The SPARROW Surface Water-Quality Model: Theory, Application and User Documentation

By G.E. Schwarz, A.B. Hoos, R.B. Alexander, and R.A. Smith

Abstract

SPARROW (SPAtially Referenced Regressions On Watershed attributes) is a watershed modeling technique for relating water-quality measurements made at a network of monitoring stations to attributes of the watersheds containing the stations. The core of the model consists of a nonlinear regression equation describing the non-conservative transport of contaminants from point and diffuse sources on land to rivers and through the stream and river network. The model predicts contaminant flux, concentration, and yield in streams and has been used to evaluate alternative hypotheses about the important contaminant sources and watershed properties that control transport over large spatial scales.

This report provides documentation for the SPARROW modeling technique and computer software to guide users in constructing and applying basic SPARROW models. The documentation gives details of the SPARROW software, including the input data and installation requirements, and guidance in the specification, calibration, and application of basic SPARROW models, as well as descriptions of the model output and its interpretation. The documentation is intended for both researchers and water-resource managers with interest in using the results of existing models and developing and applying new SPARROW models.

The documentation of the model is presented in two parts. Part 1 provides a theoretical and practical introduction to SPARROW modeling techniques, which includes a discussion of the objectives, conceptual attributes, and model infrastructure of SPARROW. Part 1 also includes background on the commonly used model specifications and the methods for estimating and evaluating parameters, evaluating model fit, and generating water-quality predictions and measures of uncertainty. Part 2 provides a user's guide to SPARROW, which includes a discussion of the software architecture and details of the model input requirements and output files, graphs, and maps. The text documentation and computer software are available on the Web at *http://water.usgs.gov/nawqa/sparrow/sparrow-mod.html*

Acknowledgment

We would like to express our utmost appreciation to Timothy A. Cohn, Stephen D. Preston and Richard B. Moore for their many helpful and insightful comments on this manuscript.

Part 1: A theoretical and practical introduction to SPARROW

1.1 Introduction

SPARROW (SPAtially Referenced Regressions On Watershed attributes; Smith and others, 1997) is a watershed modeling technique that uses a hybrid statistical and process-based approach to estimate pollutant sources and contaminant transport in watersheds and surface waters. SPARROW employs a statistically estimated nonlinear regression model with contaminant supply and process components, including surface-water flow paths, non-conservative transport processes, and mass-balance constraints. Parameters of the regression equation are estimated by correlating generally available stream water-quality records, such as those from State and Federal monitoring programs, with GIS (Geographic Information System) data on pollutant sources (e.g., atmospheric deposition, fertilizers, human and animal wastes) and climatic and hydrogeologic properties (e.g., precipitation, topography, vegetation, soils, water routing) that affect contaminant transport. The statistical estimation of parameters in SPARROW provides measures of uncertainty in model coefficients and water-quality predictions.

A unique feature of SPARROW is its model infrastructure, which consists of a detailed stream reach network with digital elevation model (DEM)-delineated watersheds to which all monitoring stations and GIS data on watershed properties are spatially referenced. This spatially distributed model structure allows separate statistical estimation of land and water parameters that quantify the rates of pollutant delivery from sources to streams and the transport of pollutants to downstream locations within the stream network (i.e., reaches, reservoirs, and estuaries). This mechanistic separation in the model of the terrestrial and aquatic features of large watersheds and emphasis on parameter estimation techniques represents an important advancement in the use of water-quality models to objectively evaluate alternative hypotheses about the major contaminant sources and watershed properties that control transport over large spatial scales. Spatial referencing and the mechanistic structure in SPARROW have been shown to improve the accuracy and interpretability of model parameters and the predictions of pollutant loadings as compared to those estimated in conventional linear regression approaches (e.g., Smith and others, 1997; Alexander and others, 2000).

SPARROW has been previously applied to the analysis of sources and transport of surface-water nutrients, pesticides, suspended sediment, organic carbon, and fecal bacteria, and is applicable to other measures of water quality, stream biology and streamflow. Recent applications of SPARROW have provided reasonably accurate estimates of nutrient sources and the long-term rates of nutrient removal in surface waters (e.g., Smith and others, 1997; Alexander and others, 2000; 2001; Alexander, Elliott, and others, 2002; Alexander, Johnes, and others, 2002). The model has demonstrated particular utility for quantifying the long-distance transport and delivery of nutrients to sensitive downstream locations (e.g., estuaries, reservoirs, drinking water intakes). Federal and State environmental managers are currently using SPARROW to assess the sources of nutrient loadings in streams, including its use for targeting nutrient reduction strategies in the Chesapeake Bay watershed (Preston and Brakebill, 1999) and in waters of the State of Kansas (Kansas Dept. Health and Environment, 2004) as well as for developing TMDLs (Total Maximum Daily Loads) in the Connecticut River Basin (NEIWPCC, 2004). The earliest version of the SPARROW model was developed to describe contaminant transport in surface waters of the State of New Jersey (Smith and others, 1994). Subsequently, applications were developed for the Chesapeake Bay watershed (Preston and Brakebill, 1999), New England watersheds (Moore and others, 2004), New Zealand river basins (Alexander, Elliott, and others, 2002; Elliott and others, 2005), North Carolina coastal watersheds (McMahon and others, 2003), and watersheds in Tennessee and Kentucky (Hoos, 2005). Models are currently under development for the Delaware River Basin and are being planned for selected regions of the U.S. (United States) as part of the U.S. Geological Survey (USGS) National Water-Quality Assessment (NAWQA) Program.

This publication documents the SPARROW modeling technique and computer software for constructing and applying basic SPARROW models. Details of the SPARROW software include the input data and installation requirements, guidance in the specification, calibration, and application of basic SPARROW models, and descriptions of the model output and its interpretation. The documentation is intended for both researchers and water-resource managers with interests in using the results of existing models and developing and applying new SPARROW models. The SPARROW software is written in SAS (Statistical Analysis System) IML (Interactive Matrix Language); however, only a very basic knowledge of SAS is required to develop most

standard versions of SPARROW models used to date. A more complete knowledge of SAS/IML is required to make extensive modifications to the model code.

The SPARROW documentation is presented in two parts. Part 1 provides a theoretical and practical introduction to SPARROW modeling techniques. This includes a discussion of the conceptual attributes and model infrastructure of SPARROW, background on the commonly used model specifications and the methods for estimating and evaluating parameters, evaluating model fit, and generating water-quality predictions and measures of uncertainty. Part 2 provides a user's guide to SPARROW. This includes a discussion of the software architecture and details of the model input requirements and output files, graphs, and maps. Throughout this report, we use a SPARROW model for total nitrogen, based on an application to a national data set for the United States, to illustrate model concepts and the components and output of SPARROW models that are supported by the computer software.

A number of technical results related to the SPARROW methodology described in this documentation have not previously been reported in the literature. We found it necessary, therefore, to expand certain sections of the manuscript in order to properly derive these results. Unfortunately, providing this level of detail may burden the reader who simply wishes to get the "big picture." With this goal in mind, such readers may elect to skip the more onerous technical sections identified as containing "advanced" material (the advanced material is shaded to assist in its identification).

1.2 SPARROW Modeling Concepts

This conceptual introduction to SPARROW modeling is intended to orient the reader with respect to the capabilities and limits of a SPARROW analysis. We first describe the key research and management modeling objectives that SPARROW can be used to address. We then articulate the general features of the mass balance approach used in the SPARROW model and the advantages it offers. This is followed by discussions of the appropriate time and space scales for developing and applying SPARROW models, and issues related to the accuracy and complexity of these models. Finally, we provide a conceptual description of how the SPARROW modeling structure compares with other types of water-quality models that are commonly applied to watersheds.

1.2.1 Model objectives

The primary objective of constructing a SPARROW model is to establish a mathematical relation between water-quality measurements made at a network of monitoring stations and attributes of the watersheds containing the stations. Once constructed, the model may be used to satisfy a variety of water-quality information objectives.

1.2.1.1 Water-quality description

One common objective is to describe past or present water-quality conditions for a state or region on the basis of monitoring data. The underlying challenge is to extrapolate from a sample of water-quality measurements made at a finite number of stream and river locations (i.e., a monitoring network) to an area containing the sampling stations and a large number of un-sampled locations. The usual limitations in doing this are: (1) sparse sampling, reflecting the high cost of monitoring; and/or (2) unrepresentative (i.e., non-random or targeted) sampling undertaken to satisfy the separate and competing objective of characterizing water quality at specific locations, especially those suspected of having water-quality problems. In the absence of an interpretive model such as SPARROW, a single monitoring design cannot be optimal for these two distinct objectives (i.e. spatially representative sampling to characterize the general water quality of a region, and targeted sampling to characterize specific water chemistry at "suspect" locations). Because the Federal Clean Water Act requires state governments to collect and report both types of information, the distinction between the two types of monitoring is of great practical importance. "Probabilistic" monitoring has been promoted by the U.S. Environmental Protection Agency (USEPA) (Yoder, 1997) to obtain a spatially unbiased, broad overview of water-quality conditions in State waters. An important limitation of this approach to assessment is that the monitoring data alone do not provide detailed information on the geography of water-quality conditions and give little understanding of the factors (i.e., sources and processes) that explain those conditions. Targeted monitoring has been used extensively by the States to identify specific streams with water-quality problems and to gage compliance with State water-quality regulatory standards and criteria. These data, however, provide a

spatially biased description of water-quality conditions in watersheds. Modeling tools, such as SPARROW, are needed to integrate samples from these different monitoring approaches to provide both a geographically representative description of water-quality conditions as well as insight into the sources and watershed processes that control water quality. A properly calibrated SPARROW model can assist with these problems and objectives in the following ways. First, by including data on watershed characteristics (e.g. the size and location of pollution sources), the model enlarges the information base and reduces the problem of sparse monitoring. The information gain allows for a more accurate description of the water quality of a region than that provided by water-quality monitoring data (even spatially representative data) alone. Second, by using the model to predict water-quality conditions at a representative set of locations, the effects of spatially biased sampling may be greatly reduced or eliminated. Finally, the model may also be used to identify the specific locations where poor or unsatisfactory water quality is present. Targeted sampling may then be used more efficiently to confirm such predictions in cases where model accuracy is estimated to be less than some predetermined tolerance.

There are numerous published examples of the use of SPARROW models to describe water-quality conditions either in map or tabular form (Smith and others, 1997; Alexander and others, 2000; 2001; Alexander, Elliott, and others, 2002; Smith and Alexander, 2000; Preston and Brakebill, 1999; Smith and others, 2003; Moore and others, 2004; Smith and others, 2004). Maps of stream reaches may be variously colored to indicate contaminant loads, yields, or concentrations and have the advantage of displaying the general regional pattern of water-quality conditions along with important details such as the specific locations of extreme values. A tabular form of presentation, on the other hand, allows for a combination of regional summary statistics compiled for a set of point predictions and including such statistics as mean values, quantiles, the proportions of locations at which specified standards or other thresholds are exceeded, and measures of error for all of the above. One interesting and useful feature of tabulated regional statistics is that the accuracy and precision of regional statistics can be seen to increase with the size of the region (measured as the number of point predictions; Smith and others, 1997). This pattern is a reflection of the fact that the quantity of information that is used in making regional predictions increases with the size of the region.

1.2.1.2 Contaminant source analysis

Another objective of SPARROW modeling is to identify and quantify the sources of pollution that give rise to in-stream water-quality conditions predicted by the model. In describing pollution sources, we distinguish between “source categories,” such as point sources, atmospheric sources, and animal agriculture, and “individual sources” defined as the rate of supply of contaminant of a particular category originating in the watershed and draining to a specific stream reach. As with the descriptive water-quality applications discussed above, information on pollution sources may be desired either for an individual stream location or may be summarized for an area defined as a set of stream locations. The ability to develop quantitative information on pollution sources in SPARROW models stems from the ability to trace, for each contaminant category, the predicted in-stream flux through a given stream reach to the individual sources in each of the upstream reach watersheds contributing contamination to that reach. Sources may be quantified either in mass units or in terms of their percent contribution to the total contaminant flux to the reach. A national summary of nutrient (total nitrogen and total phosphorus) sources for the Reach File 1 (RF1) reach watersheds (Alexander, and others, 1999) is presented by Smith and Alexander (2000).

An example application of SPARROW in quantifying pollution sources is in TMDL (total maximum daily load) analyses (e.g., McMahon and Roessler, 2002). In general, the Clean Water Act requires TMDL analyses for any stream reach in which the concentration of a contaminant exceeds the applicable water-quality standard when all pollution discharge limits are met. The ultimate objective of TMDL-related modeling is to establish a hypothetical waste-load allocation for all individual sources affecting the reach in question that would cause water quality at that location to meet the standard. In theory, an infinite number of hypothetical load allocations will satisfy the standard, and choosing the official allocation requires comparing many possible solutions in search of a least-cost or other optimum solution. Because the model is used in such cases to describe hypothetical scenarios, the modeling objective is considered to be a simulation exercise (see below). Prior to conducting the hypothetical analyses, however, a great deal of preliminary quantitative information on the actual (i.e. baseline) relations between watershed sources and in-stream conditions is useful in preparing for the discussion of alternative allocations. The percentage contributions (shares) of individual point sources are commonly of particular interest because point sources are usually the only sources subject to direct regulation.

The share contributions of individual nonpoint sources are also of interest as a means of identifying the important stakeholders to include in discussions of voluntary pollution reductions.

1.2.1.3 Water-quality simulation

Simulation refers to the use of a calibrated model to predict conditions on the basis of a set of altered (usually hypothetical) model inputs. The ability to portray counterfactual conditions for specified inputs is one of the most powerful uses of and reasons for constructing models; there are often no alternative methods for conducting controlled experiments on complex systems. In the water-quality arena, model simulations often depict the in-stream effects of changes in contaminant sources associated with alternative future pollution-control strategies. By linking water-quality conditions to pollution reduction, such simulations provide a critical step in the analysis of the costs and benefits of pollution control. Simulating alternative waste-load allocations in a TMDL analysis (see above) is a prime example of water-quality simulation with this objective. The use of SPARROW in support of TMDL development for the Neuse River Basin, NC, is described by McMahon and others (2003) and McMahon and Roessler (2002).

In another example of the use of a SPARROW model in simulation mode (Smith and others, 2004), historical pollution rates rather than future pollution reductions provide the hypothetical model inputs. The study examines the effects on stream bacteria levels of regional changes in U.S. livestock manure production caused by structural changes in the agriculture industry between 1982 and 1997. Whereas the agricultural pollution inputs were based on actual manure production data for the period, other bacterial sources were held constant at their historical mean levels. The advantage of handling the model inputs in this way is that the effects of historical changes in one contaminant source category can be viewed without the additional variability caused by historical changes in other sources.

Reducing cultural (i.e., anthropogenic) pollution sources in models to zero is one of the few methods available for simulating natural, unpolluted environments of the more distant past. The primary alternative to simulating natural environments in the water-quality area is the direct study of relatively pristine “reference” sites in remote regions. A SPARROW-based study by Smith and others (2003) combines these two approaches. Nutrient measurements from 63 USGS reference sites in 14 nutrient ecoregions were used to characterize small watershed total nitrogen (TN) and total phosphorus (TP) yields as functions of runoff and the natural vegetation cover. The small watershed models were then used to provide the nutrient inputs for national-level SPARROW simulations of natural TN and TP conditions. In applying the nutrient input models, the vegetation cover of the conterminous U.S. was altered to include only natural vegetation classes.

1.2.1.4 Hypothesis testing: examining the importance of explanatory factors and processes

One common feature of the previous modeling objectives is that they each make use of predictions of the dependent variable by a calibrated model. Another class of modeling objectives focuses on the calibration process and its results directly. As with multiple regression modeling in general, the SPARROW estimation process explores the predictive value of a set of potential explanatory variables and may also compare alternative mathematical forms. The selection of a final set of predictors and mathematical form usually has the primary objective of maximizing the accuracy of model predictions of the dependent variable, but an alternative modeling objective may be to test one or more hypotheses about the nature and importance of factors and processes that may have influenced water quality at the locations where samples were collected. Hypothesis tests are performed for each of the model parameters estimated in the calibration, and these serve as indicators of an empirical relation between the independent variables associated with each parameter and the dependent variable of the model. Because the coefficients in a SPARROW model are specified to conform to physical processes, and the potential explanatory variables are selected on the basis of some theoretical or logical connection to the dependent variable, a statistically strong parameter provides evidence of a physical relation to the dependent variable (see additional discussion in section 1.5.4.2).

For example, in cases in which multiple categories of potential sources of the contaminant are being modeled, the model estimation process provides useful hypothesis tests on the importance of each category. “Importance” is measured by the correlation between contaminant inputs from the source category and downstream monitored loads of the contaminant. That correlation will tend to be stronger when the mass contribution of the category to the total mass of contaminant flowing past many of the monitoring stations is large, but model estimation may also indicate a source category is important even when the mass contribution is

small, provided the amount contributed to stream loads by each unit of source is consistent from place to place. An important case in point is the significance of point sources of total nitrogen and total phosphorus observed in estimating the national-scale SPARROW models (Smith and others, 1997). Hypothesis tests for point sources of the two nutrients were each highly significant ($p < 0.005$), but the average mass contributions of point sources to total mass flux at the Hydrologic Unit outlets were only 2 percent and 9 percent, respectively—although point sources may dominate in specific reaches. Point sources have also been found to be statistically significant in most of the regional SPARROW models (Preston and Brakebill, 1999; Alexander, Elliott, and others, 2002; McMahon and others, 2003; Moore and others, 2004; see also the discussion in section 1.5.4), despite their small overall contributions to stream nutrient flux.

The importance of factors and processes potentially related to the transport of contaminants from sources to stream channels and within stream channels may be tested through calibration of the "land to water" and "in-stream decay" terms in the model. These terms dictate the fraction of the contaminant mass that completes the terrestrial and aquatic phases of transport within the watershed draining to each stream reach. The land-to-water terms describe the land-surface characteristics that influence both overland and subsurface transport from sources to stream channels. Similarly, the in-stream decay terms describe the effects of channel characteristics on downstream transport. In addition to their statistical significance, the mathematical form of these terms, and the signs of their estimated coefficients, provide useful empirical information on the processes that affect water quality. The form and coefficient sign, for example, indicate whether the effect of a particular land-surface variable is positively or negatively correlated with contaminant transport. Beginning with the first SPARROW models (Smith and others, 1997), for example, soil permeability has consistently been found to be negatively related to TN transport in SPARROW models, which is indicative of the long-term storage or permanent removal (i.e., denitrification) of nitrogen in soils and the subsurface. Similarly, stream channel size, measured as either the channel depth or mean flow rate, has consistently been found to be inversely related to the first-order total nitrogen decay rate in stream channels (Alexander and others, 2000; see discussion in section 1.4.4). In sum, a frequent objective in building and calibrating SPARROW models is to gain insight and to test hypotheses concerning the role of specific contaminant sources and hydrologic processes in supplying and transporting contaminants in watersheds.

1.2.1.5 Design of sampling networks

Government programs for collecting water-quality samples from a large network of regularly visited sites on rivers and streams have existed for approximately a century (Dole, 1909). For most of this history, monitoring sites have been chosen on a multi-objective basis, balancing the continuous need for a generally representative picture of regional water-quality conditions with a periodic need for more detailed information on specific locations or problems. SPARROW models are atypical in the realm of water-quality modeling because they are designed specifically to interpret the data collected at a network of monitoring sites (see Smith and others, 1997). Once a SPARROW model has been constructed for a monitoring network, some commonality likely exists between the objectives of monitoring and those of SPARROW modeling, at least with respect to the contaminants that have been modeled. It then becomes logical to consider using the model to choose sampling locations on a more objective basis so as to optimize the ability to achieve the common monitoring/modeling objectives.

An example of using a SPARROW model to identify high-priority sites for future monitoring is described by McMahon and others (2003). They point out that when the objective is to establish where contaminant levels exceed a specific threshold, it makes sense to collect data where the uncertainty about exceeding the threshold is greatest. Because they are statistically calibrated, SPARROW models are able to provide quantitative information on model prediction uncertainty for each stream reach and, thus, indicate where additional sampling would best support that objective. Other objectives may call for addressing other aspects of model uncertainty. For example, if information on the future effects of reducing a certain source of contamination is desired, it would likely be most beneficial to collect data so as to reduce the uncertainty of the model parameter that is associated with that source rather than addressing overall prediction uncertainty. In conclusion, it is important to note that usually there are multiple objectives of water-quality monitoring and that in order to design an optimal monitoring network it is necessary to precisely state the objectives and their relative importance in quantitative terms. Once this is done, however, a SPARROW model can form the infrastructure of an algorithm for optimizing network design. A great deal of work in this area is expected in the future.

1.2.2 Mass balance approach

SPARROW models constructed to date, like most water-quality models and many other models describing some aspect of the physical world, are expressed in the form of a mass balance. Such models describe the movement of mass in space and/or the change of mass in time. The law of conservation of mass implies that certain basic accounting rules must apply to a mass balance, water-quality model, such as: (1) the sum of fluxes entering the confluence of two streams equals the flux leaving the confluence; (2) the sum of the fluxes attributable to each contaminant source must equal total flux; and (3) a doubling of all sources in the model results in an exact doubling of the predicted flux at each location. Because the dependent variable in SPARROW models (the mass of contaminant that passes a specific stream location per unit time) is, in mathematical terms, linearly related to all sources of contaminant mass in the model, all accounting rules relating to the conservation of mass will apply.

Mass accounting in SPARROW models is also supported by the explicit spatial structure defined by the stream network. Most other empirically based surface water-quality and streamflow models, however, use a spatially inexplicit approach that does not support mass balance (Larson and Gilliom, 2001; Mueller and others, 1997; Tasker and others, 1996; Vogel and others, 1999). Typically, these models relate measurements of stream water quality and streamflow to spatially averaged basin characteristics via a log-linear functional relation. Because the log-linear relation is non-additive, the sum of fluxes entering a confluence, as predicted by separate sets of basin characteristics, does not equal the flux leaving a confluence, as predicted by a single set of aggregated basin characteristics.

There are a number of advantages to the mass balance approach. Because of the linear relation between flux and sources, there is an expectation that the estimation of flux over spatial scales smaller and larger than that covered by the model's sample data will yield reasonably accurate results. This would not generally be an expectation for the log-linear basin-oriented models. The imposition of mass balance greatly improves the interpretability of model coefficients. For example, because of the assumption of mass balance, the coefficient associated with the reach time-of-travel variable is interpreted as a first-order decay rate and, because point-source loadings are delivered directly to the stream network, a reasonable null hypothesis for coefficients associated with point sources is that they equal 1.0 (Smith and others, 1997; see the discussion in sections 1.4.1, 1.4.2, and 1.5.4.2). The enhanced interpretability of the model coefficients in turn facilitates the comparison of coefficient estimates from the model with other estimates described in the literature. These comparisons have been generally favorable, especially for the model components that quantify in-stream nitrogen decay rates (Alexander and others, 2000; Alexander, Elliott, and others, 2002; McMahon and others, 2003; see section 1.4.4), nutrient and suspended sediment removal rates in reservoirs (Alexander, Elliott, and others, 2002; Schwarz and others, 2001; see section 1.4.5), and the nutrient export associated with various land uses and pollutant sources (e.g., Alexander and others, 2001, 2004; Alexander, Elliott, and others, 2002; McMahon and others, 2003; see sections 1.4.2 and 1.5.4).

Mass balance provides a basis for flux accounting, whereby flux can be allocated to its various sources, both spatially and topically (that is, according to the location and type of source—for example, fertilizer, atmospheric deposition, etc.). For example, mass balance makes it possible to attribute nutrients discharged to the Gulf of Mexico to specific sources within the Mississippi basin (Alexander and others, 2000), thereby providing guidance in managing the reduction of this discharge.

There are at least two ways in which mass balance can be imposed in a hydrologic model. The first approach, which might be called the traditional approach, imposes mass balance dynamically. This approach is implemented by requiring that the mass of water entering a reach from upstream at time t, plus the mass discharged at time t into the reach from sources within the reach's incremental watershed, equals the mass of water leaving the reach at time $t + d$, where d is the time of travel through the reach. Such restrictions are commonly used in deterministic models such as TOPMODEL (Wolock, 1993) and models in the USEPA's BASINS system (Better Assessment Science Integrating Point and Nonpoint Sources; USEPA, 2001; *http://www.epa.gov/docs/ostwater/BASINS/*). The restrictions are useful because they make explicit the dominant fluvial pathways—by knowing what the water is exposed to, it is possible to infer what is in it. The complication arising from this approach, however, is the extensive infrastructure required to support it. All data must be referenced with respect to both space and time, making it difficult to construct large-scale regional models. Moreover, many of these models are over-parameterized (that is, the parameters of the model are under

identified), causing the coefficients describing contaminant supply, transformation, and transport processes to be subject to large uncertainties (see additional discussion in section 1.2.5).

An alternative approach is to impose mass balance on average, removing entirely the temporal component. In this case, the model is described in terms of summary statistics that must balance over time. The simplest summary statistic to work with is the mean—for example, the long-term annual mean flux or the long-term mean flux during a given season. This is the statistic SPARROW has been designed to model and for which all applications to date have been developed. As discussed in the subsequent section, this approach is sensitive to the effects of natural and human-related processes that supply and remove contaminants from watersheds over long periods. Thus, this approach de-emphasizes the quantification of short-term cycling and transformation processes, which are often central to the functioning of many dynamic mechanistic models, in favor of processes that have a long-term impact on elemental budgets in aquatic ecosystems. Scientific and management interests in understanding the nature of these persistent impacts have grown considerably in recent years, with particular attention focused on nutrients (e.g., Howarth and others, 1996; Vitousek and others, 1997; Carpenter and others, 1998; Preston and Brakebill, 1999; National Research Council, 2000; Boyer and others, 2002; Moore and others, 2004).

Other summary statistics also could be modeled using a mass balance approach. For example, a mass balance approach could be developed to explain the variance of flux over time. Such a model, used in combination with a model of the mean value of flux for a specified period, could then yield estimates of the distribution of flux over time—information that would be of considerable utility in evaluating many water-quality standards promulgated by the USEPA. Models of higher order statistics could be developed to fine tune the estimate of the flux distribution. Moreover, by expanding the analysis to one that determines the bivariate distribution of contaminant flux *and streamflow*, it becomes possible to determine the distribution of concentration over time—another important distribution for addressing USEPA water-quality standards. The feasibility of each of these extensions of the basic SPARROW model remains speculative pending the outcome of future research.

Technically, a mass balance approach focused on long-term conditions would preclude the analysis of water-quality and streamflow data expressed in terms of frequency of occurrence. Analyses of these types of variables have been the domain of the basin-oriented statistical models cited above. The limitation of the mass-balance approach in this context refers to a lack of simultaneity across all reaches for the particular frequency condition being studied—that is, if a particular reach is at the 70th percentile with regard to some water-quality or streamflow condition, it cannot be expected that all other upstream reaches are simultaneously experiencing the same percentile condition.

Finally, it should be pointed out that nothing in the SPARROW methodology or the program code described in this manual precludes the modeling of dependent variables that are not expressed as a mass flux. Indeed, many important water-quality variables are expressed in non-flux units, and brief mention of a few of these may serve to encourage future experimentation with SPARROW models in which the usual mass balance characteristics would either be modified or dispensed with entirely; for example, results of a preliminary SPARROW model of fecal coliform bacteria in which the dependent variable was the number of bacteria colonies passing a stream location per unit time were recently reported by Smith and others (2004). Because the flux of bacteria colonies in a stream would be expected to vary in approximate linear proportion to the mass of animal waste and other bacteria sources released per unit time in the surrounding watershed, the model retains a strong mass-balance character. If, however, the dependent variable of interest is the concentration of a contaminant in benthic sediments or in fish tissues for example, which do not flow with the water and cannot be expressed in flux units, the mass balance nature of the model is more tenuous. If the dependent variable makes no reference to mass or mass-proportional quantities whatsoever, as in species diversity for example, then clearly the model does not represent a mass balance at all. Nevertheless, all of the above dependent variables are important measures of water quality and share an important characteristic with those that have been successfully modeled with SPARROW—they may be strongly correlated with contaminant sources or other spatially referenced attributes of watersheds. Thus, attempting to model them with the methods and software described herein may be worthwhile. The absence of a mass balance relation in SPARROW models may make it more difficult to interpret the physical meaning of the predictor variables, but the spatially explicit relation between the dependent and predictor variables in such models may enhance interpretability of the model in comparison to non-spatially referenced regression models.

1.2.3 Time and space scales of the model

Previously developed SPARROW models and the basic version of the model described in the accompanying computer code and supported in this documentation are structurally designed to explain *spatial variability* in the long-term mean-annual or mean-seasonal flux (mass per unit time) of contaminants in streams, the response variable of the model. Spatial variability in the long-term mean flux is modeled as a function of natural and human-related properties of watersheds that influence the supply and transport of contaminants. Estimates of the long-term mean flux are developed from water-quality and streamflow monitoring data that are regularly collected at fixed locations on streams and rivers. The basic form of SPARROW models presented in this report is structured to describe the long-term, *steady-state* water-quality and flow conditions in streams. Contaminant source inputs are assumed to be in balance with the estimated sinks and measured riverine output (i.e., in-stream water-quality load) such that there is conservation of mass among the model components that describe the source inputs, sinks, and the in-stream flux of contaminants. The principal objective supported by this model structure is the quantification of the location and rates (and statistical uncertainties) of the supply, transport, and fate of contaminants within the terrestrial and aquatic ecosystems of watersheds. In the current specifications of SPARROW models, *temporal variability* in contaminant loads, including intra- and inter-annual variations in water quality and streamflow reflected in the monitoring data, are explicitly modeled and accounted for in a step prior to modeling spatial variability in loads with SPARROW. We comment later in this section on SPARROW model specifications that would be of interest in modeling variations in contaminant loads in both space and time.

The computation of a long-term mean-annual or mean-seasonal flux, the SPARROW response variable, requires the prior application of a water-quality flux-estimation model constructed on the basis of streamflow and water-quality records from regularly monitored stream locations (see section 1.3.1 for details). The flux-estimation model explicitly accounts for temporal variability in contaminant loads related to streamflow, season of the year, and trends (either continuous or abrupt) with time. In previous applications of SPARROW, the trend term has been used to remove time trends in contaminant loads by detrending the mean-annual contaminant load to a particular base year. A base-year load estimate ensures that the stream water-quality loads and the contaminant source data (which are commonly reported only periodically, e.g., the U.S. Agricultural Census reports agricultural conditions every five years) are contemporaneous. Therefore, the mean-annual loads used to calibrate SPARROW models (and also the SPARROW predictions of contaminant load) describe the mean load that would be expected to occur during a particular base year under long-term mean streamflow conditions. Provided there are sufficient data describing the contaminant sources in watersheds, one could alternatively estimate long-term averages in both stream water-quality loads and source inputs for a contemporaneous period of record without explicitly detrending the estimates to a base year.

The *steady-state* mass-balance structure of the basic SPARROW model quantifies the long-term net effects of biogeochemical and hydrologic processes on contaminant transport in terrestrial and aquatic ecosystems. Modeling the effects of these processes is typically of greatest interest for non-conservative chemical, physical, and biological properties of water, such as nutrients, pesticides, fecal coliform bacteria, organic carbon, and suspended sediment. Many of these constituents are subject to chemical transformations or degradation during transport and may be stored over short or long periods. For example, large quantities (greater than 75 percent) of the nitrogen input to watersheds from various sources over annual time scales are either permanently removed or stored in the terrestrial and aquatic ecosystems of many eastern U.S. watersheds (Howarth and others, 1996). In SPARROW models that are estimated using long-term water-quality records (multi-year to decadal periods), biogeochemical cycling and storage processes that temporarily immobilize or remove contaminants from flow paths are generally in steady state with those processes that mobilize or release contaminants from storage. Hence, the effects of transformation and removal processes (e.g., mineralization, nitrification, uptake by stream algae) that operate on relatively short intra-annual time scales (e.g., daily, weekly, seasonal), or over multi-year time scales that are less frequent than the period of model estimation, are not likely to be detected as contaminant losses in the steady-state form of SPARROW models.

The emphasis on long-term mean conditions in SPARROW allows the model to be especially sensitive to detecting the effects of biogeochemical processes that either remove contaminants from watersheds permanently or result in their storage over time scales that are longer than the period of the monitoring records used to estimate the model. Permanent removal processes may include denitrification—a biologically mediated conversion of inorganic nitrogen to gaseous nitrogen, the mortality of pathogens (e.g., fecal coliform indicator bacteria) from exposure to ultraviolet light or high salinity, and the natural degradation of pesticides and

formation of by-products or lower-order chemical forms. SPARROW models have been used previously (e.g., Smith and others, 1997; Alexander and others, 2000; Alexander, Elliott, and others, 2002) to quantify net losses of the total forms of phosphorus and nitrogen in terrestrial and aquatic systems over nearly decadal or longer periods. These losses would be expected to reflect the long-term net effects of the rates of such processes as denitrification, immobilization, mineralization, nitrification, and particulate settling and resuspension. The losses may include the net loss or storage of nutrients (e.g., sediment-bound phosphorus, organic nitrogen) on land (e.g., soils, ground water) and in reservoirs and streams and their nearby floodplains.

Limitations in our knowledge of temporal lags in chemical transport and their causes create uncertainties in the periods over which steady-state conditions apply. For example, the time scales for the transport of sediment in streams reflect erosion, storage, and transport processes that operate from seasonal to decadal or even longer (e.g., century) periods (Trimble and Crosson, 2000). Dissolved substances, such as nitrate, are readily transported in the subsurface and may be permanently removed from watersheds via denitrification, but may also be attenuated along lengthy subsurface flow paths over periods of years to decades (e.g., Bohlke and Denver, 1995). Recent SPARROW nitrogen models are generally sensitive to the effects of subsurface properties that affect water and nitrogen movement; the models indicate that the largest nitrogen losses occur in watersheds with permeable soils and topographic properties that accentuate water infiltration and movement into the subsurface. The available information on the age of stream waters is limited, but indicates that streams may contain an appreciable contribution of older waters with ages in excess of 10 or more years. For example, based on available tritium measurements at a small number of U.S. river locations (Michel, 1992), older waters (greater than 1 year in age) may constitute an average of nearly 50 percent of total river flow; the estimated mean residence times of these waters ranges from 10 to 20 years (Michel, 1992; Focazio and others, 1997). Temporal lags, ranging from 2 to 9 years, have also been estimated in the mean-annual nitrogen flux near the outlet of the Mississippi River (McIsaac and others, 2001). Moreover, recent studies of European rivers have noted many instances where, despite substantial declines in agricultural fertilizer use and livestock manure production, such as those evidenced after the fall of the former Soviet Union, decreases in riverine nutrient concentrations were not observed in subsequent sampling periods that ranged from 5 to 10 years (Stalnacke and others, 2003). In general, where the in-stream response of contaminant flux to changes in watershed sources is believed to be delayed because of long subsurface residence times (e.g., Bachman and others, 1998; McIssaic and others, 2001; Lindsey and others, 2003), the mean estimates of the in-stream flux and source inputs should be based on data for a correspondingly long period to more accurately reflect steady-state conditions. Water and contaminant residence times in ground water are highly uncertain (e.g., Focazio and others, 1997), however, and the temporal lags in the response of stream loads to changes in source inputs are unknown for most watersheds. As a result, multi-year temporal lags in contaminant transport have not been explicitly modeled in most surface-water contaminant transport models. In developing SPARROW models, to address uncertainties over time lags in transport, preference should be given to the use of data from lengthy multi-year periods to estimate water-quality loads and source inputs.

In addition to developing spatial models of mean stream water-quality conditions, temporal changes in mean conditions also can be explicitly modeled in SPARROW using the current model structure and software, although little work has been done on testing such models. One approach is to estimate multiple steady-state models that explain mean-annual or mean-seasonal stream contaminant loads for separate multi-year periods. Alternatively, the current software permits the estimation of a single SPARROW model having time-dimensioned dependent and independent variables—each assumed to pertain to different steady-state conditions. In this approach, temporal changes in mean-annual stream contaminant flux may be modeled as a function of temporal changes in contaminant sources, land use, and climatic factors (e.g., precipitation, streamflow, and temperature) over similar multi-year periods as those used to estimate stream loads at each monitoring station. The coefficients in this model can be time dependent or restricted to take common values over the full period of the analysis. This type of model structure has considerable data demands that require the development of historical data on contaminant sources and climatic/hydrologic variability in the watersheds and reaches (including stream velocity, which is sensitive to average streamflow and thus varies in response to changes in mean streamflow across different periods). Such a model would allow users to explicitly test for temporal changes in model parameters to determine whether changes are evident in process rates or contaminant export coefficients with time.

As an alternative to the explicit estimation of a SPARROW model with time-dimensioned coefficients, existing SPARROW models can be used to simulate changes in mean-annual stream water-quality loads as a

function of changes in source inputs. This approach assumes that the process rates and contaminant export coefficients of the model do not change with time. One recent application of this approach was developed to evaluate the effect on fecal bacteria concentration in streams from changes in the wastes emitted from confined and unconfined livestock operations (Smith and others, 2004).

The development of dynamic SPARROW models to describe short-term, intra-annual variability in water-quality loads, the long-term storage of contaminants, or short-term cycling processes that involve separate nutrient forms requires more advanced software and model specifications than those described in this report. Non-spatially referenced statistical models of time-varying mean-annual nitrogen loads as a function of contemporaneous and antecedent nitrogen source inputs were developed for the outlets of the Mississippi River basin (McIsaaic and others, 2001) and basins draining to Long Island Sound (Mullaney and others, 2002). A spatially referenced extension of these models would require the specification of storage pools within the interior reach catchments of the basin—a feature that is not part of the current SPARROW model structure and software. Dynamic versions of SPARROW models that simulate inter- and intra-year variability in stream contaminant loads for seasonal, monthly, or daily time scales could ultimately provide the most detailed temporal descriptions of water quality. However, the similarity of the structure of this type of model to existing dynamic simulation models, such as the HSPF and SWAT watershed models, may argue for simply replacing current ad hoc methods used to fit these models with traditional parameter-estimation methods from the statistics literature (e.g., using nonlinear parameter-estimation software such as PEST; Doherty, 2004). In keeping with SPARROW approaches, this would entail simplifying these simulation models to only the set of sensitive parameters (i.e., those for which the predictions are the most responsive); most existing simulation models are over-parameterized and executed with many non-sensitive and sometimes highly collinear parameters.

1.2.4 Accuracy and complexity of SPARROW models

It is generally the case for SPARROW, as for any model with statistically estimated parameters, that model accuracy (bias and precision) and complexity (number of statistically significant or sensitive parameters) are dependent on the “information content” of the water-resources data used in the model calibrations. Investigations of hydrologic models have demonstrated that both the *quantity* and *quality* of the calibration data define the information content and have important effects on parameter estimation and precision (e.g., Gupta and Sorooshian, 1985; Yapo and others, 1996). Increasing the *quantity* of data can improve the precision provided the data give new, independent information about the values of the model parameters. Data *quality*, as defined by Gupta and Sorooshian (1985), generally increases as the data become more “representative” of the range of watershed properties that affect transport and the range of conditions present in the sampled watersheds. An orthogonal or independent set of measurements are preferred for estimating parameter values, such that the data reflect the most extreme combinations of watershed conditions for the various properties. This often provides the best information for assessing a parameter’s statistical significance.

These general statistical guidelines have implications for the time and space scales required to develop SPARROW data sets and accurate models. First, a sufficiently large number of water-quality monitoring stations are required. In SPARROW models, the monitoring-station loads serve as the response variable observations in the nonlinear spatial regressions. The number of stations has a demonstrable effect on the statistical power of the regression—i.e., the capacity of the model to detect the effect of an explanatory factor on stream loads. Models with more stations generally have greater power, which typically supports more complex models—i.e., models with a larger number of statistically significant parameters and functional components. Table 1.1 and figure 1.1 illustrate this relation for previously developed national and regional SPARROW total nitrogen models. The number of statistically significant parameters in these models (fig. 1.1a, 1.1c) generally increases with the number of monitoring stations used in the calibrations. For example, the number of parameters increases from typically 6 or 7 in the regional models having fewer than about 60 stations to about 10 parameters in the Chesapeake Bay and national New Zealand models, which have nearly 80 stations, and finally to 18 parameters in the U.S. national model, which has more than 300 stations.

Table 1.1. Station characteristics and model performance data for national and regional SPARROW total nitrogen models.

[km^2, square kilometers; RMSE, the root mean square error, expressed in percent, is a measure of the average percentage error in a prediction of mean-annual concentration at a given reach]

Model Location	Station Characteristics			Model Performance	
	Drainage Area (km^2)	Number of sites	Site Density (km^2 / site)	Accuracy [RMSE] (percent)	Complexity [number of statistically significant parameters[1]]
United States	6,057,000	375	16,590	56	18
New Zealand national	136,000	77	1,766	34	11
Chesapeake Bay	125,000	79	1,582	41	10
Northeast U.S.	40,000	65	615	40	6
North Carolina Coast	26,000	44	590	47	7
New Zealand-Waikato River	14,000	37	378	38	6
Tennessee/Kentucky	100,100	36	2,781	19	7

[1] Statistical significance for an alpha of 10 percent.

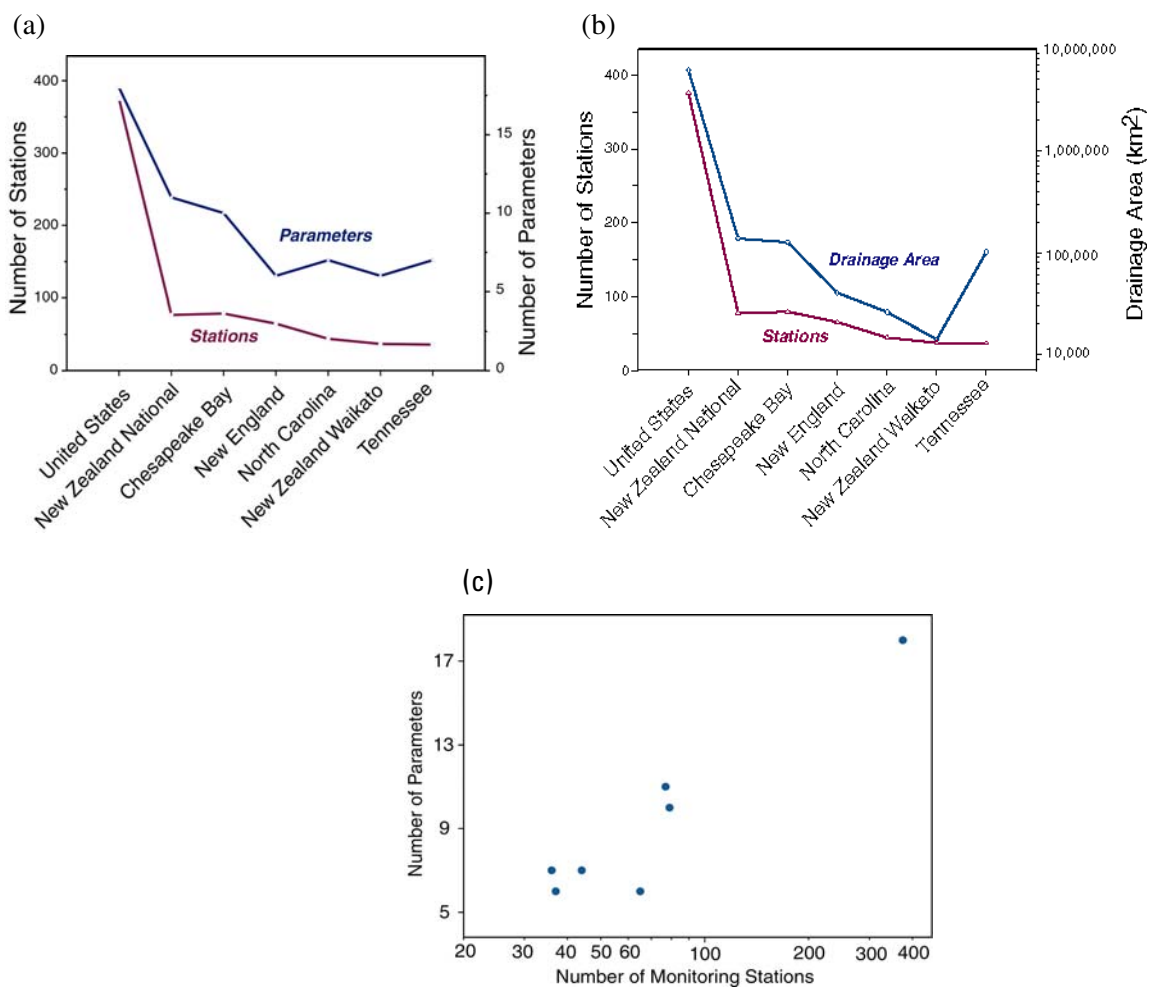

Figure 1.1. Characteristics of the national and regional SPARROW total nitrogen models: (a) number of statistically significant model parameters, (b) drainage area, (c) model complexity (number of parameters tested) and number of stations.

Second, the amount of spatial variability in the stream monitoring data and explanatory factors should reflect as broad and representative a range of watershed conditions as possible. The most complex SPARROW models typically have been developed for regions that have relatively large spatial variability (greater than one order of magnitude) in the watershed properties that affect contaminant transport. Watershed properties that vary over a wide range within a modeled region generally provide more information about the response of stream loads to different levels of a given watershed property and are more likely to be statistically significant in SPARROW models.

Consider, for example, the case of mean-annual estimates of nitrate deposition (fig. 1.2), an important source of nitrogen that is included in most SPARROW total nitrogen models. The SPARROW models for which a statistically significant coefficient has been estimated for atmospheric deposition include the national U.S. model and regional models developed for the northeastern U.S., Chesapeake Bay, and Tennessee/Kentucky. All of these models were developed for geographic areas in which the mean-annual rates of deposition vary considerably over generally short distances. In general, it is difficult to estimate a statistically significant atmospheric-deposition source variable in models where there is little spatial variability in mean-annual rates of deposition, such as the rates observed in many of the watersheds in the western and southeastern U.S. We have also found that only the largest drainage areas (i.e., associated with the U.S. and New Zealand national models) typically have sufficient spatial variation in mean-annual precipitation for the models to detect a statistically significant response in the mean-annual stream nutrient loads. Developing time-dependent models could potentially improve the responsiveness of the regional models to properties such as atmospheric deposition and precipitation.

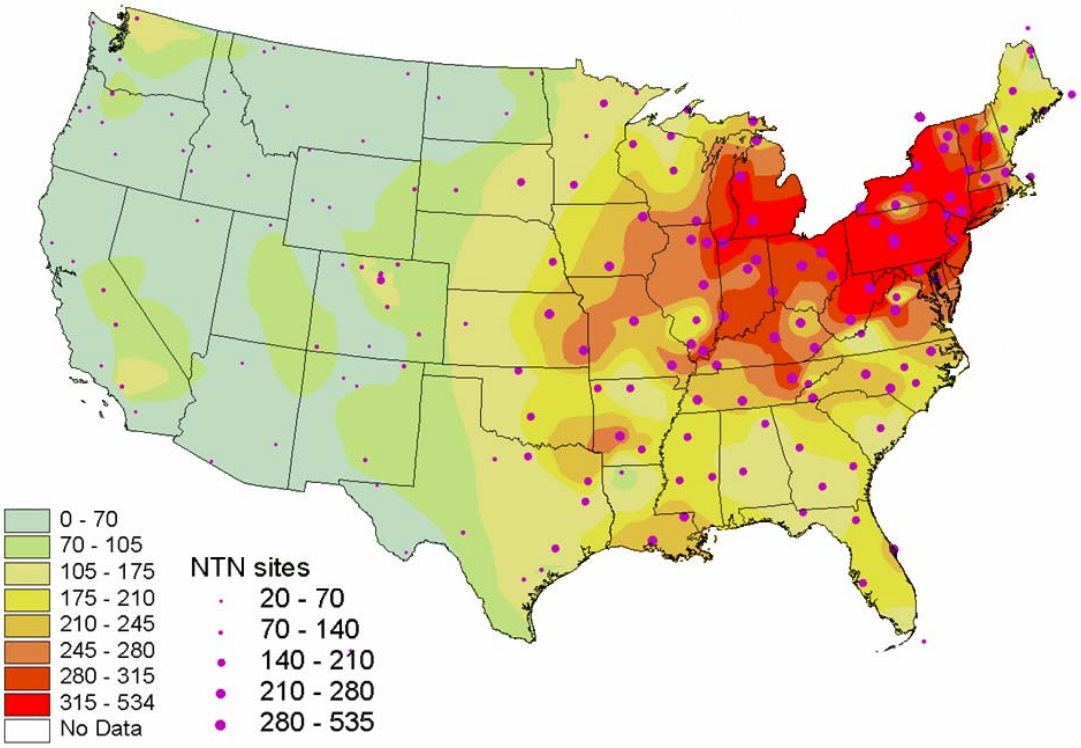

Figure 1.2. Spatially interpolated mean-annual estimates of wet nitrate deposition, based on measurements from National Atmospheric Deposition Program (NADP) sites. Units of deposition are kilograms kilometer^{-2} year^{-1}. [Image from J. W. Brakebill, U.S. Geological Survey, written comm., 2002.]

The spatial variability of a given variable is most readily increased in SPARROW models by expanding the spatial domain of the model to include larger drainage areas. This, of course, has the effect of increasing the number of monitoring stations, which also contributes to the potential for greater model complexity. Both of these effects can be seen in the greater complexity of the U.S. national, Chesapeake Bay, and New Zealand national nitrogen models (table 1.1, and fig. 1.1b). In comparison to the other regional models, these have expanded descriptions of terrestrial and aquatic loss processes and a broader description of the land uses and contaminant sources that supply nutrients to streams. In selecting water-quality monitoring sites for modeling, it is especially important to obtain sites that are located on a wide range of stream sizes (especially small streams) and are inclusive of impoundments of varying size. Nutrient removal rates are highly responsive to the hydraulic characteristics of streams and reservoirs (e.g., flow, velocity; Alexander and others, 2000; Alexander, Elliott, and others, 2002). Accounting for the wide variation in these properties can provide more accurate estimates of nutrient removal and improve the separation of land and water effects on transport in the model.

The accuracy of SPARROW model predictions as measured by the root mean square error (RMSE), a measure of the average model error, is also highly influenced by the range of stream water-quality and watershed conditions reflected in the calibration data. The level of model accuracy shows a general relation to the size of the model domain (i.e., total monitored drainage area) and the level of spatial variability in watershed properties within the domain (table 1.1). For example, the model accuracy is lowest (highest RMSE of 56 percent) for the U.S. national model, which also has by far the largest monitored drainage area used to estimate the model. By contrast, the regional models have higher accuracy with RMSE less than 47 percent, but are typically less complex, with six or seven significant parameters. The nutrient concentrations and watershed properties of the regional models typically span fewer orders of magnitude as compared with the national model. Despite the exceptionally large monitored drainage area of the Tennessee regional model (approximately 100,000 km^2), this model is the most accurate (RMSE of 19 percent) among all models. The relatively small variation in total nitrogen yields of only one order of magnitude in the watersheds (compared to two or more for the other models) contributes to this result; the watershed also shows generally small variation in certain ecological properties that may affect the supply and transport of nitrogen.

The accuracy of SPARROW model predictions is also intrinsically linked to the accuracy of the monitoring station flux estimates (i.e., the response variable of the SPARROW model), determined in a prior step to SPARROW modeling (see section 1.3.1). In evaluating water-quality monitoring stations for use in modeling, users should consider the characteristics of the water-quality records that affect the accuracy of monitoring station estimates of mean-annual contaminant flux. As described in table 1.3 of section 1.3.1.5, the accuracy of the mean-annual flux is correlated with various properties of the watersheds and the water-quality data. Important water-quality properties include the number of observations and the extent of coverage of the streamflow hydrograph by the observations. Smaller estimation errors often result from a larger number of observations and a more complete sampling of high-flow conditions. In general, estimation errors tend to be larger for the mean fluxes of smaller watersheds, which generally respond faster to changes in precipitation and thus have more variable streamflow. Estimation errors are much larger for constituents that are most affected by high flows, such as total phosphorus, suspended sediment, and fecal coliform bacteria. These constituents are generally more difficult to measure and exhibit larger variability in concentrations in streams, which can produce less precise estimates of the mean-annual flux. By contrast, dissolved substances, such as nitrate, sulfate, and total dissolved solids exhibit less variability with flow, and their fluxes can generally be more accurately estimated. Note that differences in the accuracy of the mean-annual flux estimates among the sites can be accounted for in SPARROW by executing a weighted SPARROW calibration (see discussion in section 1.5.2). Nevertheless, as pointed out in section 1.3.1, we recommend that users apply reasonable filters to the monitoring station data in advance of SPARROW modeling to eliminate stations with unacceptably few observations, those for which the range of streamflow conditions is poorly represented, and cases where the flux-estimation model fits the observed data poorly (e.g., as expressed by a high RMSE for the estimate).

On the basis of experience from national and regional modeling efforts, we recommend that stream fluxes, watershed data on pollutant sources, and watershed properties affecting contaminant transport vary over at least one order of magnitude. Stream water-quality records should be at least 2 years in length (at least 24-30 observations) and preferably include long-term continuous (i.e., daily) measures of streamflow, although the records may contain some gaps. The discussion in section 1.3.1 contains additional recommendations regarding the acceptability of monitoring data for estimating long-term flux.

The network of monitoring stations should consist of more than about 20 sites. To satisfy this requirement, SPARROW users may need to consider increasing the size of the study area (i.e., the watershed domain of the model) to increase the number of monitoring stations and also to possibly expand the variability reflected in the water-quality concentrations and explanatory factors in the model. Increasing the number of monitoring stations may be more easily accomplished if users are relying on historical stream water-quality monitoring data collected by state and local agencies (e.g., USGS National Water Information System (NWIS) or USEPA Storage and Retrieval (STORET) system). Expanding the study area to include nearby watersheds is likely to provide a much broader range of watershed conditions for the model, which will assist parameter estimation. An expansion of the study area may be especially needed to model contaminant sources or watershed properties that tend to show only modest variations over space on annual time scales, such as atmospheric nitrogen deposition or precipitation.

If increasing the size of the study area is not feasible, an alternative approach is to nest the smaller scale model inside an existing regional or national model that has greater variability in watershed characteristics and more monitoring stations. This approach, described in some detail in section 1.4.6, combines the watershed and monitoring data from a smaller study area with similar data available for a larger, existing study to form a hybrid model, one that includes the coefficients estimated from the larger-scale data along with a set of "adjustments" to those coefficients for which the smaller-scale data demonstrate there is a statistically significant difference.

1.2.5 Comparison of SPARROW with other watershed models

A wide variety of hydrologic and water-quality models have been used to describe contaminant sources and transport in watersheds and surface waters. These models can be characterized on the basis of their process complexity and the temporal and spatial scales that are used in the models (e.g., Singh, 1995). The level of complexity or process detail represented by model descriptions of hydrologic and biogeochemical processes commonly varies with the extent to which "deterministic" (i.e., mechanistic) and "statistical/empirical" methods are used to describe and estimate these processes (fig. 1.3; e.g., see Alexander, Johnes, and others, 2002). All models reflect some blend of these methods, but most place greater emphasis on one or the other type of model structure and process specification.

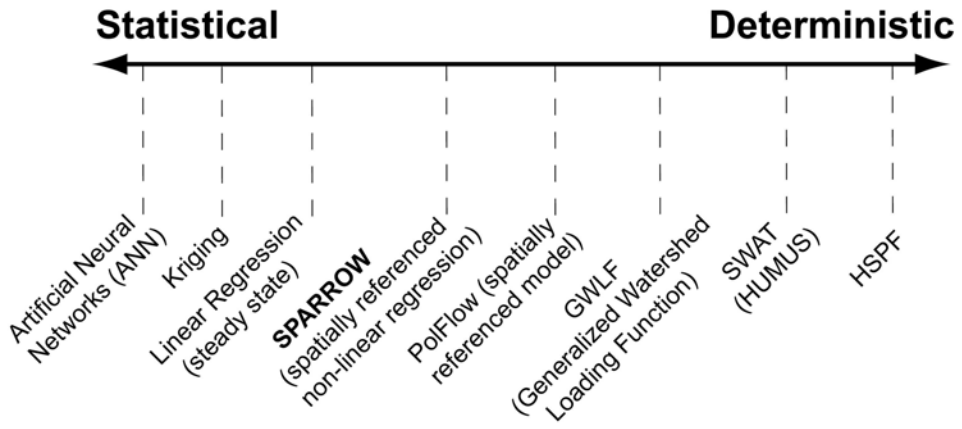

Figure 1.3. A simple continuum of model types based on the level of statistical and mechanistic descriptions of contaminant sources and biogeochemical processes.

In general, purely statistical models tend to reflect more simplistic model constructs. These models have a simple correlative mathematical structure and typically assume limited *a priori* knowledge of various processes. Conventional versions of these models (i.e., "linear regression") are expressed as simple linear (or log linear) correlations of stream measurements with watershed sources and landscape properties (e.g., Caraco and others, 2003; Peierls and others, 1991; Howarth and others, 1996). The methods have the advantage of being readily applied in large watersheds (often relying on generally available stream monitoring records) and can readily quantify the errors in model parameters and predictions. Simple correlative approaches, however, offer

little mechanistic explanation of contaminant sources and transport. They generally lack spatial detail on the distribution of sources and sinks within watersheds, rarely account for nonlinear interactions between sources and loss processes, and do not impose mass-balance constraints on contaminant transport. The most purely statistical approaches are found in artificial neural network and kriging techniques. These models commonly provide an excellent fit to the observations, but have the disadvantage of providing little understanding of the processes that affect contaminant transport.

By contrast, mechanistic water-quality models have a highly complex mass-balance structure that simulates hydrologic and contaminant transport processes, often according to relatively fine temporal scales (e.g., HSPF, Bicknell and others, 2001; SWAT, Srinivasan and others, 1993; INCA, Whitehead and others, 1998; AGNPS, Young and others, 1995). The components of these models frequently provide a highly detailed temporal description (e.g., daily, hourly) of the response of stream contaminants to climatic variability; the effects of coarser temporal variations in land use and management activities are often superimposed on the more detailed climatic variations). The mathematical descriptions of these responses are frequently based on *a priori* assumptions about the dominant processes and their reaction rates. The complexity of mechanistic simulation models creates intensive data and calibration requirements, which generally limits their application to relatively small watersheds. If the observed data for estimating the models are insufficient, reaction rates and other process components cannot be directly estimated in the models, but must be assumed. Because mechanistic water-quality models are frequently calibrated manually, robust measures of uncertainty in model parameters and predictions cannot be quantified. Two of the most commonly applied mechanistic water-quality models in large watersheds are HSPF (Hydrologic Simulation Program—Fortran; Bicknell and others, 2001) and SWAT (Soil Water Assessment Tool; Arnold and others, 1990). These models are now part of the USEPA's BASINS system (Better Assessment Science Integrating Point and Nonpoint Sources; U.S.EPA, 2001; *http://www.epa.gov/docs/ostwater/BASINS/*) for use in development of TMDL (Total Maximum Daily Load) assessments. The models are a common choice among water-resource managers to address various water-quality assessment needs.

Despite the common use of highly complex mechanistic models, especially among water-resource managers, there are growing concerns about whether sufficient water-resources data and knowledge of biogeochemical processes exist to reliably support the general use of such highly complex descriptions of processes (Beven, 2002; Jakeman and Hornberger, 1993). The intensive data and calibration requirements of these models are costly and can rapidly exceed project budgets. Without sufficient data, there is limited ability to apply formal parameter-estimation techniques, which are needed to quantify model uncertainties and to identify unique models having parameters that are sensitive and uncorrelated. Non-unique models are those for which nearly identical model predictions result from the use of different parameter sets and values. These models may have large uncertainties in the interpretability of the parameters and their characterization of the effects of specific processes (e.g., Wade and others, 2002), such as denitrification in streams (Boyer and others, in press; Filoso and others, 2004). Recently, there has been considerable discussion among hydrologists (e.g., Hill, 1998; Bevin, 2002; Duan and others, 2003) of the problems with non-unique models (i.e., statistical models with under-identified parameters), which have commonly been reported in the literature. As a result, there is growing recognition that increases in model complexity beyond certain limits results in only marginal improvements in model accuracy and interpretability (fig. 1.4; Hill, 1998). The notion is that a large fraction of the total variability in the observations can frequently be explained by a relatively small number of model parameters—increases in the number of parameters beyond these limits are likely to have only marginal increases in explanatory value (fig. 1.4). In some cases, if erroneous values of parameters are selected (e.g., "default" parameter values), model accuracy could potentially decline. The estimation of parameters using nonlinear optimization or inverse modeling techniques, such as those in PEST (Doherty, 2004) and UCODE (Poeter and Hill, 1998), and the development of other optimization methods (Duan and others, 2003) have been promoted as ways of identifying how much model complexity can be supported by the data used for parameter estimation. The use of parameter estimation has been growing with ground-water models, such as MODFLOWP (Hill, 1992, 1998). There also is increasing recognition of the value of using relatively simple statistical models for purposes of conducting TMDL assessments (National Research Council, 2001a) and investigating the sources, transport, and fate of contaminants over large spatial scales (National Research Council, 2000; Bricker and others, 1999; Alexander and others, 2000).

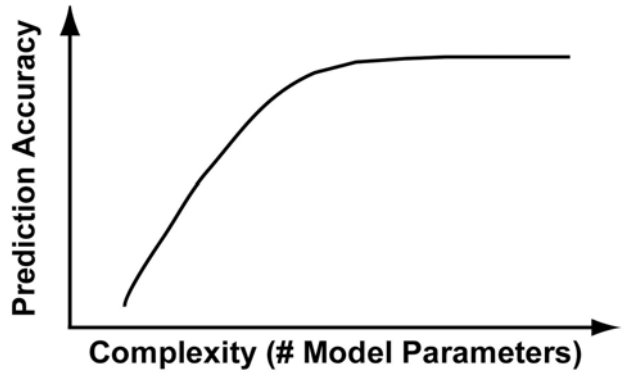

Figure 1.4. Hypothetical illustration of the relation of model accuracy to model complexity.

By comparison to other types of water-quality models, SPARROW may be best characterized as a hybrid process-based and statistical modeling approach for estimating pollutant sources and contaminant transport in surface waters. The mechanistic mass transport components of SPARROW include surface-water flow paths (channel time of travel, reservoirs), non-conservative transport processes (i.e., first-order in-stream and reservoir decay), and mass-balance constraints on model inputs (sources), losses (terrestrial and aquatic losses/storage), and outputs (riverine nutrient export). Separating land and water components provides estimates of the rates of pollutant delivery from point and diffuse sources to stream reaches and locations along downstream reaches, including reservoirs and estuarine waters. The statistical features of the model involve the use of nonlinear parameter-estimation techniques. Parameters are estimated by spatially correlating stream water-quality records with geographic data on pollutant sources (e.g., atmospheric deposition, fertilizers, human and animal wastes) and climatic and hydrogeologic properties (e.g., precipitation, topography, vegetation, soils, water routing). Parameter estimation ensures that the calibrated model will not be more complex than can be supported by the data. This provides an objective statistical approach for evaluating alternative hypotheses about important contaminant sources and controlling transport processes over large spatial scales in watersheds. Indeed, the model has been shown to improve the accuracy and interpretability of model parameters and the predictions of nutrient loadings and sources in streams as compared with conventional statistical modeling approaches (Smith and others, 1997; Alexander and others, 2000; Alexander, Elliott, and others, 2002; Alexander, Johnes, and others, 2002). As with any model that relies upon parameter estimation, SPARROW may be most reliably used in watersheds that have appreciable spatial variation in water-quality concentrations and explanatory factors as well as large numbers of stream monitoring stations (see section 1.3.1), such as those from state and local monitoring programs. The statistical estimation of parameters in SPARROW is critically important to provide robust measures of uncertainty in both the model coefficients and water-quality predictions.

Few other source-transport watershed models have used spatially referenced river networks in large watersheds together with simple process-based descriptions of sources and transport such as those used in SPARROW. One exception is the regional-scale spatially distributed watershed model, PolFlow (De Wit, 2000, 2001), which has been used to describe mean-annual total nitrogen flux in large European watersheds of the Rhine and Elbe Rivers. The model (De Wit, 2000) was recently expanded (De Wit, 2001) to include water and nutrient routing components similar to that of SPARROW. The model accounts for various natural and anthropogenic nitrogen sources, storage and permanent loss of nitrogen in soils and ground water, and nitrogen delivery to surface waters. The functional relation of nitrogen removal to stream size, channel slope, and water velocity is generally similar to that of SPARROW (Boyer and others, in press); however, the specification of the aquatic decay variables does not provide first-order rate expressions, meaning that units are neither reciprocal time (i.e., reaction-rate coefficient) or length per time (i.e., mass-transfer coefficient), so that estimated coefficients cannot be directly compared with literature or SPARROW rates (see, the SPARROW comparisons in section 1.3.1.3). Moreover, PolFlow model coefficients typically have been estimated by using manual "trial-and-error" type approaches (De Wit, 2001) rather than using formal optimization routines as in SPARROW, and thus do not provide measures of the uncertainties in model parameters and predictions.

A number of inter-model comparisons have been previously conducted that compare the performance and properties of SPARROW to those of other water-quality models. These include national comparisons with

SWAT (Soil Water Assessment Tool) and comparisons in the Chesapeake Bay watershed with HSPF (see Alexander and others, 2001). Inter-model comparisons also have been conducted with statistical and quasi-deterministic models in watersheds of the northeastern U.S. (see Alexander, Johnes, and others, 2002; Seitzinger and others, 2002). Finally, evaluations of the SPARROW technique also are included as part of a number of recent National Research Council (NRC) reports (2000, 2001a, 2001b, 2002a, 2002b). At least two of the NRC reports (2000, 2001a) have noted the advantages of statistical modeling approaches, such as SPARROW, for general water-quality assessment and use in TMDL assessments.

1.3 SPARROW model infrastructure

This section describes the major components of the spatial infrastructure that support the calibration and application of SPARROW models. The parameters of SPARROW models are statistically estimated with nonlinear regression techniques by spatially correlating water-quality flux estimates at monitoring stations with watershed data on sources, and landscape and surface-water properties that affect transport. The calibrated models are then used to predict flux, total and disaggregated by contributing source, for stream reaches throughout a river network. A "flow" diagram is shown in figure 1.5 below to illustrate the functional linkages between the major spatial components of SPARROW models. Details of the SPARROW model specifications and mathematical equations supported by these components are presented in section 1.4.

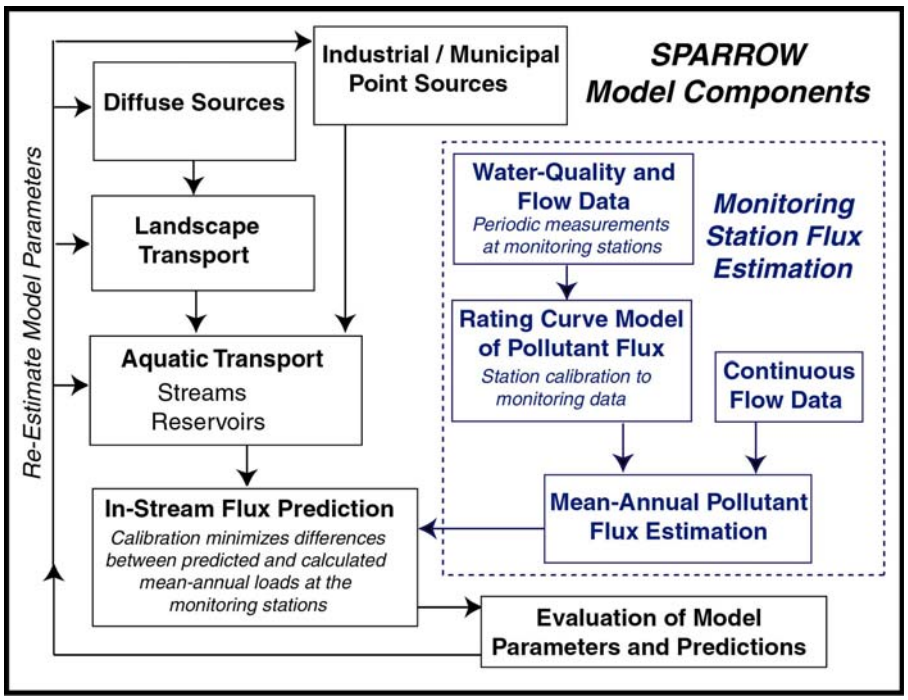

Figure 1.5. Schematic of the major SPARROW model components. [Modified from Alexander, Elliott, and others (2002).]

In the following subsections, we briefly outline the pre-processing steps that are required to develop *reach-level* information for the major components of the SPARROW model infrastructure shown in figure 1.5. *Monitoring station flux estimation* refers to the estimates of long-term flux used as the response variable in the model. Flux estimates at monitoring stations are derived from station-specific models that relate contaminant concentrations from individual water-quality samples to continuous records of streamflow and time. The methods required to derive these estimates are described in section 1.3.1. The *stream reach* (and its incremental contributing drainage basin) is the most elemental spatial unit of the infrastructure used to estimate and apply the basic SPARROW models described in this report. Stream reaches typically define the length of stream channel that extends from one stream tributary junction to another. In section 1.3.2, we provide a conceptual description of the stream reach network to support SPARROW modeling. The remaining sections describe in general terms the explanatory data required for input to SPARROW water-quality models. Explanatory data (e.g., climate, topography, land use) are frequently compiled according to geographic units that are not

coincident with the drainage basin boundaries of river reaches. These data may be collected at different spatial scales and according to spatial units that reflect political (e.g., counties) or other non-hydrologic features of the landscape. Basic methods are discussed for the assignment of explanatory variable characteristics to reaches, and references to other methods that use conventional GIS software to obtain watershed properties by reach are provided.

1.3.1. Monitoring station flux estimation

The estimation of a SPARROW model requires estimates of long-term mean flux from a spatially distributed set of monitoring stations within the study area. To produce such estimates, a monitoring station must have a sufficiently long period of record. Because water quality is rarely measured with the frequency necessary to directly estimate long-term mean flux, an indirect method is needed to estimate flux from available measurements to account for hydrologic conditions during periods when water-quality measurements are not made. The estimation procedure must account for the fact that water-quality measurements are not always collected in a random manner, implying the arithmetic average of flux is not a reliable estimate of mean flux. Rather, the measurements of flux must be related in some way to other information, collected on a continuous basis, which exhibits a close relation to flux. Invariably, this other information is streamflow. The limitation, therefore, in constructing a mass-balance model using long-term mean flux is that only water-quality measurements that can be associated with a continuous record of streamflow will be suitable for model estimation.

In order to reliably estimate the effects of processes affecting contaminant transport, it is necessary to include data from a diverse set of monitoring stations, inclusive of a wide range of spatial scales and expressing considerable variation in predictor variable conditions. Unfortunately, a problem often arises in developing such a set of stations: monitoring stations commonly have different periods of record. Long-term variations in hydrologic conditions, combined with long-term trends in water quality, imply mean fluxes computed over different periods may not be directly comparable. If there are trends in water quality, the estimate of long-term mean water-quality flux will depend on the period or window through which the water-quality data are acquired.

The problem of incomparable periods of record is perhaps best understood in the context of two nested stations—the situation in which one station is located upstream from another station. In SPARROW, the flux estimate for an upstream station is used to determine the flux at a downstream station (see section 1.4.1 for details). This conditioning of downstream flux on upstream flux serves to minimize error and reduces the correlation of errors between nested basins. If the two stations exhibit trend, however, and there are differences in the period of record, the measured mean flux between the two stations could be gained or lost for reasons that are unrelated to intervening processes affecting flux. Under ideal conditions, the incomparability of station record simply creates additional noise in the model. But bias could result if, as is typically the case, record length is a function of stream characteristics. For example, it is common for stations on larger rivers to have longer periods of record than stations on smaller rivers.

Another context in which trend in water quality creates problems for the mass balance approach arises in relating water quality to its predictor variables, principally the source variables. Ideally, the source variables will consist of a time series of estimates, and these source variables would be included in SPARROW as long-term averages in the same way the flux estimates are included. A problem with this formulation is that source information is rarely available for multiple periods. Even if source information is available over time, the period over which it is compiled will rarely match the period of the flux information; for example, point-source and land-use data are infrequently available due to the high cost of compiling this information. Other information, such as atmospheric deposition and soil erosion, is compiled on a periodic schedule, but the periods covered by these data generally differ from those covering the water-quality data.

To address the problem of incompatibility in periods of record, a SPARROW model is typically specified for a single base year; that is, all water-quality and source information is assumed to pertain to a given point in time. The severest constraint on the period of the analysis is imposed by source information that is available only for a single year. Consequently, the base year is generally selected to be in the middle of the range of available years for this ‘one-time’ information. For other information that is available over a range of years, such as contaminant flux, the base year estimate is constructed by detrending the data.

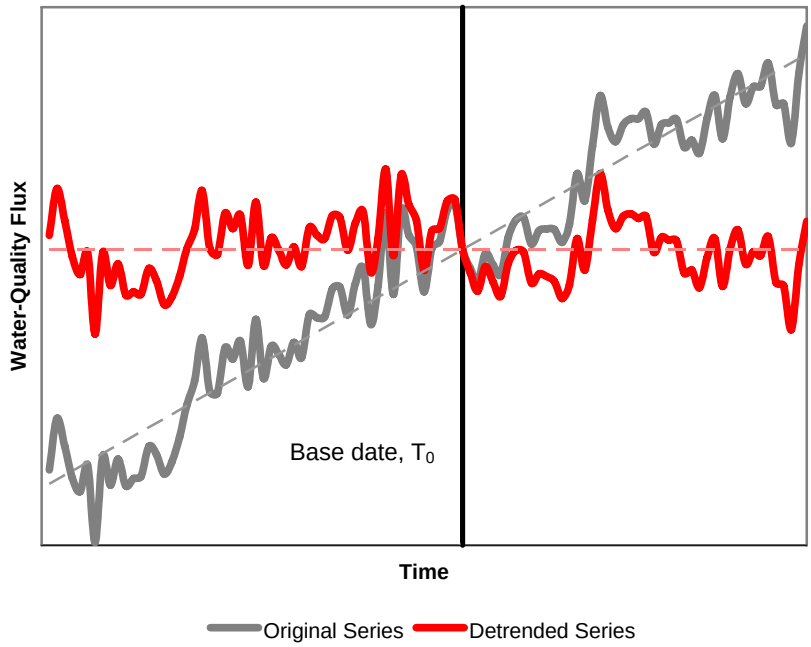

Figure 1.6. A graphical depiction of a detrended series.

The effect of detrending data is shown graphically in figure 1.6. The original time series, shown in gray, displays an erratic upward trend. The first step in the process of detrending is to fit a simple function of time, in this case a straight line, through the data (see the upward sloping line in the figure). A constant is then subtracted from the function of time to create an adjustment function to be applied to the original series; the constant used to create the adjustment function is chosen to make the adjustment function equal to zero for a particular point in time, T_0, referred to as the 'base date.' The detrended series is then obtained by simply subtracting the adjustment function, a function of time only, from the original series. In the figure, the subtraction of the adjustment function produces the erratic but non-trending series shown in red. Note that the effect of the adjustment is to shift the original positive trending series upward for periods before the base date T_0 and downwards for periods after the base date T_0; the value of the series is unchanged for the base period, T_0. The detrended series can be interpreted as the series that would have been observed if the dynamic factors causing trend, whatever they might be, were held constant throughout the entire period, equal to the values they had on the base date T_0; all other dynamic factors determining the short-term variations in the series are left unchanged. This is a purely counterfactual scenario, a 'what-if' exercise in which the factors determining trend are frozen for all time at their base period values. Thus, for example, peak flow events affecting the original series would remain in the detrended series; however, gradual improvements in water quality resulting from the adoption of improved technology over time would be substituted for technology that was in place on the base date, T_0.

In mathematical terms, the process of detrending can be described as follows. Let $h(t)$ be the function of time used to describe trend through the original series $X(t)$. The detrended series $X^*(t)$ is given by

$$X^*(t) = X(t) - \left(h(t) - h(T_0)\right). \tag{1.1}$$

In equation (1.1), the term $h(t) - h(T_0)$ is the adjustment function and the constant $h(T_0)$ is the constant that causes the adjustment to equal zero for the base date, T_0.

It is important that the function of time $h(t)$ used to detrend a series not be so variable that it removes cyclical or short-term variations. In the parlance of spectral analysis, detrending only filters out the power in the

lowest frequencies of the series, the frequencies corresponding to the long-term variation in the data; detrending should not affect the power of the higher frequencies associated with the short-term variations in the data. Simple functions that meet this requirement are the linear and quadratic functions of time. A step function in time would also be acceptable, as long as the number of steps is not too large. Another common filter for removing trend is the first difference, in which the previous value of the series is subtracted from the current value. This method is not typically used for detrending water-quality data because the data are not sampled with a fixed frequency.

Because the function that describes trend, $h(t)$, must be estimated from the data, the act of detrending necessarily introduces uncertainty into the analysis. This uncertainty can be reduced by estimating the trend function as accurately as possible. Generally, this implies the estimation of $h(t)$ in the context of a model that explains as much as possible the variation in the original series, $X(t)$. For many applications, the requirement that a model be specified in order to remove trend introduces little additional burden into the analysis; if $X(t)$ is a water-quality variable, a model is already required in order to extrapolate the infrequently sampled water-quality data to other periods so as to obtain a better estimate of long-term mean flux. The only additional consideration arising from trend removal is that the specified model must include a time component, $h(t)$. For other analyses, such as the detrending of streamflow, the data are often continuously available. In these cases, the specification and estimation of a model represents an additional step that would otherwise not be required.

The detrending of water-quality flux serves another purpose that is perhaps more useful than its role in equalizing station records. In forming an estimate of the long-term mean flux, it is advantageous to base the estimate on a long and rich history of hydrologic events. Many streamflow stations, in fact, have very long records to support this objective. The same is not generally true, however, for water-quality stations. A conflict arises if the relatively short water-quality record expresses a trend: How reliably can the trend estimated from a short water-quality record be extrapolated to the entire period for which streamflow data are available? Detrending provides a conservative means of extrapolating the water-quality model without having to accept a potentially misspecified relation with trend. If the trend relation is not misspecified, detrending reduces the sampling error arising from excessive extrapolation of the model beyond the conditions under which it is estimated. Conversely, if trend is misspecified, the process of detrending effectively erases the bias induced by improper extrapolation of trend. The final result is an estimate of flux having reduced error.

1.3.1.1 Model specification for monitoring station flux estimation

The extrapolation of infrequently sampled water-quality data and removal of trend dictates the specification of a model of flux. To be used for extrapolation, the model must relate infrequently measured data to variables that are measured continuously over time, and to accommodate detrending, the model must include a function of time. For water-quality applications, these requirements suggest a model that relates infrequently measured concentration to the variables streamflow and time. The inclusion of streamflow as an explanatory factor serves another purpose in the analysis. Because flux is typically positively related to streamflow, it has become common practice to bias water-quality sampling towards high flow events. The inclusion of flow in the water-quality model effectively conditions the estimation of flux so as to remove the effects of high-flow sampling bias.

The estimation of mean flux by a station-specific model need not account for all the processes affecting flux within a basin; this task is assigned to the SPARROW model. All that is required is that the estimated mean flux at a station be reflective of long-term average processes within the basin. This association with process could be realized by explicitly accounting for processes via the inclusion of explanatory variables; but it is also implicitly revealed in the values taken by the estimated coefficients of a less refined station-specific flux model. Thus, the station-specific flux models need not be structurally accurate; they need only be predictively accurate. A causal explanation of flux is ultimately obtained through application of a SPARROW model, whereby variations in mean flux conditions across stations, estimated either explicitly or implicitly, are correlated with variations in basin attributes across space.

There are a number of ways to specify the relation between water-quality data and the explanatory variables streamflow and time. Cohn, Caulder, and others (1992) have suggested a simple seven-parameter

model in which the logarithm of contaminant concentration $\tilde{c}$ is related via a linear model to an intercept, the logarithm of flow $\tilde{q}$, the square of the logarithm of flow, decimal time T (decimal time is time expressed as a decimal number, with whole numbers representing the year and decimal numbers representing the fraction of the year), decimal time squared, and a seasonal harmonic consisting of two trigonometric terms—the sine and cosine of 2π times decimal time. The quadratic flow term is included to account for nonlinearity in the water-quality/flow relation; for example, it is common to observe the percentage increase in sediment concentration from a percentage increase in flow become greater for very high flows.

The seven-parameter model for period *t* is written as

$$\tilde{c}_t = \gamma_0 + \gamma_q \tilde{q}_t + \gamma_{q^2} \tilde{q}_t^2 + \gamma_T T_t + \gamma_{T^2} T_t^2 + \gamma_s \sin(2\pi T_t) + \gamma_c \cos(2\pi T_t) + e_t, \tag{1.2}$$

where γ_0, γ_q, γ_{q^2}, γ_T, γ_{T^2}, γ_s, and γ_c are coefficients to be estimated, and e_t is an independent, normally distributed error term, uncorrelated with each of the predictor variables, having mean zero and variance σ_e^2. Higher order harmonics, given by $\sin(j2\pi T_t)$ and $\cos(j2\pi T_t)$ terms, where *j* is an integer greater than 1, could be included in equation (1.2) to account for more complex seasonal patterns.

Vecchia (2000) has argued that there are important long-term lags affecting the relation between water quality and flow. He suggests a specification that relates the log of contaminant concentration to a set of compound flow terms consisting of moving averages of flows, of various lengths, in addition to time trend terms. A specification that generalizes both the seven-parameter and the lag flow models takes the form

$$\tilde{c}_t = \mathbf{M}(\mathbf{Q}_t)\boldsymbol{\beta}_Q + \mathbf{h}(T_t)\boldsymbol{\beta}_T + \mathbf{X}_t\boldsymbol{\beta}_X + e_t, \tag{1.3}$$

where $\mathbf{Q}_t$ is a *p*-element row vector consisting of current and lagged logarithms of flow, $\mathbf{Q}_t = \{\tilde{q}_t, \ldots, \tilde{q}_{t-p+1}\}$; $\mathbf{M}(\mathbf{Q}_t)$ is a vector function that transforms the *p*-element logged flows into a K_q element row vector; $\boldsymbol{\beta}_q$ is a K_q element vector consisting of coefficients associated with the transformed flow terms; $\mathbf{h}(T_t)$ is a K_T element row vector function of decimal time; $\boldsymbol{\beta}_T$ is a K_T element vector of coefficients associated with the transformed decimal time terms; $\mathbf{X}_t$ is a K_X element row vector of other exogenous variables affecting water quality; $\boldsymbol{\beta}_X$ is a K_X element vector of coefficients associated with the other exogenous variables; and e_t is the normally distributed error term, independent over time and uncorrelated with each of the predictor variables, having a mean of zero and variance σ_e^2. In the sequel, a model variable displayed in bold font pertains to a vector or matrix.

The function $\mathbf{M}(\mathbf{Q}_t)$ is quite general and can accommodate many different specifications of the flow variables. For example, in the seven-parameter model *p* equals one, $\mathbf{Q}_t$ equals $\tilde{q}_t$, and $\mathbf{M}(\mathbf{Q}_t)$ is a two-element row vector function with elements $\{\tilde{q}_t, \tilde{q}_t^2\}$. In the compound flow model developed by Vecchia (2000), the function $\mathbf{M}(\mathbf{Q}_t)$ forms various moving averages of the current and lagged flow terms. The only restriction to be placed on $\mathbf{M}(\mathbf{Q}_t)$ is that it be continuous in its arguments.

The function $\mathbf{h}(T_t)$ is included to account for long-run variations in water quality that are unrelated to flow. Typically, these changes are assumed to arise from management changes that affect water quality—for example, the construction of new wastewater treatment plants or the imposition of regulations that affect the runoff of contaminants from farms. In the case of the seven-parameter model, $\mathbf{h}(T_t)$ consists of two elements, with $\mathbf{h}(T_t) = \{T_t, T_t^2\}$. Alternatively, $\mathbf{h}(T_t)$ could specify a step change function to account for one-time

permanent changes in water quality within a basin. Note that $\mathbf{h}(T_t)$ does not represent cyclical or other short-term variations in time, such as the sine and cosine terms in equation (1.2); these variations are assumed to be included in the other exogenous variables, $\mathbf{X}_t$.

An important feature of equation (1.3) is that it is linear in its parameters. If no censored observations are included in the water-quality data, the model can be estimated easily using ordinary least squares. If censored observations are present, Cohn, Gilroy, and others (1992) suggest using the maximum likelihood method for parameter estimation. For Type I censored data, corresponding to data in which the censoring threshold is known, the appropriate maximum likelihood method is given by the Tobit model (Cohn, Gilroy, and others, 1992). The Tobit model is nonlinear and must be estimated using iterative methods (see section 1.5 for a description of nonlinear estimation methods in the context of SPARROW model estimation). Consequently, the small sample properties of the coefficient estimates are not easily obtained. However, being a maximum likelihood method, the estimated parameters are consistent and efficient in large samples. Moreover, it has been shown (see Maddala, 1983) that the model with linear coefficients, as in equation (1.3), has a single maximum, corresponding to its global maximum, assuring the convergence of iterative methods.

The likelihood function of the Tobit model is given as follows. Let $\mathbf{Z}_t$ represent the row vector of all explanatory variables in (1.3), so that $\mathbf{Z}_t = \left[\mathbf{M}(\mathbf{Q}_t) \quad \mathbf{h}(T_t) \quad \mathbf{X}_t\right]$, and let $\boldsymbol{\beta} = \left[\boldsymbol{\beta}_Q' \quad \boldsymbol{\beta}_T' \quad \boldsymbol{\beta}_X'\right]'$ be a $K \times 1$ vector of the combined associated coefficients; let the water-quality sample consist of *N* observations; let $\overline{c}_t$ represent the censoring threshold (in logarithm space) for observation *t*; and let d_t be an indicator variable that equals 1.0 if observation *t* is censored and equals zero otherwise. In the context of estimating equation (1.3), the Tobit log likelihood function is given by

$$\tilde{L} \propto -N \ln(\sigma_e) + \sum_t d_t \tilde{\Phi}\left(\frac{\overline{c}_t - \mathbf{Z}_t\boldsymbol{\beta}}{\sigma_e}\right) + (1-d_t)\tilde{\phi}\left(\frac{\tilde{c}_t - Z_t\boldsymbol{\beta}}{\sigma_e}\right), \tag{1.4}$$

where $\tilde{\Phi}(\cdot)$ and $\tilde{\phi}(\cdot)$ are the natural logarithms of the standard normal cumulative distribution and density functions.

Estimation by maximum likelihood requires finding the values of the parameters $\boldsymbol{\theta} = \left[\boldsymbol{\beta}' \quad \sigma_e\right]'$ that maximize equation (1.4). Because of the global convexity of $\tilde{L}$ (Maddala, 1983), this is equivalent to finding the roots of the first-order partial derivatives $\tilde{\mathbf{L}}_\theta = 0$. The covariance matrix of the maximum likelihood estimates, $\hat{\boldsymbol{\theta}}$, is given by (Cramer, 1986)

$$\mathbf{V}\left(\hat{\boldsymbol{\theta}}\right) = -\left[E\left(\tilde{L}_{\boldsymbol{\theta}\boldsymbol{\theta}'}\right)\right]^{-1} = \left[E\left(\tilde{L}_{\boldsymbol{\theta}}\tilde{L}_{\boldsymbol{\theta}}'\right)\right]^{-1}, \tag{1.5}$$

where the expectation is over alternative realizations of the dependent variable sample, $\tilde{c}$. For independent observations, the covariance matrix can be consistently estimated by

$$\hat{\mathbf{V}}\left(\hat{\boldsymbol{\theta}}\right) = \left[N^{-1}\sum_{t=1}^{N} \hat{\tilde{L}}_{t\boldsymbol{\theta}}\hat{\tilde{L}}_{t\boldsymbol{\theta}}'\right]^{-1}, \tag{1.6}$$

where,

$$\hat{\tilde{L}}_t \propto -\ln \hat{\sigma}_e + d_t \tilde{\Phi}\left(\frac{\overline{c}_t - \mathbf{Z}_t\hat{\boldsymbol{\beta}}}{\hat{\sigma}_e}\right) + (1-d_t)\tilde{\phi}\left(\frac{c_t - \mathbf{Z}_t\hat{\boldsymbol{\beta}}}{\hat{\sigma}_e}\right), \tag{1.7}$$

and $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_e$ are the maximum likelihood estimates of the parameters $\boldsymbol{\beta}$ and σ_e. Equation (1.6) is a popular estimator in applied work due to its ease of computation.

1.3.1.2 Monitoring station flux prediction (advanced)

The coefficient estimates from the water-quality sample can be used to predict flux for every day in which there is a complete observation of the explanatory variables. A complication in making this estimate is the need to retransform the estimates from logarithm space, the space in which the model is estimated, back into real space, the space that supports mass balance. Because the model is estimated in logarithm space, the retransformation to real space requires an application of the exponential function. The convexity of the exponential function implies a random variable having a zero mean in logarithm space will have an expectation that is greater than 1.0 upon transformation into real space. This deviation from 1.0 of the expected transformed random variable is called *retransformation bias*. Two sources of uncertainty contribute to retransformation bias: uncertainty in the estimated model coefficients—sometimes called sample error because the error is eliminated in large samples—and uncertainty caused by model error, the variation in flux that is unexplained by the model's explanatory variables.

To demonstrate the full nature of the retransformation bias, consider first the unbiased estimate of flux in the case that model coefficients are known. In this case, the only source of uncertainty is the model error. Because model error cannot be removed, estimates of flux are expressed as *expectations* of flux *conditioned* on known components of the model. In logarithm space, actual flux in period *t* is given by

$$f_t = \mathbf{Z}_t\boldsymbol{\beta} + \tilde{q}_t + e_t \,. \tag{1.8}$$

The logarithm of actual flux depends on a modeled component, the modeled component of concentration $\mathbf{Z}_t\boldsymbol{\beta}$, plus the logarithm of streamflow (any conversion constants, in logarithm space being represented by additive constants, are assumed to be subsumed in the model intercept, an element of $\mathbf{Z}_t$), and a non-modeled component, given by the model error.

Before deriving the conditional expectation of flux, we state a simple mathematical fact that is indispensable to understanding the retransformation problem. The fact is this: if *x* is a normally distributed variable with mean μ and variance σ_x^2, the expectation of $\exp(x)$ is $\exp\left(\mu + \sigma_x^2/2\right)$. From this fact, the conditional expectation of flux in real space is given by

$$\overline{F}_t = E\left(e^{f_t}\right) = \exp\left(\mathbf{Z}_t\boldsymbol{\beta} + \tilde{q}_t\right) E\left(\exp\left(e_t\right)\right) = \exp\left(\mathbf{Z}_t\boldsymbol{\beta} + \tilde{q}_t\right)\exp\left(\sigma_e^2/2\right). \tag{1.9}$$

The conditional expectation of flux in real space is given by the conditional expectation of flux in logarithm space, transformed to real space (the first exponential term in the third equality in equation (1.9)), times a retransformation factor that depends on the variance of the model residual σ_e^2 (the second exponential term of the third equality in equation (1.9)). Because the error variance σ_e^2 is necessarily positive, the retransformation factor is necessarily greater than 1.0, and will be equal to 1.0 only if the model has no error. Consequently, unless the retransformation factor is included in the prediction, the estimated flux is biased downwards—a result that is obtained even if the coefficients of the model are known without error.

We now extend the analysis by assuming the coefficients of the model are not known but must be estimated. Assume the water-quality data contain no censored values so that the water-quality model given by equation (1.3), with explanatory variables denoted by the row vector $\mathbf{Z}_t$, can be estimated efficiently without bias using ordinary least squares. Let $\tilde{\mathbf{c}}$ denote the $N \times 1$ vector of logarithm transformed water-quality concentration measurements, $\mathbf{Z}$ the $N \times K$ matrix of explanatory variables, and $\mathbf{e}$ the $N \times 1$ vector of independent,

identically normally distributed errors with variance $\sigma_e^2\mathbf{I}$. The predicted flux for period *t* in logarithm space is given by

$$\begin{aligned}\hat{f}_t &= \mathbf{Z}_t\hat{\boldsymbol{\beta}} + \tilde{q}_t = \mathbf{Z}_t\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\tilde{\mathbf{c}} + \tilde{q}_t \\ &= \mathbf{Z}_t\boldsymbol{\beta} + \tilde{q}_t + \mathbf{Z}_t\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{e}.\end{aligned} \tag{1.10}$$

A naïve estimate of flux in real space is simply an exponential transformation of the predicted flux $\hat{f}_t$ given in equation (1.10)

$$\hat{F}_t = \exp\left(\hat{f}_t\right) = \exp\left(\mathbf{Z}_t\hat{\boldsymbol{\beta}} + \tilde{q}_t\right), \tag{1.11}$$

with expectation given by

$$E\left(\hat{F}_t\right) = \exp\left(\mathbf{Z}_t\boldsymbol{\beta} + \tilde{q}_t\right)\exp\left(\tfrac{1}{2}\sigma_e^2\mathbf{Z}_t\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}_t'\right). \tag{1.12}$$

The ratio of the actual conditional expectation of flux given in equation (1.9) and the naïve estimate in equation (1.12) represents the inverse of the retransformation bias factor associated with the naïve estimate. The inverse retransformation bias factor is

$$\frac{\bar{F}_t}{E\left(\hat{F}_t\right)} = \exp\left(\tfrac{1}{2}\sigma_e^2\left(1 - \mathbf{Z}_t\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}_t'\right)\right). \tag{1.13}$$

The term in the exponential function on the right-hand side of the equality in equation (1.12) can be positive or negative, implying the bias factor associated with the naïve estimator can be greater than or less than 1.0—although it is likely to be positive if the model is not greatly extrapolated to conditions outside the range of explanatory variables for the sample. A somewhat more complicated expression for the inverse retransformation bias factor is obtained if the water-quality data include censored data.

The problem of retransformation bias is solved if it is possible to derive a sample-based correction factor that is independent of $\hat{f}_t$ and has an expectation equal to the right-hand side of equation (1.12). Cohn, Gilroy, and others (1992) describe a method that can be used to remove first-order retransformation bias for cases with and without censored data. The method consists of first transforming the $\hat{\boldsymbol{\beta}}$ vector so that it is uncorrelated with $\hat{\sigma}_e$. The transformation takes the form $\hat{\mathbf{w}} = \hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\hat{\sigma}_e$, where $\boldsymbol{\gamma}$ is the correlation coefficient between $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_e$ as determined from the covariances estimated in equation (1.6). For the case without censoring, the ordinary least squares estimates of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_e$ are already uncorrelated, so no adjustment is needed. In large samples, $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_e$ are normally distributed so the lack of correlation between $\hat{\mathbf{w}}$ and $\hat{\sigma}_e$ implies they are independent; because $\hat{\sigma}_e^2$ is simply the squared value of $\hat{\sigma}_e$, it must also be independent of $\hat{\mathbf{w}}$.

It can be shown that the inverse retransformation bias factor for a naïve estimate of flux based on $\hat{\mathbf{w}}$ depends critically on σ_e and σ_e^2, both unknowns. The retransformation bias problem is solved, therefore, by deriving a function $p_t\left(\hat{\sigma}_e^2\right)$ that has an expectation equal to the inverse retransformation bias factor. Finney (1941) derives the appropriate function in the case of no censored data; Cohn, Gilroy, and others (1992) derive an analogous function to be used for data that include censored observations.

The unbiased estimate of flux in real space, for period *t*, is given by

(1.14) $$\hat{F}_t^* = \exp\left(\mathbf{Z}_t \hat{\mathbf{w}}\right) p_t\left(\hat{\sigma}_e^2\right),$$

the expectation of which is equation (1.9). An estimate of mean daily flux is obtained by averaging these period *t* estimates over all N_{T^P} days in the prediction period T^P

(1.15) $$\hat{F}^* = N_{T^P}^{-1} \sum_{t \in T^P} \hat{F}_t^* .$$

To ensure that seasonal patterns in flux are equally represented in the mean estimate, it is best to include predictions for only those days in which a complete year of estimates is available.

Gilroy and others (1990) and Cohn, Gilroy, and others (1992) derive the standard error of the mean flux as

(1.16) $$\sigma_{\hat{F}^*} = N_{T^P}^{-1} \sqrt{\sum_{t \in T^P} \sum_{s \in T^P} \mathrm{Cov}\left(\hat{F}_t^*, \hat{F}_s^*\right)},$$

where

(1.17) $$\begin{aligned} \mathrm{Cov}\left(\hat{F}_t^*, \hat{F}_s^*\right) &= E\left(\hat{F}_t^* \hat{F}_s^*\right) - E\left(\hat{F}_t^*\right) E\left(\hat{F}_s^*\right) \\ &= E\left(\exp\left(\left(\mathbf{Z}_t + \mathbf{Z}_s\right)\hat{\mathbf{w}}\right)\right) E\left(p_t\left(\hat{\sigma}_e^2\right) p_s\left(\hat{\sigma}_e^2\right)\right) - \exp\left(\left(\mathbf{Z}_t + \mathbf{Z}_s\right)\boldsymbol{\beta} + \sigma_e\right). \end{aligned}$$

The second equality in (1.17) follows because (approximately) $\hat{\mathbf{w}}$ and $\hat{\sigma}_e^2$ are independent and the estimator $\hat{F}_t^*$ is unbiased. Cohn, Gilroy, and others (1992) derive explicit expressions for the expectation terms in the second equality.

The retransformation problem discussed above will be encountered again in section 1.6 where we discuss the prediction methodology for the SPARROW model. Many of the concepts developed here will reappear in that discussion, although a different method will be proposed for resolving the problem. The discussion of the standard error of the mean flux estimate given in equation (1.16) will also have relevance to SPARROW model estimation and prediction, as will be shown in sections 1.5.3.5 and 1.6.6.1.

1.3.1.3 Mean detrended flux (advanced)

There are two sources of trend in the standard water-quality model given by equation (1.3) and both must be evaluated in order to obtain estimates of mean detrended flux. The primary source is the time trend term in the model, represented above by the function $\mathbf{h}(t)$. The second source is streamflow, the current and lagged values of which affect water quality through the function $\mathbf{M}(\mathbf{Q}_t)$.

Let the base year for detrending be denoted $T_0 \equiv T_{t_0}$. The detrending of the time trend term in the water-quality model is straightforward; simply replace $\mathbf{h}(T_t)$ with the constant $\mathbf{h}(T_0)$ in the $\mathbf{Z}_t$ vector used to predict water quality in period *t*. Because the standard error of the mean flux also depends on $\mathbf{Z}_t$, use of the modified $\mathbf{Z}_t$ vector in the standard error calculation (equations (1.16) and (1.17)) fully accounts for the effect of detrending the time trend term on the uncertainty of mean flux.

The removal of trend due to streamflow is more complicated. One approach, which is consistent in large samples, is to replace the streamflow vector $\mathbf{Q}_t$ in the function $\mathbf{M}(\mathbf{Q}_t)$ with a detrended streamflow vector

$\mathbf{Q}_t^D$. This requires the estimation of a streamflow model that includes trend terms. Let the equation for the logarithm of streamflow in period *t* be given by

$$\tilde{q}_t = \mathbf{g}(T_t)\boldsymbol{\alpha}_T + \mathbf{x}_t\boldsymbol{\alpha}_x + u_t, \tag{1.18}$$

where $\mathbf{g}(T_t)$ is a K_g-element row vector of deterministic functions of time, excluding seasonal variations; $\mathbf{x}_t$ is an exogenous process (possibly deterministic as well), including an intercept; and u_t is a normally distributed, possibly serially correlated residual.

The efficient estimation of the streamflow model coefficients requires specification of the serial correlation structure of the residual. Generally, it is our experience that a high-order autoregressive process is sufficient to capture the serial correlation in the daily streamflow residuals. A Box-Ljung chi-square test based on the estimated autocorrelations can be used to determine if a sufficient order for the autoregressive process has been specified (see SAS, 1993). If the streamflow record contains data gaps, then the estimation method may use a Kalman filtering procedure (see Hamilton, 1994).

Let the matrix $\mathbf{G}_t$ represent a $p \times K_g$ matrix consisting of the current and lagged row vectors of $\mathbf{g}(T_t)$, and let $\mathbf{G}_0$ be the $\mathbf{G}_t$ matrix evaluated at the base period t_0. We have,

$$\mathbf{G}_t = \begin{bmatrix} \mathbf{g}(T_t) \\ \vdots \\ \mathbf{g}(T_{t-p+1}) \end{bmatrix}, \mathbf{G}_0 = \begin{bmatrix} \mathbf{g}(T_{t_0}) \\ \vdots \\ \mathbf{g}(T_{t_0-p+1}) \end{bmatrix}. \tag{1.19}$$

Using equation (1.1), given estimates of the coefficient vector $\hat{\boldsymbol{\alpha}}_T$, the detrended streamflow vector $\mathbf{Q}_t^D$ is given by

$$\mathbf{Q}_t^D = \mathbf{Q}_t + (\mathbf{G}_0 - \mathbf{G}_t)\hat{\boldsymbol{\alpha}}_T. \tag{1.20}$$

Let $\mathbf{Z}_t^D = \left\{\mathbf{M}(\mathbf{Q}_t^D), \mathbf{h}(T_0), \mathbf{X}_t\right\}$ represent the row vector of detrended water-quality explanatory variables. Detrended flux for period *t* is obtained simply by substituting the detrended explanatory variables $\mathbf{Z}_t^D$ for $\mathbf{Z}_t$ in equation (1.14), and the mean detrended flux is obtained by using the detrended period *t* fluxes to estimate equation (1.15).

The substitution of detrended streamflow for actual streamflow in the determination of mean detrended water-quality flux introduces two effects on the standard error of mean detrended flux. The first and most important effect is the change in the standard error calculation (equations (1.16) and (1.17)) caused by uncertainty in the $\hat{\boldsymbol{\beta}}_Q$ coefficients interacting with the change in $\mathbf{Z}_t$ caused by substituting the series of detrended streamflow vectors $\mathbf{Q}_t^D$ for actual streamflow vectors $\mathbf{Q}_t$. Because equation (1.17) is conditioned on $\mathbf{Z}_t$, the effect of this substitution is fully accounted for in the estimation of flux uncertainty.

A secondary effect on uncertainty in detrended water quality arises because the $\boldsymbol{\alpha}_T$ vector used to compute $\mathbf{Q}_t^D$ is estimated with uncertainty. There are two reasons why this effect is likely to be negligible relative to the first effect. First, the streamflow model is generally estimated with considerably more observations than the water-quality model, implying the estimates of $\hat{\boldsymbol{\alpha}}_T$ are quite precise relative to the estimates of $\hat{\boldsymbol{\beta}}_Q$. Second, if the streamflow equation consists of a simple linear time trend, the variance of the estimated trend coefficient is of order $O\left(N_q^{-2}\right)$, where N_q is the number of streamflow observations, as

compared to $O\left(N^{-1}\right)$ for coefficients associated with non-trending variables. For these reasons, we do not derive a more complete estimate of uncertainty for the mean detrended streamflow to account for sample error in the $\hat{\alpha}_T$ coefficients.

1.3.1.4 Tools for flux estimation

If the water-quality data do not include any censored observations, the water-quality models described above can be easily estimated using ordinary least squares. For cases in which the water-quality data include censored observations, Cohn, Gilroy, and others (1992) suggest estimation via adjusted maximum likelihood. In large samples, the method of maximum likelihood is consistent and efficient (Cohn, Gilroy, and others, 1992). The standard maximum likelihood method to apply to type I censored data (that is, data for which the censoring threshold is known) is the Tobit model, so named after its inventor, economist James Tobin (1958). The method of adjusted maximum likelihood combines the Tobit model with an adjustment to correct for first-order bias in the coefficient estimates caused by estimation using a small sample.

The method of adjusted maximum likelihood is implemented within the recently released USGS program LOADEST 2000 (Runkle and others, 2004). In addition to estimates of the parameters and their covariance matrix, the program uses the retransformation methods described above to produce unbiased estimates of daily and annual flux. A simple averaging of the daily or annual estimates over all days or years yields an estimate of long-term mean flux. Unfortunately, the program does not compute a standard error for the long-term mean estimate, and detrending the estimates requires additional processing of the daily results.

The estimation of the model used to detrend flow requires a maximum likelihood method capable of correcting for serial correlation in the errors. This capability is included in the PARMA model developed by Vecchia (2000), but can also be implemented using standard statistical packages such as SAS (SAS, 1993).

More recently, a program called Fluxmaster developed by G.E. Schwarz (principal author of this documentation) includes methods to estimate the time-series flow model using maximum likelihood, detrend flow, and estimate the water-quality model via adjusted maximum likelihood. The program also computes unbiased detrended estimates of long-term mean flux, and provides an estimate of the associated standard error. The exact methods used to implement adjusted maximum likelihood and correct for the retransformation bias in Fluxmaster differ from those used in LOADEST 2000, but they are a close approximation and exactly the same if there are no censored observations. The difference pertains to parameter estimation arising from the correction of maximum likelihood coefficient estimates for first-order bias (the uncorrected parameter estimates of Fluxmaster and LOADEST 2000 are the same aside from slight differences due to differences in numerical optimization methods). The correction for first-order bias depends on an estimate of the parameter covariance matrix. Fluxmaster estimates the covariance matrix using a numerical method to estimate the expectation of the numerically approximated Hessian matrix; conversely, LOADEST 2000 estimates the covariance matrix using analytic derivatives of the likelihood function for individual observations evaluated at sample data values.

The similarity between the LOADEST 2000 and Fluxmaster methods can be demonstrated via a Monte Carlo experiment. Figure 1.7 and table 1.2 present results from a Monte Carlo analysis consisting of 1,000 cases in which 365 days of streamflow and water-quality values are randomly generated. The logarithm of streamflow is generated from a standard normal distribution; the logarithm of water quality equals the logarithm of streamflow plus a standard normal random error. All values of the logarithm of water quality below zero are censored, resulting in a 50 percent censoring threshold. For each case, a sample consisting of 52 water-quality values, one value per week, was selected and used to estimate annual flux using the Fluxmaster and LOADEST 2000 flux estimation algorithms. Figure 1.7 shows close agreement between the two algorithms. As is evident from table 1.2, both methods exhibit virtually identical bias and precision.

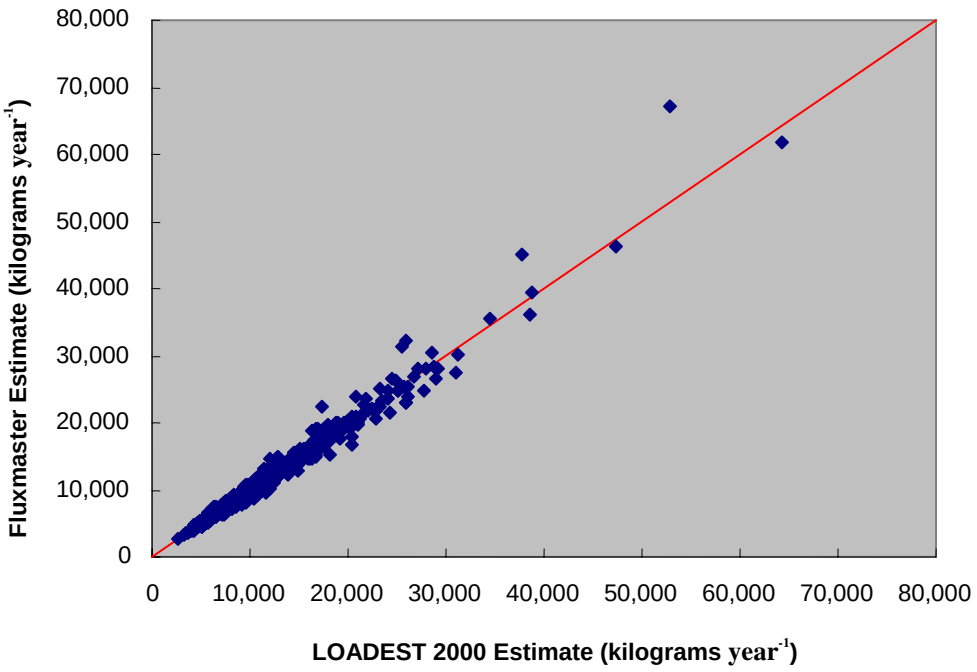

Figure 1.7. Monte Carlo analysis comparing flux estimates from the programs Fluxmaster and LOADEST 2000. The comparison consists of 1,000 replications of simulated streamflow and water-quality data, where the logarithm of streamflow is generated from a standard normal distribution and the logarithm of water quality equals the logarithm of streamflow plus a standard normal random variable, the sum of which is censored at zero (approximately 50 percent censoring).

Table 1.2. Monte Carlo evaluation of Fluxmaster and LOADEST 2000 annual flux estimates.

[Analysis consists of 1,000 repetitions, with 52 observations per case, 50 percent of which are censored; the logarithm of streamflow is generated from a standard normal distribution; the logarithm of water quality equals the logarithm of streamflow plus a standard normal random variable]

	Fluxmaster	LOADEST 2000
Bias (percent)	-4.1	-4.0
Standard Error (percent)	66.4	65.3
Ratio Standard Error to Mean Estimated Standard Error (percent)	120	114