c}^*\right)+V\left(\tilde{q}^*\right)+2Cov\left(\tilde{c}^*,\tilde{q}^*\right)\right)\right)}{\exp\left(E\left(\tilde{q}^*\right)+\frac{1}{2}V\left(\tilde{q}^*\right)\right)} \\ &= \exp\left(E\left(\tilde{c}^*\right)+\tfrac{1}{2}V\left(\tilde{c}^*\right)+Cov\left(\tilde{c}^*,\tilde{q}^*\right)\right),\end{aligned} \tag{1.22}$$

where $\tilde{c}_t^*$ and $\tilde{q}_t^*$ are the natural logarithms of detrended concentration and streamflow in period *t*.

The limit of mean, detrended, time-weighted concentration is given by

$$\lim_{T\to\infty}\overline{c}^* = \lim_{T\to\infty} T^{-1}\sum_{t=0}^{T-1} c_t^* = \exp\left(E\left(\tilde{c}^*\right)+\tfrac{1}{2}V\left(\tilde{c}^*\right)\right). \tag{1.23}$$

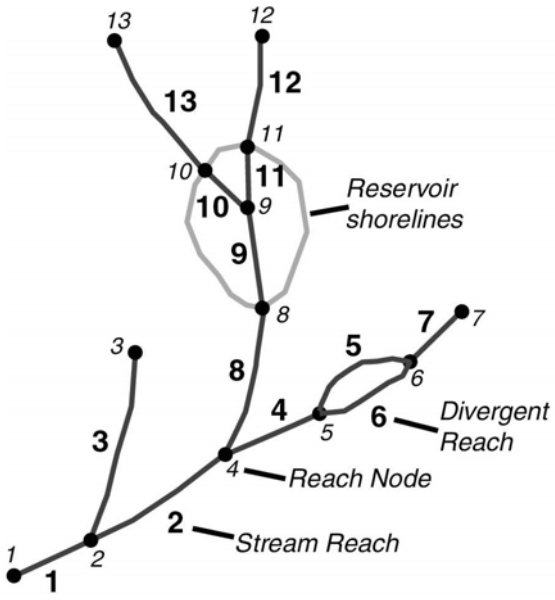
From equations (1.22) and (1.23), it is apparent that if the covariance between the logarithms of detrended concentration and streamflow is known, a consistent estimator of mean, detrended, time-weighted concentration is

$$\hat{\overline{c}}^* = \overline{c}_q^* \exp\left(-Cov\left(\tilde{c}^*,\tilde{q}^*\right)\right). \tag{1.24}$$

The utility of this relation obviously depends on the availability of estimates of the covariance term. For a SPARROW model, this requires an estimate of covariance for every reach. One approach is to obtain estimates of covariance at each monitoring station using the water-quality and streamflow sample records. The estimate of covariance for any reach could then be obtained by taking the covariance estimate from the station for which basin characteristics are closest to those of the reach. The viability of this approach is a subject of future research.

1.3.2 Stream network topology

A vector- or raster-based digital representation of the stream and river network topology is the most fundamental component of the spatial infrastructure that supports the SPARROW model (fig. 1.8). Vector-type representations are based on point and line (i.e., arc) geographic features, whereas raster representations are based on a cell (i.e., areal) structure. Whether based on vector or raster-cell topology, a stream reach network explicitly defines surface-water flow paths that spatially connect contaminant sources and landscape features with observations of water quality at downstream monitoring stations. In a vector-reach topology, a stream reach represents the length of stream channel that extends from one tributary junction to another. Reach nodes are point features that are associated with the location of tributary junctions (i.e., the ends of a reach). Reach nodes may also occur (and, in fact, are preferred) at the locations where reaches overlay with the shorelines of impoundments (reservoirs, lakes) and stream water-quality monitoring sites. In the case of a raster representation of streams, nodes may also define various intermediate locations along the stream reach between tributary junctions.



Reach Node Attribute Table

Reach Number	Up-stream Node	Down-stream Node	Diversion Fraction	Reach-Type Indicator
13	13	10	1.0	0
10	10	9	1.0	1
12	12	11	1.0	0
11	11	9	1.0	0
9	9	8	1.0	2
8	8	4	1.0	0
7	7	6	1.0	0
5	6	5	0.7	0
6	6	5	0.3	0
4	5	4	1.0	0
2	4	2	1.0	0
3	3	2	1.0	0
1	2	1	1.0	0

Figure 1.8. Schematic illustrating a vector stream reach network with node topology and water/contaminant reach-node routing table. The reach-type indicator has possible values of "0" (stream reach), "1"(impoundment reach), and "2" outlet reach for impoundment.

Whether the stream reach network is digitally represented in vector or raster form, the reach topology must define either a set of reach nodes or raster cells that are linked hydrologically to indicate the direction of water flow. The node topology must be defined according to an upstream (from-node) and downstream (to-node) node attribute table (fig. 1.8). This tabular listing of the surface-water flow paths is required for the routing of water and contaminants through the river network by SPARROW navigation software during estimation and application of the model. Thus, the reach-node table defines the fundamental data infrastructure of the model.

The SPARROW model structure supports the presence of distributary reaches or reach diversions in the stream network; these may include braided channels or reaches where water is diverted to canals or other waterbodies. The SPARROW model assumes that contaminants are diverted in proportion to water flow. Therefore, an estimate is required for each reach of the "diversion fraction"—a measure of the fraction of the stream flow that is diverted in distributary reaches. A diversion fraction of 1.0 is assigned to reaches without water diversions. In the example distributary reaches shown in figure 1.8, the fraction of the flow in reach 7 that is diverted to the two downstream reaches is 0.7 for reach 5 and 0.3 for reach 6. The diversion fraction is a key

reach-level metric that is needed to make accurate computations of the downstream routing of water and mass, and is directly incorporated into the SPARROW mass-balance equation [see equation (1.27)].

The reach topology and SPARROW model also support the separate designation of impoundments (e.g., lakes, reservoirs) that are associated with stream reaches. This designation is used in SPARROW to separately estimate the contaminant attenuation in reservoirs and lakes. In the example in figure 1.8, reaches 9-11 are associated with a reservoir. A "reach-type" indicator is used to identify reaches associated with impoundments separately from conventional stream reaches (see reach type in figure 1.8). The outlet reach of an impoundment is coded separately from other interior impoundment reaches to facilitate the pollutant attenuation calculations.

The complete listing of the ancillary properties of the stream reaches that are required for SPARROW modeling is given in table 2.1 of the user's guide in Part 2 of this document. In addition to the properties listed in the table in figure 1.8, the reach length and an estimate of the mean-annual streamflow of the reach is also required. Estimates of mean water velocity are required to estimate in-stream contaminant attenuation as a function of the water time of travel (alternatively, in-stream attenuation can be estimated as function the reach length; e.g., see Alexander, Elliott, and others, 2002). Measures of the areal water load are needed for impoundments (i.e., lakes, reservoirs) to use in estimating the contaminant attenuation in these water bodies; the areal water load is computed as the quotient of the outflow to surface area of the water body (see Alexander, Elliott, and others, 2002). Digital representations of the drainage basin boundaries associated with river reaches are also needed to estimate the incremental and total drainage area of the reaches and to support the digital overlay of drainage boundaries with polygonal boundaries (vector or raster) that define the locations of contaminant sources (point and diffuse) and various landscape properties.

The assessment of pollutant loadings to coastal estuaries requires an expanded reach network that includes shoreline features and the identification of reaches that terminate at estuaries. The node points of shoreline reaches include the downstream nodes of the terminating reaches. Shoreline reaches are used to define coastal drainage areas—areas that discharge runoff directly to the estuary without transport through a stream reach. Because the discharge for a shoreline reach does not accumulate from any upstream location, the diversion fraction for these features is set to zero. Shoreline reaches also have no stream attenuation so traveltime along the reach is set to zero.

Examples of vector-based stream networks determined from digital line graph (DLG) hydrography include the 1:500,000 scale River Reach File 1 (RF1) hydrography developed by the USEPA (U.S. Environmental Protection Agency, 1996) and the 1:100,000 scale National Hydrologic Dataset (NHD; U.S. Geological Survey, 1999). An enhanced RF1 (ERF1) file was previously developed to support national SPARROW modeling in the conterminous U.S. (ERF1 version 2.0, see Nolan and others, 2002; ERF1 version 1.2, Alexander and others, 1999). An enhanced NHD network was recently developed to assist with SPARROW modeling in the northeastern U.S. (Moore and others, 2004). A comparison of the two reach networks is given in figure 1.9 for the northeastern U.S. ERF1 includes approximately 62,000 river reaches covering more than 1 million kilometers of streams in the conterminous U.S. The median value for reach drainage area is 67 kilometers^2 and the mean-annual streamflow is 62 feet^3 second^{-1}. Drainage basin boundaries and contributing areas were determined from digital elevation models (DEMs), based on the HYDRO 1K (*http://edcdaac.usgs.gov/gtopo30/hydro/namerica.asp*) using conventional geographic information system (GIS) methods (Nolan and others, 2002; see also Brakebill and Preston, 1999).

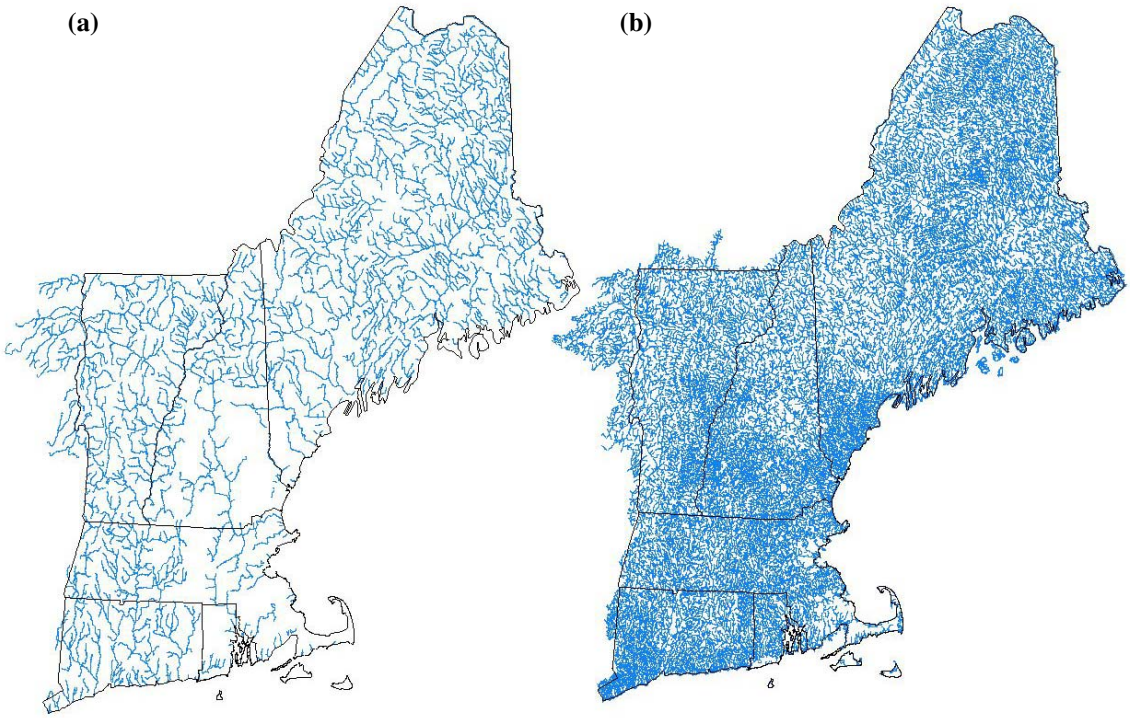


Figure 1.9. Digital river reach networks for the northeastern United States, based on (a) the 1:500,000 scale River Reach File 1 (RF1), and (b) the 1:100,000 scale National Hydrologic Dataset (NHD). The number of stream reaches shown in the maps is 2,462 and 42,000 for RF1 and NHD, respectively. [Image from R. Moore, U.S. Geological Survey, written comm., 2005.]

The task of *"spatial referencing"* is a critical step in SPARROW modeling that is necessary to construct the watershed attribute data used as explanatory variables in the models and to verify the accuracy of the hydrologic connectivity of the stream reaches (which governs routing and accumulation of mass in the model). Spatial referencing entails the use of GIS techniques (e.g., point-arc or polygon-polygon intersections) to digitally establish the geographic relation between stream reaches in the river network and the various watershed attributes that are used to develop and estimate the model. Watershed attribute data sets (e.g., land use, climate, fertilizer use) are frequently compiled and reported according to areal geographic units that are not coincident with the drainage basin boundaries of river reaches; they may instead be collected at finer spatial scales or according to spatial units that reflect cultural (e.g., counties, states) or other non-hydrologic features of the landscape (e.g., interpolated contour intervals for atmospheric deposition). Watershed attributes may also be geographically located according to precise locations coded according to latitude and longitude, which must be digitally linked to a watershed or nearby river reach. For example, this may include the locations of stream monitoring gages or municipal wastewater treatment outfalls. Linking point locations to river reaches requires the use of GIS operations that intersect points and arcs (i.e., when using a vector-based river network). Wastewater facilities and stream monitoring gages can be spatially referenced to stream reaches by using GIS operations, such as ARC NEAR, to link the latitude and longitude of the facility or gage with a nearby river reach. Additional verification of these automated GIS operations is often required to ensure accurate spatial referencing; this includes, for example, comparisons of the river name of a gage with that of a river reach or the comparison of the reported drainage area of a gage with that estimated for a river reach.

Verification of the connectivity of river reaches and nodes commonly entails the use of GIS routing software that allows a user to pick a reach or grid location and trace upstream or downstream within the network to visually determine that all reaches and grid cells are connected. Once node connections have been visually evaluated and verified, a check for proper node connectivity in the drainage basins of selected reaches may also be executed by comparing the known drainage area at monitoring sites or other locations with the total drainage area of their associated reaches. The executable program—"assign_hydseq.exe" (described in Appendix B)—

can be used to compute the total drainage area above the reaches in a stream network (this assumes that the drainage basin area of individual reaches has been determined in a prior step). Drainage area comparisons can be made for reaches where monitoring stations have been spatially referenced to the reach network. The computer program "assign_hydseq.exe" can also be used to derive the incremental drainage area for multiple reaches that lie between user-selected reach locations, such as the incremental drainage areas associated with a network of monitoring sites.

Finally, although not discussed in this document, the SPARROW node and routing architecture also can fully support the modeling of contaminant transport along "off-reach" (i.e., "landscape") flow paths according to flow directions defined by landscape topography as reflected, for example, in digital elevation models (DEMs). Such architecture could facilitate the incorporation of high-resolution spatial data sets that delineate sources and other explanatory variables at scales finer than an incremental drainage area for the reach (see the discussion of a raster-based model of land processes in section 1.4.3).

1.3.3 Watershed sources and explanatory variables

The specific explanatory variables evaluated in SPARROW models should reflect current knowledge of natural and human-related sources and the important physical, chemical, or biological properties of the terrestrial and aquatic ecosystems that affect the supply and transport of contaminants in watersheds in combination with practical considerations of the types of digital data that are available to the user. Point- and diffuse-source variables may include direct measures of the introduction or supply of contaminant mass to the landscape and streams and reservoirs (e.g., municipal and industrial wastewater effluent, atmospheric nitrogen deposition, fertilizer application, nitrogen fixation in terrestrial ecosystems). Alternatively, source variables may be selected to serve as surrogate indicators of the contaminant mass supplied by point and diffuse sources in watersheds, such as land-use/land-cover data or census data on human and livestock populations. More discussion of the specific types of explanatory variables for sources and landscape properties that can be used in SPARROW models is given in subsequent sections on model specifications (see section 1.4.1).

To develop the inputs of explanatory variables for SPARROW water-quality models, users are required to spatially reference watershed data on sources and other properties (e.g., climate, topography, land-use) to the drainage basins and stream reaches of the SPARROW river network as described in the previous section. Figure 1.10 shows the relation between areas of the source polygons and the incremental drainage basin of a hypothetical stream reach. Estimates of the diffuse sources associated with the drainage basins of stream reaches can be obtained using GIS operations (e.g., polygon-polygon intersections) to digitally overlay the drainage basin boundaries of stream reaches with the polygonal areas associated with the diffuse source data. Once this GIS procedure is used to obtain quantitative measures of the overlap in watershed areas (fig. 1.10), estimates of the area-weighted sum of source characteristics ($S_{n,j}$) for reach j and source type n can be calculated as

$$S_{n,j} = \sum_{k \in P(j)} S_{n,k}(A_{j,k} / A_k^*), \tag{1.25}$$

where $P(j)$ is the set of all source-related polygons that intersect the incremental drainage polygon for stream reach j, $S_{n,k}$ is the quantity of source-type n associated with polygon k, $A_{j,k}$ is the sub-area of reach j's incremental drainage that intersects the source-related polygon k, and A_k^* is the total area of the source-related polygon k.

If it is known that a particular contaminant source is associated with a specific land use (or some grouping of land uses)—for example, fertilizer is associated with cultivated land—and if the spatial scale of land-use information is similar to the scale at which watersheds are delineated, then the method described above can be modified to obtain a more refined estimate of sources within a reach incremental drainage area. The enhanced method requires only that the area term $A_{j,k}$ in equation (1.25) be redefined to represent the area of the associated land use that intersects the reach j incremental drainage and source polygon k, and that the area term A_k^* be redefined to equal the total area of the associated land use within the source polygon k. Otherwise, the evaluation of the source area-weighted sum is identical to equation (1.25).

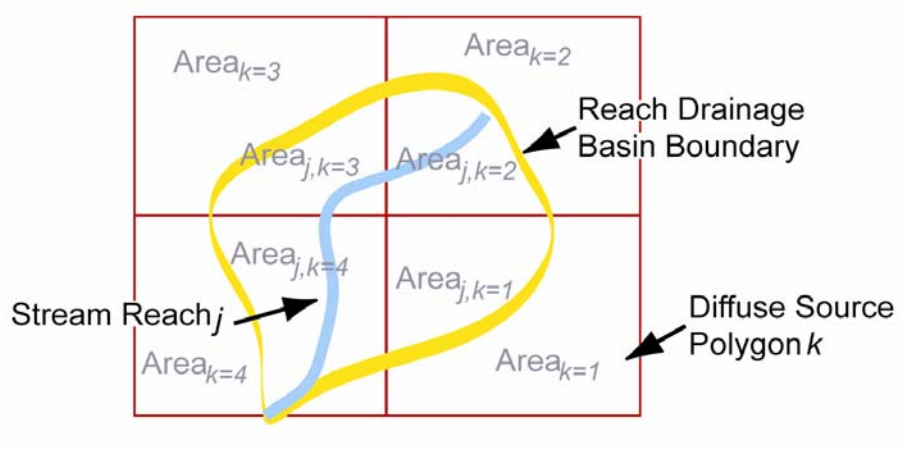


Figure 1.10. Schematic illustrating digital overlay of stream reach drainage area and polygonal areas associated with diffuse sources.

Climatic and landscape properties that affect contaminant transport may include measures of water-balance terms (e.g., solar radiation, precipitation, evaporation, evapotranspiration), soil characteristics (e.g., permeability, moisture content), water-flow path properties (e.g., slope, topographic index), or management practices and activities (e.g., tile drains, conservation tillage, Best Management Practices (BMPs)). Estimates of various climatic and landscape characteristics or properties are often calculated as an area-weighted mean estimate ($Z_{i,j}$) for stream reach *j* and landscape property *i* according to

$$Z_{i,j} = \sum_{k \in P(j)} \tilde{Z}_{i,k} (A_{j,k} / A_j^*) , \tag{1.26}$$

where *P(j)* is the set of all land-characteristics polygons that intersect the incremental drainage polygon for stream reach *j*, $\tilde{Z}_{i,k}$ is the landscape property *i* associated with polygon *k*, $A_{k,i}$ is the sub-area of reach *j*'s incremental drainage that intersects the landscape property's polygon *k*, and A_j^* is the total drainage area of the reach watershed *j*. Figure 1.11 below shows the relation between areas of landscape polygons and the incremental drainage basin of a hypothetical stream reach that is described in equation (1.26). Note that the area-weighted mean estimate defined in equation (1.26) differs from the area-weighted sum given by equation (1.25) in that the area ratio terms sum to 1.0 in equation (1.26) but not in equation (1.25).

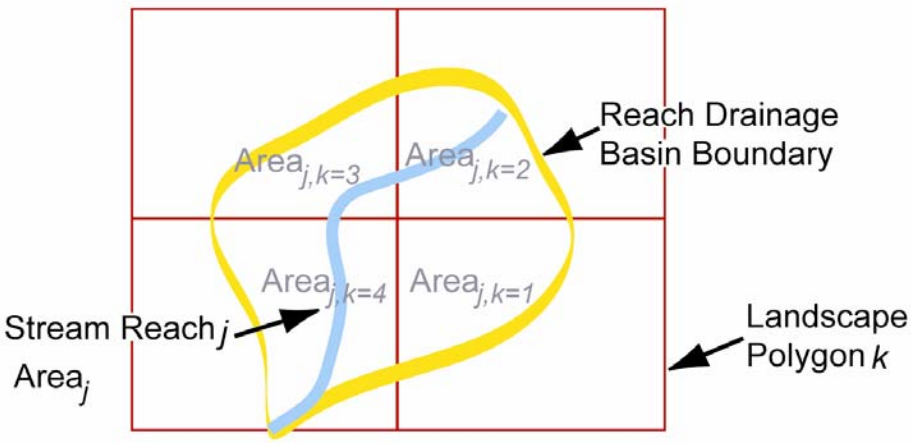


Figure 1.11. Schematic illustrating digital overlay of stream reach drainage area and polygonal areas associated with landscape properties.

The basic SPARROW models described in this report are designed to model long-term mean contaminant loadings in streams (see the next section 1.3.1 for details), implying explanatory variables in the model should be computed to reflect long-term conditions. Variables that describe contaminant sources and landscape characteristics may be averaged over multiple years, corresponding to the available period of record for water-quality monitoring data, or may reflect the conditions during a specified base year used to estimate stream water-quality loadings. The discussion in section 1.3.1 describes additional considerations for obtaining multi-year averages, including detrending the series to a base period.

1.4 Model specification

The specification of a SPARROW model consists of identifying the explanatory variables and functional forms for the associated processes the model is to include. The subsequent sections describe model specification in generic terms, followed by specific suggestions for variables and processes that have been used in previous SPARROW applications.

1.4.1 Model equation and specification of terms

Conceptually, the contaminant load or flux leaving a reach is the sum of two components:

Load leaving the reach	=	Load generated within upstream reaches and transported to the reach via the stream network	+	Load originating within the reach's incremental watershed and delivered to the reach segment

The first component represents the load that is delivered to the reach from upstream reaches via the stream network. Losses of flux from the stream network may occur at points where flow is diverted; additionally, in moving through the reach, flux will generally be attenuated by stream or reservoir processes. The second component consists of source flux that is generated within the reach's incremental watershed and first delivered to the stream network somewhere along the reach segment. A number of source-dependent processes, in addition to stream attenuation processes, affect the amount of source flux reaching the stream network and transported to the reach's downstream outlet node. For flux originating on the landscape, the processes affecting delivery to the stream network are called land-to-water delivery processes, and may include both surface and sub-surface elements.

A conceptual illustration of the pertinent spatial relations is given in figure 1.12. A connecting reach is generally defined as a stream segment that connects the confluence of two stream segments; a headwater reach is a reach that is defined without an upstream confluence. Figure 1.12 shows five complete reaches, two of which are headwater reaches, with one of the headwater reaches classified as a reservoir. Each reach is embedded within a color-coded area representing the reach's incremental drainage—the area that drains directly to the reach without passing through another reach. Because there are five reaches, there are five incremental drainage areas. Land-to-water delivery processes determine the amount of contaminant generated within an incremental drainage area, excluding contaminant generated directly on a reach, which is then delivered to the area's corresponding reach. In the figure, two monitoring stations, X and Y, form the boundaries of a nested basin, defined as all reaches above a monitored reach *i*, containing monitoring station X, but exclusive of reaches above and including all upstream monitoring stations—in this case the single monitoring station Y. Although not resolved within the figure, all monitoring stations are located within a reach segment, a segment that excludes points of confluence. In-stream attenuation processes associated with headwater reaches affect only the transport of contaminants from their incremental drainage; in-stream attenuation processes for all non-headwater reaches affect the transport of contaminants from their own incremental drainage and also from the incremental drainage of all upstream reaches (including reaches in upstream nested basins).

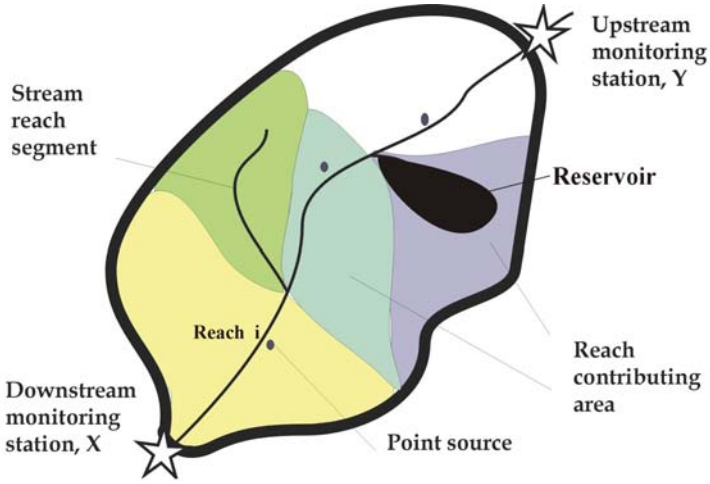


Figure 1.12. Conceptual illustration of a reach network for five incremental watersheds. Model equation (1.27) describes the supply and transport of load within an individual reach and its incremental watershed. [From McMahon and others, (2003).]

This conceptual model can be formalized through a mathematical equation. Let F_i^* be the model-estimated flux for contaminant leaving reach *i*. This flux is related to the flux leaving adjacent reaches upstream of reach *i*, denoted by F_j' where *j* indexes the set $J(i)$ of adjacent reaches upstream of reach *i*, plus additional flux that is generated within the incremental reach segment *i*. In most cases, the set of adjacent upstream reaches $J(i)$ will consist of either two reaches, if reach *i* is the result of a confluence, or no reaches if reach *i* is a headwater reach (see figure 1.12). The functional relations determining reach *i* flux are given by

$$(1.27)\quad F_i^* = \left(\sum_{j\in J(i)} F_j'\right)\delta_i A\left(\mathbf{Z}_i^S, \mathbf{Z}_i^R; \boldsymbol{\theta}_S, \boldsymbol{\theta}_R\right) + \left(\sum_{n=1}^{N_S} S_{n,i}\alpha_n D_n\left(\mathbf{Z}_i^D; \boldsymbol{\theta}_D\right)\right) A'\left(\mathbf{Z}_i^S, \mathbf{Z}_i^R; \boldsymbol{\theta}_S, \boldsymbol{\theta}_R\right).$$

The first summation term represents the amount of flux that leaves upstream reaches and is delivered downstream to reach *i*, where F_j' equals measured flux, F_j^M, if upstream reach *j* is monitored or, if it is not, is given by the model-estimated flux F_j^*. δ_i is the fraction of upstream flux delivered to reach *i*. If there are no diversions, then δ_i is set to 1. In most applications, this fraction is defined by the fraction of streamflow leaving upstream reaches that is delivered to reach *i*. If the streamflow delivery fraction is not known, it is possible to define it as a parametric function of stream characteristics that can be used to model and estimate the amount of flow diverted from the channel. For example, useful information for determining the amount of flux diverted at reach *i* would be whether or not the reach network has a diversion at reach *i* and whether or not the diversion reconnects with the main channel downstream (in which case the diversion is part of a braided channel). $A(\cdot)$ is the stream delivery function representing attenuation processes acting on flux as it travels along the reach pathway. This function defines the fraction of flux entering reach *i* at the upstream node that is delivered to the reach's downstream node. The factor is a function of measured stream and reservoir characteristics, denoted by the vectors $\mathbf{Z}^S$ and $\mathbf{Z}^R$, with corresponding coefficient vectors $\boldsymbol{\theta}_S$ and $\boldsymbol{\theta}_R$. If reach *i* is a stream, then only the $\mathbf{Z}^S$ and $\boldsymbol{\theta}_S$ terms determine the value of $A(\cdot)$; conversely, if reach *i* is a reservoir then the terms that determine $A(\cdot)$ consist of $\mathbf{Z}^R$ and $\boldsymbol{\theta}_R$.

The second summation term represents the amount of flux introduced to the stream network at reach *i*. This term is composed of the flux originating in specific sources, indexed by $n = 1,\ldots,N_S$. Associated with each source is a source variable, denoted S_n. Depending on the nature of the source, this variable could

represent the mass of the source available for transport to streams, or it could be the area of a particular land use. The variable α_n is a source-specific coefficient. This coefficient retains the units that convert source variable units to flux units. The function $D_n(\cdot)$ represents the land-to-water delivery factor. For sources associated with the landscape, this function along with the source-specific coefficient determines the amount of contaminant delivered to streams. The land-to-water delivery factor is a source-specific function of a vector of delivery variables, denoted by $\mathbf{Z}_i^D$, and an associated vector of coefficients $\boldsymbol{\theta}_D$. For point sources that are described by a measured (in the same units as flux) discharge of mass directly to the stream channel (e.g., municipal wastewater effluent measured in kilograms year^{-1}), the delivery factor should be 1.0, with no underlying factors acting as determinants, and the source-specific coefficient should be close to 1.0. The last term in the equation, the function $A'(\cdot)$, represents the fraction of flux originating in and delivered to reach *i* that is transported to the reach's downstream node. This function is similar in form to the stream delivery factor defined in the first summation term of the flux equation; however, the default assumption in SPARROW models is that if reach *i* is classified as a stream (as opposed to a reservoir reach), the contaminants introduced to the reach from its incremental drainage area receive the square root of the reach's full in-stream delivery. This assumption is consistent with the notion that contaminants are introduced to the reach network at the midpoint of reach *i* and thus experience only half of the reach's time of travel. For reaches classified as reservoirs, the default assumption is that the contaminant receives the full attenuation defined for the reach.

The nonlinear model structure in equation (1.27) contains several key features. The additive contaminant source components and multiplicative land and water transport terms are conceptually consistent with the physical mechanisms that explain the supply and movement of contaminants in watersheds. Total modeled flux for a reach is shown to be decomposed into its individual sources. Because this same decomposition is done for all upstream reaches, it is possible in this framework to perform flux accounting, whereby total flux is attributed to its source components. All processes are spatially referenced with respect to the stream network according to the reach in which they operate. This means, for example, that a reservoir at reach *i* affects the transport of all contaminants entering the reach network upstream (but not downstream) of reach *i*. The additive source components also provide a mathematical structure in the model that preserves mass. This can be seen by noting that a doubling of each of the source variables $S_{n,i}$, along with a doubling of all upstream sources, as represented by a doubling of F_j', results in an exact doubling of modeled flux F_i^*. Finally, the modeled flux at any reach *i* is conditioned on monitored fluxes entering the stream network anywhere upstream of reach *i*. This approach to nested basins serves to isolate errors introduced in any upstream basin from incremental errors that arise in a downstream basin, making it defensible to treat nested basins as independent observations. Formally, the approach is similar to the way time series models condition predictions on past realizations of the series.

1.4.2 Contaminant sources

The selection of suitable contaminant source variables to evaluate in SPARROW models initially depends upon a user's particular knowledge of a watershed as well as inferences that can be derived from the research literature about the major sources that contribute pollutants to watersheds. For example, in the case of nutrients, the most commonly modeled pollutants with SPARROW, information from catchment- and regional-scale mass-balance or budget studies (e.g., Puckett, 1995; Boyer and others, 2002; Howarth and others, 1996, 2002; see figure 1.13) is often useful in identifying the principal pollutant sources and in determining the appropriate specification for the sources in SPARROW models.

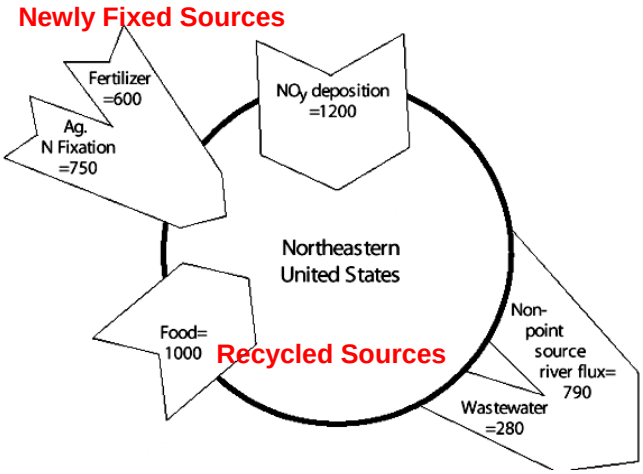


Figure 1.13. Regional nitrogen budget for watersheds of the northeastern United States. [Modified from Howarth and others (2002).]

Knowledge of the geography of these sources, based on direct measures or surrogate indicators of the contaminant mass supplied to the land surface and surface waters, is also critical to provide an effective use of the spatially distributed structure in SPARROW to model the origin and fate of contaminants in watersheds. The compilation and availability of large digital spatial data sets has become much more common (e.g., Brakebill and Preston, 1999; also see figure 1.14) and the advances in many types of digital topographic and stream network data (e.g., NHD) have made spatially distributed modeling more feasible. The inclusion of detailed information in the SPARROW model on the geographic locations where contaminants are released in watersheds has particular relevance to the potential policy and management applications of the model. Consider, for example, the specification of nitrogen in SPARROW models, a substance that undergoes extensive recycling in the production of food and animal feed (Jordan and Weller, 1996). Although nitrogen inputs to watersheds from newly fixed sources—i.e., fertilizers and atmospheric deposition—are recycled and exported from watersheds in food and animal feed (Jordan and Weller, 1996; Howarth and others, 1996; fig. 1.13), the mechanics of these nutrient transfers, which often involve considerable uncertainties, have not been typically included in SPARROW models. Instead, the geography of the releases of these recycled nutrients to watersheds, which has primary interest from a management perspective, is often explicitly accounted for by directly including animal and human (municipal wastewater effluent) wastes as sources in the model. Detailed spatial data are commonly available for describing the geography of these sources in watersheds (Smith and others, 1997; Brakebill and Preston, 1999). Refinements to the source data or the use of more complex model specifications to account for the removal of nitrogen in harvested crops (or to account for other details of the agricultural production system) can be accommodated in SPARROW models, provided spatial data are available that describe these activities and processes. Details of the agricultural production system may perhaps be best specified as land-to-water delivery factors (see discussion in the next section). In fact, we have previously

evaluated specifications in SPARROW nutrient models that included the available estimates of residual fertilizer applications, based on assumed exports in harvested crops and volatilization losses (Kellogg and others, 2000); however, these have been found to provide little to no additional explanatory power in the models. This highlights the need to place initial emphasis on developing simple SPARROW model specifications that accurately account for the principal source inputs. The model infrastructure can then be used to test more complex specifications that incorporate other details of the transport pathways and subtleties of the management system in cases where the data are available.

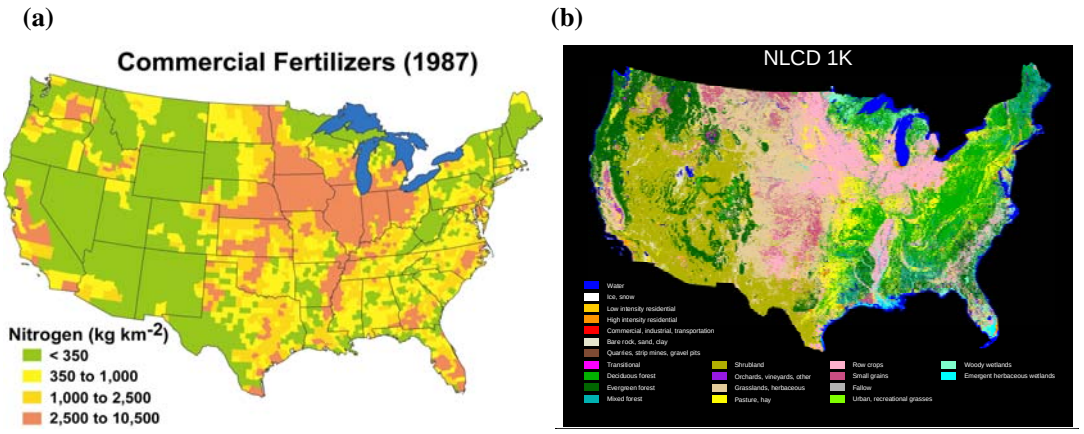


Figure 1.14. Examples of (a) intensive and (b) extensive measures of nitrogen sources in United States watersheds. [County fertilizer use is from Puckett (1995); National Land Cover Dataset (NLCD) land-use data (Vogelmann and others, 2001) are aggregated to one-kilometer areas.]

Note that there can be statistical limits to building more highly complex models of similar contaminant sources. The individual coefficients of sources with similar or correlated spatial distributions (e.g., confined and unconfined animal wastes; nitrate and ammonia wet deposition) may be difficult to statistically estimate in a SPARROW model. Difficulties can also arise if individual components of a source (e.g., total fertilizer) contribute relatively small quantities of pollutant mass to streams (e.g., non-farm fertilizer use). Moreover, additional complications can arise if the monitoring stations are located too far downstream to detect the effects of a sub-component of a major source. One alternative (and more deterministic) approach is to combine the mass quantities from multiple sources and statistically estimate a single coefficient in the model. This provides a mechanism in an estimated version of the model for simulating the effects of changes in the inputs from the individual sources that were combined.

Overall, the availability of digital spatial data on pollutant sources can also place practical limits on the set of feasible specifications of the source variables in SPARROW models. For example, estimation of the nitrogen supplied to watersheds by N fixation can involve considerable uncertainties (Jordan and Weller, 1996; Boyer and others, 2002) because the spatial measurements are relatively sparse and highly variable among different vegetation types. Moreover, there are frequently considerable difficulties in obtaining reliable measurements of wastewater treatment plant effluent loadings (Knopman and Smith, 1993); spatially detailed population statistics from the Census Bureau may need to be used as surrogate measures of human wastes when other information is not available or is unreliable.

Table 1.4. Potential types of contaminant-source variables in SPARROW models.

[P, phosphorus; kg, kilograms; km, kilometers; yr, year]

		Examples for total phosphorus (TP) as a response variable	
Source type	**Explanation**	**Source variable**	**Model coefficient units**
Intensive	Direct measures of contaminant mass	Nutrient mass in fertilizer, livestock waste, and atmospheric deposition	Dimensionless—mean fraction of source input delivered to streams
Extensive	Measures of area or population; surrogate indicators of mass	Land area in crops, forest, or impervious cover	kg P km^{-2} yr^{-1}
		Population served by septic or sewer disposal systems	kg P $person^{-1}$
Predictions from terrestrial and atmospheric mechanistic models	Measures of contaminant mass or surrogate indicators of mass	National Resources Inventory (NRI) Universal Soil Loss Equation (USLE) estimated erosion	kg P per kg sediment
		Net primary production from a terrestrial biogeochemical model	kg P per kg carbon
		Erosion Productivity Impact Calculator (EPIC) agricultural model estimate of P runoff	Dimensionless

A summary of the types of source variables, $S_{n,j}$, that can be included in SPARROW models is given in table 1.4. The source terms used in the model can be generally classified as *intensive* and *extensive* measures of contaminant mass (see examples in figure 1.14). The former measures are typically descriptive of direct measures of pollutant mass, such as fertilizer application, livestock waste, atmospheric deposition, or sewage-effluent loadings. In these cases, the source-specific parameter (α_n) is expressed as a dimensionless coefficient that, together with standardized expressions of the land-to-water delivery factor, describes the proportion or fraction of the source input that is delivered to streams (note that source and land-to-water delivery coefficients that are standardized in relation to the mean values of the land-to-water delivery variables are necessary to compare and interpret the physical meaning of source coefficients; see discussions in the next section). This fraction would be expected to be less than 1.0 but greater than zero, reflecting the removal of contaminants in soils and ground water. In the case of fertilizer, for example, the coefficient would include the aggregate effects of a variety of processes and human activities that remove nitrogen from agricultural lands and subsurface flow paths, such as the volatilization of ammonia fertilizer forms, the removal of nitrogen in harvested crops, and the long-term immobilization and storage of N or denitrification in soils and ground waters.

Exceptions to the required upper bound of 1.0 on source coefficients with dimensionless units may occur, however. One such case is that in which the measured input of a source that is used as an explanatory variable in the model is less than the total amount of the contaminant that is likely to be introduced to the watershed, implying the source specified in the model is a surrogate for the total mass input. For example, spatially detailed measurements of atmospheric deposition are available only for wet inorganic nitrogen forms; other forms of nitrogen (organic and dry) are measured too sparsely to be used as an explanatory variable in a national SPARROW model. In this case, a greater than unity coefficient is potentially estimated and will reflect

additional contributions from wet deposition of organic N and dry deposition of inorganic N, which are not included in the inputs to the model. This result is expected, provided that these unmeasured quantities are correlated with the measured wet deposition, which is commonly the case (Alexander and others, 2001). A second exception is that in which direct measures of point-source loadings (e.g., in municipal wastewater effluent), that have identical units as the response variable, are used in the model. In this case, the point-source coefficient is expected to be close to 1.0 (and can even slightly exceed 1.0) and the confidence interval is expected to contain 1.0. A significant deviation of the coefficient from unity is not expected because point source facilities discharge directly to water bodies without any loss of contaminant mass. In fact, the confidence intervals of the point-source coefficient for many of the regional SPARROW models frequently contain 1.0 (see the discussion in section 1.5.4.2). Appreciable deviations in the coefficient value from 1.0 may indicate a poor specification of the model or inaccurately measured point-source effluent data.

Extensive measures of contaminant mass also may be used in SPARROW models. These are surrogate indicators of contaminant mass and include measures of watershed properties such as specific land-use area and sewered population that are considered to be proportional to the actual mass loadings generated by a general type of a contaminant source. The empirical estimates of the source coefficients in the model provide a quantitative measure of the proportion of the mass loading that is associated with a specified source. If extensive measures are used, the associated model coefficients are expressed as the contaminant mass generated per unit of the source type (e.g., kilograms kilometer^{-2} year^{-1}; kilograms person^{-1} year^{-1}). If combined with the land-to-water delivery factor and expressed as a standardized source coefficient (see the discussion in section 1.4.3), the coefficient indicates the mean quantity of contaminant mass per unit of the surrogate source measure that is delivered to streams. For land-use terms, the standardized coefficient gives what is frequently cited as an "export" coefficient (Johnes and others, 1996; Beaulac and Reckhow, 1982). Observed values of export coefficients have been reported in the literature for various land-use/land-cover types, such as crop, urban, and forested lands (Beaulac and Reckhow, 1982); the literature coefficients often compare favorably with those estimated by SPARROW for these land-use/land-cover types (e.g., Alexander and others, 2004). For surrogate measures based on sewered population, the standardized coefficient gives a per capita rate of contaminant discharge to streams. For nitrogen, we commonly estimate SPARROW per capita rates of 2-5 kilograms person^{-1} year^{-1}, which is similar to the reported range of per capita values associated with discharges of wastewater effluent from municipal treatment plants (Alexander and others, 2001).

The predictions of contaminant mass from mechanistic terrestrial and atmospheric spatial models may also be used to define the source input terms in SPARROW models. For example, these include measures of net primary productivity from biogeochemical models such as BIOME-BGC (BioGeochemical Cycles) as an indicator of organic carbon availability (Running and Hunt, 1993), estimates of wet and dry atmospheric deposition of oxidized and reduced nitrogen forms from atmospheric models such as Regional Modeling System for Aerosols and Deposition (REMSAD; ICF Consulting, 2002), sediment erosion estimates generated from the Universal Soil Loss Equation (USLE) as part of the U.S. Department of Agriculture (USDA) National Resources Inventory (NRI; Natural Resources Conservation Service, 2005), and edge-of-field deliveries of nutrients from the agricultural model Environmental Policy Integrated Climate (EPIC; Sharpley and Williams, 1990), which describes the effects of agricultural management on crop productivity and erosion. SPARROW provides a spatial framework that can be used to evaluate whether the additional complexity reflected in such model predictions offers greater explanatory power than the more simple intensive and extensive measures of contaminant sources.

The discussion above pertains to the use of SPARROW to model water quality. An equally valid exercise, however, would be to apply SPARROW to streamflow, in which case a principal source variable is atmospheric precipitation. In modeling streamflow, it may be necessary to introduce the concept of a negative source—a "source" that represents removal of flux from the stream network. Examples of a negative source would be the consumptive use of water by various water users, such as irrigation, public supply, industrial and commercial users (Solley and others, 1998).

1.4.3 Landscape variables

Landscape variables in SPARROW describe properties of the landscape that relate to climatic, or natural- or human-related terrestrial processes affecting contaminant transport. These typically include properties for which there is (1) some conceptual or empirical basis for their importance in controlling the rates of contaminant processing and transport, and (2) broad-scale availability of continuous measurements of the properties for use in model estimation and prediction. The model structure allows the user to test hypotheses about the influence of specific features of the landscape on contaminant transport. Landscape variables may include water-balance terms (e.g., precipitation, evapotranspiration) related to climate and vegetation, soil properties (e.g., organic content, permeability, moisture content, see figure 1.15), topographic water flow-paths variables (e.g., TOPMODEL overland flow, topographic index, and slope) (Beven and Kirkby, 1979), or management practices and activities, including tile drainage, conservation tillage practices, and BMPs related to stream riparian properties. Particular types of land-use classes, such as wetlands or impervious cover, may also be potentially used to describe transport properties of the landscape.

Overall, users of SPARROW models should select for evaluation landscape properties that are known to be associated with water inputs and with overland and subsurface transport as well as with biogeochemical processes in vegetation, soils and ground waters that control the long-term transport of contaminants. The model specification assumes steady-state conditions so that the landscape properties are indicative of the long-term aggregate, net effects of processes such as denitrification, mineralization, or long-term storage in ground waters. Thus, landscape properties are used to infer the net effects of a variety of processes on nutrient transport and delivery to aquatic ecosystems. The sign of the landscape delivery coefficients commonly provide important information for inferring the nature of the relation to in-stream flux; however, the magnitude of estimated coefficients in the model cannot be specifically interpreted as is possible with the aquatic decay terms that are discussed in subsequent sections of the report.

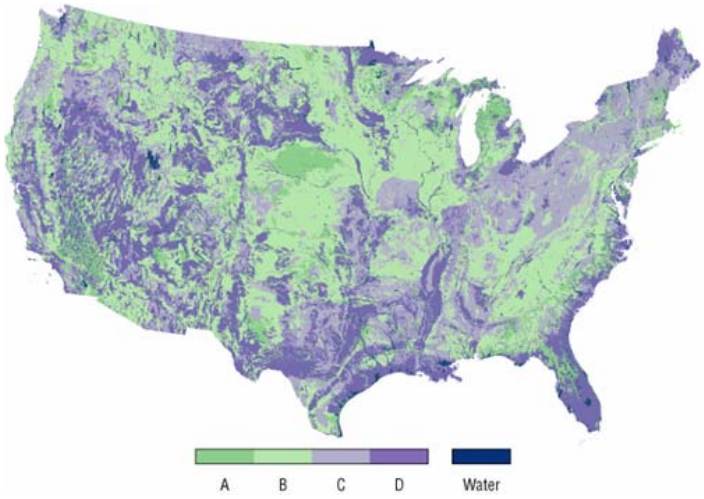


Figure 1.15. Mapped hydrologic soils groups from the STATSGO database reflecting the water infiltration and saturation properties of soils. [From Miller and White (1998).]

In general, it is advisable to select landscape variables that describe properties of the landscape exclusive of those of the aquatic systems (i.e., streams, reservoirs, and lakes). For example, the use of "precipitation" as a delivery factor would be generally preferable to using "runoff", which frequently is estimated using water gage data, and would not generally provide climatic information that is independent of the in-stream velocity data used to estimate in-stream decay rates in SPARROW models.

The interaction of particular land-to-water delivery factors with individual sources may also be important to consider in SPARROW models, and the software can accommodate the investigation of these interactions. The software allows each source and land-to-water delivery factor interaction to be set to an "off" or "on" position so that such interactions can be evaluated. Several examples are described in detail in Part 2 of this report to assist users with these investigations. Example interactions include the effects of tile drainage or conservation tillage practices on nitrogen export from row-cropped lands or on the delivery of fertilizer nitrogen to streams. These agricultural management activities would not typically interact with other types of diffuse sources in the model, such as urban land export or atmospheric sources deposited on nonagricultural lands. In addition, measures of soil permeability or porosity would not generally interact with contaminant mass released in areas with impervious surfaces, such as commercial or high-density urban land uses. Because municipal wastewater loadings discharge directly to stream channels, point-source terms have no interaction with land-to-water delivery factors in the SPARROW model specification.

The land-to-water delivery factor in equation (1.27), $D_n\left(\mathbf{Z}_i^D;\boldsymbol{\theta}_D\right)$, is a source-specific function of a vector of delivery variables, denoted by $\mathbf{Z}_i^D$, and an associated vector of coefficients $\boldsymbol{\theta}_D$. In basic SPARROW models, the land-to-water delivery factor has been commonly expressed according to an exponential functional form. For source *n*, the fraction of contaminant mass that is generated in the incremental drainage area of the reach and delivered to the reach (excluding the source coefficient term) is estimated as

$$D_n\left(\mathbf{Z}_i^D;\boldsymbol{\theta}_D\right)=\exp\left(\sum_{m=1}^{M_D}\omega_{nm}Z_{m\,i}^D\theta_{Dm}\right), \tag{1.28}$$

where $Z_{m\,i}^D$ represents delivery variable *m* for the incremental drainage of reach *i*, θ_{Dm} is its corresponding coefficient, ω_{nm} is an indicator variable that is 1.0 if delivery variable *m* affects source *n* and zero otherwise, and M_D is the number of delivery variables. In our experience, we have found that log-transformed delivery variables, such as the logarithm of soil permeability, may provide an improved fit to the data in certain cases. Under the log-transformation, the land-to-water coefficient is interpreted as the percent change in flux delivered to streams, derived from all sources to which the land-to-water variable is applied, from a one-percent increase in the land-to-water delivery variable.

The full effect of landscape variables on the delivery of source *n* to streams is determined by the product of the delivery factor in equation (1.28) and the source coefficient α_n. Therefore, a modification of delivery factor specification can be expected to change the mean of the delivery factor, resulting in a corresponding counter-adjustment of the source coefficient α_n. In order to improve the interpretability of the source coefficient, and to provide some stability in its value across alternative specifications of the land-to-water delivery factor, it is recommended that the delivery variables $Z_{m\,i}^D$ in equation (1.28) be expressed as differences from their mean value over all reaches. The SPARROW code allows the user to select this option to compute and report standardized delivery and source coefficients as part of the model output.

Other functional forms of the landscape variables have been used in previous SPARROW models (e.g., Smith and others, 1997; Alexander and others, 2000). In these cases, the delivery factor was specified by using an imbedded negative sign in the exponential function, and variables having a positive effect on delivery were expressed as reciprocals (that is, as $\left(Z_{m\,i}^D\right)^{-1}$). This specification allows all delivery coefficients to be reported as positive values, but requires that delivery variables having a positive effect on delivery be transformed to their reciprocal form prior to model execution, implying the user has *a priori* information regarding the direction of the effect of a particular delivery variable on the delivery factor. In general, we recommend use of the expression in equation (1.28)—i.e., the non-reciprocal form—to allow a more transparent reporting of the sign of the land-to-water delivery coefficient. This approach also eliminates the need for the user to identify the direction of the flux and delivery variable relation in advance of model estimation.

It is also feasible for SPARROW to operate with a more spatially detailed specification of the land-to-water delivery term to account for transport-altering properties along "off-reach" flow paths within the incremental reach watersheds, although this functionality has not been previously evaluated in the model. In this

approach, an incremental watershed is subdivided into τ_i individual cells, each cell having high-resolution information concerning the cell's elevation, source contributions, and other attributes. A generalization of the model source and delivery terms in equation (1.27) as applied to M_D land-to-water delivery variables results in a first-order approximation of the source *n* delivery factor for reach *i* given by

$$D_n\left(Z_i^D;\theta_D\right)=\exp\left(\sum_{m=1}^{M_D}\omega_{nm}\theta_{Dm}\bar{z}_{mi}^D\right)\left(1+\sum_{m=1}^{M_D}\omega_{nm}\theta_{Dm}\tau_i\mathrm{Cov}\left(s_{n,i},z_{m\,i}^D\right)\right), \tag{1.29}$$

where $\bar{z}_{mi}^D$, representing the cell equivalent of $Z_{m\,i}^D$ in (1.28), is the average across all τ_i cells comprising the incremental drainage for reach *i* of the individual cell values $z_{m\,i}^D$ of the m^{th} delivery variable, $s_{n,i}$ are the individual cell values of the source n for reach *i*, and $\mathrm{Cov}\left(s_{n,i},z_{m\,i}^D\right)$ is the cell-based covariance between source *n* and delivery variable *m* in the incremental watershed *i*.

1.4.4 Stream transport

Stream attenuation processes that act on contaminant flux as it travels along stream reaches are frequently modeled according to a first-order reaction rate process (Chapra, 1997). A first-order decay process implies that the rate of removal of the contaminant from the water column per unit of time is proportional to the concentration or mass that is present in a given volume of water (note that a zero-order process would correspond to a constant rate of removal per unit of time). According to a first-order decay process, the fraction of contaminant removed over a given stream distance is estimated as an exponential function of a first-order reaction rate coefficient (expressed in reciprocal time units) and the cumulative water time of travel over this distance.

A reaction rate expression is estimated on a volumetric basis, and therefore is expected to be dependent on properties of the water column that are proportional to water volume, such as streamflow and water-column depth (Stream Solute Workshop, 1990). Accordingly, in basic forms of the SPARROW model, the fraction of the contaminant mass originating from the upstream node and transported along reach *i* to its downstream node is estimated as a function of the mean water time of travel ($T_{c\,i}^S$; units of time) in reach *i* and stream class *c* defined according to discrete intervals of mean streamflow or depth (in this case, $\mathbf{Z}_i^S=\left\{T_{c\,i}^S\right\}$, $c=1,\ldots,C_S$, where $T_{c\,i}^S$ is nonzero only for the streamflow class corresponding to reach *i*), and a stream-size dependent loss rate coefficient (θ_{Sc}; units of time^{-1}) such that

$$A\left(\mathbf{Z}_i^S,\mathbf{Z}_i^R;\boldsymbol{\theta}_S,\boldsymbol{\theta}_R\right)=\exp\left(-\sum_{c=1}^{C_S}\theta_{Sc}T_{c\,i}^S\right). \tag{1.30}$$

Mean water time of travel is estimated as the quotient of the reach channel length and mean water velocity. The most accurate estimates of water velocity are obtained from time-of-travel dye studies (Jobson, 1996), but in some cases may be obtained from instantaneous measurements of water velocity taken during flow gage site visits. Empirical geomorphic relations, which use regression methods to relate time-of-travel measurements to channel and basin properties (e.g., streamflow, slope), also are available for regions of the U.S. and other countries (e.g., U.S. Environmental Protection Agency, 1996; Alexander and others, 1999; Jobson, 1996; Jowett, 1998) and can be used to estimate the time-of-travel of stream reaches for a given river network. Alternatively, channel length may be used in equation (1.30) where water velocity estimates cannot be obtained for the stream network. This assumes that the water time of travel is proportional to the channel length, and the estimates of the removal rate are expressed as reciprocal length (e.g., see McMahon and others, 2003; Alexander, Elliott, and others, 2002).

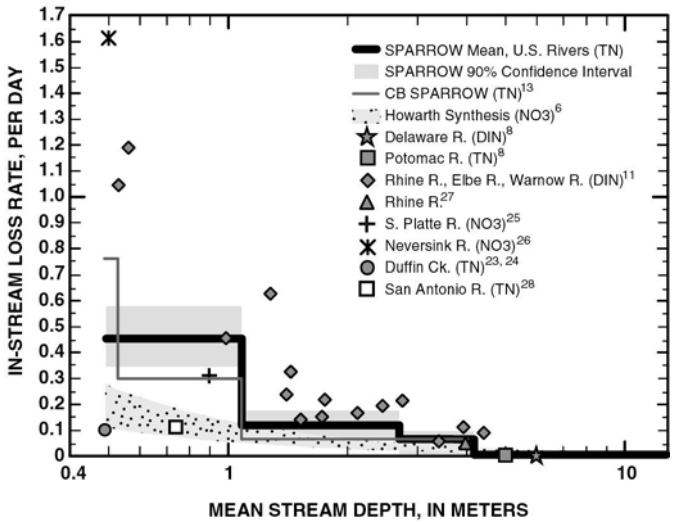


Figure 1.16. Estimates of in-stream nitrogen removal from SPARROW and experimental studies. [TN, total nitrogen; NO3, nitrate; DIN, dissolved inorganic nitrogen; CB, Chesapeake Bay; from Alexander and others (2000).]

The specification of a stream-size dependent loss rate coefficient in equation (1.30) is appropriate for modeling the transport of many non-conservative contaminants in which water depth or volume are expected to influence in-stream removal processes. For nitrogen, this includes processes such as denitrification and particulate settling and storage (e.g., Alexander and others, 2000; Peterson and others, 2001; Stream Solute Workshop, 1990). A decline in the rate of contaminant removal in streams with increasing water depth (as illustrated for nitrogen in figure 1.16) is generally consistent with theories about the physical and biological mechanisms that explain solute loss from the water column (Stream Solute Workshop, 1990). For nutrients, the rates of removal are controlled by the contact and exchange of stream waters with the benthic sediments (e.g., hyporheic zone), where denitrification and heterotrophic and autotrophic uptake occur (Peterson and others, 2001; Harvey and Wagner, 2000; Bohlke and others, 2004; Alexander and others, 2000). The rate at which these removal processes occur is generally expected to decline with increasing stream depth and flow, corresponding to an increase in the water volume relative to the surface area of the benthic sediment.

The in-stream loss rate coefficients estimated by SPARROW in equation (1.30) reflect the *mean-annual* rates of contaminant removal in the stream reach. The mean-annual rates reflect the steady-state form of the basic SPARROW model, based on the use of a detrended flow-adjusted mean-annual flux as the response variable. The in-stream loss rate represents a net annual removal rate that would be expected to describe relatively permanent removal processes. For example, in the case of nutrients, these removal processes may include denitrification and any long-term storage of particulate phosphorus or organic particulate nitrogen that may occur over long periods in the stream channel or flood plain. The estimated loss rate should not be sensitive to seasonal removal processes (e.g., algal uptake), which would not be expected to cause a net removal of nutrients over an annual period because of the seasonal re-cycling of nutrients in stream biota.

In recent SPARROW nitrogen models, the estimates of the nitrogen loss in streams display an exponential decline in magnitude with increasing water depth (fig. 1.16). Comparisons with literature rates of nitrogen removal, which reflect permanent loss (i.e., denitrification) or long-term storage, are generally similar in magnitude to the SPARROW rates and also show a decline with increasing channel size. By comparison, these removal rates are only 1 to 10 percent of the removal rates associated with the temporary removal of nitrogen via heterotrophic uptake (Alexander, Elliott, and others, 2002), based on Lotic Intersite Nitrogen eXperiment (LINX) stream measurements (Peterson and others, 2001). Evidence for similar inverse relations between in-stream removal and stream size has been obtained in SPARROW models for TP (Smith and others, 1997; Alexander and others, 2004) and fecal coliform bacteria (Smith and others, 2004). Settling processes and mortality from ultraviolet light exposure (for fecal coliform) are mechanisms that may explain the relation of fecal bacterial losses to stream size. By contrast, a recent SPARROW model of suspended sediment flux (Schwarz and others, 2001) indicates strong empirical evidence that streams act as a sediment source, presumably related to channel erosion (i.e., a negative decay coefficient)—a result that is generally consistent

with current understanding of long-term sediment deposition and transport based on catchment-scale observations (Trimble and Crosson, 2000). The sediment model indicates that most of the long-term aquatic removal of sediment occurs in reservoirs rather than in streams.

The choice of the appropriate number of loss-rate coefficients and steam-size classes in SPARROW nutrient models has been typically based on an empirical evaluation of model fit (e.g., mean square error), the presence of a positive sign on the rate coefficients among competing models, and evidence of an inverse relation with stream size (e.g., see Alexander and others, 2000; Alexander, Elliot, and others, 2002). In previous SPARROW studies, rate coefficients have been typically estimated for two or more stream size classes defined on the basis of mean streamflow. Properties such as water depth or measures of Strahler stream order may also be used in SPARROW models to separate steams according to their size. Initial streamflow breakpoints may be selected to reflect order-of-magnitude separations and to ensure that a sufficient number of stream reaches and monitoring sites are represented in the flow classes.

Alternatively, in-stream contaminant removal may be modeled as a continuous function of streamflow by re-expressing equation (1.30) as

$$A\left(\mathbf{Z}_i^S, \mathbf{Z}_i^R; \boldsymbol{\theta}_S, \boldsymbol{\theta}_R\right) = \exp\left(-\theta_{S1} Q_i^{\theta_{S2}} T_i^S\right), \tag{1.31}$$

where Q_i is the mean-annual streamflow of reach *i*, and θ_{S1} and θ_{S2} are estimated coefficients. This type of continuous function has been used in SPARROW (Alexander, Elliott, and others, 2002; Elliott and others, 2005) and other watershed models (Cooper and Bottcher, 1993) to estimate nutrient removal in New Zealand streams. The rates, obtained empirically in these models, are generally consistent with experimental measurements of nutrient removal in New Zealand streams, which has been shown to be exponentially related to streamflow (Rutherford and others, 1987; Alexander, Elliott, and others, 2002). A negative θ_{S2} coefficient is consistent with an inverse relation between the in-stream removal rate and stream size.

Based on our experiences in previous studies (e.g., Alexander and others, 2000; Alexander, Elliott, and others, 2002; Elliott and others, 2005) and in recent updates and evaluations of the U.S. national nutrient models, both the *discrete* and *continuous* in-stream loss functions generally provide nearly equally acceptable statistical fits to the observed data, as measured by the RMSE. Some differences are apparent, however, in the estimates of the mean and variance of the loss rates over a range of stream sizes. Figure 1.17 illustrates how the two loss specifications compare for their estimated in-stream removal rates of total nitrogen over a range of stream sizes based on mean-annual streamflow. These rates were estimated on the basis of an analysis of the U.S. stream data used as the illustration dataset in this report, with source and land-to-water delivery variables as described in Alexander and others (2000) and a reservoir removal rate specified according to equation (1.34) given in section 1.4.5. The model was applied to the Enhanced Reach File 1 (ERF1) version 2.0 infrastructure described in Nolan and others (2002). The statistical fit is virtually identical for the two functions (RMSE of 0.6229 versus 0.6211 for the discrete and continuous models, respectively—i.e., less than 0.2 percent difference in model error). The discrete loss rates approximate the general shape and magnitude of the loss rates associated with the continuous function—the closest agreement in the decay rates is observed in mid-size rivers (approximately 1,000 to 10,000 feet^3 second^{-1}). Somewhat larger continuous loss rates are generally estimated for the smallest streams (less than 100 feet^3 second^{-1}) and the largest rivers (greater than 10,000 feet^3 second^{-1}). The model compensates for these differences in the loss rates, in part, by upwardly adjusting the export for nonagricultural diffuse sources in the continuous loss rate model. Despite these differences in loss rates, the rates of both functional forms generally agree well with the available literature estimates of in-stream nitrogen removal as given in figure 1.16 above, although there is evidence that the in-stream removal rates for nitrogen in large rivers based on the continuous function are somewhat larger than those reported in the literature.

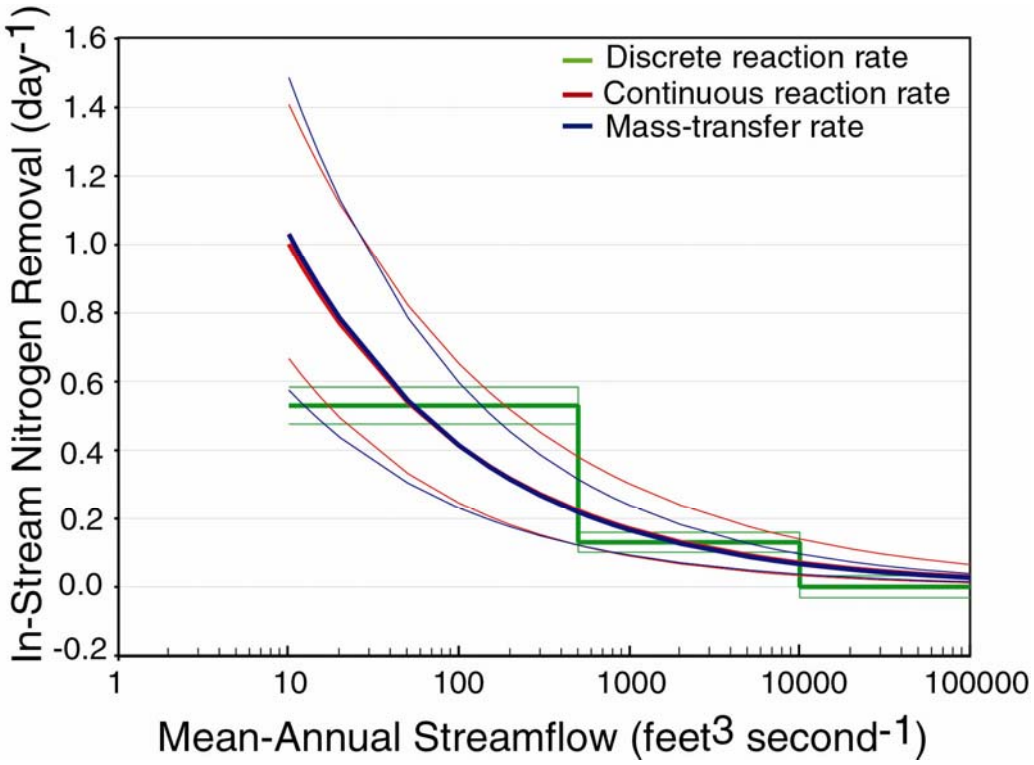


Figure 1.17. Estimates of in-stream total nitrogen removal from SPARROW. Mean rate (bold lines) and standard errors were estimated from the illustration national dataset prepared for this report.

Larger differences are apparent in the estimates of the variance of the in-stream loss rates (i.e., coefficient standard error of estimate) for the discrete and continuous loss functions shown in figure 1.17. The standard errors of the discrete function are typically smaller and generally uniform in absolute units over the range of flows as compared to those associated with the continuous function. In contrast to the standard errors of the discrete loss rates, those associated with the continuous function decline in absolute units with an increase in stream size. The standard errors associated with both the discrete and continuous loss functions increase in relative units (i.e., expressed as a percentage of the mean) with increasing stream size.

The uniformity of the standard errors of the *discrete* loss function suggests that the particular choice of flow intervals shown here provides a generally similar level of precision in absolute units for estimating the loss coefficients. Differences in the t-statistics and the statistical significance of the coefficients are, therefore, largely a function of the magnitude of the coefficient estimates. Other choices of flow intervals result in different levels of precision, especially in the tails of the flow distribution; for example, the choice of narrower flow intervals to define the smallest and largest rivers would reduce the number of reaches and monitoring stations associated with the smallest and largest flow classes, and lower the precision of the coefficients.

By contrast, the decline in the absolute variance of the in-stream loss rate with increasing flow, based on the *continuous* loss function, suggests that the most precise estimates of nitrogen loss and smallest range of uncertainty in absolute units exist for the loss rates of large rivers. Conceptually, this description of our knowledge of the uncertainties in the rates of in-stream removal is perhaps reasonable for contaminants where the removal processes are explained by water volume/surface area ratios and hyporheic interactions. For example, in large rivers where relatively little of the water column interacts with the benthic sediment, in-stream removal of nitrogen would be expected to be negligible, and the empirically estimated loss rates would be constrained in the vicinity of zero. Conversely, in small streams, where hyporheic interactions with the water column are more prominent, removal rates for nitrogen can vary over a large range as a function of variation in a wide range of non-volume related rate-controlling properties of the benthic sediment (e.g., organic carbon and oxygen content, sediment grain size, denitrifying bacteria density) and water column (e.g., nitrate and oxygen concentration). This notion is generally supported by the observed decline in the variability of the literature rates of nitrogen removal with increasing stream size (e.g., fig. 1.16; Alexander and others, 2000).

In general, it is difficult to identify either the *discrete* or *continuous* loss functions as having a unique advantage over the other for estimating the mean rates of contaminant removal; both merit evaluation and comparison in SPARROW modeling studies. On the one hand, the discrete function provides the most flexible functional form to estimate the loss coefficients over a wide range of streamflow. It also provides a somewhat direct measure of the information content in the highest and lowest flow classes (e.g., number of reaches and monitoring station observations) and the effect of this information on parameter estimation (i.e., mean and standard error) within specific intervals of stream size. Having estimates of the contaminant removal rates that are specific to the smallest and largest rivers can be helpful in defining unique breakpoints of the discrete function because these waters typically contain the least amount of information for fitting the in-stream loss function (i.e., few or no monitoring stations may be available for the smallest streams and, in the largest rivers, the rates of in-stream loss are relatively small and typically statistically indistinguishable from zero). The discrete function, however, is limited by the subjectivity of defining the streamflow class intervals and the simplicity of assuming that a single loss rate is applicable throughout a given interval of flow. The continuous function, although more constrained in its functional form, perhaps has the advantage of providing a more integrated view of contaminant removal across the full range of stream sizes. The continuous function also provides excellent complementary information about the slope of the loss relation that can be useful in defining stream size classes to support the use of the discrete loss function.

As an alternative to volume-dependent first-order reaction rates as given in equations (1.30) and (1.31), contaminant removal in streams can also be described as a flux to the benthic sediment as measured according to a volume- or depth-independent *mass-transfer coefficient* (expressed in units of length per time). The mass-transfer coefficient quantifies the vertical velocity at which a contaminant migrates from the water column into the sediment. This expression can be especially useful (as described below) for evaluating the effects of non-volume-related or non-hydrologic (i.e., biologic-related) properties on in-stream contaminant removal. The fraction of contaminant removed according to this expression is proportional to an estimated coefficient for the mass-transfer rate and the measured hydraulic load or water velocity in a stream or lake. Therefore, we can re-express equation (1.30), such that

$$A\left(\mathbf{Z}_i^S, \mathbf{Z}_i^R; \boldsymbol{\theta}_S, \boldsymbol{\theta}_R\right) = \exp\left(-\theta_S^{MT}\, T_i^S \,/\, D_i^S\right), \tag{1.32}$$

where θ_S^{MT} is the mass transfer coefficient in units of length per time and D_i^S is the water depth in units of length. This expression intrinsically accounts for the effects of water depth on in-stream processing of contaminants. The reciprocal of the $T_i^S \,/\, D_i^S$ term is the hydraulic loading—also a measure of the water displacement or water velocity in the reach (note that this term is mathematically equivalent to the reciprocal of the areal hydraulic loading, described in the next section as the ratio of the reservoir discharge to the reservoir surface area).

A mass-transfer coefficient can be used to calculate a continuous set of depth-dependent reaction rates (units of reciprocal time) for a range of stream channel depths by dividing the mass-transfer coefficient by stream depth. The resulting depth-dependent reaction rates are virtually identical to those estimated using a continuous reaction-rate function. For example, the SPARROW estimated mass-transfer coefficient (59 meters year^{-1}) for our illustration data set in figure 1.17 corresponds to a series of depth-dependent reaction rates that range from 1 day^{-1} at 10 feet^3 second^{-1} to 0.06 day^{-1} at 10,000 feet^3 second^{-1}. These reaction rates derived from a SPARROW estimated mass-transfer coefficient are virtually the same as those estimated using the continuous reaction-rate function (fig. 1.17). The estimates of variance for the mass-transfer coefficient also decline with increases in stream size, but are slightly smaller than those associated with the reaction-rate function for rivers of intermediate and large sizes.

Mass-transfer coefficients are commonly used by researchers in experimental tracer studies when comparing removal rates among streams of differing size (Stream Solute Workshop, 1990; Peterson and others, 2001) and are frequently cited in discussions of stream nutrient dynamics and nutrient spiraling concepts (Newbold and others, 1981; Stream Solute Workshop, 1990; Thomas and others, 2001; Peterson and others, 2001). Experimental mass-transfer coefficients are computed as the quotient of the measured areal rate of nutrient removal (related to heterotrophic or autotrophic processes) and the nutrient concentration in the water column. The mass-transfer coefficient is considered to be more descriptive of the intrinsic effects of non-hydrologic processes on nutrient removal (e.g., sediment grain size, organic carbon, dissolved oxygen content, and microbial population densities) than a volumetric-based reaction rate that includes the effect of water depth.

This expression of the in-stream loss rate is generally well suited for use in SPARROW to evaluate hypotheses about the effects of non-hydrologic processes on the aquatic removal of contaminants (see discussion below).

Note that the mass-transfer coefficient has identical units as the *apparent settling velocity* coefficient that has been used to describe the benthic flux of nutrients in lakes and reservoirs; this flux is related to denitrification and particulate settling losses from the water column (see discussion in the subsequent section). The settling velocity coefficient has been frequently used in empirical mass transport models for lakes and reservoirs to estimate the rate of nutrient removal from the water column (Chapra, 1997).

In estimating contaminant removal in streams in SPARROW models, no definitive *statistical* case can be conclusively made for using a depth-dependent *reaction rate* versus a depth-independent *mass-transfer rate* [although it is of note that estimation of a mass-transfer rate function uses fewer degrees of freedom as it typically requires fewer model parameters than a similar application of a depth-dependent reaction rate loss function]. One important consideration in selecting a removal rate expression is which approach best satisfies the objectives of the study (e.g., evaluation of hydrological or biological-related processes on in-stream removal) or perhaps provides removal rates that are most readily compared with literature rates. As noted previously, there are also differences in the magnitude of the variance of the removal rates. In previous SPARROW applications, we have frequently estimated multiple depth/flow-dependent reaction rates (e.g., Smith and others, 1997; Alexander and others, 2000; Alexander, Elliott, and others, 2002) as a means of evaluating and quantifying their relation to stream size. In fact, the effect of water volume on reaction rates had not been previously demonstrated over a broad range of stream and river sizes prior to the initial SPARROW studies of nutrients (Smith and others, 1997; Alexander and others, 2000).

The mass-transfer rate expression has particular merit for investigations of the effects of non-hydrologic processes on in-stream contaminant removal. Conceptually, a mass-transfer coefficient describes in-stream nutrient loss as a benthic flux process and may provide a generally more realistic description of nutrient removal mechanisms that occur as the result of interactions with the benthic sediment and hyporheic zone (related to denitrification or settling/storage processes). The expression of the mass-transfer function in SPARROW, equation (1.32), provides an estimate of the net removal rate that is adjusted for the effects of stream water depth and water time of travel (i.e., the hydraulic loading). Hence, the estimated mass-transfer coefficient is generally more descriptive of the intrinsic effects of non-hydrologic factors on nutrient removal than a reaction rate measure that is dependent on depth. Mass-transfer rate expressions in SPARROW can provide an important approach for directly testing hypotheses about the effects of biologic controlling properties, such as concentration (i.e., saturation kinetics) and temperature, on in-stream contaminant removal. One limitation of using the discrete depth-dependent reaction rate to evaluate non-hydrologic effects is that the implicit mass-transfer rate associated with the estimated reaction rates changes with water volume or depth (i.e., within each discrete flow/depth class the implied mass-transfer rate increases with stream size; note that the implied mass-transfer rate based on the continuous reaction rate expression increases only slightly with stream size). Therefore, the mass-transfer rate provides a generally more precise mechanistic separation in the model of the effects of hydrologic and non-hydrologic properties on contaminant processing.

The effects of non-hydrologic properties on in-stream removal of contaminant mass can be evaluated in SPARROW using discrete functional forms of the mass-transfer function by estimating separate mass-transfer coefficients for stream reaches that are classified according to discrete intervals of these properties. In evaluating temperature effects, for example, streams could be classified according to two classes—those having high and low water temperature—for which two separate mass-transfer coefficients are estimated. A larger mass-transfer coefficient for streams with warmer water temperatures than that estimated for cooler water temperatures would be theoretically consistent with the expectation that reaction kinetics increase with temperature. Alternatively, continuous functions describing the effects of non-hydrologic properties on contaminant loss can also be used in combination with equation (1.32)—see, for example, the temperature expression used in modeling contaminant removal rates in reservoirs in the subsequent section (1.3.1.4). SPARROW can also support the use of continuous functions of non-hydrologic properties, such as temperature, in combination with the discrete and continuous depth-dependent reaction rate expressions in equations (1.30) and (1.31).

As a final note, the discrete removal function in equation (1.30) can theoretically be modified in SPARROW to address the subjectivity of manually evaluating breakpoints for the stream size classes (i.e., based on flow or depth). The modification involves estimating separate reaction rate coefficients for the smallest and largest rivers according to the discrete functional form in equation (1.30) and a continuous function for intermediate-sized streams, based on a mass-transfer rate expression. In this formulation, SPARROW explicitly

estimates the appropriate stream size class breakpoints defined according to the stream depth. Note that whereas this technique performs well on simulation datasets, we have found that real datasets lack sufficient information to reliably estimate the flow breakpoints. We nevertheless present this approach to illustrate the flexibility of SPARROW model specifications. The process includes an additive coefficient (reaction rate) and an additional mass-transfer coefficient multiplied by stream depth. The functional form of the relation is

$$A\left(\mathbf{Z}_i^S, \mathbf{Z}_i^R; \boldsymbol{\theta}_S, \boldsymbol{\theta}_R\right) = \begin{cases} \exp\left[-\left(\theta_{S0} + \dfrac{\theta_{S1}}{D_{\text{low}}}\right) T_i^S\right], & \text{if } D_i^S \le D_{\text{low}}^S \\ \exp\left[-\left(\theta_{S0} + \dfrac{\theta_{S1}}{D_i^S}\right) T_i^S\right], & \text{if } D_{\text{low}}^S < D_i^S \le D_{\text{hi}}^S \\ \exp\left[-\left(\theta_{S0} + \dfrac{\theta_{S1}}{D_{\text{hi}}}\right) T_i^S\right], & \text{if } D_i^S > D_{\text{hi}}^S \end{cases} \tag{1.33}$$

where D_i^S is stream depth for reach *i*, D_{low}^S and D_{hi}^S are lower and upper cut-offs for stream depth (representing coefficients to be estimated), and θ_{S0} and θ_{S1} are the reaction rate and mass-transfer coefficients, respectively. The decay rate is multiplied by the negative of reach time of travel and exponentiated to determine the fraction of flux entering the reach from upstream that is delivered to the downstream node.

1.4.5 Reservoir/lake transport

Attenuation processes that act on contaminant mass as it travels through a lake or reservoir are frequently modeled according to a net removal process, with the loss coefficient expressed as either a first-order *reaction rate* or a *mass-transfer coefficient* (also referred to as an *apparent settling velocity* in the lake literature) (Chapra, 1997). Both of these expressions of the loss processes have been used in empirical mass-balance lake models for phosphorus (e.g., Vollenweider, 1976; Reckhow and Chapra, 1983) and nitrogen (e.g., Kelly and others, 1987; Molot and Dillon, 1993), although preference has been generally given to the use of the mass-transfer rate expression. These mass-balance models typically assume steady-state and uniformly mixed conditions in the waterbody. The *reaction rate* is expressed in reciprocal time units and its estimation and use is dependent on knowledge of the water depth and surface area of the waterbody (i.e., water volume). The *apparent settling velocity* is expressed in units of length per time and is estimated as a function of the ratio of the outflow rate for the reservoir and the surface area of the reservoir sediments (which are assumed to be equivalent to the surface area of the waterbody); this ratio, denoted q_i^R for reach *i*, is termed the areal hydraulic load and is a measure of the water displacement or velocity in the reservoir. The term "apparent" indicates that the settling velocity measures the net effect of various processes that remove the contaminant from the water column and deliver it to the sediments, and processes that may add contaminant to the water column (e.g., nitrogen fixation; nitrogen mineralization; phosphorus dissolution and resuspension).

Although the *reaction rate* and *settling velocity* expressions can be shown to be mathematically equivalent (and have been found to be statistically indistinguishable based on evaluations of the literature data on lakes; Reckhow and Chapra, 1983), preference has frequently been given to the use of a settling velocity expression in many of the empirical models (Chapra, 1997). This is because this loss expression generally provides a more accurate conceptual description of the mechanics of the flux (i.e., mass transfer) of nutrients and many other contaminants to the benthic sediments and their removal from the water column (e.g., via particulate settling, algal uptake and burial, and benthic denitrification). Moreover, as discussed in the previous section, this depth-independent measure of the removal rate provides a more explicit description of the effect of biological properties on contaminant removal. For practical purposes in SPARROW modeling, the use of a measure of contaminant loss that is independent of the water-column volume or depth of a lake or reservoir is advantageous because this information can sometimes be difficult to obtain from public databases. Data on the surface area of reservoirs are typically more readily available from public databases and, perhaps most importantly, can be digitized directly from topographic maps or easily obtained from DEMs.

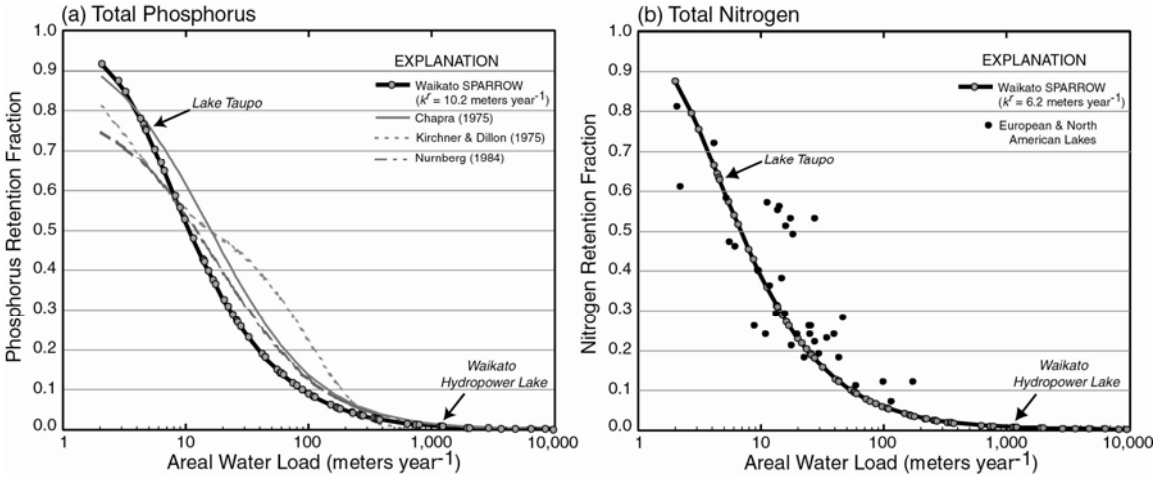


Figure 1.18. Nutrient removal in lakes and reservoirs of the Waikato River Basin, New Zealand, expressed as a fraction of the nutrient inputs to the water body in relation to the areal hydraulic load for (a) total phosphorus and (b) total nitrogen. [Modified from Alexander, Elliott, and others (2002).]

In SPARROW models, we have more recently used an *apparent settling velocity* expression to describe the loss of contaminant mass in lakes and reservoirs. Several recent SPARROW models illustrate (fig. 1.18) the use of this expression for total nitrogen (Alexander, Elliott, and others, 2002; McMahon and others, 2003; Moore and others, 2004) and total phosphorus (Alexander, Elliott, and others, 2002; Alexander and others, 2004). The fraction of nutrient inputs that are removed in reservoirs as shown in figure 1.18 is estimated as a function of an estimated settling velocity rate and the measured areal hydraulic load; nutrient removal declines in more rapidly flushed reservoirs and shows generally close agreement with the predictions of other models and measurements of the nitrogen removal fraction reported for selected lakes in the literature. Estimates of the settling velocities for total nitrogen and total phosphorus in U.S. and New Zealand SPARROW models are less than 20 meters year^{-1}, which agree well with the reported range of literature values. The SPARROW removal rates reflect the net annual removal of nutrients, and therefore, reflect the net balance of processes that supply (e.g., nitrogen fixation, phosphorus resuspension) and remove nutrients from the water column of reservoirs and lakes. For nitrogen, settling velocities of less than about 10 meters year^{-1} have frequently been observed in lakes where denitrification (rather than algal uptake) is known to be the predominant removal process; algal uptake is generally a more efficient removal process than denitrification and is typically consistent with settling rates of more than 25 meters year^{-1} (Alexander, Elliott, and others, 2002). Higher reservoir settling velocities (150-300 meters year^{-1}) have been estimated with SPARROW for suspended sediment (Schwarz and others, 2001) and fecal bacteria (Smith and others, 2004), which are consistent with the settling rates reported for clays/colloids and algal cells with sizes similar to those of pathogenic indicator bacteria (Chapra, 1997).

Other types of loss expressions may be more appropriate for other modeled contaminants and conditions and can be easily applied in SPARROW models, including plug-flow type expressions (Chapra, 1997). These may be more appropriate to describe the exponential loss of contaminant with distance that can occur in reservoirs with large length to width ratios (e.g., Higgins and Kim, 1981). Loss expressions may also be modified in SPARROW to include the potential effects of temperature or other reservoir properties as described below.

In the current SPARROW model specification, the fraction of the contaminant mass originating from the upstream reach node and transported through the reservoir segment of reach *i* to its downstream node is estimated as a function of the reciprocal of the areal hydraulic load $\left(q_i^R\right)^{-1}$ (units of length time^{-1}) for the reservoir associated with reach *i* and an apparent settling velocity coefficient (θ_{R0} ; units of length time^{-1}) (e.g., see Reckhow and Chapra, 1983), such that

(1.34) $$A\left(\mathbf{Z}_i^S, \mathbf{Z}_i^R; \boldsymbol{\theta}_S, \boldsymbol{\theta}_R\right) = \frac{1}{1+\theta_{R0}\left(q_i^R\right)^{-1}}.$$

The areal hydraulic load is computed as the quotient of the outflow discharge to surface area of the impoundment. This expression assumes that the surface area of the impoundment accurately reflects the surface area of the benthic sediments.

Equation (1.34) can be modified to evaluate how the apparent settling rate coefficient might change in response to the effects of non-hydrologic properties of reservoirs, such as temperature, concentrations of dissolved oxygen or nutrients, or the organic carbon content of the sediments—all of which can regulate bio-chemically mediated reactions such as benthic denitrification. *Discrete* functions of these properties can be evaluated by estimating separate settling rate coefficients for reservoirs classified according to the specified properties of interest. For example, reservoirs could be classified according to two classes—those having high and low water temperature—for which two separate settling rate coefficients are estimated. A larger settling rate in reservoirs with warmer water temperatures than that for cooler water temperatures would be consistent with the expectation that reaction kinetics increase with temperature.

Alternatively, equation (1.34) can be modified to account for the *continuous* effects of water temperature on contaminant loss, such that

(1.35) $$A\left(\mathbf{Z}_i^S, \mathbf{Z}_i^R; \boldsymbol{\theta}_S, \boldsymbol{\theta}_R\right) = \frac{1+\theta_{R1}^{C_i^R-20}}{1+\theta_{R0}\left(q_i^R\right)^{-1}},$$

where $1+\theta_{R1}^{C_i^R-20}$ is a simplified temperature-dependent expression of the Arrhenius equation (Chapra, 1997) for observed values of temperature C_i^R in units of Centigrade for the reservoir segment of the i^{th} reach and θ_{R1} is the estimated Arrhenius coefficient. Values of θ_{R1} above zero indicate a positive relation between the loss rate and temperature (the expected sign of most temperature-dependent reactions); values below zero indicate a negative relation. Note that the Arrhenius coefficient is typically evaluated in relation to unity rather than zero; a numerical value of 1.0 has been added to equation (1.35) so that the t-statistic accurately evaluates deviations of the coefficient from zero. Also note that the Arrhenius expression used here may also be applied to the stream decay functions discussed in the previous section of this report.

Previous SPARROW models of total nitrogen (Alexander, Elliott, and others, 2002; McMahon and others, 2003) estimated the nutrient transport in reservoirs as an exponential expression of the apparent settling velocity coefficient and areal hydraulic load, such that $A\left(\mathbf{Z}_i^S, \mathbf{Z}_i^R; \boldsymbol{\theta}_S, \boldsymbol{\theta}_R\right) = \exp\left(-\theta_{R0}\left(q_i^R\right)^{-1}\right)$. This expression was necessary because of the path-dependent calculations of stream reach properties required in earlier versions of the SPARROW infrastructure (i.e., earlier models required linear specifications of the aquatic loss terms in the exponential because reach properties, such as water time of travel and areal hydraulic load, were summed along the entire flow path between monitoring stations and upstream reaches). Reach-level flow path calculations are now used in the current version of SPARROW infrastructure and the model can support the nonlinear specification given in equations (1.34) and (1.35).

Contaminant removal in reservoirs may also be modeled according to a *depth-dependent reaction rate* expression similar to that in equation (1.30). Accordingly, the fraction of the contaminant mass originating from the upstream reach node and transported to its downstream node through the reservoir associated with reach *i* is estimated as a function of the mean water residence time ($T_{c\,i}^R$; units of time) in the reservoir associated with reach *i* and the reservoir size class *c* defined according to discrete intervals of the mean outflow rate of the reservoir ($T_{c\,i}^R$, $c = 1, \ldots, C_R$, is nonzero only for the class corresponding to the outflow rate of reservoir reach *i*), and a reservoir-size dependent loss rate coefficient (θ_{Rc}; units of time^{-1}) such that

$$A\left(\mathbf{Z}_i^S, \mathbf{Z}_i^R; \boldsymbol{\theta}_S, \boldsymbol{\theta}_R\right) = \exp\left(-\sum_{c=1}^{C_R} \theta_{Rc} T_{c\ i}^R\right). \tag{1.36}$$

This type of expression was evaluated in the earliest SPARROW model specifications for nutrients (Smith and others, 1997); however, an expression that used the much shorter mean water time of travel in the reaches associated with reservoirs was found to provide a better fit to the data for total phosphorus than an expression based on the available estimates of the water residence time in the reservoirs.

1.4.6 Regional model coefficients and nested model designs

SPARROW models can be used to evaluate the regional influence of watershed properties on in-stream pollutant flux. The estimated coefficients in the previous models are designed to provide measures of the net rates of contaminant supply and processing that apply to the entire modeled watershed area. There are instances, however, in which users may want to test whether the model coefficients associated with a given watershed property vary geographically over large and generally contiguous regions within the modeled area. These regions may describe the spatial distribution of explanatory properties of watersheds (e.g., soils, geology) that can be broadly grouped into contiguous geographic units. SPARROW can be used to test for geographic variations in the aggregate effect of various process rates (e.g., denitrification, storage) that may be associated with a particular watershed property and are reflected by the model coefficients. Separate coefficients may also be estimated and evaluated for geographic regions for which local management applications of the model are of interest. The spatially distributed model structure in SPARROW can support the development and testing of a spatially distributed set of parameters that is required to address these needs.

An example of the former case is that in which a physiographic region, such as a coastal plain or mountainous region, is believed or known to be associated with soil or geologic properties that potentially affect the removal and transport of a pollutant, such as nitrogen (McMahon and others, 2003; S. Preston, U.S. Geological Survey, written comm., 2004). Physiographic regions also may represent a geologic source of a pollutant, such as phosphorus in the case of the Tennessee/Kentucky regional SPARROW model (Hoos, 2005). These properties can be specified in SPARROW models as discrete explanatory variables that correspond to contiguous or nearly contiguous regional areas. These discrete variables are sometimes referred to as “indicator” or “dummy” variables, and are typically used in *analysis of covariance* regression methods (see Helsel and Hirsch, 1992). Such variables are incorporated into the model as binary (“0” or “1”) variables that represent different classifications of an explanatory factor and its associated model coefficient. Although regional specifications of SPARROW models are most reliably developed and tested using direct spatial measurements of watershed properties (e.g., soils, geology, physiography), regional patterns in model prediction errors—i.e., areas of consistent over- or under-prediction—may potentially provide insight into other explanatory factors and regional specifications of those factors that are not accounted for, but need to be included, in the model (e.g., McMahon and others, 2003).

One notable use of regionally specific coefficients in SPARROW models is to accommodate regional applications of the model (see figure 1.19), including the use of the model in a specific State, particular watersheds within a State, or by a regional multi-state management authority (e.g., New England Interstate Water Pollution Control Commission). These more geographically focused modeling interests have previously been addressed by developing regional SPARROW models, such as those in New England (Moore and others, 2004), the Chesapeake Bay (Preston and Brakebill, 1999), North Carolina coastal watersheds (McMahon and others, 2003), and watersheds in the states of Tennessee and Kentucky (Hoos, 2005). These models are typically estimated using local USGS and State stream monitoring data and may include spatial data on pollutant sources that have been developed by local political and scientific organizations.

As an alternative to these regional modeling approaches, users may also develop and test regional specifications within the existing national SPARROW models or even more locally specific specifications within regional models. We refer to these types of applications as *nested model designs* in which the local/regional model and its application are nested within a larger (e.g., national) spatial model structure. The advantage of this approach is twofold. First, this approach uses a much larger number of stream monitoring stations to estimate the model. For example, the local stream monitoring station flux estimates may be combined with the larger number of station flux estimates used to estimate the national model. The statistical power of such a model is generally much higher than can be achieved by using a conventional regional model based

exclusively on the local monitoring data. The estimation of regionally specific coefficients on selected explanatory variables in the model evaluates whether there is sufficient statistical evidence so that the local/regional response of in-stream flux to watershed properties differs from that estimated in the national model. Second, this approach enhances the ability of a user to account for the effects of an important watershed property for which there is too little regional variation to explicitly estimate its effects in a regional model. For example, some regions may contain too little spatial variation in mean-annual measures of atmospheric nitrogen deposition to statistically estimate a coefficient in a regional model, although this source may contribute appreciable quantities of nitrogen to streams in the region. An estimate of the contribution of the source in regional watersheds can be obtained directly from the national model. Therefore, a coupling of the regional and national models can provide a more comprehensive accounting of the sources of pollutants than can be achieved through the independent use of a regional model.

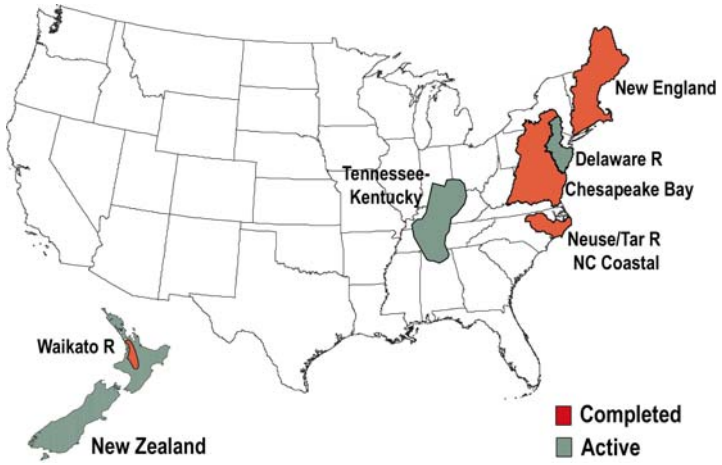


Figure 1.19. Completed and active SPARROW regional modeling applications.

The following example illustrates how a regional coefficient for in-stream contaminant removal would be specified in a national SPARROW model. In the U.S. national total nitrogen model, the estimated depth-dependent reaction rates (as shown in figure 1.17) describe the nitrogen removal rates that apply nationally to three different sizes of streams. There may be reason to believe that these removal rates might vary regionally in response to high dissolved organic carbon loads in low-gradient southeastern rivers or might differ in the northeastern U.S. where particular TMDL uses of the model are of interest (e.g., Moore and others, 2004). To evaluate such potential regional differences, the specification of the stream attenuation function described in equation (1.30) would be re-written as

$$A\left(\mathbf{Z}_i^S, \mathbf{Z}_i^R; \boldsymbol{\theta}_S, \boldsymbol{\theta}_R\right) = \exp\left(-\sum_{c=1}^{C_S}\left(\theta_{Sc} + \theta_{S\,c}^{\Delta}\eta_{c\,i}^{S}\right)T_{c\,i}^{S}\right), \tag{1.37}$$

where $\eta_{c\,i}^{S}$ is a regional binary variable for stream size class c in which the value of the variable for each reach is coded as "1" for reaches in the region of interest and "0" for all reaches in other geographic areas; and $\theta_{S\,c}^{\Delta}$ is the regional incremental stream attenuation coefficient expressed in units of day^{-1} that indicates the incremental difference in the rate of nitrogen removal within the streams of the region of interest (note that the regional nitrogen removal rate is computed as $\theta_{Sc} + \theta_{S\,c}^{\Delta}$, where $\theta_{S\,c}^{\Delta}$ can be either positive or negative). A formal hypothesis test (i.e., evaluation of t and p statistics; see subsequent section 1.5.4.1) of each of the regional incremental coefficients, $\theta_{S\,c}^{\Delta}$, would provide explicit information about whether statistically significant

differences are observed in the rates of in-stream removal of nitrogen within the specified region. The regional model specification would be based on a modified national estimation data file in which the regional stream monitoring flux estimates have been added to reaches. Moreover, if regional data are available for any of the explanatory variables (e.g., fertilizer use, soil permeability), then new explanatory variables would be created that reflect a combination of the national and regional data in which the regional data supersede the national data in the estimation file.

1.5 Model estimation

The SPARROW model equation, given in equation (1.27), is clearly a nonlinear function of its parameters. As such, the model must be estimated using nonlinear techniques. The errors of the model are assumed to be independent across observations and have zero mean; the variance of each observation may be observation specific. A general method commonly used for these types of problems, one in which it is not necessary to assume the precise distribution of the residuals, is nonlinear weighted least squares (NWLS). This is the estimation method used by SPARROW.

There are a number of algorithms in use for obtaining nonlinear weighted least squares estimates. This documentation will not delve into the details of any of these other than to identify the particular algorithm implemented in SAS to estimate the SPARROW model, the Levenberg-Marquardt Least-Squares Method, and to direct the reader to SAS documentation that provides additional information (SAS, 1999). In the following section, however, we describe how to think about nonlinear estimation in terms of an iterative linear estimation process. This discussion is useful for obtaining an intuitive understanding of the nonlinear estimation methodology and the diagnostic statistics generated by the SPARROW model, particularly as they relate to ordinary least squares.

Unlike linear models, the statistical properties of the estimated parameters for nonlinear models are not precisely known in finite samples. For this reason, much of the theory of nonlinear estimation has focused on characterizing asymptotic properties—the statistical properties of the coefficient estimates as the sample size goes to infinity. A discussion of these properties and of the restrictions of the model necessary to obtain them is given in section 1.5.2. Section 1.5.3 follows on this discussion by providing a description of bootstrap methods that are used to infer some of the small sample properties of the estimates—allowing the user to increase or decrease their confidence in the reliability of the asymptotic properties as they pertain to their particular SPARROW application.

An important topic for the practical user of SPARROW is how to interpret SPARROW model estimation results. Section 1.5.4 provides details on what to look for in the model estimates and how to interpret what is found. Sections 1.5.5 and 1.5.6 impart additional insight, of a practical nature, into the lessons to be learned from plotting the residuals and how to know if the model has adequate fit.

1.5.1 A guide to nonlinear estimation

The estimation method of nonlinear weighted least squares is not generally as well known as that for ordinary least squares (OLS). In order to improve interpretation of nonlinear model estimation, the following describes the close relation between nonlinear weighted least squares and the ordinary least squares method. In particular, it is shown that nonlinear weighted least squares can be interpreted simply as an iterative application of ordinary least squares. As such, many of the concepts developed from ordinary least squares analysis are directly transferable to the estimation of nonlinear models. This understanding is useful for understanding the derivation of the covariance matrix and leverage factor under nonlinear weighted least squares.

1.5.1.1 Brief review of ordinary least squares

A useful way to begin the discussion is to briefly review the results of ordinary least squares analysis. Consider a linear model, expressed in matrix notation, defined as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}\,, \tag{1.38}$$

where $\mathbf{y}$ is an N by 1 vector containing the dependent variable data, $\mathbf{X}$ is an N by K matrix containing the explanatory variable data, where K is the number of explanatory variables, $\boldsymbol{\beta}$ is a K by 1 vector of unknown

coefficients, and **e** is an *N* by 1 vector of errors assumed to be mutually uncorrelated with zero mean and variance σ_e^2. The estimated values of the coefficients are given by

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}. \tag{1.39}$$

Associated with the least squares methodology are a number of statistics that are useful for interpreting the statistical significance of the estimates and diagnosing potential problems in the data. These statistics include the covariance matrix and the leverage factor. The estimated covariance matrix of the estimated coefficients is given by

$$V\left(\hat{\boldsymbol{\beta}}\right) = s^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}, \tag{1.40}$$

where s^2 is the estimated error variance σ_e^2, or mean squared error given by

$$s^2 = \frac{\mathbf{y}'\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}\right)\mathbf{y}}{N-K}, \tag{1.41}$$

where **I** is the $N \times N$ identity matrix.

The square root of the diagonal elements of the covariance matrix represent the standard errors of the coefficients and are used to compute the *t*-statistics and probability values. The *t*-statistic for coefficient *k*, denoted $t\left(\hat{\beta}_k\right)$, is given by the ratio of the coefficient estimate to its standard error

$$t\left(\hat{\beta}_k\right) = \frac{\hat{\beta}_k}{\sqrt{V_{kk}\left(\hat{\boldsymbol{\beta}}\right)}}, \tag{1.42}$$

where $V_{kk}\left(\hat{\boldsymbol{\beta}}\right)$ corresponds to the k^{th} diagonal element of the covariance matrix $V\left(\hat{\boldsymbol{\beta}}\right)$. Under the hypothesis that the true value of $\hat{\beta}_k$ is zero, the two-sided probability value for $\hat{\beta}_k$, denoted $p\left(\hat{\beta}_k\right)$, corresponds to the probability of obtaining a *t*-statistic having an absolute value that is greater than or equal to the absolute value of $t\left(\hat{\beta}_k\right)$. If the errors **e** are normally distributed, and the hypothesis that $\hat{\beta}_k$ equals zero is true, the *t*-statistic in equation (1.42) is distributed Student's *t* with $N-K$ degrees of freedom and the two-sided probability value is given by

$$p\left(\hat{\beta}_k\right) = 2\left(1 - \mathrm{T}_{(N-K)}\left(\left|\frac{\hat{\beta}_k}{\sqrt{V_{kk}\left(\hat{\boldsymbol{\beta}}\right)}}\right|\right)\right), \tag{1.43}$$

where $\mathrm{T}_{(N-K)}(\cdot)$ is the Student's *t* cumulative distribution function for *N–K* degrees of freedom.

The confidence interval for β_k is defined as the random interval having a probability P_c, called the coverage probability or confidence level, of including the true coefficient value. Under the hypothesis that the errors are normally distributed, the statistic $\left(\hat{\beta}_k - \beta_k\right) \Big/ \sqrt{V_{kk}\left(\hat{\boldsymbol{\beta}}\right)}$ is distributed $\mathrm{T}_{(N-K)}(\cdot)$. Let $\mathrm{T}^{-1}_{(N-K)}(p)$ be the quantile of the Student's *t* distribution correspond to the probability *p*. Upon careful manipulation of the

probability statement $\Pr\left(\left|\left(\hat{\beta}_k-\beta_k\right)\Big/\sqrt{V_{kk}\left(\hat{\boldsymbol{\beta}}\right)}\right|\le \mathrm{T}^{-1}_{(N-K)}\left(\left(1-P_c\right)/2\right)\right)=P_c$, we obtain the confidence interval for β_k given by

$$\Pr\left(\hat{\beta}_k-\sqrt{V_{kk}\left(\hat{\boldsymbol{\beta}}\right)}\mathrm{T}^{-1}_{(N-K)}\left(\left(1+P_c\right)/2\right)\le\beta_k\le\hat{\beta}_k-\sqrt{V_{kk}\left(\hat{\boldsymbol{\beta}}\right)}\mathrm{T}^{-1}_{(N-K)}\left(\left(1-P_c\right)/2\right)\right)=P_c. \tag{1.44}$$

The leverage statistic is used to determine observations that have undue influence in the determination of model coefficients. Formally, the leverage of observation *i* is given by

$$h_i=\frac{V\left(\hat{e}_i-e_i\right)}{\sigma^2}=\frac{V\left(\mathbf{X}_i\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\mathbf{X}_i'\right)}{\sigma^2}=\mathbf{X}_i\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}_i'. \tag{1.45}$$

The leverage statistic is always positive and has an average value over all observations equal to K/N. A "rule of thumb" criterion for identifying a high influence observation is one in which $h_i>3K/N$.

1.5.1.2 Nonlinear weighted least squares

Each of the statistical results stated above has an analog in a nonlinear weighted least squares analysis. In order to better understand the derivation of those results, we show how the method of nonlinear weighted least squares can be framed as an iterative application of ordinary least squares.

To focus the discussion, the extensive notation in equation (1.27) is greatly simplified. The objective of estimation is to explain measured flux, so the abbreviated notation explicitly references only monitored reaches $i\in I$. For convenience, we also refer to the monitoring stations associated with monitored reach *i* as monitoring station *i*. Let measured flux at reach *i* continue to be denoted as F_i^M. The abbreviated notation for modeled flux is $F_i^*\left(\boldsymbol{\beta}\right)$, where $\boldsymbol{\beta}=\left\{\boldsymbol{\alpha}',\boldsymbol{\theta}_D',\boldsymbol{\theta}_S',\boldsymbol{\theta}_R'\right\}'$ is a vector of all the model coefficients. Included as determinants of $F_i^*\left(\boldsymbol{\beta}\right)$, but not explicitly stated, are the source, land-to-water delivery and in-stream attenuation variables that determine flux within the nested basin of monitoring station *i*. Also included are any measured fluxes upstream of reach *i* that define the boundary of monitoring station *i*'s nested basin.

The relation between measured (F_i^M) and monitored ($F_i^*\left(\boldsymbol{\beta}\right)$) flux for monitored reach *i* is assumed to be given by

$$\ln\left(F_i^M\right)=\ln\left(F_i^*\left(\boldsymbol{\beta}\right)\right)+e_i, \tag{1.46}$$

where e_i is a random error that is independent across monitored basins $i=1,\ldots,N$, with mean zero and variance σ_i^2. The assumption that the random error is additive in logarithm space, implying it is multiplicative in real space, is reasonable given the non-negativity of flux. Less defensible is the assumption that the error terms are independent across monitored basins. There are, undoubtedly, unobserved factors, such as unmeasured climate factors, that would cause errors between basins to be correlated. Unfortunately, the additional model structure required to incorporate that correlation is beyond the reach of the current SPARROW model. For now, we remark that correlated residuals do not lead to biased estimates of the coefficients, although they can potentially bias the covariances of the estimates and make the estimation methodology described below inefficient. Note also that equation (1.46) is not a conventional log-linear regression on basin attributes. The

logarithm function on the right hand side of equation (1.46) pertains to an overall transformation of all the additive terms on the right hand side of equation (1.27). As such, mass balance continues to hold.

An appropriate estimation methodology for models with independent, heteroscedastic (that is, not identically distributed) errors is nonlinear weighted least squares (Judge and others, 1985, Gallant, 1987), the objective of which is the minimization of

$$Q = \sum_{i \in I} w_i e_i^2 , \tag{1.47}$$

with respect to $\boldsymbol{\beta}$, where w_i is the weight applied to monitored reach *i*. [Technically, the presence of upstream monitored fluxes in $F_i^*(\boldsymbol{\beta})$ makes the estimation problem more complicated than it appears here. Some of these complications are addressed in section 1.5.2.1.]

To show the relation with ordinary least squares, we simplify the notation further and write the model for all *N* monitored basins as

$$\mathbf{f}^M = \mathbf{f}^*(\boldsymbol{\beta}) + \mathbf{e}, \tag{1.48}$$

where $\mathbf{f}^M$ and $\mathbf{f}^*(\boldsymbol{\beta})$ are each $N \times 1$ vectors having elements $\ln\left(F_i^M\right)$ and $\ln\left(F_i^*(\boldsymbol{\beta})\right)$, $i = 1, \ldots, N$, and $\mathbf{e}$ is an $N \times 1$ vector of the corresponding error terms. The estimation of the nonlinear model in equation (1.48) using OLS methods is achieved by taking a first-order Taylor approximation around an initial parameter vector $\boldsymbol{\beta}_0$

$$\mathbf{f}^M = \mathbf{f}^*(\boldsymbol{\beta}_0) + \mathbf{f}_{\boldsymbol{\beta}'}^*(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \mathbf{e}, \tag{1.49}$$

where $\mathbf{f}^*(\boldsymbol{\beta}_0)$ is the $N \times 1$ vector function evaluated at the initial coefficient estimate $\boldsymbol{\beta}_0$, and $\mathbf{f}_{\boldsymbol{\beta}'}^*(\boldsymbol{\beta}_0)$ is the $N \times K$ matrix of partial derivatives of $\mathbf{f}^*$ with respect to $\boldsymbol{\beta}$, also called the gradient of $\mathbf{f}^*$, evaluated at the initial coefficient estimate $\boldsymbol{\beta}_0$.

An assumption of the SPARROW model is that e_i, the error of observation *i*, has an observation-specific variance σ_i^2. One implication of this assumption is that the ordinary least squares method of estimation is inefficient. Additionally, if the observation-specific variance is correlated with any of the explanatory variables, the standard estimation of the coefficient covariance matrix is biased (White, 1980). An appropriate approach to estimation in this case is weighted least squares or, in the context of the nonlinear model, weighted nonlinear least squares. The optimal weight for observation *i* is proportional to the inverse of the variance of observation *i*'s error, such that $w_i \propto 1/\sigma_i^2$ (Judge and others, 1985). To retain the average variance of the error in the weighted procedure, the proportionality constant used to compute w_i is the average of the error variance, resulting in

$$w_i = \frac{\sum_{j=1}^{N} \sigma_j^2}{N \sigma_i^2}. \tag{1.50}$$

The close link between weighted nonlinear least squares and ordinary least squares becomes clear by applying weights to both sides of equation (1.49) and transforming terms as follows

$$\tilde{\mathbf{y}}_0 = \mathbf{D}(\mathbf{w})^{1/2}\left(\mathbf{f}^M - \mathbf{f}^*(\boldsymbol{\beta}_0) + \mathbf{f}^*_{\boldsymbol{\beta}'}(\boldsymbol{\beta}_0)\boldsymbol{\beta}_0\right),$$

(1.51) $$\tilde{\mathbf{X}}_0 = \mathbf{D}(\mathbf{w})^{1/2}\mathbf{f}^*_{\boldsymbol{\beta}'}(\boldsymbol{\beta}_0), \text{ and}$$

$$\tilde{\mathbf{e}} = \mathbf{D}(\mathbf{w})^{1/2}\mathbf{e},$$

where $\mathbf{D}(\mathbf{w})^{1/2}$ is a diagonal matrix with the vector of the square roots of the weights w_i along the diagonal. Given $\boldsymbol{\beta}_0$, the next iteration's estimate of $\boldsymbol{\beta}$ is obtained by estimating the transformed equation

(1.52) $$\tilde{\mathbf{y}}_0 = \tilde{\mathbf{X}}_0\boldsymbol{\beta} + \tilde{\mathbf{e}}.$$

Equation (1.52) is linear in $\boldsymbol{\beta}$ and can be easily estimated using ordinary least squares to obtain the next iteration's estimate of the coefficient vector, denoted $\boldsymbol{\beta}_{+1}$,

(1.53) $$\hat{\boldsymbol{\beta}}_{+1} = \left(\tilde{\mathbf{X}}_0'\tilde{\mathbf{X}}_0\right)^{-1}\tilde{\mathbf{X}}_0\tilde{\mathbf{y}}_0.$$

The above discussion suggests two iterative procedures for obtaining the nonlinear weighted least squares estimates, the first to be used if the weights are known and the second to be used if the weights are unknown. Both procedures begin with a selection of initial coefficient estimates $\boldsymbol{\beta}_0$. If the weights are known, the initial coefficient estimates and known weights are used to compute the transformations given in equation (1.51) and estimate a new coefficient vector $\hat{\boldsymbol{\beta}}_{+1}$ by estimating equation (1.52) using ordinary least squares. The coefficient vector $\hat{\boldsymbol{\beta}}_{+1}$ is then used to define a new set of initial estimates $\boldsymbol{\beta}_0$, reevaluate the transformations in equation (1.51) and obtain a new coefficient vector $\hat{\boldsymbol{\beta}}_{+1}$. This procedure is continued until the change in the estimated vector $\hat{\boldsymbol{\beta}}_{+1}$ between iterations meets a specified convergence criterion.

If the weights are unknown, it may be possible to estimate them from an estimate of the un-weighted residuals $\hat{\mathbf{e}}$. In the first step, the nonlinear model is estimated with weights set to 1.0 to obtain consistent estimates of the errors,

(1.54) $$\hat{\mathbf{e}} = \mathbf{f}^M - \mathbf{f}^*\left(\hat{\boldsymbol{\beta}}\right),$$

where $\hat{\boldsymbol{\beta}}$ is the vector of parameter estimates obtained from the converged iterative procedure without weighting. In a second step, the estimated errors can be used to estimate a functional relation with additional explanatory variables **W**. Let the relation between error variance and the explanatory variables be given by

(1.55) $$\sigma_i^2 = g(\mathbf{W}_i, \boldsymbol{\gamma}) + v_i,$$

where $\boldsymbol{\gamma}$ is a vector of unknown parameters and $\mathbf{W}_i$. It is assumed the errors v_i are independent of $\mathbf{W}_i$ and of the gradient $\mathbf{f}^*_{\boldsymbol{\beta}'}(\boldsymbol{\beta})$ evaluated at the actual coefficient values. The $\boldsymbol{\gamma}$ coefficients in equation (1.55) can be estimated by nonlinear regression, with $\hat{e}_i^2$ substituting for σ_i^2 as the dependent variable. The second step of the procedure is to estimate the nonlinear weighted least squares model using weights

$$\hat{w}_i = \frac{\sum_{j=1}^{N} g\left(\mathbf{W}_j, \hat{\boldsymbol{\gamma}}\right)}{N g\left(\mathbf{W}_i, \hat{\boldsymbol{\gamma}}\right)}. \tag{1.56}$$

The two-step procedure is consistent and efficient in large samples (White, 1980).

White (1980) suggests a general procedure that does not require an explicit model of error variance. Under his approach, the modeled component of error variance takes the form $g\left(\mathbf{W}_i, \boldsymbol{\gamma}\right) = \mathbf{W}_i \boldsymbol{\gamma}$, where $\mathbf{W}_i$ is a row vector consisting of the row vector $\mathbf{f}^*_{\boldsymbol{\beta}'}\left(\hat{\boldsymbol{\beta}}_0\right)$ and all unique convolutions of $\mathbf{f}^*_{\boldsymbol{\beta}'}\left(\hat{\boldsymbol{\beta}}_0\right)$, where $\hat{\boldsymbol{\beta}}_0$ is a vector of coefficient estimates from the first step. The selection of weights in this way results in consistent and an asymptotically efficient estimator for the coefficients. In practice, however, it is generally impossible to fully specify the $\mathbf{W}_i$ vector in this way due to the large number of terms implied by the convolution. A practical approach, therefore, may be to limit the convolutions to those terms showing the greatest correlation with the squared residuals.

1.5.1.3 The asymptotic covariance matrix

Following the least squares analogy, the estimated covariance matrix of the nonlinear weighted least squares coefficients is given by

$$\mathbf{V}\left(\hat{\boldsymbol{\beta}}\right) = s^2 \left(\tilde{\mathbf{X}}'_* \tilde{\mathbf{X}}_*\right)^{-1}, \tag{1.57}$$

where $\tilde{\mathbf{X}}_*$ is the transformed value of $\tilde{\mathbf{X}}$ given in equation (1.51) evaluated at the converged parameter estimates $\hat{\boldsymbol{\beta}}$, and

$$s^2 = \frac{\sum_{i=1}^{N} \hat{w}_i \left(f_i^M - f_i^*\left(\hat{\boldsymbol{\beta}}\right)\right)^2}{N} \tag{1.58}$$

is the mean squared weighted error for the model (colloquially referred to as the mean squared error or MSE).

1.5.1.4 Estimation of leverage in the nonlinear model (advanced)

The leverage statistic, which is used in ordinary least squares to identify observations that exert disproportionate influence on the estimated values of the coefficients, has an asymptotic analog under nonlinear least squares. For the nonlinear model, leverage is derived analogously to equation (1.45) by considering the large sample limit of the sampling variance of the weighted estimated error

$$\lim_{N\to\infty} N h_i = \lim_{N\to\infty} \frac{V\left(\sqrt{N} w_i \left(\hat{e}_i - e_i\right)\right)}{\sigma_N^2} = \lim_{N\to\infty} \frac{V\left(\sqrt{N} w_i \left(f_i^*\left(\hat{\boldsymbol{\beta}}\right) - f_i^*\left(\boldsymbol{\beta}\right)\right)\right)}{\sigma_N^2}, \tag{1.59}$$

where $\sigma_N^2 = 1/N \sum_{i=1}^{N} \sigma_i^2$. The large sample limit of equation (1.59) can be determined by taking a first-order Taylor expansion of $\sqrt{N} f_i^*\left(\hat{\boldsymbol{\beta}}\right)$ about $\boldsymbol{\beta}$, yielding

$$\sqrt{N}\left(f_i^*\left(\hat{\boldsymbol{\beta}}\right) - f_i^*\left(\boldsymbol{\beta}\right)\right) = f^*_{i\boldsymbol{\beta}'}\left(\tilde{\boldsymbol{\beta}}\right) \sqrt{N}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right), \tag{1.60}$$

where $\tilde{\beta}_k = \lambda_k \hat{\beta}_k + (1-\lambda_k)\beta_k$, for some $\lambda_k \in [0,1]$, $k = 1,\ldots,K$. Under large sample theory, given certain regulatory conditions (see section 1.5.2), $\hat{\boldsymbol{\beta}}$ is consistent and the asymptotic distribution of $\sqrt{N}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)$ is normal with mean zero and covariance matrix $\lim_{N\to\infty} N\sigma_N^2 \left(\tilde{\mathbf{X}}_*'\tilde{\mathbf{X}}_*\right)^{-1}$. The consistency of $\hat{\boldsymbol{\beta}}$ implies the probability limit of $f_{i\boldsymbol{\beta}'}^*\left(\hat{\boldsymbol{\beta}}\right)$ equals $f_{i\boldsymbol{\beta}'}^*\left(\boldsymbol{\beta}\right)$ (see Rao, 1973, p. 122), which, by the definition of $\tilde{\boldsymbol{\beta}}$, implies the probability limit of $f_{i\boldsymbol{\beta}'}^*\left(\tilde{\boldsymbol{\beta}}\right)$ also equals $f_{i\boldsymbol{\beta}'}^*\left(\boldsymbol{\beta}\right)$. Consistency also implies that the probability limit of the estimated weights $\hat{w}_i$ is w_i. Consequently,

$$\lim_{N\to\infty} \frac{V\left(\sqrt{N} w_i \left(f_i^*\left(\hat{\boldsymbol{\beta}}\right) - f_i^*\left(\boldsymbol{\beta}\right)\right)\right)}{\sigma_N^2} = \lim_{N\to\infty} \tilde{\mathbf{X}}_{*i}\left[N\left(\tilde{\mathbf{X}}_*'\tilde{\mathbf{X}}_*\right)^{-1}\right]\tilde{\mathbf{X}}_{*i}', \tag{1.61}$$

where $\tilde{\mathbf{X}}_{*i}$ is the i^{th} row of the $\tilde{\mathbf{X}}_*$ gradient matrix.

It is straightforward to show that $\lim_{N\to\infty} N\sum_{i=1}^{N} h_i = K$. Thus, asymptotically, the nonlinear leverage statistic

$$h_i = \tilde{\mathbf{X}}_{*i}\left(\tilde{\mathbf{X}}_*'\tilde{\mathbf{X}}_*\right)^{-1}\tilde{\mathbf{X}}_{*i}' \tag{1.62}$$

retains the same interpretation as the leverage statistic defined under ordinary least squares, where the weighted gradient matrices $\tilde{\mathbf{X}}_*$ serve as nonlinear analogs to the design matrix $\mathbf{X}$.

1.5.1.5 Estimation of gradients (advanced)

As is evident from equation (1.51), the method of nonlinear optimization requires the computation of gradients, and the particular optimization routine used in SPARROW requires the computation of second-order derivatives. A SPARROW model represents a complex function of the coefficients making analytic first- and second-order derivatives $\mathbf{f}_{\boldsymbol{\beta}'}^*\left(\boldsymbol{\beta}_0\right)$ and $\mathbf{f}_{\boldsymbol{\beta}\boldsymbol{\beta}'}^*\left(\boldsymbol{\beta}_0\right)$ difficult to obtain. An alternative is to estimate the derivatives using a numerical method called *forward finite differences*. The first-order partial derivative of the function f_i^* with respect to β_k is given by

$$\frac{\partial f_i^*\left(\boldsymbol{\beta}_0\right)}{\partial \beta_k} \approx \frac{f_i^*\left(\boldsymbol{\beta}_0 + h_k \mathbf{e}_k\right) - f_i^*\left(\boldsymbol{\beta}_0\right)}{h_k}, \tag{1.63}$$

where $\mathbf{e}_k$ is a vector, of the same dimension as $\boldsymbol{\beta}_0$, having a 1.0 recorded in the k^{th} row and zeros in all other rows, and h_k is a step change given by $\sqrt{\eta}\left(1 + \left|\beta_{0k}\right|\right)$, where η is the machine precision (SAS, 1999). The forward difference estimate of the second-order derivative is

$$(1.64)\quad \frac{\partial^2 f_i^*(\boldsymbol{\beta}_0)}{\partial\beta_j\partial\beta_k} \approx \begin{cases} \dfrac{f_i^*(\boldsymbol{\beta}_0 + h_j'\mathbf{e}_j + h_k'\mathbf{e}_k) - f_i^*(\boldsymbol{\beta}_0 + h_j'\mathbf{e}_j) - f_i^*(\boldsymbol{\beta}_0 + h_k'\mathbf{e}_k) + f_i^*(\boldsymbol{\beta}_0)}{h_j' h_k'}, & j \neq k \\ \dfrac{f_i^*(\boldsymbol{\beta}_0 - h_j'\mathbf{e}_j) - 2f_i^*(\boldsymbol{\beta}_0) + f_i^*(\boldsymbol{\beta}_0 + h_j'\mathbf{e}_j)}{h_j'^2}, & j = k, \end{cases}$$

where h_k' is the step size for computing the second-order derivative, given by $\eta^{2/3} h_k$ (SAS, 1999).

A more accurate estimate of derivatives is possible using a central finite difference algorithm (SAS, 1999), but this method requires significantly more evaluations of the SPARROW objective function. For most applications, the forward finite difference algorithm is sufficient. In the interest of limiting computation time, this is the algorithm used.

1.5.2 Asymptotic properties of the estimators (advanced)

Unlike for linear models, the statistical properties of the estimated parameters for nonlinear models are not precisely known in finite samples. For example, in a linear model, if it is known that the errors are independent identically distributed Gaussian (that is, normal) and the predictors are non-stochastic, then the estimated parameters are normally distributed. Such a conclusion cannot be made when the parameters are estimated in a nonlinear model with a finite sample. Despite this limitation, quite a bit can be said about the statistical properties of the estimated parameters in a nonlinear model as sample size becomes large. Because these properties are attained only in large samples, they are referred to as *asymptotic properties*, and the derivation of these statistical properties is called an asymptotic analysis. Two statistical properties of the parameters require verification: consistency of the parameter estimators and the form of their asymptotic distribution. A parameter estimator is consistent if, as sample size becomes large, the estimates converge on the true parameter values. This convergence is described probabilistically: consistency is established if as sample size goes to infinity the probability goes to zero that the estimated parameters deviate from the true parameters by any specified amount $\varepsilon > 0$ (Amemiya, 1985, p. 95).

The second statistical property established in an asymptotic analysis is the distribution of a parameter estimator as sample size goes to infinity. One might think that if a parameter is consistent, then the asymptotic distribution consists of a mass of probability at the true parameter value and zero probability everywhere else. As such, the derivation of the asymptotic distribution is trivial. What is typically meant by an asymptotic distribution, however, is the distribution of a standardized estimate, which is formed as the difference between the parameter estimate and its true value, divided by the standard error of the estimate. That is, if we define a parameter estimator $\hat{\beta}_{kN}$ as the estimate of β_k from a sample of size *N*, then the standardized statistic $\left(\hat{\beta}_{kN} - \beta_k\right) \Big/ \sqrt{V_{kk}\left(\hat{\boldsymbol{\beta}}_N\right)}$ is the ratio of two terms that under certain conditions each go to zero as *N* goes to infinity. Because under general conditions both the numerator and denominator of this ratio have the same rate of convergence to zero as sample size gets large, the ratio itself has a limiting distribution that is non-degenerate. Indeed, one of the great achievements in statistics theory is the finding that this limiting distribution is the standard normal. This is a powerful result because it makes it possible, in the context of a nonlinear model, to use standard hypothesis tests based on the normal distribution to evaluate the estimated coefficients.

The attainment of asymptotic consistency and normality is not guaranteed but depends on underlying properties of the nonlinear model, properties that can be conservatively assessed in terms of a set of general sufficient conditions. These conditions are technical and in practice rarely given much attention: more commonly, as an afterthought, they are simply assumed to hold. There are, however, certain features of a hydrologic model that test the limits of these conditions. It is necessary, therefore, that we expend some effort discussing them.

The salient features of the SPARROW model that affect the manner of conditions that must be satisfied in order to attain consistency and asymptotic normality are these:

(a) the general form of a SPARROW model is nonlinear in terms of its parameters;

(b) a SPARROW model is estimated with restrictions on the values of its coefficients, it being possible for a true coefficient value to exactly satisfy its restriction;
(c) the error term is heteroscedastic;
(d) a SPARROW model is accumulative of flux within a given basin, with accumulation according to a dendritic pattern;
(e) the uncertainty associated with the accumulation of flux is such that it does not go to zero at the outflow of any basin and may scale with the size of the basin; and
(f) the modeled flux may depend on flux measurements from upstream basins.

Features (a) and (c) have already been discussed. Feature (b) refers to the fact that there are restrictions on some model coefficients; for example, source coefficients are generally restricted to be non-negative. Features (d), (e) and (f) are suggestive of a time series analysis in which the value of the dependent variable depends on an accumulated history of events, some of which can be summarized by previous values taken by the dependent variable, and uncertainty regarding the future value of the dependent variable becomes larger the further the future is extended. The comparison with time series analysis is only suggestive, however, owing to the dendritic nature of the accumulation.

We are not aware of any formal technical analysis in the literature that establishes sufficient conditions for consistency and asymptotic normality in the context of a hydrologic model having features (a)-(f). In lieu of this, we appeal to rather general conditions that have been established in the literature for nonlinear dynamic models in time, a particularly general statement of which has been made by Gallant and White (1988). The specific assumptions required to establish consistency and asymptotic normality as stated by Gallant and White are quite technical, serving to set limits such that it is possible to apply a law of large numbers—a general law that states that the sum of independent random numbers goes to a normal distribution. Here we give a non-technical account of what these assumptions accomplish.

First, it is assumed that the set of possible parameter values is compact, meaning that they are restricted to a definite range defined by closed operators such as '$\leq$' and '$\geq$' as opposed to open operators such as '$<$' and '$>$'. This is a technical restriction that can be imposed on the SPARROW model with little loss of generality.

A number of the assumptions stated by Gallant and White (1988) pertain to nonlinear models in general, models without a dynamic element. For example, these assumptions require the modeled flux function f^* to be compact over the range of admissible parameter values and to be continuously differentiable of order two for all possible values of the parameters, a condition that is also required in order to implement the iterative algorithm used to minimize the nonlinear weighted least squares objective function. The modeled flux function and its first and second derivatives are also assumed to be Lipschitz, a condition that bounds the smoothness of these functions in a somewhat stronger way than continuity. Generally, these conditions can be satisfied by any SPARROW model by appropriate specification of the admissible values of the parameters.

Consistency requires that the objective function in equation (1.47) is assumed to have identifiably unique minimizers $\boldsymbol{\beta}$ within the set of admissible parameters. Asymptotic normality requires the identifiably unique minimizers are in the *interior* of the set of admissible parameters. These are standard assumptions for nonlinear least squares estimators. The interior restriction would seem to rule out parameter constraints, or at least the attainment of these constraints by the estimated parameters. The restriction can be avoided, however, by assuming parameters estimated to exactly meet a constraint are treated as *known* coefficients fixed at their constrained values, in which case the parameter space is reduced to include only the unconstrained parameters. The remaining parameters would then be subject to the interior condition, but not the constrained parameter.

The weighted squared difference between measured flux and modeled flux, $w_i\left(f_i^M - f_i^*\left(\boldsymbol{\beta}\right)\right)^2$, and its first and second derivatives are also required to be stochastically bounded or 'dominated.' This is a technical restriction that places limits on the extent of variance in the heteroscedastic errors, as well as on the variation of scales of the watersheds included in the model. Because watersheds are physically bounded, this restriction essentially places limitations on the amount of variation allowed in the errors. An appropriate choice of weights should suffice to assure heteroscedasticity is appropriately bounded.

The other assumptions imposed by Gallant and White (1988) deal with limitations in the amount of dependence between temporally separated stochastic variables as the separation becomes large. Generally, these assumptions require that this dependence must go to zero at a sufficient rate. This would seem to have a bearing on hydrologic models applied to conservative contaminants, in which case the dependence between upstream

and downstream fluxes would not be expected to go to zero as distance between them is increased. As we show below, there are important reasons, of an asymptotic nature, that a hydrologic model be estimated on data that include non-nested basins, although including some nested basins is allowed. Because the dependence problem created by conservative contaminants is purely a consequence of nesting, the restriction that the model include non-nested basins helps alleviate the dependence problem if it can be assumed that other, non-modeled factors causing dependencies between basins, such as unobserved climate variables, satisfy the types of limitations assumed by Gallant and White. These assumptions thus reduce to the same types of assumptions that would be required to estimate any model composed of correlated unobserved factors. The assumption made for SPARROW in section 1.5.1 is that the errors across nested and non-nested basins are stochastically independent. This is an extreme assumption that is probably unfounded (see Qian, 2005) but suffices in the present context to ensure that fluxes across non-nested basins are also independent, thereby satisfying the assumptions of dependence imposed by Gallant and White.

1.5.2.1 Implications of asymptotic normality (advanced)

If the assumptions described above are satisfied, Gallant and White (1988) show that $\sqrt{N}\left(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}\right)$, where $\hat{\boldsymbol{\beta}}_N$ is the coefficient vector that minimizes equation (1.47), is asymptotically distributed normal with a vector mean of zero and covariance matrix given by

$$\lim_{N\to\infty} NV\left(\hat{\boldsymbol{\beta}}_N\right) = \lim_{N\to\infty} s_N^2 \left(N^{-1} \sum_{i\in I} w_i f_{i,\boldsymbol{\beta}}^*\left(\hat{\boldsymbol{\beta}}_N\right) f_{i,\boldsymbol{\beta}}^*\left(\hat{\boldsymbol{\beta}}_N\right)' \right)^{-1}, \tag{1.65}$$

where $s_N^2 = N^{-1}\sum_{i\in I} \hat{w}_i \left(f_i^M - f_i^*\left(\hat{\boldsymbol{\beta}}_N\right)\right)^2$, I represents the set of N monitored reaches, indexed by i, and $f_{i,\boldsymbol{\beta}}^*\left(\hat{\boldsymbol{\beta}}_N\right)$ is the derivative of the function defining modeled flux at reach i with respect to the coefficient vector $\boldsymbol{\beta}$, evaluated at the nonlinear weighted least squares coefficient estimates $\hat{\boldsymbol{\beta}}_N$.

This result can be used to justify estimates of the p-value and confidence interval based on the normal distribution. Given asymptotic normality, the standardized statistic for the k^{th} coefficient, $\left(\hat{\beta}_{kN} - \beta_k\right) \Big/ \sqrt{V_{kk}\left(\hat{\boldsymbol{\beta}}_N\right)}$, is distributed standard normal in large samples. Therefore, the p-value for the k^{th} coefficient is given by

$$p\left(\hat{\beta}_{kN}\right) = 2\left(1 - \Phi\left(\left|\hat{\beta}_{kN}\right| \Big/ \sqrt{V_{kk}\left(\hat{\boldsymbol{\beta}}_N\right)}\right)\right), \tag{1.66}$$

and the confidence interval with coverage probability P_c has lower and upper bounds $\underline{\hat{\beta}_{kN}}$ and $\overline{\hat{\beta}_{kN}}$ given by

$$\begin{aligned} \underline{\hat{\beta}_{kN}} &= \hat{\beta}_{kN} - \sqrt{V_{kk}\left(\hat{\boldsymbol{\beta}}_N\right)}\Phi^{-1}\left((1+P_c)/2\right), \text{ and} \\ \overline{\hat{\beta}_{kN}} &= \hat{\beta}_{kN} - \sqrt{V_{kk}\left(\hat{\boldsymbol{\beta}}_N\right)}\Phi^{-1}\left((1-P_c)/2\right), \end{aligned} \tag{1.67}$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function and $\Phi^{-1}(\cdot)$ is the inverse of the cumulative standard normal probability distribution.