Biogeochemistry **57/58:** 295–339, 2002.

A comparison of models for estimating the riverine export of nitrogen from large watersheds

RICHARD B. ALEXANDER[1,*], PENNY J. JOHNES[2], ELIZABETH W. BOYER[3] & RICHARD A. SMITH[1]
[1]*U.S. Geological Survey, 413 National Center, 12201 Sunrise Valley Drive, Reston, Virginia 20192, U.S.A.;* [2]*University of Reading, Aquatic Environments Research Centre, Department of Geography, Whiteknights Reading RG6 6AB, U.K.;* [3]*State University of New York, College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY 13210, U.S.A.*
*(*author for correspondence, e-mail: ralex@usgs.gov)*

Key words: error analysis, model validation, nitrogen, watershed models

Abstract. We evaluated the accuracy of six watershed models of nitrogen export in streams (kg km^2 yr^{-1}) developed for use in large watersheds and representing various empirical and quasi-empirical approaches described in the literature. These models differ in their methods of calibration and have varying levels of spatial resolution and process complexity, which potentially affect the accuracy (bias and precision) of the model predictions of nitrogen export and source contributions to export. Using stream monitoring data and detailed estimates of the natural and cultural sources of nitrogen for 16 watersheds in the northeastern United States (drainage sizes = 475 to 70,000 km^2), we assessed the accuracy of the model predictions of total nitrogen and nitrate-nitrogen export. The model validation included the use of an error modeling technique to identify biases caused by model deficiencies in quantifying nitrogen sources and biogeochemical processes affecting the transport of nitrogen in watersheds. Most models predicted stream nitrogen export to within 50% of the measured export in a majority of the watersheds. Prediction errors were negatively correlated with cultivated land area, indicating that the watershed models tended to over predict export in less agricultural and more forested watersheds and under predict in more agricultural basins. The magnitude of these biases differed appreciably among the models. Those models having more detailed descriptions of nitrogen sources, land and water attenuation of nitrogen, and water flow paths were found to have considerably lower bias and higher precision in their predictions of nitrogen export.

1. Introduction

Nitrogen inputs to terrestrial systems approximately doubled in the latter half of the 20th century, and riverine flux to coastal waters, where nitrogen is most limiting to primary production, increased by a similarly large factor (Vitousek et al. 1997). These increases have caused the eutrophication of many coastal and estuarine ecosystems worldwide, leading to chronic hypoxia, reductions in species abundance, and stressed fisheries resources. Although these problems are clearly related to cultural sources (Vitousek et al. 1997; Nixon et al. 1995), knowledge of the effects of specific nitrogen sources and watershed processes on riverine export is needed to better understand the likely consequences for coastal ecosystems of anticipated increases in N inputs related to population growth and economic development. The large increases that are expected to occur globally in multiple sources of nitrogen by early in the 21st century (i.e. a doubling of fertilizer and fossil-fuel N fixation; Galloway et al. 1994) underscore the need for improved understanding of nitrogen transport in large coastal watersheds. However, the increased complexity of sources and controlling processes in large watersheds potentially limits understanding at these scales. Nitrogen inputs to watersheds are removed at widely varying rates in streams and reservoirs and on the landscape through storage, denitrification, and interbasin transfers of agricultural products. Estimates of nitrogen removal in streams vary over two orders of magnitude (Seitzinger & Kroeze 1998; Alexander et al. 2000). The rates of nitrogen flux vary considerably in forested catchments (Johnson 1992), pastoral and agricultural catchments (Johnes 1996; Beaulac & Reckhow 1982), and in larger watersheds of mixed land use (Howarth et al. 1996; Caraco & Cole 1999; Seitzinger & Kroeze 1998). Knowledge of transport is also limited by the focus of most watershed studies on a relatively narrow range of environmental conditions in small watersheds.

Despite these difficulties, recent progress has been made in developing models of the mean-annual riverine yield or export (kg ha^{-1} yr^{-1}) from large watersheds. These models rely on relatively simple assumptions and descriptors of nitrogen sources and landscape characteristics. The models explain from 50 to 90 percent of the spatial variability in stream nitrogen export based on reported R^2 statistics. Nitrogen export varies by more than three orders of magnitude in major rivers of the world (Seitzinger & Kroeze 1998; Caraco & Cole 1999) and in individual countries such as the United States (Alexander et al. 2001). However, R^2, although frequently used as a measure of model performance, does not reliably describe the accuracy (bias and precision) of predictions of stream nitrogen export. R^2 is sensitive to various statistical properties of the explanatory and response variables (Montgomery & Peck 1982), making comparisons of model performance unreliable. The models also differ in their levels of spatial resolution and process complexity, which may affect the accuracy of model predictions of nitrogen export, processing rates, and source contributions to streams. Recent global (Seitzinger & Kroeze 1998; Caraco & Cole 1999) and regional (Howarth et al. 1996) models suggest that variations in export can largely be explained as a function of nutrient sources despite the large variability that occurs in nitrogen controlling processes in watersheds. By contrast, recent statistical (Smith et al. 1997; Preston & Brakebill 1999; Alexander et al. 2000, 2001) and export-coefficient models (Johnes 1996; Johnes & Heathwaite 1997) indicate that knowledge of spatial variations in watershed properties that influence nitrogen processing (e.g. surface water flow paths, soils, climate) can significantly improve the accuracy of estimates of stream export and source contributions over a broad range of watershed scales.

These export models rarely have been compared in systematic manner over a range of climatic conditions, sources, and watershed sizes. Such comparisons are needed to improve understanding of how model accuracy varies in different environmental settings and with the complexity of model descriptions of N sources and biogeochemical processes that affect nitrogen transport. This analysis provides an initial comparison of the performance of several prominent models for estimating riverine export of nitrogen. We begin by applying each of the selected models to 16 watersheds in the northeastern United States (see Figure 1). These watersheds have climatic conditions and nitrogen sources that lie within the range of watershed conditions used to calibrate the original models, and therefore, provide an appropriate collection of environmental settings for evaluating the models. The northeastern watersheds are the focus of SCOPE (Scientific Committee On Problems of the Environment) investigations of nitrogen cycling (e.g. see Boyer et al. 2002 and Van Breemen et al. 2002) for which detailed estimates of natural and cultural sources are available. This study is complementary to previous SCOPE investigations of N transport in watersheds (Howarth et al. 1996) and the SCOPE analyses in this volume that use certain of the models. The study also builds on previous studies that have evaluated a more limited number of models (Seitzinger & Kroeze 1998; Caraco & Cole 1999; Stacy et al. 2001; Alexander et al. 2001; National Research Council 2000).

The analysis is presented in eight sections. Following the introduction, section two summarizes the range of approaches that have been used to model nutrient export from large watersheds, noting their principle features and assumptions. Section three presents the methods and data used to evaluate the selected nutrient export models. We describe methods for quantifying the accuracy (bias, variability) of the model predictions, including the use of an error modeling technique to identify prediction biases caused by deficiencies in the export model descriptions of nitrogen sources and processes in watersheds. Section four presents the results of the error analysis. A discussion of the error analysis is given in section five. Model predictions of source contributions to stream export are compared in section six. Section seven

Figure 1. Location of 16 watersheds in the northeastern United States.

describes the model estimates of watershed attenuation of N, and the final section presents the summary and conclusions.

2. Background

A variety of deterministic, statistical, and hybrid methods have been used to model the transport of nitrogen in rivers basins. We provide a brief review of various models that have been applied to large watersheds, frequently several thousands of square kilometers in size or larger. We selected a subset of these models for use in this analysis (see section 3.1).

The simplest deterministic approaches (Howarth et al. 1996; Jaworski et al. 1992) are mass balance models that provide a static accounting of nitrogen inputs (e.g. fertilizer application, atmospheric deposition) and outputs (e.g. river export, crop N fixation and removal). Where major sources or sinks are difficult to measure independently (e.g. groundwater storage, denitrification), estimates are often obtained as a difference of measured terms. Recent refinements (Jordan & Weller 1996) have been made in budget methods to account for watershed imports and exports of nitrogen in food and feed. Additional sources and sinks have been quantified for the northeastern watersheds by several studies reported in this volume (e.g. Van Breemen et al. 2002; Boyer et al. 2002; Seitzinger et al. 2002) including N fixation and uptake in forests and N attenuation in streams and reservoirs. Selected results from these studies are used in this investigation. Where source inputs and sinks cannot be spatially referenced, budget terms are assumed to be uniformly distributed within watersheds with loss processes operating equally on all sources. In the absence of source-specific flux rates, estimates of the relative contributions of sources to surface waters must also be assumed to be proportional to the nitrogen inputs to watersheds. In extending these methods to large watersheds, uncertainties may exist over the rates of N supply and loss, which are based on the extension of literature estimates from experimental studies.

More complex deterministic models of nitrogen flux (e.g. HSPF, Bicknell et al. 1997; SWAT, Srinivasan et al. 1993; INCA, Whitehead et al. 1998; AGNPS, Young et al. 1995) describe transport and loss processes in detail by simulating nitrogen availability, transport, and attenuation processes according to mechanistic functions that include descriptions of the spatial and temporal variations in sources and sinks in watersheds. Simulation models of landscape processing of nitrogen, such as the biochemical dynamics of nitrogen in soils (e.g. CENTURY, Parton et al. 1988), have also been developed. The complexity of deterministic models often creates intensive data and calibration requirements, which generally limits their application in large watersheds; these models have been more commonly applied in small catchments. One deterministic surface-water nutrient model (SWAT, Srinivasan et al. 1993) has been recently applied in the watersheds of major regions of the United States (see Alexander et al. 2001), although this agricultural model does not account for all sources of nitrogen (e.g. atmosphere). In general, there are uncertainties involved in aggregating the components of fine-scale deterministic models in watershed applications (Rastetter et al. 1992) and in extrapolating the results of catchment models and field-scale measurements to larger spatial scales. Deterministic methods also often lack robust measures of uncertainty in model coefficients and predictions, which might otherwise be used to assess their accuracy in modeling N transport.

One common approach to estimating stream export (e.g. Delwiche & Haith 1983; Fisher & Oppenheimer 1991) has been to apply the reported yields (mass of N per unit drainage area) from small, homogeneous watersheds to the variety of land types contained within larger heterogeneous basins ('export coefficient' method). Because the nitrogen yields for given land types are highly variable (Beaulac & Reckhow 1982; Frink 1991; Johnson 1992; Johnes 1996), reflecting variations in climatic conditions, nutrient sources, and terrestrial and aquatic loss processes, these methods can produce imprecise and potentially biased estimates of export when extrapolated to other areas and larger scales (Beaulac & Reckhow 1982). Refinement of the export coefficient method in the United Kingdom has produced more robust models capable of accurately simulating the nutrient export response to temporal changes in nutrient source inputs and land and waste management practices at the watershed and regional scales (Johnes 1996; Johnes et al. 1996; Johnes & Heathwaite 1997). The model calculates the total N and P load delivered to a waterbody separately by source type as a function of the rates of nutrient input and the export potential of a watershed. The export potential is estimated according to the location of sources in watersheds and the landscape and climatic conditions. The export coefficients are expressed as a percentage of nutrient inputs (rather than mass per unit area as in conventional export methods), allowing the simulation of the effect of historical land-use changes and management in watersheds. One limitation of this model is that detailed information is required on the export potential of landscapes and the types and location of nutrient sources. Nevertheless, in the U.K., where source and monitoring information are available from the 1860s to date, the model has been extensively validated and applied nationally to all watersheds in England and Wales. The model accurately allows the scaling up of plot-scale experimental measurements to the watershed and regional scales, explaining up to 98% of the spatial variations in stream export.

Other watershed models have been developed that represent a mixture of deterministic and export-coefficient approaches. A recent model of watershed export combined the deterministic budget approach with literature-based export coefficients for different land types and source inputs (Castro et al. 2001). This model produced mixed results in applications to the drainages of 34 major estuaries of the United States (Castro et al. 2001; Stacy et al. 2001) with the method tending to overestimate the riverine export from agricultural watersheds. Other types of models described as 'loading functions' (GWLF; Haith & Shoemaker 1987) represent a compromise between the export coefficient method and the complexity offered by simulation models. These models have mechanistic water and sediment transport components (Howarth et al. 1991) with nutrient dynamics often described by simple empirical relations

(Haith & Shoemaker 1987). Model parameters may be obtained from the literature or statistically estimated if sufficient data are available. Applications have been made to eastern U.S. watersheds as large as several thousands of square kilometers (e.g. Lee et al. 2000; Howarth et al. 1991).

Statistical approaches to modeling nitrogen flux in coastal basins have their origins in simple correlations of stream nitrogen measurements with watershed sources and landscape properties. These methods assume limited *a priori* knowledge of biogeochemical processes, but provide empirical estimates of the aggregate supply and loss of nitrogen through the use of conventional stream monitoring data, which are often readily available in large watersheds with mixed land use. Anthropogenic N sources constitute the principle predictor variables in these models. Some of the models (e.g. Howarth et al. 1996) make use of literature rates of N processing (crop N fixation, N removal in crops) to estimate agricultural source inputs. Recent examples include regressions of nitrogen export from large watersheds on population density (Peierls et al. 1991), net anthropogenic sources (Howarth et al. 1996), atmospheric deposition (Jaworski et al. 1997; Howarth et al. 1996), and measures of per capita energy consumption by humans (Meybeck 1982). These models explain up to 80 percent or more of the spatial variations in nitrogen export. In contrast to complex deterministic models, these statistical methods have the advantage of being readily applied in large watersheds. Moreover, statistical approaches are capable of quantifying errors in model parameters and predictions. These simple correlative models are limited, however, in that they consider sources and sinks to be homogeneously distributed in space, do not separate terrestrial from in-stream loss processes, and rarely account for nonlinear interactions between sources and loss processes.

A recently developed hybrid approach (SPARROW; SPAtially Referenced Regression On Watershed attributes; Smith et al. 1997) expands on conventional regression methods by using a mechanistic model structure in correlating measured nitrogen flux in streams with spatial data on nitrogen sources, landscape characteristics (e.g. soil permeability, temperature), and stream properties (e.g. streamflow, water time of travel). The model, which separately estimates the quantities of nitrogen delivered to streams and the outlets of watersheds from point and diffuse sources, has been applied nationally in the United States (Smith et al. 1997) with separate studies of nitrogen flux in the Chesapeake Bay watershed (Preston & Brakebill 1999), the Mississippi River and its tributaries (Alexander et al. 2000), the watersheds of major U.S. estuaries (Alexander et al. 2001), and watersheds of New Zealand (McBride et al. 2000). By spatially referencing nitrogen sources and watershed attributes to surface water flow paths, defined according to a digital drainage network, and imposing a mass-balance constraints, the model has been shown to improve the accuracy of predictions of stream export and the interpretability of model coefficients (Smith et al. 1997; Alexander et al. 2000, 2001). Model estimates of in-stream nitrogen loss and stream nutrient export from watersheds of various land-use types are generally consistent with literature rates (Alexander et al. 2000). By comparison to the simple correlative approaches, this method requires more spatially detailed data on watershed characteristics, such as river drainage attributes (e.g. surface-water flow paths) and point and diffuse nitrogen sources.

Quasi-empirical models (Seitzinger & Kroeze 1998; Caraco & Cole 1999) of nitrate-nitrogen export from the largest rivers of the world were recently developed using empirical regression methods and literature-based rate coefficients. These models were developed for estimating N budgets at the continental scale and evaluating the effects of cultural sources on N export in some of the largest river basins of the world. These models indicate that the large variations in nitrate export among rivers worldwide can be largely explained by several major nitrogen sources and relatively simple descriptors of nitrogen removal on the landscape and in rivers. Nitrate export is modeled as a function of point sources (i.e. urban population and an assumed per capita discharge rate of 1.85 kg-N year^{-1}), the diffuse inputs of fertilizer and atmospheric deposition, statistically calibrated runoff (discharge per unit drainage area) coefficients, and a literature-based in-stream loss coefficient of 30%. Model predictions of export were highly correlated (r-squared > 80%) with measured export for 35 of the largest rivers of the world. Some of the principle outliers in the models were associated with watersheds containing reservoirs (Caraco & Cole 1999), which were not explicitly included in the models. The stronger correlations of nitrate with population density than observed for total dissolved nitrogen has been suggested as an indication that nitrate measurements more readily display the effects of anthropogenic activities on river export in these rivers (Caraco & Cole 1999). However, the accuracy of these global models is not well known for smaller watersheds and for rivers with higher ammonium and organic nitrogen loads.

3. Methods

3.1. *Export models*

We compared the performance of six empirical and quasi-empirical nitrogen watershed models (see Table 1) in 16 watersheds of the northeastern United States (see Figure 1). Watershed models classified as strictly deterministic (e.g. SWAT, HSPF) were not evaluated in this analysis (see comparisons in Alexander et al. 2001). Table 1 presents details of the model equations, the types of data required by each model, and characteristics of the calibration data. The data used in applying the SPARROW model are given in Alexander et al. (2000, 2001). Two of the models (SPARROW, HOWARTH) predict mean-annual total nitrogen export in streams. The remaining models predict mean-annual nitrate-N export; method performance is only evaluated for nitrate-N for these models. Two previously unpublished statistical models were estimated by applying ordinary least squares (LS) to data from the largest rivers of the world separately for 29 rivers (LS1-GLOBAL; Seitzinger & Kroeze 1998) and for 35 rivers (LS2-GLOBAL; Caraco & Cole 1999). These models offer several potential advantages. First, rather than assuming that the diffuse sources (i.e. fertilizer and atmospheric deposition) are supplied and transported at identical rates as in the quasi-empirical global models based on these data (Seitzinger & Kroeze 1998; Caraco & Cole 1999), the statistical models estimate separate rate coefficients for the diffuse sources, allowing any differences in the rates of supply and processing to be empirically determined. Second, the models allow the uncertainty of the model coefficients to be empirically determined and used to evaluate model fit. Finally, an intercept term can be evaluated in the models to determine whether any of the remaining variability in nitrogen export is potentially explained by a constant source. This source may potentially represent natural or background sources of nitrogen that are unexplained by the major cultural nitrogen sources explicitly specified by the models. Results for these two statistical models are described in section 4.1.

3.2. *Watershed characteristics*

Selected characteristics for the northeastern watersheds are presented in Table 2, including estimates of total nitrogen and nitrate-N stream export, drainage area, runoff (discharge per unit of drainage area), and land use. The data are compiled for selected years during the 1988 to 1993 time period (see Boyer et al. 2002, for details). A mixture of land use is represented in the drainages, but the watersheds are predominantly forested (range of 48 to 87%), with small to moderate amounts of cultivated land area (range from 2 to 40%). Developed land ranges from less than one percent to about 20% of the basin drainage area, but is commonly less than about 3%. Wetlands typically cover less than about 3% of the basin area. Population density ranges over nearly two orders of magnitude (Boyer et al. 2002). The drainage areas of the watersheds span more than two orders of magnitude from 475 to 70,000 km^2. Stream nitrogen export and runoff of the watersheds typically span about a factor of two to five. Nitrate nitrogen generally represents less than half of the total nitrogen in streams in most of the watersheds (median = 43%; interquartile range = 30 to 53%) although nitrate is the predominant fraction

Table 1. Stream nitrogen export models of total nitrogen (TN) and nitrate (NO_3)

Model Name	Reference	Export (kg km^2 yr^{-1})	Calibration Data Set	Export Equation	R^2
SPARROW	Alexander et al. (2000, 2001); Smith et al. (1997)	TN	374 sites in conterminous United States	$TN_i = \{\sum_{n=1}^{N} \sum_{j \in J(i)} S_{n,j} \beta_n \exp(-\alpha' Z_j)\exp(-k' T_{i,j})\}\varepsilon_i A_i^{-1}$ j is the reach belonging to the set of reaches (J(i)) located upstream of the downstream monitoring site in reach i, for which the export prediction is made $S_{n,j}$ = mass from source n (belonging to a total of N sources) input to the drainage of reach j β_n = source-specific coefficient (fertilizer use, livestock wastes, wet atmospheric deposition, nonagricultural land (diffuse runoff), industrial-municipal point sources)[a] $\exp(-\alpha' Z_j)$ = exponential function[b] giving the proportion of available nitrogen mass delivered to reach j as a function of land-to-water loss coefficients (defined by vector α) and associated landscepe characteristics (soil permeability-cm hr^{-1}; drainage density-km^{-1}; temperature-^{0}F.), Z_j, in the drainage to reach j $\exp(-k' T_{i,j})$ is the proportion of nitrogen mass in reach j transported to downstream reach i as a function of a first-order rate of nitrogen loss (k'), defined according to a vector of four discrete classes of channel size in units of reciprocal water travel time ($T_{i,j}$; days) ε is a multiplicative error term A_i is the drainage area of the basin (km^2)	0.88
HOWARTH	Howarth et al. (1996)	TN	9 regions (Canada, U.K., western Europe, eastern U.S.)	TN = –120 + 0.79 NO_y + 0.11 NAI NO_y = atmospheric deposition (wet and dry oxidized forms – NO_3, HNO_3; kg km^{-2} yr^{-1}) NAI = net anthropogenic inputs (fertilizer + crop N fixation + food/feed imports – food exports; kg km^{-2} yr^{-1})	0.89

PEIERLS	Peierls et al. (1991)	NO_3	34 global large rivers	$NO_3 = 10^{1.51(0.64\log(PD))}$ PD = population density (people km^{-2})	0.51
GLOBAL	Seitzinger & Kroeze, (1998); Caraco & Cole, (1999)	NO_3	35 largest rivers in the world (only runoff coefficients calibrated)	$NO_3 = R_{export}\ (PI + WS_{export}\ (F + NO_y))$ R_{export} = 0.7; the average in-stream transport based on literature studies (fraction of N inputs to streams) PI = point source inputs (kg km^{-2} yr^{-1}); product of population density (people km^{-2}), fraction of population in urban areas, and a per capita N release of 1.85 kg N yr^{-1} WS_{export} = N loaded to streams (kg km^{-2} yr^{-1}); modeled as a function of runoff (R; discharge per unit drainage area; m yr^{-1}), where $WS_{export} = 0.4\ R^{0.8}$ F = fertilizer use (kg km^{-2} yr^{-1})	0.84–0.89[c]
LS1-GLOBAL	Model fit with data from Caraco & Cole (1999)	NO_3	35 largest rivers in the world	$NO_3 = 2.81\ PI' + \exp(-0.2785/R)\ (0.0816\ F + 0.2969\ NO_y)$ PI' = population density (people km^{-2}) in urban areas (fraction)	0.90
LS2-GLOBAL	Model fit with data from Seitzinger & Kroeze (1998)	NO_3	29 largest rivers in the world	$NO_3 = 0.34\ PI' + \exp(-0.2960/R)\ (0.2316\ F + 0.3762\ NO_y)$	0.83

[a]All units are kg yr^{-1} except nonagricultural land area (ha). The land-to-water delivery function is equal to one for point-source inputs. Atmospheric deposition contributions to stream export are based on wet-fall deposition; land-to-water delivery fractions exceed unity indicating that additional atmospheric forms of nitrogen (e.g. ammonium, organic) are included (see Alexander et al. 2001).

[b]The product of the land-to-water delivery function (and its associated coefficients) and the nonpoint-source coefficients quantifies the fraction of the diffuse source inputs delivered to rivers.

[c]R^2 from correlation of predicted stream nitrogen export with measured stream export; only the runoff coefficients are statistically calibrated in the model.

in many of the more developed watersheds. The lowest nitrate contributions to total stream N export are found in the northern watersheds, where nitrate represents less than a third of total export. In the largest rivers of the world (Caraco & Cole 1999), the proportions of organic N and nitrate have been found to be roughly equivalent. Inputs of nitrogen sources, including fertilizer, atmospheric deposition (NO_y, the total wet and dry oxidized components of deposition – NO_3, HNO_3), and net anthropogenic sources, typically vary over a factor of two to three based on the interquartile range of the distribution, with the most extreme inputs of fertilizer and net anthropogenic nitrogen differing by a factor of about 10 (Boyer et al. 2002). NO_y-N is estimated to be about 65% of the total deposition in many of these watersheds (Boyer et al. 2002).

The cultural sources and climatic conditions of the northeastern watersheds lie within the range of conditions used to calibrate the original stream export models, and thus, the watersheds provide an appropriate set of locations for evaluating the models. Figure 2 compares conditions in the northeastern watersheds (i.e. NE US) with those in the calibration watersheds for selected explanatory variables of the models (i.e. runoff, population density, fertilizer use) and for stream nitrogen export (total nitrogen, nitrate-nitrogen). The range of the original data used to calibrate the various stream export models is inclusive of the full range of characteristics of the northeastern watersheds. Appreciable similarities exist in the watersheds given the considerable overlap in the interquartile ranges of many of the basin conditions (Figure 2). The drainage sizes of the northeastern watersheds are generally at the lower end of the range of watershed sizes used in the calibrations of most of the stream export models. Models that were applied to the largest rivers of the world (Peierls et al. 1991; Seitzinger & Kroeze 1998; Caraco & Cole 1999) used watersheds frequently larger than 0.2 million km^2; however, the calibration data include smaller river basins, such as the Susquehanna, Delaware, and Hudson, which are included in the set of 16 northeastern watersheds. The SPARROW model calibration used 374 U.S. watersheds ranging in size from about 10 to 2.9 million km^2 with an interquartile range of about 3,000 to 37,000 km^2 (approximately 10% of the watersheds are located in the northeastern United States, and include nine of the 16 northeastern watersheds). The HOWARTH model was calibrated for nine large regional watersheds in northern Europe and the eastern half of the United States and Canada, ranging in size from 0.3 to 3.2 million km^2.

3.3. *Error analysis*

We evaluated the performance of the models through an analysis of the errors in the predictions of stream nitrogen export. Prediction errors ($E_{i,k}$)

Table 2. Nitrogen export and watershed characteristics for the sites in the northeastern United States [from Boyer et al. 2002; TN = total nitrogen]

River Name	Drainage Area (km^2)	TN Export[a] (1988–1993) ($kg\ km^2\ yr^{-1}$)	NO_3 Export[a] (1988–1993) ($kg\ km^2\ yr^{-1}$)	Ratio NO_3/TN Export	Runoff ($m\ yr^{-1}$)	Developed Land (%)	Cultivated Land (%)	Forested Land (%)	Other Land[b] (%)
Penobscot	20,109	317	66	0.21	0.59	0.4	1.5	83.8	14.4
Kennebec	13,994	333	87	0.26	0.57	0.9	5.9	79.6	13.6
Androscoggin	8,451	404	112	0.22	0.64	1.1	4.8	84.6	9.5
Saco	3,349	389	81	0.21	0.67	0.8	3.6	87.4	8.2
Merrimack	12,005	499	155	0.31	0.59	8.7	7.7	74.7	8.9
Charles	475	644	335	0.53	0.58	22.2	8.3	59.3	10.2
Blackstone	1,077	1,140	496	0.43	0.65	17.6	8.0	63.3	11.1
Connecticut	25,019	538	233	0.43	0.64	4.0	9.0	79.0	8.0
Hudson	11,942	428	222	0.53	0.62	2.7	10.4	80.8	6.0
Mohawk	8,935	826	427	0.53	0.55	4.7	28.0	63.1	4.2
Delaware	17,560	961	620	0.63	0.55	3.3	16.7	74.7	4.4
Schuylkill	4,903	1,755	1,419	0.83	0.49	10.2	38.4	48.1	3.3
Susquehanna	70,189	977	742	0.77	0.49	2.4	28.5	66.7	2.4
Potomac	29,940	897	392	0.43	0.33	2.6	34.6	60.8	2.0
Rappahannock	4,134	470	231	0.50	0.36	1.4	35.9	61.3	1.3
James	16,206	314	107	0.35	0.41	1.4	15.6	80.6	2.4
25th percentile	*4,711*	*400*	*111*	*0.30*	*0.49*	*1.3*	*7.3*	*61.2*	*3.0*
Median	*11,974*	*518*	*232*	*0.43*	*0.57*	*2.7*	*9.7*	*74.7*	*7.0*
75th percentile	*18,197*	*913*	*445*	*0.53*	*0.63*	*5.7*	*28.1*	*80.7*	*10.0*

[a] Mean-annual export.
[b] Includes open water, barren land, shrub land, and wetlands.

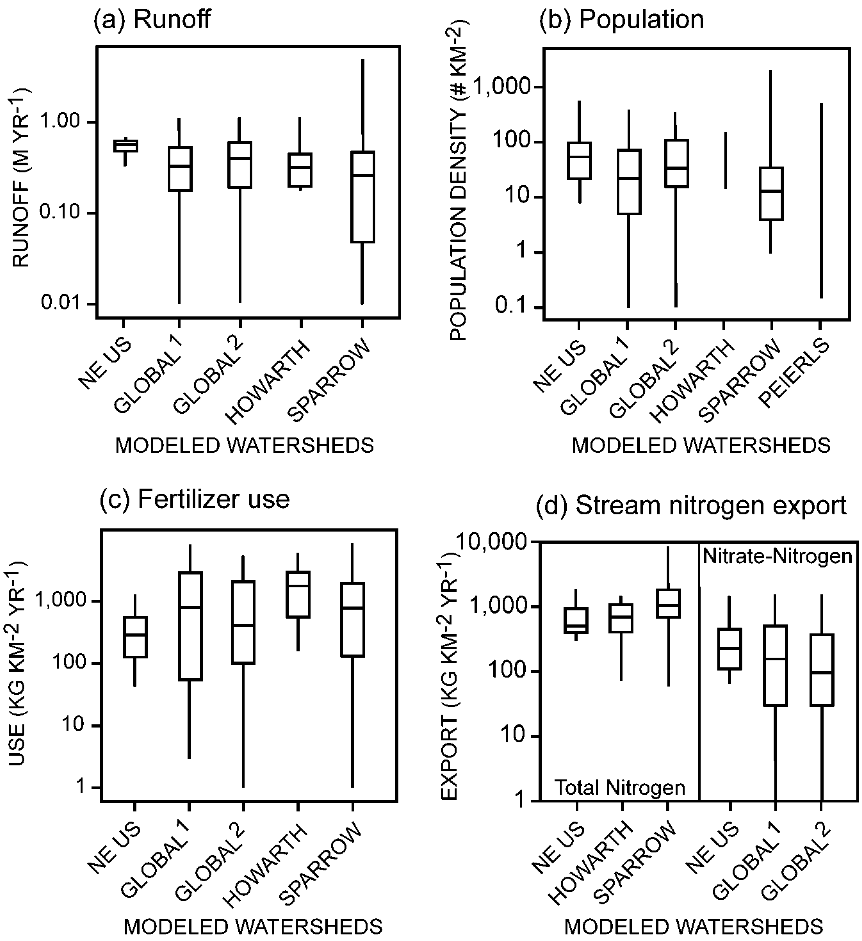

Figure 2. Box and whisker plots of the characteristics of the northeastern U.S. watersheds (NE US) in comparison to those of watersheds used to calibrate the original stream nitrogen export models: (a) runoff (discharge per unit drainage area), (b) population density, (c) fertilizer use, and (d) nitrogen export. Each box graphs the quartiles with the lower and upper edges representing the 25th and 75th percentiles, respectively. The midline plots the median. The upper and lower whiskers are drawn to the minimum and maximum values. The vertical lines in (b) give the range of the data. The models $GLOBAL^1$ (Caraco & Cole 1999) and $GLOBAL^2$ (Seitzinger & Kroeze 1998) were developed from data for the largest rivers of the world. The models are described in Table 1. Minimum values of zero are shown as 0.1 in (b) for the two global models and as 1.0 in (c) for the SPARROW and $GLOBAL^2$ models.

are computed for the ith watershed and kth model as the difference between the predicted ($P_{i,k}$) and 'measured' ($M_{i,k}$) stream nitrogen export, expressed as a percentage of the measured export, according to

$$E_{i,k} = \left[\frac{P_{i,k} - M_{i,k}}{M_{i,k}}\right] \cdot 100 \tag{1}$$

Errors with a positive sign indicate an over prediction of the measured export, whereas a negative sign indicates an under prediction of the measured stream export. The measured stream nitrogen export is based on the use of a rating curve technique (see Boyer et al. 2002; Cohn et al. 1989) to estimate the mean-annual nitrogen export for the period 1988 to 1993. The technique correlates instantaneous measurements of nitrogen flux (product of concentration and stream flow) with corresponding daily values of streamflow. By integrating the relation over the full record of daily flow, this method provides more precise estimates of stream export than can be obtained from a simple averaging of the instantaneous nitrogen export measurements. The standard error of the mean- annual export estimates for the 16 sites in this study was typically 5 to 10% of the mean. This magnitude of error is considered in forming conclusions about the comparative accuracy of the stream export models in Table 1.

The error analysis consists of two evaluations of the distribution of prediction errors for the 16 watersheds and six export models. First, we quantified the bias and variability of the model predictions of stream export according to robust statistical measures. The bias for each of the models was defined as the median of the prediction errors for the 16 watersheds from equation (1). The symmetry of the 25th and 75th percentile values of error about zero and the cumulative frequencies of the errors were also used as approximate indicators of the presence of bias in the prediction errors. The variability of the prediction errors was defined for each of the models as the interquartile range (IQR; i.e. difference between the 75th and 25th percentiles) of the prediction errors from equation (1). The model precision was defined as the reciprocal of the interquartile range. For selected models for which the original calibration data were available, we also quantified the bias and variability of the models for the original calibration set of watersheds, and compared these estimates with those observed for the 16 northeastern watersheds (see section 4.1).

Second, using multiple regression, we correlated errors in the predictions of stream nitrogen export with watershed characteristics that are associated with nitrogen sources and reflect terrestrial and aquatic processes affecting N mobilization and transport. The regression relation identifies 'factor-related' prediction biases that indicate potential deficiencies in the model descriptions of the supply and processing of nitrogen in the northeastern watersheds. Such deficiencies in model specification can be caused by watershed properties (sources or biogeochemical processes) that are not explicitly included in the export models or by model coefficients that poorly describe how N is supplied, mobilized, or transported in the northeastern watersheds. Either case is of interest and serves to highlight potential deficiencies in model performance. We fit separate regression models to the prediction errors associated with each of the six nitrogen export models by regressing the errors ($E_{i,k}$) from equation (1) on four characteristics of the watersheds that are measured with relatively high accuracy, according to the relation

$$E_{i,k} = \beta_{0,k} + \beta_{1,k} C_i + \beta_{2,k} D_i + \beta_{3,k} R_i + \beta_{4,k} BA_i + \varepsilon_{i,k} \tag{2}$$

where C_i is the cultivated land area (percent of total basin area) for the ith watershed, D_i is the developed (i.e. urban) land area (percent of total basin area), Ri is runoff (discharge per unit of drainage area; cm yr^{-1}), BA_i is the basin drainage area (km^2), $\beta_{o,k} \ldots, \beta_{4,k}$ are the regression parameters, and $\varepsilon_{i,k}$ describes the regression model error, assumed to be independent across watersheds. Cultivated and developed land area serve as indicators for a variety of nitrogen sources (e.g. fertilizer application, N fixation, urban washoff, and wastewater discharges) and landscape processes (e.g. N storage in forest biomass, crop harvesting, impervious cover) associated with these land uses. Evidence of a correlation between forestland and related processes in the error models would be expressed as a negative correlation with the cultivated and developed land area. Runoff and drainage basin area are potentially related to various biogeochemical processes that influence the mobilization and transport of nitrogen in terrestrial and aquatic environments. These may include biochemical reactions involving denitrification that are affected by water travel times through the subsurface and in channels. Runoff may reflect the effects on nitrogen export of such factors as climate (i.e. precipitation, evaporation), soil texture and moisture content, subsurface flow paths, stream properties (e.g. channel density and morphology, water velocity and flow), and water storage (e.g. lakes, reservoirs, groundwater). Climate-related effects on nitrogen flux may also reflect differences in vegetation that affect the natural supply of nitrogen to watersheds, although the land-use factors would be expected to reflect most major N sources. Drainage basin area provides a measure of the effects of spatial scale on nitrogen transport in watersheds, including scale-dependent properties of the rates of water flow, storage, and loss (e.g. length of surface and subsurface flow paths, water velocity, evaporation).

4. Results of the error analysis

We present the error analysis in three subsections. In section 4.1, we report the accuracy (bias, variability) of five of the nitrogen export models based exclusively on the original calibration data sets. Section 4.2 describes the bias and variability of the predictions of N export based on the application of the stream export models to the 16 northeastern watersheds and the calculation of prediction errors in equation (1). Section 4.3 gives the results of the regression-based models of the prediction errors (equation 2) associated with the application of each stream export model to the northeastern watersheds.

4.1. *Accuracy of the original model calibrations*

Table 3 gives the model parameters for the global models, LS1-GLOBAL, LS2-GLOBAL, based on a statistical fit to the two global data sets (Table 1). The overall fit of the models as measured by the R^2 values is similar or slightly improved in comparison to that of the original global models (Table 1). The model coefficients display weak (p = 0.13 to 0.19) to high ($p < 0.03$) levels of statistical significance indicating a range of uncertainty in the coefficient estimates. Point sources and atmospheric deposition are highly significant ($p < 0.03$) in the LS1-GLOBAL model, whereas fertilizer and atmospheric deposition sources are highly significant ($p < 0.02$) in the LS2-GLOBAL model. The mean coefficient values for runoff and atmospheric deposition are similar for the two models. By contrast, coefficient values for point sources and fertilizer differ greatly for the two models. Both models show consistently larger coefficient values for atmospheric deposition than for fertilizer suggesting a greater delivery of atmospheric N than fertilizer-related sources in the watersheds. According to the models, atmospheric inputs are delivered to watershed outlets from 1.5 (LS2-GLOBAL) to 3.5 (LS1-GLOBAL) times the rate of fertilizer-related source inputs. A previous export model for North Atlantic watersheds (Howarth et al. 1996; Howarth 1998; NRC 2000) also found higher rates of delivery of atmospheric sources relative to other anthropogenic sources. Insignificant intercept terms ($p > 0.6$) were estimated for both models in preliminary regressions suggesting that virtually all of the nitrogen sources are accounted for by the three sources (i.e. point, fertilizer, and atmospheric) specified in the models. Additional sources of nitrogen that are accounted for by an intercept term represent apparently negligible quantities of nitrogen (e.g. <8% for the LS1-GLOBAL model). An intercept was not included in the final models because the standard errors of some of the source coefficients were inflated by its inclusion.

The prediction errors, based on the original calibration data sets, are presented separately for five of the export models in Table 4. The original

Table 3. Regression model parameters for the global data sets. The units of the regression coefficients are: fertilizer use (dimensionless), atmospheric deposition (dimensionless), runoff (m yr^{-1}), and point sources (kg yr^{-1})

Model	R^2	No. Sites	Model Coefficients (*p – statistical significance*)			
			Point Sources	Runoff	Fertilizer Use	Atmospheric Deposition[a]
LS1-GLOBAL[b]	0.90	35	2.805	0.2785	0.0816	0.2969
			(*<0.01*)	(*0.13*)	(*0.17*)	(*0.03*)
LS2-GLOBAL[c]	0.83	29	0.3442	0.2960	0.2316	0.3762
			(*<0.19*)	(*0.01*)	(*0.02*)	(*0.02*)

[a]NO_y total wet and dry oxidized components of atmospheric deposition (Caraco & Cole 1999; Seitzinger & Kroeze 1998).
[b]Model calibrated with data from Caraco & Cole 1999.
[c]Model calibrated with data from Seitzinger & Kroeze 1998.

data for large rivers of the world used to calibrate the PEIRLS model were not available. Several of the models have relatively unbiased predictions of export, and include the HOWARTH, SPARROW, and LS2-GLOBAL models. Median errors for these models are less than 5% and the interquartile range of each is relatively symmetrical, typically ranging from about –30% to 40%. Median errors in the other export models are larger by a factor of six or more, and indicate that the models tend to over predict export at most of their calibration sites. A tendency for over prediction is noted for the Seitzinger and Kroeze (1998; median = 46%; IQR = –1 to 66%) and the Caraco and Cole (1999; median = 18%; IQR = –25 to 82%) global models. Although the LS1-GLOBAL model has a very small median error (–0.6), the interquartile range is unbalanced (–34 to 61%) suggesting a tendency for over prediction at many of the sites. The lowest variability in prediction errors based on the calibration data is found for the HOWARTH (IQR = 27%), SPARROW (IQR = 65%), and GLOBAL (IQR = 66% based on the Seitzinger and Kroeze data) models. The variability of the global model prediction errors, based on the Caraco and Cole (1999) riverine data set of 35 sites, tends to be higher than that of the other global models, based on the Seitzinger and Kroeze (1998) data set.

4.2. *Accuracy of the nitrogen export models – northeastern US watersheds*

The errors in model predictions of stream nitrogen export are shown separately for each method in Figure 3 with summary statistics for the prediction errors listed in Table 5. Figure 3(a) shows the cumulative frequencies of the

Table 4. Errors in model predictions of stream nitrogen export for the original calibration data sets. Errors are computed as the difference between the predicted and measured values of stream nitrogen export expressed as a percentage of the measured export

Model	median	IQR[a]	Prediction Errors (%)			
			min.	25th percentile	75th percentile	max.
SPARROW	3.2	65.2	–95.2	–27.4	37.8	1210.6
HOWARTH	2.8	26.6	–29.3	–6.0	20.7	66.3
GLOBAL[b]	18.0	107.6	–77.3	–25.6	82.0	1204.8
LS1-GLOBAL	–0.6	94.8	–83.6	–34.0	60.8	1960.7
GLOBAL[c]	46.5	66.6	–83.6	–1.1	65.5	861.5
LS2-GLOBAL	–3.3	71.6	–80.6	–30.3	41.3	469.6

[a]Interquartile range (difference between the 25th and 75th percentiles of the distribution of errors).
[b]Caraco & Cole 1999; only runoff coefficients were calibrated in the original models.
[c]Seitzinger & Kroeze 1998; only runoff coefficients were calibrated in the original models.

prediction errors for each model. The prediction errors for the 16 watersheds were ranked in ascending order for each model and assigned a cumulative frequency value according to a Cunnane plotting position quantile (Cunnane 1978). Better performing models (i.e. those with smaller prediction errors) are associated with values of the percent error that plot closest to zero (i.e. the horizontal line in Figure 3(a)) and display a relatively shallow slope; the median error for a model corresponds to a cumulative frequency of 0.5. The box and whisker plots in Figure 3(b) display summary statistics (Table 5) for the prediction errors, including the median, interquartile range (difference between the 25th and 75th percentiles), and minimum and maximum values. The length of the box and the whiskers (i.e. interquartile range, and data range) provides information about the variability of the model predictions, whereas the location of the median and symmetry of the interquartile range relative to zero describe the magnitude of bias in model predictions for the watersheds.

Based on comparisons of the median and symmetry of the interquartile range of the model errors (Figure 3(b), Table 5) and their cumulative frequencies (Figure 3(a)), all of the methods show at least small amounts of bias. The median errors are within about 15% for five of the six models. The GLOBAL, SPARROW and LS1-GLOBAL models have the smallest median errors of less than about 5%. Prediction errors of this magnitude are within the standard error of the measured values of export (about 5 to 10%), and are therefore

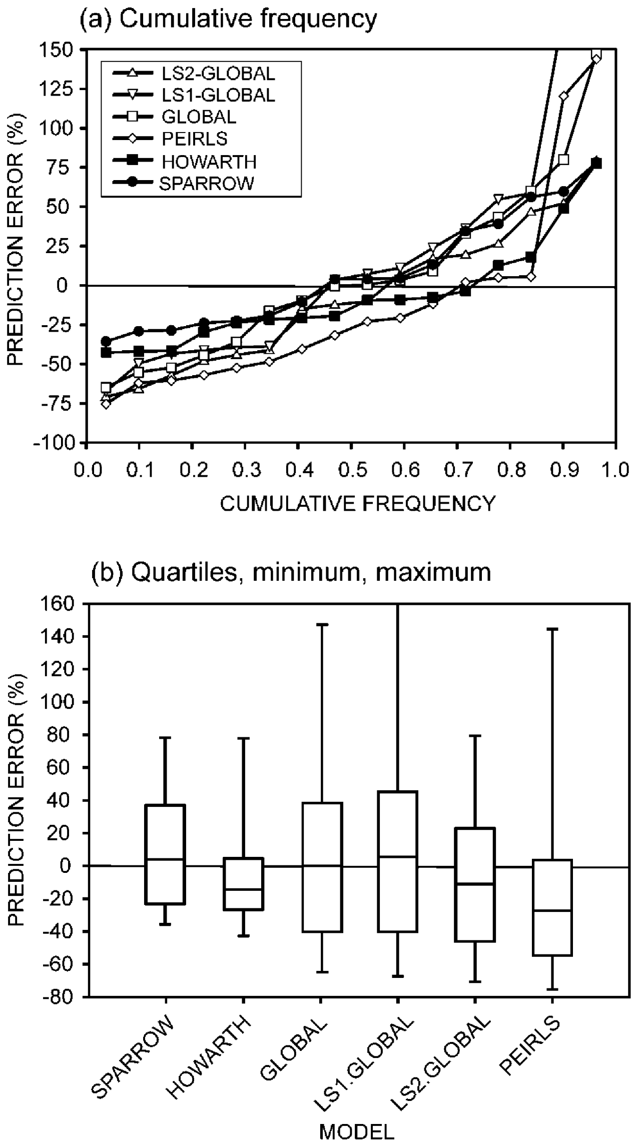

Figure 3. Errors in the predictions of stream nitrogen export from the application of the six watershed models to the 16 northeastern basins: (a) cumulative frequency plots (values of 174% and 373% for the LSI-GLOBAL model plot above the scale), and (b) box and whisker plots (the maximum of 373% for the LSI-GLOBAL model plots above the scale). Each box graphs the quartiles with the lower and upper edges representing the 25th and 75th percentiles, respectively. The midline plots the median. The upper and lower whiskers are drawn to the minimum and maximum values.

indistinguishable from zero. The LS2-GLOBAL and HOWARTH models have median errors of –11% and –14%, respectively. The greatest symmetry of the IQR about zero, an approximate measure of model bias, was observed for three of the models: GLOBAL (–43% to 35%), SPARROW (–23% to 36%), and LS1-GLOBAL (–43% to 48%). The HOWARTH, PEIERLS, and LS2-GLOBAL models all have a greater tendency to under predict stream export (the HOWARTH and PEIERLS models under predict export in 12 of the 16 watersheds; see Figure 3(a)). However, the HOWARTH model under predicts by a relatively small magnitude in comparison to other models. The large negative bias associated with the PEIERLS model predictions (median = –27%) is indicative of the tendency of this model to under predict stream export in the northeastern watersheds. The SPARROW model under predicts by the smallest magnitude of all of the models as evidenced by the proximity of the cumulative frequency curve to the zero line over the negative range of the prediction errors (i.e. probability <0.5) in Figure 3(a). The asymmetry in the 75th percentile (36%) as compared with the 25th percentile (–23%) indicates that the SPARROW model tends to over predict export by a somewhat larger magnitude than it under predicts export.

The smallest variability (i.e. highest precision) in the prediction errors as indicated by the magnitude of the interquartile range (IQR; Table 5) is found for the HOWARTH model (IQR = 26%), followed by the SPARROW (59%) and PEIERLS (60%) models. By comparison, the variability in prediction errors of the global models is about 30% higher than that for the SPARROW and PEIERLS models and about twice that of the HOWARTH model. The various global models differ only modestly in the magnitude of the variability in the prediction errors, with IQRs ranging from 66% to 80%. The LS2-GLOBAL model shows the smallest IQR, which is about 20% lower than that of the LS1-GLOBAL model.

Comparisons of the prediction errors for the northeastern watersheds (Table 5) with those for the original calibration data (Table 4) indicate that the greatest similarities among individual models are found in the magnitude of the variability of the prediction errors. For example, the variability (i.e. IQR) for the HOWARTH model predictions is 26.6% (Table 4) and 25.6% (Table 5), respectively, for the calibration and northeastern watersheds; variability in the prediction errors for the SPARROW model is 65.2% and 61.6%, respectively. Among the various global models, those based on the Seitzinger and Kroeze (1998) data display the most similar variability in prediction errors (GLOBAL IQR = 63.8 and 73.8; LS2-GLOBAL IQR = 71.6 and 66.3). Other comparisons of the model errors for the calibration and northeastern watersheds in Tables 4 and 5 are also of note. For example, distributions of prediction errors for the SPARROW model have a similar median (3.2% and

Table 5. Errors in model predictions of stream nitrogen export from applications in the northeastern watersheds. Errors are computed as the difference between the predicted and measured values of stream nitrogen export expressed as a percentage of the measured export (equation 1). The prediction errors are plotted in figure 3

Model	median	IQR[a]	Prediction Errors (%)			
			min.	25th percentile	75th percentile	max.
SPARROW	4.1	58.6	–35.6	–22.8	35.8	78.2
HOWARTH	–14.4	25.6	–42.7	–25.1	0.5	77.7
PEIERLS	–27.2	60.1	–75.3	–55.8	4.3	143.9
GLOBAL	0.1	73.8	–64.8	–42.8	35.3	147.2
LS1-GLOBAL	5.5	80.4	–67.3	–43.0	47.9	373.0
LS2-GLOBAL	–11.2	66.3	–70.7	–47.7	24.7	79.3

[a]Interquartile range (difference between the 25th and 75th percentiles of the distribution of errors).

4.1%) and quartile values. The various global models generally over predict stream export for the original calibration watersheds, whereas the GLOBAL and LS2-GLOBAL models tend to slightly under predict stream export in the northeastern watersheds. For the HOWARTH model, the prediction errors for the original calibration data display a slight positive bias (median = 3%; IQR from –6% to 21%) suggesting a tendency of the model to over predict export for the calibration watersheds. By contrast, this model has a somewhat greater tendency to under predict stream export in the northeastern watersheds.

4.3. *Factor-related model biases – northeastern US watersheds*

The results of the error analysis based on an application of the regression-based error models in equation 2 are shown in Table 6. For each model, we report the R^2, mean square error (MSE), regression coefficients for each of the explanatory variables, and the statistical significance (p-value) of the coefficients. In contrast to interpretations of conventional regression models, a high R^2 (also low MSE and small p values) for the error models identifies inaccuracies (i.e. ‘factor-related’ biases) in the specification of the original stream export models related to deficiencies in the model descriptions of nitrogen supply and processing. The coefficients of the error models quantify the magnitude of the change in prediction error that is expected to occur in response to a unit change in a watershed property, and may be directly compared among the six models. The interpretation of these coefficients in terms of the tendency for models to over or under predict stream nitrogen

Table 6. Results of the regression-based error models for the northeastern watersheds. The regression coefficients indicate the change in the percent prediction error per unit change in the land-area percentage (cultivated land or developed land), runoff (cm yr^{-1}), or drainage basin area ($\times 10^{-3}$ km^2). The intercept is expressed in units of percent prediction error

Model	R^2 (*p; statistical sig.*)	Mean Square Error (MSE)	Regression Coefficients (*p*; *statistical significance*) Cultivated Land	Developed Land	Runoff	Basin Area	Intercept
SPARROW	0.36	1127.6	–2.49	0.42	–1.85	–0.29	150.9
	(*0.260*)		(*0.053*)	(*0.785*)	(*0.220*)	(*0.621*)	(*0.130*)
HOWARTH	0.43	854.4	–1.68	–0.066	–3.29	–0.437	206.4
	(*0.157*)		(*0.120*)	(*0.961*)	(*0.022*)	(*0.397*)	(*0.026*)
PEIERLS	0.72	1490.7	–4.30	6.45	–3.21	–0.22	200.5
	(*0.005*)		(*0.008*)	(*0.003*)	(*0.075*)	(*0.741*)	(*0.085*)
GLOBAL	0.72	1196.5	–4.84	3.89	–3.79	–0.65	280.4
	(*0.004*)		(*0.002*)	(*0.029*)	(*0.025*)	(*0.293*)	(*0.013*)
LS1-GLOBAL	0.79	3439.8	–7.10	13.01	–5.51	–0.30	378.5
	(*0.001*)		(*0.005*)	(*<0.001*)	(*0.048*)	(*0.770*)	(*0.038*)
LS2-GLOBAL	0.75	695.9	–3.93	–2.11	–2.64	–0.858	224.1
	(*0.002*)		(*0.001*)	(*0.102*)	(*0.038*)	(*0.082*)	(*0.010*)

export in response to changes in watershed characteristics will depend on the range of the prediction errors and their symmetry about zero. For models with less overall bias (as measured by the median error), a positive coefficient indicates that the prediction errors generally become more positive and larger in magnitude (tendency to over predict export) in response to increases in the magnitude of the factor. However, in the case of the PEIERLS and HOWARTH models, where the majority of the prediction errors are negative (i.e. under prediction), a positive coefficient, for example, would indicate that over much of the range of the explanatory factor the prediction errors become less negative and smaller (i.e. less under prediction) in response to increases in the watershed property. Statistical significance of the model coefficients is based on the application of a t-test. The residuals of the error models generally comply with the regression assumptions; they are approximately normal with relatively constant variance.

The R^2 of the error models span a large range, indicating that from 36% to 79% of the spatial variability in the prediction errors can be explained by the four watershed properties (Table 6). The MSE varies in response to how well the watershed characteristics explain the prediction error, but also as a function of the magnitude of the variability in the prediction errors (i.e. the precision of the model predictions). The SPARROW and HOWARTH error models have the lowest R^2 values (0.36 and 0.43, respectively) among the various models with explanatory variables displaying moderate to very weak significance (p = 0.02 to 0.96). The small MSE for the HOWARTH model reflects the low variability (IQR) in the prediction errors. The other models show much higher R^2 values (0.72–0.79) with correspondingly larger coefficient values and higher levels of statistical significance. The signs of the explanatory coefficients are generally consistent among most of the models indicating that the direction of the biases are similar; however, the magnitudes of the coefficients vary indicating differences in the importance of these factors in potentially explaining prediction errors in the nitrogen export models.

Cultivated land area is found to be statistically significant in all models, but the magnitude of the effect based on the size of the coefficient value differs appreciably among the models. Cultivated land area is inversely correlated with the prediction errors, indicating that there is a tendency for most of the export models to under predict stream export in watersheds that are more highly agricultural and over predict in watersheds with less cropland and larger amounts of forested lands. The partial residual plots (Montgomery & Peck 1982) in Figure 4 display the relation between the prediction errors and cultivated land area for each of the models. The slope of the lines in Figure 4 corresponds to the slope coefficient of the cultivated land area term in the error models reported in Table 6. These plots show variations in prediction errors, adjusted for the variability that is explained by the other predictor variables in the model (i.e. developed land, runoff, and basin area); the partial error residuals (vertical axis) are also adjusted for the mean so that the residuals are centered about zero. The plots in Figure 4 illustrate the relative strength of these relations for the various export models. The strength varies in proportion to the magnitude of the slope of the fitted line relative to the variability of the partial residuals about the line. The fitted lines also illustrate for the various models how the prediction errors change in magnitude over the full range of cultivated land area in the watersheds. The PEIERLS and various global models show the strongest relations between the model prediction errors and cultivated land area as evidenced by the relatively steep slopes of the fitted lines and their relatively high statistical significance (Table 6). The coefficients for cultivated land in these models are typically one and a half to five times larger than that observed for the HOWARTH and SPARROW models, which display coefficients of –1.67% and –2.49% error per unit change in cultivated land percentage, respectively. Thus, the PEIERLS and global models under predict export by a much larger magnitude

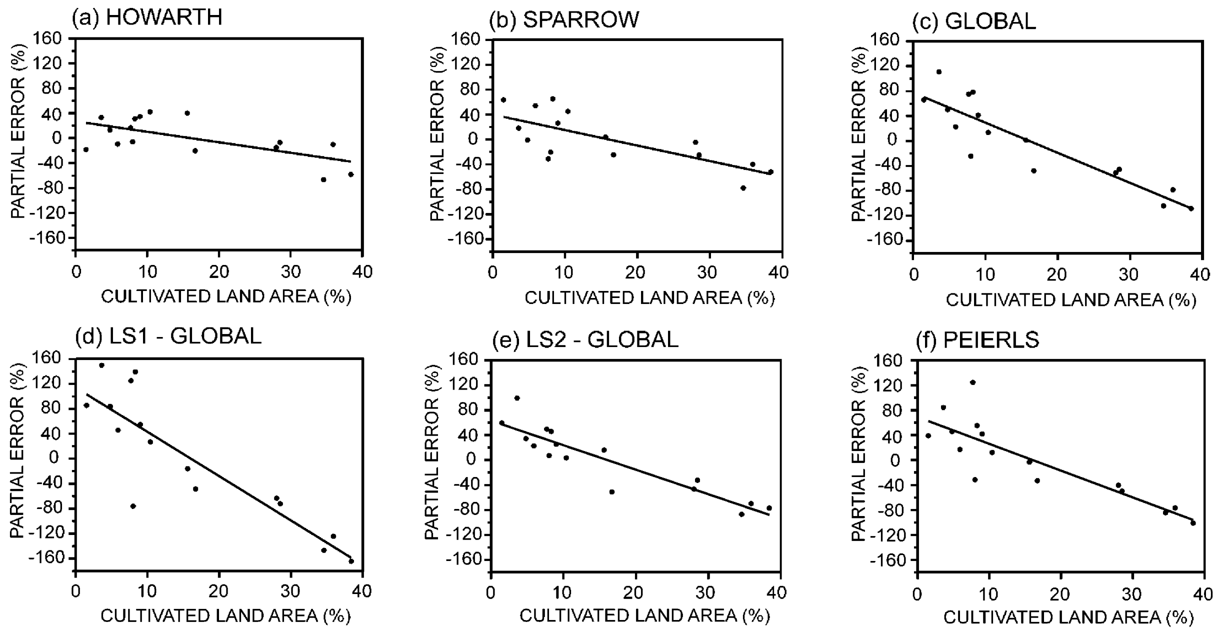

Figure 4. Partial residual plots of the relation between prediction errors and cultivated land area for the six error models: (a) HOWARTH, (b) SPARROW, (c) GLOBAL, (d) LSI-GLOBAL, (e) LS2-GLOBAL, and (f) PEIERLS. The slope of the lines corresponds to the slope coefficient of the cultivated land area variable in the error models in Table 6. The partial errors are adjusted for variability explained by the other model predictors (developed land, runoff, and basin area), and are centered about zero by subtracting the mean. The smaller slopes for the HOWARTH and SPARROW error models indicate that cultivated land area explains smaller amounts of the over and under prediction of nitrogen export in the watershed models.

in more highly cultivated watersheds than do the other models. In less cultivated basins, the global models also over predict export by a larger magnitude than the other models. The smaller slopes for the HOWARTH and SPARROW error models indicate that cultivated land area or related factors potentially explain smaller amounts of the over and under prediction of nitrogen export by the watershed models.

Developed land area is also found to be significant in explaining prediction errors for the PEIERLS and various global models. A positive relation is observed suggesting that there is a tendency for the stream export models to under predict in less developed basins (i.e. also basins with more area in forest and cultivated land). The magnitude of the coefficient values of the PEIERLS and LS1-GLOBAL error models is considerably larger than observed for cultivated land area in these models (i.e. 6.5 and 13% error per unit change in percentage land area). Developed land area is not statistically significant in the SPARROW ($p = 0.79$) and HOWARTH ($p = 0.96$) models.

Runoff is moderately to highly significant in all of the error models except SPARROW, which is statistically insignificant ($p = 0.22$). The negative correlation with the prediction errors suggests that there is a tendency for the stream export models to under predict in basins with high runoff and over predict in basins with low runoff. The coefficients for the SPARROW, LS2-GLOBAL, and HOWARTH error models are among the smallest (–1.85, –2.32, and –2.48% change in prediction error per unit change in runoff – cm yr^{-1}, respectively). Coefficients for the PEIERLS and global models range from –3.2 to –5.5%.

Drainage basin size is generally insignificant or only moderately significant in most of the error models. Four of the six models display the smallest change in prediction error per unit change in basin size (i.e. about –0.2% to –0.4% error per unit change in drainage area – 1000 km^2 – for the SPARROW, HOWARTH, PEIERLS, and LS1-GLOBAL models). None of these coefficients are statistically significant ($p > 0.30$). Two of the remaining models (GLOBAL, LS2-GLOBAL) display larger coefficients (–0.6% and –0.9% error per unit change in runoff – cm yr^{-1}, respectively), but have weak levels of statistical significance ($p > 0.08$).

5. Discussion of the error analysis

In evaluating the accuracy of the predictions of stream nitrogen export for the six watershed models, we presented two measures of model bias. First, we reported the overall bias (Table 5; Figure 3). This measure is defined as the median prediction error, and provides a measure of the typical bias in the model predictions of nitrogen export for the entire set of northeastern watersheds. The overall bias indicates whether there is evidence that the model predictions are systematically too high or low among the 16 basins (see Figure 3). Second, we reported the factor-related bias as quantified by the regression-based model coefficients (equation 2) in Table 6. These coefficients describe the magnitude of the change in the prediction error that occurs in response to unit changes in various watershed properties. Such changes are indicative of biases in the model predictions of nitrogen export caused by model mis-specification, including sources or delivery processes that are not explicitly included in the models or model coefficients that inaccurately describe the supply and transport of nitrogen. This information is useful for assessing model performance over a range of specific environmental conditions. Strong evidence of factor-related bias implies that improvements are feasible in the accuracy of the model predictions through improved calibrations or modifications of the model structure. As an additional note, a model with relatively little overall bias can display significant factor-related biases – a pattern that is evident for several of the models examined here. For example, although the GLOBAL and LS1-GLOBAL models display relatively small median prediction errors (<6%), large factor-related biases are detected that are related to specific watershed properties.

Based on the magnitude and statistical significance of the coefficients of the error models (Table 6), the prediction errors of the stream export models appear to be most strongly related to land use and runoff. The land-use factors indicate limitations in how well the stream export models account for nitrogen sources and/or attenuation processes associated with cultivated, developed or forested lands, which may be related to such factors as nitrogen fixation, N uptake by vegetation, or cultural inputs of N. Prediction errors are negatively correlated with cultivated land percentage in most of the error models indicating that the stream export models over predict nitrogen export in less agricultural basins and under predict in more agricultural basins. The magnitude of the prediction errors vary widely among the models over the range of cultivated land area in the watersheds (see Figure 4). The HOWARTH and SPARROW models display the smallest changes in prediction errors in response to changes in the cultivated land area; changes in the prediction errors of these models are at least 50% smaller than those of the other models. The lower cultivated land-related bias in the HOWARTH model may reflect the value of directly accounting for food/feed imports and exports and nitrogen fixation in crops, which are not explicitly included in the other models. The inclusion of N in livestock wastes in the SPARROW model accounts for feed consumption, which may help explain the model's relatively low cultivated land-related bias. The tendency of most of the models to over predict nitrogen export in less agricultural basins may also reflect limitations in the model estimates of nitrogen supply and attenuation on forested lands. Any effects of forestland would be expressed in the error models as a negative correlation with the percent cultivated land area. The prediction errors are also significantly correlated with developed land area for all models except for the SPARROW and HOWARTH models. The positive relation between the prediction errors and developed land area indicates a tendency of the models to over predict stream export in more developed basins and to under predict export in less developed basins. Less developed basins include watersheds that are more highly forested as well as those with cultivated lands. The measures of cultivated and developed land area in the 16 northeastern watersheds are uncorrelated (r = –0.06; variance inflation factors = 1), and therefore, provide independent land-use information in the error models. The use of additional land-use terms (e.g. forested land) in preliminary error models led to colinearity problems (variance inflation factors >9), and thus, were not used in the final models.

Runoff and drainage basin size are included in the error models to quantify the potential effects on stream N export of nitrogen mobilization processes that are independent of land use. Although runoff may reflect climate-related effects on stream export related to the supply of nitrogen in vegetation, the land-use terms in the error models should account for variations in many of the cultural and natural sources of N. In addition, natural sources of nitrogen, such as forest N fixation, are relatively minor contributors to the N budgets of the northeastern watersheds (Boyer et al. 2002; Table 8). Runoff and basin size are potentially related to many of the same hydrologic and time-dependent properties that influence nitrogen attenuation on the landscape and in streams (e.g. water velocity, flow, water time of travel); however, the two factors are relatively uncorrelated (r = –0.27; variance inflation factor = 1) for the northeastern watersheds, thereby providing independent explanatory measures in the error models.

The negative relation between the prediction errors and runoff indicates that the stream export models tend to under predict in watersheds where runoff is high and over predict in watersheds where runoff is low. These effects are least significant in the SPARROW model and most pronounced for the various global models. The global model predictions of stream export for major rivers of the world (Caraco & Cole 1999; Seitzinger & Kroeze 1998) also show a moderately significant negative relation between the prediction errors and runoff for the original calibration data. Runoff is sensitive to the effects of climate, geology, soils, and stream morphology on the rates of surface and subsurface flow. Runoff may influence the rates of nitrogen uptake and storage, and the permanent removal of nitrogen in terrestrial and aquatic environments by affecting water residence times and water contact with sites suitable for denitrification, such as anoxic soils, benthic stream sediments, channel hyporheic and riparian zones, and wetlands (Kelly et al. 1987; Hill 1996; Sauer et al. in press; Molot & Dillon 1993). The increased mobilization of N with runoff may explain the nearly proportional positive relation that has been observed between stream nitrogen export and runoff in developed (Sauer et al. in press; Behrendt 1996) and undeveloped (Lewis et al. 1999, in press) watersheds. The negative correlation ($r = -0.57$) between runoff and estimates of the total loss of nitrogen in the northeastern watersheds (computed as the difference between major N inputs and stream N exports; Boyer et al. 2002) is consistent with the effects of these nitrogen removal processes on stream nitrogen export; larger losses of nitrogen are observed in northeastern watersheds with relatively low runoff and smaller losses occur in watersheds with high runoff. Therefore, the negative relation between prediction errors and runoff suggests that the under prediction of stream export in watersheds with high runoff may be caused, in part, by the overestimation of the rates of nitrogen attenuation in the models. Conversely, the over prediction of stream export in watersheds with low runoff may reflect the underestimation of the rates of nitrogen attenuation rates in the export models.

Drainage basin size is found to be a relatively insignificant explanatory variable in nearly all of the error models. This suggests that the prediction errors are not strongly related to scale-dependent characteristics of the watersheds as measured by drainage size. It is possible that scale-dependent properties related to the rates of water flow and storage, such as water travel time, may be more clearly described by runoff for the northeastern watersheds, which may partially explain its importance in many of the error models. In contrast to the other models, drainage basin area is moderately significant in the LS2-GLOBAL error model, and is much larger in magnitude than observed for the other error models. Thus, this model displays direct evidence of a scale-dependency in the prediction errors related to basin size with a tendency of the model to under predict stream nitrogen export in large watersheds.

The error analysis indicates that the watershed models with more complex descriptions of nitrogen sources and attenuation processes have appreciably lower bias and higher precision in their predictions of nitrogen export. The two models with the most detailed descriptions of nitrogen sources, land and water N attenuation, and water flow paths (HOWARTH, SPARROW) show smaller factor-related biases in the predictions of stream nitrogen export. This is supported by both the smaller magnitude and statistical significance of the coefficients in the associated error models. These findings suggest that model complexity has a beneficial effect on prediction accuracy. The HOWARTH model gives a detailed accounting of agricultural sources, including fertilizer use, crop N fixation, and the import and export of nitrogen in food and feeds. The SPARROW model spatially references stream monitoring data, point and diffuse nitrogen sources, and landscape properties to surface water flow paths defined by a digital drainage network. Agricultural sources include fertilizers and livestock wastes. The model explicitly quantifies the rates of nitrogen removal on the landscape and in streams though the use of spatial referencing and mass-balance constraints. By contrast, the PEIERLS model lacks explicit point and diffuse source terms, and relies solely on population density as a predictor variable. This may contribute to the strong tendency of the model to under predict in more undeveloped (i.e. less populated) watersheds. In the global models, agricultural nitrogen sources are quantified exclusively as a function of fertilizer use and runoff; point-source contributions are a function of population density and urban land area. Neither the PEIERLS nor the global models account for the location of sources and water travel times in the watersheds.

6. Model predictions of source contributions

The predictions of source contributions to stream nitrogen export are reported for the stream export models in Table 7 and Figure 5. The classification of source contributions and the model assumptions and forms of nitrogen differ considerably among the various models, which affect the comparisons of model estimates. Unlike the previous evaluations of stream nitrogen export, there is no known measure of the magnitude of source contributions to streams, and thus, the predictions of nitrogen sources can only be compared among the models. Although the error models could be used to correct for model biases in the estimates of source shares, this would require assumptions about how the explanatory factors in the error models, especially developed and cultivated land, correspond to the source variables in the stream export models. The inverse relation between the prediction errors of most of the models and cultivated land area as previously discussed generally implies that the stream export models have a tendency to under predict in more agricultural watersheds and to over predict in more forested and less agricultural watersheds. This allows for the possibility that the effect of agricultural sources may be somewhat underestimated in more agricultural basins. It is also possible that prediction biases related to developed land area may affect the accuracy of the point source contributions estimated by some the models.

Estimates of nitrogen inputs from major watershed sources based on nitrogen budgets (see Boyer et al. 2002) are presented in Table 8 for comparison with the model predictions of source contributions at the watershed

Table 7. SPARROW predictions of source contributions to stream nitrogen export from the northeastern watersheds

River Name	Source Contributions to Export (% of export)[a]					Watershed Attenuation		
	Point	Fertilizer Use	Livestock Wastes	Atmosphere[b]	Nonagr. Nonpt.	In-stream Loss[c] (% of stream inputs)	In-stream Loss[d] (% of basin loss)	Landscape Loss[e] (% of basin loss)
Penobscot	1	5	2	31	61	47	54	46
Kennebec	2	3	5	36	55	46	37	63
Androscoggin	3	5	5	34	54	40	30	70
Saco	1	2	2	34	60	35	25	75
Merrimack	18	5	4	29	44	34	15	85
Charles	74	2	1	9	14	30	7	93
Blackstone	37	6	5	20	32	31	23	77
Connecticut	6	8	7	38	39	47	28	72
Hudson	4	9	8	39	41	45	23	77
Mohawk	13	12	15	36	25	40	21	79
Delaware	15	14	9	35	28	38	29	71
Schuylkill	37	17	15	17	13	28	17	83
Susquehanna	7	17	24	30	23	45	25	75

Table 7. Continued

River Name	Source Contributions to Export (% of export)[a]					Watershed Attenuation		
	Point	Fertilizer Use	Livestock Wastes	Atmosphere[b]	Nonagr. Nonpt.	In-stream Loss[c] (% of stream inputs)	In-stream Loss[d] (% of basin loss)	Landscape Loss[e] (% of basin loss)
Potomac	4	26	23	31	17	59	34	66
Rappahannock	1	26	23	27	23	59	18	82
James	1	11	14	39	35	58	18	82
25th Percentile	*2*	*5*	*4*	*28*	*23*	*35*	*18*	*71*
Median	*5*	*9*	*8*	*32*	*33*	*43*	*24*	*76*
75th Percentile	*16*	*15*	*15*	*35*	*47*	*47*	*30*	*82*

[a]Expressed as a percentage of the predicted stream nitrogen export from the watersheds.
[b]Atmospheric deposition contributions to stream export are based on wet-fall deposition. Land-to-water delivery fractions exceed unity indicating that additional atmospheric forms of nitrogen (e.g. ammonium, organic) are included (see Alexander et al. 2001).
[c]The in-stream loss of nitrogen in RF1 (river reach file 1) reaches.
[d]The mass of nitrogen removed in streams is estimated as (M/(1-S)) * S, where M is the measured stream nitrogen export in equation (1) and S is the in-stream nitrogen loss estimated by SPARROW and expressed as a fraction of the stream inputs. The mass of nitrogen removed in streams is expressed as a percentage of the total loss in the watershed (estimated as the differences between the watershed inputs – fertilizer, total atmospheric deposition, crop and forest N fixation, and net food/feed imports – and riverine N export; see Boyer et al. 2002).
[e]Computed as the complement of the in-stream loss percentage.

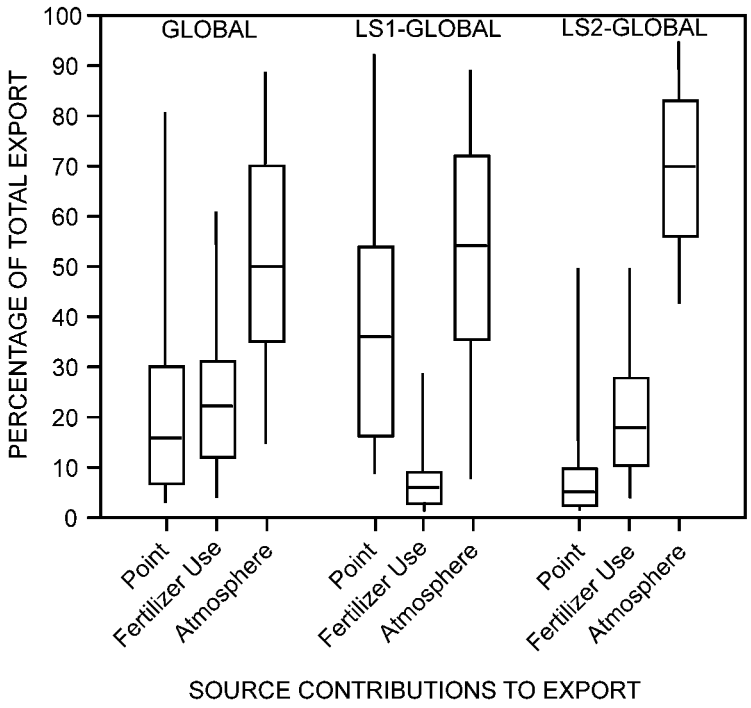

Figure 5. GLOBAL, LSI-GLOBAL, and LS2-GLOBAL model predictions of source contributions to stream nitrate-nitrogen export for the northeastern watersheds. Sources contributions are expressed as a percentage of the stream nitrate-nitrogen export. Each box graphs the quartiles with the lower and upper edges representing the 25th and 75th percentiles, respectively. The midline plots the median. The upper and lower whiskers are drawn to the minimum and maximum values.

outlets. The budget estimates would accurately describe nitrogen source contributions to stream export at the watershed outlets only if sources are uniformly distributed in the watersheds and loss processes are assumed to operate equally on all sources. Nevertheless, the budget estimates provide useful information on the major inputs of N for comparison with the model predictions. Atmospheric nitrogen typically represents from about 29% to 59% of the total inputs to the watersheds, based on the interquartile range (Table 8). With the exception of forest N fixation, which typically is less than 6% of the total inputs of nitrogen, much of the remaining portions of the nitrogen inputs originate from agricultural-related sources and products either applied directly to fields in fertilizers or consumed in food and feeds. Nitrogen inputs from food/feed imports are typically less than about 15% of the total inputs from major sources although N imports represent more than a third of the total nitrogen inputs in the more populated watersheds. Nitrogen in food imports make their way to streams via sewered and unsewered waste systems.

The source predictions of the various global models (GLOBAL, LS1-GLOBAL, LS2-GLOBAL) are shown in Figure 5. These models classify sources according to three types: point, atmospheric, and fertilizer (agricultural). Because these models predict nitrate rather than total nitrogen export as a function of the specified sources, use of the model to characterize sources in the northeastern watersheds assumes that the source shares for nitrate are equivalent for other nitrogen forms. In addition, other sources of nitrogen, such as natural fixation, are not explicitly described by the models, and would tend to be included by the three model sources. The GLOBAL model indicates that atmospheric deposition represents the predominant source in stream export (median = 50%; IQR = 35 to 70%), especially in the northern watersheds of the Penobscot, Kennebec, Androscoggin, and Saco where atmospheric contributions exceed 75%. Contributions from agricultural sources are second in importance in most of the watersheds (median = 22%; IQR = 12 to 31%). Agricultural sources are relatively large in all watersheds south of the Mohawk, but only rarely exceed 50%. Point sources (median = 16%; IQR = 7 to 30%) are only slightly smaller than agricultural sources, but contribute the largest shares (>50%) in the two smallest watersheds, the Charles and Blackstone, and in the Merrimack. The LS1-GLOBAL model suggests that atmospheric deposition has a similar, but slightly higher relative contribution than in the GLOBAL model (median = 54%; IQR = 35 to 72%). Larger differences are noted in the other sources, where point sources are typically higher (median = 36%; IQR = 16 to 54%) and agricultural sources are generally lower (median = 6%; IQR = 3 to 9%) than reported for the GLOBAL model. The LS2-GLOBAL model typically shows higher atmospheric contributions (median = 70%; IQR = 56 to 83%) than either of the other global models. The higher contributions are offset by lower contributions from both point and fertilizer-related sources; however, the LS2-GLOBAL model predictions of contributions from the fertilizer-related sources are generally similar to those predicted by the GLOBAL model.

The SPARROW model indicates that atmospheric sources and non-agricultural diffuse sources each contribute about one third of the nitrogen to stream export in most of the watersheds (Table 7). Some of the largest contributions from atmospheric nitrogen (>35%) are found in the Kennebec, Connecticut, Hudson, Mohawk, and James watersheds. Most watersheds have atmospheric contributions that are only slightly lower that this, and typically lie in the range from 28 to 35%. Agricultural sources (fertilizer plus livestock waste sources) are typically less than 30%, but contribute nearly 50% of the nitrogen in the Susquehanna, Potomac, and Rappahannock

Table 8. Estimates of nitrogen inputs from major sources and nitrogen losses in the northeastern watersheds

River Name	Nitrogen Inputs to Watershed[a] (% of total inputs)					In-Stream Loss[c]	In-Stream Loss[d]	Landscape Loss[e]
	Atmosphere[b]	Fertilizer Use	Net Feed & Food Imports	Agric. N Fixation	Forest N Fixation	(% of stream inputs)	(% of basin loss)	(% of basin loss)
Penobscot	70	11	2	9	7	68	130	–
Kennebec	64	5	11	15	5	63	74	26
Androscoggin	61	6	15	12	5	52	48	52
Saco	73	3	7	8	9	47	41	59
Merrimack	42	7	34	10	7	61	45	55
Charles	23	4	64	4	5	37	10	90
Blackstone	31	9	43	9	8	53	57	43
Connecticut	45	13	21	17	5	66	61	39
Hudson	55	11	9	20	5	58	38	62
Mohawk	34	13	11	40	2	60	48	52
Delaware	43	19	8	24	7	60	72	28
Schuylkill	22	23	28	23	4	52	48	52
Susquehanna	31	17	17	31	5	76	97	3

Table 8. Continued

River Name	Nitrogen Inputs to Watershed[a] (% of total inputs)					In-Stream Loss[c]	In-Stream Loss[d]	Landscape Loss[e]
	Atmosphere[b]	Fertilizer Use	Net Feed & Food Imports	Agric. N Fixation	Forest N Fixation	(% of stream inputs)	(% of basin loss)	(% of basin loss)
Potomac	17	23	27	27	6	69	53	47
Rappahannock	21	24	14	34	7	58	17	83
James	34	13	15	25	13	72	33	67
25th percentile	*29*	*7*	*10*	*10*	*5*	*53*	*40*	*37*
Median	*38*	*12*	*15*	*18*	*6*	*60*	*48*	*52*
75th percentile	*57*	*17*	*27*	*25*	*7*	*67*	*63*	*60*

[a]Expressed as a percentage of the total nitrogen inputs from major watershed sources (Boyer et al. 2002).
[b]NO_y total wet and dry oxidized components of atmospheric deposition (Boyer et al. 2002).
[c]The in-stream loss of nitrogen in RF1+RF3 scale reaches based on the RivR-N model (see Seitzinger et al. 2002).
[d]The mass of nitrogen removed in streams is estimated as (M/(1-S)) * S, where M is the measured stream nitrogen export in equation (1) and S is the in-stream nitrogen loss estimated by the RivR-N model (Seitzinger et al. 2002) and expressed as a fraction of the stream inputs. The mass of nitrogen removed in streams is expressed as a percentage of the total loss in the watershed (estimated as the differences between the watershed inputs – fertilizer, total atmospheric deposition, crop and forest N fixation, and net food/feed imports – and riverine N export; see Boyer et al. 2002).
[e]Computed as the complement of the in-stream loss percentage.

watersheds. Point sources (industrial and municipal) typically contribute less than 15%. Much larger point-source contributions (37 to 74%) are found in the more highly populated watersheds of the Charles, Blackstone, and Schuylkill. Non-agricultural diffuse sources are largest in the highly forested northern watersheds, contributing from 44 to 61%; in other watersheds, the contributions are less than about 30%. This source category is proportional to non-agricultural land area, and accounts for remaining sources of nitrogen that are not specified by the other source inputs in the model. These sources may include nitrogen in the surface and subsurface flows from urban, forested, wetlands, and barren lands. Nitrogen from forested lands may include biotic N fixation. Groundwater nitrogen is implicitly included in the agricultural and non-agricultural sources specified in the model, and may include older waters from a mixture of sources.

A comparison of the global and SPARROW models shows that the estimates of agricultural contributions are similar among the GLOBAL, LS2-GLOBAL, and SPARROW models. Atmospheric sources consistently contribute less to stream export according to the SPARROW model (about 30 to 40% less) than predicted by the various global models. Point source contributions are similar in the SPARROW and LS2-GLOBAL model, which are about one half of the magnitude of the point source shares predicted by the GLOBAL model. The SPARROW model classifies about a third of the nitrogen contributions as non-agricultural diffuse sources, which may include a mixture of sources in the groundwater and runoff from urban and rural lands. The point source shares from SPARROW and GLOBAL model are highly correlated with the percentage of land classified as developed ($r = 0.95$ for both models). The percentage of cultivated land is highly correlated with SPARROW estimates of agricultural contributions ($r = 0.92$); the GLOBAL model shows somewhat less correlation ($r = 0.53$). Municipal point sources in the SPARROW model expressed on a per capita basis have a median of 3.3 kg-N person^{-1} with an interquartile range from 1.8 to 5.8 kg-N person^{-1} (Alexander et al. 2001), and compare with a per capita rate of 1.85 kg-N person^{-1} for the GLOBAL model. Per capita rates for residential wastewater effluent in the United States have been previously estimated to range from 2.2 to 7 kg-N person^{-1} (Thomann 1972; US EPA 1980).

The estimation of source contributions to streams using the HOWARTH model is difficult because of uncertainty over how the intercept should be apportioned to each source and the lack of separate point and cultural diffuse inputs to the model. Although the model intercept of –120 provides a reasonably accurate adjustment to total stream export for additional N sources and watershed attenuation, adjustments for these factors cannot be reliably made to individual source terms. For example, the model does not ensure that the individual sources (atmospheric deposition, net antropogenic inputs) have positive nitrogen mass or that the mass contributions for sources are less than total stream export. This is not resolved by the use of any of several assumptions about how the intercept might be apportioned to the sources, including the assumption that the intercept is distributed to each source in proportion to the source's share of the net inputs of nitrogen to each watershed. Under this assumption, atmospheric contributions range from 67 to 115% and net anthropogenic sources range from –15 to 38%.

7. Model predictions of nitrogen attenuation in watersheds

We made separate estimates of the rates of nitrogen loss in streams and on the landscape (see Table 7) using empirically derived N loss coefficients in the SPARROW model (Alexander et al. 2000, 2001) and estimates of the total N loss in the northeastern watersheds, based on the difference between major N inputs and stream nitrogen export (see Boyer et al. 2002). The other nitrogen export models examined here lack explicit coefficients that quantify the rates of N removal in watersheds. We compared these estimates with those generated by an application of the RivR-N model to the northeastern watersheds (see Table 8; Seitzinger et al. 2002). RivR-N is a statistical in-stream loss model that was calibrated using literature observations from mass balance and denitrification studies for North American and European lakes and streams. The model was used to estimate the removal of nitrogen in streams and lakes of the northeastern watersheds as a function of the physical and hydraulic characteristics (i.e. depth, time of travel) of the water bodies.

SPARROW estimates of stream nitrogen losses, when expressed as a percentage of the total quantities of nitrogen removed in the watersheds (Boyer et al. 2002), range from 7% to 54% (median = 24%; IQR = 18 to 31%; Table 7). These estimates suggest that a majority – typically about 75% – of the nitrogen loss in watersheds can be explained by attenuation processes on the landscape (median = 76%; IQR = 71 to 82%). By comparison, RivR-N estimates of in-stream nitrogen loss are higher. When expressed as a percentage of the total quantities of nitrogen removed in the watersheds, are typically about 48% (IQR = 40 to 63%; Table 8). Thus, according to this model, landscape processes would typically account for 37 to 60% (IQR; median = 52%) of the total quantities of N removed in the watersheds.

The higher estimates of in-stream loss by the RivR-N model may primarily reflect differences in the spatial scale of the river networks used to derive the estimates. The RivR-N model includes additional nitrogen losses in streams smaller than RF1 streams (River Reach File; 1:500,000 scale; see Seitzinger et al. 2002) to which SPARROW is applied. The RivR-N model includes

1:100,000-scale reaches that are located upstream of RF1 streams. The RivR-N loss estimates, expressed as a percentage of external inputs to streams, are about 25 to 60% higher than the SPARROW estimates (median = 39%). By contrast, the RivR-N loss estimates for the 1:500,000-scale RF1 streams are only modestly larger than those for SPARROW (median ratio of RivR-N to SPARROW loss percentage = 1.11; IQR = 0.95 to 1.16; r = 0.75). Inclusion of the smaller 1:100,000-scale streams increases the loss percentages by about 10 to 20 percentage points (Seitzinger et al. 2002). As much as about 25% of the difference in stream loss between the two methods may relate to non-uniformities in the geographic distribution of sources in the watersheds. When SPARROW estimates of in-stream loss are derived under an assumed uniform spatial distribution of sources (identical to that assumed by the RivR-N model), the estimated losses are typically larger (median = 11%; IQR = 5 to 25%) than those in Table 7, which are based on the actual reach locations of sources. This suggests that a larger proportion of the point and diffuse sources are probably located in the lower portions of the watersheds (e.g. urban sources) and undergo less decay during the shorter travel times to watershed outlets.

8. Summary and conclusions

We evaluated the accuracy (bias and precision) of six nitrogen export models having varying levels of spatial resolution and process complexity and representing various empirical and quasi-empirical models that have been applied to large watersheds. Four of the models were previously described in the literature; two models were statistically calibrated in this study using published data sets for the largest rivers of the world. Many of the models were previously shown to explain large portions of the spatial variability in nitrogen export from rivers in major continents of the world according to reported R^2 statistics. However, the accuracy (bias and precision) of model predictions of stream export, which R^2 alone does not reliably measure, has not been previously reported and compared among the models. This study illustrates the value of using more reliable methods than R^2 to evaluate model performance. We validated the models using detailed data on stream nitrogen export, land-use, and natural and cultural inputs of nitrogen for 16 northeastern watersheds in the United States. The watersheds cover a sufficient portion of the range of the conditions present in the original calibration watersheds so as to provide an appropriate set of locations for evaluating the models. The analysis improves understanding of how the models perform over a range of environmental settings and how model complexity affects prediction accuracy.

Most of the models predicted stream nitrogen export to within 50% of the measured export in a majority of the watersheds; however, all models showed at least small amounts of bias in the model predictions. The three models with the smallest bias (SPARROW, LS1-GLOBAL, and GLOBAL) have a median prediction error of less than 5%. The PEIERLS model had the largest bias (median error = –27%) followed by the HOWARTH model (median error = –14%); both of these models under predicted nitrogen export in 12 of the 16 watersheds. The lowest variability in the prediction errors (i.e. most precise estimates of stream export) was observed for the HOWARTH model, followed by the SPARROW and PEIERLS models.

We developed regression-based models of the prediction errors to determine whether biases in model predictions are potentially caused by mis-specification of the models in relation to various watershed characteristics (i.e. 'factor-related' bias). Such biases may be caused by sources or delivery processes that are not explicitly included in the models or model coefficients that inaccurately describe the supply and transport of nitrogen. This measure of bias provides information about the performance of the nitrogen export models in specific environmental settings. Evidence of factor-related bias implies that improvements are feasible in the accuracy of the model predictions through improved calibrations or modifications of the model structure.

The two nitrogen export models with the smallest factor-related biases (SPARROW, HOWARTH), as evidenced by small coefficient values for each of the four watershed properties (cultivated land, developed land, runoff, drainage size) evaluated in the error models, had prediction biases that were at least 50% and smaller. The prediction biases were also less statistically significant than those detected for the other models. Because these models have more detailed descriptions of nitrogen sources, land and water attenuation, and water flow paths than the other models, the results suggest that model complexity has a beneficial effect on the accuracy of the predictions of stream export. The HOWARTH model gives a detailed accounting of agricultural sources, including crop N fixation and the import and export of foods and feeds. SPARROW spatially references stream monitoring data, point and diffuse nitrogen sources, and landscape properties to surface water flow paths and imposes mass-balance constraints to empirically estimate the rates of nitrogen transport on the landscape and in streams.

The evaluations of factor-related biases indicated that the prediction errors of all of the export models are inversely correlated with cultivated land area. Thus, there is a tendency for the models to under predict stream export in watersheds that are more highly agricultural and over predict in watersheds with less cropland and larger amounts of forested lands. The lower cultivated land-related bias for the HOWARTH model may reflect the value of a more detailed accounting of the supply and transport of nitrogen in agricultural watersheds, including nitrogen fixation in crops, feed imports, and crop exports. These are not explicitly accounted for in the various global models. The inclusion of the livestock waste source in the SPARROW model provides additional specification of agricultural sources that may account for its relatively low cultivated land-related bias.

All of the export models except for SPARROW showed a statistically significant negative correlation between prediction errors and runoff. In view of the effects of runoff on stream export and nitrogen attenuation in watersheds, this finding suggests that the models may need to account more effectively for nitrogen loss processes (e.g. denitrification, storage) at the watershed scale related to the rates of water transport through surface and subsurface pathways. The rates of nitrogen removal on the landscape and in streams may be mediated by various hydrogeologic factors related to runoff (e.g. channel density, stream morphology, water velocity, soil texture, groundwater storage). More explicit descriptions of these factors in the models may improve prediction accuracy.

Comparisons of the model predictions of source contributions to stream export displayed the greatest consistency in the results for agricultural sources; notable differences were found in the estimates of point sources and atmospheric contributions. Although there are uncertainties as to the specific effects of the prediction biases on the estimates of source contributions, the error analysis suggests that many of the models may underestimate the contributions of agricultural sources in more highly agricultural watersheds. Some of the models may overestimate N contributions from point sources in more highly developed watersheds.

The study represents an initial effort to validate the reliability of several prominent stream nitrogen export models, and provides information for guiding future applications and enhancements to the models. The regression-based error analysis illustrated here can be readily applied in future evaluations of stream export models. It provides a reasonable approach for validating and possibly correcting watershed models. The method also identifies factor-related biases that can potentially be eliminated through improved model calibrations. Future assessments of model errors would benefit from evaluations of additional stream export models, such as GWLF (Haith & Shoemaker 1987) and export-coefficient models (Johnes 1996), as well as the inclusion of larger numbers of watersheds representing a more diverse range of climate, land uses, and nitrogen sources. Because of the importance of landscape attenuation and nitrogen processing related to specific land uses, future error assessments should make use of deterministic landscape models in evaluating stream export models. Improvements in the modeling of landscape sources and sinks may yield important gains in prediction accuracy of regional export models, and provide insight into ways to scale up catchment fluxes more reliably. Research in this area may also lead to improvements in the ability to combine mechanistic descriptions of processes in deterministic models with statistical methods of empirically estimating flux rates at the watershed scale.

Acknowledgements

This work was initiated as part of the International SCOPE Nitrogen Project, which received support from both the Mellon Foundation and from the National Center for Ecological Analysis and Synthesis. Support was provided by the U.S. Geological Survey National Water Quality Assessment Program (NAWQA). We thank LJ Puckett and M Focazio of the USGS and two anonymous reviewers for comments on the manuscript.

References

Alexander RB, Smith RA & Schwarz GE (2000) Effect of stream channel size on the delivery of nitrogen to the Gulf of Mexico. Nature 403: 758–761

Alexander RB, Smith RA, Schwarz GE, Preston SD, Brakebill JW, Srinivasan R & Pacheco PA (2001) Atmospheric nitrogen flux from the watersheds of major estuaries of the United States: An application of the SPARROW watershed model. In: Valigura R, Alexander R, Castro M, Meyers T, Paerl H, Stacey P & Turner RE (Eds) Nitrogen Loading in Coastal Water Bodies: An Atmospheric Perspective, American Geophysical Union Monograph 57, pp 119–170

Beaulac MN & Reckhow KH (1982) An examination of land use – nutrient export relationships. Wat. Res. Bull. 18: 1013–1024

Behrendt H (1996) Inventories of point and diffuse sources and estimated nutrient loads – a comparison for different river basins in central Europe. Water Sci. Tech. 33: 99–107

Bicknell BR, Imhoff JC, Kittle JL Jr., Donigian AS Jr & Johanson RC (1997) Hydrological simulation program – fortran user's manual for release 11. U.S. Environmental Protection Agency, Environmental Research Laboratory, Athens, Georgia, U.S.A. EPA/600/R–97/080

Boyer EW, Goodale CL, Jaworski NA & Howarth RW (2002) Anthropogenic nitrogen sources and relationships to riverine nitrogen export in the northeastern U.S.A. Biogeochemistry 57/58: 137–169

Caraco NF & Cole JJ (1999) Human impact on nitrate export: An analysis using major world rivers. Ambio. 28: 167–170

Castro MS, Driscoll C, Jordan TE, Reay W, Seitzinger S, Stiles R & Cable J (2001) Assessment of the contribution made by atmospheric nitrogen deposition to the total nitrogen load to thirty-four estuaries on the Atlantic and Gulf coasts of the United States, In: Valigura R, Alexander R, Castro M, Meyers T, Paerl H, Stacey P & Turner RE (Eds)

Nitrogen Loading in Coastal Water Bodies: An Atmospheric Perspective, American Geophysical Union Monograph 57, pp 77–106
Cohn TA, DeLong LL, Gilroy EJ, Hirsch RM & Wells DK (1989) Estimating constituent loads. Wat. Resour. Res. 25: 937–942
Cunnane C (1978) Unbiased plotting positions – a review. J. Hydrology 37: 205–222
Delwiche LL & Haith DA (1983) Loading functions for predicting nutrient losses from complex watersheds. Water Resour. Bul. 19: 951–959
Fisher DC & Oppenheimer M (1991) Atmospheric nitrogen deposition and the Chesapeake Bay estuary. Ambio. 20: 102–108
Frink CR (1991) Estimating nutrient exports to estuaries. J. Environ. Qual. 20: 717–724
Galloway JN, Levy II H & Kasibhatla PS (1994) Year 2020: consequences of population growth and development on the deposition of oxidized nitrogen. Ambio. 23: 120–123
Haith DA & Shoemaker LL (1987) Generalized watershed loading functions for stream flow nutrients. Water Resources Bulletin 23: 471–478
Hill AR (1996) Nitrate removal in stream riparian zones. J. Envir. Qual. 25: 743–755
Howarth RW (1998) An assessment of human influences on fluxes of nitrogen from the terrestrial landscape to the estuaries and continental shelves of the North Atlantic Ocean. Nutrient Cycling in Agroecosystems 00: 1–11
Howarth RW, Billen G, Swaney D, Townsend A, Jaworski N, Lajtha K, Downing JA, Elmgren R, Caraco N, Jordan T, Berendse F, Freney J, Kudeyarov V, Murdoch P & Zhu Zhao-liang (1996) Regional nitrogen budgets and riverine N & P fluxes for the drainages to the North Atlantic Ocean: natural and human influences. Biogeochem. 35: 75–139
Howarth RW, Fruci JR & Sherman D (1991) Inputs of sediment and carbon to an estuarine ecosystem: Influence of land use. Ecological Applications 1: 27–39
Jaworski NA, Groffman PM, Keller AA & Prager JC (1992) A watershed nitrogen and phosphorus balance: the Upper Potomac River basin. Estuaries 15: 83–95
Jaworski NA, Howarth RW & Hetling LJ (1997) Atmospheric deposition of nitrogen oxides onto the landscape contributes to coastal eutrophication in the northeast United States. Environ. Sci. Techno. 31: 1995–2004
Johnes PJ (1996) Evaluation and management of the impact of land use change on the nitrogen and phosphorus load delivered to surface waters: the export coefficient modelling approach. J. of Hydrology 183: 323–349
Johnes PJ & Heathwaite AL (1997) Modelling the impact of land use change on water quality in agricultural catchments. Hydrological Processes 11: 269–286
Johnes PJ, Moss B & Phillips GL (1996) The determination of water quality by land use, livestock numbers and population data – testing of a model for use in conservation and water quality management. Freshwater Biology 36: 951–473
Johnson DW (1992) Nitrogen retention in forest soils. J. Envir. Qual. 21: 1–12
Jordan TE & Weller DE (1996) Human contributions to terrestrial nitrogen flux: assessing the sources and fates of anthropogenic fixed nitrogen. Bioscience 46: 655–664
Kelly CA, Rudd JWM, Hesslein RH, Schindler DW, Dillon PJ, Driscoll CT, Gherini SA & Hecky RE (1987) Prediction of biological acid neutralization in acid-sensitive lakes. Biogeochem. 3: 129–141
Lee KY, Fisher TR, Jordan TE, Correll DL & Weller DE (2000) Modeling the hydrochemistry of the Choptank River Basin using GWLF and Arc/Info: 1. Model calibration and validation. Biogeochem. 49: 143–173
Lewis WM Jr, Melack JM, McDowell WH, McClain M & Richey JE (1999) Nitrogen yields from undisturbed watersheds in the Americas. Biogeochemistry 46: 149–162

Lewis WM Jr (2002) Yield of nitrogen from minimally disturbed watersheds of the United States. Biogeochemistry 57/58: 375–385

McBride GB, Alexander RB, Elliot AH & Shankar U (2000) Regional scale modelling of water quality. Water and Atmosphere, Vol. 8 (pp 29–31). National Institute of Water and Atmospheric Research (NIWA), Auckland, New Zealand

Meybeck M (1982) Carbon, nitrogen and phosphorus transport by world rivers. Amer. J. of Sci. 282: 401–450

Molot LA & Dillon PJ (1993) Nitrogen mass balances and denitrification rates in central Ontario Lakes. Biogeochem. 20: 195–212

Montgomery DC & Peck EA (1982) Introduction to linear regression analysis. John Wiley & Sons, New York, NY

National Research Council (2000) Clean coastal waters: Understanding and reducing the effects of nutrient pollution, Ocean Studies Board and Water Science and Technology board, National Academy of Sciences, Washington, DC

Nixon SW (1995) Coastal marine eutrophication: a definition, social causes, and future concerns. Ophelia 41: 199–219

Parton WJ, Stewart JWB & Cole CV (1988) Dynamics of C, N, P, and S in grassland soils: a model. Biogeochem. 5: 109–131

Peierls BL, Caraco NF, Pace ML & Cole JJ (1991) Human influence on river nitrogen. Nature 350: 386–387

Preston SD & Brakebill JW (1999) Application of spatially referenced regression modeling for the evaluation of total nitrogen loading in the Chesapeake Bay watershed. U.S. Geological Survey Water Resources Investigations Report 99–4054

Rastetter EB, King AW, Cosby BJ, Hornberger GM, O'Neill RV & Hobbie JE (1992) Aggregating fine-scale ecological knowledge to model coarser-scale attributes of ecosystems. Ecological Apps. 2: 55–70

Sauer TJ, Alexander RB, Brahana JV & Smith RA (in press) The importance and role of watersheds in the transport of nitrogen. In: Follett, RF & JL Hatfield (Eds) Nitrogen in the Environment: Sources, Problems, and Management. Elsevier Science Publishers, the Netherlands

Seitzinger SP & C Kroeze (1998) Global distribution of nitrous oxide production and N inputs in freshwater and coastal marine ecosystems. Global Biogeochem. Cycles 12: 93–113

Seitzinger SP, Renee VS, Boyer EA, Alexander RB, Billen G, Howarth RW, Bernhard MB & van Breemen N (2002) Nitrogen retention in rivers: Model development and application to watersheds in the northeastern U.S.A. Biogeochemistry 57/58: 199–237

Smith RA, GE Schwarz & RB Alexander (1997) Regional interpretation of water-quality monitoring data. Wat. Resour. Res. 33: 2781–2798

Srinivasan R, Arnold JG, Muttiah RS, Walker D & Dyke PT (1993) Hydrologic unit modeling of the United States (HUMUS). In: Yang S (Ed) Advances in Hydro-Science and Engineering, Vol. I, Part A (pp 451–456). Washington, D.C., U.S.A.

Stacy PE, Greening HS, Kremer JN, Peterson D & Tomasko DA (2001) Contributions of atmospheric nitrogen deposition to U.S. estuaries: Summary and conclusions. In: Valigura R, Alexander R, Castro M, Meyers T, Paerl H, Stacey P & Turner RE (Eds) Nitrogen Loading in Coastal Water Bodies: An Atmospheric Perspective, American Geophysical Union Monograph 57, pp 187–226

Thomann (1972) Systems analysis and water quality management. Environmental Research and Applications, Inc., New York, NY

United States Environmental Protection Agency (1980) Design manual: Onsite wastewater treatment and disposal systems. Office of Water Programs Operations, Washington, DC, EPA 625/1-80-012

Van Breemen N, Boyer EW, Goodale CL, Jaworski NA, Paustian K, Seitzinger SP, Lajtha K, Mayer B, van Dam D, Howarth RW, Nadelhoffer KJ, Eve M, & Billen G (2002) Where did all the nitrogen go? Fate of nitrogen inputs to large watersheds in the northeastern U.S.A. Biogeochemistry 57/58: 267–293

Vitousek PM, Aber JD, Howarth RW, Likens GE, Matson PA, Schindler DW, Schlesinger WH & Tilman DG (1997) Human alteration of the global nitrogen cycle: sources and consequences. Ecological Applications 7: 737–750

Whitehead PG, Wilson EJ & Butterfield D (1998) A semi-distributed Integrated Nitrogen model for multiple source assessment in Catchments (INCA): Part I – model structure and process equations. Sci. of the Total Environ. 210/211: 547–558

Young RA, Onstad CA & Bosch DD (1995) AGNPS: An agricultural nonpoint source model. In: Singh VP (Ed) Computer Models of Watershed Hydrology. Water Resource Publications, Highlands Ranch, Colorado