Identifying the appropriate scales to model nitrogen flows from land to water

Egon Dumont

Promotoren:
Prof. dr. R. Leemans, hoogleraar Milieusysteemanalyse, Wageningen Universiteit
Prof. dr. A. Stein, hoogleraar Toegepaste Statistiek, Wageningen Universiteit

Co-promotoren:
Dr. C. Kroeze, universitair hoofddocent bij de leerstoelgroep Milieusysteemanalyse, Wageningen Universiteit
Dr. A.F. Bouwman, werkzaam bij het Milieu en Natuur Planbureau, Bilthoven
Dr. E.J. Bakker, universitair docent bij de leerstoelgroep Toegepaste Statistiek, Wageningen Universiteit

Promotiecommissie:
Prof. dr. ir. M.F.P. Bierkens, Universiteit Utrecht
Prof. dr. H.J.P. Eijsackers, Wageningen Universiteit
Prof. dr. M. Scheffer, Wageningen Universiteit
Prof. dr. J.J. Middelburg, Universiteit Utrecht & Nederlands Instituut voor Ecologie

Dit onderzoek is uitgevoerd binnen onderzoeksschool SENSE.

Identifying the appropriate scales to model nitrogen flows from land to water

Egon Dumont

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof. Dr. M. J. Kropff,
in het openbaar te verdedigen
op dinsdag 5 juni 2007
des namiddags te vier uur in de aula

Dumont, E.L., 2007

Identifying the appropriate scales to model nitrogen flows from land to water

PhD thesis Wageningen University, with summaries in English and Dutch

ISBN 90-8504-651-6

Table of contents

1 Introduction

1.1 Background

Human activities cause leaking of nitrogen (N) into streams and rivers. This has many serious consequences, including acidification and loss of biodiversity in lakes and streams [*Vitousek et al.*, 1997], and hypoxia, loss of biodiversity, and habitat degradation in coastal waters [*Rabalais et al.*, 2002]. These nitrogen problems need to be alleviated. This is difficult because determining the impacts of the sources involves complex processes with feedbacks. Also sources are often spatially and temporally separated from the impacts.

Understanding of the complexity of N related problems, estimates of future trends, and assessments of possible solutions can be achieved by proper modeling. Several models have been developed, ranging from simple empirical models to complex process based models [*Andersen et al.*, 2003]. All N-flow models have their specific limitations. One of the major weaknesses is the use of inappropriate scales. This thesis, therefore, addresses identification of appropriate modeling scales. This can assist modelers in developing better models that deal with local, regional and continental N issues in wetlands, rivers and coastal waters.

Before we can start to develop such understanding, we need first to discuss and define some important theories, concepts and approaches, which concentrate on:

- Definitions and issues related to scale (Section 1.2).
- Nitrogen, its flow and the processes involved (Section 1.3)
- Modeling and different approaches to modeling N flow (Section 1.4).

This thesis centers on selecting the appropriate scales to model N flows.

1.2 Scale issues

In the following the concept of scale is explained with two examples: a terrestrial and a river system.

<u>A terrestrial example</u>: A land surface of a large watershed is observed at a position of 1000 km above the earth surface. One might be able to distinguish between different biomes present in the watershed, like forest or grassland. When zooming in on a part of the watershed one might notice that what is recognized depends on the degree to which is zoomed in. At first, urban areas, natural land and cultivated land can probably be distinguished. When further zooming in, individual parcels can be distinguished. When increasing the degree of zooming in one could subsequently encounter: biomes, ecosystems, land use types, parcels, trees, plants, leaves, etc. This simple example illustrates that zooming in at one level can reveal structures on the earth surface that can not be recognized when zooming in at a different level. It also illustrates that the smaller structures, that can be recognized when zooming in more,

often appear in the same area where the larger structures can be recognized when zooming out.

A river system example: A network of rivers and small streams draining the same large watershed is observed from 1000 km above the earth surface. One might be able to distinguish a large lake, a delta and the largest most downstream branch of the river. When zooming in on the watershed one would probably first see that most of the watershed is covered by reaches of the river. When continuously further zooming in, more and smaller river reaches would keep appearing, until having zoomed in so far that points where springs emerge from the soil are distinguishable.

There are some similarities in the terrestrial and river system example. Both show additional detail when zooming in. The size of the smallest detail that be distinguished at a certain degree of zooming in is hereafter referred to as the **support** [consistent with *Beckie*, 2001; *Bierkens et al.*, 2000; *Heuvelink*, 1998]. On the terrestrial land surface the support can be quantified as the smallest terrestrial surface area that can still be distinguished from its surroundings. In the river system a support can be quantified as the narrowest river width that can still be distinguished. In both examples the total viewed surface area is likely to decrease when zooming in. For the terrestrial example this was the total area of land surface, and for the river system example this was the total surface water area. In both cases the total viewed area will hereafter be referred to as **extent** [consistent with *Bierkens et al.*, 2000].

There are also some differences between the two examples. When zooming in on the terrestrial land surface one can predict at what levels of zooming in structures such as ecosystems or parcels will be recognized. But it can not be predicted (without extra information about the watershed) where in the view the smaller structures will appear when zooming in. However, they will appear within structures that could be distinguished before. When looking at the river system it is possible to predict where smaller structures (river reaches) will emerge in the view when zooming in more: they will emerge upstream of the smallest reaches that can be distinguished at the present level of zooming in. Such smaller reaches that emerge when further zooming in are of a smaller **stream order** [*Strahler*, 1964].

Another important difference between the two examples is that one can always see new structures emerging when zooming in on the stream network, no matter how small the increase in level of zooming in is. When zooming in on the terrestrial land surface, however, there are levels without identifiable structure: at those levels all structures are either too small or too large. Each terrestrial structure also has a typical level of zooming at which it can be best recognized as this structure. For example, in the case of observing a tree, the view of the earth surface should be larger than the size of a tree. If the extent would become much larger than this, one would be less able to distinguish important details of the shape and surface of the tree that contribute to its recognition and ones ability to describe the tree. When the extent would become smaller than this view, one would not be able to see the complete shape of the tree. The diameter of this "optimal extent" is often referred to as **characteristic length** [*Dent and Grimm*, 1999; *Schneider*, 1994; *Torgersen et al.*, 2004].

The level of zooming can be referred to as **scale** [*Schneider*, 1994]. Looking at a large watershed from a certain distance is a metaphor for the research that is done on watersheds in which information is gathered not only by looking but also by measuring things that cannot be readily seen. Measurements are done on different scales: When measuring river mouth discharge (by a gauge) or the surface area of desert (by remote sensing), one is hardly zooming in. Therefore these measurements are referred to as **coarse scale measurements** When measuring N concentration in a brook or denitrification potential of soil columns we refer to **fine scale measurements** [*Gibson et al.*, 2000; *Lijklema*, 1998].

If zooming in would not be possible, but if one would still want to have an idea about what can be seen at finer scales of the large watershed, one needs to resort to **downscaling** [*Bierkens et al.*, 2000; *Habersack*, 2000], which means that one uses what is seen on a coarse scale (e.g. 1000km distance) and knowledge about the relation between coarse and fine scale watershed properties to estimate what can be seen at finer scales in this particular watershed. The opposite of downscaling is **upscaling** [*Bierkens et al.*, 2000; *Habersack*, 2000]. One would upscale if one would walk on the watershed surface and look around to observe fine scale properties and if these fine scale observations would be combined with general knowledge on the relation between fine scale and coarse scale watershed properties to estimate what can be seen at the coarse scale. The terms upscaling and downscaling also apply to properties that can not be seen but are detected otherwise, such as or soil nitrogen content or nitrogen uptake. The issue of scale, upscaling and downscaling is also frequently reported in the geostatistical literature [see *Bierkens et al.*, 2000; *Stein et al.*, 2001]. This becomes relevant when a data analysis is to be carried out.

What can be observed at a coarse scale is in fact an aggregate of causes that can be observed at finer scales [*Desbarats*, 1991; *McBratney*, 1998]. This also applies to nitrogen flow within large watersheds. Nitrogen compounds generally flow in the direction of the outlet of a watershed. This flow is partly generated by N dissolving in water that subsequently flows in the drainage direction. But in addition to drainage of water, many fine scale biological and chemical processes affect flow of N molecules.

Coarse scale processes generally vary slower (i.e. over coarse time scales) because these processes move within larger areas or volumes, but may have speeds of the same magnitude as fine scale processes (e.g. [*Stalnacke et al.*, 2004] versus [*Durand et al.*, 1994]). The reason is that the per unit area available energy (ultimately from solar radiation) for a coarse scale process is usually of the same magnitude as that for fine scale processes. An example is aquifer N concentration, which varies much slower and over larger areas and volumes than soil N concentration, driven by quickly varying processes such as plant growth and lateral runoff occurring in smaller areas.

The previous discussion has focused much on concepts related to spatial scale of detected phenomena. These concepts also apply to temporal scale. More importantly they also apply to the scale of modeling. Similar to the scale of detection, modeling can take place on coarser and finer scales. A given set of model equations, however, can only reliably describe processes on a limited scale range. In order to describe these processes they need to use input of an appropriate scale [*Caraco et al.*, 2003; *Costanzo et al.*, 2003; *De Wit*, 1999; *Dent and Grimm*, 1999; *Jarvie et al.*, 1999; *Jarvis*, 1995; *Jordan et al.*, 2003; *McClain et al.*, 2003; *Meybeck*, 2002; *Quinn*, 2004; *Schneider*, 1994; *Seitzinger et al.*, 2006; *Seitzinger et al.*, 2002; *Torgersen et al.*, 2004; *Van Herpe et al.*, 2002; *Wagenet*, 1998; *Wolfert*, 2001]. We will call the scale of used input the modeling scale. Similarly to detection scale, modeling scale can be characterized by support and extent. Model equations can also be upscaled or downscaled. This means that the scales where they are reliable are shifted to coarser or finer scales, respectively.

Apart from reliable process description, several other factors affect the appropriate modeling scale. These can be clustered into three groups:

(1) Availability of useful data [*Eckhardt et al.*, 2003; *Joao*, 2002].

(2) Required effort [*Becker and Braun*, 1999; *Bellamy and Loveland*, 2001b; *Beven*, 1995; *Bierkens et al.*, 2000; *Hassanizadeh and Gray*, 1979; *Li et al.*, 1999; *Schneider*, 1994; *Sposito*, 1998].

(3) Usefulness of model predictions [*Meybeck et al.*, in prep.; *Omernik*, 2003; *Omernik and Bailey*, 1997; *Sherman*, 1991].

1.3 Nitrogen flow from land to water

Nitrogen (N) inputs to streams and rivers due to human activities, cause problems. N-flow models may help to understand them and develop solutions for them. Like other models, N-flow models have appropriate scales. Before we focus on N-flow models, we will first introduce N, N flow, and the problems that it causes.

Nitrogen compounds in nature can be divided into two groups: non-reactive and reactive. The most important non-reactive form of N is dinitrogen (N_2), the main constituent of the atmosphere. Reactive N includes all biologically, photochemically, and radiatively active N compounds in the atmosphere and biosphere. Thus, reactive N includes inorganic reduced forms of N such as ammonia (NH_3) and ammonium (NH_4^+), inorganic oxidized forms such as nitrate (NO_3^-) and nitrite (NO_2^-), and organic forms (e.g., urea, amines, proteins, and nucleic acids). Nitrogen forms in aquatic systems can be further subdivided in dissolved N (DN) and particulate N (PN). DN consists of dissolved inorganic N (DIN) and dissolved organic N (DON) [*Seitzinger et al.*, 2005].

The availability of reactive nitrogen in ecosystems under natural conditions is affected by inputs of reactive N (e.g. N_2 fixation, atmospheric deposition), N accumulation, and losses of reactive N (e.g. denitrification, and transport of reactive N out of the system). N_2 fixation is the conversion of N_2 to organic N [*Galloway et al.*, 2004]. It can be mediated by lightning, by some blue-green algae, by free living soil bacteria, and by soil bacteria that live enclosed in nodules in the roots of certain leguminous plants [*Vitousek et al.*, 2002]. N losses include denitrification and transport of N out of the system. Denitrification is the conversion of NO_3^- to N_2 or N_2O. Denitrification is largely carried out by bacteria in anoxic conditions [*Zumft*, 1997]. Within ecosystems, certain transformations between forms of reactive nitrogen occur. In oxic conditions, a conversion of NH_4^+ to NO_2^- and NO_3^- can be performed primarily by soil-living bacteria (nitrification). Uptake is the transformation of NO_3^- and NH_4^+ into living organic N. Organic N can be decomposed to NH_4^+ (ammonification). In this thesis, we will refer to reactive N in the biosphere as N.

The current worldwide production of N by humans exceeds the natural N_2 fixation [Galloway et al. 2003]. The global increase in N production has three main causes: (1)

increased cultivation of legumes, rice, and other crops that promote conversion of N_2 to organic N through biological N_2 fixation; (2) combustion of fossil fuels, which converts both atmospheric N_2 and fossil N to reactive nitrogen oxides (NO_x); and (3) the Haber-Bosch process, which converts non-reactive N_2 to reactive NH_3 to sustain food production and some industrial activities [*Smil*, 2002].

A worrisome consequence is the increased availability of N in the biosphere, and increased transport through the environment, leading to accumulation of N in, for example, aquatic systems. This increase is expected to continue [*Millennium Ecosystem Assessment*, 2005]. The inputs of N in many aquatic systems exceed rates of N removal through denitrification to non-reactive N_2 [e.g. *Quynh et al.*, 2005; *Stalnacke et al.*, 2004; *Turner and Rabalais*, 1991]. N in aquatic systems is responsible (together with sulfur) for acidification and loss of biodiversity in lakes and streams in many regions of the world [Vitousek et al. 1997]. Further, N is responsible for eutrophication, hypoxia, loss of biodiversity, and habitat degradation in coastal ecosystems. It is considered one of the largest pollution problems in coastal waters [e.g., Howarth et al. 2000, NRC 2000, Rabalais 2002].

1.4 Modeling of nitrogen flow

To anticipate undesired effects of N in aquatic systems it is necessary to model the effects of different scenarios and mitigation options on the flow of N from terrestrial systems to aquatic systems [*Thoman and Linker*, 1998]. Models are used in two general ways to asses N flow to aquatic systems [*Andersen et al.*, 2003]. First they can be used to predict flows, concentrations, and dynamics of different forms of N in aquatic systems where these quantities are not measured. Second they can be used to predict future conditions with altered nutrient management strategies, land use changes, water discharge regulations or waste water treatment [*Meybeck*, 2002]. Modeled nutrient management strategies can involve changes in the amount and timings of N input by application of fertilizer and manure. Modeled land use changes could be change from agriculture to nature, or vice versa. Modeled water discharge regulations could be constructions in surface waters to control high water levels, for water conservation or for rewetting areas.

Several types of models quantifying N flow between land and water on a regional to global scale have been developed during the last decade [*Shoemaker et al.*, 1997]. These models were intended for different purposes. As a result, they differ in their complexity, their resolution in time and space, and they need data with differing levels of detail [*Andersen et al.*, 2003]. Also models quantifying N flow between land and water are able to describe either parts or all of the dominant processes that govern nutrient cycling at catchment scale [*Borah and Bera*, 2004]. Models quantifying N flow between land and water are usually only appropriate for a specific region [*Andersen et al.*, 2004]. Most models have only been applied to a specific part of the world, which means that they may not be able to handle the gradient in climate (e.g. frozen soils), hydrology (e.g. shallow groundwater), land use and/or agricultural practices existing in other parts of the world. Problems with the acquisition of input data to the different models can also severely limit their application to particular parts of the world [*Bellamy and Loveland*, 2001a].

We define a model as a set of equations where each individual equation is coupled to at least one other equation by input that it receives from or provides to that other equation. Models for land to water N flow can be subdivided in a number of categories, based on the characteristics of the most important model parts. We may distinguish between: process based models, conceptual models, and empirical models (Box 1.1). Process based models normally require large amounts of input data at a very small temporal and spatial support. In many cases, such detailed data may not be available, at least not at the larger extent, requiring some assumptions or default values to be made, or transfer functions to be developed. Empirical and conceptual models may in such cases be viable alternatives. Even in this category there is a large variability in complexity [e.g. *Howarth et al.*, 1996; *Smith et al.*, 1997a]. Many empirical and conceptual models, however, have as a limitation that they may not be able to describe the dynamics in the flows. This trade-off between the complexity and applicability of these two approaches has been discussed by several authors [e.g. *De Vries*, 1994].

Process based, conceptual and empirical model parts can be further categorized according to the following categories: steady state or dynamic; geographically explicit or lumped; and used to model land, water and/or the interaction between both.

The suitability of a specific type of model depends on the purpose of that model. Modeling effects of historical trends or accumulation of N requires dynamic modeling. Process based models are generally appropriate for analyses of effects of processes, or if a particularly robust model is required that is valid in any type of scenario. Empirical and conceptual models are generally suitable for analyses of regions or scenarios in which conditions are similar to the conditions for which the model was derived [e.g. *Smith et al.*, 1997b]. Geographically explicit models are generally appropriate if the model needs to describe spatial dispersion of N and the relative location of critical sources and sinks of N [e.g. *Beaujouan et al.*, 2001]. Modeling processes known from experiments in for example ponds, fields or brooks may require process based and geographically explicit modeling [e.g. *Neitsch et al.*, 2002]. For modeling processes known from sub-watershed scale models, conceptual and geographically explicit modeling is appropriate.

Box 1.1 Overview of three types of models: process based, conceptual, and empirical

Process based models

Process based models describe causal relations [e.g. *Billen et al.*, 1994; *Neitsch et al.*, 2002; *Wolf et al.*, 2003]. Developers of process based approaches need to be aware of process scales. Often different processes driving N flow on different scales are correlated to each other. Such correlation is often due to the fact that these processes are ultimately affected by the same variable [*Habersack*, 2000]. For example, many processes that drive N flow which occur in both large and small areas (and in both large and small river reaches) are influenced by variations in temperature. It is, however, not temperature that causes the variations in N flow, but the processes that are influenced by temperature. These processes are different on different scales, i.e. they occur over different surface areas and in river reaches of different stream orders.

Conceptual models

A conceptual model is a representation of the most current understanding of the major relationships in a particular environment. However these relationships do not need to be exclusively direct causal relationships. Conceptual models are usually to some degree based on expert judgment or other existing conceptual models. They are commonly used for modeling coarse scale systems [e.g. *Anthony et al.*, 1996; *Boyer et al.*, 2002; *Green et al.*, 2004; *Howarth et al.*, 1996; *Peierls et al.*, 1991; *Seitzinger and Kroeze*, 1998; *Smith et al.*, 1997a]. Conceptual models for coarse scale systems are developed at the coarse scale using coarse scale data such as remote sensing data and governmental administrative databases. An important advantage of conceptual models is that they can account for emergent properties at regional and global scales.

Empirical models

Empirical models are statistically derived by combining those variables in a model that are correlated with the output [e.g. *Grimvall and Stålnacke*, 1996; *Johnes et al.*, 1998]. The predictive capacity of empirical models relies on correlation of model terms and N flow that holds when the model is applied at different locations or times.

Hereafter we will refer to a model of N flow between land and water as “N-flow model”. N-flow models often have the possibility to use input of different scales. We will refer to the scale of model input as modeling scale. An N-flow model may be applied on different modeling scales. Existing N-flow models vary widely in their modeling scale. A distinction can be made between N-flow models that describe spatial detail (spatially explicit), N-flow models that can describe temporal detail (temporally explicit) and those that are both spatially and temporally explicit.

Process based N-flow models are often both spatially and temporally explicit [e.g. *Billen et al.*, 1994; *Neitsch et al.*, 2002; *Wolf et al.*, 2003]. They usually distinguish individual upstream sub-watersheds and river reaches and commonly use time steps as small as an hour.

Some N-flow models are temporally explicit but not spatially explicit. An example is a conceptual N-flow model by Ruiz et al. [2002b] of a 12 km^2 catchment to investigate NO_3^- concentrations in its stream water. Model calculations were lumped over the whole catchment, because the spatial behavior of N flow was regarded as being unknown. Water drainage and N leaching were modeled on a daily basis. The reason for the daily basis was that more than 90% of the NO_3^- is exported with base flow which can be well represented with such a temporal modeling scale.

Finally, several N-flow models are spatially explicit but not temporally explicit. For example, a conceptual N-flow model by Smith et al. [1997a] was aimed solely at producing spatially explicit information on N (and P) flows, being maps of (1) N concentrations in relatively large river reaches occurring in the United States, and of (2) sources and processes involved in N transport to these reaches. Model calculations were made for individual river reaches and individual sub-watersheds draining to these reaches.

For many N-flow models, the chosen modeling scales are not well justified in the contemporary literature. Reasons for using different scales for different model parts are often not reported. In addition, usually not all factors affecting the appropriateness of a chosen model scale are addressed. Finally, there is usually uncertainty about the appropriate scale of N-flow models [*Kroeze et al.*, 2003]

One of the reported reasons for the choice of spatial support is that the resulting spatial units of model calculation (spatial support units) have relatively homogeneous properties. This is preferred because modeled processes can then be assumed constant within such spatial support units. Examples are areas with relatively homogeneous land use and soil [e.g. *van Griensven and Brauwens*, 2001], climate and geology [e.g. *Narula et al.*, 2003], or watershed parts with relatively homogeneous intensity of erosion (upstream), transport and sedimentation (downstream) of PN [e.g. *Hollander et al.*, 2006].

Another reported reason for a chosen spatial model scale is that the modeler is specifically interested in inflow, outflow or storage of N in specific compartments in the modeled system. The model scale then must be fine enough to enable the model to distinguish these compartments. Examples of such compartments are dammed reservoirs [e.g. *Charles and Berrien*, 1991], valuable ecosystems [*Hunsaker and Levine*], or important agricultural N sources [e.g. *Merete and Børge*, 1993; *Van Herpe et al.*, 2002]

Most model parts of contemporary N-flow models have scales in line with available input data. Availability of input data, however, is not the only factor that affects appropriateness of modeling scale (Section 1.2). Further, current geostatistical methods [e.g. *Burrough*, 1981; *Lajaunie and Wackernagel*, 2000] can only be used to identify appropriate modeling scales if data are available that are suitable for this purpose. Currently there is no guidance in transparently and comprehensively assessing all factors affecting appropriateness of scales of important model parts. Therefore guidance is required in transparently and comprehensively identifying appropriate temporal and spatial modeling scales of important model parts.

1.5 Research objectives

The objective of this thesis is to increase the ability to identify appropriate spatial and temporal scales for N-flow models in a transparent and comprehensive way. The main focus will be on coarse scale models of one or more watersheds. In order to meet the objective, the following sub-objectives will be addressed:

I. To develop and apply a model for global N flows from land to water in a spatially explicit way.

This model serves as an illustrative example of a model for which comprehensive identification of appropriate modeling scale is needed.

II. To develop a framework for identifying the appropriate spatial and temporal scales of N-flow models.

III To apply this framework to a model of global N flow.

The framework is developed for coarse scale models of one or more watersheds. We also aim to assess the applicability of this framework to models of smaller systems. Therefore, the fourth objective is:

IV To assess the applicability of this framework to models of N flow between floodplains and rivers.

These sub-objectives will be obtained as follows:

First, models of land-to-water N flow in different systems are studied. Based on this, a spatially explicit model of nitrogen flow from land to coastal waters of the world is developed and applied (sub-objective I). This model is described in Chapter 2.

Second, a framework for identification of appropriate spatial and temporal scales of N-flow models (FAMOS) is developed. (sub-objective II, Chapter 3). It is developed to be transparent and comprehensive. This requires the possibility to consider all important factors without dependence on availability of suitable data. For the latter reason quantitative analyses of spatial variation using contemporary geostatistics is not included in FAMOS. FAMOS is applied in two case studies. First, FAMOS is applied to the global model described in Chapter 2 (sub-objective III, Chapter 4). In the second case study, the applicability of FAMOS is investigated for model equations specifically describing N exchange between rivers and floodplains. The applicability of FAMOS to take into account the scale dependent validity of such equations is studied (sub-objective IV, Chapter 5).

Finally, the results are discussed and conclusions drawn about the appropriate modeling scales of N-flow models and how these can be identified using the approach developed in this thesis (Chapter 6).

2 Global distribution and sources of dissolved inorganic nitrogen export to the coastal zone: Results from a spatially explicit, global model[a]

Abstract

Here we describe, test, and apply a spatially explicit, global model for predicting dissolved inorganic nitrogen (DIN) export by rivers to coastal waters (NEWS-DIN). NEWS-DIN was developed as part of an internally consistent suite of global nutrient export models. Modeled and measured DIN export values agree well (calibration R^2 =0.79), and NEWS-DIN is relatively free of bias. NEWS-DIN predicts: DIN yields ranging from 0.0004 to 5217 kg N km^{-2} y^{-1} with the highest DIN yields occurring in Europe and South East Asia; global DIN export to coastal waters of 25 Tg N y^{-1}, with 16 Tg N y^{-1} from anthropogenic sources; biological N_2 fixation is the dominant source of exported DIN; globally, and on every continent except Africa, N fertilizer is the largest anthropogenic source of DIN export to coastal waters.

[a] **This chapter has been published as:**
Dumont, E., J. A. Harrison, C. Kroeze, E. J. Bakker, and S. P. Seitzinger (2005), Global distribution and sources of dissolved inorganic nitrogen export to the coastal zone: Results from a spatially explicit, global model, Global Biogeochemical Cycles, 19, GB4S02, doi:10.1029/2005GB002488.

2.1 Introduction

In many watersheds, nutrient delivery by rivers to coastal waters has been increasing as a result of human activities such as increased use of artificial fertilizer in agriculture, population growth and cultivation of legumes [e.g. *Carpenter et al.*, 1998; *Galloway et al.*, 2004; *Vitousek et al.*, 1997]. This increased nutrient delivery has been blamed for the formation and increased extent of coastal "dead zones", as well as loss of seagrass habitat, decreases in coastal biodiversity, and increased frequency and severity of harmful and nuisance algae blooms [*Diaz et al.*, 2003]. One of the most important nutrients in this respect is nitrogen (N), because nitrogen is often the most limiting nutrient [*Justic et al.*, 1995; *Turner and Rabalais*, 1994; *Vince and Valiela*, 1973]. Dissolved inorganic nitrogen (DIN) is often the most abundant and bioavailable form of nitrogen, and therefore contributes significantly to coastal eutrophication [*Veuger et al.*, 2004].

To date, three models have been applied at the global scale to model river DIN export in a spatially explicit manner [*Green et al.*, 2004; *Seitzinger and Kroeze*, 1998; *Smith et al.*, 2003]. These models have greatly improved our understanding of global patterns and magnitudes of DIN export. Smith et al. [2003] presented an empirical multiple regression model predicting DIN export simply as a function of runoff and population density. Seitzinger and Kroeze [1998] developed a river DIN export model at 1 x 1° resolution that included some DIN sources and sinks. However, important DIN sources such as N_2 fixation and manure were not included in this model. More recently, Green et al. [2004] have developed a model at the 0.5 x 0.5° scale that includes a wider range of input parameters as well as loss terms. However, like the other two models, the model of Green et al. [2004] was calibrated using data from multiple decades, leading to large differences between measurement times of inputs and outputs of these models. Furthermore, none of the three have been validated with global data not used in model calibration. Finally, none of these models were developed in a way that allows for easy comparison of output with predicted export of other nutrients or nutrient forms.

Here we describe, evaluate, and apply a spatially explicit, global DIN export model called NEWS-DIN. NEWS-DIN was developed as part of a multi-investigator effort to model river export of multiple bioactive elements and element forms (particulate/dissolved, organic/inorganic) called Global Nutrient Export from WaterSheds (Global NEWS; see other papers in this issue [*Beusen et al.*, 2005; *Harrison et al.*, 2005a; *Harrison et al.*, 2005b]). To the extent possible, models developed within Global NEWS use the same input data (e.g. population, runoff, and basin delineation data), are calibrated or developed using data for the 1990s, have the same resolution, and have the potential to run scenarios for future nutrient export to coastal waters. Eventually, we hope to use NEWS-DIN, in conjunction with other NEWS models to compare global spatial patterns of export of different nutrients and nutrient forms. To date, this has not been possible due to the disparate input data requirements and structures of existing global nutrient export models. Here, we use NEWS-DIN to estimate global DIN export for 1995 in a spatially explicit manner, and indicate dominant sources and sinks of DIN export.

2.2 Methodology

Building on past work by Caraco and Cole [1999] and Seitzinger and Kroeze [1998], we developed a new model for predicting DIN export from watersheds (NEWS-DIN). NEWS-DIN includes several input variables missing from the original model reported by Seitzinger and Kroeze [1998] (N-model) such as manure N and biological N_2 fixation. NEWS-DIN also includes retention and loss terms that were absent from the original N-model, including N retention in river networks, N retention in dammed reservoirs, N loss via consumptive water use, and N loss via harvesting and grazing. NEWS-DIN also includes a more sophisticated treatment of sewage point sources than was included in the original N-model, incorporating estimates of sewage treatment, sewage connectivity, and variable N-excretion rates. The spatial resolution of input data and basin delineations was increased from 1 x 1° to 0.5 x 0.5° grids. We also used an enhanced dataset for calibration and validation; the model has now been calibrated primarily for one specific period of time: 1990 to 1997, rather than for data spanning over three decades of measurements. This restriction gives a better temporal match between input data and DIN export data used in calibration. Finally, we use more river basins for calibration than was used in formulation of the original N-model (61 vs. 35 basins).

2.2.1 Model Form and Input Data

NEWS-DIN

NEWS-DIN can be summarized as:

$$DIN = FE_{riv} \cdot [DIN_{sew} + (FE_{ws} \cdot TN_{diff})] \qquad (1)$$

where DIN is modeled DIN yield per river basin (kg N km^{-2} y^{-1}), DIN_{sew} is DIN from sewage point sources (kg N km^{-2} y^{-1}) and TN_{diff} is total nitrogen (TN) from diffuse sources that is mobilized from the watershed soils and sediments (kg N km^{-2} y^{-1}). FE_{riv} is a river export fraction representing the fraction (0-1) of total point and diffuse DIN inputs to the river that is exported as DIN, FE_{ws} is a watershed export fraction representing the fraction (0-1) of TN from diffuse sources in the watershed that leaches to rivers as DIN (Appendix 2.2). To estimate total DIN export per basin (kg N basin^{-1} y^{-1}) we multiplied DIN yield by basin area (km^2, derived from Vörösmarty et al. [2000a; 2000b]).

Sub-models and input data used to calculate the terms in Equation 1 are described in Sections 2.1.2. to 2.1.6. Most input data (Table 2.1) were available as a 0.5 x 0.5° grid and were selected to be representative of the years between 1990 and 1997. Before using this gridded data in the model it was averaged over river basins delineated from an updated version of the 0.5 x 0.5° STN30-p global river network [*Vörösmarty et al.*, 2000a; *Vörösmarty et al.*, 2000b], unless stated otherwise. Many input data sources, such as those for land use, basin delineations, runoff, discharge, population, dam properties, fertilizer, manure, harvesting losses, and sewage are the same as those used for other Global NEWS models (see other papers in this issue) including NEWS-DOC, NEWS-DON, NEWS-DOP [*Harrison et al.*, 2005b], NEWS-DIP [*Harrison et al.*, 2005a] and a particulate export model by Beusen et al. [2005].

Table 2.1 Overview of data that is used as inputs and for calibration of NEWS-DIN, including their resolution and source.

Data	Resolution	Source
Land use	0.5°	Bouwman et al. [2005b]
Stream network and basin delineations	0.5°	STN30 [*Vörösmarty et al.*, 2000b]
Runoff	0.5°	Fekete et al. [2000]
Discharge	river basin	USGS [*Alexander et al.*, 1996], European Environmental Agency [*EEA*, 1998], and Meybeck & Ragu [1995]
Population data	Country	United Nations [1998] FAO [2001] and WorldBank [2001]
NO_3^--N and NH_4-N concentration	river basin	USGS [*Alexander et al.*, 1996], the European Environmental Agency, and Meybeck & Ragu [1995]
Anthropogenic river water removal	river basin	Dynesius and Nilsson [1994]
Dam properties	subbasin	Vörösmarty et al. [2003]
Manure N addition	0.5°	Bouwman et al. [2005b]
Fertilizer N addition	0.5°	Bouwman et al. [2005b]
Biological N_2 fixation	0.5°	Green et al. [2004]
Atmospheric NO_3^--N deposition	5 x 3.75° interpolated to 0.5°	Dentener [2004]
Crop N export	0.5°	Bouwman et al. [2005b]
Sewage point sources	Country	Bouwman et al. [2005a]

Aquatic retention

Aquatic retention ($1\text{-}FE_{riv}$) is the retention of DIN (0-1) in reservoirs and in the STN-30 basin river network. The fraction of river DIN inputs exported to the coastal zone as DIN, FE_{riv}, is defined as:

$$FE_{riv} = (1\text{-}L_{den})\cdot(1\text{-}Q_{rem})\cdot(1\text{-}D) \tag{2}$$

where ($1\text{-}L_{den}$) is the fraction of DIN not retained in river reaches, ($1\text{-}D$) is the fraction of DIN not retained in dammed reservoirs, and ($1\text{-}Q_{rem}$) is the fraction that is not diverted to other basins or removed for irrigation.

DIN loss by denitrification during transport throughout entire river networks (L_{den}) is estimated according to Seitzinger et al. [2002] as:

$$L_{den} = c \cdot \ln(A) - d \tag{3}$$

where A is basin area (km^2) from Vörösmarty et al. [2000a; 2000b] and c and d are fitted coefficients equal to 0.0605 and 0.0443, respectively (r^2= 0.88), using data on A and modeled L_{den} based on 16 rivers in the north-eastern U.S. as in Seitzinger et al. [2002]. Estimation of c and d was done using rivers with an L_{den} smaller than 0.65. Therefore the maximum L_{den} was set at 0.65 to avoid extrapolation error.

The impact of anthropogenic removal of river water (containing DIN) was estimated as:

$$Q_{rem} = \frac{Q_{irr} + Q_{div}}{Q_{nat}} \tag{4}$$

where Q_{rem} is the fraction of DIN retained, due to the anthropogenic removal of river water (containing DIN), Q_{div} is the amount of discharged water lost from the river by anthropogenic transfer of water out of the basin ($km^3\ y^{-1}$), Q_{irr} is the amount of discharge removed for irrigation, minus the amount of irrigation water that ultimately flows back into the river (i.e. Q_{irr} is extracted irrigation water that evaporates on irrigated fields) ($km^3\ y^{-1}$), and Q_{nat} is the river discharge before any direct human manipulations on the river system ($km^3\ y^{-1}$). Values for Q_{irr}, Q_{div}, and Q_{nat} were from Dynesius and Nilsson [1994]. If these values were unavailable then Q_{rem} was assumed to equal 0.039, the average of Q_{rem} for the 115 basins where Q_{irr} and Q_{div} were available.

DIN retention in reservoirs (D) within a river basin is modeled according to Seitzinger et al. [2002] as:

$$D = \frac{1}{Q}\cdot\sum_{i=1}^{n} 0.8845\cdot\left(\frac{DEPT_i}{Rt_i}\right)^{-0.3677}\cdot Q_i \tag{5}$$

where $DEPT_i$ is reservoir depth (m), Rt_i is water residence time (y), and i is the reservoir identification number within a basin ({i=1….i=n}). The retention in each reservoir of a river basin is aggregated to a basin average (D) by taking a weighted arithmetic average of the retention in reservoirs. The modeled retention in a reservoir is weighted by the fraction of total basin discharge that the reservoir intercepts. Q is total basin discharge and Q_i is the discharge intercepted by dam i. Maximum D is set at 0.965, because that is the maximum D in the set of basins used for calibration. Per reservoir values for $DEPT_i$, Rt_i, and Q_i were from Vörösmarty et al. [2003].

Point sources

DIN input into rivers from point sources is estimated by first estimating the amount of TN from human excreta and industrial wastewater in sewage effluents, and by subsequently multiplying this amount with an estimated fraction of TN that is DIN in sewage effluents.

TN in sewage effluents is estimated as described by Bouwman et al. [2005a]:

$$TN_{sew} = H \cdot (1 - T_N) \cdot I \cdot E_N \qquad (6)$$

where TN_{sew} is the annual sewage TN discharged to surface water per km^2 of basin (kg N km^{-2} y^{-1}), H is population density (individuals km^{-2}) from United Nations [1998], FAO [2001] and WorldBank [2001], T_N is a country by country fraction of TN removed by wastewater treatment compiled by Bouwman et al. [2005a], I is a country by country fraction of population connected to sewer systems compiled by Bouwman et al. [2005a], E_N is the per capita human TN emission estimated as described by Bouwman et al. [2005a] (kg N individual^{-1} y^{-1}). For those countries where I and T_N were not available on a country by country basis, we used regional estimates according to Bouwman et al. [2005a].

Sewage effluent DIN emitted to rivers is estimated as:

$$DIN_{sew} = TN_{sew} \cdot \left[0.485 + \frac{T_N}{\max(T_N)} \cdot 0.255 \right] \qquad (7)$$

where DIN_{sew} is DIN in sewage effluents for a basin (kg N km^{-2} y^{-1}), *max* is the maximum of all countries for which T_N is known (max(T_N) = 0.8 in Finland), 0.485 is an estimate of the fraction of TN that is DIN in sewage effluent [*Seitzinger,* 1995] without treatment, and 0.255 is the maximum increase in DIN to TN ratio that can be achieved by sewage treatment [*Seitzinger,* 1995].

Watershed export

The watershed export fraction, FE_{ws}, in Equation 1 is the fraction of TN from diffuse sources that is transported as DIN from soils to rivers. In NEWS-DIN, FE_{ws} is defined as;

$$FE_{ws} = e \cdot R \qquad (8)$$

Where *R* is runoff (m y^{-1}) and, *e*, the watershed export coefficient, is a calibrated parameter that defines the slope of the assumed linear relationship between *R* and FE_{ws}. Values for *R* were obtained from Fekete et al. [2000]. FE_{ws} was not allowed to exceed a value of 1.

This way of defining transport of DIN from soils to rivers is consistent with Caraco and Cole [1999] who used a similar relationship for transport of NO_3^- from soils to rivers. Runoff has been found to be an important predictor of DIN transport through soils, brooks and small river reaches [*Behrendt and Opitz*, 2000; *Goolsby et al.*, 2000].

Diffuse sources

The total amount of diffuse source nitrogen that is mobilized annually (TN_{diff}) (kg N km^{-2} y^{-1}) is estimated in NEWS-DIN as:

$$TN_{diff} = TN_{am} + TN_{fe} + TN_{dep} + TN_{fix} - TN_{exp} \qquad (9)$$

where TN_{am} is animal manure N addition (kg N km^{-2} y^{-1}), TN_{fe} is fertilizer N addition (kg N km^{-2} y^{-1}), TN_{dep} is atmospheric NO_y deposition (kg N km^{-2} y^{-1}), TN_{fix} is biological N_2 fixation (kg N km^{-2} y^{-1}) and TN_{exp} is N in crops and grassland that is removed from the land by harvesting and grazing (kg N km^{-2} y^{-1}). TN_{am}, TN_{fe} and TN_{exp} were calculated as in Bouwman et al. [2005b]. TN_{dep} is annual average deposition of atmospheric NO_y-N modeled for 1995 [*Lelieveld and Dentener*, 2000]. TN_{fix} includes both natural and agricultural biological N_2 fixation. Biological N_2 fixation data were obtained from Green et al. [2004]. Subtraction of TN_{exp} in Equation 9 makes it so that TN_{diff} represents only the fraction of N from diffuse sources that is directly available for leaching to surface water, thereby avoiding double counting of N inputs. For example, in Equation 9 we do not count fertilizer N that is removed by harvesting or grazing, and that is subsequently available for leaching to surface water after reapplication as animal manure or as point source from human sewage.

Equation 9 is separately applied to each 0.5 x 0.5° grid cell. In individual cases, calculated TN_{diff} may be negative because crop and grazing N export exceeds N additions. However, we set negative TN_{diff} values to zero, because there is generally not a net flux of N from rivers to agricultural fields. Subsequently TN_{diff} for each grid cell was averaged over the basin and this average TN_{diff} was used in Equation 1.

In order to avoid double counting of deposited NH_x that has volatilized from manure produced in the same year and basin, we did not include NH_x deposition in Equation 9.

Source contributions

In evaluating model output, we separately estimated the contributions of atmospheric NO_y deposition, fertilizer addition, manure addition and agricultural and non-agricultural biological N_2 fixation to DIN export. This was done by removing the point sources term (DIN_{sew}) from Equation 1 and by replacing TN_{diff} with the amount of N mobilized from atmospheric NO_y-N deposition, fertilizer N addition, manure N addition, agricultural N_2 fixation or non-agricultural N_2 fixation, respectively. The contribution to exported DIN from manure, fertilizer and agricultural N_2 fixation inputs was obtained by multiplying the amount of N from each of these sources by the fraction remaining after harvest and grazing. This fraction (*G*) was calculated as:

$$G = 1 - \frac{TN_{\exp}}{TN_{am} + TN_{fe} + (0.23 \cdot TN_{fix})} \qquad (10)$$

where 0.23 is an estimate of the fraction of TN_{fix} occurring on agricultural land, calculated based on Galloway et al. [2004]. We assumed that NO_y deposition and natural N_2 fixation were not subject to removal of N by harvesting or grazing.

We also estimated the contribution of point sources to model output by removal of TN_{diff} from Equation 1.

Source contributions are either expressed in Tg N y^{-1} or as a fraction of model output, where the source with the highest fraction is referred to as the dominant source.

Measured DIN export from watersheds

Discharge, DIN concentration, and basin surface area data for 61 basins were compiled from several sources for use in model calibration and validation (Appendix 2.1). Together, these 61 basins account for 33 % of global exoreic discharge. We restricted our dataset to include only long term (>4 y) annual averages with at least 85% of the measurements taken between 1990 and 1997. These annual averages were obtained from Meybeck and Ragu [1995] (median), EEA [1998] (median), and USGS [2003] (arithmetic average). We also excluded basins encompassing fewer than 10 grid cells due to uncertainties associated with smaller basins [*Harrison et al.*, 2005a; *Vörösmarty et al.*, 2000b]. Despite these requirements for inclusion, both calibration and validation datasets include basins with a broad range of sizes, land uses, climates, topographies, and ecosystems (Figure 2.1). Data on DIN yield have an inter-annual variation ranging from a factor 2 to 13 [*USGS*, 2003].

Measured DIN yield was obtained as follows:

$$DIN_{meas} = ([NO_3^-] + [NH_4^+]) \cdot \frac{Q}{A} \qquad (11)$$

DIN_{meas} is DIN yield near the river mouth (kg N km^{-2} y^{-1}). $[NO_3^-]$ and $[NH_4^+]$ are measured NO_3^--N and NH^{4+}-N concentrations, respectively, measured near the river mouth (kg N km^{-3}). *Q* is measured basin discharge (km^3 y^{-1}) and *A* is basin area (km^2).

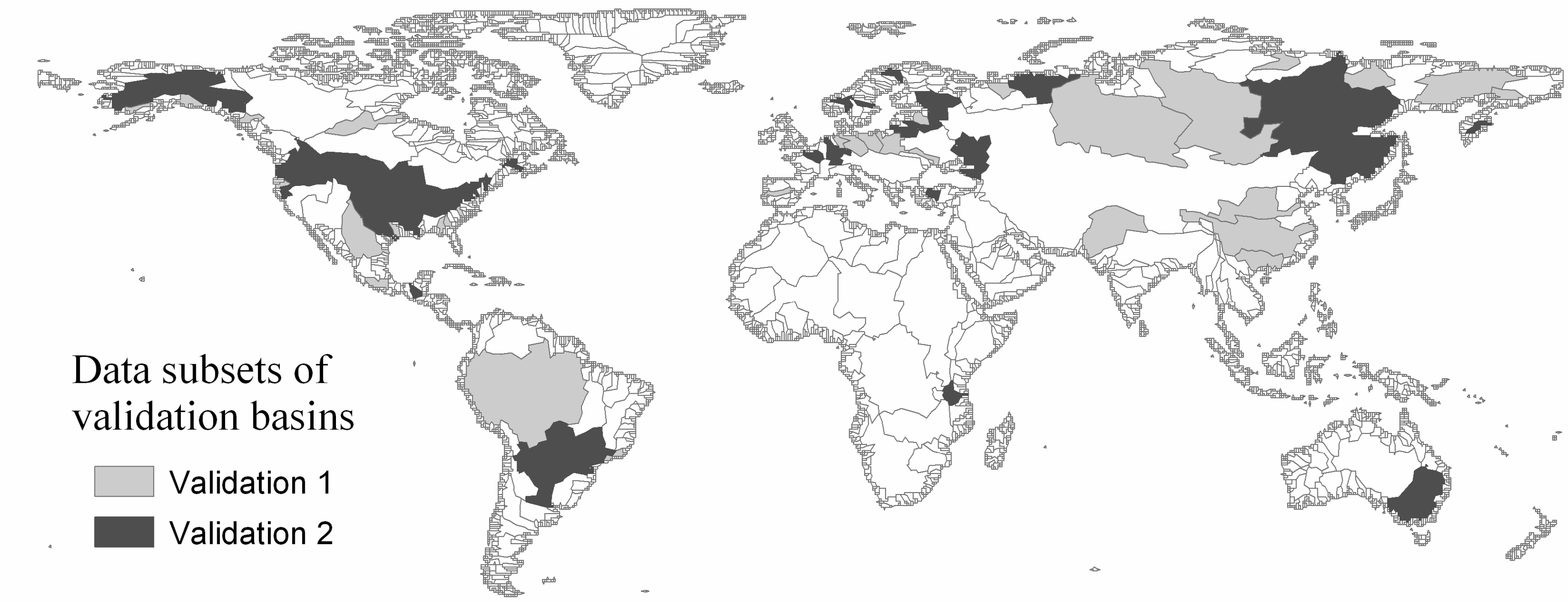

Figure 2.1 Geographic distribution of basins used to calibrate and validate NEWS-DIN.

2.2.2 Calibration and model analyses

Model fit

Two indicators of model fit are used in this paper: model efficiency and model error. The quality of the one to one linear relationship between the logarithm of measurements and model estimates is expressed as model efficiency (R^2, distinct from r^2, the coefficient of determination) [*Nash and Sutcliffe*, 1970].

Model error, *ME* (%), is expressed for the i^{th} basin according to Alexander et al. [2002]:

$$ME_i = \frac{Mod_i - Obs_i}{Obs_i} \cdot 100 \tag{12}$$

Where *Obs* is observed DIN export for a basin and *Mod* is modeled DIN export for the same basin.

Model calibration and validation

Equations 3 and 8 contain calibrated parameters. The watershed export coefficient (*e* in Equation 8) was calibrated separately from parameters c and d (Equation 3). Equation 3 was calibrated using model-predicted % N removed in river networks [Seitzinger et al., 2002]. Parameter e in Equation 8 was calibrated by optimizing the linear one to one relation between log measurements and log model outputs for 61 basins, hereafter referred to as the calibration basins.

To validate, the set of available DIN export measurements (n = 61) was split randomly into two subsets of 31 and 30 measurements. These two data subsets (subsequently referred to as Data subset 1 and Data subset 2) were given a basin area frequency distribution that was as equal as possible. As a result the basins belonging to each subset covered similar surface areas and each basin area class is equally represented in the two subsets (Figure 2.1). In the first validation (Validation 1), Data subset 2 was used to recalibrate parameter e (Equation 8) and Data subset 1 was used to evaluate model fit. In the second validation (Validation 2), data subset 1 was used to recalibrate parameter e and Data subset 2 was used to evaluate model fit. This approach was taken in order to avoid equifinality [*Beven*, 1993]. The recalibrated values of parameter e for Validation 1 and 2 are hereafter referred to as e1 and e2, respectively.

Model efficiency and sensitivity analyses

We conducted a model efficiency analysis, in order to test the relative contribution of NEWS-DIN's model parts in explaining DIN export. In this analysis, we evaluated how model efficiency (R^2) changed when individual model parts, such as D or TN_{dep}, were removed sequentially. Changes in R^2 are used to calculate the fraction of otherwise unexplained variation that is explained by inclusion of the removed model part.

To evaluate model sensitivity we calculated the change in modeled DIN yield for each basin resulting from 5% changes in model inputs and parameters. We then used this to calculate the global mean and maximum per basin change.

2.3 Results and Discussion

2.3.1 NEWS-DIN performance

Despite substantial uncertainties associated with model inputs, a comparison of modeled versus measured DIN yield and export indicates that NEWS-DIN's predictive capacity is quite high. NEWS-DIN explains 54-78% of the variability in DIN yield (kg N km^{-2} y^{-1}), and 72-83% of DIN export (kg N $basin^{-1}$ y^{-1}) in validation basins (Table 2.2). Median model error ranges between -21 and 19 % and the interquartile range (IQR) between 109 and 168% in validation basins (Table 2.2). The values of the calibrated parameter, e_1 and e_2, differed only slightly after being calibrated on either of the two independent data subsets, suggesting that the uncertainty in this parameter is low. Model performance of Smith et al. [2003] and Seitzinger and Kroeze [1998] when validated on Data subset 1 (Val.1) and data subset 2 (Val. 2) was lower than found for NEWS-DIN, regarding fit (R^2), precision (IQR) and bias (absolute median error). The model of Green et al. [2004] had smaller bias, equal precision, and lower fit, compared to NEWS-DIN when validated on Data subsets 1 and 2.

Fit, precision and bias obtained during calibration on the calibration basins are similar to those obtained during validation. Parameter e was set to 1.1 by calibration. On calibration basins, 70% of the variability in DIN yield (kg N km^{-2} y^{-1}) (Figure 2.2a), and 79% of DIN export (kg N $basin^{-1}$ y^{-1}) (R^2 = 0.79) (Figure 2.2b) was explained. Distribution of residuals in Figures 2a and 2b suggests that NEWS-DIN yield and export estimates are relatively unbiased. The median percent error of 7% also suggests a low model bias. The calibration R^2 value for DIN yield (0.70) is slightly greater than calibration values reported for DIN yield for two other comparable calibrated global models; 0.59 [*Smith et al.*, 2003], 0.68 [*Green et al.*, 2004].

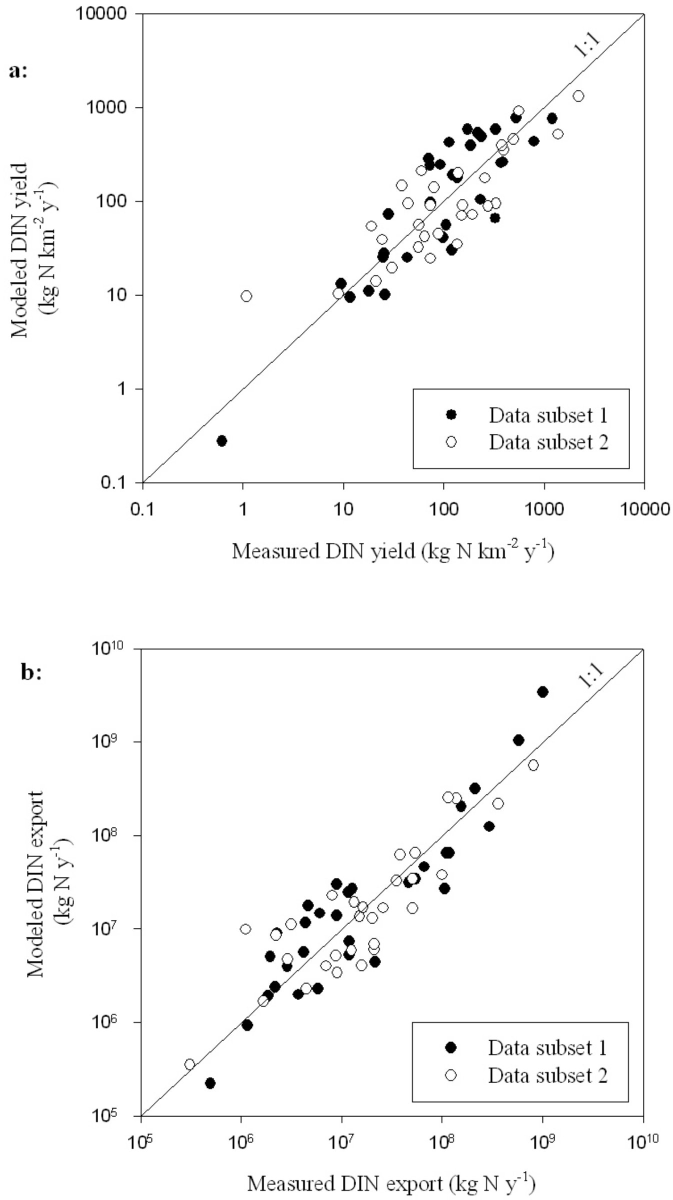

Figure 2.2 A: Modeled versus measured DIN yield (kg N km^{-2} y^{-1}). R^2= 0.70. B: Same as A, but for DIN export (kg N y^{-1}). R^2= 0.79.

2.3.2 Predicted spatial distributions

DIN export

Predicted DIN yields spanned 7 orders of magnitude, ranging from 0.0004 to 5217 kg N km^{-2} y^{-1} (Figure 2.3). NEWS-DIN's highest DIN yields were typically predicted for tropical humid basins such as the Amazon and the Zaire, densely populated basins with a high GDP (Gross Domestic Product) such as the Rhine and the Thames (England), and basins with much intensive agriculture such as the Ganges and the Chang Jiang (Eastern China). Lowest yields were predicted for basins with low population densities such as most basins at high latitudes, and also for basins in arid regions such as the Tamanrasett (Africa) or the Nile.

Largest DIN export by basin was predicted for the Amazon (3.4 Tg N y^{-1}), followed by the Ganges (2.2 Tg N y^{-1}), Chang Jiang (1.0 Tg N y^{-1}), Zaire (0.8 Tg N y^{-1}) and Mississippi (0.6 Tg N y^{-1}). Together, these high export-basins account for 32% of the global total.

NEWS-DIN-predicted distribution of high and low DIN yields is somewhat different from that predicted by other DIN export models. For example, Seitzinger and Kroeze [1998] predicted a much lower DIN yield in humid tropical areas than is predicted by NEWS-DIN. This can be explained by the fact that their model did not account for biological N2 fixation. Green et al. [2004] also predicted much lower DIN yields in tropical areas than NEWS-DIN. This is probably due the fact that the Green et al. [2004] model includes a positive relationship between N retention and temperature, whereas NEWS-DIN does not.

Table 2.2 Metrics of model performance during validation of NEWS-DIN and three other global DIN export models on two validation subsets (after being calibrated on independent subsets of rivers). Datasets Val. 1 and Val. 2 contained 30 and 31 basins, respectively. Model errors (%) and R^2 are computed as defined in Section 2.2.1. for DIN export (kg N y^{-1}) and DIN yield (kg N km^{-2} y^{-1}).

			R^2		Model errors (%)					
model	Validation	Fitted Par.	Yield	Export	IQR[a]	Min.	25th	Median Error	75th	Max.
NEWS-DIN	Val. 1	1.2[b]	0.78	0.83	168	-78	-41	19	127	332
	Val. 2	1.1[c]	0.54	0.72	109	-74	-49	-21	60	781
Smith et al. [2003]	Val. 1	n.a.	0.36	0.71	130	-88	-51	29	79	1513
	Val. 2	n.a.	0.20	0.59	146	-81	-45	34	102	488
Seitzinger and Kroeze [1998]	Val. 1	n.a.	0.40	0.56	214	-97	-84	6	130	2155
	Val. 2	n.a.	0.33	0.52	124	-95	-74	-23	51	1070
Green et al. [2004]	Val. 1	n.a.	0.54	0.52	157	-96	-58	6	99	541
	Val. 2	n.a.	0.25	0.25	162	-99	-51	3	111	623

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

[b] e_1

[c] e_2

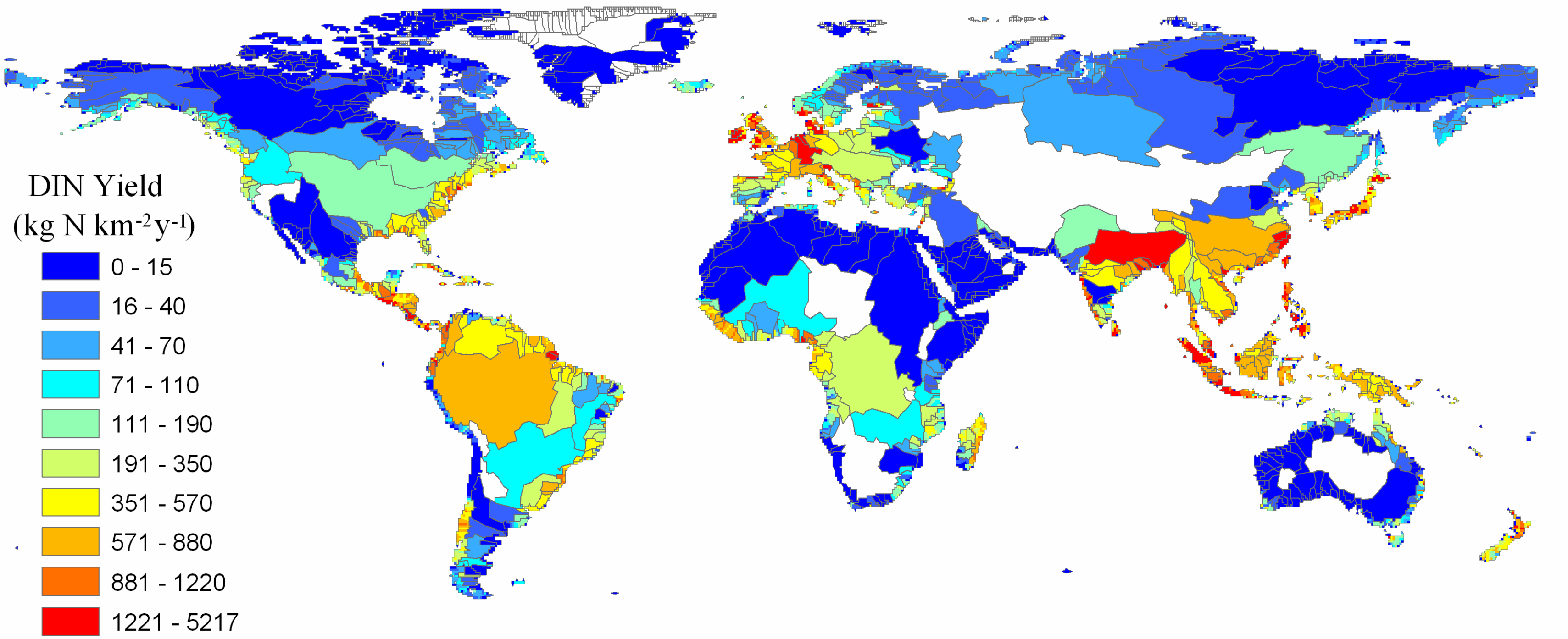

Figure 2.3 Modeled DIN yield by exoreic basin in kg N km^{-2} y^{-1}.

Source contributions to river DIN export

According to NEWS-DIN, biological N_2 fixation is the dominant source of exported DIN (not to be confused with N loading onto land) over much of the earth's surface (Figure 2.4). N_2 fixation constitutes the dominant source of DIN export to the coast in many tropical, subtropical and boreal basins, including basins in Canada, Russia, central Africa, Indonesia and Brazil (Figure 2.4). Anthropogenic N, especially from fertilizer, is the dominant source of DIN export in southern and eastern Asia, western Europe, and the central U.S.. Fertilizer is the dominant source of exported DIN in two thirds of the basins with DIN yields exceeding 1000 kg N km^{-2} y^{-1}. Sewage point sources are predicted to dominate DIN export mainly in arid basins (in e.g. Mexico, north Africa, west Australia, Arabia). This is because arid basins have low modeled export of DIN from diffuse sources, due to a low FE_{ws} (Equation 8). Sewage point sources are also often the dominant source of DIN export in numerous small (≤ 15000 km^2) densely populated basins, due to high human N emission and their relatively low predicted aquatic retention (Section 2.3.3). NEWS-DIN predicts that manure addition is mainly the dominant source of exported DIN in parts of the eastern U.S. (e.g. Florida, Georgia), south-eastern Australia and Argentina.

The magnitude of predicted basin DIN export is related to the NEWS-DIN-predicted dominant DIN source. Largest DIN export is generally predicted to occur in basins where fertilizer dominates DIN export, whereas lowest DIN export is often predicted to occur in basins in which point sources dominate DIN export.

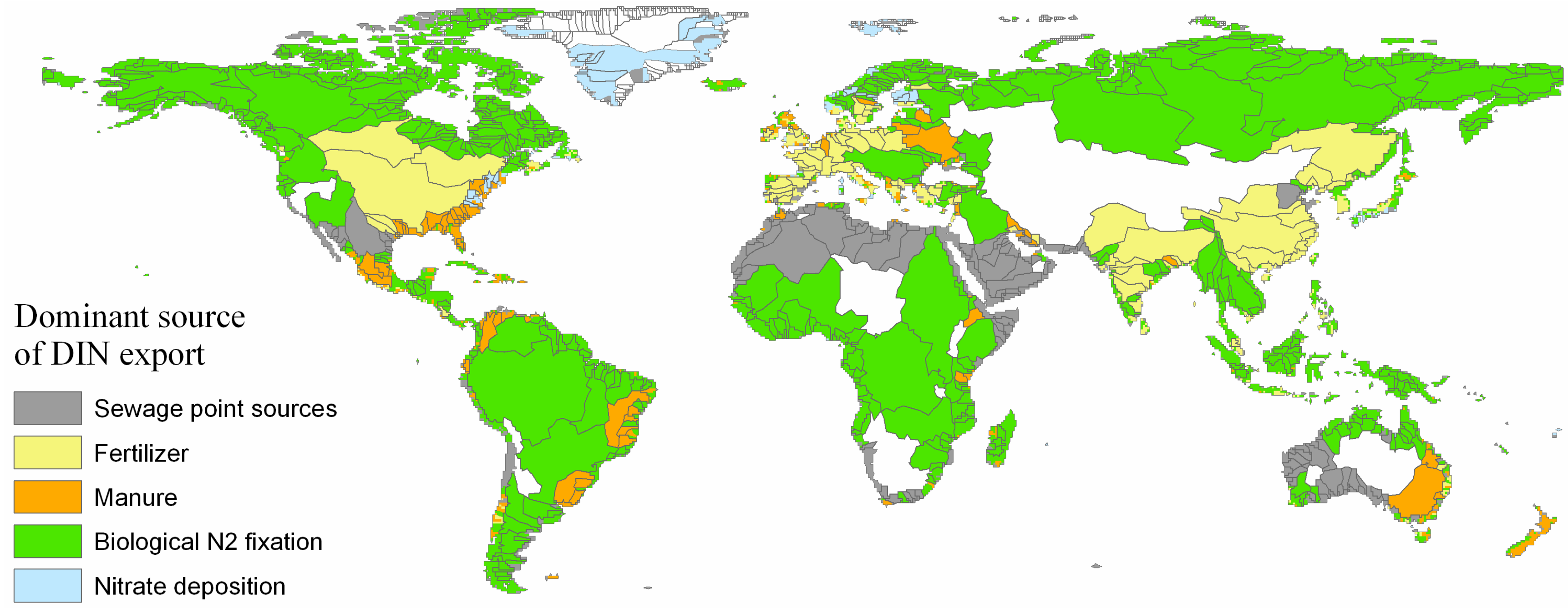

Figure 2.4 NEWS-DIN-predicted dominant sources of DIN export in exoreic basins.

2.3.3 DIN retention

Global patterns of N retention and their underlying causes are relatively uncertain [*Nixon et al.*, 1996; *Seitzinger*, 1988; *Van Breemen et al.*, 2002]. NEWS-DIN enables us to make predictions regarding global patterns of DIN retention in river systems.

Due to model assumptions, NEWS-DIN predicts that the highest aquatic retention is in basins with dams on their main stem such as the Rio Grande or Huang He. Lowest aquatic retention was generally predicted in small basins close to oceans.

Dams greatly increased predicted DIN retention in some rivers. For example in dam-influenced rivers such as the Colorado, Rio Grande, Orange, and Huang He, the predicted dam induced retention ranged from 16% to 97% percent of total aquatic retention. However, according to NEWS-DIN, dam induced retention has a relatively small impact on DIN retention at the global scale. NEWS-DIN predicts that just 6% of the DIN retained in aquatic systems globally is due to dams. This is consistent with an analysis by Seitzinger et al. [2002] indicating average contribution of 2% by dams to aquatic retention in 16 north eastern U.S. watersheds with small to medium sized dammed reservoirs. Nevertheless, dams will likely become more important in the future because the number of dams is projected to increase much in the coming decades [*Vörösmarty et al.*, 2003].

The percentage of N loaded onto basins from point and non-point sources (both anthropogenic and natural) exported as DIN ranged from 0.0001% to 43% for basins larger than ten 0.5 x 0.5° grid cells (Figure 2.5). Lowest retention rates were predicted for basins with high runoff (>1 m y^{-1}), because FE_{ws} is one in these basins (Section 2.1.4). Highest retention rates were predicted for basins with extensive damming or very low annual precipitation and runoff. According to NEWS-DIN, in exoreic dry areas of Africa, Asia, Australia and North America where runoff is less than 0.1 m y^{-1} (30% of the exoreic world), 95% (on average) of N applied to watersheds is not exported to coastal waters as DIN. Export from these arid systems is predicted to be low because there is very little runoff to transport soluble nitrogen (nitrate) from the soil and unsaturated zone to streams.

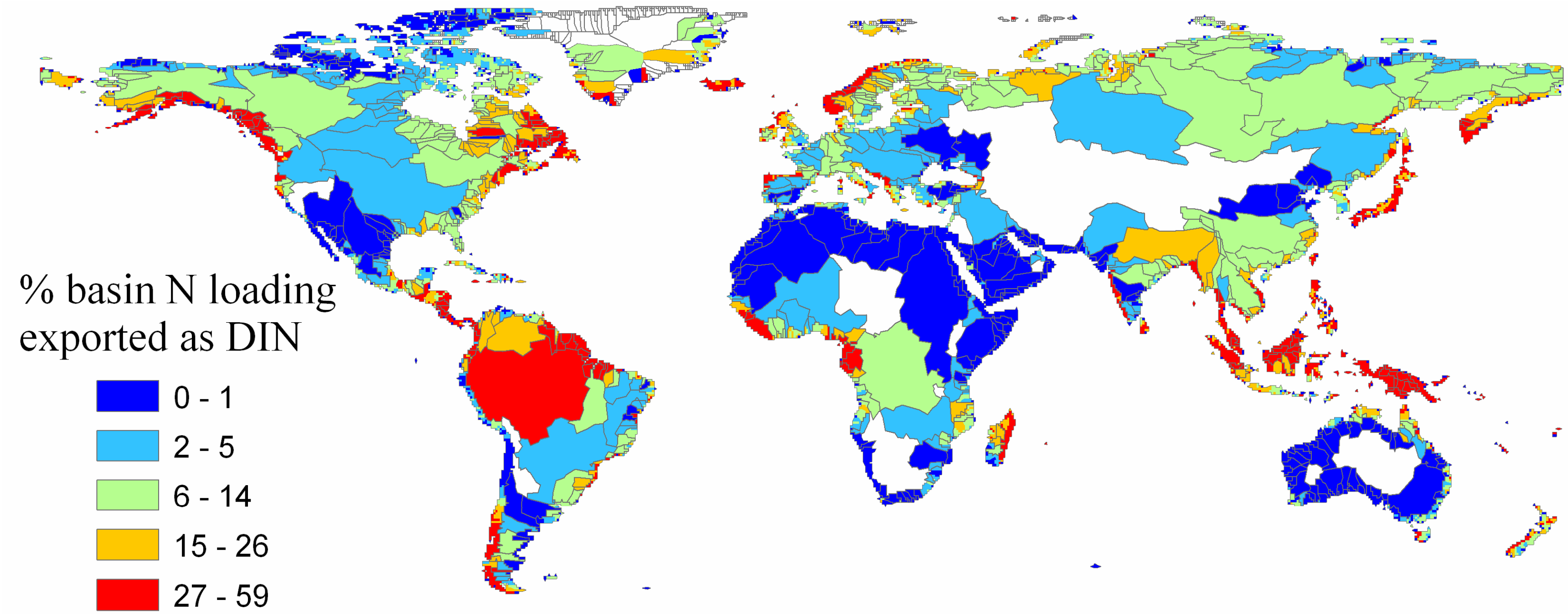

Figure 2.5 Percentage of N loaded onto basins from point and non point sources that NEWS-DIN predicts is exported to coastal waters as DIN.

2.3.4 Global and regional analyses

NEWS-DIN predicts a global rate of river DIN export of 24.8 Tg N y^{-1}. NEWS-DIN's estimate is somewhat higher than other estimates for the global total of DIN export: 20.8 Tg N y^{-1} by *Seitzinger and Kroeze* [1998], and 14.5 Tg N y^{-1} by Green et al. [2004]. The ratio of NEWS-DIN's prediction of global DIN export to predictions of global TN export (48.7 Tg N y^{-1} [*Galloway et al.*, 2004], 54 Tg N y^{-1} [*Van Drecht et al.*, 2003], 40 Tg N y^{-1} [*Green et al.*, 2004]) suggests that DIN is responsible for 46-62% of global TN export by rivers.

NEWS-DIN predicts that of the 24.8 Tg N y^{-1} exported by rivers, 15.8 Tg N y^{-1} is anthropogenic. Meybeck [1982] estimated global anthropogenic DIN export as 7 Tg y^{-1} for the late 1970s. This suggests an approximate doubling of anthropogenically derived DIN export between the late 1970s and the mid 1990s. This is reasonable considering that sources of N to surface waters have increased dramatically, including a 1.6 fold increase in world population [*UN*, 1998], increase in animal populations [*FAO*, 1996], 2.6 fold increase in rates of N fertilizer application [*Bouwman et al.*, 2005b], and the widespread development of enhanced sewer systems, which more efficiently export sewage, and hence DIN, to rivers. Our estimate of DIN export from natural sources (9.3 Tg y^{-1}) is higher than, previous estimates of non-anthropogenically derived DIN export (5 Tg N y^{-1} [Meybeck 1982; Seitzinger and Kroeze 1998] and 2.4 Tg N y^{-1} [*Green et al.*, 2004]). The much lower estimate of natural DIN export by Green et al. [2004] is probably due to an assumption that a very small fraction of modeled organic non point sources (N_2 fixation and manure) is exported as DIN. Furthermore, as mentioned previously, Green et al. [2004] assume that DIN delivery to coastal waters is reduced by high temperatures, reducing their predicted DIN export from tropical basins where natural sources are relatively important. The proportion of NEWS-DIN-modeled riverine DIN from natural sources (36 %) is slightly lower than found for TN by Van Drecht et al. [2003] (51%). This difference is consistent with the notion, supported by existing data [*Van Breemen*, 2002], that natural systems export N mainly in forms other than DIN.

NEWS-DIN predicts that of all the continents, Asia exports the most DIN to its coasts, due in part to its large surface area, but also to its large population and cultivated land area. This is consistent with predictions by previous DIN export models [e.g. *Seitzinger and Kroeze*, 1998]. The relative share of DIN export from Africa and South America in global DIN export predicted by NEWS-DIN are higher than predicted by Seitzinger and Kroeze [1998]. Absolute NEWS-DIN predicted DIN export from Africa and South America is also higher than predicted by Seitzinger and Kroeze [1998]. This is probably due to the fact that N_2 fixation which is a relatively important source of DIN export on these two continents, is not explicitly represented in the model used by Seitzinger and Kroeze [1998].

According to NEWS-DIN, biological N_2 fixation is the dominant DIN export source for the majority of continents, including North America, South America, Africa and the aggregate of Australia, Indonesia and Oceania, where it is predicted to account for 39, 73, 72 and 70 % of exported DIN, respectively. There is substantial variation in other predicted DIN sources between continents. Contributions to DIN export by manure and fertilizer vary most between continents. Contribution to continental DIN export by manure and fertilizer spans from 10 to 30% and from 1 and 39%, respectively (Figure 2.6).

According to NEWS-DIN, natural N_2 fixation is the single largest source of river exported DIN globally, accounting for 36% (9 Tg N y^{-1}) of the total global DIN exported from watersheds. Inorganic fertilizer application is the second most important source of river-exported DIN, accounting for 21% (5.3 Tg N y^{-1}) of the total global export. Manure application, agricultural N_2 fixation, NO_y deposition, and sewage point sources, are less important, though still significant, sources of DIN to coastal waters, accounting for 18, 15, 8, and 2% of global DIN export, respectively (4.5, 3.8, 1.9, and 0.4 Tg N y^{-1}). NEWS-DIN predicts that agricultural sources of DIN (inorganic fertilizer, animal manure and agricultural N_2 fixation) account for about half of the total DIN export globally.

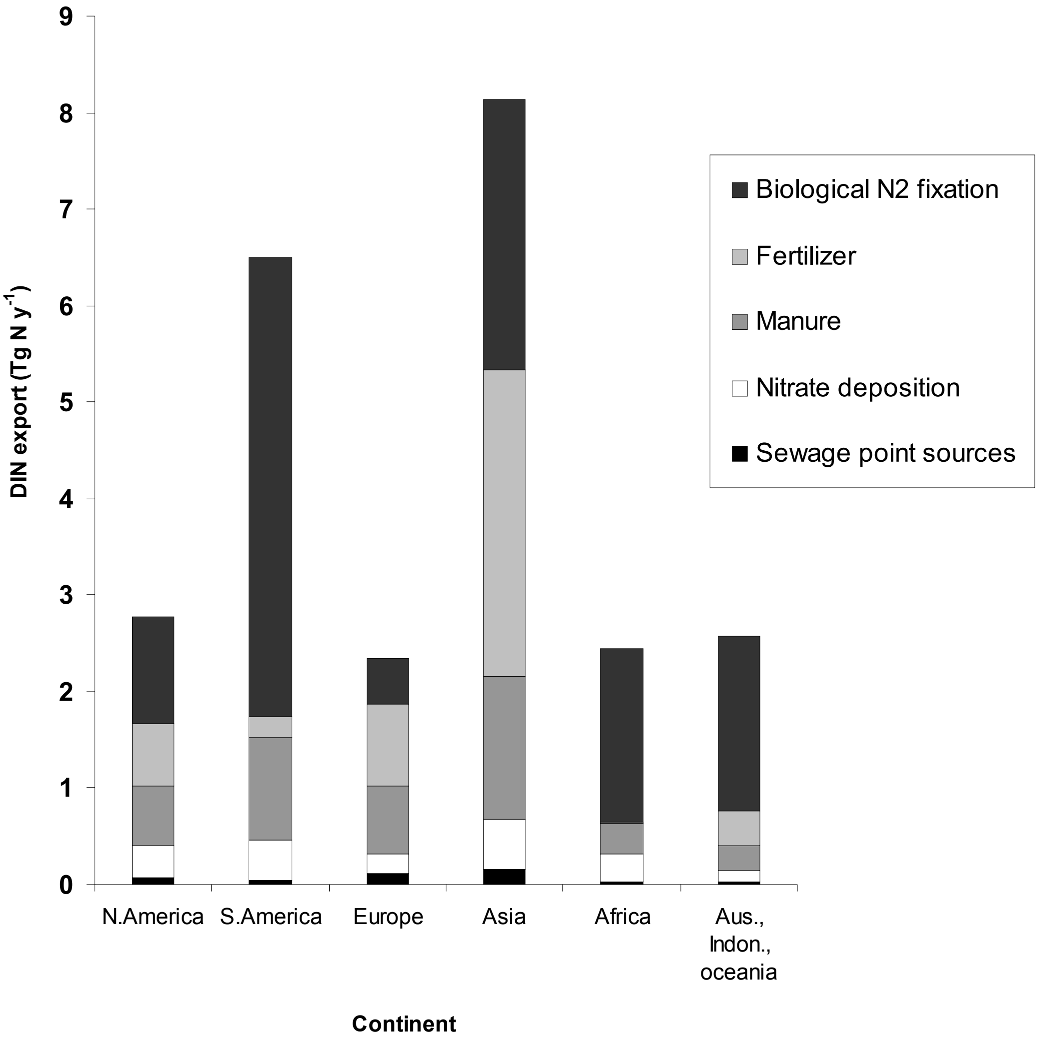

Figure 2.6 Modeled DIN export to the coastal zone of each continent by modeled source.

2.3.5 Efficiency and sensitivity of model parts

We analyzed the change in model efficiency upon removing different parts of NEWS-DIN (2.2.3). For the modeled DIN yield per basin, the change in model efficiency ranged from 0 to 87 % depending upon the portion of the model removed (Table 2.3). The largest decrease in model efficiency occurred when *R* or river network retention terms were removed from NEWS-DIN. This suggests that future improvements to NEWS-DIN may depend much upon a better understanding of river network retention and the factors controlled by runoff within watersheds.

Table 2.3 Decrease in NEWS-DIN model efficiency as a function of removal of individual model components (described in Section 2.2.3).

Model Component	% decrease in model efficiency resulting from component removal
River network retention (fraction)	61
Anthropogenic removal of river water (fraction)	2
Dam reservoirs (fraction)	36
Sewage point sources (kg N $km^{-2}y^{-1}$)	0
e (Equation 8)	1
Runoff (m y^{-1})	87
Atm. NO_3^--N dep. (kg N $km^{-2}y^{-1}$)	12
Net biol. N_2 fixation (kg N $km^{-2}y^{-1}$)	46
Fertilizer (kg N $km^{-2}y^{-1}$)	63
Manure (kg N $km^{-2}y^{-1}$)	36
Harvesting and grazing N removal (kg N km^{-2} y^{-1})	18

Removal of the sewage point source term from NEWS-DIN decreased model efficiency only slightly, which is consistent with the low model sensitivity for sewage (Table 2.4). Removing the calibrated watershed export coefficient, *e*, from NEWS-DIN also decreased model efficiency only slightly, suggesting that NEWS-DIN is relatively insensitive to uncertainty in this calibrated parameter. The small change in model efficiency resulting from removal of irrigation and river diversion terms from NEWS-DIN suggests that anthropogenic river water removal has a relatively small impact on DIN export at the global scale. However, at a local scale, anthropogenic river water removal can have an important impact on DIN export. For example, in the Colorado (south west U.S.) and the Eastmain (east Canada), anthropogenic river water removal accounts for 99.5% and 93% of DIN retention, respectively.

The effect of 5% changes in inputs and parameters on the mean percent change in output (basin DIN export or yield) ranges from 0.4 to 9.6 % (Table 2.4). Highest mean percentage change in output was found for coefficient *c* (Equation 3). Sensitivity analyses suggests that NEWS-DIN predictions are quite sensitive to treatment of its retention terms (Table 2.4), which is consistent with results of the efficiency analyses. In this regard, NEWS-DIN is similar to previous DIN export models (e.g. Seitzinger et al., [2002]; Smith et al., [1997a]). Despite the sensitivity of DIN models to river N retention, estimates of river N retention are quite uncertain, due largely to a lack of N-retention studies at the appropriate scale. To improve these estimates it will be necessary to make measurements of denitrification throughout river networks at appropriate temporal and spatial scales to refine understanding of the magnitude of denitrification and controlling factors in river networks.

The sensitivity of NEWS-DIN to parameters and inputs in individual basins can be much larger than its average sensitivity of DIN export in all basins of the world. This is especially true for model parts related to aquatic retention such as coefficient *c*, irrigation and river diversion and dam reservoirs parts. The frequency of occurrence of basins, with high sensitivity to inputs related to aquatic retention such as damming, irrigation and river water transfer, is expected to increase in the future [*Dynesius and Nilsson*, 1994; *Vörösmarty et al.*, 2003]. Therefore, it is important to improve our understanding of the relationship between damming, anthropogenic river water removal and DIN export for future predictions.

Table 2.4 Results of sensitivity analysis: mean and maximum change in modeled DIN yield per basin as a function of changing inputs and model parameters by +5% and –5%

Input or parameter	Mean output change (%)	Max. change (%)
River network retention (fraction)	5.0	5
c (Equation 3)	9.6	31
d (Equation 3)	0.4	1
Anthropogenic removal of river water (fraction)	2.5	995
Dam reservoirs (fraction)	0.6	25
Sewage point sources (kg N km^{-2} y^{-1})	0.7	5
e (Equation 8)	3.8	5
Runoff (m y^{-1})	3.8	5
Atm. NO_3^- dep. (kg N km^{-2} y^{-1})	0.7	5
Net boil. N_2 fixation (kg N km^{-2} y^{-1})	2.7	5
Fertilizer (kg N km^{-2} y^{-1})	0.9	7
Manure (kg N km^{-2} y^{-1})	1.1	6
Harvesting and grazing N removal (kg N km^{-2} y^{-1})	1.2	7

Acknowledgements

We are grateful to members of Global Nutrient Export from WaterSheds (Global NEWS) and staff of the Environmental Systems Analyses (ESA) group for useful discussion during development of NEWS-DIN. We also thank Lex Bouwman for providing us with data on sewage effluents and net N input via fertilizer and manure, Frank Dentener for data on atmospheric NO_3^--N deposition, Charles Vörösmarty for data on basin delineations, runoff and dams and Pamela Green for data on biological N_2 fixation. Egon Dumont is grateful to Netherlands Organization for Scientific Research (NWO) for funding his research and for the facilities made available by Wageningen University and Institute of Marine and Coastal Sciences. Sybil Seitzinger and John Harrison's work on this project was supported in part by NOAA Sea Grant (NJSG-05587). This paper is a contribution to the Global NEWS project that receives support from UNESCO-IOC and the U.S. National Science Foundation.

Appendix 2.1. River basin data used for calibration and validation.

River	Continent	Basin area (km^2)	Measured DIN yield (kg N km^{-2} y^{-1})	Validation subset ID[a]	Data source[b]
Altamaha	N. America	41450	113.2	1	2
Amazon	S. America	5833000	172.5	1	1
Amur	Asia	1748000	79.7	2	1
Anabar	Asia	98550	11.7	1	1
Appalachicola	N. America	54660	235.1	1	2
Balsas	N. America	122600	73.1	1	1
Brazos	N. America	124600	56	2	2
Bug	Europe	68980	28.3	1	3
Chang Jiang	Asia	1788000	327.5	1	1
Churchill (Hudson Bay)	N. America	302400	9.5	1	1
Colorado (Texas)	N. America	120800	24.2	2	2
Columbia	N. America	729300	74.1	2	1
Copper	N. America	66990	325.2	1	2
Dalalven	Europe	29820	56.7	2	1
Daugava	Europe	83160	151.3	2	3
Don	Europe	421600	19.1	2	1
Elbe	Europe	148000	795.4	1	1
Glama	Europe	47310	191.8	2	1
Huang He	Asia	890500	120.5	1	1
Hudson	N. America	43070	381.1	2	1
Indus	Asia	1139000	136.9	1	1
Kamchatka	Asia	50370	88.8	2	1
Klamath	N. America	32080	71	1	2
Kolyma	Asia	663200	18	1	1
Kuban	Europe	63630	330.9	2	1
Kuskowin	N. America	115400	136.9	2	2
Lena	Asia	2433000	21.1	2	1
Mezen	Europe	75430	24.8	1	1
Mississippi	N. America	3191000	255.6	2	1
Murray	Australia	1028000	1.1	2	1
Narva	Europe	58010	73.3	1	3
Nemanus	Europe	96630	138.4	2	1
Neva	Europe	283500	74.1	2	1
Nushagak	N. America	35300	105.3	1	2
Ob	Asia	3015000	98	1	1
Odra	Europe	119400	389.8	1	1
Paraiba do Sul	S. America	62760	185.2	1	1
Parana	S. America	2654000	43.9	2	1
Pechora	Europe	313100	64.7	2	1
Pee Dee	N. America	27640	219.3	1	2
Penzhina	Asia	85540	25.5	1	1
Potomac	N. America	38300	395.7	2	1
Rhine	Europe	164500	2200.4	2	1
Rio Grande (US)	N. America	801900	0.6	1	1
Rufiji	Africa	186100	275.9	2	1
Sacramento	N. America	58690	38.1	2	1
Saint John	N. America	52850	59.8	2	1

Sakarya	Asia	56830	155	2	1
Seine	Europe	73190	1364.9	2	1
Stikine	N. America	51170	233	1	2
Susquehanna	N. America	71860	493.2	2	1
Tejo	Europe	73090	122.8	1	1
Tornionjoki	Europe	34510	9	2	1
Trinity	N. America	47380	92.2	1	2
Usumacinta	N. America	67890	562.2	2	1
Weser	Europe	45470	1204.7	1	1
Wisla	Europe	180000	371.8	1	1
Yana	Asia	224200	25.9	1	1
Yenisei	Asia	2569000	43.1	1	1
Yukon	N. America	852700	30.6	2	1
Zhujiang	Asia	407100	523.3	1	1

[a] Validation subsets 1 and 2 as described in Section 2.2.4

[b] 1 = Meybeck & Ragu [1995], 2 = USGS [*Alexander et al.*, 1996], 3 = European Environmental Agency [*EEA*, 1998]

Appendix 2.2 Definitions of variables and parameters

Symbol	Definition
A	Basin area (km^2)
c	Fitted parameter in relationship between basin area (A) and river network retention (L_{den}); in NEWS-DIN coefficient c was set to equal 0.0605.
d	Fitted parameter in relationship between basin area (A) and river network retention (L_{den}); in NEWS-DIN coefficient d was set to equal 0.0443.
D	Fraction of DIN retained in dammed reservoirs (0-1)
$DEPT_i$	Depth of reservoir i (m)
DIN	DIN yield modeled per river basin (kg N km^{-2} y^{-1})
DIN_{meas}	DIN yield measured per river basin (kg N km^{-2} y^{-1})
DIN_{sew}	DIN from sewage point sources along rivers (kg N km^{-2} y^{-1})
e	Coefficient defining the relationship between runoff and EF_{ws} if runoff is less than 0.91 (m y^{-1}); in NEWS-DIN coefficient e was set to equal 1.1.
E_N	Per capita human N emission (kg N $individual^{-1}$ y^{-1})
FE_{riv}	Fraction of total point and non point DIN inputs to the river that is exported as DIN(0-1)
FE_{ws}	Fraction of N from diffuse sources in the watershed that leaches to the river as DIN (0-1)
G	Fraction of N from manure, fertilizer and agricultural N_2 fixation available for leaching after harvest and grazing
H	Population density (individuals km^{-2})
I	Fraction of the population connected to sewerage systems (0-1)
L_{den}	Fraction of DIN lost in the basin river network (0-1)
Q	Basin discharge (km^3 y^{-1})
Q_{div}	Amount of discharged water lost from the river by anthropogenic transfer of water out of the basin, mostly by artificial channels (km^3 y^{-1}) [a]
Q_i	Discharge intercepted by dam i (km^3 y^{-1}) [b]
Q_{irr}	Amount of discharge removed for irrigation, minus the amount of irrigation water that ultimately flows back into the river, i.e. extracted irrigation water that evaporates on irrigated fields (km^3 y^{-1}) [a]
Q_{nat}	Amount of discharge if Q_{irr} and Q_{div} did not occur (km^3 y^{-1}) [a]
Q_{rem}	Fraction of DIN retained, due to the anthropogenic removal of (DIN containing) river

	water (0-1)
R	Precipitation minus evaporation (m y^{-1})
Rt_i	Water residence time of reservoir i (y)
T_N	Fraction of N removed by wastewater treatment (0-1)
TN_{am}	Addition of manure N (kg N km^{-2} y^{-1})
TN_{dep}	Atmospheric deposition of NO_y-N (kg N km^{-2} y^{-1})
TN_{diff}	N from diffuse sources that is mobilized from the watershed soils and sediments (kg N km^{-2} y^{-1})
TN_{exp}	N in crops and grassland that is removed from the land by harvesting and grazing (kg N km^{-2} y^{-1})
TN_{fe}	Addition of fertilizer N (kg N km^{-2} y^{-1})
TN_{fix}	Natural and agricultural biological N_2 fixation (kg N km^{-2} y^{-1})
TN_{sew}	N from sewage effluents discharged to surface water (kg N km^{-2} y^{-1})

[a] from Dynesius and Nilsson [1994]

[b] from Fekete et al. [2000]

3 A framework to identify appropriate spatial and temporal scales for modeling N flows from watersheds[b]

Abstract

We describe a framework to identify the appropriate spatial and temporal scales for nitrogen (N) flow models. The framework has been developed for models that predict N export from large watersheds and the contribution of N sources and N sinks to this N export. With this framework, modelers can identify the appropriate scale for model predictions and for independently scalable model parts.

The framework determines the appropriateness of model scales using a set of indicators. These indicators are to be specified by the modeler and are associated with four criteria. The criteria require modeling scales to correspond with (A) data and scenarios, (B) model assumptions, (C) available resources for modeling, and (D) appropriately scaled predictions.

We conclude that the framework can contribute substantially to a well-balanced and comprehensive identification of appropriate modeling scales.

[b] **This chapter has been submitted for publication as:**
Dumont, E., C. Kroeze, A. F. Bouwman, E. J. Bakker, A. Stein, and R. Leemans, A framework to identify appropriate spatial and temporal scales for modeling N flows from watersheds, submitted to Ecological Modelling.

3.1 Introduction

Many different large-scale watershed nitrogen (N) flow models exist [*Andersen et al.*, 2003], all of which simulate processes related to the horizontal movement of N through large drainage networks of river basins. An important property of such models is their modeling scale. The modeling scale affects the processes that can be described, the required input data, the size of watersheds that can be modeled, the scenarios that can be simulated, and the scale of resulting predictions. It can be measured as a combination of support, extent, and stream order of model parts. Spatial model support is the size of the areas represented by single values of input variables used in model calculations [*Heuvelink*, 1998]. Temporal model support is duration of the times represented by single values of input variables used in model calculations. If the model is stochastic then model support applies to single input distributions instead of single input values. Model extent is the total range of time or space within which processes are modeled. The spatial extent of N-flow models is typically a watershed or a group of adjacent watersheds. The temporal model extent is usually between a few months and a few decades. Stream order is a measure of the size of river reaches that are modeled.

The reason for selecting a particular modeling scale is usually not explicitly reported, making the appropriateness of a particular modeling scale difficult to judge. Modelers often have no clear guide for selecting appropriate spatial and temporal modeling scales for predicting N flows in large river basins.

We argue that it is necessary to consider different model parts when assessing the appropriateness of scales of watershed N-flow models. The reason is that watershed N-flow models can be composed of different parts, each of which can have a different nature. For example, some parts may be classified as empirical or conceptual, while other parts of the model may be process-based. Alternatively, model parts may be classified as steady state or dynamic.

N-flow models relevant for environmental impact assessments are often used to predict N export for specific river basins. Such models can have different spatial and temporal supports and extents. For example, the Riverstrahler model [*Billen et al.*, 1994] has been applied with various temporal supports to river basin areas ranging from 100 to 100,000 km^2 and with spatial model supports ranging from 1st to 5th order upstream basins ranging in area between 1 and 5000 km^2 [*Sferratore et al.*, 2005]. Other models, however, were developed to cover groups river basins up to a global scale coverage with varying temporal scales. Hence, with models ranging from basin-specific to global and with different temporal scales, it is interesting to asses what the appropriate scale is.

The appropriateness of modeling scale depends on a number of factors, such as model assumptions, available resources for modeling, the scale of required predictions, and properties of data, mitigation options, and scenarios. These factors and their relative importance are affected by the research objective. For example, if the objective is to trace the relative change in global river export to coastal seas, then the spatial model scale could be that of continents and the temporal scale could be one year. However, if the objective is to determine the effect of seasonal variations of river N export on the risk of algal blooms in a specific estuary or coastal sea, the spatial scale would be the river basin and the temporal scale a month.

There is empirical information on the appropriateness of scales of N-flow models and methodological information on the identification of such appropriateness in the literature. We will first review existing empirical literature, followed by existing methodological literature.

Empirical information on the validity of sets of coupled equations of existing models at different supports has been obtained directly (i) from model developers (e.g. Table 3.1), (ii) from reported tests of prediction accuracy of N-flow models with the same equations at different supports [*Beaujouan et al.*, 2001; *Bellamy and Loveland*, 2001a; *Caraco et al.*, 2003; *Curmi et al.*, 1998; *FitzHugh and Mackay*, 2000; *Jha et al.*, 2004; *Johnes and Butterfield*, 2002; *Mamillapalli et al.*, 1996; *Sferratore et al.*, 2005], and (iii) from tests of the smallest detail in modeled scenarios that still has a meaningful effect on model predictions [*Eckhardt et al.*, 2003; *Joao*, 2002]. Other empirical literature can be used as a basis for estimation of appropriateness of model scales. An example is literature reporting differences between scales in the expected contribution or predictive value of processes to watershed N-flow. Such literature can be used to check if processes described by watershed N-flow model equations agree with the processes that are reported to have high contribution or predictive value at the scale at which these equations are applied. We distinguish (i) expert judgment [*Meybeck*, 2002; *Wagenet*, 1998], (ii) validation of watershed N-flow models with different process descriptions on the same scale [*De Wit*, 1999], (iii) radioactive tracers indicating processes affecting exported N from watersheds on different scales [*Costanzo et al.*, 2003], and (iv) an approach called minimum information requirement where all processes that do not contribute to prediction accuracy on a scale considered are removed from a detailed watershed N-flow model developed on a fine scale [*Quinn*, 2004; *Van Herpe et al.*, 2002]. Empirical literature on the scale at which individual N processes emerge may also be used as a basis for estimation of appropriateness of scales of watershed N-flow models. Such literature can report results of empirical research on the sizes of patches of N processes [*Dent and Grimm*, 1999; *Jarvie et al.*, 1999; *Torgersen et al.*, 2004; *Wolfert*, 2001] or of expert judgment [*McClain et al.*, 2003; *Seitzinger et al.*, 2006]). Also some empirical information exists on the scale of required predictions of watershed N-flow models [*Meybeck et al.*, in prep.; *Omernik*, 2003; *Omernik and Bailey*, 1997; *Sherman*, 1991].

Table 3.1 Indications of appropriate modeling scale reported for seven models of N flow in river basins compiled from Andersen et al. [2003].

Model name	Temporal support of predictions	Spatial support of modeled processes	Model reference
NL-CAT	≥ day	Farm level for modeling terrestrial processes	Wolf et al. [2003]
N-LES CAT	≥ year	Field level for modeling terrestrial processes	Simmelsgaard et al. [2000]
MONERIS	≥ year	Terrestrial and aquatic processes in $50km^2$ or larger basins	Behrend at al. [2002]
TRK	≥ day	- [a]	EPA [1997]
SWAT	≥ day	- [a]	Neitsch et al. [2002]
EVENFLOW	≥ day	- [a]	Anthony at al. [1996]
NOPULU	≥ year	- [a]	Agency/IFEN [2000]

a : No indications on appropriate spatial support were given in Andersen et al. **[2003]**

There is methodological literature that supports the identification of appropriate scales of N-flow model types or parts for which the appropriate scales cannot be reliably estimated from empirical literature. Some of this methodological literature supports research on the effects of model scale on model validity, such as methods aiming at identifying scales where the modeled system is deterministic [*Bruneau et al.*, 1995; *Habeeb et al.*, 2005; *Robin et al.*, 1995; *Wood et al.*, 1988] and uncertainty analyses [*Beven*, 1995; *Heuvelink*, 1998; *Lilburne et al.*, 2004; *Vachaud and Chen*, 2002]. Other methodological literature supports the identification of the ranges of measurement scale at which different nitrogen processes can be observed, such as wavelet analyses [*Platt and Denman*, 1975] and variogram analyses [*Isaaks and Srivastava*, 1998; *Kotliar and Wiens*, 1990]. Also methodological literature exists supporting the deduction of scales at which different N processes can be observed, such as methods using knowledge of scale-specific feedbacks controlling these processes [*Easterling and Kok*, 2002; *Gibson et al.*, 2000; *Holling*, 1992; *Levin*, 1992] or knowledge of fractal properties making the processes apparent over a range of scales [*Burrough*, 1981; *Schneider*, 1994; *Schroeder*, 1991; *Sposito*, 1998].

Empirical and methodological information in current literature is not suited to comprehensively and generically identify appropriate modeling scale of N-flow models. Empirical literature on the appropriateness of scales of N-flow models is not comprehensive because it usually only focuses on one measure of scale (either support, extent, or stream order, and either spatial or temporal scale) and on either the scale of processes, modeled scenarios, a single model part, or predictions. Methods reported in current literature require too much time and data for most practical applications. Further, most of these reported methods focus on only one measure of scale and on either the scale of processes or single model parts.

The purpose of this paper is to present a comprehensive framework to identify the appropriate spatial and temporal scale for N-flow models. Here models are defined as coupled sets of equations that can be applied at any scale. We refer to this framework as FAMOS (Framework for Appropriate MOdeling Scale). FAMOS has been developed for models that can predict N export from large watersheds and the contribution of N sources and sinks to this N export. With FAMOS, modelers can identify appropriate scales for model predictions and for independently scalable model parts. FAMOS may also assist in reporting the rationale behind the scale of a model's application. A preliminary version of FAMOS has been published in Dumont et al. [2006].

In the following, the general structure of FAMOS is discussed (Section 3.2). This is followed by a description of indicators that are the basis of FAMOS (Section 3.3). Subsequently, the sequence of use of indicators and SMPs is described (Section 3.4).

3.2 Overview of FAMOS

In this section, we describe FAMOS (Framework for Appropriate MOdeling Scale). It is developed to assist model users in identifying appropriate modeling scales in a well-balanced and comprehensive way. FAMOS aims at minimizing the prediction bias by ensuring sufficient validity of model assumptions while simultaneously ensuring the feasibility of model application, agreement with scales of data and scenarios, acceptable scaling error, and useful predictions.

FAMOS indicates the appropriate ranges of up to seven measures of spatial and temporal modeling scale. These are mean spatial support (*sups*), mean temporal support (*supt*), mean stream length (*sl*), spatial extent (*exts*), temporal extent (*extt*), smallest stream order (so_{min}), and largest stream order (so_{max}). The appropriate ranges of these seven measures are identified largely independent of each other. In the following, any of these seven measures will be indicated by *S* (scale). The user can to some degree choose which measures of *S* are considered in FAMOS.

Appropriate ranges of *S* are estimated for a number of SMPs. An SMP is a set of either model equations or model predictions. It has a unique *S* in existing model applications, and its S can be adjusted. Figure 3.1A and B are an example of a representation of possible SMP delineations and SMP scales. Defining SMPs is a crucial element in FAMOS.

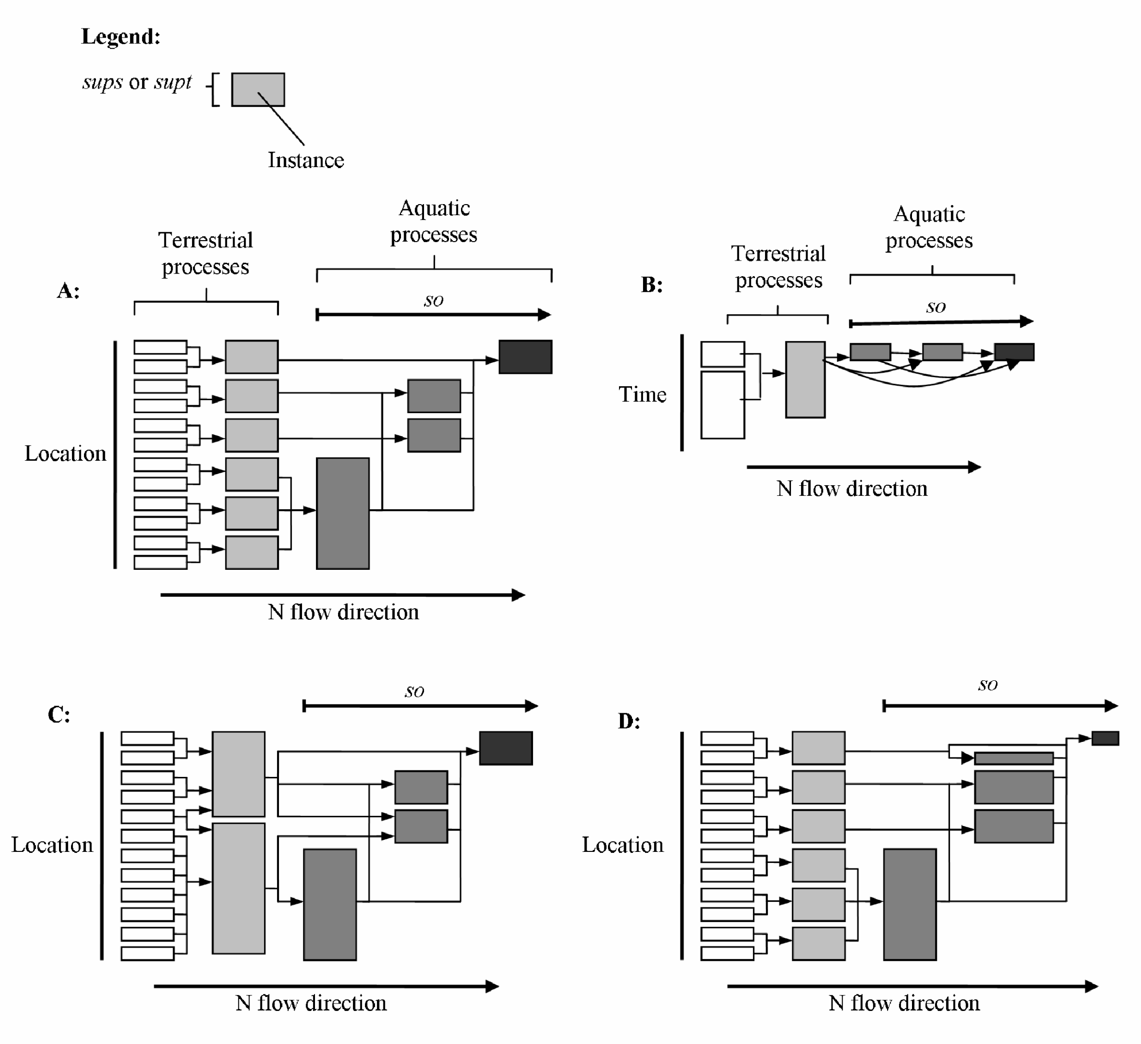

Figure 3.1 Possible delineation of SMPs, SMP instances, and their relations in a watershed N-flow model. Axes indicate the stream order and support of the SMPs and the degree to which locations and times of SMPs overlap. Rectangles are SMP instances. The grey tone indicates the SMP to which an instance belongs. Designation of SMPs to instances is arbitrary. Thin arrows indicate the flow of information. A: Possible delineation of SMPs and their spatial support (*sups*) and represented locations. B: Possible delineation of SMPs and their temporal support (*supt*) and represented times. C: The light grey SMP has increased its sups, compared to the dark grey and black SMP, leading to the necessity to downscale output of the light grey SMP, as indicated by diverging connections with these subsequent SMPs. D: The dark grey SMP has broadened the range of stream orders that it represents. This has necessitated connections with an instance of the light grey SMP that it was not connected to in A.

Four criteria are used to assess the appropriateness of modeling scale. These four criteria require modeling scales to correspond with (A) data and scenarios, (B) model assumptions, (C) available resources for modeling, and (D) appropriately scaled predictions. Fulfillment of each criterion is indicated with indicators that are the basis for the evaluation of appropriateness of each of the seven measures of scale of a scalable model part (SMP). The indicators to be used are selected by the user of FAMOS (Table 3.2). SMP scales are considered appropriate if indicator values exceed pre-specified thresholds. FAMOS enables its users to evaluate appropriateness of different modeling scales according to their preferences and ideas on appropriateness as reflected by user-defined weights for criteria and indicators. There are three reasons that the 12 indicators in Table 3.2 are suitable to test the model scale based on the four criteria: first they are relatively easy to quantify, second each can be applied to almost any N-flow model, and third together they generally cover all aspects that are required to judge the appropriateness of model scale.

As a first step, a user of FAMOS will have to decide on the model parts to be considered. Due to the complexity of most N-flow models, it is often almost impossible to consider all instances and all modeled N sources and sinks. An instance is a part of an SMP modeling an area, time or stream length for which input is homogeneous. Examples of possible N sources described by an SMP are fertilizer or sewage, and examples of N sinks described by an SMP are riparian zones or lakes. The application of FAMOS can be limited to a restricted set of instances, sources, and sinks represented by an SMP. Similarly, the analyses by FAMOS can be limited to only a subset of all SMPs.

There are several reasons for considering only a particular part of the model in the application of FAMOS. For instance, only those model parts can be included that are relatively important in the modeled N transport. Their importance can be determined by their expected contribution to outlet N flux and potential for mitigation

Table 3.2 Description of criteria A, B, C, and D (*c*) and indicators (*j*) that can be used in FAMOS.

c	*j*	**Description of criterion or indicator**
A		**Requires SMP scale to correspond with that of data and scenarios**
	1	Degree to which artifacts in available input data affect model representation of process variability [1]
	2	Measure of the fraction of the spatial or temporal variability of anthropogenic influences described in scenarios that is represented by an SMP [1]
B		**Requires SMP scale to correspond with model assumptions**
	3	Measure of the fraction of actual N flow variability on the scale of the considered SMP that is accountable to drivers of N flow variability described by the considered SMP [1]
	4	Measures of the degree to which an SMP with one or more calibrated coefficients describes processes occurring on the scale for which it uses input. [1]
	5	Measures if there is sufficient relation between modeled N source emissions and outlet N flow in the temporal scale of the SMP
	6	Measure of the validity of the assumptions of the SMP considered about horizontal N dispersion[1,]
	7	Measure of the fraction of total input variation caused by error [1]
C		**Requires adequate resources for modification of the SMP**
	8	Indication of the effort needed for downscaling to obtain inputs from other SMPs, and provide input to other SMPs
	9	Measure of the resources required for an SMP to use input for certain areas, depending on the range of stream orders that dominates in these areas [2]
	10	Measures whether the effects of differences between upstream N emissions in adjacent time steps on outlet N flow are large enough to consider in the model
D		**Requires the prediction scale to correspond to requirements of model users**
	11	Measure of the usefulness of predictions for impact studies and policy makers
	12	Measure of the usefulness of river N export predictions and predictions of source and sink contributions for a fundamental scientific objective of the modeler

[1] This indicator does not apply to SMPs consisting of predictions

[2] This indicator applies only to SMPs modeling aquatic processes

Before application of FAMOS it is necessary that the user sets the upper and lower limits of possible SMP scales (*S*) for which the appropriateness is to be identified. Beyond these limits the user of FAMOS considers scales to be inappropriate. So he/she expects the most appropriate scale to lie within these limits. These limits are necessary for the application of indicators (Section 3.3).

Appropriateness of SMP scale is the result of four nested functions describing: (1) appropriateness of modeling scale according to all criteria, (2) appropriateness of modeling scale according to each criterion, (3) fulfillment of criteria, and (4) indicators.

Appropriateness of modeling scale according to all criteria

Appropriateness according to all four criteria (A-D) is estimated as:

$$A(S) = \prod_{c=A}^{D} a_c(S) \qquad (1)$$

where:

$A(S)$ – is the appropriateness of scale S for all four criteria (0 or 1),
$a_c(S)$ – is the appropriateness of modeling scale S for criterion c (0 or 1).

Here, $A(S)$ is 1 for appropriate scales S and 0 for inappropriate scales S. Appropriateness $A(S)$ is identified largely independent for the different measures of scale, except for the appropriateness of so_{min} and so_{max} which are dependent. Appropriateness $a_c(S)$ takes the value of 1 if S is appropriate for criterion c and 0 if S is inappropriate for criterion c, $c \in \{A,B,C,D\}$, being indices for the four criteria used in FAMOS. The four values of appropriateness of modeling scale S for each criterion c ($a_c(S)$) are multiplied to obtain $A(S)$ (Equation 1).

Equation 1 is used to identify ranges of S that are appropriate. If $A(S)$ is 0 for all *sups*, *supt*, *sl*, *exts, extt*, so_{min}, or so_{max} within the user-defined limits then this may indicate that the considered SMP is not adequate given the research question, resources and available data.

Appropriateness of modeling scale according to each criterion

Estimation of appropriateness according to each of the four criteria ($a_c(S)$) can be summarized as;

$$a_c(S) = B(F_c(S) > T(c)) \qquad (2)$$

where:

$B(..)$ – is a Boolean function. *B(expression)* equals 1 (if *expression* is true), or 0 (if *expression* is false),
$F_c(S)$ – is the fulfillment of criterion c for scale measure S. $F_c(S)$ is real valued on [0,1].
$T(c)$ – is a threshold for appropriateness according to criterion c (0 to 1).

Threshold $T(c)$ is user defined and indicates the degree of fulfillment of criterion c that is considered to be appropriate by the user. Thresholds $T(c)$ are the same for all considered measures of S.

Fulfillment of criteria

Fulfillment of criteria ($F_c(S)$) can be estimated using indicator values. This can be summarized as follows:

$$F_c(S) = \sum_{j=1}^{N_c} v_j \cdot f_{c,j}(S) \qquad (3)$$

where:

$f_{c,j}(S)$ – is the value of indicator j, being the j[th] real valued function of S with values between 0 and 1,
v_j – is a weight indicating the relative importance of indicator j for $F_c(S)$,
N_c – is the number of indicators that is used to calculate the fulfillment of criterion c,

The weights v_j are to be defined by the user of FAMOS. The sum of v_j used in the calculation of $F_c(S)$ is 1.

Indicators

Indicator values ($f_{c,j}(S)$) can be estimated as follows:

$$f_{c,j}(S) = \sum_{i=1}^{N_j} w_{j,i} \cdot g_{j,i}(S) \qquad (4)$$

where:

$g_{j.i}$ – is goal function i of indicator j, the value of which (0 or 1) depends on S,
$w_{j,i}$ – is a weight to indicate relative importance of function $g_{j.i}$ for $f_{c,j}(S)$,
N_j – is the number of instances of $g_{j.i}$ used in the calculation of $f_{c,j}(S)$.

Indicators and goal functions to be used are selected by the user of FAMOS. Indicator j is used to test the fulfillment of a criterion c. We suggest two possible indicators to test the fulfillment of criterion A, five for B, three for C, and two for D (Table 3.2). Only the relevant indicators listed in Table 3.2 need to be used. A user of FAMOS may add case-specific indicators. A goal function ($g_{i,j}$) can only be 0 or 1. Weights $w_{j,i}$ may be defined by the user using methods described in Section 3.3. The sum of $w_{j,i}$ used in the calculation of $f_{c,j}(S)$ is 1.

Five of the indicators presented in this paper cannot be summarized by Equation 4. These are indicators A1, B4, C8, C9, and D12 (Section 3.3).

3.3 Indicators

FAMOS is based on four criteria. We present twelve indicators to test each of these criteria. The user of FAMOS may use these, or a selection of these, to identify appropriate modeling scales. An example of an application of one of these indicators can be found in Chapter 4.

The presented indicators vary between 0 and 1 as a function of *sups*, *supt*, *sl*, *exts*, *extt*, so_{min} and/or so_{max}. Here *sups*, *supt*, and *sl* are the mean surface area, duration, and river section length represented by single values of input variables of an SMP. If the model is stochastic then *sups*, *supt*, and *sl* apply to single input distributions instead of single input values. In FAMOS we use the mean of surface area, time duration and river section length to determine *sups*, *supt*, and *sl* because there may be some variation in these measures in individual inputs. For inputs consisting of grids or regular time series *sups* or *supt* equal the grid cell size or time step, respectively. Measures *exts* and *extt* are the total area and duration, respectively, for which an SMP is applied. Measures *sups* and *exts* are measured as mean diameters. Variables so_{min} and so_{max} are measures of the minimum and maximum sizes of river reaches, respectively, to which an SMP is applied [*Strahler*, 1964] (see also Box 3.1).

3.3.1 Criterion A: Agreement with data and scenarios

Criterion A requires the SMP scale to correspond with that of data and scenarios. It applies to the scale of any SMP except that of model predictions. Criterion A can be tested with indicators A1 and A2 (Table 3.2).

Indicator A1: Artifacts in data

Indicator A1 measures the degree to which artifacts in available input data influence the model representation of process variability. Such artifacts can be the data resolution, density of underlying measurements, strata used in underlying stratified sampling or scales of underlying models. These artifacts cause artificial characteristic scales in the data (see Appendix 1 for an explanation of characteristic scales). The value of indicator A1 increases with the number of such artificial characteristic supports or stream lengths in data that are smaller than the support or stream length of the SMP considered (Equation 1-1 and 1-2, Box 3.2) because inputs then better represent the variability of processes. Indicator A1 is one for extents smaller or equal to the maximum extent where the spatial and temporal coverage of the model fully overlaps with that of the input data (Equation 1-3) because then input data are available for the whole modeled area or time period. Otherwise it is zero. Further, it is one for modeled stream order ranges for which all input data is available (Equation 1-4).

Indicator A2: Scenarios

Indicator A2 measures the fraction of actual spatial and temporal variability of anthropogenic influences described in scenarios simulated with an SMP. A scenario can be a projection of a future situation. Indicator A2 considers only changes that are directly caused by future developments or changing policies at a particular organizational level. The reason is that changes resulting from such drivers are expected to have relatively distinctive characteristic scales in contrast to those resulting from changes in biophysical conditions, which usually take place over a broader scale range [*Holling*, 1992; *Schroeder*, 1991]. Examples of organizational levels in ascending order are persons, households, companies, municipalities, provinces, states, and international organizations or multinationals [*Gibson et al.*, 2000; *Leemans*, 2005]. Although characteristic scales of future changes may also depend on the scale of biophysical conditions, indicator A2 focuses only on the scale of the organizational levels affecting future change because organizational levels usually correspond to a typical range of scale. As organizational level increases, the surface area, time frame and stream order over which measures are implemented increase (Figure 3.2).

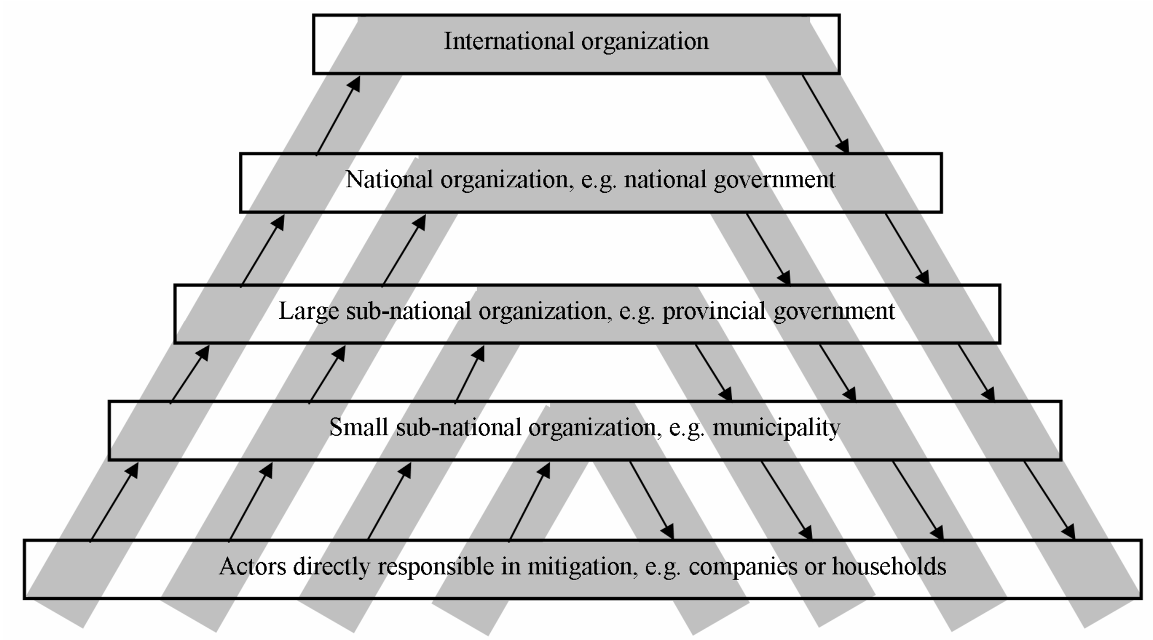

Figure 3.2 Scheme to illustrate that mitigation options involving higher organizational levels require more information transfer with various other organizational levels. Boxes indicate organizational levels. Gray areas highlight each of five alternative paths of information transfer between organizational levels. Upward arrows indicate information transfer from lower to higher organizational levels such as stakeholder involvement, monitoring, and negotiation. Downward arrows indicate information transfer from higher to lower organizational levels such as implementation, education, and maintenance.

The value of indicator A2 increases with the number of projections of a future development for which the characteristic support or stream length is larger than *sups*, *supt* or *sl* (Equation 2-1, Box 3.2). This characteristic support or stream length is the smallest distance between two locations for which a projection of a future development can explain the difference in input value. The value of A2 increases with the number of projections of a future development the extent of which is larger or equal to *exts* or *extt* because this means that more information regarding future developments is available for the whole modeled area or time period (Equation 2-2). The value of indicator A2 also increases with the number of future developments which stream orders are completely modeled by the SMP considered (Equation 2-3). We advise to make weight $w_{2,i}$ used in calculating the values of indicator A2 (Equation 4), proportional to the distinctiveness of the characteristic scale of effects of changes in driver *i* on SMP inputs and its relative importance for predicted N flow.

3.3.2 Criterion B: Validity of model assumptions

Criterion B requires the SMP scale to correspond with model assumptions on processes. It applies to the scale of any SMP except that of model predictions. The fulfillment of criterion B can be tested using indicators B3 to B7 (Table 3.2).

Indicator B3: Represented processes

Indicator B3 measures the fraction of the variability of N flows on the scale of the SMP considered that is actually described by equations of the SMP. This indicator has a lower value if the effect of modeled drivers is obscured by trends or noise that cannot be not be explained by the SMP [in line with *Habeeb et al.*, 2005]. Such trends or noise can occur if a model is applied on a scale where certain drivers are not modeled explicitly or where modeled drivers are less important. Indicator B3 is higher if modeled processes explain a larger part of the N-flow variability. This indicator can therefore be best determined if the considered SMP is process-based.

The value of indicator B3 increases with the number of modeled processes that have predictive value at the support, *sl*, or stream order range of the SMP. It decreases with the number of not explicitly modeled processes that influence predicted N flow at the support, *sl*, or stream order range of the SMP (Equation 3-1 and 3-2,).

The weights involved in the calculation of indicator B3 (Equation 4) are negative if the considered process is not modeled by the SMP and positive if it is. Furthermore, the value of the weights is also proportional to the magnitude of the flow of N generated by the process considered and to the relative degree to which the SMP is nonlinear in the input necessary to model the process considered (Appendix 3.2).

Indicator B4: Validation

Indicator B4 measures the degree to which SMPs with one or more calibrated coefficients can describe processes occurring on the scale for which they use input. Models may make reliable predictions only if the scale of application is equal to the scale at which validation provided good agreement. This is because most existing models for N export from river basins are nonlinear to some degree and, hence, scale specific (Appendix 3.2). If a calibrated SMP was validated independently of other SMPs then indicator B4 only applies to this SMP. Otherwise it applies to all SMPs that were validated as a whole.

For each measure of *S*, indicator B4 has only one function of *S* that is based on the scale of the validation where the lowest prediction error was found. This function takes the value of one if (i) supports or stream lengths are within the range of SMP supports or stream lengths involved in validation (Equation 4-1, Box 3.3) or if (ii) so_{min} and so_{max} that are within the range of the smallest and largest stream orders of the validation data of the SMP, respectively (Equation 4-2). For other *S*, indicator B4 is equal to zero. We advise to make the weight of this indicator in testing criterion B, v_5, proportional to the degree to which the considered SMP is nonlinear (Appendix 2). Weight v_5 is zero for SMPs that did not affect the validation.

Indicator B5: Travel time of N

Indicator B5 measures whether inertia of N flow with respect to modeled N source emissions agrees with the temporal scale of the SMP (Figure 3.3). For steady state models, N emitted by modeled sources must be able to reach the SMP considered within the duration of *supt*. For such models, the value of indicator B5 increases with the number of SMPs modeling N sources from which the system modeled by the SMP considered receives a larger fraction of N within the duration of *supt* (Equation 5-1).

For dynamic models, the N emitted from sources modeled by the SMP considered must be able to reach the prediction locations within *extt*. For such models, the value of indicator B5 increases with the number of SMPs representing predictions where a larger fraction of N emitted from the SMP considered reaches the prediction location within *extt* (Equation 5-2).

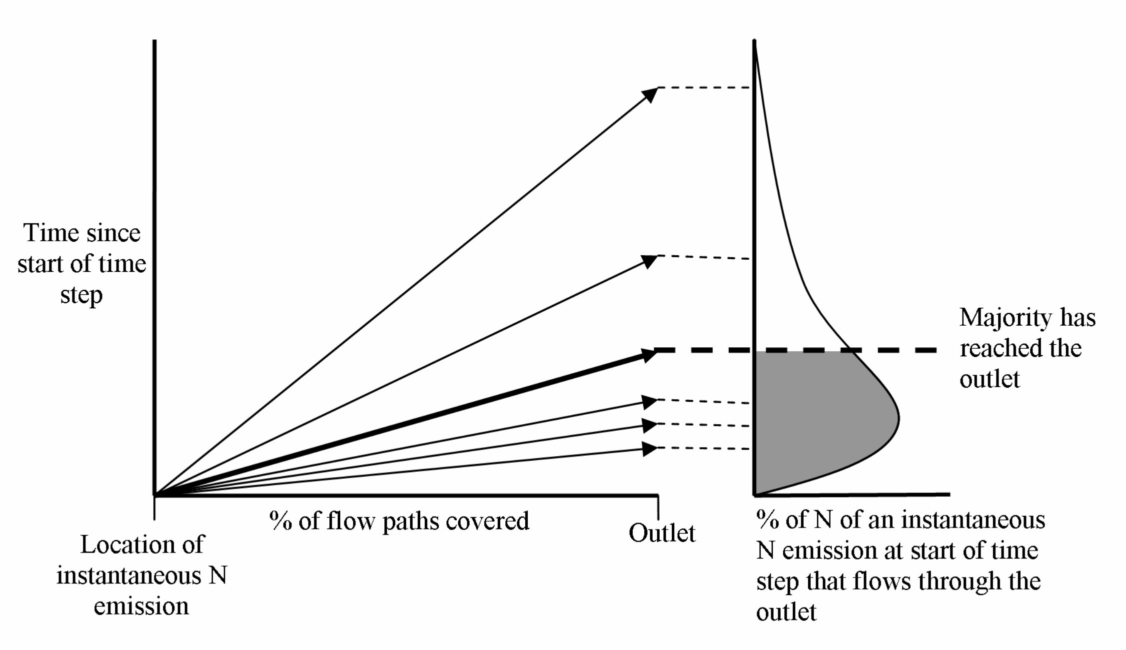

Figure 3.3 Illustration of the times at which different fractions of N, instantaneously emitted from an upstream source, may reach the outlet. Horizontal axis of the left plot is the % of flow paths covered by N affected by an instantaneous upstream N emission and the vertical axis the time that has passed since the instantaneous local emission occurred. Each arrow represents a different fictive path of N through the modeled watershed.

The value of indicator B5 also depends on the stream orders of the SMP considered (Equation 5-1) and of the model predictions (Equation 5-2) because if these are larger then inertia of N flow with respect to modeled N source emissions is larger. Therefore the user must previously select a range of stream orders for the SMP considered based on $a_D(so_{min},so_{max})$ that must be identified before (Equation 2). In case of statically modeled N flow (Equation 5-1), $w_{5,i}$ used in Equation 4 increases with the relative amount of N entering the watershed part represented by the considered SMP from the sources represented by SMP *i*. In case of dynamically modeled N flow (Equation 5-2), $w_{5,i}$ increases with the relative amount of N flowing from the considered SMP to predictions represented by SMP *i*.

Indicator B6: Horizontal N dispersion

Indicator B6 tests whether the assumption of the SMP about horizontal N dispersion is valid. We distinguish three broad classes of assumptions of SMPs about horizontal N dispersion that together cover all possible assumptions about N dispersion. Each of these classes is valid at a different range of *sups* or *exts*. Validity of a dispersion assumption is affected by *sups* if the assumption is about N exchange between areas represented by spatial support units, and validity of a dispersion assumption is affected by *exts* if the assumption is about N exchange across model boundaries. The classes of assumptions that we distinguish are that dispersion is; unidirectional (1), can be approximated as a diffusion process (2), or is negligible (3).

Unidirectional dispersion can be assumed at the smallest *sups* and *exts*, and negligible dispersion can be assumed at the largest *sups* and *exts*. A diffusion approximation is usually valid at intermediate *sups* and *exts*. The transition between scales of validity of assumed unidirectional flow and assumed diffusion behavior is explained by many studies that have shown that the collective behavior of unidirectional flows of matter or organisms can be approximated as a diffusion process [*Levin*, 1992]. The transition between scales of validity of assumed diffusion behavior and assumed negligible dispersion occurs because at sufficiently coarse scales many N flows with inverse directions cancel out each other's effect [*Easterling and Kok*, 2002]. The absolute *sups* or *exts* values in the ranges where these three assumptions are valid differ between N types. They increase with the most common horizontal travel distances of the N type considered. If *sups* or *exts* are within the range of the most common travel distances then probably N dispersion can be approximated as a diffusion process. If *sups* or *exts* are smaller then this range then dispersion can probably be assumed to be unidirectional and if larger then it can probably be assumed to be negligible. An example of a list of N types that is ordered by increasing *sups* and *exts* of validity of the previously described dispersion assumptions are N in; litter fall, eroding soil particles, cattle biomass, floodplain sediment pores, sewerage pipes, wetland surface water, groundwater, salmon biomass, lower atmosphere, and higher atmosphere.

The value of indicator B6 increases with the number of modeled N types where *sups* or *exts* is such that the SMP assumption on horizontal N dispersion is valid (Equation 6-1 and 6-2). An assumption of unidirectional flow can only be used if the *supt* or *extt* are smaller then the time needed for a significant flow direction change of the N type considered [*Dumont et al.*, submitted-a] and the range of *sups* and *exts* where a diffusion approximation or assumed negligible flow is valid increases with decreasing *supt* and *extt* (Equation 6-1 and 6-2). To increase ease of use of indicator B6, the appropriate *supt* and *extt* are identified with other indicators before indicator B6 is applied (Section 3.4). We advise to increase the weight of an N type used to calculate indicator B6 ($w_{6,i}$) with the relative magnitude of spatial variation in horizontal dispersion of the considered N type. We advise to increase the weight of indicator B6 in testing criterion B (v_6) with the degree of nonlinearity of the SMP considered (Appendix 3.2).

Indicator B7: Input estimation uncertainty

Indicator B7 measures the fraction of the total input variation caused by error. The effect of some errors on uncertainty in model inputs may decrease if the SMP support is increased (aggregation) beyond particular thresholds. An example of such errors is a classification error in remote sensed images that may be largely removed by aggregating multiple pixels, where pixels with positive classification errors compensate for those with negative errors. Another example are fine scale errors in data resulting from multi-scale models where model calculations at a large support resulted in boundary conditions for model calculations with a smaller support, thus propagating their errors to these smaller supports [*De Floriani and Magillo*, 2002; *De Koning et al.*, 2000]. Averaging out of such errors applies to space, time, or both. The value of indicator B7 increases with the number of SMP inputs for which considered errors are averaged out. (Equation 7-1, Box 3.3). Indicator B7 considers only error sources leading to errors that average out at relatively small *sups*.

Scale measures *sups* and *supt* interact in affecting the value of indicator B7 (Equation 7-1). Thus for simplicity the appropriate *supt* is identified with other indicators before indicator B7 is applied (Section 3.4). We advise to increase the weights of inputs used in the calculation of indicator B7 ($w_{7,i}$) with the fraction of total variation of the input that is caused by the considered error.

3.3.3 Criterion C: Resources for modification

Criterion C requires that available resources are sufficient for modification of the SMP. This criterion can be tested using indicators C8, C9 and C10 and is used only when the appropriate scale cannot be determined with criteria A, B, and D. This is generally the case when the SMP is not process-based or has a linear response to variations in input data. Threshold *T(C)* has a relatively high value for such SMPs, whereas *T(A)*, *T(B)*, and *T(D)* are low.

Indicator C8: Resources for downscaling between SMPs

Indicator C8 is an indication of the effort needed for downscaling to obtain inputs from other SMPs and to provide input to other SMPs. Downscaling is needed if outputs of coarse scale SMPs are used as input in fine scale SMPs. This downscaling requires data and time, and the level of these requirements is measured by indicator C8. For example, an increase in spatial SMP support can cause an increase in the need for downscaling of SMP outputs (Figure 3.1C). This would also be the case for a change of temporal SMP support. As a consequence, changed combinations of *sups* and *supt* of SMPs can require different investments in data and time needed for downscaling.

The value of indicator C8 is high if no downscaling is required and low if many resources are needed for downscaling. To indicate how much the support of the considered SMP contributes to required data and time for downscaling, it is necessary to first identify what support will be used for the connected SMPs. Furthermore, it is only useful to quantify indicator C8 for an SMP for which the scale can be adjusted to that of other SMPs when this improves the overall appropriateness of its scale (*A(S)*). This is generally the case only if criterion C is the most important one for this SMP (expressed by relatively high *T(C)*), i.e. when the appropriate scale depends less on the criteria A, B and D for SMPs that are not process-based or when variations in scale of the SMPs considered are not expected to cause large changes in prediction error (for linear or only slightly nonlinear SMPs).

Indicator C8 can only be applied if the appropriate scales according to criteria A, B, and D have already been identified for all connected SMPs. The value of indicator C8 increases with the number of connected SMPs for which probably no downscaling is required to provide input to the considered SMP or to obtain input from the considered SMP (Equation 8-1 to 8-3, Box 3.4). Weights $w_{8,r}$ and $w_{8,p}$ indicate the relative importance of each SMP connected to the considered SMP. They increase with the amount of effort needed in downscaling output from a connected SMP or in downscaling to provide input to a connected SMP.

Indicator C9: Resources for adjusting modeled stream order ranges

Indicator C9 measures the resources that are required for SMPs to change the *sl* and range of stream orders that they model. It applies only if the considered model has already been applied in the area of study, and if FAMOS is used to improve the scale of model application. Indicator C9 is zero if resources are needed to adjust the connections to other SMPs. Otherwise it has a value of one (Equation 9-1, Box 3.4).

Particular stream orders drain particular parts of the land area. Therefore, an aquatic SMP representing a particular range of stream orders is connected only to those instances of a terrestrial SMP that represent the part of the land surface that drains to these stream orders. Comparison of Figure 3.1A and D shows that if the range of stream orders of the dark grey SMP is changed, it receives input from the upper instance of the light grey terrestrial SMP, whereas this was not so before. Depending on the type of model database and software that is used, the latter type of alterations in connections between SMPs can require a relatively large effort.

Indicator C9 applies only to an SMP which scale can be expressed in stream order or *sl*.

Indicator C10: Mixing

Indicator C10 measures whether the effects of differences between upstream N emissions in adjacent time steps on outlet N flow are large enough to consider in the model. These effects are small if the temporal scales of N emission are fast and travel times differ to a large extent due to e.g. mixing (Figure 3.4) [*Sterman*, 2000]. Examples of places in watersheds where N fluxes generated at different times can mix are reservoirs, converging streams, aquifers, etc. In larger watersheds more mixing can take place along N flow paths. Therefore, dynamics in outlet N flow of larger watersheds can generally be better explained with N source emissions on a larger *supt*. The value of indicator C10 increases with the number of dynamically modeled N flows having a range travel times to prediction locations that is narrower then *supt* (Equation 10-1). This range is the interval of travel times around the mean travel time and is given by a user defined fraction (*UDF*) of N flow from the SMP considered that reaches prediction locations.

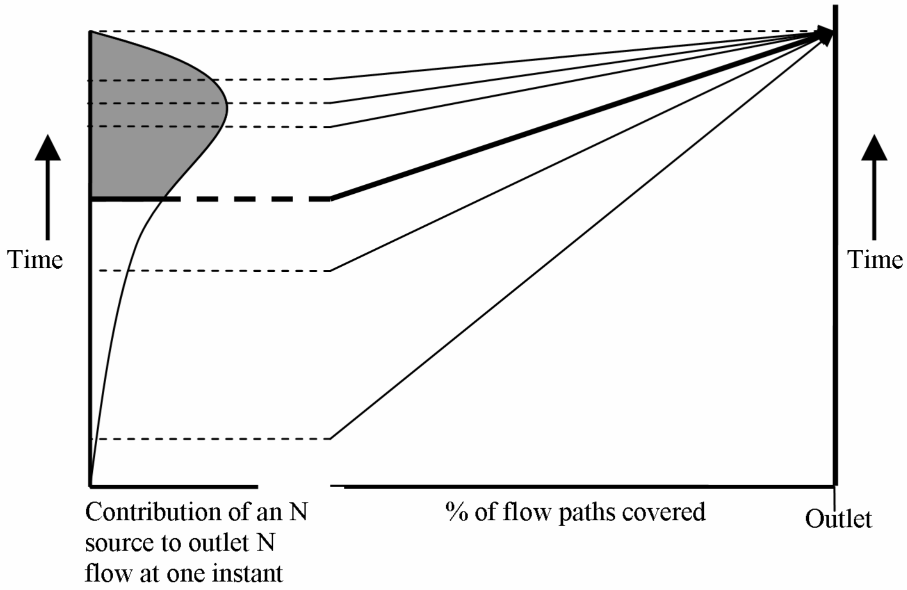

Figure 3.4 Illustration of the degree to which different times of upstream N emissions may affect outlet N flow at one point in time. Horizontal axis of the right-hand plot is the % of flow path covered by N from upstream N emissions and the vertical axis the time that has passed since the N emissions occurred. Each arrow represents a different fictive path of N through the modeled watershed.

The value of indicator C10 also depends on the stream orders of the model predictions (Equation 10-1) because if these are larger, then larger *supt* is required for the differences between the effects of N emission in adjacent time steps on predicted N flux to be large enough. Therefore the user must previously select a range of stream orders for the SMP considered based on $a_D(so_{min},so_{max})$ (Equation 2) that must be identified before (Section 3.4). Weight $w_{10,i}$ used in Equation 4 increases with the relative amount of N flowing from the considered SMP to prediction locations represented by SMP *i*.

3.3.4 Criterion D: Usefulness of predictions

Criterion D requires that the prediction scale is such that model predictions are sufficiently useful. Criterion D can be tested using indicators D11 and D12.

The usefulness of river N export predictions depends on the type of use. Two types of users are distinguished. Indicators D11 and D12 focus on a different use: impact studies and policy development using N flow predictions to assess environmental impacts (indicator D11); scientists interested in fundamental questions related to the processes in river basins (indicator D12). For these two uses, the usefulness of predictions can be tested with indicators D11 and D12.

Indicator D11: Usefulness for impact studies and policy development

Indicator D11 is a measure of the usefulness of river N export predictions for impact studies and policy development at different scales.

An impact study can be a known research project that uses the predictions of river N export as input or it can be a research project that uses information on the predictions that is made available in public media. The latter prediction users are unknown, so in this case the user of FAMOS needs to assess the probable scale at which scientists investigating aquatic impacts require input data on N inflow from rivers. Policy makers may directly use N flow predictions for their policy. Policy makers require predictions of the scale at which legislation allows and conventions drive mitigation of impacts [*Sherman*, 1991].

For SMPs that are predictions, indicator D11 increases with the number of prediction users for whom predictions have a smaller *sups*, *supt* or *sl* than the grain of their study or interest (Equation 11-1, Box 3.5). Grain is the smallest scale at which a prediction user responds to predicted patch structure by differentiating among patches [*Kotliar and Wiens*, 1990]. Grain increases with the organizational level of the prediction user (e.g. individual, municipal, or national level). Indicator D11 increases with the number of prediction users requiring predicted N flow over an area or time period smaller than *exts* and *extt* of predictions (Equation 11-2, Box 3.5). Also it increases with the number of prediction users requiring predicted N flow into aquatic systems directly receiving N from stream orders within the range of stream orders for which predictions are made (Equation 11-3, Box 3.5).

For SMPs that can model mitigation options, indicator D11 measures the fraction of spatial and temporal variability of relevant mitigation options that can be represented by the SMP considered. Characteristic scales of mitigation options usually depend on the scale of the organizational level that coordinates the mitigation (see description if indicator A2). For SMPs modeling mitigation options, indicator D11 increases with the number of prediction users which study mitigation options having characteristic supports or stream lengths larger than *sups*, *supt* or *sl* (Equation 11-1, Box 3.5). The characteristic support or stream length of a mitigation option is the smallest distance between two locations for which a mitigation option can generally explain the difference in input values. The value of D11 increases with the number of prediction users which study mitigation options exclusively in an area or time period that is smaller or equal to *exts* or *extt* because this means that more of the required information about future effects of mitigation is modeled (Equation 11-2). The value of indicator D11 also increases with the number of prediction users which study mitigation options that mostly take place in stream orders modeled by the SMP considered (Equation 11-3). For each considered SMP modeling mitigation options, only the mitigation options are considered that can be modeled by this SMP. For such SMPs, we advise to make $w_{11,i}$ used in calculating the values of indicator D11 (Equation 4) proportional to relative contribution to predicted N flow of mitigation options studied by prediction user *i*.

Indicator D12: Usefulness for fundamental research

Indicator D12 is a measure of the usefulness of river N export predictions and predictions of source and sink contributions for a fundamental scientific objective. We assume that predictions are more valuable for fundamental research if they can describe variation on more scales. As indicated in the literature [e.g. *Meybeck*, 2002], new variability generally appears at about a constant rate if the logarithm of support is decreased at a constant rate or if the logarithm of extent is increased at a constant rate.

Therefore the value of indicator D12 decreases linearly with the logarithm of support and increases linearly with the logarithm of extent (Equation 12-1 and 12-2, Box 3.5).

3.4 Use of FAMOS

The indicators need to be used in a particular sequence because some indicators require information on the appropriateness of certain measures of scale that should be supplied by certain other indicators that have been applied earlier. FAMOS is first applied for process-based and nonlinear SMPs. For such SMPs, first indicators A1, A2, B3, B4, D11 and D12 are applied, followed by indicators B5 and C10, and finally indicators B6 and B7 enabling the calculation of *A(S)*. This is followed by the application of FAMOS to linear or non-process-based SMPs in the same sequence except that indicators C8 and C9 are used in addition for such SMPs. The reasons for this sequence of use of indicators and SMPs in FAMOS are:

- Stream order (so_{min} and so_{max}), and *supt* interact in determining the value of indicators B5 and C10, resulting in a different value of these indicators for every combination of stream order and *supt*. Therefore so_{min} and so_{max} are set to a fixed value, being the most appropriate combination of so_{min} and so_{max} according to preliminary values of $F_c(S)$. These values need to be based on other indicators that have been applied earlier.
- Measures of temporal and spatial modeling scale interact in determining the value of indicators B6 and B7, resulting in a different value of these indicators for every combination of temporal and spatial modeling scale. Therefore *supt* and *extt* are set to fixed values being the most appropriate *supt and extt* according to preliminary values of $F_c(S)$. These values need to be based on other indicators that have been applied earlier.

Appropriateness *A(S)* needs to be known for all process-based and nonlinear SMPs before indicators C8 and C9 can be applied to the remaining SMPs that are linear or non-process-based. Therefore *A(S)* is first identified for SMPs that are process-based and nonlinear.

3.5 Concluding remarks

We present an extension of current methods for identification of the appropriate modeling scales of N-flow models. Current methods are often based on only one criterion for appropriateness of scale and often focus on only one SMP and only one measure of scale. This may lead to an identified appropriate scale with respect to only one aspect of the model quality at the expense of another aspect of model quality that is not taken into account. FAMOS includes all aspects of model quality that are affected by modeling scale and thereby enables its user to rigorously judge the appropriateness of modeling scale. It also prevents the modeler from focusing on only one criterion by making him aware of other criteria of appropriateness, other SMPs, and other measures of scale. To this aim, FAMOS uses four criteria to identify appropriate spatial and temporal scales. Those, in turn, are tested by 12 user-defined indicators. The four criteria constitute a comprehensive basis for identifying the appropriateness of the modeling scale of any watershed N-flow model. Weights can be assigned to each criterion thus allowing FAMOS to take into account many different objectives. The indicators proposed in this paper appear to be generally applicable as well as a sufficient basis to identify the appropriate scale.

The distinction in FAMOS between SMPs increases the reliability of estimates of appropriateness of modeling scales, because it enables FAMOS to account for interactions between SMPs. These interactions are that the scale of an SMP affects the appropriateness of scales of other SMPs (indicator C8). Also the distinction between different measures of spatial and temporal scale in FAMOS increases reliability of estimates of appropriateness of modeling scales. It enables FAMOS to account for interactions between these measures. These interactions are that the appropriateness of values of a measure of SMP scale is usually affected by values chosen for another measure of SMP scale (indicators B5, B6, B7, C10).

FAMOS is sensitive to user-defined weights ($w_{i,j}$, v_j) and thresholds ($T(c)$). Appropriate values of these parameters cannot be known from experiments or measurements, but require thorough knowledge of the model and FAMOS. If knowledge of either of them is incomplete, then the sensitivity of FAMOS to these weights and thresholds will make estimates of appropriate scales unreliable. The sensitivity of estimated appropriateness of different scales ($A(S)$) to weights and thresholds is especially high if the values of two indicators or goal functions indicate ranges of appropriate scales that do not overlap. In that case, small changes in weights may cause scales to be identified as appropriate that were previously identified as inappropriate and vice versa. For example, if indicator B7 indicates that inputs with spatial supports smaller than 10 km are too uncertain and indicator B3 indicates that the dominating processes can be well modeled within a certain range of supports smaller than 10 km, then the ratio between v_3 and v_7 determines whether spatial supports under 10 km are appropriate or inappropriate according to these two indicators. So in such a case, it is important that the user of FAMOS can compare losses of model quality due to uncertainty at supports smaller than 10 km and the model quality gains due to better modeling of dominant processes at these supports. This requires sufficient knowledge about the development of model equations and methods that were used to generate input datasets.

The thresholds *T(A)* to *T(C)* are able to inform the modeler on how the model needs to be improved. In case the model scales are not all appropriate, a threshold that is not exceeded is an indication to what methods need to be used to improve the model. If it is *T(A)* then better scaled input data are needed; if it is *T(B)* then some model equations need to be scaled; and if it is *T(C)* it means that more appropriate knowledge or software needs to be obtained for scaling.

Appendix 3.1: Characteristic scale

Characteristic scales are those with a maximum in their change rate of variability as a function of scale (Figure A1c). Characteristic scale can usually be estimated with more certainty for processes that are more clumped or patchy. Patchiness can occur in time [*Platt and Denman*, 1975], space [*Holling*, 1992] and over stream lengths [*Torgersen et al.*, 2004]. We define a patch as a region or time interval that differs from its surroundings or neighboring times, respectively, but is not necessarily internally homogeneous [*Kotliar and Wiens*, 1990]. Characteristic scales can be identified for time, space, and stream lengths. Characteristic spatial supports, characteristic stream lengths, and characteristic spatial extents are those equaling a patch size with a relatively high spatial coverage. Characteristic temporal supports and characteristic temporal extents are those equaling a patch duration with a relatively high temporal coverage.

Examples of a process at a number of its characteristic supports in space are:

- N fertilizer application per private property, municipality, province, and country (Figure A1).
- N uptake per crop, agricultural parcel, and agricultural area.

An example of a process with a number of characteristic supports in time is N discharge in waste water which may have three characteristic temporal supports influenced by daily waste discharge cycles, seasonal waste discharge cycles, decadal construction/abandonment of waste discharge points.

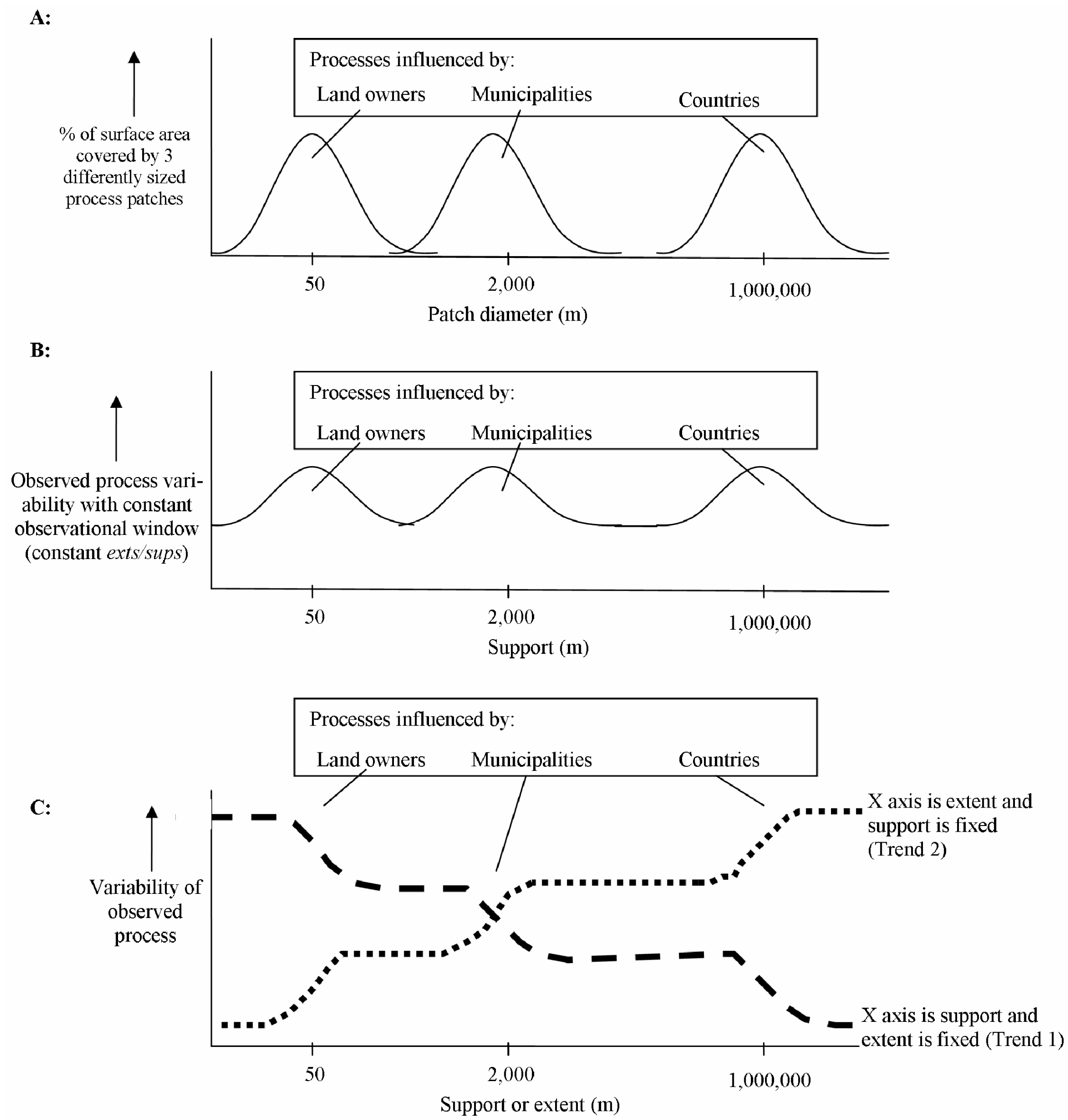

Figure A1 Illustration of the characteristic scales of a process influenced by three organizational levels: land owners, municipalities and countries. A) Trend in percentage coverage of likely process patches as a function of their diameter. Characteristic scales are those patch diameters with the largest coverage. B) Trend in variability of observed process levels as a function of support while maintaining a constant observational window (ratio of extent to support). Characteristic supports are those corresponding to the 3 dominant patch sizes. C) Trend in observed process variability as a function of support when extent is fixed (Trend 1), and as a function of extent when support is fixed (Trend 2). Characteristic scales can be recognized by a relatively large decrease in slope in Trend 1 and a large increase in slope in Trend 2.

Appendix 3.2: Relation between SMP support and prediction bias

Generally, data connections between SMPs in watershed N-flow models involve aggregation of SMP outputs to obtain inputs for other SMPs. The reason is that the modeling of N flow in watersheds requires equations to describe flow paths of N starting at various locations on the land surface that eventually converge at one point, i.e. the outlet. These convergences of N-flow paths, as the outlet is approached, are reflected in N-flow models as aggregations of SMP outputs. If SMPs are applied to a support for which they were not developed, the output may be incorrect. However, the aggregation of these outputs to obtain inputs for another SMP removes much of the error if the former SMP is linear. However, if the SMP for which the output has been aggregated is nonlinear, the obtained input for the subsequent SMP is biased [in line with *Bierkens et al.*, 2000; *Heuvelink*, 1998]. An SMP is nonlinear if it contains input variables that are correlated to each other, raised to a power, or part of a logarithm. We consider an entire SMP to be nonlinear if it is nonlinear in at least one input. Nonlinearity causes SMP support to have a larger effect on aggregated output error, depending on the effect of SMP support on the probability density function (PDF) of its inputs [*Isaaks and Srivastava*, 1998] (Figure A2). This, in turn, leads to a change in the mean output of nonlinear SMPs (Figure A3) and hence to a change in error in mean outputs used as predictions or as inputs to other SMPs.

Changed SMP support does not always lead to changed input PDFs because this also depends on the measurement support being the average surface area or time duration over which a measurement is made [*Beckie*, 2001]. If the measurement support of an input is larger than the SMP support, then the PDF of the input can only be changed by changing the measurement support of this input (Figure A4). However, if the SMP support is larger than the measurement support of its input, then the PDF of the input can be adjusted by changing the SMP support. This can be understood from Figure A5, where the PDF of the inputs is narrower and higher than the PDF of the underlying measurements due to an SMP support that is larger than the measurement support of its input.

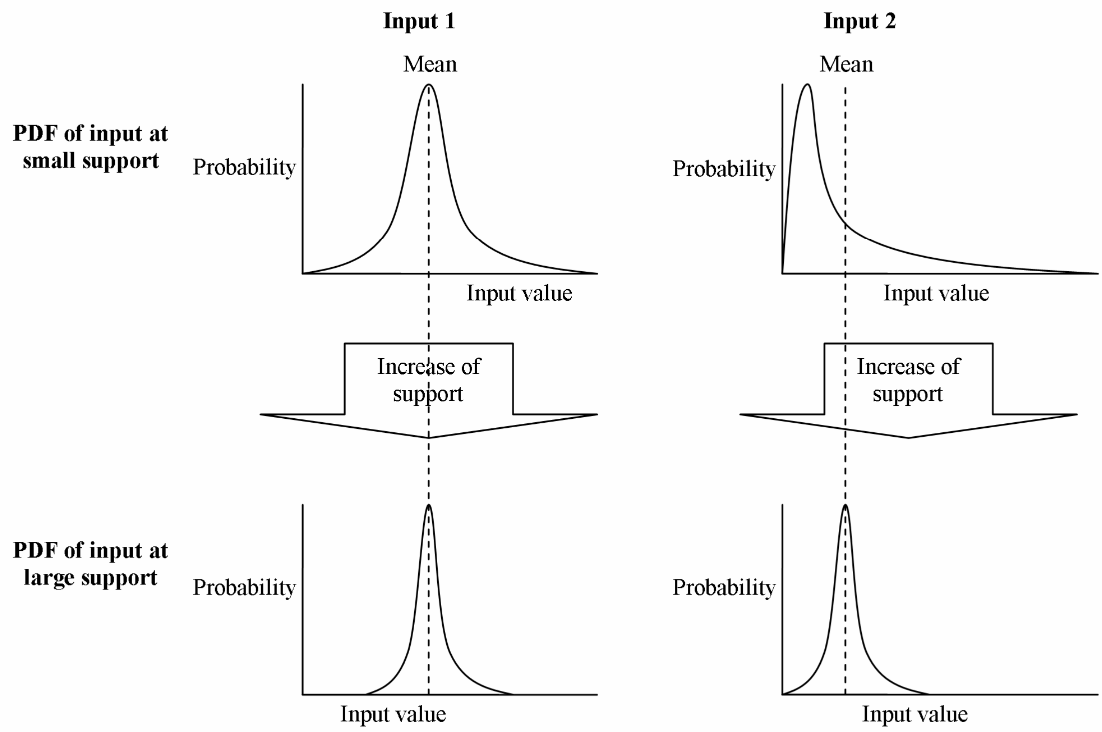

Figure A2 Illustration of how change of model support can change possible probability density functions (PDFs) of model inputs. Many real phenomena have PDFs similar to those of inputs 1 and 2. Examples are temperature and river discharge in a river basin, respectively. The PDF of Input 1 becomes narrower and that of Input 2 becomes both less skewed and narrower if support is increased.

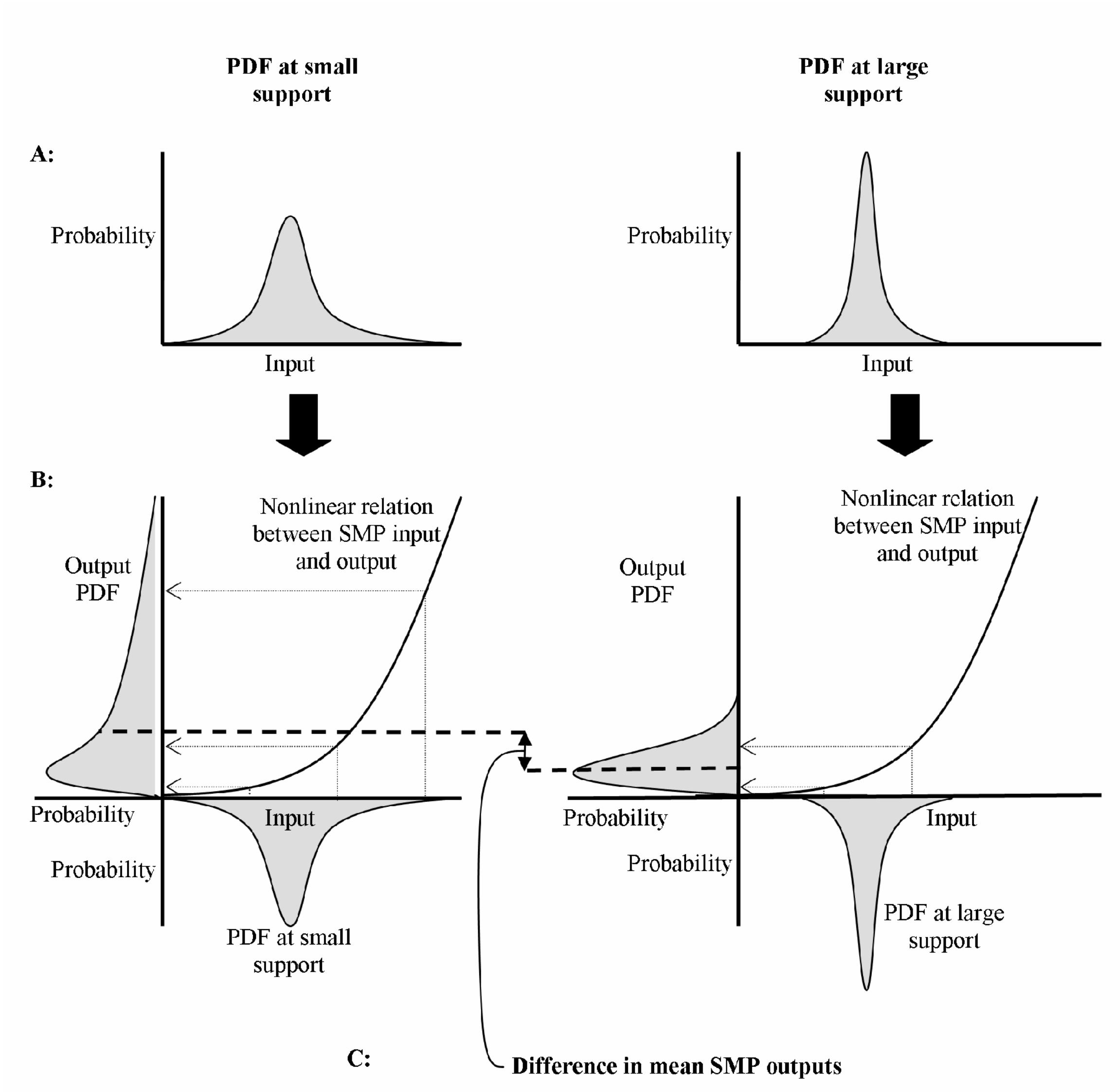

Figure A3 Illustration of how support changes of a nonlinear SMP can cause change in mean output. A) PDFs of inputs that differ due to differing input supports: The PDF for the small support is broader then that for the larger support. B) The different input PDFs in A lead to different output PDFs of a nonlinear SMP. C) These differing output PDFs have differing means.

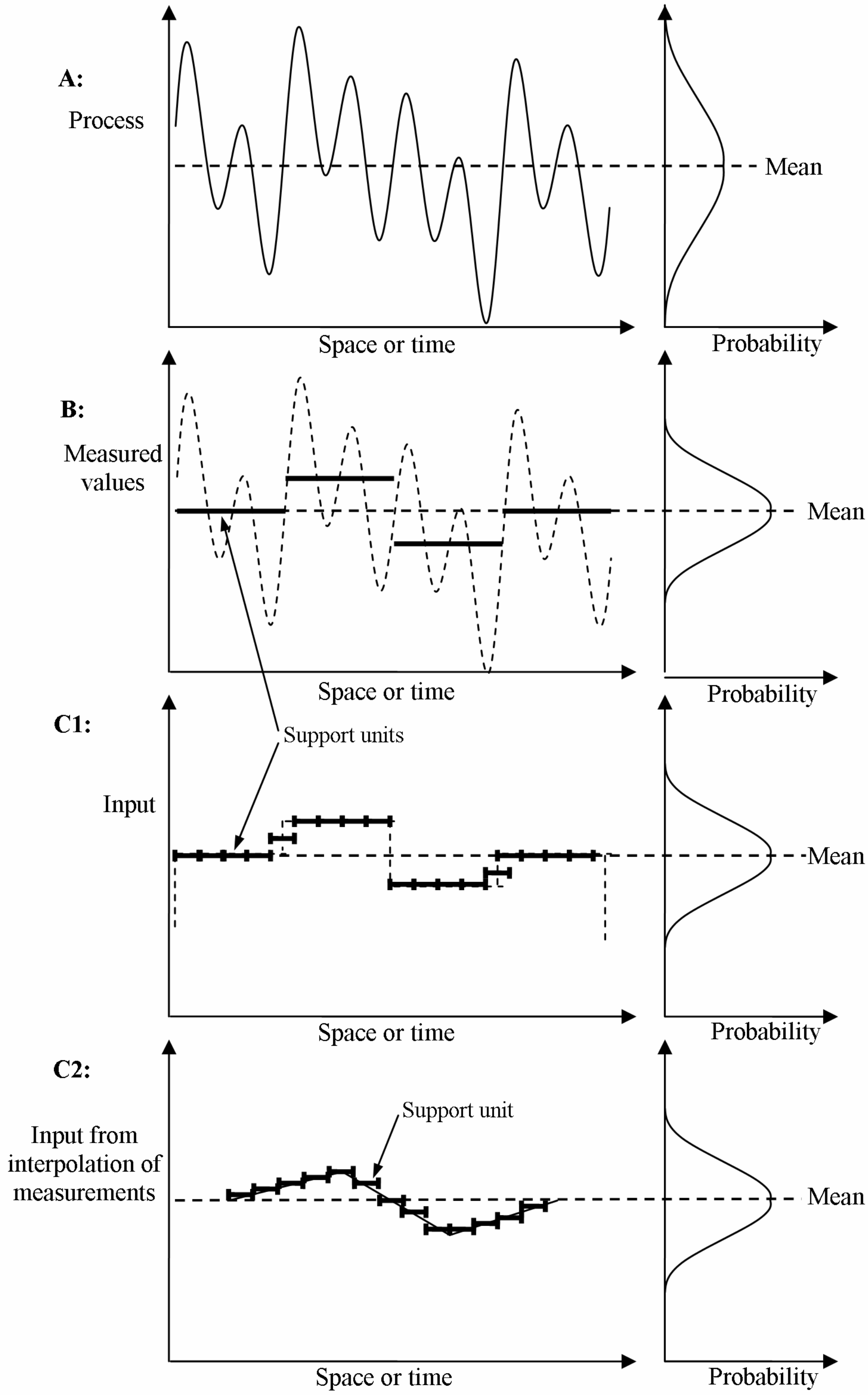

Figure A4 Illustration of the dependence of the relation between PDFs for processes, measurements, and model inputs, on the support used for measurements and inputs. The PDF of input values does not differ from PDF of measurements, from which these inputs are derived, because the support of the inputs is smaller then the support of the measurements. The support units are indicated with fat horizontal lines. A: Levels of a process. B: Measurements of this process. C: Input levels derived from these measurements. D: Input levels derived from the measurements after interpolation.

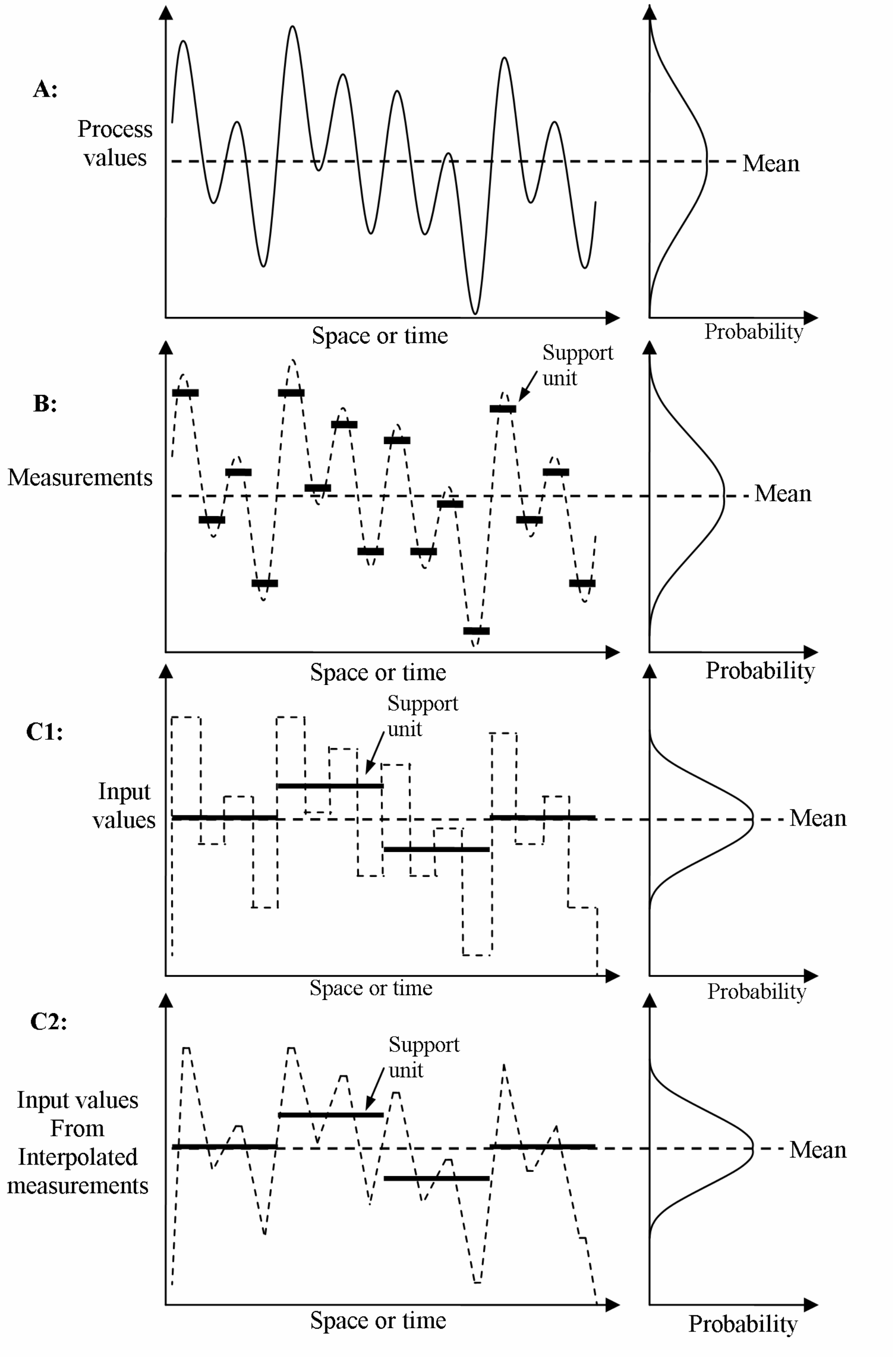

Figure A5 Illustration of the dependence of the relation between PDFs for processes, measurements, and model inputs, on the support used for measurements and inputs. In contrast to
Figure A4, the PDF of input values becomes more slim and high than the PDF of measurements from which these inputs are derived because here the support of the inputs is larger then the support of the measurements. The support units are indicated with fat horizontal lines. A: Levels of a process. B: Measurements of this process. C: Input levels derived from these measurements. D: Input levels derived from the measurements after interpolation.

Appendix 3.3: Boxes

Box 3.1 Definition of measures of scale

sups	=	Mean spatial support of inputs of the considered SMP, where spatial support is measured as the diameter
supt	=	Mean temporal support of inputs of the considered SMP
sl	=	Mean length of stream represented by individual input values of the SMP considered
exts	=	Spatial extent of inputs of the considered SMP, measured as the diameter
extt	=	Temporal extent of inputs of the considered SMP
so_{min}	=	If data on stream properties is used in modeling of N flow in streams: Smallest stream order for which the considered SMP uses input If only proxies for stream properties are used in modeling N flow in streams: Smallest stream order that commonly occurs in the areas from which the considered SMP receives input
so_{max}	=	If data on stream properties is used in modeling of N flow in streams: Largest stream order for which the considered SMP uses input If only proxies for stream properties are used in modeling N flow in streams: Largest stream order that commonly occurs in the areas from which the considered SMP receives input
$range(so)$	=	The set of stream orders between and including so_{min} and so_{max}
S	=	Modeling scale of an SMP, measured by $sups$, $supt$, sl, $exts$, $extt$, so_{min}, or so_{max}

Box 3.2 Mathematical formulation of indicator A1, and goal functions of indicator A2 used in testing criterion A.

Indicator A1: Artifacts in data

Remark: Applies not to SMPs representing predictions

$$f_{A,1}(S) = \frac{1}{N^{data}} \cdot \sum_{i=1}^{N^{data}} \frac{1}{N_i^{art}} \cdot \sum_{h=1}^{N_i^{art}} B\left(S \geq S_{i,h}^{data}\right) \text{ for } S = sups\text{, or } supt \quad (1\text{-}1)$$

$$f_{A,1}(sl) = \frac{1}{N^{data}} \cdot \sum_{i=1}^{N^{data}} \frac{1}{N_i^{art}} \cdot \sum_{h=1}^{N_i^{art}} B\left(sl > sl_{i,h}^{data}\right) \quad (1\text{-}2)$$

$$f_{A,1}(S) = \prod_{i=1}^{N^{data}} B(S \leq S_i^{data}) \text{ for } S = exts\text{, or } extt \quad (1\text{-}3)$$

$$f_{A,1}(so_{min} so_{max}) = \prod_{i=1}^{N^{data}} B(range(so) \subset range(so_i^{data})) \quad (1\text{-}4)$$

$S_{i,h}^{data}$	=	Characteristic spatial or temporal data support in available input dataset *i* due to artifact *h*
$B(..)$	=	Boolean function. *B(expression)* equals 1 (if *expression* is true), or 0 (if *expression* is false)
N^{data}		Number of input datasets considered in the calculation of indicator A1
N_i^{art}		Number of artifacts considered for input dataset *i*
$sl_{i,h}^{data}$	=	Characteristic stream length in available input dataset *i* due to artifact *h*
S_i^{data}	=	Maximum spatial or temporal extent where the spatial coverage of the model fully overlaps with that of input dataset *i*
$range(so_i^{data})$	=	Range of stream orders for which input dataset *i* is available

Indicator A2: Scenarios

Remarks:
Equations 2-1 to 2-3 need to be inserted into equation 4 to calculate the value of indicator A2. Applies only changes described by scenarios partly occurring in the system represented by the considered SMP. Applies not to the SMP representing predictions.

$$g_{2,i}(S) = B\left(S \geq S_i^{a,scen}\right) \text{ for } S = sups\text{, } supt\text{, or } sl \quad (2\text{-}1)$$

$$g_{2,i}(S) = B\left(S \leq S_i^{b,scen}\right) \text{ for } S = exts \text{ or } extt \quad (2\text{-}2)$$

$$g_{2,i}(so_{min}, so_{max}) = B(range_i(so^{scen}) \subset range(so)) \quad (2\text{-}3)$$

Continued

Box 3.2 continued		
i	=	unique combination of anthropogenic driver and its institutional level
$S_i^{a,scen}$	=	Characteristic spatial support, temporal support or stream length of changes in variables of the considered SMP resulting from each *i* that is described in a scenario. (for *S* = *sups*, *supt* or *sl*, respectively)
$S_i^{b,scen}$	=	area or time period where the influence of *i* on the system represented by the SMP is projected or known (for *S* = *exts* or *extt*, respectively)
$range_i(so^{scen})$	=	Range of stream orders where inputs of the considered SMP are most affected by changes in *i* described in the considered scenario

Box 3.3 Mathematical formulation of goal functions of indicators B3 to B7 used in testing criterion B.

Indicator B3: Represented processes

Remarks:
Equations 3-1 and 3-2 need to be inserted into Equation 4 to calculate the value of indicator B3. Applies only to SMPs that are not calibrated. Applies not to the SMP representing predictions

$$g_{3,i}(S) = B\left(S \in range(S_i^{proc})\right) \text{ for } S = sups, supt, \text{ or } sl \tag{3-1}$$

$$g_{3,i}(so_{min}, so_{max}) = B(range(so) \subset range(so_i^{proc})) \tag{3-2}$$

$range(S_i^{proc})$	=	Range of spatial support, temporal support or stream length where process *i* has predictive value (for *S* = *sups*, *supt* or *sl*, respectively)
$range_i(so^{proc})$	=	Stream order range where process *i* has predictive value

Indicator B4: validation

Remark: Indicator B4 applies to al SMPs that, individually or commonly, affected the outcome of a validation. Applies not to the SMP representing predictions

$$f_{B,4}(S) = B\left(S \in range\left(S^{val}\right)\right) \text{ for } S = sups, supt, \text{ or } sl \tag{4-1}$$

$$f_{B,4}(so_{\min}, so_{\max}) = B\left(range(so) = range\left(so^{val}\right)\right) \tag{4-2}$$

$range\left(S^{val}\right)$	=	Range of surface areas, time durations or stream lengths for which an observation was used during validation (for *S* = *sups*, *supt*, or *sl*, respectively)
$range\left(so^{val}\right)$	=	Total range of stream orders occurring in the areas for which the considered SMP received input during validation

Indicator B5: Travel time of N

Remarks:
Equations 5-1 and 5-2 need to be inserted into equation 4 to calculate the value of indicator B5.

Equation 5-1 applies for N flow paths modeled statically:

$$g_{5,i}(supt, range(so)) = F_i^{in}(supt, range(so)) \tag{5-1}$$

$F_i^{in}(supt, range(so))$	=	Fraction of N eventually reaching the locations modeled by the SMP considered from sources modeled by SMP *i* that has reached these locations within the duration of *supt*. Scale measure *range(so)* should have been determined in advance (Section 3.4) because it affects the location of the system modeled by the considered SMP
i	=	Index for SMPs modeling systems through which N flows before it reaches the location modeled by the SMP considered

Continued

Box 3.3 continued

Equation 5-2 applies for N flow paths modeled dynamically:

$$g_{5,i}(extt, range(so)) = F_i^{out}\left(extt, range(so)\right) \quad (5\text{-}2)$$

$F_i^{out}\left(extt, range(so)\right)$ = Fraction of N eventually reaching location *i* where model predictions are required that can flow within the duration of *extt* from the location modeled by the SMP considered to location *i*. Scale measure *range(so)* should have been determined in advance (Section 3.4) because it affects the location of the system modeled by the considered SMP.

i = Index for locations where model predictions are required

Indicator B6: Horizontal N dispersion

Remarks:
Equation 6-1 and 6-2 need to be inserted into equation 4 to calculate the value of indicator B6.
Applies not to SMPs representing predictions.

$$g_{6,i}(sups, supt) = B\left(sups \in range_i^{assump}\left(supt\right)\right) \quad (6\text{-}1)$$

$$g_{6,i}(exts, extt) = B\left(exts \in range_i^{assump}\left(extt\right)\right) \quad (6\text{-}2)$$

$range_i^{assump}\left(supt\right)$ Range of *sups* where the assumption of the SMP about dispersion of N type *i* between areas represented by spatial support units is valid for a fixed *supt*. *supt* interacts with *sups* in affecting validity of dispersion assumptions but for simplicity its value is based on other indicators as described in Section 3.4.

$range_i^{assump}\left(extt\right)$ = Range of *exts* where the assumption of the SMP about dispersion of N type *i* across model boundaries is valid for a fixed *extt*. *extt* interacts with *exts* in affecting validity of dispersion assumptions but for simplicity its value is based on other indicators as described in Section 3.4.

Indicator B7: Input estimation uncertainty

Remark:
Equation 7-1 needs to be inserted into equation 4 to calculate the value of indicator B7.
Applies not to SMPs representing predictions.

$$g_{7,i}(sups, supt) = B\left(sups > sups_i^7\left(supt\right)\right) \quad (7\text{-}1)$$

$sups_i^7(supt)$ = *sups* at which errors of a considered error source in SMP input *i* with a given *supt* average out. Only error sources are considered that lead to errors that average out at relatively small *sups*. Scale measure *supt* interacts with *sups* in affecting whether errors of a particular error source average out but for simplicity value of *supt* is set to a fixed value that is based on other indicators as described in Section 3.4.

Box 3.4 Mathematical formulation of indicators C8 and C9 and of the goal function of indicator C10 used in testing criterion C.

Indicator C8: Resources for downscaling between SMPs

$$g_{8,r}(S) = \max\left(\min\left(\frac{S - S_r^{\min}}{S_r^{\max} - S_r^{\min}}, 1\right), 0\right) \text{ for } S = sups \text{ or } supt \quad (8\text{-}1)$$

$$g_{8,p}(S) = \max\left(\min\left(1, \frac{S_p^{\max} - S}{S_p^{\max} - S_p^{\min}}\right), 0\right) \text{ for } S = sups \text{ or } supt \quad (8\text{-}2)$$

$$f_{C,8}(S) = \sum_{r=1}^{N^{rec}} g_{8,r}(S) \cdot w_{8,r} + \sum_{p=1}^{N^{prov}} g_{8,p}(S) \cdot w_{8,p} \text{ for } S = sups \text{ or } supt \quad (8\text{-}3)$$

$g_{8,r}(S)$	=	Goal function indicating the probability (0-1) that there is no need to downscale input *received* from SMP r for use as input for the SMP considered
$g_{8,p}(S)$	=	Goal function indicating the probability (0-1) that there is no need to downscale output of the SMP considered before it can be used as input for SMP p
$S_r^{\max}$	=	Largest appropriate spatial or temporal support according to criteria A, B, and D of SMP r from which the considered SMP *receives* input (for $S = sups$ or $supt$, respectively)
$S_r^{\min}$	=	Smallest appropriate spatial or temporal support according to criteria A, B, and D of SMP r from which the considered SMP *receives* input (for $S = sups$ or $supt$, respectively)
$S_p^{\max}$	=	Largest appropriate spatial or temporal support according to criteria A, B, and D of SMP p to which the considered SMP *provides* input (for $S = sups$ or $supt$, respectively)
$S_p^{\min}$	=	Smallest appropriate spatial or temporal support according to criteria A, B, and D of SMP p to which the considered SMP *provides* input (for $S = sups$ or $supt$, respectively)
N^{rec}	=	Number of process based or nonlinear SMPs r from which the SMP considered *receives* input and for which the appropriate spatial or temporal support can be identified with criteria A, B, and D
N^{prov}	=	Number of process based or nonlinear SMPs p to which the SMP considered *provides* input and for which the appropriate spatial or temporal support can be identified with criteria A, B, and D
$w_{8,r}$, $w_{8,p}$	=	Weights of goal functions. They indicate the resources required for downscaling of information received from SMP r or provided to SMP p, $\sum_r w_{8,r} + \sum_p w_{8,p} = 1$

Continued

Box 3.4 continued

Indicator C9: Resources for adjusting stream order ranges

Remark: Applies only if the considered model is already applied in the area of study, and if FAMOS is used to improve the scale of model application. Also it only applies to aquatic SMPs

$$f_{c,9}(S) = B\left(S = S^{old}\right) \text{ for } S = sl,\ so_{min} \text{ or } so_{max} \tag{9-1}$$

S^{old} = sl, so_{min}, or so_{max} of the SMP before application of FAMOS (for $S = sl$, so_{min} or so_{max}, respectively)

Indicator C10: Mixing

Remark: Equation 10-1 applies for N flow paths modeled dynamically

$$g_{5,i}(supt, range(so_i^{pred})) = B\left(supt > IntTT(UDF, range(so_i^{pred}))\right) \tag{10-1}$$

i = Index for locations where model predictions are required

$range(so_i^{pred})$ = Range between so_{min} and so_{max} of model predictions at location i

$IntTT_i(UDF, range(so_i^{pred}))$ = Duration of the interval around the mean travel time within which a user defined fraction (UDF) of N flow from the SMP considered arrives at location i. $range(so_i^{pred})$ should have been determined in advance (Section 3.4) because it affects the degree of mixing that can take place before prediction location i is reached.

Box 3.5 Mathematical formulation of indicator D12 and of the goal functions of indicator D11 used in testing the fulfillment of criterion D.

Indicator D11: Usefulness for impact studies and policy development

Remarks:
Equations 11-1 to 11-3 need to be inserted into Equation 4 to calculate the value of indicator D11. Characteristic scales of mitigation options are only taken into account if they are the result of the organizational level to which the people belong that consider these mitigation options.
Indicator D11 considers only SMPs that either can represent mitigation options or are predictions. The meaning of the symbols in Equations 11-1 to 11-3 depends on whether they are applied to SMPs consisting of model equations or predictions (explained below).

$$g_{11,i}(S) = B\left(S < S_i^{a,user}\right) \text{ for } S = sups, supt \text{ or } sl \quad (11\text{-}1)$$

$$g_{11,i}(S) = B\left(S \geq S_i^{b,user}\right) \text{ for } S = exts \text{ or } extt \quad (11\text{-}2)$$

$$g_{11,i}(so_{\min} so_{\max}) = B(range(so) \supset range(so_i^{user})) \quad (11\text{-}3)$$

$S_i^{a,user}$ = If the considered SMP can model mitigation options: Characteristic spatial support, characteristic temporal support or characteristic stream length of change in mitigation options that are studied by user *i* and that can be represented by the considered SMP (for *S* = *sups*, *supt* or *sl*).
If the considered SMP consists of predictions: Largest *sups*, *supt*, or *sl* that is appropriate for prediction user *i* (for *S* = *sups*, *supt* or *sl*).

$S_i^{b,user}$ = If the considered SMP can model mitigation options: Total area or total time period within which mitigation options are considered by user *i* (for *S* = *exts* or *extt*)
If the considered SMP consists of predictions: Smallest spatial or temporal extent that is appropriate for prediction user *i* (for *S* = *exts* or *extt*)

$range(so_i^{user})$ = If the considered SMP can model mitigation options: Range of stream orders where most change in mitigation measures studied by prediction user *i* take place
If the considered SMP consists of predictions: Range of stream orders that typically discharges into the aquatic systems considered by user *i*

Indicator D12: Usefulness for fundamental research

$$f_{D,12}(S) = \left(\ln(S^a_{\max}) - \ln(S)\right) / \left(\ln(S^a_{\max}) - \ln(S^a_{\min})\right) \text{ for } S = sups, supt, \text{ or } sl \quad (12\text{-}1)$$

$$f_{D,12}(S) = \left(\ln(S) - \ln(S^b_{\min})\right) / \left(\ln(S^b_{\max}) - \ln(S^b_{\min})\right) \text{ for } S = exts \text{ or } extt \quad (12\text{-}2)$$

$$f_{D,12}(so_{\min}) = 1 - \left(so_{\min} - (so_{\min})_{\min}\right) / \left((so_{\min})_{\max} - (so_{\min})_{\min}\right) \quad (12\text{-}3)$$

$$f_{D,12}(so_{\max}) = 1 - \left((so_{\max})_{\max} - so_{\max}\right) / \left((so_{\max})_{\max} - (so_{\max})_{\min}\right) \quad (12\text{-}4)$$

$S^a_{\max}$ = Upper boundary of range of *sups, supt,* or *sl* within which the appropriateness is determined by the FAMOS user (Section 3.2), for *S* = *sups*, *supt*, or *sl*, respectively

Continued

Box 3.5 continued		
$S^{a}_{\min}$	=	Lower boundary of range of *sups, supt* or *sl* within which the appropriateness is determined by the FAMOS user (Section 3.2), for *S* = *sups*, *supt*, or *sl*, respectively
$S^{b}_{\max}$	=	Upper boundary of range of *exts* or *extt* within which the appropriateness is determined by the FAMOS user (Section 3.2), for *S* = *exts* or *extt*, respectively
$S^{b}_{\min}$	=	Lower boundary of range of *exts* or *extt* within which the appropriateness is determined by the FAMOS user (Section 3.2), for *S* = *exts* or *extt*, respectively
$(so_{min})_{min}$	=	Lower boundary of range of so_{min} within which the appropriateness is determined by the FAMOS user (Section 3.2)
$(so_{min})_{max}$	=	Upper boundary of range of so_{min} within which the appropriateness is determined by the FAMOS user (Section 3.2)
$(so_{max})_{min}$	=	Lower boundary of range of so_{max} within which the appropriateness is determined by the FAMOS user (Section 3.2)
$(so_{max})_{max}$	=	Upper boundary of range of so_{max} within which the appropriateness is determined by the FAMOS user (Section 3.2)

4 Identification of appropriate modeling scales for a global model of nitrogen export from land to coastal zones[c]

Abstract

Decision makers need to know the effects of alternative watershed management strategies on export of nitrogen (N) to coastal waters. They use this knowledge to reduce undesirable effects of excess N in coastal waters. Various models exist to predict the effects of watershed management on export of N. This paper describes a decision framework to identify the appropriate spatial and temporal scales for using N flux models. The framework is developed for existing models that predict N export from large watersheds and the contribution of N sources and N sinks to this N export. With this framework, modelers can identify the appropriate scale for model predictions and independently scalable model parts. The framework bases the appropriateness of model scales on indicators, which are to be specified by the modeler and which are associated with four criteria. The four criteria require modeling scales to correspond with (A) data, mitigation options, and scenarios, (B) model assumptions, (C) available resources for modeling, and (D) requirements of prediction users. A successful application of the framework is illustrated for a global model of dissolved inorganic nitrogen (DIN) export from watersheds to coastal waters. Ranges of appropriate scales are determined for model output and five independently scalable model parts, which model the (1) surface N balance, (2) point sources, (3) N flow in sediments and small streams, (4) retention in dammed reservoirs, and (5) riverine DIN retention. We conclude that the framework can contribute substantially to a well-balanced and comprehensive identification of appropriate modeling scales.

---

[c] **This chapter is a revision of:**
Dumont, E., C. Kroeze, E. J. Bakker, A. Stein, A. F. Bouwman (2006) Application of a framework to identify the appropriate scale for a model developed to predict global river export of dissolved inorganic nitrogen. In: Caetano, M. and Painho, M. (eds.) Accuracy 2006: proceedings of the 7th international symposium on spatial accuracy assessment in natural resources and environmental sciences, pp. 730-739, Lisbon.

4.1 Introduction

Many different large scale watershed N flow models exist [*Andersen et al.*, 2003]. These models have in common that they simulate processes related to horizontal movement of nitrogen through large drainage networks of river basins. An important property of such models is their modeling scale. Modeling scale can be viewed as the combination of support, extent, and stream order of independently scalable model parts, and model outputs. Model support is the temporal or spatial range in which modeled processes are assumed to be homogeneous. Model extent is the total range of time or space within which processes are modeled. The spatial extent of N flow models is typically a basin or a group of adjacent basins. The temporal model extent is usually between a few months and a few decades. Modeling scale affects many important properties of N flow models such as the processes that can be described, the required input data and the size of watersheds that can be modeled. A model may only make good predictions if the scale of application is the same as the scale at which validation provided good agreement. This is because existing models for N export from river basins are nonlinear to some degree and, hence, scale specific.

N flow models relevant for environmental impact assessments often require predictions for specific river basins. Such models can have different spatial and temporal supports and extents. For example, the Riverstrahler model [*Billen et al.*, 1994] has been applied with various temporal supports on watershed surface areas ranging from 100 to 100000 km^2 and with spatial model supports ranging from 1st to 5th order upstream basins ranging in area between 1 and 5000 km^2 [*Sferratore et al.*, 2005].

The reason for selecting a particular modeling scale is usually not explicitly reported, and the appropriateness of a particular modeling scale is therefore difficult to judge. Modelers have no clear guide for selecting appropriate spatial and temporal modeling scales for predictions of N flows in large river basins.

The purpose of this paper is to present a comprehensive framework to identify the appropriate spatial and temporal scale for N flow models. The framework is developed for models that predict N export from large watersheds and the contribution of N sources and sinks to this N export. With this framework, modelers can identify appropriate scales for model predictions, and for independently scalable model parts. This decision framework may also assist in reporting the rationale behind the scale of a model application. The framework focuses explicitly on existing N flow models that serve as predictive tools. In this paper an application of the framework to a global N flow model is summarized.

4.2 Description of FAMOS

In this section, we describe FAMOS (Framework for Appropriate MOdeling Scale). It is developed to assist model users in identifying appropriate modeling scales in a well-balanced and comprehensive way. FAMOS aims at minimizing the prediction bias by ensuring sufficient validity of model assumptions while simultaneously ensuring the feasibility of model application, agreement with scales of data and scenarios, acceptable scaling error, and useful predictions.

FAMOS indicates the appropriate ranges of up to seven measures of spatial and temporal modeling scale. These are mean spatial support (*sups*), mean temporal support (*supt*), mean stream length (*sl*), spatial extent (*exts*), temporal extent (*extt*), smallest stream order (so_{min}), and largest stream order (so_{max}). The appropriate ranges of these seven measures are identified largely independent of each other. In the following, any of these seven measures will be indicated by *S* (scale). The user can to some degree choose which measures of *S* are considered in FAMOS.

Appropriate ranges of *S* are estimated for a number of SMPs. An SMP is a set of either model equations or model predictions. It has a unique S in existing model applications, and its S can be adjusted. Figure 4.1 A and B are an example of a representation of possible SMP delineations and SMP scales. Defining SMPs is a crucial element in FAMOS.

Four criteria are used to assess the appropriateness of modeling scale. These four criteria require modeling scales to correspond with (A) data and scenarios, (B) model assumptions, (C) available resources for modeling, and (D) appropriately scaled predictions. Fulfillment of each criterion is indicated with indicators that are the basis for the evaluation of appropriateness of each of the seven measures of scale of a scalable model part (SMP). The indicators to be used are selected by the user of FAMOS (Table 4.1). SMP scales are considered appropriate if indicator values exceed pre-specified thresholds. Similar criteria have been suggested for testing the appropriateness of N-flow models [e.g. *Andersen et al.*, 2003]. FAMOS enables its users to evaluate appropriateness of different modeling scales according to their preferences and ideas on appropriateness as reflected by user-defined weights for criteria and indicators. There are three reasons that the 12 indicators in Table 4.1 are suitable to test the model scale based on the four criteria: first they are relatively easy to quantify, second each can be applied to almost any N-flow model, and third together they generally cover all aspects that are required to judge the appropriateness of model scale.

As a first step, a user of FAMOS will have to decide on the model parts to be considered. Due to the complexity of most N-flow models, it is often almost impossible to consider all instances and all modeled N sources and sinks. An instance is a part of an SMP modeling an area, time or stream length for which input is homogeneous. Examples of possible N sources described by an SMP are fertilizer or sewage, and examples of N sinks described by an SMP are riparian zones or lakes. The application of FAMOS can be limited to a restricted set of instances, sources, and sinks represented by an SMP. Similarly, the analyses by FAMOS can be limited to only a subset of all SMPs.

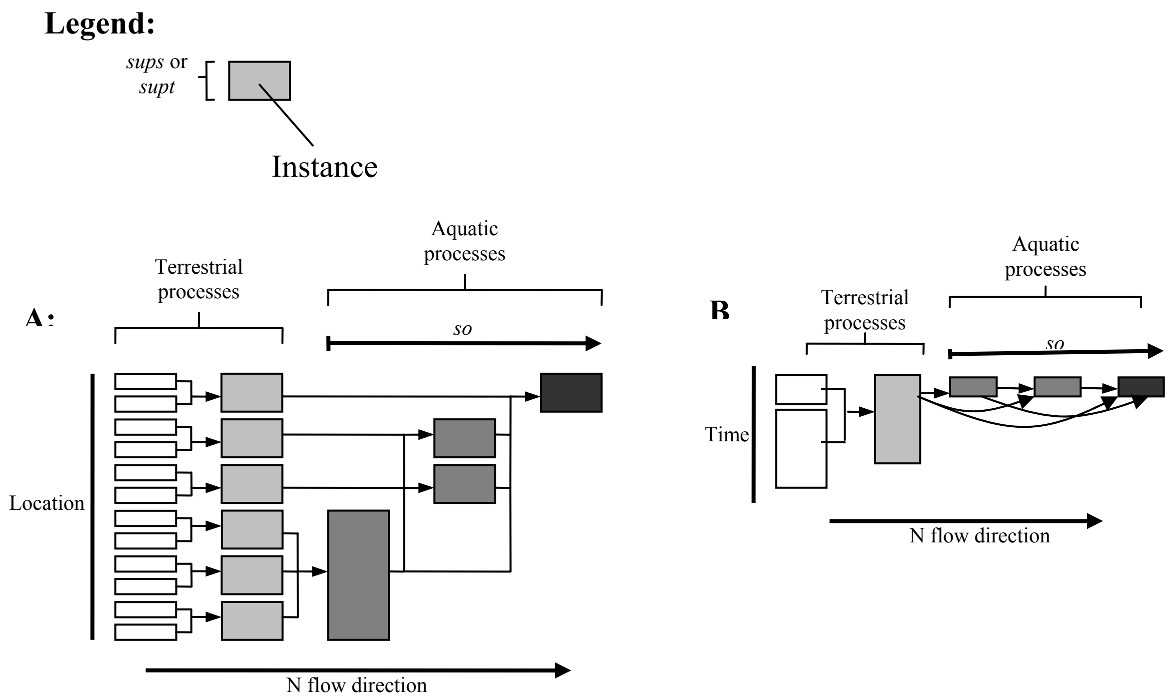

Figure 4.1 Possible delineation of SMPs, SMP instances, and their relations in a watershed N flow model. Axes indicate the stream order and support of the SMPs and the degree to which locations and times of SMPs overlap. Rectangles are SMP instances. The grey tone indicates the SMP to which an instance belongs. Designation of SMPs to instances is arbitrary. Thin arrows indicate the flow of information. A: Possible delineation of SMPs and their relative location and spatial support (*sups*). B: Possible delineation of SMPs and their relative time and temporal support (*supt*).

There are several reasons for considering only a particular part of the model in the application of FAMOS. For instance, only those model parts can be included that are relatively important in the modeled N transport. Their importance can be determined by their expected contribution to outlet N flow and potential for mitigation.

Before application of FAMOS it is necessary that the user sets the upper and lower limits of possible SMP scales (*S*) for which the appropriateness is to be identified. Beyond these limits the user of FAMOS considers scales to be inappropriate. So he/she expects the most appropriate scale to lie within these limits. These limits are necessary for the application of indicators.

Appropriateness of SMP scale is the result of four nested functions describing: (1) appropriateness of modeling scale according to all criteria, (2) appropriateness of modeling scale according to each criterion, (3) fulfillment of criteria, and (4) indicators.

Appropriateness of modeling scale according to all criteria

Appropriateness according to all four criteria (A-D) is estimated as:

$$A(S) = \prod_{c=A}^{D} a_c(S) \tag{1}$$

where:

$A(S)$ – is the appropriateness of scale S for all four criteria (0 or 1),
$a_c(S)$ – is the appropriateness of modeling scale S for criterion c (0 or 1).

Here, $A(S)$ is 1 for appropriate scales S and 0 for inappropriate scales S. Appropriateness $A(S)$ is identified largely independent for the different measures of scale, except for the appropriateness of so_{min} and so_{max} which are dependent. Appropriateness $a_c(S)$ takes the value of 1 if S is appropriate for criterion c and 0 if S is inappropriate for criterion c, $c \in \{A,B,C,D\}$, being indices for the four criteria used in FAMOS. The four values of appropriateness of modeling scale S for each criterion c $a_c(S)$ (0 or 1) are multiplied to obtain $A(S)$ (Equation 1).

Equation 1 is used to identify ranges of S that are appropriate. If $A(S)$ is 0 for all *sups*, *supt*, *sl*, *exts, extt*, so_{min}, or so_{max} within the user-defined limits then this may indicate that the considered SMP is not adequate given the research question, resources and available data.

Appropriateness of modeling scale according to each criterion

Estimation of appropriateness according to each of the four criteria ($a_c(S)$) can be summarized as;

$$a_c(S) = B(F_c(S) > T(c)) \tag{2}$$

where:

$B(..)$ – is a Boolean function. *B(expression)* equals 1 (if *expression* is true), or 0 (if *expression* is false),
$F_c(S)$ – is the fulfillment of criterion c for scale measure S. $F_c(S)$ is real valued on [0,1],

$T(c)$ – is a threshold for appropriateness according to criterion c (0 to 1).

Threshold $T(c)$ is user defined and indicates the degree of fulfillment of criterion c that is considered to be appropriate by the user. Thresholds $T(c)$ are the same for all considered measures of S.

Fulfillment of criteria

Fulfillment of criteria ($F_c(S)$) can be estimated using indicator values. This can be summarized as follows:

$$F_c(S) = \sum_{j=1}^{N_c} v_j \cdot f_{c,j}(S) \tag{3}$$

where:

$f_{c,j}(S)$ – is the value of indicator j, being the j^{th} real valued function of S with values between 0 and 1,
v_j – is a weight indicating the relative importance of indicator j for $F_c(S)$,
N_c – is the number of indicators that is used to calculate the fulfillment of criterion c.

The weights v_j are to be defined by the user of FAMOS. The sum of v_j used in the calculation of $F_c(S)$ is 1.

Indicators

Indicator values ($f_{c,j}(S)$) can be estimated as follows:

$$f_{c,j}(S) = \sum_{i=1}^{N_j} w_{j,i} \cdot g_{j,i}(S) \quad (4)$$

where:
$w_{j,i}$ – is a weight to indicate relative importance of function $g_{j.i}$ for $f_{c,j}(S)$,
$g_{j.i}$ – is goal function i of indicator j, the value of which (0 or 1) depends on S,
N_j – is the number of instances of $g_{j.i}$ used in the calculation of $f_{c,j}(S)$.

Indicators and goal functions to be used are selected by the user of FAMOS. Indicator j is used to test the fulfillment of a criterion c. We suggest two possible indicators to test the fulfillment of criterion A, five for B, three for C, and two for D (Table 4.1). Only the relevant indicators listed in Table 4.1 need to be used. A user of FAMOS may add case-specific indicators. A goal function ($g_{i,j}$) can only be 0 or 1. Weights $w_{j,i}$ may be defined by the user using methods described in Section 3.3. The sum of $w_{j,i}$ used in the calculation of $f_{c,j}(S)$ is 1.

Five of the indicators presented in this paper cannot be summarized by Equation 4. These are indicators A1, B4, C8, C9, and D12 (Table 4.1).

Table 4.1 Description of criteria A, B, C, and D (*c*) and indicators (*j*) that can be used in FAMOS.

c	*j*	**Description of criterion or indicator**
A		**Requires SMP scale to correspond with that of data and scenarios**
	1	Degree to which artifacts in available input data affect model representation of process variability [1]
	2	Measure of the fraction of the spatial or temporal variability of anthropogenic influences described in scenarios that is represented by an SMP [1]
B		**Requires SMP scale to correspond to model assumptions**
	3	Measure of the fraction of actual N flow variability on the scale of the considered SMP that is accountable to drivers of N flow variability described by the considered SMP [1]
	4	Measures of the degree to which an SMP with one or more calibrated coefficients describes processes occurring on the scale for which it uses input. [1]
	5	Measures if there is sufficient relation between modeled N source emissions and outlet N flow in the temporal scale of the SMP
	6	Measure of the validity of the assumptions of the SMP considered about horizontal N dispersion[1,]
	7	Measure of the fraction of total input variation caused by error [1]
C		**Requires adequate resources for modification of the SMP**
	8	Indication of the effort needed for downscaling to obtain inputs from other SMPs, and provide input to other SMPs
	9	Measure of the resources required for an SMP to use input for areas, depending on the range of stream orders that dominates in these areas [2]
	10	measures whether the differences between the effects of N emission in adjacent time steps on outlet N flow is large enough to consider in the model
D		**Requires the prediction scale to correspond to requirements of model users**
	11	Measure of the usefulness of predictions for impact studies and policy makers
	12	Measure of the usefulness of river N export predictions and predictions of source and sink contributions for a fundamental scientific objective of the modeler

[1] This indicator does not apply to SMPs consisting of predictions
[2] This indicator applies only to SMPs modeling aquatic processes

4.3 Application of FAMOS

We used FAMOS to identify the appropriate scale for SMPs of an existing model for river export of dissolved inorganic nitrogen (DIN) to coastal zones of the world [*Dumont et al.*, 2005b]. This model, called NEWS-DIN, takes into account both diffuse and point sources of N. FAMOS is applied to all parts of NEWS-DIN because this is feasible due to the limited complexity of this model. We divided the coupled set of equations of NEWS-DIN into five SMPs, and distinguished an additional SMP that is the predicted DIN export to coastal zones. Our subdivision of NEWS-DIN equations and predictions into six SMPs was based on the condition that these model parts were applied on a unique *S* by Dumont et al. [2005a]. Additionally, it was based on the condition that it is feasible to adjust the individual scales of these SMPs. The SMPs are equations of (1) the surface N balance, (2) point source emissions, (3) N flow in sediments and small streams, (4) dammed reservoir retention, (5) riverine retention, and (6) predictions of DIN flow to coastal zones (Table 4.2). The six distinguished SMPs had four different spatial supports, three different temporal supports, and three different stream order ranges when applied by Dumont et al. [2005a] (Table 4.2). FAMOS is not used to determine the appropriate *sl* because *sl* can not be adjusted within NEWS-DIN. All indicators of FAMOS (Table 4.1) are used, except indicator B7 due to insufficient information on spatial and temporal autocorrelation of input error.

As an example of indicator estimation, the estimation of indicator B5 for SMP 6 is described here. Equation 5 is used for estimation of indicator B5:

$$g_{5,i}(supt, range(so)) = F_i^{in}(supt, range(so)) \tag{5}$$

Where $F_i^{in}(supt, range(so))$ is the fraction of N eventually reaching the locations modeled by the SMP considered from sources modeled by SMP *i* that has reached these locations within the duration of *supt*, *i* is an index for SMPs modeling sources from which N flows before it reaches the location modeled by the SMP considered, *range(so)* is the range of stream orders modeled by the SMP considered. Scale measure *range(so)* should have been determined in advance. It affects the location of the system modeled by the considered SMP.

SMP 6 consists of the NEWS-DIN predictions of DIN export to coastal zones. We use indicator B5 to measure the fraction of DIN eventually reaching the coastal zones from sources modeled by NEWS-DIN that has reached these coastal zones within the duration of *supt* of SMP 6. Only SMP 1 and SMP 2 model N sources. Sources modeled by SMP 1 are located on land surfaces. Sources modeled by SMP 2 are mainly households connected to rivers by sewerage. Values of *supt* of SMP 6 with low values of indicator B5 are expected to cause insufficient relation between modeled N emissions from sources and N flow to coastal zones during the time of prediction. Values of indicator B5 for SMP 6 are higher if a larger fraction of N emitted from modeled sources reaches coastal zones within the duration of *supt*. Indicator B5 is determined for an exhaustive number of *supt* values between one day and 100 years. This range is chosen because appropriate scales of SMP 6 are not expected beyond this range. Indicator B5 for SMP 6 is calculated using various N residence times in different N stores between sources modeled by SMP 1 and 2 and coastal zones (Table 4.3 and Table 4.4). Calculation of indicator B5 for SMP 6 requires the most appropriate range of stream orders for this SMP to be known (Equation 5). This stream order range is 8 to 12, being the range with the highest $F_D(so_{min}, so_{max})$ (Equation 3). Measure $F_D(so_{min}, so_{max})$ was calculated in advance based on indicators D11 and D12.

Table 4.2 Scales (*S*) and data connections of six SMPs of the NEWS-DIN model as applied by Dumont et al. [2005a].

SMP index	1	2	3	4	5	6
Modeled process	Surface N balance	Point source emissions	N flow in sediments and small streams	Dammed reservoir retention	Riverine retention	Riverine DIN export [a]
sups (km)	55[b]	55[b]	150[c]	135[d]	150[c]	150[c] - globe
supt (y)	3	10	10	10	30	10
exts (km)	globe	globe	globe	globe	globe	globe
extt (y)	3	10	10	10	30	10
so_{min}	0	0	unclear	5	5	6
so_{max}	0	0	unclear	10	12	12
Receives input from	Datasets	Datasets	SMP 1, and datasets	Datasets	SMPs 3 and 4, and datasets	SMP 5
Provides output to	SMP 3	SMP 5	SMP 5	SMP 4	SMP 6	Prediction user

a Model prediction.

b average diameter of a 0.5 × 0.5º grid cell

c average diameter of a basin discharging to a coastal zone

d average diameter of a sub-basin discharging into a dammed reservoir

Table 4.3 Values used when estimating the value of $g_{5,1}(supt, range(so))$ for SMP6 of NEWS-DIN: Estimated residence times of N in temporary N stores downstream of N sources on the land surface modeled by SMP1, and the fraction of N from these land surface sources reaching the coastal zones that has flown through each of the N stores considered. Goal functions $g_{5,i}(supt, range(so))$ are used in the calculation of $f_{A,5}(S)$.

N store	Estimated residence time (y)[a]	Fraction of N from the sources on the land surface reaching the coastal zones that has flown though the considered N store	SMP that models the considered N store [b]
Forest biomass (percolation)	$2.0 \cdot 10^2$	0.05	SMP 1
Humus (percolation)	10^2	0.10	SMP 1
Peat (percolation)	10^4	0.05	SMP 1
Fine textured mineral soils (percolation)	40	0.40	SMP 1
Coarse textured mineral soils (percolation)	1.0	0.40	SMP 1
Groundwater (groundwater flow)	10	0.29	SMP 3
Land surface (overland flow)	$1.4 \cdot 10^{-3}$	0.14	SMP 3
Shallow sediments (through flow)	0.011	0.57	SMP 3
Dammed reservoir	0.25	0.25	SMP 4
No dammed reservoir	0.0	0.75	SMP 4
River surface water	$2.7 \cdot 10^{-3}$	0.91	SMP 5
River sediment	0.5	0.09	SMP 5
Seawater circulations	0.083	1.0	SMP 6

a Expert judgment

b See Table 4.2.

Table 4.4 Values used when estimating the value of $g_{5,2}(supt, range(so))$ for SMP6 of NEWS-DIN: Estimated residence times of N in temporary N stores downstream of N emission by households modeled by SMP1, and the fraction of N from households reaching coastal zones that has flown through each of the N stores considered.

N store	Estimated residence time (y)[a]	Fraction of N from households reaching the coastal zones that has flown through the considered N store	SMP that models the considered N store [b]
Sewage treatment	0.05	0.50	SMP 2
Sewerage without treatment	$2.7 \cdot 10^{-3}$	0.50	SMP 2
Dammed reservoir	0.25	0.25	SMP 4
No dammed reservoir	0.0	0.75	SMP 4
River surface water	$2.7 \cdot 10^{-3}$	0.91	SMP 5
River sediment	0.50	0.09	SMP 5
Seawater circulations	0.083	1.0	SMP 6

a Expert judgment
b See Table 4.2.

Table 4.5 Ranges of appropriate scales for SMPs of NEWS-DIN as identified with FAMOS

SMP index	1	2	3	4	5	6
sups (km)	42–194	10–67	42–244	10–133 [b]	42–244	42–300
supt (y)	1.1–10	0.003–26	1.1–10	1.0–17	n.a. [c]	n.a. [c]
exts (km)	3,000–30,000 [a]	2,000–30,000 [a]	30,000 [a]	13,000–30,000 [a]	24,000–30,000 [a]	13,000–30,000 [a]
extt (y)	10–50 [b]	26–50 [b]	unclear	23–50 [b]	47–50 [b]	23–50 [b]
so_{min}	0	0	unclear	7	8	8
so_{max}	0	0	unclear	8	10	12

[a] 30,000 signifies global extent
[b] Scale of NEWS-DIN as applied by Dumont et al. [2005] (Table 4.2) is outside this range
[c] No appropriate values of *supt* are identified using FAMOS

Sixty pathways of N from land surfaces to coastal zones are considered for the calculation of $F_1^{in}(supt, range(so))$ (Equation 5). Eight pathways of N from households to coastal zones are considered for the calculation of $F_2^{in}(supt, range(so))$. These pathways are all the possible combinations of the N stores given in Table 4.3 and Table 4.4, respectively. First, the fraction of N passing through each considered pathway is calculated by multiplication of the fraction of N passing through each individual N store, and the travel time through each pathway is calculated as the summation of the residence times of each of the N stores through which the pathway goes (Table 4.3 and Table 4.4). For example, a pathway of N from the land surface through coarse textured mineral soil surface, shallow sediments, a dammed reservoir, river surface water and sea water has a travel time of 1.35 years ($1+0.011+2.7\cdot10^{-3}+0.083$) and transports a fraction of 0.05 of all N emitted from land surfaces that eventually reaches coastal waters ($0.4\cdot0.57\cdot0.25\cdot0.91\cdot1$) (Table 4.3). A plot of travel time to coastal zones versus fraction of N for N from land surfaces and N from households was made using the latter method of calculation (Figure 4.2). This plot can be used to identify fractions $F_1^{in}(supt, range(so))$ and $F_2^{in}(supt, range(so))$ considering N from sources modeled by SMP1 and 2, respectively, that reach the coastal zone (SMP6) within the duration of *supt*. Weights $w_{5,1}$ and $w_{5,2}$ are the expected fractions of N reaching the coastal zones that originate from sources modeled by SMP 1 and SMP 2, respectively. We set $w_{5,1}$ at 0.94 and $w_{5,2}$ at 0.06 because we expect 94% of DIN exported to coastal zones to originate from land surfaces and 6% to originate from households [based on *Dumont et al.*, 2005b]. We use Figure 4.2 and Equation 5 to identify the value of indicator B5 for values of *supt* between one day and 100 years. Indicator B5 is 0 for *supt* smaller than 0.09 years. Indicator B5 exceeds 0.1 at *supt* larger then 1 year, 0.4 at *supt* larger then 11 years, and 0.6 at *supt* larger then 40 years. Indicator B5 is zero for *supt* below 0.09 years because no N entering the watershed from land surfaces and households is expected to reach the coastal waters within the duration of such small *supt*. Indicator B5 starts to increase rapidly when *supt* exceeds 1.09 years because then *sups* starts to equal travel times from land surfaces which have a much higher weight in the calculation then travel times from households. An increase in indicator B5 from 0.47 to 0.62 occurs when *supt* exceeds 40 years because at travel times larger then 40 years the N from land surfaces on fine textured mineral soils is expected to reach the coastal zone (Table 4.3), as opposed to only the N from land surfaces on course textured mineral soils. The inertia that we expect for coastal zone responses to land surface processes, shown in Figure 4.2, are in line with those found by Stalnacke et al. [2004].

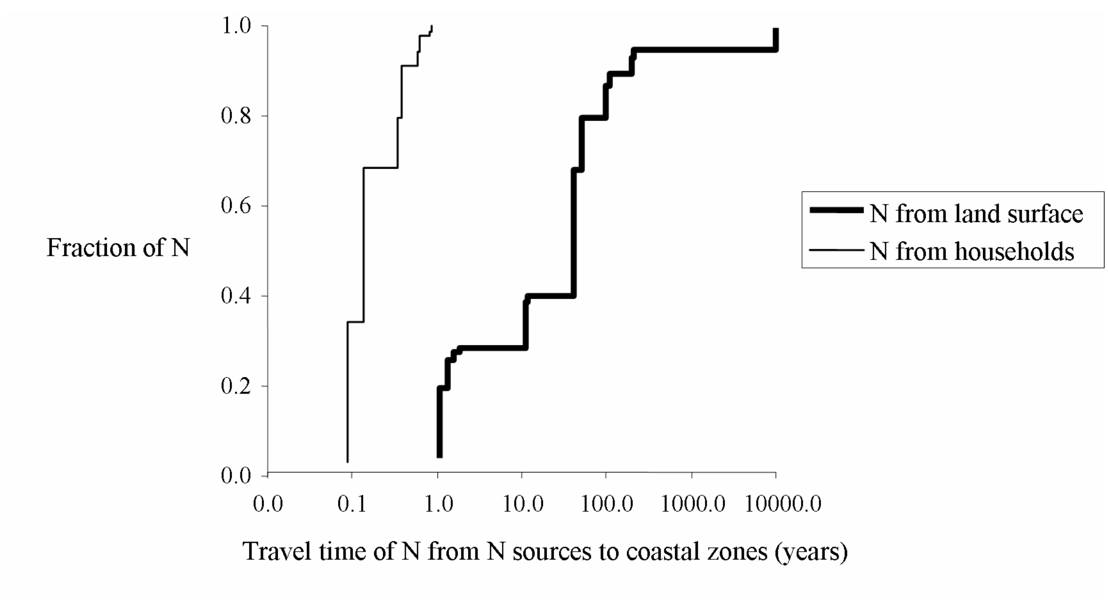

Figure 4.2 Estimated fractions of N reaching the coastal zones from land surfaces and households that had travel times below the travel times that are indicated on the horizontal axis. These relations apply to the system that should be modeled by NEWS-DIN according to FAMOS, being all basins in the world having stream orders from 8 to 12.

In the following steps, Equations 1 to 3 were used to combine the values of indicator B5 with those of other indicators to find the appropriate scales for SMP 6 (Table 4.5). First values of indicator B5 are combined with those of indicators B3 and B6 in Equation 3 to find $F_B(S)$. This fulfillment of criterion B is then inserted in Equation 2 to identify $a_B(S)$, which is inserted in Equation 1 to identify $A(S)$. These steps are also taken for the other SMPs resulting in the appropriate SMP scales reported in Table 4.5.

Values for *exts*, so_{min} and so_{max} of SMPs of NEWS-DIN as applied by Dumont et al. [2005] were within the ranges identified as appropriate using FAMOS (Table 4.2 versus Table 4.5). The values of *sups* and *extt* were not always in the ranges identified as appropriate using FAMOS. For the SMPs modeling riverine retention and predicted riverine DIN export (SMPs 5 and 6, respectively) no appropriate *supt* values were found. The *sups* of the SMP modeling dammed reservoir retention in Dumont et al. [2005] (SMP4) was slightly larger then the range identified as appropriate by FAMOS. The *extt* values of all SMPs in Dumont et al. [2005] were smaller then the ranges identified as appropriate by FAMOS. No appropriate *supt* values were found for SMPs 5 and 6 because predictions with a *supt* smaller than one year are desired but available knowledge and data are not sufficient for this. Based on these FAMOS results, we advise (1) to obtain knowledge and data needed to model riverine processes on a *supt* smaller the one year and (2) to adjust the *extt* of NEWS-DIN to the ranges identified as appropriate by FAMOS by a hind-cast or modeling scenarios of the future.

4.4 Discussion and Conclusion

This paper describes the development of a decision framework (FAMOS) to identify appropriate spatial and temporal scales for modeling N flows. It is based on four criteria, which in turn are fed by 12 indicators. A central concept in the framework is the scalable model part. Scalable model parts are distinguished in the framework to be able to fully account for all aspects of model scale.

We illustrate the use of FAMOS by applying it to NEWS-DIN, an existing model for river export of dissolved inorganic nitrogen to coastal zones. Ranges of appropriate scales were identified for all scalable model parts. FAMOS results indicate that most of the scales used in the original application of NEWS-DIN (Chapter 2) are appropriate. The *sups* and *extt* of some SMPs are not appropriate. No appropriate *supt* values were found for the SMPs describing riverine retention and predicted riverine DIN export. Such results of applying FAMOS appear to be a useful basis for improved application of NEWS-DIN and many other N-flow models.

Application of FAMOS requires profound knowledge of input data production, model development and characteristic scales of processes in the study area. Also the application of the decision framework for all scalable model parts of a model is only feasible if the model is relatively simple. Otherwise a selection of scalable model parts must be made.

We conclude that now we have been able to systematically and transparently address the issue of scale selection for N flow models. This attempt appears to be generic and general and is a useful basis for improved application of NEWS-DIN and many other N flux models.

Acknowledgements

We are grateful to Rik Leemans of the Environmental Systems Analyses (ESA) group for reviewing drafts of this paper. Egon Dumont is grateful to Netherlands Organization for Scientific Research (NWO) for funding his research and for the facilities made available by Wageningen University.

5 Identification of appropriate spatial and temporal scale for modeling N flows in reconnected floodplains[d]

Abstract

Levees along rivers are increasingly removed to increase flood water storage. Such floodplain reconnection often causes increased nitrogen (N) inflows into floodplain waters, which may have undesired effects. Models for N flows into reconnected floodplains are developed to assess such effects. N flow models can have different spatial and temporal scales. We recently developed a framework (FAMOS) to identify the appropriate scales for large-scale N flow models using a set of indicators. In the current study, we assess the applicability of FAMOS to models of N flow from rivers to aquatic systems in reconnected floodplains.

For applying FAMOS to floodplain N flow models, we added three new indicators for the appropriateness of modeling scale: (1) the reliability of empirical relationships for modeling the aggregated effect of alternating conditions on N flow, (2) the validity of particular nonlinear model equations, and (3) the effect of inertia on model reliability. These new indicators are applied to a hypothetical model of an existing reconnected floodplain to illustrate their use, output and effect on appropriate model scales. We show that the new indicators provide a feasible basis for a comprehensive identification of appropriate model scales.

[d] **This chapter has been submitted for publication as:**
Dumont, E., T. Hein, S. Preiner. Appropriate spatial and temporal scale for modeling N flows in reconnected floodplains. Submitted to Water Resources Research.

5.1 Introduction

Natural floodplains have disappeared along many regulated rivers [*Sparks*, 1995]. For example, about eighty percent of the floodplain in the Danube river basin is now hydrologically decoupled from the river [*Hohensinner et al.*, 2004]. As a result, many rivers lack sufficient storage capacity for water in case of peak flow. To reduce the associated risk for flooding during peak discharges, levees are nowadays partly removed in some river basins to restore the flood water storage capacity of floodplains [*Stanley and Doyle*, 2003].

Floodplain reconnection usually leads to a net nitrogen (N) flow from rivers to floodplain waters because the N concentration in rivers is generally higher than those in the surface waters of their floodplains [*Hein et al.*, 2004b; *Jolankai and Pataki*, 2005]. Floodplain managers may need models to assess the effect of such N inflow on many important floodplain functions related to N cycling. Examples of affected floodplain functions are the habitat function for ecologically important species [*Chovanec et al.*, 2000] and the N retention capacity of floodplains related to microbial processing [e.g. *McClain et al.*, 2003].

Floodplains may be increasingly reconnected to rivers in the future for two reasons. First, annual precipitation and its distribution during the year may change and the frequency of rain events with high intensity may increase as a consequence of climate change [*IPCC*, 1998]. This may increase risks of river flooding, and floodplain reconnection may be necessary to reduce these risks. Second, awareness of undesired effects of levees protecting floodplains increases [*Parish*, 2004; *World Commission on Dams*, 2000]. Examples include loss of biodiversity [*Ward et al.*, 1999], excess sedimentation in floodplains [*Schiemer et al.*, 1999], and reduced nursery value for commercial fish species [*Roni et al.*, 2005].

As a consequence, there is an increasing interest in modeling N flow in reconnected floodplains, and, associated with that, an increasing need for guidance in the selection of a modeling scale. Modeling scale is an important issue because it affects model validity, the effort required to apply the model, usefulness of model predictions, the data needed for the model, and the type of mitigation options and scenarios that can be modeled [*Dumont et al.*, submitted-b]. The choice of a modeling scale may be difficult because of differences in scales of data, processes, research questions and management questions. For example, a model may describe denitrification at the spatial scale of a few m^2, while the data on for example soil properties required by this model may only be available on spatial scales of one or a few hectares. In such cases, the question at which spatial scale denitrification should be modeled is not easily answered. This type of question is common in N flow modeling [e.g. *Bellamy and Loveland*, 2001a; *FitzHugh and Mackay*, 2000; *Jha et al.*, 2004; *Mamillapalli et al.*, 1996; *Quinn*, 2004; *Ruiz et al.*, 2002b; *Sferratore et al.*, 2005; *Vachaud and Chen*, 2002].

Recently the Framework for Appropriate MOdeling Scale (FAMOS), has been developed for identifying appropriate spatial and temporal scales for modeling N flows from watersheds to large aquatic systems, like river reaches and estuaries [*Dumont et al.*, submitted-b]. FAMOS is the first comprehensive guide to achieve this. It considers measures of scale, scalable model parts (SMPs), and criteria and indicators for appropriateness of scales. An SMP is a set of either model equations or model predictions. It has a unique scale in existing model applications, and its scale can be adjusted [*Dumont et al.*, submitted-b]. FAMOS helps the modeler to quantify and combine all important indicators needed to identify the most appropriate scale in a structured and comprehensive way. FAMOS has been applied to a model of N flow to large coastal waters [*Dumont et al.*, 2006]. This resulted in the identification of appropriate spatial and temporal scales for all six SMPs of this model. Applicability of FAMOS to N-flow models for smaller and upstream aquatic systems has not been investigated yet.

In this study we assess the applicability of the existing version of FAMOS to models of N flow from rivers to aquatic systems in reconnected floodplains. We first extend FAMOS for application to these models and subsequently we apply FAMOS to modeling average nitrate (NO_3^--N) concentration in reconnected side arms of the Danube near Regelsbrunn (Austria).

An overview of FAMOS and the need for its extension with new indicators is presented in Section 5.2. This is followed by a description of a set of new indicators to make FAMOS more suitable to floodplain models (Section 5.3). The use of the new indicators is discussed in Section 5.4. A general discussion and conclusions are in Section 5.6.

5.2 Overview of FAMOS

5.2.1 The framework

FAMOS is a framework developed to guide in identification of the appropriate spatial and temporal scales for models of N flow. FAMOS identifies appropriate ranges of scale for one or more SMPs. FAMOS uses four criteria which require modeling scales to correspond with (A) data and scenarios, (B) model assumptions, (C) available resources for modeling, and (D) appropriately scaled predictions. Fulfillment of each criterion for an SMP is tested with indicators which are selected by the user. Users can thus include their own knowledge and experience in FAMOS. Descriptions of 12 indicators can be found in Dumont et al. [submitted-b] (Table 3.2). Finally, the values of all indicators are combined into a single index of appropriateness [*Dumont et al.*, submitted-b].

Identification of the appropriate spatial and temporal scale for each selected SMP is done in FAMOS in five steps: selection and identification of SMPs to be considered by FAMOS (1), and subsequent estimation of indicator values (2), fulfillment of criteria (3), appropriateness according to each criterion (4), and finally appropriateness according to all criteria (5) (for details, see Section 3.2).

Table 5.1 Description of criteria A, B, C, and D (*c*) and indicators (*j*) that can be used in FAMOS. Indicators are listed below the criterion that they can test. Indicators labeled 1 to 12 are described in Dumont et al. [submitted-b] and are part of the existing version of FAMOS. Indicators B8, B3a, and B5a are new indicators that are introduced in this paper.

c	*j*	**Description of criterion or indicator**
A		**Requires SMP scale to correspond with that of data and scenarios**
	1	Degree to which artifacts in available input data affect model representation of process variability [1]
	2	Measure of the fraction of the spatial or temporal variability of anthropogenic influences described in scenarios that is represented by an SMP [1]
B		**Requires SMP scale to correspond to model assumptions**
	3	Measure of the fraction of actual N flux variability on the scale of the considered SMP that is accountable to drivers of N flux variability described by the considered SMP [1]
	3a	*Indication of the validity of particular nonlinear model equations*
	4	Measure of the degree to which an SMP with one or more calibrated coefficients describes processes occurring on the scale for which it uses input. [1]
	5	Measures if there is sufficient relation between modeled N inputs and predicted N flow in the temporal scale of the SMP
	5a	*Indication of the effect of inertia on model reliability*
	6	Measure of the validity of the assumptions of the SMP considered about horizontal N dispersion[1,]
	7	Measure of the fraction of total input variation caused by error [1]
	8	*Indication of the reliability with which empirical relations can model the aggregated effect of alternating conditions on N flow*
C		**Requires adequate resources for modification of the SMP**
	8	Indication of the effort needed for downscaling to obtain inputs from other SMPs, and provide input to other SMPs
	9	Measure of the resources required for an SMP to use input for areas, depending on the range of stream orders that dominates in these areas [2]
	10	Measures whether the differences between the effects of N emission in adjacent time steps on outlet N flux is large enough to consider in the model
D		**Requires the prediction scale to correspond to requirements of model users**
	11	Measure of the usefulness of predictions for impact studies and policy makers
	12	Measure of the usefulness of river N export predictions and predictions of source and sink contributions for a fundamental scientific objective of the modeler

[1] This indicator does not apply to SMPs that are model predictions

[2] This indicator applies only to SMPs modeling aquatic processes

Application of FAMOS results in an indication of appropriate ranges of up to seven measures of spatial or temporal modeling scale. These measures of modeling scale are mean spatial support (*sups*), mean temporal support (*supt*), mean stream length, spatial extent (*exts*), temporal extent (*extt*), and the smallest and largest stream order. The measure *sups* is the mean size of the areas represented by single values of input variables [based on *Heuvelink*, 1998], and *supt* is the mean duration of the times represented by single values of input variables. For stochastic models *sups* and *supt* apply to single input distributions instead of single input values. For inputs consisting of grids or regular time series, *sups* or *supt* equal the grid cell size or time step, respectively. Stream length is the length of river sections represented by single values (or distributions) of input of an SMP. Measures *exts* and *extt* are the total area and duration, respectively, for which an SMP is applied. Smallest and largest stream order are measures of the minimum and maximum sizes of river reaches, respectively, to which an SMP is applied [based on *Strahler*, 1964]. In FAMOS the mean support is used because there may be some variation in these measures in individual inputs. Measures *sups* and *exts* are measured as mean diameters.

5.2.2 The need for new indicators

FAMOS has been developed and tested for models of N flow in one or more watersheds. Here we explore the possibility to apply FAMOS to models of N flow between rivers and reconnected floodplains. The FAMOS indicators for the fulfillment of criteria A, C and D (Table 3.2) can most probably be quantified for such models. Five FAMOS indicators can be used to test the fulfillment of criterion B for testing the correspondence of model scale with model assumptions. They consider the scale of processes (B3), validation (B4), inertia (B5), N dispersion (B6) and input errors (B7).

Although these indicators for criterion B can be applied to floodplain models, the assessment of criterion B needs to be improved. It is needed to allow for describing the effect of model scale on the reliability of model predictions of the effect of flooding. For this purpose we introduce one new indicator (B8) for flooding effects and modify indicators B3 and B5 to improve their applicability for floodplain models.

New indicator B8

Intermittent flooding is common in reconnected floodplains because of the alternate presence and absence of surface connection with river water. During surface connection N exchange between river and floodplain is relatively fast, and floodplains are inundated. This is not the case if there is no surface connection. The temporal pattern of N transformation processes in reconnected floodplains depends on the alternating surface connection and disconnection. Each flooding triggers a series of N transformations. For example, shortly after surface connection aerobic nitrification activity may increase in floodplain lakes, followed by increased algal productivity. After subsequent cessation of surface connection there is increased organic matter retention, followed by anoxia which may increase denitrification activity [Dent and Grimm, 1999]. Indicator B8 for alternating conditions is used to assess the degree to which the effect of flooding on such sequences of processes can be modeled depending on model scale (Section 5.3.1).

Adjusted indicator B3a

The original indicator B3 (Table 5.1) is used to test if the model scale corresponds to the scale of the processes described by nonlinear model equations. B3 recognizes that the processes driving the N flows change with *sups*. Indicator B3 is especially suitable for models developed for the watershed or larger scale. At such scales, empirical data usually are more scarce and unreliable than at finer scales [*Corwin et al.*, 2006]. As a result, model equations for the watershed scale (or larger) may be relatively unreliable because (1) they are often the result of upscaling or (2) they are derived using the relatively scarce and unreliable data available at these coarse scales [*Corwin et al.*, 2006].

Equations in floodplain models are often derived using relatively reliable data and knowledge available at the floodplain scale. Therefore, these equations are generally more reliable than the coarser scale models for which FAMOS was originally developed. This relatively high precision of both input data and model equations allows for a more reliable assessment of scale-dependent validity of nonlinear model equations than can be achieved with B3. For this reason indicator B3 is adjusted for floodplain models (Section 5.3.2).

Adjusted indicator B5a

The original indicator B5 tests the relationship between modeled N inputs and predicted N flow for different temporal scales of the SMP (Table 3.2). It applies to ranges of model scale where slope is the main factor determining the direction of diffuse N flow. Since slopes generally do not change over time, the flow direction can be assumed to be constant and the inertia of the effect of distant N sources on N in an aquatic system is estimated in indicator B5 as a function of travel time.

In floodplains the direction of N flow is determined by fine-scale processes such as uptake by plants, sedimentation, migration of animals, and water flows driven by variations in river water level and ground water level. As a result, the delay between inflow and outflow for a system can not be approximated as a "pipeline delay" where all N that has followed the same path has the same travel time [*Sterman*, 2000]. Instead, the travel time needs to be characterized as a delay typical for perfectly mixed systems, where the order at which the N molecules entered the modeled system is irrelevant for the order of exit [*Sterman*, 2000]. To account for the inertia in floodplains, we introduce indicator B5a to select the appropriate model scale accounting for this inertia (Section 5.3.3).

5.3 New indicators of appropriate modeling scale

5.3.1 Indicator B8: Alternating conditions

Indicator B8 tests the reliability of empirical relationships used to model the aggregated effect of alternating flooding and drainage on N flow and N cycling in river floodplains. In floodplains, conditions under which N cycling takes place may vary repeatedly in time. Shortly after the change to a new condition the N cycling will move towards a new equilibrium. The new equilibrium may not be reached when conditions change again. Therefore, knowledge of the temporal pattern of alternation between conditions is required, and such knowledge is often not available over a larger temporal extent. Models describing N cycling over larger temporal extents are often based on empirical equations lumping the effect of alternating conditions on N cycling over prolonged time periods. An input of such a lumped equation is an aggregate of the alternating conditions.

Indicator B8 assesses the validity of aggregates of alternating conditions that are used to model N cycling. Indicator B8 is a function of *supt* and increases along with the correlation with other aggregates of the temporal pattern. We expect that this correlation increases with *supt* [in line with *Blöschl and Sivapalan*, 1995; *Wood et al.*, 1988] and that it indicates the potential of the lumped equation to explain the N flow. We will illustrate this for a case study (Section 5.4.1).

Temporal patterns of alternating conditions can vary within a reconnected floodplain. For example, floodplain soils that are higher are inundated during shorter periods and less frequently than soils in depressions. Therefore a comprehensive calculation of indicator B8 needs to be based on a representative sample of the different temporal patterns of such alternating conditions in different locations in the modeled floodplain. This is done as follows:

$$f_{B8}(supt) = \sum_{j=1}^{N^{loc}} w_j^{B8} \cdot \sum_{i=1}^{N^{aggr}} w_{j,i}^{B8} \cdot r_{j,i}^2(supt) \qquad (1)$$

where $f_{B8}(supt)$ is the value of indicator B8, $r_{j,i}^2(supt)$ is the r^2 between the value of the input of the considered empirical relationship and other possible aggregates *i* of the temporal pattern of an alternating condition on location *j*, N^{aggr} is the number aggregates *i* considered for location *j*, $w_{j,i}^{B8}$ is a user defined weight indicating the relative importance of aggregate *i* on location *j*, N^{loc} is the number of considered locations, w_j^{B8} is a user defined weight indicating the relative importance of location *j* for the modeler. The aggregates *i* are user defined. The weights sum to one, i.e. $\sum_{j=1}^{N^{loc}} w_j^{B8} = 1$ and $\sum_{i=1}^{N^{prox}} w_{j,i}^{B8} = 1$ (for all *j*).

5.3.2 Indicator B3a: nonlinear model equations

Indicator B3a tests the validity of a nonlinear equation in two cases, i.e. (1) varying patchiness or (2) piecewise nonlinearity between scales. For the first case of a patchy distribution of an input variable in space or time at the conceptual scale of the respective nonlinear equation, we delineate patches which are assigned a value indicating the corresponding state of the input variable. The resulting delineated patches are used to obtain the fraction of the modeled area or time where the nonlinear model equation is expected to be valid for different values of *sups* or *supt*, respectively. The nonlinear equation is expected to be valid for individual input values representing areas or times within a patch [in line with *Beven*, 1995]. The fraction of input values for which this is the case is denoted by *F*. Fraction *F* decreases with increasing *sups* or *supt*, respectively. In the calculation of indicator B3a, *F* is estimated as the probability that a randomly located area of size *sups* or time period of duration *supt* falls within the previously delineated patches. Fraction *F* can be calculated by taking a moving average, with a window size equaling the *sups* or *supt*, of the map or time series of delineated patches. Then *F* is the fraction of the modeled area or time where this moving average has the exact value assigned to one of the delineated patches. If there are no fine-scale measurements of the input variable considered, the fraction *F* needs to be based on knowledge of the size of the patches.

All calculated or estimated values of *F* for a given *sups* or *supt* are used as follows in the calculation of indicator B3a:

$$f_{B3a}(S) = \sum_{i=1}^{N^{NLeq}} w_i^{B3a} \cdot F_i(S) \qquad \text{for } S = sups \text{ or } supt \tag{2}$$

Where $f_{B3a}(S)$ is the value of indicator B3a as a function of *sups* or *supt*, w_i^{B3a} is a user defined weight indicating the relative importance of nonlinear equation *i* in the SMP considered, N^{NLeq} is the number of nonlinear equations in the SMP considered that are used in the estimation of $f_{B3a}(S)$, $F_i(S)$ is fraction *F* for a given *sups* or *supt*. The sum of all w_i^{B3a} is one. We advise to increase the value of w_i^{B3a} with the fraction of the modeled area or time to which $F_i(S)$ applies, the sensitivity of the model to equation *i*, the nonlinearity of equation *i*, and the certainty with which $F_i(S)$ could be identified.

In the second case, when the input variable of a nonlinear model equation is not patchy but the equation is piecewise linear, a fine-scale map or time series of the input variable is used to delineate areas or time periods for which only one linear section of the piecewise linear relationship applies. The size of delineated areas are used to identify the fraction *F* (i.e. the fraction of the modeled area or time where the equation is expected to be valid) for different values *sups* or *supt*. This is done using a moving average, with a window size of *sups* or *supt*, respectively, as described previously. If there are no fine-scale measurements of the input variable considered, fraction *F* needs to be based on knowledge of the approximate sizes of areas or time periods within which only one linear section of the piecewise linear relationship applies.

5.3.3 Indicator B5a: Inertia

Indicator B5a tests the effect of inertia on model reliability. For static SMPs indicator B5a assesses the appropriateness of *supt* and for dynamic SMPs the appropriateness of *extt*. The aim of indicator B5a is similar to that of the original indicator B5, being the reduction of uncertainty due to ignorance of N inflows into the modeled area prior to the modeled time period. The adjustment of indicator B5a compared to the original B5 is that it accounts for the time-dependent directions of many N flows in and out of N stores in reconnected floodplains. N stores are places where N is stored for a certain time before it is released. An example of such an N store is a reconnected floodplain lake receiving N from the river during high water levels by inflowing surface water, and discharging N to the river during declining river water levels. Another example is vegetation taking up N from the soil during spring and summer, and releasing N to the soil during autumn and floods, by die-off and decomposition. Indicator B5a is more appropriate than indicator B5 if N stores are affected by N flows with time dependent directions.

The calculation of indicator B5a is based on turnover time. Turnover time is chosen as a basis for indicator B5a because it is a quantity that can often be estimated from general knowledge of processes in the N store, without a need for direct measurements. We define turnover time of an N store as the time period during which most of the N in an N store has been replaced by new N.

Indicator B5a can consider different N stores modeled by an SMP individually. The N stores considered have spatial boundaries coinciding with boundaries of areas represented by single values calculated by an SMP (spatial support units [*Bierkens et al.*, 2000]). Therefore, indicator B5a represents the reliability of SMP output for individual spatial support units (as opposed to aggregates of output over multiple spatial support units). The value of indicator B5a increases with the number of such N stores having turnover times of N smaller than *supt* or *extt* at all times.

The value of indicator B5a varies with *sups* and *supt* or *extt* as summarized in Equation 3;

$$f_{B5a}(S, sups) = \sum_{i=1}^{N^{store}} w_i^{B5a} \cdot g_i(S, sups) \qquad \text{for } S = supt \text{ or } extt \tag{3}$$

where $f_{B5a}(S,sups)$ is the value of indicator B5a, S is *supt* in case the SMP considered is static or *extt* in case the SMP considered is dynamic, N^{store} is the number of modeled N stores i considered, w_i^{B5a} is a user defined weight indicating the relative importance of N store i for the N flow predicted by the SMP considered. The sum of w_i^{B5a} for all i is one. $g_i(S,sups)$ is a goal function for N store i defined in equation 4;

$$g_i(S,sups) = \begin{cases} 1 & for \quad S \geq \max\left(s_i(sups,t) / r_i(t) \right) \\ 0 & for \quad S < \max\left(s_i(sups,t) / r_i(t) \right) \end{cases} \qquad \text{for } S = supt \text{ or } extt \tag{4}$$

where $g_i(S,sups)$ is one for values of S that are larger or equal to the maximum turnover time ($max(s_i(sups,t)/r_i(t))$), S is *supt* in case the SMP considered is static or *extt* in case the SMP considered is dynamic, t is an instant in time that is likely to be within the modeled time period, or which probably has values of turnover time that are similar to those in the modeled time period, $s_i(sups,t)$ is the amount of N on time t in modeled N store i with a size of *sups*, and $r_i(t)$ is the total rate of all N flows in or out of modeled N store i on time t. Rate $r_i(t)$ is the total of both N flows to neighboring N stores and N transformations to N forms that are not modeled (e.g. denitrification/N fixation or immobilization/mobilization). Ratio $s_i(sups,t)/r_i(t)$ is an estimate of turnover time in N store i on time t.

Assumptions underlying equation 4 are that there is instantaneous and complete mixing of all N inflows to the considered N stores and that N inflow equals N outflow. Equation 4 differs for static and dynamic SMPs. Static SMPs ignore N that was stored before the time period represented by each single calculated value (*supt*). For static SMPs we therefore expect that ignoring this becomes unacceptable if the maximum turnover time of the N within modeled N stores is longer then *supt*. Dynamic SMPs ignore the N that was stored prior to the period that is simulated. For such SMPs we therefore expect that this ignorance becomes unacceptable if the maximum turnover time of the N within modeled N stores is longer than the duration of the time period of the simulation (*extt*).

Indicator B5a can be estimated using values of maximum turnover time based on general knowledge of processes in the N stores considered (e.g. Section 5.4.3). When a suitable time series of measurements is available, the values of $s_i(sups,t)$ and $r_i(t)$ in equation 4 may be calculated individually. Equation 4 is then only valid if $r_i(t)$ is temporally autocorrelated over time durations similar or larger then those of $s_i(sups,t)$. Otherwise we advise to smooth the time series of $r_i(t)$ with a moving average using a window size of the time over which $s_i(sups,t)$ is temporally autocorrelated.

The measure *sups* is inversely related to the value of indicator B5a (Equation 4), because it increases turnover time in N stores with a surface area of *sups*. The reason is that for a large *sups*, a smaller part of the area of such an N store is close to its boundaries where N exchange takes place. To simplify the application of indicator B5a, *sups* is assigned a fixed value equal to the appropriate *sups* according to other indicators in FAMOS (Table 3.2) that are applied before indicator B5a.

5.4 Application of the new indicators

The three new indicators are used to test the appropriateness of *supt* and *extt* between one day and one year. The indicators are applied to a hypothetical scalable model part (SMP) of a reconnected side arm of the Danube near Regelsbrunn (Austria) (Figure 5.1). The aim of the SMP is to assess the N concentration in the side arm.

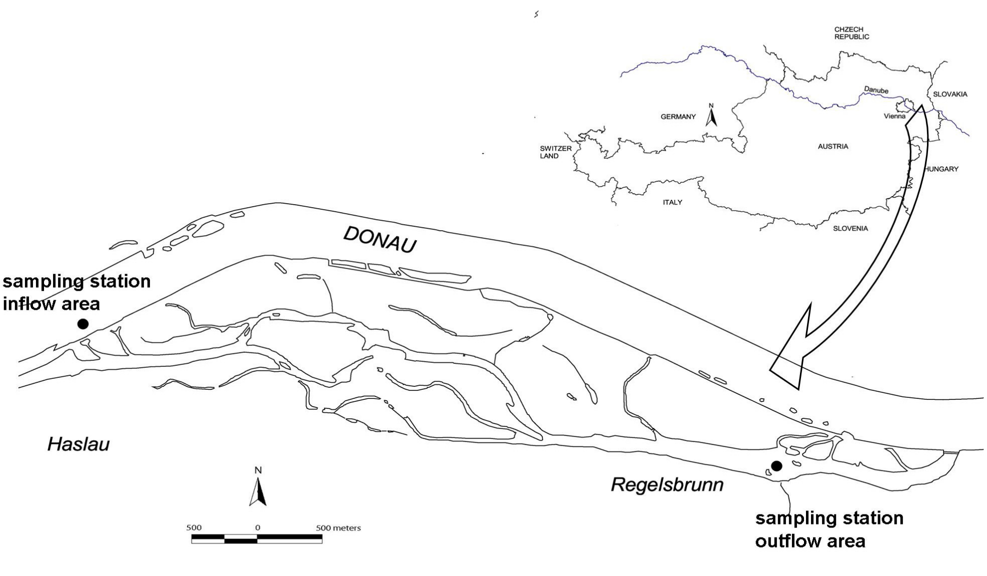

Figure 5.1 Reconnected side arm of the Donau near Regelsbrunn and its location in Austria.

5.4.1 Indicator B8

We will use indicator B8 to test the reliability of an empirical relationship between the total duration of surface connection and average measured nitrate (NO_3^--N) concentrations in reconnected side arms. The pattern of alternating conditions determines the shape of this relationship. Two different locations with such alternating conditions are distinguished. Location 1 is the part of the side arm that is always filled with water. Location 2 is littoral zone around the side arm where sediments are temporarily inundated. Alternating conditions affecting side arm NO_3^--N concentrations differ between these two locations.

In location 1, the following sequence of processes affecting side arm NO_3^--N concentration occurs after the start of each event of surface connection: (1) increased aerobic nitrification activity, (2) increased algal productivity, and (3) an increased organic matter retention after cessation of surface connection followed by (4) anoxia increasing denitrification [*Dent and Grimm*, 1999].

In location 2 (littoral zone), inundation of soils after the start of each event of surface connection will lead to a decrease in oxygen and onset of anaerobic conditions. Before these anaerobic conditions are reached, the following sequence of processes affecting side arm NO_3^--N concentration occurs: (1) rapid leaching of nutrients from decomposition and mineralization of leaf litter resulting in an initial pulse of C, N, and P, leading to (2) an increase of aerobic microbial activity such as nitrification, which causes (3) increased biological productivity. This leads to (4) anoxia increasing denitrification, and after subsequent cessation of inundation (5) aerobic nitrification and mineralization of organic matter during drying of floodplain soils, followed by (6) bacterial mortality and release of N caused by cell lysis after desiccation, and if time allows (7) colonization of exposed sediments by terrestrial plants leading to N accumulation in plant biomass [*Baldwin and Mitchel*, 2000].

The frequency of surface connection (location 1), or inundation (location 2) determines whether all of these sequential processes occur. It also influences the rate of processes in the following event, because it affects the conditions under which processes in the following event take place. Thus a causal relation exists between the temporal pattern of the alternating conditions at locations 1 and 2, and the side arm NO_3^--N concentration. This causal relation can not be explicitly modeled if the *supt* becomes larger then duration of the alternating conditions. This problem can be solved by using a lumped empirical relationship having an aggregate of this temporal pattern as input. For example, a positive relation exists between total duration of surface connection and average measured NO_3^--N concentrations in reconnected side arms (Figure 5.2) (in agreement with Figure 5.7 and Hein et al. [2004a]).

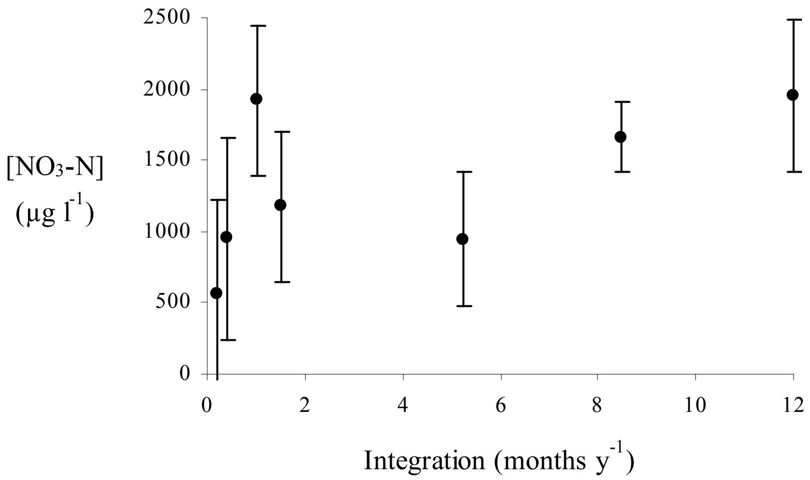

Figure 5.2 Measured NO_3^--N concentration versus total duration of surface connection (months per year with surface connection between river and different side arms). Mean and standard deviations of NO_3^--N concentrations are plotted. NO_3^--N concentration is measured in the main channel of the Danube and in different reaches of side arms near Regelsbrunn and Orth, both before and after reconnection [*Hein et al.*, 2004a]

We use indicator B8 to assess what would be the most appropriate *supt* for using this empirical relation in an SMP to model average NO_3^--N concentration in the reconnected side arm of the Danube near Regelsbrunn (Figure 5.1). We use a time series consisting of hourly water level measurements from January 2000 to May 2004 in the Danube near the inflow point of the reconnected side arm (Figure 5.3) [*Schiemer and Reckendorfer*, 2004]. In the side arm (location 1), surface connection with the river is established at water levels above 166 cm (50 cm below long term mean water level). In this application we assume that inundation of most littoral sediments (location 2) is established above long term mean water level.

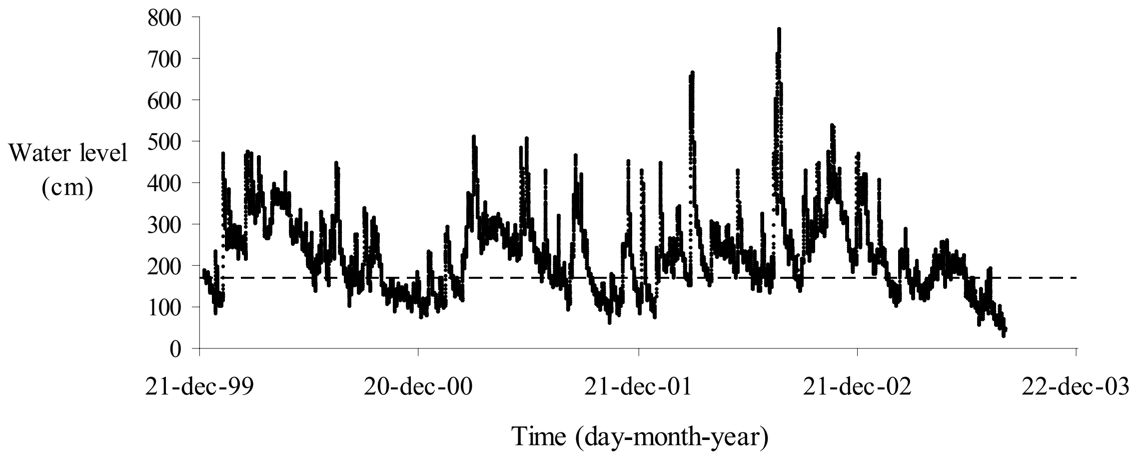

Figure 5.3 Hourly water levels in the Danube near a reconnected side arm in Regelsbrunn between December 1999 and May 2004 provided by the Austrian River Authority (water level Wildungsmauer). Dashed line indicates the water level of inflow areas of the Regelsbrunn side arm. Above this level there is surface connection between the Danube and the reconnected side arm in Regelsbrunn (166 cm).

Values of indicator B8 are calculated for different values of *supt*. This is done by calculating the correlation coefficient (r^2) between total duration of surface connection and other aggregates of the temporal pattern in both locations 1 and 2 (Equation 1). A higher correlation indicates that total duration of surface connection is a better aggregate. The aggregates which we will compare with the total duration of surface connection are frequency of surface connection in locations 1 and 2, average duration of surface connection events in location 1 and average duration of inundation events in location 2. The required values of the aggregates are calculated using the hourly water level time series and a moving window with a window size of *supt*. Weights $w_{j,i}^{B8}$ are set to equal values because we have no reason to expect that one of the two considered aggregates in any of the two locations is more important than the other. Also the weights w_j^{B8} of considered locations 1 and 2 were set to equal values.

Resulting values of $f_{B8}(supt)$ are plotted for different values of *supt* between 100 and 8500 hours (Figure 5.4). This range is chosen because $f_{B8}(supt)$ does not vary with *supt* below a *supt* of 100 hours and *supt* becomes too large relative to the size of the used time series if *supt* is above 8500 hours. The value of indicator B8 increases with *supt* values. The slight decrease in appropriateness beyond 6000 hours may be an artifact because the window size used in the window operation then is almost a quarter of the size of the used time series.

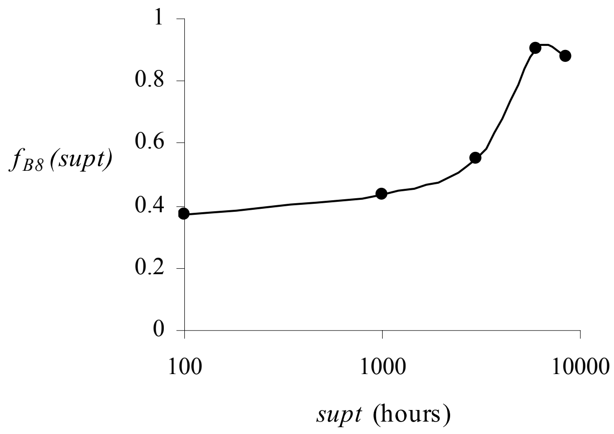

Figure 5.4 Values of indicator B8 versus *supt* for a supposed empirical SPM modeling NO_3^--N concentration in a reconnected side arm of the Danube near Regelsbrunn (Austria). The SMP uses an empirical function to relate total duration of surface connection of the Regelsbrunn side arm to NO_3^--N concentration in this side arm. The higher the value of $f_{B8}(supt)$, the more appropriate the SMP scale (*supt*)

Thus indicator B8 indicates for the considered SMPs that in the tested range of *supt* (between four days and one year), large *supt* values are more appropriate than small ones.

5.4.2 Indicator B3a

Here we use the piecewise linear approach (Section 5.3.2) for indicator B3a. The piecewise linear equation considered describes the effect of the Danube water level on the difference between NO_3^--N concentration of water flowing in and out of the reconnected side arm near Regelsbrunn (through the inflow area and outflow area, respectively, in Figure 5.1). This equation describes the relationship shown in Figure 5.5, of concentration differences in the side arm versus Danube water level. Figure 5 shows that there are three approximately linear parts, being horizontal at river discharges below 166 cm (levels without surface connection) and above 240 cm. Between 166 and 240 cm it shows an approximately linear decrease. We suppose that this equation is part of an SMP.

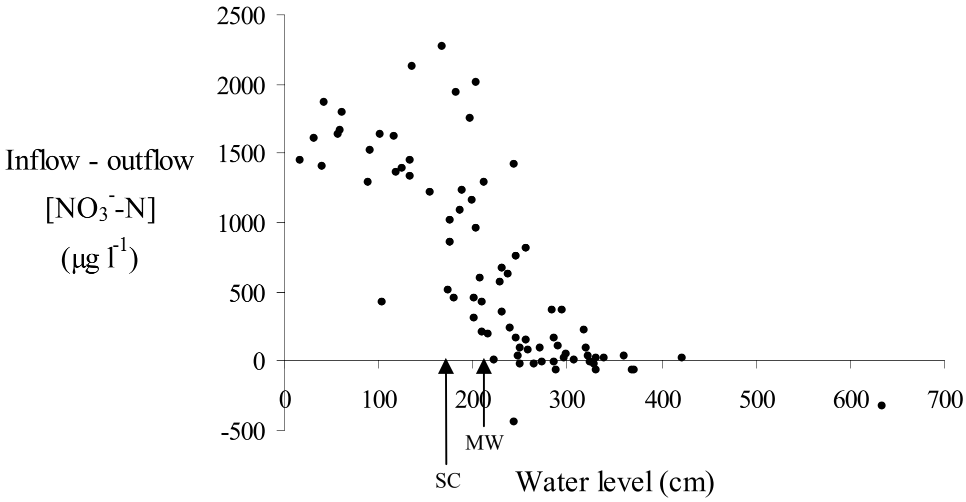

Figure 5.5 Difference between NO_3^--N concentration of water flowing in and out of the Regelsbrunn side arm versus water level in the Danube near the side arm. MW = mean water level at the inflow point of the Regelsbrunn side arm. Surface connection (SC) between Danube and side arm is established at water levels above 166 cm (50 cm below MW).

The three linear parts correspond to three different states of the side arm. For water levels that are below the level of the inflow point of the Regelsbrunn side arm (166 cm) the dominant mechanism of N exchange with the Danube is N leaching. The difference between NO_3-$^-$N concentration of in and outflow area is then relatively high due to low water discharge in the side arm, causing a long retention time (first linear part). If river water levels are above the level of the inflow point, N exchange by surface water flow occurs [*Hein et al.*, 1999]. If the water level increases further then surface water flow becomes larger leading to decreasing residence time between in and outflow areas. This decreases the difference between the NO_3^--N concentrations in these areas (second linear part). If the water level exceeds 240 cm, the surface water inflow is so large that N retention in the side arm is negligible compared to N inflow. This leads to differences between the NO_3^--N concentration of in and outflow areas that are close to zero (third linear part).

Fraction $F_i(supt)$ (Equation 2) for the previously described relationship can be calculated using an hourly time series of water level measurements in the Danube near the inflow point of the side arm (Figure 5.3). In this application, $F_i(supt)$ is the probability that a randomly selected time interval with the duration of *supt* exclusively coincides with a time period with water levels below 166 cm, between 166 and 240 cm, or above 240 cm. The first step in calculating $F_i(supt)$ consists of converting the hourly water level time series into a time series having a value of 1 at hours with water levels below 166 cm, 10 at water levels between 166 and 240 cm, and 100 at water levels above 240 cm. This transformed time series is smoothed using a moving average with a window size of *supt*. The fraction of time where the resulting smoothed time series is either 1, 10 or 100 for a particular *supt* is $F_i(supt)$. Is this application, $F_i(supt)$ equals $f_{B3a}(supt)$ because N^{NLeq} is one (Figure 5.6).

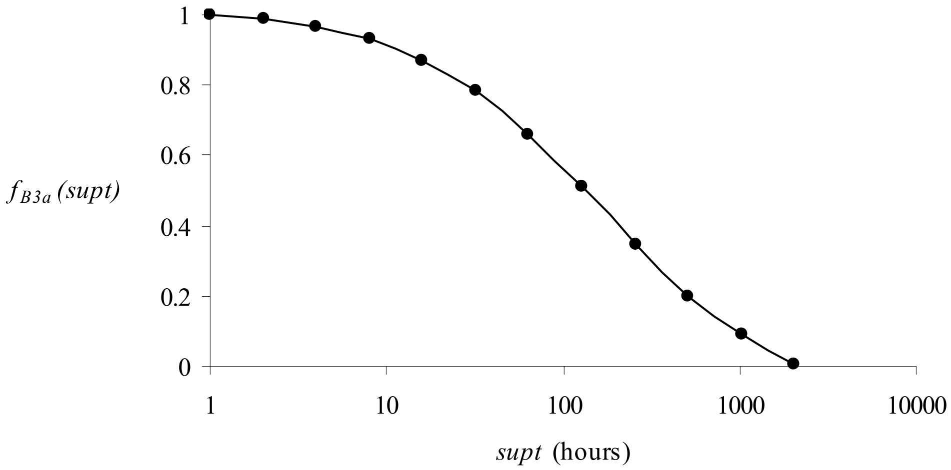

Figure 5.6 The value of indicator B3A ($f_{B3a}(supt)$) for different temporal SMP scales (*supt*). The higher the value of $f_{B3a}(supt)$, the more appropriate the *supt*.

Figure 5.6 shows that $f_{B3a}(supt)$ decreases with increasing *supt*, indicating that the previously considered model equation is valid for a smaller part of the modeled time period if *supt* is larger. For example, the equation considered is valid for at least 80 % of the time period simulated when $f_{B3a}(supt)$ is at least 0.8. Figure 5.6 shows that this is true for *supt* shorter than about 100 hours, i.e. about 4 days.

Thus indicator B3a indicates for the SMPs considered that in the *supt* range between 1 and 2050 hours, small *supt* values are more appropriate than large ones. Below one hour it is one and above 2050 hours it is zero.

5.4.3 Indicator B5a

The appropriate values of *supt* or *extt* for a supposed static or dynamic SMP, respectively, for describing the water column NO_3^--N concentration in the reconnected side arm near Regelsbrunn, is tested by Equations 3 and 4. The only N store that we will consider is the water column of the side arm. Therefore, both the values of N^{store} and w_i^{B5a} equal one. We assume that other indicators of FAMOS suggest that a *sups* corresponding to the size of the whole reconnected side arm is appropriate. Therefore we will fix *sups* to this size in Equation 4.

No suitable measurements are available to calculate the values of $s_i(sups,t)$ and $r_i(t)$ (Equation 4). Therefore, the maximum turnover time, $max(s_i(sups,t)/r_i(t))$, is estimated using knowledge of processes in the water column of the side arm and time series data containing simultaneous measurements of NO_3^--N concentration and water age [*Hein et al.*, 2004a] at the outflow point of the side arm at 78 points in time (*t*) between 1997 and 2004 [*Hein et al.*, 2004a; *Preiner*, 2003]. The time periods between pairs of subsequent measurements vary in length. Water age is an adapted metric of residence time measuring water exchange in multi input systems [*Baranyi et al.*, 2002].

NO_3-N inflow in surface water and groundwater may become so low that most NO_3^--N entering the side arm in surface and groundwater does not reach the outflow point before it is taken up or denitrified. Figure 5.7 indicates that these circumstances are reached at water ages of about 30 days because then the side arm NO_3^--N concentration in outflowing water seems to be independent of water age. Therefore, 30 days is our estimate of $max(s_i(sups,t)/r_i(t))$. As a result indicator B5a has a value of one for *supt* or *extt* longer then 30 days, indicating that these are appropriate scales for a static or dynamic SMP, respectively, for describing side arm NO_3^--N concentration.

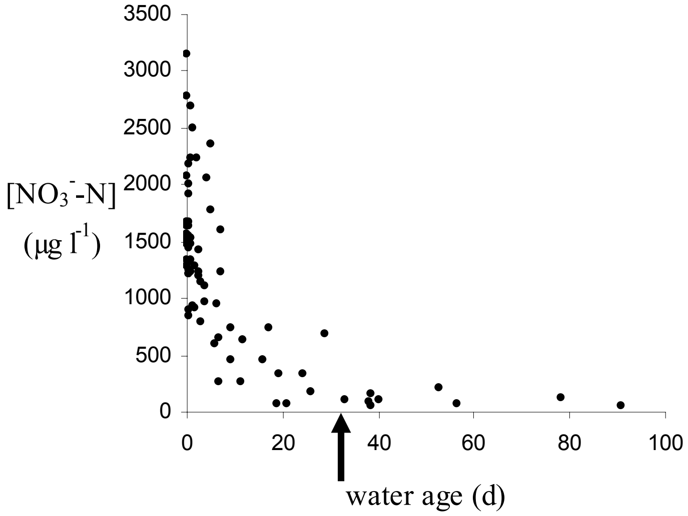

Figure 5.7 NO_3^--N concentration versus water age at the outflow point of the reconnected side arm Regelsbrunn between 1997 and 2004 (n=78). Water age is an adapted metric of residence time measuring water exchange in multi input systems [*Baranyi et al.*, 2002]. The bold arrow indicates the water age that approximately equals the maximum turnover time of NO_3^--N in the side arm. Indicator B5a ($f_{B5a}(S, sup\,s)$) as applied in Section 5.4.3 is one for values of *supt* above this value. Otherwise it is zero.

5.4.4 Implications of the new indicators in FAMOS

Scales for which the new indicators B8, B3a and B5a have high values are more likely to be identified as appropriate by FAMOS (Equations 5 to 7) than those scales for which they have low values. Using the three new indicators increases the probability that the following temporal scales of the supposed SMP calculating the NO_3^--N concentration in the side arm are appropriate: *supt* values larger than 8 months (static SMPs), *supt* values of 1 to 3 hours (dynamic SMPs), and *extt* values larger then 1 month (dynamic SMPs). It should be noted that we have given equal weights to the indicators (Equation 7).

5.5 Discussion and conclusions

In this study we adjusted the framework for identification of appropriate model scales (FAMOS) to make it suitable for application to models of N flow between rivers and reconnected floodplains. Three new indicators for the appropriateness of modeling scale are defined: (B8) the reliability of empirical relationships for modeling the aggregated effect of alternating conditions on N flow, (B3a) the validity of particular nonlinear model equations, and (B5a) the effect of inertia on model reliability.

The new indicators may be useful for identifying the appropriate modeling scale for other systems, too. For example, indicator B3a applies to all systems that are modeled with relatively reliable equations and input data. This condition can not only be fulfilled by models of reconnected floodplains, but also by models of N flow in, for example, small watersheds, or through buffer zones between agricultural land and wetlands. Indicators B8 and B5a apply to all systems where N exchange occurs, as opposed to unidirectional N flow. This may also occur in systems other than reconnected floodplains, such as estuaries where directions of N flow are diverse and vary with tide [*Shaw et al.*, 1998].

The uncertainties associated with the three new indicators depend on the data or process knowledge used in their application, the concept underlying their equations, and the values assumed by the FAMOS user for the weights (*w*). Another source of uncertainty is that the FAMOS user may apply an indicator to a part of an SMP. For example, the user may apply Equation 2 only to those SMP equations for which it is known how they were derived. Equation 3 is likely to be applied only to those parts of the SMP representing subsystems for which N budgets are relatively well known, and subsystems that can be clearly distinguished from other subsystems. Equation 1 is more likely to be used for processes that are relatively well known on a finer time scale compared to the modeling scale. Selecting SMP parts thus depends on the specific knowledge of a FAMOS user, and this may lead to different results.

It should be noted that our application of indicators B8 and B3a for modeling NO_3^--N concentration in the Regelsbrunn side arm is exceptional because often there is no suitable data available for a finer scale than the modeling scale. In such cases the scale-dependent parameters of indicator B8 and B3a ($r_{j,i}^2(supt)$ and $F_i(S)$, respectively) need to be estimated based on expert judgment. In developing indicators B8, B3a, and B5a we allowed for using expert judgment as a basis.

The three new indicators can generally be quantified with existing data and relatively simple calculations. Therefore, we expect that they can be successfully applied with acceptable effort in models of N flow between rivers and reconnected floodplains, and provide a basis for making transparent judgments of appropriateness of spatial and temporal scales of floodplain models. They inform the modeler about factors that determine the appropriateness of scales of the model considered. Moreover, they aid in communicating the rationale behind those scale choices. When merged with FAMOS, these new indicators warrant a comprehensive argumentation behind a scale choice.

Acknowledgements

We thank Carolien Kroeze, Lex Bouwman, Alfred Stein, Evert Jan Bakker, and Rik Leemans for reviewing drafts of this paper. We are grateful to Jan Sendzimir of the Risk and Vulnerability Program (IIASA) for useful discussion during development of this paper. Egon Dumont is grateful to Netherlands Organization for Scientific Research (NWO) for funding his research and for facilities made available by Wageningen University and the International Institute for Applied Systems Analyses (IIASA).

6 Discussion and conclusions

6.1 Discussion

This section includes a discussion of the framework and the analyses performed. First, FAMOS is compared to other approaches. This is followed by a discussion of the applicability of FAMOS.

Comparison to other approaches to identify appropriate modeling scales

Some contemporary approaches indicating appropriateness of modeling scales identify effects of model scale on model validity. Examples are methods aiming at identifying scales where the modeled system is deterministic [*Bruneau et al.*, 1995; *Habeeb et al.*, 2005; *Robin et al.*, 1995; *Wood et al.*, 1988] or methods based on uncertainty analyses [*Beven*, 1995; *Heuvelink*, 1998; *Lilburne et al.*, 2004; *Vachaud and Chen*, 2002]. Other approaches support the identification of the ranges of measurement scale at which different nitrogen processes can be observed, such as wavelet analyses [*Platt and Denman*, 1975] and variogram analyses [*Isaaks and Srivastava*, 1998; *Kotliar and Wiens*, 1990]. Also approaches exist supporting the deduction of scales at which different nitrogen processes can be observed. Examples are approaches using knowledge of scale-specific feedbacks controlling these processes [*Easterling and Kok*, 2002; *Gibson et al.*, 2000; *Holling*, 1992; *Levin*, 1992] or knowledge of fractal properties making the processes apparent over a range of scales [*Schneider*, 1994; *Schroeder*, 1991; *Sposito*, 1998]. A disadvantage of most approaches reported in existing methodological literature is that they require too much time and data for most practical applications. Furthermore, most of these reported methods focus on only one measure of scale and on either the scale of processes or single model parts. FAMOS does not have these limitations.

Only one other generic approach to identify appropriate scales of multiple model parts exists, called ScaleMatcher [*Lilburne et al.*, 2004]. ScaleMatcher implicitly considers criteria for appropriateness of scale that are similar to criteria A, B and D (Chapter 3) of FAMOS: correspondence of model scale with (A) data and scenarios, (B) model assumptions, and (D) appropriately scaled predictions. ScaleMatcher considers only spatial support and spatial extent and does not distinguish the interactions between the scales chosen for different SMPs, and the resources required for adjusting their scales. It is structured as a flowchart guiding the user trough a sequence of tests of appropriateness of scale. The disadvantage of this structure is that it difficult for a user to omit or add tests if required. FAMOS is more flexible. Also ScaleMatcher strongly relies on demanding sensitivity and uncertainty analyses to identify the effect of considered scales on model output. These analyses are not included in FAMOS because they require too many resources from the FAMOS user unless the model is extremely simple.

We learned that temporal scales and stream orders interact with the appropriateness of spatial support and spatial extent. The fact that Lilburne et al. [2004] do not take into account temporal scales and stream order is not only problematic because it provides the user with less information, but especially because it neglects these important interactions.

Applicability of the developed framework

FAMOS is developed for coarse scale models of one or more watersheds. However, we also aimed to assess the applicability of FAMOS to models of smaller systems. The applicability of FAMOS has been assessed for model equations derived on the global scale (NEWS-DIN) in Chapter 4 and for model equations derived on the floodplain scale in Chapter 5. In both chapters indicators for the appropriateness of modeling scale could be quantified with existing knowledge or data, and relatively simple calculations. Therefore, we expect that FAMOS can be successfully applied with acceptable effort to other models.

The application of FAMOS in Chapters 4 and 5 indicates that FAMOS can be used for a wide range of models. Model equations considered varied from process based to empirical (Box 1.1) and were derived on very different scales. FAMOS was directly applicable to the global NEWS-DIN model (Chapter 4). In Chapter 5, however, we showed that three new indicators were needed to make FAMOS applicable to floodplain modeling. These indicators are adjusted to the specific case of floodplain models. Because of the variety of currently considered model equations and the possibility to add indicators, we expect that FAMOS is sufficiently flexible to be adapted to many other types of N-flow models.

The relative simplicity of NEWS-DIN allowed us to consider all its model equations in FAMOS. However this may not be possible for models with more components and equations. In these models, FAMOS can probably only focus on part of the model equations. Ideally, those are the equations that have the largest effect on the appropriateness of model scales. The effect on the results of selecting and analyzing only part of the model equations in FAMOS has not yet been tested but that can be accomplished.

6.2 Conclusions

The main objective of this thesis was to increase our ability to identify appropriate spatial and temporal scales for N-flow models in a transparent and comprehensive way. In order to meet this objective four sub-objectives were addressed (Section 1.5). Our conclusions regarding the four sub-objectives and the main objective are presented here.

Global modeling of N flows

- A model (NEWS-DIN) was developed that predicts dissolved inorganic nitrogen (DIN) export to coastal waters of the world with higher fit, precision, and less bias than contemporary models.

- DIN export from watersheds to coastal waters is modeled as a function of N inputs to watersheds including manure N, fertilizer N, NO_x deposition, and biological N_2 fixation, as well as sewage. NEWS-DIN also includes retention and loss terms, including N retention in river networks, N retention in dammed reservoirs, N loss through consumptive water use, denitrification in rivers, and N loss through harvesting and grazing.

- Modeled DIN yield ranges from 0 to 5217 kg N km^{-2} y^{-1} with the highest DIN yields occurring in European and South East Asian basins. Modeled global DIN export to coastal waters is 25 Tg N y^{-1}, with 16 Tg N y^{-1} from anthropogenic sources. Biological N_2 fixation is the dominant source of exported DIN. Globally, and on every continent except Africa, N fertilizer is the largest anthropogenic source of DIN export to coastal waters.

- The percentage of N loaded onto basins from point and non-point sources (both anthropogenic and natural) exported as DIN ranges from 0.0001% to 43% for basins larger than ten 0.5 × 0.5 º grid cells. Lowest retention rates were calculated for basins with high runoff, highest retention rates were calculated for basins with extensive damming or very low annual precipitation and runoff. According to NEWS-DIN, in exoreic dry areas of Africa, Asia, Australia and North America (30% of the exoreic world), 95% (on average) of N applied to watersheds is not exported to coastal waters as DIN.

- The scales of NEWS-DIN application were chosen for pragmatic reasons. Scales at which NEWS-DIN was applied were the same as the scales at which NEWS-DIN was validated and calibrated. Datasets and model program files used for the NEWS-DIN application were largely the same as those used for validation and calibration. For practical reasons, the NEWS-DIN application was such that no new data or new model program files were needed.

- The supports of NEWS-DIN during calibration and validation were mainly data driven. Spatial and temporal supports during calibration and validation were chosen as fine as possible for the sake of better process description. The finest support that was possible was determined by (1) the scale of available data that could be related to N flow with existing knowledge or (2) the scale of existing knowledge to relate available data to N flow. The available data was limited because the data needed to have a global coverage.

A framework for identifying the appropriate spatial and temporal scales of N-flow models (FAMOS)

- A framework (FAMOS) was developed which is an improvement on current methods used to identify appropriate modeling scales of N-flow models. In contrast to contemporary methods FAMOS is based on multiple criteria for appropriateness of scale, multiple scalable model parts (SMPs), and multiple measures of scale. We expect that FAMOS identifies all important measures of appropriate spatial and temporal modeling scales with respect to all aspects of the model quality.

- FAMOS has four criteria that appear to be a comprehensive basis for identifying the appropriateness of the modeling scale of any N-flow model. These criteria require modeling scales to correspond with (A) data and scenarios, (B) model assumptions, (C) available resources for modeling, and

(D) appropriately scaled predictions. User-defined weights can be assigned to each criterion. They make FAMOS a flexible system that can take into account many different research objectives.

- To estimate the fulfillment of each criterion different indicators can be used. These indicators are variables which values increase with the degree of fulfillment of a criterion. Fifteen indicators are proposed in this thesis. These are generally applicable as well as a sufficient basis to identify the appropriate modeling scale. Moreover they can be easily substituted by user-defined alternatives or used in combination with new indicators that may be developed in the future. User-defined weights are to be assigned to each indicator.

- FAMOS is a flexible system that can take into account different indicators and weights. When testing scales different input data, research objectives, users, and N-flow models can be considered. The selection of indicators by the FAMOS user to estimate the fulfillment of a criterion depends on the system in which land to water N flow is modeled, the type of model equations, the available data, the resources for modeling, the research objective, and the prediction users.

- An important characteristic of FAMOS is that it is not used to identify appropriate scales for models, but for scalable model parts (SMPs). An SMP is a set of either model equations or model predictions. It has a unique scale in existing model applications, and its scale can be adjusted. Distinction between SMPs in FAMOS increases the reliability of estimates of appropriateness of modeling scales, because it enables FAMOS to account for interactions between SMPs. These interactions are that the scale of an SMP affects the appropriateness of scales of other SMPs. Further, the appropriateness of scales of an SMP is affected by the scales of other SMPs.

- Distinction between different measures of spatial and temporal scale in FAMOS increases reliability of estimates of appropriateness of modeling scales. It enables FAMOS to account for interactions between these measures. These interactions are that the appropriateness of values of a measure of SMP scale are usually affected by values chosen for another measure of SMP scale.

Application of FAMOS to NEWS-DIN, a model of global N flows

- FAMOS was applied to each of the six SMPs of NEWS-DIN. Ranges of appropriate scales could be identified for each of these SMPs.

- Most of the scales used in the original application of NEWS-DIN (Chapter 2) are appropriate. The spatial support and temporal extent of some SMPs are not appropriate. For the SMPs describing riverine DIN retention and riverine DIN export no appropriate values of temporal support were found.

- The application of FAMOS has indicated that scale of application of NEWS-DIN could be improved if (1) knowledge or data sufficient to model riverine processes on a smaller temporal support is obtained or if (2) if NEWS-DIN is used to hindcast or model scenarios of the future. This may be the key to better predictions of regional to global DIN export.

Applicability of FAMOS for models of N flow between rivers and floodplains

- Three additional indicators of appropriateness of modeling scale were required to make FAMOS suitable for models of N in reconnected floodplain systems. These indicators are: (1) the reliability of empirical relationships for modeling the aggregated effect of alternating conditions on N flow, (2) the validity of particular nonlinear model equations, and (3) the effect of inertia on model reliability. These indicators improved the ability of FAMOS to estimate the effect of different modeling scales on validity of such models.

- The three new added indicators are important for estimating the appropriateness of scales for modeling reconnected floodplains because the conditions that justify the introduction of these indicators all occur in reconnected floodplain systems. These three new indicators may also make FAMOS suitable for identifying the appropriate scale of models for other relatively small systems, such as small watersheds, estuaries, or buffer zones between agricultural land and wetlands.

- Application of the three new indicators shows that they can be used to identify scales that are appropriate for modeling of nitrate concentrations in a side arm in a reconnected floodplain of the Danube.

Our understanding of identification of appropriate spatial and temporal scales for models of N flow from land to water

The overall aim of this thesis was to increase our understanding of identification of appropriate spatial and temporal scales for N-flow models. We now present some overall conclusions on identifying appropriate scales for N-flow models.

First, using the most appropriate scales in modeling ensures an optimal balance between available data and scenarios, validity of model assumptions, effort needed for modeling, and usefulness of model output. This is independent of the modeled system and the type of model equations.

Second, appropriateness of model scales is affected by many factors. Therefore, identification of appropriate modeling scales requires a comprehensive overview of these factors, being various properties of certain model equations, datasets, knowledge, software and research objectives. Furthermore, it requires an appraisal of the sensitivity of scale appropriateness to individual factors in order to make a proper selection of these factors.

Third, appropriate scales need to be identified for scalable model parts (SMPs), not for a whole model. It should be realized that scales of SMPs affect the appropriateness of scales of other SMPs. These relations between SMPs must be taken into account when identifying appropriate modeling scale.

Fourth, within an SMP, the value chosen for a measure of scale always affects the appropriateness of other measures of scale. Such relations must be taken into account when identifying appropriate modeling scale and can be either positive or negative. They can involve spatial support, temporal support, spatial extent, temporal extent and stream order.

Fifth, the selection of FAMOS indicators to be used depends to a large extent on the characteristics of the modeled system and the type of model equations.

Finally, FAMOS has proved to be a useful tool to identify the appropriate scales of a wide range of models. It is the first method that leads to a comprehensive identification of the most appropriate combination of spatial and spatial and temporal scales of a model. Also it is the first method to account for all major factors affecting appropriateness of model scales. Future applications of FAMOS are expected to further prove the quality of this framework and to result in improved application of N-flow models.

References

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

Alexander, R. B., J. R. Slack, A. S. Ludtke, K. K. Fitzgerald, and T. L. Schertz, Data from selected U.S. Geological Survey national stream water quality monitoring networks (WQN). U.S.Geological Survey., 1996.

Andersen, H. E., S. Anthony, B. Arheimer, A. Barr, H. Behrendt, F. Bouraoui, H. Ejhed, P. Groenendijk, M. Jeuken, H. Johnsson, B. Kronvag, G. Le Gall, A. Murdock, A. Lo Porto, L. Price, O. Schoumans, M. Silgram, R. Smit, E. Varanou, and U. Zweynert, Review and literature evaluation of quantification tools for the assessment of nutrient losses at catchment scale, *no 1*, 120, Norwegian Institute for Water Research, Norway, 2003.

Andersen, H. E., S. Anthony, B. Arheimer, A. Barr, H. Behrendt, F. Bouraoui, H. Ejhed, P. Groenendijk, M. Jeuken, H. Johnsson, B. Kronvang, G. Le Gall, A. Murdock, A. Lo Porto, L. Fawcett, O. Schoumans, M. Silgram, R. Smit, E. Varanou, and U. Zweynert, Modelling approaches: Model parameterisation, calibration and performance assessment methods in the EUROHARP project, *EUROHARP 8-2004*, 18, NIVA, Oslo, 2004.

Anthony, S. G., P. Quinn, and E. I. Lord, Catchment scale modelling of nitrate, *Aspects of Applied Biology*, 23-32, 1996.

Baldwin, D. S., and A. M. Mitchel, The effects of drying and re-flooding on the sediment and soil nutrient dynamics of lowland river-floodplain systems: a synthesis, *Regulated Rivers: Research & Management*, *16*, 457-467, 2000.

Baranyi, C., T. Hein, C. Holarek, K. S., and F. Schiemer, Zooplankton biomass and community structure in a Danube River floodplain system: effects of hydrology, *Freshwater Biology*, *47*, 473-482, 2002.

Beaujouan, V., P. Durand, and L. Ruiz, Modelling the effect of the spatial distribution of agricultural practices on nitrogen fluxes in rural catchments, *Ecological Modelling*, *137*, 93-105, 2001.

Becker, A., and P. Braun, Disaggregation, Aggregation and spatial scaling in hydrological modelling, *Journal of hydrology*, *217*, 239-252, 1999.

Beckie, R., A comparison of methods to determine measurement support volumes, *Water Resources Research*, *37*, 925, 2001.

Behrendt, H., P. Huber, M. Kornmilch, D. Opitz, O. Schmoll, G. Scholz, and R. Uebe, Estimation of the nutrient inputs into river basins - experiences from German rivers, *Regional Environmental Changes*, 2002.

Behrendt, H., and D. Opitz, Retention of nutrients in River systems: dependence on specific runoff and hydraulic load, *Hydrobiologia*, *410*, 111-122, 2000.

Bellamy, P. H., and P. J. Loveland, Effects on model output of different data sources, SWANCATCH and SWATCATCH, 34-56, Soil Survey & Land Research Centre, Cranfield University, Silsoe, 2001a.

Bellamy, P. H., and P. J. Loveland, Upscaling, predictive models and catchment water quality, *ENV4-CT97-0439*, 34-56, Commission of the European Communities, Cranfield, 2001b.

Beusen, A. H. W., A. L. M. Dekkers, A. F. Bouwman, W. Ludwig, and J. Harrison, Estimation of global river transport of sediments and associated particulate C, N and P, *Global Biogeochemical Cycles*, *Submitted*, 2005.

Beven, K., Linking parameters across scales: subgrid parameterizations and scale dependent hydrological models. In: Scale issues in hydrological modelling, in *Scale issues in hydrological modelling*, edited by J. D. Kalma, and M. Sivaplan, 9-48, John Wiley, Chichester, 1995.

Beven, K. J., Prophecy, reality and uncertainty in distributed hydrological modelling, *Advances in water resources*, *16*, 41-51, 1993.

Bierkens, M. F. P., P. A. Finke, and P. De Willigen, *Upscaling and downscaling methods for environmental Research*, Kluwer Academic Publishers, Dordrecht, 2000.

Billen, G., J. Garnier, and P. Hanset, Modelling phytoplankton development in whole drainage networks: the Riverstrahler model applied in the Seine river system, *Hydrobiologia*, *289*, 119-137, 1994.

Blöschl, G., and M. Sivapalan, Scale issues in hydrological modelling: a review, in *Scale issues in hydrological modelling*, edited by J. D. Kalma, and M. Sivaplan, 1995.

Borah, D. K., and M. Bera, Watershed-scale hydrologic and nonpoint-source pollution models: Review of mathematical bases, *Transactions of the ASAE*, *47*, 789-803, 2004.

Bouwman, A. F., G. Van Drecht, J. M. Knoop, A. H. W. Beusen, and C. R. Meinardi, Exploring changes in river nitrogen export to the world's oceans, *Global Biogeochemical Cycles*, *19*, 2005a.

Bouwman, A. F., G. Van Drecht, and K. W. Van der Hoek, Nitrogen surface balances in intensive agricultural production systems in different world regions for the period 1970-2030, *Pedosphere*, *15*, in press, 2005b.

Boyer, E. W., C. L. Goodale, N. A. Jaworski, and R. W. Howarth, Antropogenic nitrogen sources and relationships to riverine nitrogen export in the north eastern USA, *Biogeochemistry*, *57/58*, 137-169, 2002.

Bruneau, P., C. Gasquel-odoux, P. Robin, P. Mérot, and K. J. Beven, Sensitivity analyses to time and space resolution on an hydrological modelling based on digital elevation model, *Hydrological Processes*, 69-81, 1995.

Burrough, P. A., Fractal dimensions of landscapes and other environmental data, *Nature*, *294*, 240-242, 1981.

Caraco, N. F., and J. J. Cole, Human impact on nitrate export: an analysis using major world rivers, *Ambio*, *28*, 167-170, 1999.

Caraco, N. F., J. J. Cole, J. E. Likens, G. M. Lovett, and K. C. Weathers, Variation in NO_3 Export from Flowing Waters of Vastly Different Sizes: Does One Model Fit All?, *Ecosystems*, *6*, 344-352, 2003.

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

Charles, J. V., and M. Berrien, Modeling basin-scale hydrology in support of physical climate and global biogeochemical studies: An example using the Zambezi River, *Surveys in Geophysics*, *V12*, 271-311, 1991.

Chovanec, A., P. Jäger, M. Jungwirth, V. Koller-Kreimel, O. Moog, S. Muhar, and S. Schmutz, The Austrian way of assessing the ecological integrity of running waters: a contribution to the EUWater Framework Directive, *Hydrobiologia*, 445-452, 2000.

Corwin, D. L., J. Hopmans, and G. H. de Rooij, From Field- to Landscape-Scale Vadose Zone Processes: Scale Issues, Modeling, and Monitoring, *Vadose Zone Journal*, *5*, 129-139, 2006.

Costanzo, S. D., M. J. Donohue, and W. C. Dennison, Assessing the Seasonal Influence of Sewage and Agricultural Nutrient Inputs in a Subtropical River Estuary, *Estuary*, *26*, 857-865, 2003.

Curmi, P., P. Durand, C. Gascuel-Odoux, P. Merot, C. Walter, and A. Taha, Hydromorphic soils, Hydrology and waterquality: spatial distribution and functional modelling at different scales, *Nutrient Cycling in Agroecosystems*, *50*, 127-142, 1998.

De Floriani, L., and P. Magillo, Regular and irregular multi-resolution terrain models: a comparison, in *10th ACM international symposium on Advances in geographic information systems* 143 - 148, ACM Press, McLean, Virginia, USA, 2002.

De Koning, G. H. J., A. Veldkamp, K. Kok, N. De Ridder, L. O. Fresco, and J. Schoorl, Multi-scale land use modelling with the clue modelling framework, *410200053*, Wageningen University, Wageningen, 2000.

De Vries, W., Soil response to acid deposition at diferent regional scales, Ph.D. thesis, Wageningen, The Netherlands, 1994.

De Wit, M., Nutrient fluxes in the Rhine and Elbe basins, Utrecht University, Utrecht, 1999.

Dent, C. L., and N. B. Grimm, Spatial heterogeneity of stream water nutrient concentrations over successional time, *Ecology*, *80*, 2283–2298, 1999.

Dentener, F., Global Patterns and Magnitudes of Nitrogen Deposition, a Revised Approach, *Global Biochemical Cycles*, *submitted*, 2004.

Desbarats, A. J., Spatial averaging of hydraulic conductivity in three-dimensional heterogeneous porous media, *Mathematical Geology*, *24*, 249-267, 1991.

Diaz, R. J., J. Nestlerode, and M. L. Diaz, A Global Perspective on the Effects of Eutrophication and Hypoxia on Aquatic Biota, in *7th Annual Symposium on Fish Physiology, Toxicology and Water Quality*, edited by G. L. Rupp, and M. D. White, U.S. Environmental Protection Agency, Ecosystems Research Division, Estonia, 2003.

Dumont, E., J. Harrison, C. Kroeze, E. J. Bakker, and S. Seitzinger, Global distribution and sources of DIN export to the coastal zone: results from a spatially explicit, global model, *Global Biogeochemical Cycles*, 2005a.

Dumont, E., J. A. Harrison, C. Kroeze, E. J. Bakker, and S. P. Seitzinger, Global distribution and sources of dissolved inorganic nitrogen export to the coastal zone: Results from a spatially explicit, global model, *Global Biogeochemical Cycles*, *19*, doi:10.1029/2005GB002488, 2005b.

Dumont, E., H. Hein, and S. Preiner, Identification of appropriate spatial and temporal scale for modeling N flows in reconnected floodplains, submitted-a.

Dumont, E., C. Kroeze, E. J. Bakker, A. Stein, and A. F. Bouwman, Application of a framework to identify the appropriate scale for a model developed to predict global river export of dissolved inorganic nitrogen in *Accuracy 2006: proceedings of the 7th international symposium on spatial accuracy assessment in natural resources and environmental sciences*, edited by M. Caetano, and M. Painho, 730-739, Lisbon, 2006.

Dumont, E., C. Kroeze, A. F. Bouwman, E. J. Bakker, A. Stein, and R. Leemans, A framework to identify appropriate spatial and temporal scales for modeling N flows from watersheds, *Ecological modelling*, submitted-b.

Durand, P., C. Neal, H. A. Jeffery, G. P. Ryland, and M. Neal, Major, minor and trace element budgets in the Plynlimon afforested catchments (Wales): general trends, and effect of felling and climate variations, *Journal of hydrology*, *157*, 139-156, 1994.

Dynesius, M., and C. Nilsson, Fragmentation and flow regulation of river systems in the northern third of the world, *Science*, *266*, 753-762, 1994.

Easterling, W. E., and K. Kok, Emergent Properties of scale in global environmental modeling - Are there any?, *Integrated assessment*, *3*, 233-247, 2002.

Eckhardt, K., L. Breuer, and H. G. Frede, Parameter uncertainty and the significance of simulated land use change effects, *Journal of hydrology*, *273*, 164-176, 2003.

EEA, Europe's Environment. Statistical compendium for the second assessment, Copenhagen, 1998.

European Environment Agency/IFEN, Calculation of nutriment surplus from agricultural sources. Statistics spatialisation by means of CORINE Land Cover. Application to the case of Nitrogen., *51*, 2000.

FAO, World livestock production systems. Current status, issues and trends, *127*, Food and Agriculture Organization of the United Nations, Rome, 1996.

FAO, edited by FAO, FAOSTAT database collections, Rome, 2001.

Fekete, B. M., C. J. Vorosmarty, and W. Grabs, Global, Composite Runoff Fields based on Observed River Discharge and Simulated Water Balances, 2000.

FitzHugh, T. W., and D. S. Mackay, Impacts of input parameter spatial aggregation on an agricultural nonpoint source pollution model, *Journal of hydrology*, *236*, 2000.

Galloway, J. N., F. J. Dentener, D. G. Capone, E. W. Boyer, R. W. Howarth, S. P. Seitzinger, G. P. Asner, C. C. Cleveland, P. Green, E. A. Holland, D. M. Karl, A. F. Michaels, J. H. Porter, A. R. Townsend, and C. J. Vörösmarty, Nitrogen cycles: past, present and future, *Biogeochemistry*, *70*, 153-226, 2004.

Gibson, C. C., E. Ostrom, and T. K. Ahn, The concept of scale and the human dimensions of global change: a survey, *Ecological Economics*, *32*, 217-239, 2000.

Goolsby, D. A., W. A. Battaglin, B. T. Aulenbach, and R. P. Hooper, Nitrogen flux and sources in the Mississippi River Basin, *Science of the Total Environment*, *248*, 75-86, 2000.

Green, P. A., C. J. Vorosmarty, M. Meybeck, J. N. Galloway, B. J. Peterson, and E. W. Boyer, Pre-industrial and contemporary fluxes of nitrogen through rivers: a global assessment based on topology, *Biogeochemistry*, *68*, 71-105, 2004.

Grimvall, A., and P. Stålnacke, Statistical methods for source apportionment of riverine loads of pollutants, *Envirometrics*, *7*, 201-213, 1996.

Habeeb, R. L., J. Trebilco, S. Wotherspoon, and C. R. Johnson, Determining natural scales of ecological systems, *Ecological Monographs*, *75*, 467-487, 2005.

Habersack, H. M., The river-scaling concept (RSC): a basis for ecological assessments, *Hydrobiologia*, *422/423*, 49–60, 2000.

Harrison, J., S. Seitzinger, N. Caraco, L. Bouwman, A. Beussen, and C. Vörösmarty, Dissolved Inorganic Phosphorus Export to the Coastal Zone: Results From a Spatially Explicit, Global Model, *Global Biochemical Cycles*, *accepted*, 2005a.

Harrison, J. A., N. F. Caraco, and S. P. Seitzinger, Global patterns of dissolved organic carbon, nitrogen, and phosphorus export by rivers: results from a spatially explicit, global model, *Global Biogechemical Cycles*, *Submitted*, 2005b.

Hassanizadeh, M., and W. G. Gray, General conservation equations for multiphase systems: 1. Averaging technique, *Advances in water resources*, *2*, 131-144, 1979.

Hein, T., C. Baranyi, and W. Reckendorfer, Einfluss von Öffnungsmaßnahmen auf die hydrochemische Situation und die planktischen Prozesse in einem dynamischen Ausystem, *Abhandlungen der Zoologisch-Botanischen Gesellschaft in Österreich*, *34*, 31-46, 2004a.

Hein, T., C. Baranyi, W. Reckendorfer, and F. Schiemer, The impact of surface water exchange on the nutrient and particle dynamics in side-arms along the River Danube, Austria, *Science of the Total Environment*, 207–218, 2004b.

Hein, T., G. Heiler, D. Pennetzdorfer, P. Riedler, M. Schagerl, and F. Schiemer, The Danube Restauration Project: Functional aspects and planctonic productivity in the floodlain system, *Regulated Rivers: Research & Management*, *15*, 259–270, 1999.

Heuvelink, G. B. M., Uncertainty analyses in environmental modelling under a change of spatial scale, *Nutrient Cycling in Agroecosystems*, *50*, 255-264, 1998.

Hohensinner, S., H. Habersack, M. Jungwirth, and G. Zauner, Reconstruction of the characteristics of a natural alluvial river-floodplain system and hydromorphological changes following human modifications: the Danube river (1812-1991), *River Research and applications*, *20*, 25-41, 2004.

Hollander, A., M. A. J. Huijbregts, A. M. J. Ragas, and D. v. d. Meent, BasinBox: A Generic Multimedia Fate Model for Predicting the Fate of Chemicals in River Catchments, *Hydrobiologia*, *V565*, 21-38, 2006.

Holling, C. S., Cross-Scale Morphology, Geometry, and dynamics of ecosystems, *Ecological Monographs*, *62*, 447-502, 1992.

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

Hunsaker, C. T., and D. A. Levine, Hierarchical Approaches to the Study of Water Quality in Rivers, *BioScience*, *45*, 193-203, 1995.

IPCC, The Regional Impacts of Climate Change – An Assessment of Vulnerability, 517, Cambridge University Press, Cambridge, 1998.

Isaaks, E. H., and R. M. Srivastava, An introduction to Applied Geostatistics, 458-488, Oxford University Press, New York, 1998.

Jarvie, H. P., A. J. Robson, and C. Neal, Riverine mass load estimation: Paris Commission Policy, practice and implications for load estimation for a major UK river system, in *Water quality: Processes and Policy*, edited by S. T. Trudgill, D. E. Walling, and B. W. Webb, 27-45, John Wiley & Sons Ltd., Chichester, 1999.

Jarvis, P. G., Scaling processes and problems, *Plant, cell and environment*, *18*, 1079-1089, 1995.

Jha, M., P. W. Gassman, S. Secchi, R. Gu, and J. Arnold, Effect of watershed subdivision on SWAT flow, sediment, and nutrient predictions, *Journal of the American Water Resources Association*, *40*, 811-825, 2004.

Joao, E., How scale affects environmental impact assessment, *Environmental Impact Assessment Review*, *22*, 289-310, 2002.

Johnes, P. J., and D. Butterfield, Landscape, regional and global estimates of nitrogen flux from land to sea, *Biogeochemistry*, *57/58*, 429-476, 2002.

Johnes, P. J., C. Curtis, B. Moss, P. Whitehead, H. B. Bennion, and S. Patrick, Trial classification of lake water quality in England and Wales: a proposed approach, *E53*, Environment Agency, Bristol, 1998.

Jolankai, G., and B. Pataki, Description of the "Tisza River Project" and its main results. A Report on The Tisza River Project: Real-life scale integrated catchment models for supporting water- and environmental management decisions, *EVK1-CT-2001-00099*, 2005.

Jordan, T. E., D. E. Weller, and D. L. Correl, Sources of Nutrient Inputs to the Patuxent River Estuary, *Estuaries*, *16*, 226-243, 2003.

Justic, D., N. N. Rabalais, R. E. Turner, and Q. Dortch, Changes in nutrient structure of river-dominated coastal waters: stoichiometric nutrient balance and its consequences, *Estuarine, Coastal and Shelf Science*, *40*, 339-356, 1995.

Kotliar, N. B., and J. A. Wiens, Multiple scales of patchiness and patch structure: A hierarchical framework for the study of heterogeneity, *Oikos*, *59*, 253-260, 1990.

Kroeze, C., R. Aerts, N. van Breemen, D. van Dam, K. van der Hoek, P. Hofschreuder, M. Hoosbeek, J. de Klein, H. Kros, H. van Oene, O. Oenema, A. Tietema, R. van der Veeren, and W. de Vries, Uncertainties in the fate of nitrogen I: An overview of sources of uncertainty illustrated with a Dutch case study, *Nutrient Cycling in Agroecosystems*, *66*, 43-69, 2003.

Lajaunie, C., and H. Wackernagel, Geostatistical Approaches to Change of Support Problems. Theoretical Framework, *19*, Centre de Géostatistique, Fontainebleau, 2000.

Leemans, R., Scientific Challenges for Anthropogenic Research in the 21st Century: Problems of Scale, in *Earth system science in the antropocene* edited by E. Ehlers, and T. Krafft, 249-262, Springer, Bonn, 2005.

Lelieveld, J., and F. Dentener, What controls tropospheric ozone?, *Journal of Geophysical Research*, *105*, 3531-3551, 2000.

Levin, S. A., The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture, *Ecology*, *73*, 1943-1967, 1992.

Li, D., B. Beckner, and A. Kumar, A new efficient averaging technique for scaleup of multimillion-cell geologic models, in *1999 SPE Annual Technical Conference and Exhibition*, Houston, Texas, 1999.

Lijklema, L., Dimensions and scales, *Water science and technology*, *37*, 1-7, 1998.

Lilburne, L. R., T. H. Webb, and G. L. Benwell, The Scale Matcher: a procedure for assessing scale compatibility of spatial data and models, *International Journal of Geographical Information Science*, *18*, 257-279, 2004.

Mamillapalli, S., R. Srinivasan, J. G. Arnold, and B. A. Engel, Effect of spatial variability on basin scale modelling, in *Proceedings of the third international NCGIA Conference on integrated GIS and environmental modelling*, Santa Fe, 1996.

McBratney, A. B., Some considerations on methods for spatially aggregating and disaggregating soil information, *Nutrient Cycling in Agroecosystems*, *50*, 51-62, 1998.

McClain, M. E., E. W. Boyer, C. L. Dent, S. E. Gergel, N. B. Grimm, P. M. Groffman, S. C. Hart, J. W. Harvey, C. A. Johnston, E. Mayorga, W. H. McDowell, and G. Pinay, Biogeochemical Hot Spots and Hot Moments at the Interface of Terrestrial and Aquatic Ecosystems, *Ecosystems*, *6*, 301–312, 2003.

McNaughton, K. G., Effective stomatal and boundary-layer resistances of heterogeneous surfaces, *Plant, Cell and Environment*, *17*, 1061-1068, 1994.

Merete, S., and S. Børge, Modelling of N-movements on catchment scale - a tool for analysis and decision making, *Nutrient Cycling in Agroecosystems*, *36*, 7-17, 1993.

Meybeck, M., Carbon, nitrogen and phosphorus transport by world rivers, *American Journal of Science*, *282*, 401-450, 1982.

Meybeck, M., Riverine quality at the Anthropocene: Propositions for global space and time analyses, illustrated by the Seine River, *Aquatic Sciences*, *64*, 376-393, 2002.

Meybeck, M., H. H. Dürr, and S. Roussennac, Regional seas as mega-filters of the land to oceans riverine outputs, in prep.

Meybeck, M., and A. Ragu, River discharges to oceans: an assessment of suspended solids, mayor ions and nutrients, edited by UNEP, 1995.

Millennium Ecosystem Assessment, Millennium ecosystem assessment synthesis report, Island Press, Washington, D.C., 2005.

Narula, K., N. K. Bansal, A. Gosain, and F. Wendland, Area-differentiated modelling of excess nutrient and contaminant leaching risks at the river basin scale using a hydrological approach in India, *International Journal of environment and pollution*, *19*, 225-242, 2003.

Nash, J. E., and J. V. Sutcliffe, river flow forecasting through conceptual models. Part 1 - a discussion of principles, *Journal of Hydrology*, *10*, 282-290, 1970.

Neitsch, S. L., J. G. Arnold, J. R. Kiniry, J. R. Williams, and K. W. King, Soil and Water Assessment Tool Theoretical Documentation. Version 2000, Temple, Texas, 2002.

Nixon, S. W., J. W. Ammerman, L. P. Atkinson, V. M. Berounski, G. Billen, W. C. Boicourt, W. R. Boynton, T. M. Chorch, D. M. E. Ditoro, R., J. H. Garber, A. E. Giblin, R. A. Jahnke, N. J. P. Owens, M. E. Q. Pilson, and S. P. Seitzinger, The fate of nitrogen and phosphorus at the land-sea margin of the North Atlantic Ocean, *Biogeochemistry*, *35*, 141-180, 1996.

Omernik, J. M., The misuse of hydrologic unit maps for extrapolation, reporting and ecosystem management, *Journal of the American Water Resources Association*, *39*, 563-573, 2003.

Omernik, J. M., and R. G. Bailey, Distinguishing between watersheds and ecoregions, *Journal of the American Water Resources Association*, *33*, 1-15, 1997.

Parish, F., A review of river restoration experience in east Asia, in *River restoration in Asia; proceedings of the east Asia regional seminar on river restoration*, edited by F. Parish, M. Mohktar, A. R. B. Abdullah, and C. O. May, 14-23, Global Environmental Centre and Department of Irrigation and Drainage, Kuala Lumpur, 2004.

Peierls, B. L., N. F. Caraco, M. L. Pace, and C. V. Cole, Human influence on river nitrogen, *nature*, *350*, 386-387, 1991.

Platt, T., and K. Denman, Spectral analyses in ecology, *Annual Review of Ecology and Systematics*, *6*, 189-210, 1975.

Preiner, S., The importance of a re-connected side-arm for the biogeochemical cycling of the River Danube in Austria, University of Vienna, 2003.

Quinn, P., Scale appropriate modelling: representing cause-and-effect relationships in nitrate pollution at the catchment scale for the purpose of catchment scale planning, *Journal of hydrology*, *291*, 197-217, 2004.

Quynh, L. T. P., G. Billen, J. Garnier, S. Théry, C. Fézard, and C. V. Minh, Nutrient (N, P) budgets for the Red River basin (Vietnam and China), *Global Biogeochemical Cycles*, *19*, doi:10.1029/2004GB002405, 2005.

Rabalais, N. N., R. E. Turner, and D. Scavia, Beyond science into policy: Gulf of Mexico hypoxia and the Mississippi river, *BioScience*, *52*, 129-142, 2002.

Robin, P., P. Merot, and K. Beven, Sensitivity to space and time resolution of a hydrological model using digital elevation data, *Hydrological processes*, *9*, 69-81, 1995.

Roni, P., K. Hanson, T. Beechie, G. Pess, M. Pollock, and D. M. Bartley, Habitat rehabilitation for inland fisheries Global review of effectiveness and guidance for rehabilitation of freshwater ecosystems, FAO, Rome, 2005.

Ruiz, L., S. Abiven, C. Martin, P. Durand, V. Beaujouan, and J. Molénat, Effect on nitrate concentration in stream water of agricultural practices in small catchments in Brittany. II. Temporal variation and mixing processes, *Hydrology and earth system sciences*, *6*, 507-513, 2002b.

Schiemer, F., C. Baumgartner, and K. Tockner, Restoration of floodplain rivers: the Danube restoration project., *Regulated Rivers: Research & Management*, *15*, 231-244, 1999.

Schiemer, F., and W. Reckendorfer, Das Donau Restaurierungsprojekt – Ökologische Auswirkungen. The Danube Restoration Programme – ecologic consequences, *34*, 185, Vienna, 2004.

Schneider, D. S., *Quantitative Ecology, spatial and temporal scaling*, Academic Press Limited, London, 1994.

Schroeder, M., *Fractals, Chaos, Power laws. Minutes from an infinite paradise*, 429, W.H. Freeman and Company, New York, 1991.

Seitzinger, S., J. A. Harrison, J. K. Böhlke, A. F. Bouwman, R. Lowrance, B. Peterson, C. Tobias, and G. Van Drecht, Denitrification Across Landscapes and Waterscapes: a Synthesis, *Ecological Applications*, *16*, 2064-2090, 2006.

Seitzinger, S. P., Denitrification in freshwater and coastal marine ecosystems: Ecological and geochemical significance, *Limnology and Oceanography*, *33*, 702-724, 1988.

Seitzinger, S. P., Data Collection Program in Support of the Harbor-Wide Eutrophication Model for New York-New Jersey Harbor Estuary Program., U.S. EPA and ISC, New York, 1995.

Seitzinger, S. P., J. A. Harrison, E. Dumont, A. H. W. Beusen, and A. F. Bouwman, Sources and delivery of carbon, nitrogen, and phosphorus to the coastal zone: An overview of Global Nutrient Export from Watersheds (NEWS) models and their application., *Global Biogeochemical Cycles*, *19*, doi:10.1029/2005GB002606, 2005.

Seitzinger, S. P., and C. Kroeze, Global distribution of nitrous oxide production and N inputs in freshwater and coastal marine ecosystems, *Global Biochemical Cycles*, *12*, 93-113, 1998.

Seitzinger, S. P., R. V. Styles, E. W. Boyer, R. B. Alexander, G. Billen, R. W. Howarth, B. Mayer, and N. Van Breemen, Nitrogen retention in rivers: model development and application to watersheds in the northeastern U.S.A, *Biogeochemistry*, *57/58*, 199-237, 2002.

Sferratore, A., G. Billen, J. Garnier, and S. Théry, Modeling nutrient (N, P, Si) budget in the Seine watershed: Application of the Riverstrahler model using data from local to global scale resolution, *Global Biogeochemical Cycles*, *19*, doi:10.1029/2005GB002496, 2005.

Shaw, P. J., C. Chapron, D. A. Purdie, and A. P. Rees, Impacts of phytoplankton activity on dissolved nitrogen fluxes in the tidal reaches and estuary of the Tweed, UK, *Marine pollution bulletin*, *37*, 280-294, 1998.

Sherman, K., The Large Marine Ecosystem Concept: Research and Management Strategy for Living Marine Resources, *Ecological Applications*, *1*, 349-360, 1991.

Shoemaker, L., M. Lahlou, M. Bryer, D. Kumar, and Kratt.K., Compendium of Tools for Watershed Assessment and TMDL Development, United States Environmental Protection Agency, Washington, D.C., 1997.

Simmelsgaard, S. E., K. Kristensen, H. E. Andersen, R. Grant, J. O. Jørgensen, and H. S. Østergaard, An empirical model for calculation of root zone nitrate leaching., *32*, 67, Danmarks Jordbrugs Forskning, Markbrug, 2000.

Smil, V., Nitrogen and Food Production: Proteins for Human Diets, *Ambio*, *31*, 126-131, 2002.

Smith, R. A., G. E. Schwartz, and B. Alexander, Regional interpretation of water quality monitoring data, *Water Resources Research*, *33*, 2781-2781, 1997a.

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

Smith, S. V., D. P. Swaney, L. Talaue-Mcmanus, J. D. Bartley, P. T. Sandhei, C. J. McLaughlin, V. C. Dupra, C. J. Crossland, R. W. Buddemeier, B. A. Maxwell, and F. Wulff, Humans, hydrology, and the distribution of inorganic nutrient loading t the ocean., *BioScience*, *53*, 235-245, 2003.

Sparks, R. E., Need for ecosystem management of large rivers and their floodplains, *BioScience*, *45*, 162–182, 1995.

Sposito, G., *Scale dependence and scale invariance in hydrology*, Cambridge University Press, Melbourne, 1998.

Stalnacke, P., S. M. Vandsemb, A. Vassiljec, A. Grimvall, and G. Jolankai, Changes in nutrient levels in some Eastern European rivers in response to large-scale changes in agriculture, *Water science and technology*, *49*, 29-36, 2004.

Stanley, E. H., and M. W. Doyle, Trading off: the ecological effects of dam removal, *Frontiers in Ecology and the Environment*, *1*, 15-22, 2003.

Stein, A., J. Riley, and N. Halberg, Issues of scale for environmental indicators, *Agriculture, Ecosystems and Environment*, *87*, 215-232, 2001.

Sterman, J. D., *Business dynamics. Systems thinking and modeling for a complex world*, McGraw-Hill, 2000.

Strahler, A. N., Quantitative geomorphology of drainage basins and channel networks, in *Handbook of Applied Hydrology*, edited by V. T. Chow, 4-39 - 34-76, McGraw-Hill, New York, 1964.

Swedisch EPA, Nitrogen from land to sea. Main report., *4801*, Swedisch Environmental Protection Agency, Stockholm, 1997.

Thoman, R. V., and L. C. Linker, Contemporary issues in watershed and water quality modeling for eutrophication control, *Water Science Technology*, *37*, 93-102, 1998.

Torgersen, C. E., R. E. Gresswell, and D. S. Bateman, Pattern detection in stream networks: quantifying spatial variability in fish distribution, *GIS/Spatial Analyses in Fishery and Aquatic Sciences*, 405-420, 2004.

Turner, E. R., and N. N. Rabalais, Changes in Mississippi River water quality this century and implications for coastal food webs, *BioScience*, *41*, 140-147, 1991.

Turner, R. E., and N. N. Rabalais, Coastal eutrophication near the Mississippi river delta, *nature*, *368*, 619-621, 1994.

UN, World population prospects: 1998, United Nations Department for Economic and Social Information and Policy Analyses, New York, 1998.

USGS, USGS Node of the National Geospatial Data Clearinghouse, 2003.

Vachaud, G., and T. Chen, Sensitivity of a large-scale hydrologic model to quality of input data obtained at different scales; distributed versus stochastic non-distributed modelling, *Journal of hydrology*, *264*, 101-112, 2002.

Van Breemen, N., Nitrogen cycle: Natural organic tendency, *Nature*, *415*, 381 - 382, 2002.

Van Breemen, N. E. W., E. W. Boyer, C. L. Goodale, N. A. Jaworski, K. Paustian, S. P. Seitzinger, K. Lajtha, B. Mayer, V. D. D., R. W. Howarth, K. J. Nadelhoffer, M. Eve, and G. Billen, Where did all the nitrogen go? Fate of nitrogen inputs to large watersheds in the northeastern U.S.A, *Biogeochemistry*, *57/58*, 0 - 267, 2002.

Van Drecht, G., A. F. Bouwman, J. M. Knoop, A. H. W. Beussen, and C. R. Meinardi, Global modeling of the fate of nitrogen from point and nonpoint sources in soils, groundwater, and surface water, *Global Biochemical Cycles*, *17 no. 4*, 2003.

van Griensven, A., and W. Brauwens, Integral water quality modelling of catchments, *Water science and technology*, *43*, 321-328, 2001.

Van Herpe, Y. J. P., P. A. Troch, P. F. Quinn, and S. Anthony, Modelling Hydrological mobilization of nutrient pollutants at the catchment scale, *Agriculture, Hydrology and water quality*, 2002.

Veuger, B., J. J. Middelburg, H. T. S. Boschker, J. Nieuwenhuize, P. van Rijswijk, E. J. Rochelle-Newall, and N. Navarro, Microbial uptake of dissolved organic and inorganic nitrogen in Randers Fjord, *Estuarine, Coastal and Shelf Science*, *61*, 507-515, 2004.

Vince, S., and I. Valiela, The effects of ammonium and phosphate enrichment on chlorophyll a, pigment ratio, and species composition of phytoplankton of vineyard sound, *Marine biology*, *56*, 111-134, 1973.

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

Vitousek, P. M., K. Cassman, C. Cleveland, T. Crews, C. B. Field, N. B. Grimm, R. W. Howarth, R. Marino, L. Martinelli, E. B. Rastetter, and J. I. Sprent, Towards an ecological understanding of biological nitrogen fixation, *Biogeochemistry*, *57-58*, 1-45, 2002.

Vörösmarty, C. J., B. M. Feteke, M. Meybeck, and R. Lammers, Geomorphometric attributes of the global system of rivers at 30-minute spatial resolution, *Journal of Hydrology*, *237*, 17-39, 2000a.

Vörösmarty, C. J., B. M. Feteke, M. Meybeck, and R. Lammers, A simulated topological network representing the global system of rivers at 30-minute spatial resolution (STN-30). *Global Biogeochemical Cycles*, *14*, 599-621, 2000b.

Vörösmarty, C. J., M. Meybeck, B. Feteke, K. Sharma, P. Green, and J. P. M. Syvitski, Antropogenic sediment retention: major global impact from registered river impoundments, *Global and Planetary Change*, *39*, 169-190, 2003.

Wagenet, R. J., Scale issues in agroecological research chains, *Nutrient Cycling in Agroecosystems*, *50*, 23-34, 1998.

Ward, J. V., K. Tockner, and F. Schiemer, Biodiversity of floodplain river ecosystems: ecotones and connectivity, *Regulated Rivers: Research & Management*, 125–139, 1999.

Wolf, J., A. H. W. Beusen, P. Groenendijk, T. Kroon, R. Rötter, and H. Van Zeijts, The integrated modeling system STONE for calculating nutrient emissions from agriculture in the Netherlands, *Environmental modelling & software*, *18*, 597-617, 2003.

Wolfert, H. P., Geomorphological change and river rehabilitation, Wageningen University, Wageningen, 2001.

Wood, E. F., M. Sivapalan, K. Beven, and L. Band, Effects of spatial variability and scale with implications to hydrologic modeling, *Journal of hydrology*, *102*, 29-47, 1988.

World Commission on Dams, *Dams and development*, 404, Earthscan Publications, London, 2000.

WorldBank, The 2000 World Bank Development Indicators [CDROM], Washington, D.C., 2001.

Zumft, W. G., Cell biology and molecular basis of denitrification, in *Microbiology and Molecular Biology Reviews*, 533-616, 1997.

Glossary

Aggregate
: Coarse scale result of effects of fine scale variables. An aggregate can be estimated from values of fine scale variables by aggregation.

Aggregation
: Estimation of the value of a coarse scale variable from the values of fine scale variables affecting it. Aggregation can consist of summation or averaging of values of fine scale variables [e.g. *McNaughton*, 1994]. However often it is also needed to take into account fine scale interactions [e.g. *Jarvis*, 1995]. To take into account spatial or temporal interactions, spatial or temporal heterogeneity must be considered, respectively.

Coarse scale
: Having a large support and extent

Characteristic length
: Length of patches or clumps of a particular process

Characteristic scale
: Duration, length, surface area or stream order range of patches or clumps of a particular process

Downscaling
: Using coarse scale information to derive information at a finer scale

Extent
: Total range of time or space that is considered by, for example, data, a model or scenarios.

Fine scale
: Having a small support and extent

Grain
: Smallest scale at which an actor responds to patch structure by differentiating among patches [*Kotliar and Wiens*, 1990]

Instance
: A part of a scalable model part (SMP) modeling an area, time, or stream length for which input is homogeneous

Outlet
: Location where modeled N forms leave the modeled part of a watershed and for which N flow is predicted. The outlet can be a river reach, an estuary, or a coastal zone.

Patch

A region that differs from its surroundings but is not necessarily internally homogeneous [*Kotliar and Wiens*, 1990]

Stream order
: Measure of the scale of river reaches of part of a river network [*Strahler*, 1964]

Scale
: Combination of extent and support, and if applicable also stream order

SMP
: Scalable Model Part: A part of a model that calculates at a unique spatial support, temporal support, spatial extent, temporal extent or stream order range in existing model applications. An SMP is a set of either model equations or model predictions. Its scale can be adjusted.

Support
: Areas or time periods that are assumed to be homogeneous in measurements, data, models or predictions.

Support unit
: Specific case of an area or time period that is assumed to be homogeneous in a measurement, data set, model or prediction.

Upscaling
: Using fine scale information to derive information at a coarse scale

Summary

Human activities cause flow of nitrogen (N) from terrestrial to aquatic systems. This has many serious consequences that need to be alleviated. Understanding and anticipation of N flow to aquatic systems can be achieved by modeling. Several models have been developed but one of their major weaknesses is the use of inappropriate scales. Therefore, the objective of this thesis is to increase our ability to identify appropriate spatial and temporal scales for N-flow models in a transparent and comprehensive way. In order to meet this objective, the following sub-objectives are addressed:

I. To model of global N flows from land to water in a spatially explicit way.

II. To develop a framework for identifying the appropriate spatial and temporal scales of N-flow models.

III To apply this framework to a model of global N flow.

IV To assess the applicability of this framework to models of N flow between floodplains and rivers.

First, a spatially explicit, global model for predicting dissolved inorganic nitrogen export (DIN) by rivers to coastal waters (NEWS-DIN) is developed. NEWS-DIN models DIN export from watersheds to coastal waters as a function of N inputs to watersheds including manure N, fertilizer N, atmospheric N deposition, and biological N_2 fixation, as well as sewage. NEWS-DIN also includes retention and loss terms, including N retention in river networks, N retention in dammed reservoirs, N loss via consumptive water use, denitrification in rivers, and N loss via harvesting and grazing. For global watersheds DIN yields are calculated ranging from 0.0004 to 5217 kg N km^{-2} y^{-1} with the highest DIN yields occurring in Europe and South East Asia. The calculated global DIN export to coastal waters is 25 Tg N y^{-1}, with 16 Tg N y^{-1} from anthropogenic sources. Biological N_2 fixation is the dominant source of exported DIN. And globally, and on every continent except Africa, N fertilizer is the largest anthropogenic source of DIN export to coastal waters. NEWS-DIN is a global model, calculating annual DIN flows for 540 basins, while resolving equations at the scale of individual basins and sub-basins as well as a grid of 0.5 x 0.5°. These scales were mainly chosen for pragmatic reasons.

Next, a framework (FAMOS) is described to identify the appropriate spatial and temporal scales for N-flow models. FAMOS has been developed for models that predict N export from large watersheds. With FAMOS, modelers can identify the appropriate scale for model predictions and other Scalable Model Parts (SMPs). Different measures of model scale are distinguished in FAMOS. These are support, stream length, extent, and stream order. Spatial support is a measure of the size of the areas represented by single values of input variables. Temporal model support is a measure of the duration of the times represented by single values of input variables. Stream length is the length of river sections represented by single values of input variables. Model extent is the total range of time or space within which processes are modeled. Stream order is a measure of the size of river reaches that are modeled.

Using twelve indicators, FAMOS determines the appropriateness of model scales. Indicators are to be specified by the modeler and are associated with four criteria. The criteria require modeling scales to correspond with (A) data and scenarios, (B) model assumptions, (C) available resources for modeling, and (D) appropriately scaled predictions.

The applicability of FAMOS is assessed for NEWS-DIN. Ranges of appropriate scales are determined for model output and five SMPs, which model the (1) surface N balance, (2) point sources, (3) N flow in sediments and small streams, (4) retention in dammed reservoirs, and (5) riverine DIN retention. Indicators of appropriateness of modeling scale are quantified based on existing data and knowledge. A comparison is made between the scale at which NEWS-DIN was applied and the identified appropriate scale. Based on this, recommendations can be made for improvement of application of the model. The results indicate that most of the scales used in the original application of NEWS-DIN are appropriate. FAMOS identified that spatial support and temporal extent of some SMPs are inappropriate. For the SMPs modeling riverine retention and predicted riverine DIN export no appropriate values of temporal support were found. The applied scales for NEWS-DIN were for practical reasons chosen such that they agreed with the available data. We conclude that the modeling scale of NEWS-DIN could be improved if (1) knowledge or data sufficient to model riverine processes on a smaller temporal support is obtained, or if (2) if NEWS-DIN is used to hindcast or model scenarios of the future.

The applicability of FAMOS is also assessed for models of N-flow from rivers to aquatic systems in reconnected floodplains. Floodplain models describe smaller systems than the ones for which FAMOS was originally developed. We conclude that FAMOS can be applied to floodplain models, if three phenomena affecting appropriate scales of these models are considered. This requires an extension with three new indicators for the appropriateness of modeling scale: (1) the reliability with which empirical relations can model the aggregated effect of an intermittent process on N flow, (2) the validity of particular nonlinear model equations, and (3) the reliability of modeling past N inflows. These new indicators are applied to a hypothetical model of an existing reconnected floodplain to illustrate their use, output and effect on appropriate scales. It is shown that inclusion of the new indicators in FAMOS provides a feasible basis for a comprehensive identification of appropriate scales of N flow from rivers to aquatic systems in reconnected floodplains.

Novel aspects of our approach for identifying appropriate scales for N-flow models can be summarized as follows. First, the most appropriate scale results in an optimal balance between the use of available data and scenarios, validity of model assumptions, effort needed for modeling, and the usefulness of model output. Second, different indicators can be used to assess these criteria. Third, appropriate scales need to be identified for scalable model parts, not for complete models. Fourth, several interactions in affecting scale appropriateness can be identified between values of different measures of modeling scale. SMP scales affect the appropriateness of scales of other SMPs. And the appropriateness of a value of a measure of model scale is usually affected by the value chosen for another measure of model scale within the same SMP. Finally, the modeled system and the type of model equations affect the methods that are suitable to indicate if there is an optimal balance between agreement with available data, model assumptions, resources and end users. This thesis shows that FAMOS is a promising tool that can be applied to a wide range of models.

Samenvatting

Menselijke activiteiten veroorzaken stroming van stikstof (N) van terrestrische naar aquatische systemen. Dit heeft veel belangrijke consequenties waar rekening mee dient te worden gehouden. Inzicht en anticipatie van N-stroming naar aquatische systemen kan worden verkregen door modelering. Diverse modellen zijn daartoe ontwikkeld, maar een van de belangrijkste zwaktes van deze modellen is het gebruik van ongeschikte modelleerschaal tijdens de toepassing van deze modellen. Daarom is het doel van dit proefschrift om de bekwaamheid van modelleurs in het vinden van geschikte ruimtelijke en temporele modelleerschalen te vergroten. Om dit doel te bereiken, worden de volgende subdoelen nagestreefd:

I Modellering van N stroming van land naar water in de wereld op een ruimtelijk expliciete manier.

II Ontwikkeling van een leiddraad waarmee geschikte ruimtelijke en temporele schalen van N stromingsmodellen geïdentificeerd kunnen worden.

III Toepassing van deze leiddraad op een model van stikstofstroming in de wereld.

IV Bepaling van de toepasbaarheid van deze leidraad voor modellen van stikstofstroming tussen rivieren en uiterwaarden.

Ten eerste wordt een model ontwikkeld van stroming van opgelost anorganisch stikstof (DIN) via rivieren naar kustwateren in de wereld. Dit model (NEWS-DIN) beschrijft DIN stroming uit stroomgebieden die uitmonden in kustwateren als functie van N-bronnen in stroomgebieden zoals N in mest, kunstmest, en atmosferische depositie, en N-aanvoer door N-fixatie en riolen. NEWS-DIN beschrijft ook N-retentie en N-verliezen zoals N-verlies processen in rivieren en in stuwmeren, en N-verlies uit terrestrische systemen via oogsten en begrazing. De hoeveelheid N-stroming vanuit stroomgebieden die uitmonden in kustwateren verspreid over de wereld varieert tussen 0,0004 en 5217 kg N $km^{-2}y^{-1}$, met de hoogste DIN afvoer in Europa en Zuid Oost Azië. De gemodelleerde N-stroming naar de kustwateren van de wereld is 25 Tg N y^{-1}, waarvan 16 Tg N y^{-1} uit antropogene bronnen afkomstig is. Biologische N-fixatie is de dominante bron van DIN stroming naar kustwateren. Gemiddeld over de wereld, en op elk continent behalve Afrika is N-kunstmest de belangrijkste menselijke bron van DIN stroming naar kustwateren. NEWS-DIN is een model voor DIN stroming uit 540 stroomgebieden die uitmonden in kustwateren over de hele wereld. NEWS-DIN bestaat uit vergelijkingen die rekenen per basin, sub-basin en per gridcel van 0.5 x 0.5°. Deze modelleerschalen zijn vooral op basis van pragmatische overwegingen gekozen.

Vervolgens is een leiddraad (FAMOS) geschreven voor het identificeren van de geschikte ruimtelijke en temporele schalen voor N-stromingsmodellen. FAMOS is ontwikkeld voor modellen die N-stroming uit grote stroomgebieden berekenen. Met behulp van FAMOS kunnen modelleurs de geschikte schaal identificeren voor modelvoorspellingen en andere modelcomponenten waarvan de schaal veranderd kan worden (SMPs). Verschillende maten voor modelleerschaal worden onderscheiden in

FAMOS. Deze zijn support, stream length, extent en stream order. Ruimtelijke support is een maat voor de omvang van de gebieden waarvoor individuele waarden van inputvariabelen representatief zijn. Temporele support is een maat voor de tijdsduur waarvoor individuele waarden van inputvariabelen representatief zijn. Stream length is de lengte van rivier secties die door individuele waarden van input variabelen gerepresenteerd worden. Extent is de tijdsduur of gebiedsomvang waarin processen worden gemodelleerd. Stream order is een maat voor de grootte van de riviertakken die worden gemodelleerd. Met behulp van twaalf indicatoren bepaalt FAMOS de geschiktheid van modelleerschalen. Indicatoren moeten worden gespecificeerd door de modelleur en toetsen vier criteria. Volgens deze criteria moeten modelleerschalen corresponderen met (A) data en scenario's, (B) aannames in het model, (C) beschikbare middelen voor modelering en (D) de schaal waarop modelvoorspellingen bruikbaar zijn.

FAMOS is toegepast op NEWS-DIN. Geschikte modelleerschalen zijn bepaald voor model output en vijf SMPs die de volgende systemen modeleren: (1) oppervlakte balans van N, (2) puntbronnen, (3) N-stroming in sedimenten en kleine riviertakken, (4) DIN retentie in stuwmeren en (5) DIN retentie in grote riviertakken. Indicatoren voor de geschiktheid van modelleerschaal worden gekwantificeerd op basis van bestaande data en kennis. Een vergelijking wordt gemaakt tussen de schaal waarop NEWS-DIN was toegepast en de geïdentificeerde geschikte schaal. Op basis hiervan kunnen aanbevelingen gedaan worden voor verbetering van de toepassing van dit model. De resultaten laten zien dat de meeste schalen van de originele toepassing van NEWS-DIN geschikt waren. Met FAMOS is bepaald dat de ruimtelijke support en temporele extent van enkele SMPs tijdens originele toepassing van NEWS-DIN ongeschikt was. Voor de SMPs die DIN retentie in grote riviertakken en DIN export naar kustwateren beschrijven, zijn geen geschikte waarden van temporele support gevonden. De modelleerschalen tijdens de originele toepassing van NEWS-DIN waren om praktische redenen zo gekozen dat ze pasten bij de beschikbare data. We concluderen dat de modelleerschaal van NEWS-DIN verbeterd zou kunnen worden als (1) ervoor wordt gezorgd dat kennis en data toereikend zijn om aquatische processen in rivieren op een kleinere temporele support te modelleren, of als (2) NEWS-DIN wordt gebruikt om voorspellingen te doen die gelden binnen een langere tijdsduur.

De toepasbaarheid van FAMOS is bepaald voor modellen van N-stroming van rivieren naar aquatische systemen in gerenaturaliseerde uiterwaarden. Zulke modellen beschrijven systemen die kleiner zijn als de systemen waarop FAMOS geijkt is. We concluderen dat FAMOS toegepast kan worden op zulke uiterwaardmodellen, wanneer rekening gehouden wordt met drie fenomenen die de geschiktheid van modelleerschaal beïnvloeden. Hiertoe moet FAMOS uitgebreid worden met drie nieuwe indicatoren voor geschiktheid van modelleerschaal: (1) betrouwbaarheid waarmee empirische relaties het geaggregeerde effect van een discontinu proces op N-stroming kunnen modelleren, (2) de geldigheid van bepaalde niet-lineaire modelvergelijkingen, en (3) de betrouwbaarheid waarmee N-instroom gedurende het verleden gemodelleerd kan worden. Deze nieuwe indicatoren zijn ter illustratie toegepast op een hypothetisch model van een bestaande gerenaturaliseerde uiterwaard. Uit deze toepassing blijkt dat het uitbreiden van FAMOS met de nieuwe indicatoren zorgt voor een haalbare basis voor een alomvattende identificatie van geschikte schaal voor modellen van N-stroming van rivieren naar aquatische systemen in gerenaturaliseerde uiterwaarden.

Nieuwe aspecten van de gepresenteerde methode voor identificatie van geschikte modelleerschaal kunnen als volgt worden samengevat. Ten eerste resulteert de geïdentificeerde geschikte schaal in een optimale balans tussen het gebruik van beschikbare data en scenario's, geldigheid van modelaannames, aanspraak op beschikbare middelen, en de bruikbaarheid van modelvoorspellingen. Ten tweede kunnen verschillende indicatoren gebruikt worden om de deze criteria te toetsen. Ten derde worden geschikte schalen geïdentificeerd voor afzonderlijke modelcomponenten met aanpasbare modelleerschaal (SMPs) en dus niet voor modellen als geheel. Ten vierde wordt rekening gehouden met een aantal interacties tussen schalen van model componenten en geschiktheid van bepaalde andere modelleerschalen in hetzelfde model. SMP schalen beïnvloeden de geschiktheid van schalen van andere SMPs. En de geschiktheid van een bepaalde waarde van en maat voor modelleerschaal hangt meestal af van de waarde die gekozen wordt voor een andere maat voor modelleerschaal binnen dezelfde SMP. Tenslotte beïnvloeden het gemodelleerde systeem en het type modelvergelijkingen de methoden die geschikt zijn om te indiceren wat de geschikte modelleerschaal is. Dit proefschrift laat zien dat FAMOS een veelbelovende methode is die toegepast kan worden op een groot aantal modellen en modeltypen.

Acknowledgements

This thesis contains the results of PhD research that was funded by the Netherlands Organization for Scientific Research (NWO) as part of the Aspasia program.

Firstly, I would like to thank my promotors and co-promotors; Carolien Kroeze, Alfred Stein, Rik Leemans, Lex Bouwman, and Evert Jan Bakker for giving me the freedom and support to develop my approach, the very pleasant cooperation and our regular supervision meetings. We had good discussions while exploring the best approach to answer my research questions in the start of my research. Later, they were of great value by reviewing the large amounts of texts that I wrote for this thesis. I especially would like to thank Carolien Kroeze for her advice on all aspects of the research, the way she motivated both myself and other people involved in my supervision, the enormous amount of time that she spent on my supervision and her always timely and motivating responses to all my questions. I would like to thank Lex Bouwman for his rigorous checks of almost all my texts, which have much increased their understandability and removed much unnecessary complexity. Also, I would like to thank Lex for providing most of the data that was necessary for development and application of the NEWS-DIN model and for writing Chapter 2. I would like to thank Evert Jan Bakker for checking all the formulas on correctness and for very useful discussions on the method and concept if the indicators (Chapters 3 and 5). I would like to thank Rik Leemans for substantially increasing the readability of a number of texts and for his positive effect on the working atmosphere. I would like to thank Alfred Stein for his timely responses and supervision regarding many aspects of the research and for his rigorous and critical review of draft chapters of this thesis.

Writing parts of this thesis was not possible without the support from several other people. I would like to thank some people that contributed to Chapter 2, especially John Harrison (Washington State University) for his hospitality during my visit to Rutgers University, New Jersey (U.S.) and his large role in the development of NEWS-DIN. I would like to thank the Global NEWS project and its initiator and coordinator, Sybil Seitzinger, for accepting me as a contributing scientist, for introducing me to a very new and exciting field of science, and for giving me the opportunity to learn from a group of very experienced scientists.

The Young Scientists Summer Program (YSSP) at the International Institute for Applied Systems Analyses (IIASA) proved to be very useful for writing Chapter 5. Participation in this program was not possible without the financial support from the Netherlands Organization for Scientific Research (NWO), and the fantastic guidance and organization of the YSSP by Matthias Jonas, Tanja Huber, and Barbara Hauser. I would like to thank Jan Sendzimir for supervising me during the YSSP. Discussions with him were the basis for Chapter 5. He also introduced me to a number of people who later proved to be indispensable for Chapter 5. These are Thomas Hein, Stefan Preiner, Christian Baumgartner, and Iwona Cisowska. I thank Thomas Hein and Stefan Preiner for providing the data necessary for Chapter 5 and giving useful feedback on the drafts of this chapter. I thank Christian Baumgartner for the large amount of information that he provided on the anthropogenic effects on functioning of floodplains. I would like to thank Iwona Cisowska for useful discussion on the effects of floodplain reconnection on N cycling, and for being a good friend.

I thank the colleagues of the environmental Systems Analyses Group for creating a very pleasant working atmosphere. I especially thank Serge Stalpers for being a very pleasant colleague, roommate and friend, and for useful discussions about all topics that mattered. I thank Ria Cuperus for being the social backbone of the Environmental Systems Analyses Group and for her never failing friendliness.

Finally, I would like to acknowledge all my friends, family and other social relations. Especially, I would like to thank Huub and Marie Louise for their unconditional and never failing support, attention, and hospitality.

Curriculum Vitae

Egon Dumont was born on September 16th, 1977 in Heerlen, The Netherlands. He completed his secondary school education (VWO) at Sophianum in Gulpen in 1995. In the same year he started his study biology at the Radboud University Nijmegen with specialization in aquatic ecology, and graduated in 2000. During his study he spent five months at the Aquaculture Development Centre of the University College Cork, Ireland. In 2001 he worked as an ecological researcher at a consultancy firm (Natuurbalans) in Nijmegen. Later that year, he started his study geo-information science at Wageningen University with specialization in groundwater flow modeling, and graduated in 2003. In that year he also started his PhD research on "Identifying the appropriate scales to model nitrogen flows from land to water" at the Environmental Systems Analyses Group of Wageningen University. The research was funded by the Netherlands Organization for Scientific Research (NWO) as part of the Aspasia program. In 2006, Egon was selected for participation in the Young Scientists Summer Program, which is a three month fellowship at the International Institute for Applied Systems Analyses (IIASA) in Austria. In 2007, the PhD research at Wageningen University resulted in this thesis.

The SENSE Research School declares that Mr. Egon Dumont has successfully fulfilled all requirements of the Educational PhD Programme of SENSE with a work load of 30 ECTS, including the following activities:

SENSE PhD courses:

- Environmental Research in Context
- Research Context Activity: "Organization of two international workshops on the application of a GIS-based model of nutrient export from land to large marine ecosystems."
- Uncertainty Modelling and Analyses
- Interactions between Element Cycles and Ecosystems
- Basic Statistics and Advanced Statistics

Other PhD courses:

- Techniques for Writing and Presenting Scientific Papers
- Career Perspectives
- Project and Time Management
- Interaction in Communication

Activities:

- Courses and research on Complex Environmental Issues, June – August 2006, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

Oral Presentations:

- Global NEWS workshop, 24 May 2004, Paris, France
- SENSE Summer Symposium, 17 June 2004, Utrecht, The Netherlands
- GEF LME Nutrient Export Modeling Workshop, 23 January 2006, Paris, France
- Opportunities in Global Change (conference), 7 February 2006, Amsterdam, The Netherlands
- Global NEWS workshop, 11 May 2006, Paris, France
- 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 6 July 2006, Lisbon, Portugal
- IIASA Late summer workshop, 24 August 2006, Vienna, Austria
- GEF LME Nutrient Export Modeling Workshop, 18 September 2006, Paris, France
- International Young Scientists Global Change Conference, 6 November 2006, Beijing, China

Deputy director SENSE
Dr. A. van Dommelen