A hydrological programming language extension for integrated catchment models

Kumulativdissertation zur Erlangung des
akademischen Grades „Dr. rer. nat."
am Fachbereich 09, **Agrarwissenschaften, Ökotrophologie und Umweltmanagement**

der

Justus-Liebig-Universität Gießen

vorgelegt von

Philipp Kraft
(Dipl.-Geoökol.)
aus Holzminden

Gießen, den 16. Januar 2012

Zusammenfassung

Im letzten Jahrzehnt ist die Anwendung klassischer hydrologischer Modelle als Mittel zum Prozessverständnis des Wassertransports in Landschaften in die Diskussion geraten, da diese Modelle nicht gut zur Formulierung und Zurückweisung von Hypothesen geeignet sind (z.B. Beven 2002 und 2006, Seibert und McDonnell 2002, Sivapalan et al. 2003, Kirchner 2006, Tetzlaff et al. 2008). Vaché und McDonnell (2006) entwickelten ein System, mit dem Modellstrukturen falsifiziert werden können, wobei jede Modellstruktur als Hypothese der Abflussbildung verstanden wird. Die vorliegende Doktorarbeit zeigt eine generalisierte Neuentwickelung des Vaché-McDonnell Systems, das „Catchment Modelling Framework“ (Einzugsgebiets-Modell-System, CMF). Es ist als Erweiterung der Programmiersprache Python implementiert und bietet einen Baukasten für die Entwicklung einer Vielzahl verschiedener hydrologischer Modelle, basierend auf dem finiten Volumen Ansatz von Qu und Duffy (2007).

Auch Buytaert et al. (2008) und Clark et al. (2011) fordern solche modularen Systeme als Mittel für die Formulierung, Programmierung, Prüfung und Falsifizierung von Prozesshypothesen. Buytaert et al. (2008) stellen fest, dass solche Systeme portabel, verfügbar und modular sein müssen. Gleichzeitig wächst, während Hydrologen die theoretischen Anwendungsgrenzen von Abflussmodellen für Stofftransport in Landschaften diskutieren, der Bedarf an integrierten Landschaftsmodellen für die Integration von lateralem Transport in interdisziplinären Projekten, wie z.B. dem NitroEurope IP (EU). Vor diesem Hintergrund sollten modulare hydrologische Systeme so konzeptioniert sein, dass Datenaustausch während der Modellaufzeit mit Modellen aus anderen Disziplinen möglich ist.

CMF ist portabel, verfügbar und modular, da es eine quelloffene Erweiterung der Sprache Python (s. Kapitel I und II) ist, und kann für die Formulierung verschiedener Hypothesen genutzt werden (s. Kapitel III). Das Prinzip der Verbindung von CMF mit Modellen aus anderen Disziplinen wird in Kapitel IV gezeigt; Kapitel V zeigt das Potential solcher eng gekoppelter Transport- und Umsatzmodelle für die Berechnung der Emission von Treibhausgasen aus Ökosystemen.

Short summary

In the last decade, the application of classical hydrological rainfall runoff models is disputed as a valid method for understanding water transport in landscapes (eg. Beven 2002 and 2006, Seibert and McDonnell 2002, Sivapalan et al. 2003, Kirchner 2006, Tetzlaff et al. 2008), since current models lack the ability for formulation and rejection of hypotheses. Vache and McDonnell (2006) developed a framework for the rejection of model structures, where each model structure is understood as a hypothesis on runoff generation. This thesis presents a newly developed generalized form of the Vache-McDonnell rejectionist framework, called the Catchment Modelling Framework (CMF). It is an extension to the programming language Python, offering a toolkit for the set up of a wide range of hydrological models, following the finite volume approach by Qu and Duffy (2007).

Buytaert et al. (2008) and Clark et al. (2011) call also for such modular frameworks as a tool for the formulation, implementation, test and rejection of process hypotheses. Buytaert et al. (2008) demand such frameworks to be portable, accessible and modular. While hydrologists debate the theoretical application limits of runoff models for solute transport in landscapes, a growing demand of integrated landscape models for the integration of lateral transport of matter by runoff arises in interdisciplinary projects, like eg. the NitroEurope IP (EC). With this background, modular hydrological frameworks need to be designed for simplified data exchange during the model runtime for interfacing the hydrology with models from different disciplines, like CMF.

CMF is portable, accessible and modular as an open source extension to the Python language (see Chapter I and II) and can be used for the formulation of different hypotheses (Chapter III). The principle of the connection of CMF with models from different disciplines is shown in Chapter IV, and Chapter V shows the relevance of tightly connected models of transport and turnover for the emission of greenhouse gases from ecosystems.

Table of content

I. Extended Summary

In the last 40 years, since the publication of a blueprint for physically based deterministic hydrological models by Freeze and Harlan (1969), a multitude of more or less physically-based, deterministic models have been developed, for various scopes, scales and regions (cf. Singh and Frevert, 2002). If a model describing the pathways of water as a transporting media is incorporated in an integrated landscape model, a decision has to be made, which model will be used. Due to the effort of integrating the model in the broader scope, the decision might not be changeable for later applications.

Hydrological Models

A common characteristic of an integrated catchment tool is that the hydrological model serves as a lateral and vertical distributor of solutes through surface and subsurface water flows (Band et al., 2001). Hence, the selection or set up of a suitable hydrological model is an important step in the process of setting up an integrated catchment model. In principle a three dimensional variable saturated media distributed transport model seems to be most appropriate. Although such models exist, a primary example being HYDRUS 3D (Simunek et al., 2008) calculating the complete three dimensional transport in mesoscale catchments, it is not feasible due to lack of computational power, and more importantly, an inability to adequately characterize the mesoscale boundary conditions (Beven, 2002; Kirchner, 2006; Tetzlaff et al., 2008). Various techniques for simplifying the water fluxes in a landscape have been developed for hydrological models. The simplifications primarily focus on linearization and the identification of controlling processes or features.

One dimensional plot scale transport models are often utilized for situations involving transport in the unsaturated zone. While useful for a wide variety of situations, this group of models, including DRAINMOD (Gayle and Skaggs, 1978), BROOK90(Federer, 1995), HYDRUS 1D (Simunek et al., 2005) or MACRO (Jarvis et al., 1997), cannot play the role of a lateral distributor in an integrated landscape approach because of the potential importance of lateral interactions in spatially distributed

models. A second stage of complexity, designed to include a lateral dimension, is the introduction of empirical or conceptual retention functions between the outflows from a single soil column representing a partition of the study area to the drainage network. This approach is chosen by the class of lumped and semi-distributed models like HBV (Bergström, 1995), LASCAM (Sivapalan, et al., 2003), TOPMODEL (Beven and Kirkby, 1979) and SWAT (Arnold et al., 1998).

SHE (Abbott et al., 1986) introduced a connection between vertical and lateral transport through the inclusion of a series of two dimensional models. The former is a groundwater flow model defining exchange at the lower boundary of the soil models. The latter focuses on surface flows through the application of a two dimensional overland flow model which effectively connects a set of one dimensional column models laterally. MIKE-SHE (Refsgaard and Storm, 1995) extends the saturated transport of SHE by using a three dimensional model based on the groundwater model MODFLOW (Harbaugh et al., 2000). DHSVM (Wigmosta et al., 1994) and PIHM (Qu and Duffy, 2007) separate the vertical transport in the vadose zone and lateral transport in the saturated zone.

Another concept is the connection of the soil columns over their full depth, thus creating a variable saturated flow model for a two dimensional setup as in HYDRUS 2D (Simunek et al., 1999), IHDM (Beven et al., 1987), CATFLOW (Maurer, 1997; Zehe et al., 2001) and HILLFLOW (Bronstert, 1999). With the exception of HYDRUS 2D, the modelled hillslope slices are connected, thus a quasi three dimensional representation of the dominant fluxes is created. However, these models are not designed to model laterally divergent fluxes. Bronstert (1999) discussed the approach in detail.

An additional alternative to simplify three dimensional variable saturated flow systems is the application of kinematic assumptions to ease the underlying continuum equations. While kinematic assumptions are used throughout many of the models mentioned previously, they are explicitly invoked in a number of well known simulations, including TOPKAPI (Ciarapica and Todini, 2002) and TACD (Uhlenbrook et al., 2004). In this case, the flux leaving a computational element is a function of the state of the element only and not a function of a gradient.

Each simplification scheme is based upon a hypothesis involving the identification of both major and minor flow paths. The controlling flow paths are then explicitly accounted for, with the minor flow paths remaining untreated. In the SHE model, as one example, very shallow interflow on a perched water table cannot be captured. Divergent groundwater flow, on the other hand, is not possible to simulate in models comprised of coupled two dimensional hillslopes, including IHDM or CATFLOW. The challenge, then, for model users is to identify the key assumptions that are built into the models they use, but more importantly to use that information to select a model that is appropriate for the problem at hand. Two pieces of information are needed to make such a decision. The first is a perceptual model of local processes to guide model selection. The second is a set of transparent definitions of different computer models, specifically including key assumptions and key processes, and therefore facilitating an informed decision regarding the most appropriate structure. In practice, the latter is, for a variety of reasons, difficult to achieve. An alternative procedure that does not rely upon the detailed understanding of a large number of model options involves applying a set of candidate models, using the simulated results, rather than a priori understanding, to guide the model selection, application, and evaluation process as mentioned by Beven (2002).

Modular frameworks

In response to this challenge, Buytaert et al. (2008) suggest that model codes should be fully accessible, modular and portable. The incorporation of these qualities into the code is a mechanism to allow the model developer to aid model users in testing different hypotheses about flow regimes, and, ultimately, in applying the most appropriate model to different questions of interest. To the concepts of accessibility, modularity, and portability (Buytaert et al., 2008), we add the concept of connectability, i.e. that model codes also should focus on between-model connections, particularly in the case of water quality models.

The best known example of a hydrological model implementing a range of process descriptions available as building blocks, combined with an open interface for coupling with other models, is MIKE-SHE (Refsgaard and Storm, 1995). The user of the model can choose between a finite

difference application of the Richards equation and different tipping bucket approaches in the unsaturated zone, and tune the model complexity of the lateral saturated transport ranging from 3D finite difference schemes solving Darcy's law to simplified linear storage approaches for large scale applications. Despite this flexibility, the flow paths available for incorporation remain fixed to the SHE type, and cannot be extended by users because of the commercial nature of the product, and the closed nature of the source code. The ability to link MIKE-SHE to other models is accomplished by implementing the OpenMI interface. MIKE-SHE fulfills the request for connectibility and modularity, but fails at the requests of accessibility and portability. Examples of open, accessible hydrological models with a built in flexibility to be adapted to a range of process descriptions are the rejectionist framework approach presented by Vaché and McDonnell (2006) or the Penn State Integrated Hydrological Model System (PIHM) (Qu and Duffy, 2007).

Although these models can be run in different modes and a designed for extension, it is not possible to test each single underlying hypothesis of the model, as Clark et al. (2011) demand for modular modelling frameworks. Flexible frameworks for lumped models like FLEX (Fenicia et al., 2006) and FUSE (Clark et al., 2008) show a general trend towards a generalization of the underlying mathematical formulation of the runoff generating processes. This work proposes an even more generalized concept: the finite volume approach (Qu and Duffy, 2007) is applied to a general form of water storage as the structural base for building a wide range of hydrological models. Water is routed between storages and the model domain boundaries using exchangeable mathematical descriptions of water fluxes, ranging from simple linear storage-flux relationships to the Richards equation.

Modelling structure

Common concept for a wide range of water transport model types

Any transport problem is largely governed by mass conservation. According to Beven (2006), the closure of the mass balance in a hydrologic model should be the starting point of all model applications, even though the closure cannot be routinely verified by field experiments.

Mass conservative problems in space, including water and solute transport in the landscape, can be expressed as a generalized hyperbolic problem:

$$\frac{\delta \mathbf{u}}{\delta t} + \nabla f(\mathbf{u}, t) = 0 \tag{1}$$

with the state continuum **u**, and the fluxes $f(\mathbf{u},t)$. Using the finite volume method for discretizing **u** space, the partial differential equation (1) turns into a system of ordinary differential equations (ODEs) in (2), where the fluxes in and out of the finite volume i are integrated over the shell of the volume S and normalized by the size of the finite volume v_i.

$$\frac{du_i}{dt} = -\frac{\oint_{S_i} f(\mathbf{u}, t)\, dS}{v_i} \tag{2}$$

If the shell is constructed by a finite number of planes, the line integral over the shell is transformed to the sum of fluxes connecting adjacent volumes.

When (2) is applied to a water and solute problem, a wide range of model approaches can be implemented. The model approaches differ by the method used to discretize the problem domain, the flux functions (e.g. Richard's equation or linear storage function) and by the complexity of connections between the finite volumes. The CMF approach covers the flexibility of the description of volumes and fluxes by providing clearly defined and relatively independent interfaces to implement flux functions and discretization methods. A range from highly detailed one, two or three dimensional plot scale water and solute transport models to regional applications using simplified flux descriptions and discretization / connection concepts is possible.

Hydrologic programming language

Such system is not trivial to set up. Due to the general form of system description, the relationship between water storages, implicitly defined in single approach models needs to be defined, along with the parameter sets. Generally, two possibilities to setup different model options exist: graphic user interfaces (GUIs) and user provided text files. GUIs are not

only expensive to build, both in terms of money and time, but also have a tradeoff between usability and flexibility. The use of a text-based execution strategy can provide a degree of flexibility, but is limited in terms of usability. Further, it is not a strategy that is typically employed given the evolution of current programming tools, GUI development, and the availability of higher level scripting tools. This last option, the use of available scripting languages, is not well represented in available models of environmental process, even if Ousterhout (1998) called for this use of scripting languages in science over a decade ago. The modularity of the approach of extending a programming language is already discussed by Buytaert et al. (2008).

The advantages of a "hydrologically enriched" scripting language compared to setup text files includes (1) a clear and approved syntax, (2) a standard library for data access and data transformations, (3) the capacity to develop tutorials and (4) the idioms of a Turing complete programming language, including loops and conditional executions.

We chose Python as the target language, due to its common use for science applications, its third party libraries for visualization and data access and the simplicity of writing extension code for it. Python as an structural tool was adopted in the field of hydrology only recently (Karssenberg et al., 2007; Schmitz et al., 2009; Bakker and Kelson, 2009).

The core components of CMF have been developed in C++, providing object-oriented, high-level code as well as high performance. The Python interface for the C++ code was generated using SWIG (simple wrapper interface generator, www.swig.org), a scripting language interface generator for wrapping C and C++ programs. This interface generator was originally designed to allow the use of scripting languages for steering large scale scientific simulation models (Beazley and Lomdahl, 1996).

The core model components provide definitions of the model objects, like the water storages and fluxes and the numerical solution procedures, but they do not provide the run time loop. This has to be provided by the calling, user written script, allowing an output of the model tailored to the need of the user. In this run time loop, other models that implement the scripting interface can be run for the same time step and exchange

data with CMF, as a simple and flexible integrated model approach. The creation of such a wrapper for other existing models is relatively simple, as long as the code is designed following a coding paradigm of separating processes, as object oriented or functional programming schemes. The wrapper generator allows the extension of different target languages, not only Python, but e.g. Java and C#, too. An interface to the OpenMI (Gregersen et al., 2007) coupling platform is also entirely feasible.

The model implementation, both the C++ model core and the Python interface, follow an object-oriented design, using specialization / abstraction hierarchies to facilitate the communication of different sub models through common interfaces. The approach to obtain a common handling for a wide range of transport model types is based on the finite volume approach as discussed in detail by Qu and Duffy (2007). Though it incorporated somewhat less strictly, it is introduced in the following.

The abstraction hierarchy

One of the keys for an open extensible framework is the abstraction of common behavior of different objects in a model. In CMF, different layers of abstraction exist to facilitate the access to processes and data by a single interface. By definition, the abstract nature of the interface results in a situation where the type of process and how it is described is of no consequence to the simulation framework. It is of course fundamentally important to the user. Two primary abstraction hierarchies have been implemented as part of the core CMF structure: storages and fluxes.

Storages, at the most abstract layer, are defined as state variables and expose their state and a function to calculate their derivatives at a given time. The state of an individual water storage is the volume of water and the change rate is the sum of the fluxes in and out of the volume. The water budget of a general water storage, derived from (2) is given by (3).

$$\frac{dV_i}{dt} = \sum_{j=1}^{N_i} \left(-q_{i,j}\left(V_i, V_j, t\right) A_{i,j} \right) \tag{3}$$

V is the volume of stored water in the control volume, i the current control volume, N the number of connected storages to i, $q_{i,j}$ the flux from i to j, and $A_{i,j}$ is the cross sectional area of the flux.

The state of a solute storage is the amount of tracer particles in the finite volume. The change rate of the state is the sum of advective tracer fluxes and the sum of any existent source or sink flux.

$$\frac{dX_i}{dt} = \sum_{j=1}^{N_i} \left(-q_{i,j}\left(V_i, V_j, t\right)[X]_s \right) + q_{X,in}(t) - r^- X$$
$$[X]_s = \begin{cases} [X]_i & \text{if } q_{i,j} > 0 \\ [X]_j & \text{if } q_{i,j} \leq 0 \end{cases} \tag{4}$$

X_i is the amount (mols or mass) of a tracer in control volume *i*, *[X]* is the concentration of the tracer in amount per volume water and $q_{X,in}(t)$ is a source or sink flux and *r* is a decay rate. In the future, an implementation of the advection dispersion equation is planned.

A water store maintains a reference to each tracer. Figure I-1 shows the storage concept hierarchy in CMF. If needed, state variables for momentum or energy conservation models can be introduced.

To facilitate the calculation of fluxes, specializations of water storages are implemented. Open water storages provide a calculation of their fill height as a function of state, and soil water storages have additional values to characterize the position in the soil column and the water retention curve. The retention curve is an object with methods calculating conductivity and matrix potential from a given water saturation of a soil water storage.

Fluxes of water in CMF are modeled as a flow network, where the fluxes occur at the edges of the network. The nodes of the network are the water storages and boundary conditions, while the edges are represented by flux connection objects (Figure I-2). The fluxes are calculated using specializations of the flux connections, representing the mathematical models of the hydrologic behaviour of storage interaction.

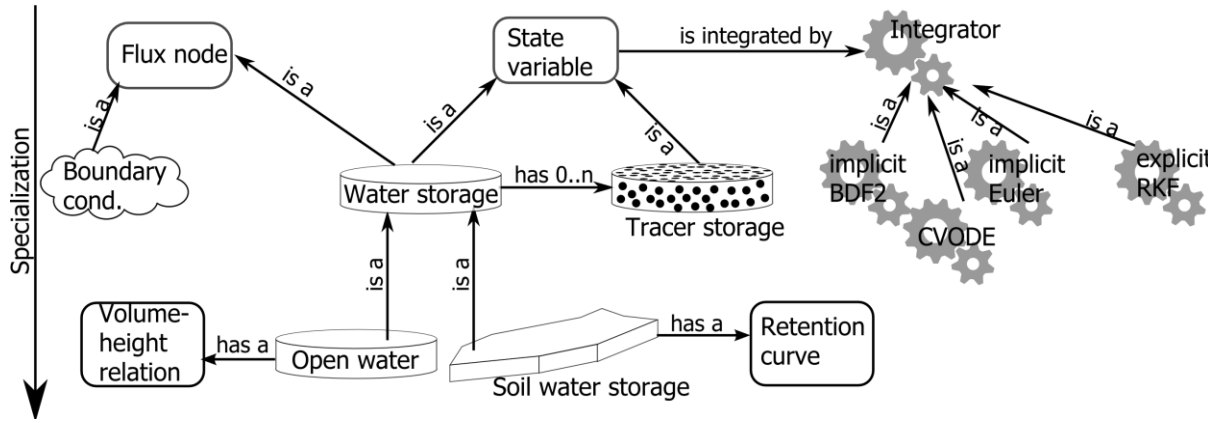

Figure I-1: Abstraction hierarchy of storages

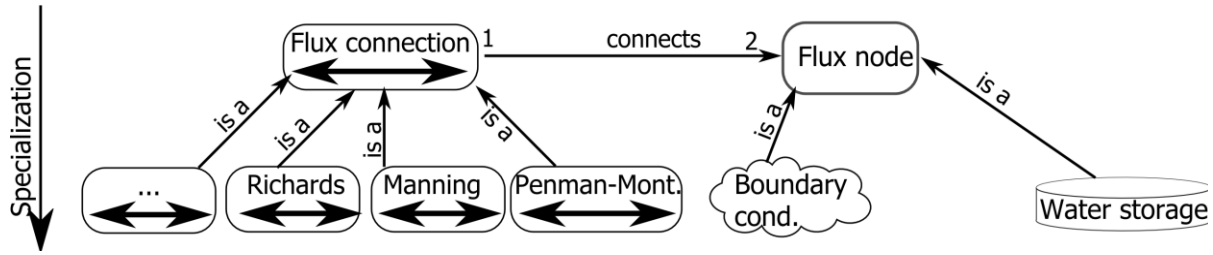

Figure I-2: Abstraction of fluxes

Each flux node maintains a method to calculate its water balance, as the sum of the in- and outgoing fluxes. For water storages, the water balance equals the derivative of the stored water volume, while the water balance of other flux node types can be queried to calculate the boundary fluxes between distinct subsystems and to calculate the area water budget.

Each specialization of a flux connection is the implementation of a flux generating process. In the current version of CMF, mostly well known and tested processes are implemented. Since the implementation of a flux generating process is relatively simple, and because a set of well-established numerical procedures for the solution of systems of differential are integrated into the core model components, only few lines of code are required to implement additional processes. The Richards equation, for example, needs only 20 lines of code, including the handling of special cases and all overhead associated with the CMF infrastructure. Thus the extension of CMF by new process descriptions, like handling macro pores, is relatively simple assuming one has a clear mathematical description of the process of interest.

Mathematical solving system

The state variables provide methods for calculating their derivatives, thus any collection of state variables forms a system of ODEs to be integrated by any well known solver for initial value problems. Different kinds of solvers are available (Table I-1).

Table I-1: Available integrators for initial value problems

Integrator	Implicit	Iteration	Order	Time step	Reference
Explicit Euler	No	N/A	1	No	(Kraft, Vache, et al., 2011)
RKF 4/5	No	N/A	4/5	error controlled	(Press et al., 1992)
Implicit Euler	Yes	Functional	1	# of iterations	(Kraft, Vache, et al., 2011)
BDF2	Yes	Functional	2	# of iterations	(Roussel and Roussel, 2003; Eckert et al., 2004)
Gears	Yes	Functional	1-4	No	(Roussel and Roussel, 2003)
CVODE	Yes	Newton-Krylov	1-5	error controlled	(Hindmarsh et al., 2005)

All solvers, except the CVODE solver, are implemented using OpenMP for parallel computation of the derivatives, while the original CVODE code is integrated into CMF as a static library. The CVODE library is also parallelized using OpenMP.

The selection of the most appropriate solver for the ODEs making up any one model is not trivial. While explicit methods, like the explicit Euler scheme and the Runge-Kutta-Fehlberg (RKF) method are computationally inexpensive per time step, many possible setups of CMF are too unstable or stiff, to be solved by an explicit method. The stiffer the system, the more complex integration method needs to be used. Stiffness of systems is indicated by the appearance of processes with highly different time scales (Tocci et al., 1997). This is especially the case if the Richards equation is to be solved and surface runoff is allowed to reinfiltrate the soil. For such systems only the CVODE solver with an appropriate preconditioner for the Newton-Krylov iteration is able to solve the system. While CMF cannot provide significant guidance in terms of the most appropriate solution procedure, having the choice between different methods, simplifies calculation time optimization of specific models.

Again, we make the argument that the most appropriate model for any unique situation may best be arrived through some degree of experimentation. The solution procedure is part of this process, and CMF is designed to explicitly recognize this fact.

Building a landscape model

Having the model building blocks, water storages and flux connections defined in a conceptual sense, the next step is to build the model for the required scale of application (point, hillslope, catchment). As catchments are the most complex scales in this sense, we show the general approach in building a catchment scale application in the following. The catchment scale model is created in three steps:

1. Discretizing the catchment horizontally into cells
2. Discretizing each cell vertically into finite volumes
3. Connecting the finite volumes with flux connections

The objects of the catchment model are stored in a container object, referred to as the project.

Horizontal discretization

The first step is the horizontal discretization of the terrain into patches, or cells. Theoretically, the shape and the size of the cells is not prescribed. In practice, size and shape constraints exist and have relationships to stability of the numerical methods and solution time steps. These constraints differ between the chosen types of lateral flux models. Typical discretization schemes involve squared raster cells, triangular networks, Voronoi polygons or definitions that follow landscape boundaries, including land cover, soils, or field definitions. For semi-distributed and lumped model approaches, landscape units derived from the hydrologic response unit concept, as in SWAT, or topographic index classes, as in TOPMODEL, can be implemented. System geometry is setup with a file containing the boundaries and the height of each cell, and derived with standard GIS software. As a second step, topological information about the vicinity of other cells and the width of the shared boundaries can be included directly, or can be generated by CMF. Either

method can be used to develop system topology, and is necessary to establish the connection of water storages in the last step. Since each submodel is in principle compatible with all others, the spatial resolution and process description depth may vary in the study area. If, for example, the riparian zone needs a higher detailed description than dry uphill regions, the framework can be set up accordingly. A hierarchy of differently detailed landscape objects, like hillslopes, patches, and micro-catchments, is created, which is comparable to the concept of spatial object hierarchies by Band et al. (2001).

The horizontal discretization is accomplished with user provided code. The general form of this code involves the explicit creation of cells and their topology, and is achieved through applications of methods defined in the core model and accessed through Python. For real applications, this approach is useful for simple numerical hillslope experiments, with provided position along the slope. For complete catchments, an automatic meshing either using the squared cells of a raster based digital elevation model (DEM) or using geometrical data of irregular cell shapes. Both automatic creation methods create cells and topological information based upon calculated topological relationships derived from the original data sources.

Vertical discretization

Each soil column is vertically divided into a series of layers. The thickness of layers may vary; although a set of simple computational constraints exist: too much variability in the layer thickness can have implications for numerical stability, and too many layers may result in unacceptably long simulation times. In addition to soil water storage, cells also reference surface water, snowpack, and vegetation storage. In all cases, helper methods exist to utilize mapped landscape data in the process of model development and the definition of appropriate storages.

Connecting storages

The water storages set up in the previous steps are then spatially related through the specification of flux connections. By creating a new flux connection object, the flux model is explicitly invoked by the simulation. The number and type of parameters needed for the connection differs

between the types. Although this explicit definition of a flux connection maintains a large degree of flexibility, it is often the case that a modeller will develop connections based upon exiting data sources. CMF implements a series of helper functions that facilitate the creation of landscape connections through the evaluation of topological relations defined in the horizontal cell mesh.

Model driver variables

The model is driven by the boundary conditions. The most important of these is, in most cases, incoming precipitation. Data from measurement stations are distributed to the cells through spatial relationships between model cells and the stations. The connection between a certain cell and a station is setup through the implementation of a standard set of rules, including nearest neighbour, similar elevation, or a combination of both. Other meteorological data is required for the calculation of evapotranspiration (ET). The meteorology object is a data provider that can be tuned to the needs of the user, creating a meteorological record for a point in space and time. It utilizes the provided measured data and can include global radiation, temperature, wind speed and vapor pressure. In addition, a meteorology object can be instantiated with default values, or with estimates based on existing data if the event measured time series are missing. For data estimation, the FAO guidelines (Allen et al., 1998) are used. Parameters outlining qualities of the vegetation cover influence the rate of ET at each cell. These values are set for each cell and can include leaf area index, canopy coverage, vegetation height and stomatal resistance.

Some model applications may be driven by special boundary conditions, like pumping, measured or modelled groundwater heads, or a river entering the system. The measured flux data series can be integrated into the model by setting up Neumann boundary conditions connected with the appropriate water storages.

Usage of storages and fluxes in the catchment model

With the different kinds of water storages, boundary conditions and flux connections the landscape model is effectively complete. The parts are put together like building blocks and resemble a water and solute flux

model incorporating user defined detail. The surface of a cell is, by default, a flux node, where all incoming water is either routed into the soil water storages or to the channel system. In the case that infiltration capacity is not sufficient to move incoming water from the surface into the soil, excess water can be directed to an appropriate target. For some models, particularly those applied in arid or semi-arid locations where overland flow is a dominant process of runoff generations, the surface environment can be treated as a storage allowing for delayed infiltration and explicit spatial surface water routing. In other cases, excess water can be directed to the near channel for surface routing through open channels.

To demonstrate the flexibility gained by these more or less generic model building blocks, combinations mimicking existing model concepts are outlined in the following section.

A one dimensional hydraulic setup

A one dimensional highly detailed model, comparable to models like BROOK90 (Federer, 1995) or Hydrus 1D (Simunek et al., 2005), is developed through the allocation of a number of soil water storages, with connections based upon a Richards equation based flux connector. In terms of boundary conditions, the precipitation is directed to the surface water. The deepest layer of the cell is connected to a newly created sink, the groundwater recharge with a Richards equation based flux connector. If the precipitation rate exceeds the infiltration capacity, or if the soil water storages are full, the exceeding flux is routed to the sink "surface runoff". Figure I-3 outlines such a setup.

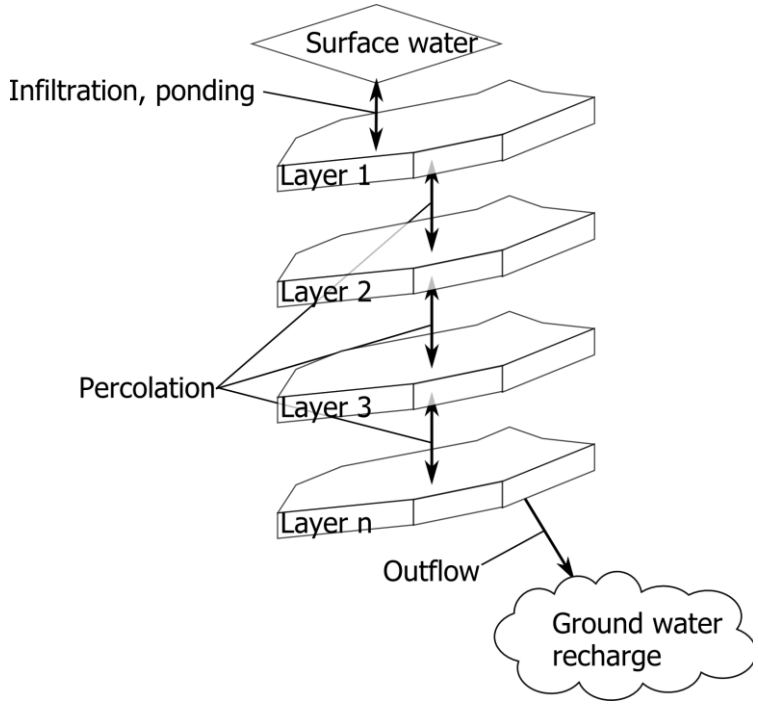

Figure I-3: A one dimensional setup using the CMF objects

Simulating macro pore flow with the given set of process description is possible and could involve the definition of additional water storages using a physically based model of macro pore flow as implemented by Jarvis et al. (1997) or by a single water storage with a dual porosity retention curve (e.g. Durner, 1994). This functionality has not been incorporated in the most current version of CMF, however, the process for doing so (along with other similar extensions) is well-defined.

Adding surface processes

The 1D model described above is only valid to use when the water input to the surface water is the canopy throughfall and no snow fall occurs. Evaporation and transpiration from vegetation surfaces is entirely neglected. To apply it for many common situations, surface water storages must also be defined (Figure I-4).

To incorporate interception losses, a canopy water storage is created and connected with the rainfall source. A snow pack water storage receives snow fall and releases the snow melt water to the surface water. For more advanced snow pack simulations a storage of liquid water in the snow pack can be incorporated.

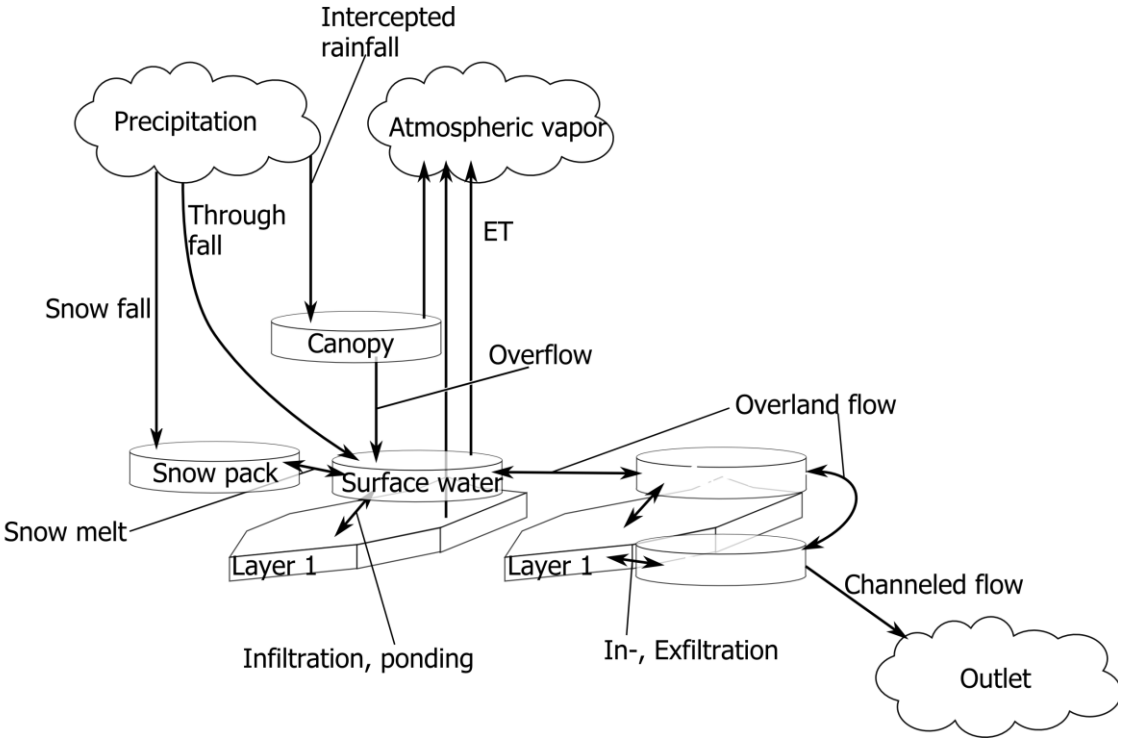

Figure I-4: Additional storage and fluxes for surface processes using CMF objects

2D hillslope setup

2-dimensional hillslope simulations are developed through the allocation of multiple 1-dimensional cells, with lateral flux connections between the layers of neighbouring cells. The lateral fluxes can be calculated with either a version of the Richards equation for unsaturated and saturated media or with the Darcy equation alone when the saturated and unsaturated zones are treated independently (as in DHSVM). The boundary conditions for unconnected sides of each hillslope element are generally defined as no flow boundaries, though this assumption can be relaxed.

A complex 3D catchment model

The extension to a three dimensional landscape model essentially amounts to a process of allocating additional cells in the same fashion as for the 2D case. Topological information from existing data, where elevation grids are a very convenient starting point, is utilized to develop the lateral connections between the soil layers and the surface water automatically. From the standpoint of the framework, this procedure simply amounts to the allocation of increased number of model elements,

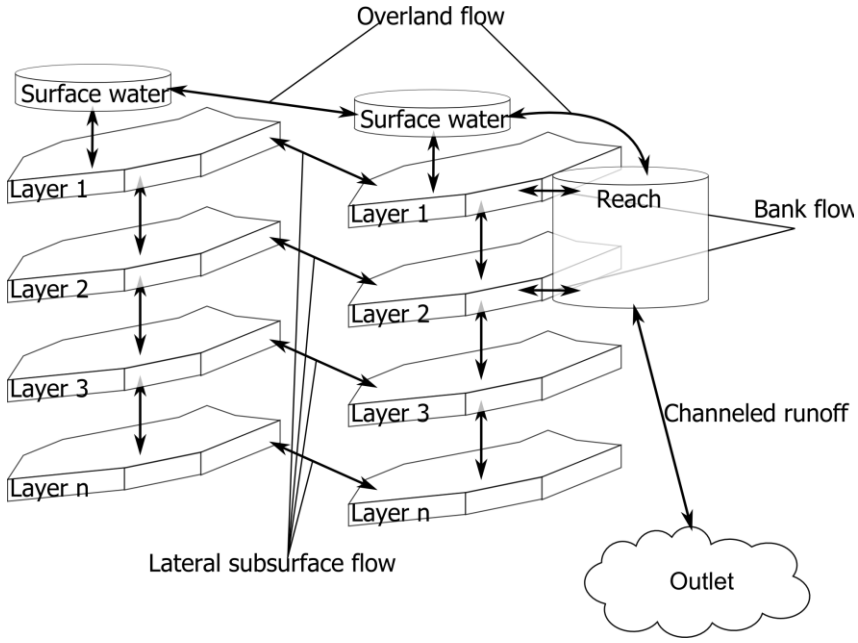

Figure I-5: A fully distributed and connected flux model using CMF objects.

although the complete definition of the full catchment model does require a number of addition types of water storage to effectively capture the variability and complexity of real world systems. These elements include water transport and storage by streams, dams, lakes and drainage pipe systems, the use of which is outlined in Figure I-5. This is not to suggest that this type of fully-defined and physically-based catchment model is not without fault – a fact that has been the focus of a significant part of the hydrologic literature since the development of the Freeze and Harlan (1969) blueprint for the development of such a model. Going beyond discussions of parameterization and scale, the complexity of this approach, within CMF, leads to a significant demand for computational power and with standard machines, is suitable only for small catchments (less than 10 ha) at this time. One mechanism to alleviate some of these shortcomings is to invoke simplifications that are appropriate for different simulation goals.

Lumped and semidistributed models

One of the key functionalities of the framework design of CMF is the ability to scale not only the spatial dimension of modelled objects, but to adjust the model complexity to both, the questions of interest and to the scale of application. Figure I-6 outlines an example of a simplified setup for regional river systems. In this case the spatial discretization is not accomplished through meshed elements, but rather by larger units such as hydrotopes or subbasins. The primary simplification behind this approach, which is often described as semi-distributed, is that the routing

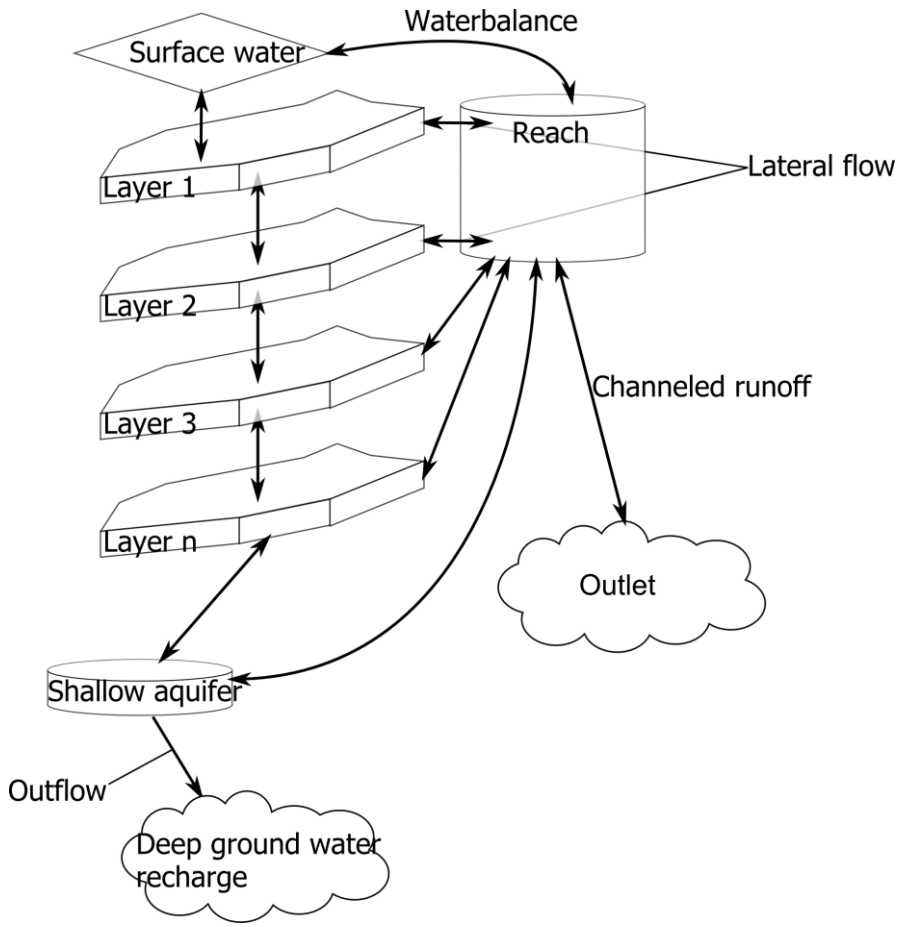

Figure I-6: A lumped model using CMF objects

of all surface flows occurs only within a network representing the stream system. The direct exchange of water between cells – overland and subsurface flow – is neglected. This approach is similar to semi-distributed models like SWAT (Arnold et al., 2005) and TOPMODEL (Beven and Kirkby, 1979), or, if whole subbasins are used in the development of a lumped catchment model, to models like HBV (Bergström, 1995). By varying the type of connection between the conceptional storages the approach of modular frameworks for lumped models like FLEX (Fenicia et al., 2006) and FUSE (Clark et al., 2008) can be adopted.

Running the model

To run the model, one of the numerical integration schemes, described above is chosen and connected with the equation system represented by the model project. The model is run by creating a time loop, where the solver is stepped forward by a time increment *dt*. This time increment is used by most of the solver types as a synchronization point. If a numerical calculation produces a local integration error that is greater

than the tolerances of the solver, additional shorter time steps are taken to ensure numerical stability. In all cases, any property of any landscape element, storage, flux or boundary condition can be written to a file, saved in memory for later usage, send to other submodels of an integrated approach or used for visualization.

Application of CMF in un- and poorly gauged catchments

Un- and poorly gauged catchments are the norm and not the exception (Sivapalan, Takeuchi, et al., 2003), but even at the end of the "IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012" no coherent methodology to deal with missing knowledge is established. Modular frameworks, like CMF can be adopted to the specific applications of CMF to real catchments, deal with limited knowledge. CMF has been applied to model different river systems, like the artificial catchment Chicken Creek, Germany (Holländer et al., 2009) and the upper Xilin catchment, (Kraft, Barthold, et al., 2011) as well as the application of an evapotranspiration sub-model of CMF in the San Francisco Catchment, Ecuador (Fig I-7.), as described below.

Figure I-7:Map of CMF applications

Chicken Creek

The Chicken Creek catchment is a small, artificial catchment, built in an open pit mining area near Cottbus, Germany. This catchment has been built up as a restoration of the former source area of the creek "Hühnerwasser" by the mining company, and was intensively monitored by the BTU Cottbus. The artificial catchment is described in detail by Holländer et al. (2009). As an experiment, different groups of modelers were invited to apply a variety of models to this catchment. The teams were provided with data the surface and soil properties, as well as climatic data but access to discharge data was blocked. Thus different model and modelers could predict discharge in a kind of ungauged catchment, despite the intensive instrumentation. Each of the 10 applied

models failed to predict discharge using only soil properties and driver data for the modeling. The preliminary version of CMF was used to create a lateral two dimensional model of groundwater flow in combination with a one dimensional variably sized unsaturated storage on an irregular grid. While groundwater levels were quite well predicted by CMF, discharge was heavily underestimated due to an overestimation of evapo-transpiration, using a not suitable parameterization of the soil surface. A complete discussion about the performance of the models used in this experiment is written by Holländer et al. (2009).

Xilin catchment

The upper Xilin catchment in Inner Mongolia, China has a size of 475 km² and is characterized by annual precipitation between 150 and 500 mm. Earlier applications of classical hydrological models, like SWAT and HBV failed to predict runoff, it was not possible to close the water balance (Schneider, 2008; Schäfer, 2009; Barthold et al., 2008). To understand the reason for failure of these applications and to understand the hydrological processes in the catchment, Kraft et al. (2011, Chapter III in this thesis) followed a multiple hypothesis approach suggested by Clark et al. (2011). CMF was used to model and test different hypothesis of subsurface connectivity of landscape elements in an advanced semi-distributed modelling approach. As a result, the hypothesis of strong external groundwater inflow into the catchment from outside could be supported, although not proven. Since discharge measurements were only incomplete available and in semi arid environments limited in their explanatory power (cf. Silberstein, 2006), different observations, ranging from “hard” directly measured data to expert knowledge and qualitative, visual observations, were used to constraint and reject models.

ET modelling for San Francisco catchment

In opposite to the Xilin river catchment, rainfall is in the catchment of the San Francisco river (75 km²) in south Ecuador with an annual sum between 2000 and 4500 mm abundant. Plesca et al. (2011) performed a model intercomparison of existing conceptional rainfall runoff models, namely HBVlight (Seibert, 1997), HBV-N-D (Lindgren et al., 2007), CHIMP (Exbrayat et al., 2010), LASCAM (Sivapalan, Ruprecht, et al., 1996), HEC-

HMS (Feldman and (US), 2000) and SWAT (Arnold et al., 2005). Although CMF was not used to model runoff for the San Francisco catchment, it was used to calculate potential Evapotranspiration and Precipitation from 5 meteorological stations on a 100x100 m² grid for the production of a height and distance weighted driver data set. The result was then for all models aggregated on the subcatchment scale, except for the distributed HBV-N-D model.

Building integrated nutrient turnover and transport models

The ability of a hydrological model not only to predict the discharge correctly but to model the flow pathways is crucial for process understanding (get the right answers for the right reasons) and, in particular, to model solute transport (Kirchner, 2006). Where and how fast water is travelling in systems can be characterized by its age and travel time, determined by the comparison of conservative tracer concentrations. McDonnell et al. (2010) provide a review of the common methods and current challenges of water age and travel time. Modelling the transport of nutrients and other non-conservative tracers in landscapes introduces challenges beyond the assessment of travel time and age. Additionally to the assessment of the pathways and flux intensity of the water the interaction of the solutes with each other, and the soil matrix need to be modelled. Due to the complexity of the microbiological, biochemical, geochemical and physical processes only a broad interdisciplinary team can create such models. Coupling existing, independent disciplinary models to an interdisciplinary “super”-model, is a way to facilitate the exchange of knowledge, governing equation and computer codes between different fields of science. There are two well-represented approaches to assembling a system of otherwise independent models, both with advantages and disadvantages: 1. the implementation of an entirely new software tool, which involves the integration of the disciplinary tools into a new integrated framework, and 2. fitting the existing models in an infrastructure designed to enable communication of states and fluxes between the models.

Examples for the first approach include nutrient flux model systems like SWAT (Arnold et al., 2005), RHESSys (Band, Tague, Groffman, et al., 2001), LASCAM (Sivapalan, Viney, et al., 1996), INCA (Whitehead et al., 1998) or

TNT2 (Beaujouan et al., 2001). The advantage of such a system is to have full control over feedback loops, high frequency data exchange and shared state and flux variables. But the freedom gained by this approach comes at a price: The reimplementation as a submodel in a greater integrated model requires significant programming. Adding to this challenge, the interpretation of older code, perhaps without adequate documentation, can be challenging for programmers not involved in the original design. Additionally, in the likely event that the re-implemented sub-model is also maintained by original developers (in its original codebase), divergence between the two versions is inevitable, and will quickly develop.

The second approach involves the application of data exchange infrastructures. Using this methodology, the integrated model makes direct use of compiled versions of well-established disciplinary models, and in this way, disciplinary experts are afforded an a priori understanding of the features of the integrated model they understand most deeply, and at the same time, the placement of their model within overall system, including its inputs, outputs and relationships to other system components, can be clearly outlined. The inner workings of the submodels, however, remain opaque under this strategy.

The simplest version of this approach involves reading the output of one model as the input into another, thus establishing fixed cause effect chains and disabling signal feedback loops (e.g. DNDC – Mike SHE case study (Cui et al., 2005). The most significant advantages of this setup of soft model coupling is simply that it can be implemented without changing the model software, and that as most models write output to disk, few of them are outside the scope of the technique. A step further is the utilization of a steering program, which defines a simple communication strategy between component models and acts as an execution supervisor, insuring temporal synchronicity between component models (e.g. LANAS (Dragosits et al., 2002). This kind of coupling is suitable for any kind of model, but requires that each component can be started from a well-defined system state. Since the model processes involved need to be completely terminated at each time step, and can require a high frequency data exchange, this methodology

can be costly in terms of computation time. Interest in this type of concurrent strategy has resulted in a number of recently developed, and increasingly robust, advances beyond the use of a simple execution supervisor. These more advanced coupling platforms, including OpenMI (Gregersen et al., 2007), PALM (Lagarde et al., 2001), are designed to facilitate the communication of models in memory. These systems represent a significant step beyond strategies that rely on files written to the disk, and while there are additional complications that go along with the increased sophistication, the increases in overall execution times far outweigh any additional set-up costs. In this case, the communication between the model processes is handled by the framework's run time environment, requiring code-level alterations to component models to correctly define the specific interfaces defined by the coupling platforms. The possibility for in memory high frequency exchange of states and fluxes between the models allow the studying of real feedback loops across the entire model domain.

An alternative to the use of specialized coupling software focuses on the use of scripting languages (promoted by Ousterhout (1998) and Sanner (1999) later for environmental models by Karssenberg (2002)). The use of scripting languages still requires the implementation of a specific interface within each component model, yet also has the potential to provide a greater degree of flexibility to the model developer. The most significant challenge in the use of a scripting language is likely that the development of a clear and well-established coupling interface designed for use with moderate modification of the component models is a non-trivial process. This, of course, is the primary reason for the continued growth of well established tools, such as PALM and OpenMI. Nevertheless, if the code defining the coupled models is consistent across models, and designed for access with scripting tools, the flexibility of the strategy may outweigh the challenges associated with its development.

CMF has been developed from scratch as an extension to Python and provides therefore an extensive programming interface for coupling with other models. Results of one model serving as an input variable of another, like eg. leaf area index, is as long as these parameters are directly accessible through the programming interface and hence simple

to change. Mass exchange of water fluxes between coupled modelling domains can be implemented as boundary conditions of CMF, while reactive solute fluxes are either modelled as an absolute or relative source/sink term of the finite volume equation. This flux oriented approach separates the model domain clearly from each other, and allows by averaging fluxes to use different time steps for the different submodels.

Coupling independent models by exchanging fluxes

This approach is used by Kraft et al. (Chapter IV) for a virtual hillslope experiment, coupling CMF with a plot scale plant growth model and a layer scale decomposition model, DECOMP (Wallman et al., 2006). The fluxes exchanged between the models are shown in Fig. I-8. It was not the goal of the study, to model reality, the models where initialized using default values; incoherencies, like DECOMP being developed for forest soils, but in this study applied in an agricultural context, were not handled. The study serves rather as a feasibility study for data exchange for biogeochemical modelling at the hillslope scale and a presentation of feedback loops. Such a feedback, fertilizer surplus at the upslope causes vegetational effects downslope, could be demonstrated by the study.

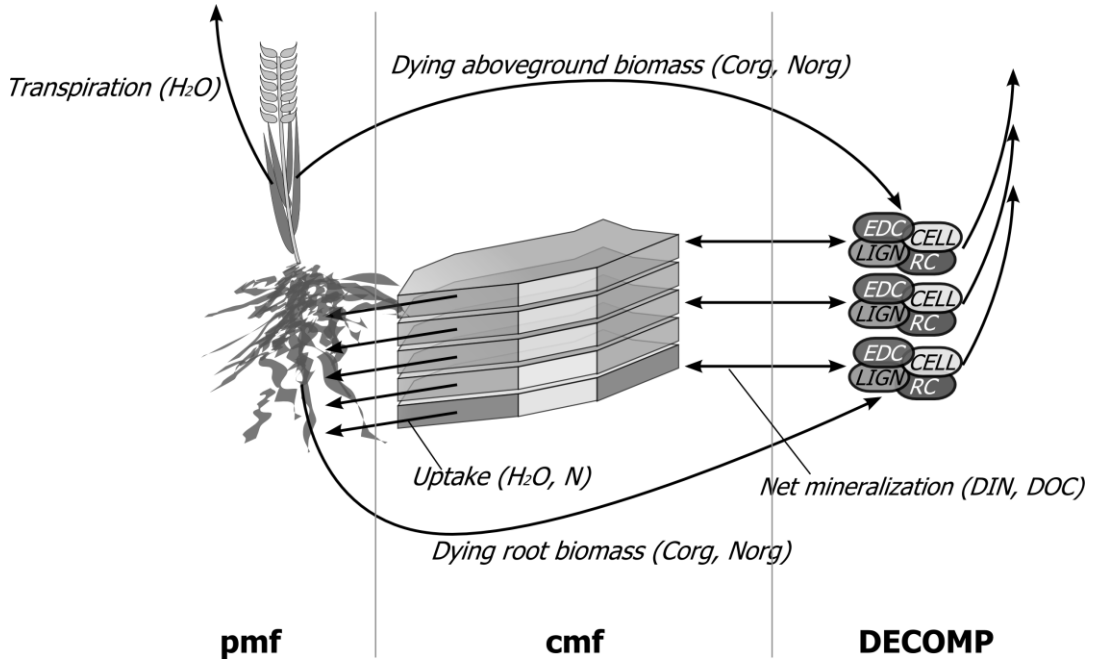

Figure I-8: Matter fluxes in one cell of the integrated model. Matter fluxes across model domain boundaries are addressed integrated setup script, while matter fluxes between cells are calculated by the transport model CMF. EDC (easily decomposing compounds), CELL (cellulose like compounds), LIGN (lignin like compounds) and RC (recalcitrant compounds) denote the four organic matter compounds in DECOMP.

Coupling independent models by consecutive overwriting of states

If state variables, like solute concentration or water content is stored by more than one submodel and fast changes of these storages can occur, the exchange of fluxes coupling approach might lead to divergating values in the storages of the models. In such a setup, typically occurring in integrated systems with submodels of higher complexity, another approach is needed. In CMF, not only external fluxes can be set by the driven Python script; state variables can be accessed directly also. Some solvers, like the CVODE solver, need to be made aware of the change of state variables, but otherwise state variables can be changed during the model run from outside. If all coupled model offer this ability, the steering script can exchange the fluxes calculated by one model by updating the state variables of the others. The advantage of this approach is a consecutive synchronization of the states, thus stabilizing the whole system. On the downside, the model domains are not as clearly separated as for the first approach. The coupling approach of LandscapeDNDC and CMF by Haas et al. (2011, Chapter V) uses the update of states method for stabilization of the coupled system. The LandscapeDNDC model is in

itself a suit of plot scale models, based on recoded internal coupled submodels with shared state variables. LandscapeDNDC is a complex carbon and nitrogen turnover model, including submodels for plant growth, soil and canopy climate, decomposition, ammonification, nitrification, denitrification and methonogenesis.

It is an extension with a new modular code basis of the DNDC model (Li et al., 1992). At the plot scale, LandscapeDNDC is designed to model the

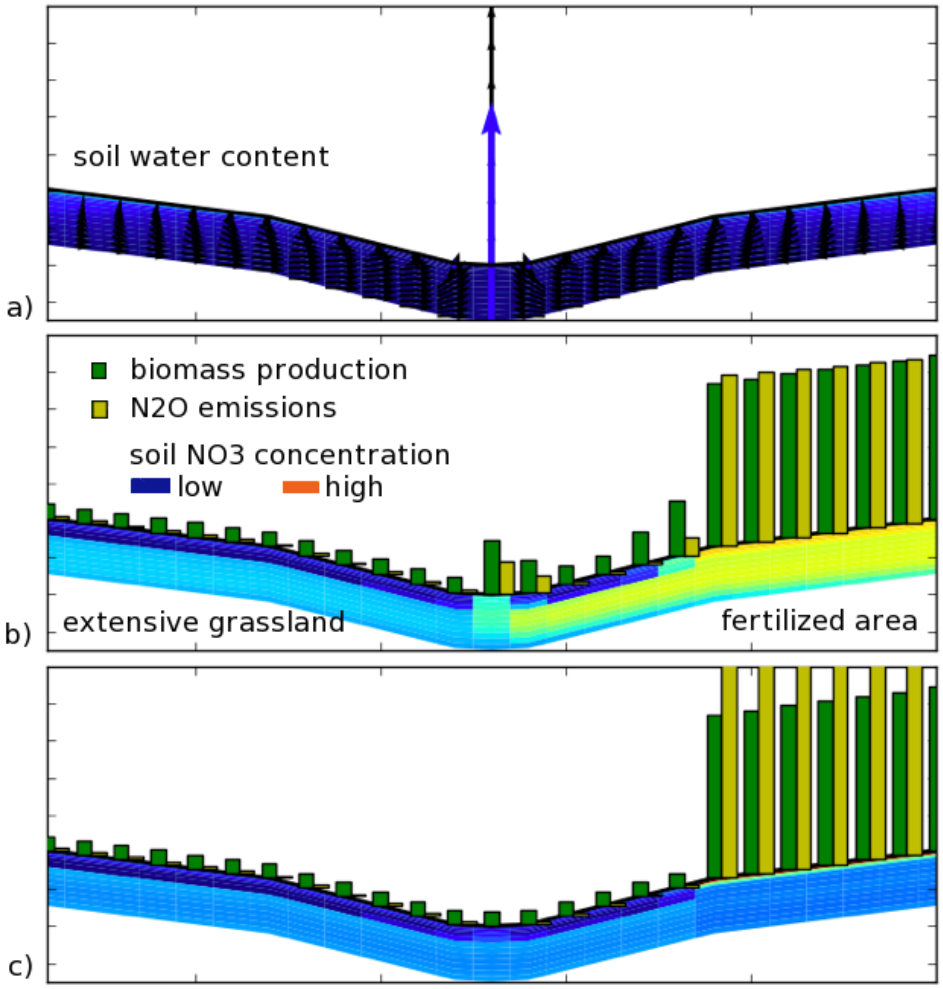

Figure I-9: Coupled LandscapeDNDC/ CMF and baseline (LandscapeDNDC stand alone) simulations of water and nitrogen fluxes along two hillslopes (left extensive grassland; right intensive fertilized rotation of winter barley, maiz, winterwheat: a) Soil water content (indicated by the blue countour plot) and the water fluxes along the hillslope as well as discharge formation (indicated by arrows) as simulated by CMF. b) Coupled simulation and c) stand alone simulaton of soil nitrate content (indicated by the contour plot of the soil layer) and accumulated plant biomass production (green bars) and N2O emissions (golden bars) before harvest in the third year of simulation.

complete nitrogen and carbon cycle in the bio- and pedosphere. However, exchange of water and solute fluxes between multiple plots is not feasible to model by design. By combining LandscapeDNDC as the model for

reactive fluxes and CMF as the model for transport fluxes, this kind of lateral connectivity can be included into the model setup. The different at virtual hillslope experiment with and without lateral transport is shown in Fig. I-9. With lateral fluxes included in the model (Fig. I-9b). N_2O emissions from the valley bottom occur as indirect emissions. If lateral solute transport is excluded, like in Fig. I-9c and in an earlier study of coupling DNDC with a landscape water model (Cui et al., 2005), the dislocation of denitrification of nitrate originating upslope does not occur.

Conclusion

CMF, as a generic modular framework for creating models of water and solute fluxes, has shown to be capable to follow a new paradigm of model development: the multiple hypotheses approach (Clark et al., 2011). Buytaert et al. (2008) demand in their comment "Why can't we do better than Topmodel" modern modeling tools to be modular, accessible, and portable, CMF fulfills these demands since it is published as free software (accessibility), written in ANSI-C++ and Python (portability) and as an extension library for the Python programming language (modularity, cf. Chapter II).

Does this mean we can "do better than Topmodel" with CMF or other generic frameworks, like Buytaert et al. (2008) and Clark et al. (2011) claim? The experience gained during finishing this thesis shows that the answer depends, on what is meant with "doing better". To fit a modeled to an observed hydrograph Topmodel and other conceptual lightweight models like HBV, HEC-HMS and LASCAM still "do better", and especially more economical: invention, implementation, test and rejection of model structures is, even if facilitated by software like CMF, time consuming and does not lead necessarily to high model efficiencies if only discharge is considered. Eg. to find a suitable model structure for the Xilin catchment (Chapter III) took much longer than planned, CMF was not part of the model intercomparison study by Plesca et al. (2011), due to this longer time. However, "out of the box" models failed completely to understand the water balance of the Xilin catchment, since such fixed model structures are not easily extended to processes like interbasin inflow as an additional source and the extensive riparian zones as an important

part of the evapotranspiration mechanisms. Understanding the underlying processes and pathways of water in the landscape in a “white box” model is crucial for the prediction of solute behavior in landscapes. CMF features the transport of conservative tracer also. However, conservative tracers do, by definition, not interact with the soil matrix and the vegetation and hence play a minor role for ecological studies. Pesticides and nutrients on the other hands react strongly with the environment. These reactions tend to be complex, most notably the interaction of the various forms of reactive nitrogen. Reactive nitrogen as one of the most important nutrients is not only essential for plant growth, but applied excessively as fertilizer, also the cause for many harmful ecological effects. It is a thread to biodiversity by eutrophication, it is effecting the climate if transformed to nitrous oxide and a source for soil acidification. Due to transport in its dissolved forms, as nitrate, ammonia dissolved organic nitrogen and nitrite cause (eg. fertilization, deposition) and effect (N_2O emission, eutrophication) are spatially and temporally displaced (cf. Chapter IV and V). Extending well established plot scale models like LandscapeDNDC with a lateral flow component like CMF seems to be a good starting point for the understanding of transport-reaction feedbacks and serve as guidance for the design of future field measurement campaigns on the hillslope and catchment scale.

Publications connected with this thesis

Submitted articles

Kraft, P., Barthold, F.K., Frede, H.-G., Breuer, L., 2011. Refining model structures in a poorly gauged catchment based on soft data, educated guesses and model failures. **Water Resource Ressources,** Manuscript #2011WR011792 (Chapter III)

Haas, E., Klatt, S., Fröhlich, A., Kraft, P., Werner, C., Kiese, R., Grote, R., Breuer, L. and Butterbach-Bahl, K.: Towards a new approach to simulating regional N2O emissions - the LandscapeDNDC Model, **Landscape Ecology**, Manuscript #LAND-S11-00551-2 , 2011. (Chapter V)

Accepted articles in peer reviewed journals

Plesca, I., Timbe, E., Exbrayat, J.F., Windhorst, D., Kraft, P., Crespo, P., Vaché, K.B., Frede, H.G., Breuer, L., 2011. Model intercomparison to explore catchment functioning: Results from a remote montane tropical rainforest. **Ecological Modelling**, in Press

Kraft, P., Vaché, K.B., Frede, H.-G. Breuer, L. 2011. A hydrological programming language extension for integrated catchment models, **Environmental Modelling & Software** 26, 828-830. (Chapter II)

Kraft, P., Multsch, S., Vaché, K.B., Frede, H.-G. Breuer, L. 2010. Using Python as a coupling platform for integrated catchment models. **Advances in Geosciences** 27, 51-56. (Chapter IV)

McDonnell, J.J., McGuire, K., Aggarwal, P., Beven, K.J., Biondi, D., Destouni, G., Dunn, S., James, A., Kirchner, J., Kraft, P., Lyon, S., Maloszewski, P., Newman, B., Pfister, L., Rinaldo, A., Rodhe, A., Sayama, T., Seibert, J., Solomon, K., Soulsby, C., Stewart, M., Tetzlaff, D., Tobin, C., Troch, P., Weiler, M., Western, A., Worman, A., Wrede, S., 2010. How old is streamwater? Open questions in catchment transit time conceptualization, modelling and analysis. **Hydrological Processes** 24, 1745-1754.

Holländer, H.M., Blume, T., Bormann, H., Buytaert, W., Chirico, G.B., Exbrayat, J.F., Gustafsson, D., Hölzel, H., Kraft, P., Stamm, C., Stoll, S., Blöschl, G. Flühler, H. 2009. Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data. **Hydrological Earth System Science** 13, 2069-2094.

Selected conference proceedings

Kraft, P., Haas, E., Butterbach-Bahl, K., Frede, H.-G., Breuer, L. 2011. Modelling nitrogen transport and turnover at the hillslope scale - a process oriented approach Nitrogen & Global Change, Key findings - future challenges, Conference Proceedings, Edinburgh, April 11 - 15, 2011

Kraft, P., Dalgaard, T., Schelde, K., Wohlfahrt, T., Frede, H.-G., Breuer, L. 2011. Predicting major subsurface transport pathways as a key to understand spatial dynamics of reactive nitrogen in stream water Nitrogen & Global Change, Key findings - future challenges, Conference Proceedings, Edinburgh, April 11 - 15, 2011

Kraft, P., Breuer, L., Vaché, K.B. Frede, H.-G. 2009. A modular framework for matter flux simulation at the catchment scale. Geophysical Research Abstracts 11, European Geosciences Union, General Assembly 2009, Vienna, Austria, 19 – 24 April 2009

Kraft, P., Vaché, K.B., Breuer, L. Frede, H.-G., 2008. A solute and water flux library for catchment models. Sànchez-Marrè, M., Béjar, J., Comas, J., Rizzoli, A., and Guariso, G. 2, 2053-2056. 2008. International Environmental Modelling and Software Society (iEMSs). iEMSs 2008: International Congress on Environmental Modelling and Software. Integrating Sciences and Information Technology for Environmental Assessment and Decision Making. 4th Biennial Meeting of iEMSs

II. A hydrological programming language extension for integrated catchment models

Philipp Kraft, Kellie B. Vaché, Hans-Georg Frede, Lutz Breuer

This chapter is published in the Journal Environmental Modelling and Software 26, pages 828-830.

Abstract

Hydrological models are created for a wide range of scales and intents. The Catchment Modelling Framework (CMF) extends the Python programming language with hydrology specific language elements, to setup specific hydrological models adapted to the scientific problems and the dominant flow processes of a particular study area. CMF provides a straightforward method to test hydrological theories and serve as a transport module in integrated, interdisciplinary catchment model approaches.

Keywords: Hydrological model, solute transport, model coupling, object orientated programming, catchment models

Software availability

Software name	Catchment Modelling Framework – CMF
Software homepage	http://www.uni-giessen.de/cms/faculties/f09/institutes/ilr/ilr-frede/download
Developers	Philipp Kraft, Kellie B. Vaché
Contact	Philipp.kraft@umwelt.uni-giessen.de
Year first available	2010
Availability	free download, GNU public license, v2
Software required	Python 2.6 with NumPy (>1.3) and Shapely (>1.2.4) on Linux and Windows
Hardware required	Multi-core CPU
Programming language	C++, Python
Program size	2.2 MB source code, 4MB documentation

Introduction

In the last 40 years, a multitude of more or less physically-based, deterministic hydrological models have been developed for various scopes, scales and regions. Despite the existence of broadly accepted governing equations in hydrology, including the Richards equation describing flow in porous media and the St. Venant equations defining open channel flows, modelling landscape-scale water movement remains hampered by heterogeneity, parameter uncertainty and lacking computational power.

In response to this challenge, Buytaert et al. (2008) suggest that model codes should be fully accessible, modular and portable. The incorporation of these qualities into the code is a mechanism to allow the model developer to aid model users in testing different hypotheses about flow regimes, and, ultimately, in applying the most appropriate model structure to different questions of interest. To these essential characteristics we add the concept of connectability, i.e. that model codes also should focus on between-model connections, particularly in the case of water quality models, to allow integration into a multi-objective landscape model approach.

Software design concept

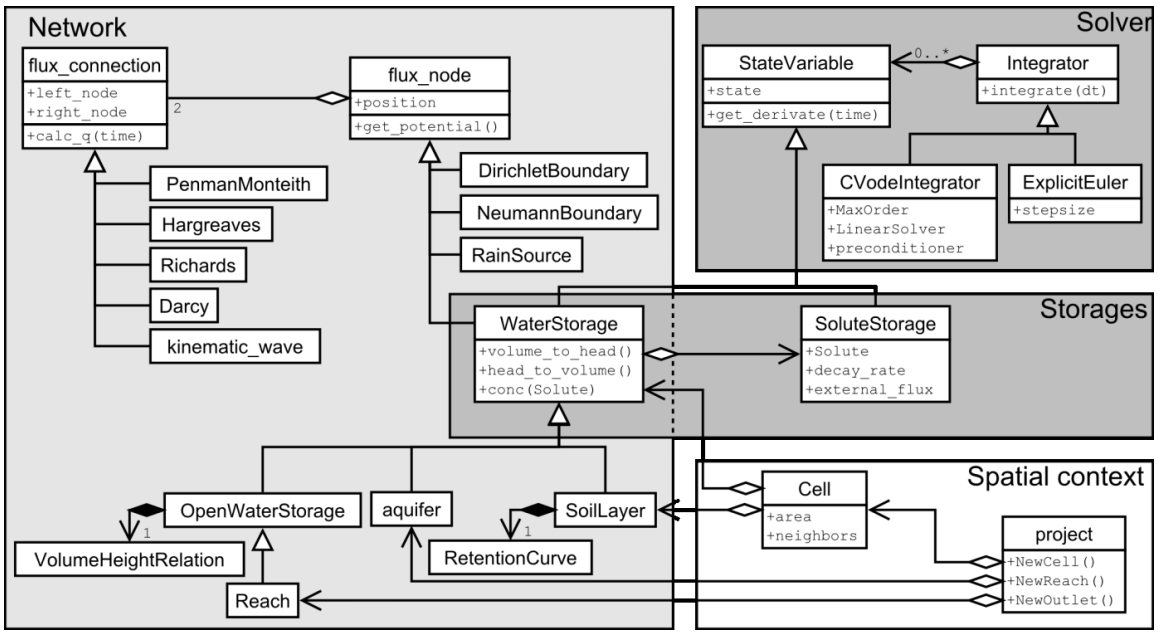

Figure II-1: Simplified UML class diagram of CMF. The left section (light gray) includes the components for defining the network of water fluxes, the upper right part (dark gray) the classes to create and solve the resulting equation system, and the lower right (white) part shows classes for the spatial context

Our approach presented here, originally based on the rejectionist framework approach by Vaché and McDonnell (2006), represents a model abstraction concept similar to the finite volume method (FVM) approach by Qu and Duffy (2007). The Catchment Modelling Framework (CMF) generalizes the FVM discretization scheme to let the user attach the finite volumes (water storages in CMF) with a variety of flow accounting equations (flux connections in CMF). The compounds of the model are assembled using the well known scripting language Python. The advantages of scripting computer languages for scientific modelling are described by Ousterhout (1998) and Karssenberg et al. (2007), namely rapid application development using very high level instructions and a dynamic typing system. To benefit from the simplicity of scripting languages on the one hand, and from the reduced computation effort of compiled languages, the core components of the framework are written in C++ as an extension for Python, using the "Simplified Wrapper Interface Generator" (SWIG) by Beazley and Lomdahl (1996).

A model built on CMF is composed of a network, a spatial context and a solver (Figure II-1). The nodes of the network correspond to the water storages and boundary conditions, and the network edges to the governing equations, like Richard's or Manning's equation, amongst others. The differential equation representing the water volume (V) in one storage-object *i* is then derived from the network, and is defined as:

$$\frac{dV_i}{dt} = \sum_{j=1}^{N_i} -q_{i,j}\left(V_i, V_j, t\right) \tag{5}$$

The flux function *q* is calculated by the flux connection objects, the edges of the network. N is the number of connected nodes of the actual node *i* and *j* denotes a node connected to *i*. Sources and sinks of water are realized by boundary objects. Solute storages are associated with each water-storage-object. The resulting system of ordinary differential equations for water volume and matter content is in most cases stiff and only integratable with an implicit solver. The CVODE solver (Hindmarsh et al., 2005) has been included in CMF for that purpose, along with other integration methods for the few cases of non-stiff setups. Any solver in CMF supports shared memory parallelism using OpenMP. Each node of

the network is part of a spatial context, either the whole study area (the project), like reaches and big aquifers, or of a horizontal discretization unit, the cell.

Using these components a wide range of hydrological catchment models can be build: Lumped conceptual models with physically based boundary conditions, semidistributed models as well as highly detailed physically based fully distributed models. Other modular framework approaches limit the range of models to be built to a specific model type. FUSE (Clark et al., 2008) and FLEX (Fenicia et al., 2006) for example are restricted to lumped and semidistributed approaches and the rejectionist framework by Vaché and McDonnell (2006) constrained to distributed topographic gradient driven approaches. The MIKE SHE model (Refsgaard and Storm 1995) covers a broader range of possible model setups. However due to the closed nature of the source code, the user cannot extend the system by his or herself.

Apart from the catchment scale, hillslope models, as well as three dimensional representations of fluxes in a lysimeters can be set up using CMF. A feasibility study concerning the coupling capabilities with biogeochemical models of the CMF approach has recently been published (Kraft et al., 2010).

III. Refining model structures in a poorly gauged catchment based on soft data, educated guesses and model failures

Philipp Kraft, Frauke Katrin Barthold, Hans-Georg Frede, Lutz Breuer

This chapter is submitted as an article to the Journal Water Ressources Research. If accepted for publishing, the copyright will be transferred to the American Geophysical Union.

Abstract

Determination of the dominant hydrological processes in catchments is difficult, even for the best examined catchments. Given the fact that most areas of the world are un- or poorly gauged and data are scarce new methods for understanding and predicting flow processes are needed. A method to create a model for the simulation of hydrological processes following a multiple hypothesis approach for a 475 km² catchment in a semiarid steppe environment of Inner Mongolia, China is presented. The model is setup according to the temporal and spatial sparse knowledge of the catchment. Apart from quantitative "hard data" like short and incomplete hydrographs, this involves mainly "soft data" sources, like results from tracer data analyses and expert knowledge like visual observations, stories from locals or analogies to other catchments.

Keywords: Semi-arid hydrology, Modular framework, Soft data

Introduction

Applying rainfall-runoff models in semi-arid environments is difficult and error prone (Silberstein et al., 1999). In such ecosystems only a small fraction, often less than 10%, of the rainfall leaves the system as runoff, while 90% and more of the rainfall is evaporated and transpired. As a consequence, a 10% error of evapotranspiration (ET) or rainfall in the model leads to a 100% error in runoff. In remote and poorly gauged catchments, like the upper Xilin river catchment (China), the error in measured rainfall and modeled ET is likely to exceed 20% easily (Schneider et al., 2007). How can one understand the water budget dynamics of such catchments, when conceptualizing hydrological

processes is still challenging even for intensively gauged hydrological systems (Tetzlaff et al., 2010)?

Clark et al. (2011) propose to formulate multiple working hypotheses which focus on the dominant processes in a specific catchment and test them by using a flexible model framework. With such a framework, the authors claim, all model components can be exchanged and varied, for testing the component inherent hypotheses. Testing these hypotheses requires "clever use of data" which may include (but are not restricted to) discharge measurements, but needs to go beyond fitting the model to the observed hydrograph. The rejections framework approach (Vache and McDonnell, 2006) uses mean transit time as a rejection criterion of hypothesis concerning the complexity of runoff generation in the shallow subsurface. This process has been advanced to a more iterative approach by Fenicia et al. (2008). Given the fact that data are scarce in many un- or poorly-gauged catchments, like the upper Xilin catchment (Barthold et al., 2010), information about mean transit time is not available. Others suggested using more soft data in the model evaluation process, including expert knowledge (Seibert and McDonnell, 2002).

The "Catchment Modeling Framework" CMF (Kraft, Vache, et al., 2011) is such a flexible framework were a wide range of model components can be arranged freely for building hydrological models based on working hypotheses. The model components form the nodes and connections of a network of storages and boundary conditions, connected by water fluxes. It is generalizing the rejectionist framework (Vache and McDonnell, 2006) using a finite volume approach similar to PIHM (Qu and Duffy, 2007). However by using a strict object oriented approach, the water flux network can be freely assembled, using the programming language Python.

In this study, working hypotheses are gained from field experience, tracer data analysis (Barthold et al., 2010, 2011), failed model applications (Barthold et al., 2008; Schäfer, 2009; Schneider, 2008), hydrogeological maps (Geological Survey, Inner Mongolia, 1980, 1:50.000, K50-3) are translated into several model architectures. This process was only possible through intensive discussions between modelers and field experts, as productive as predicted by Seibert and McDonnell (2002). The

formulation of the hypotheses / models is done in an interactive manner, step by step, and is tested against qualitative and quantitative descriptors of hydro system functioning.

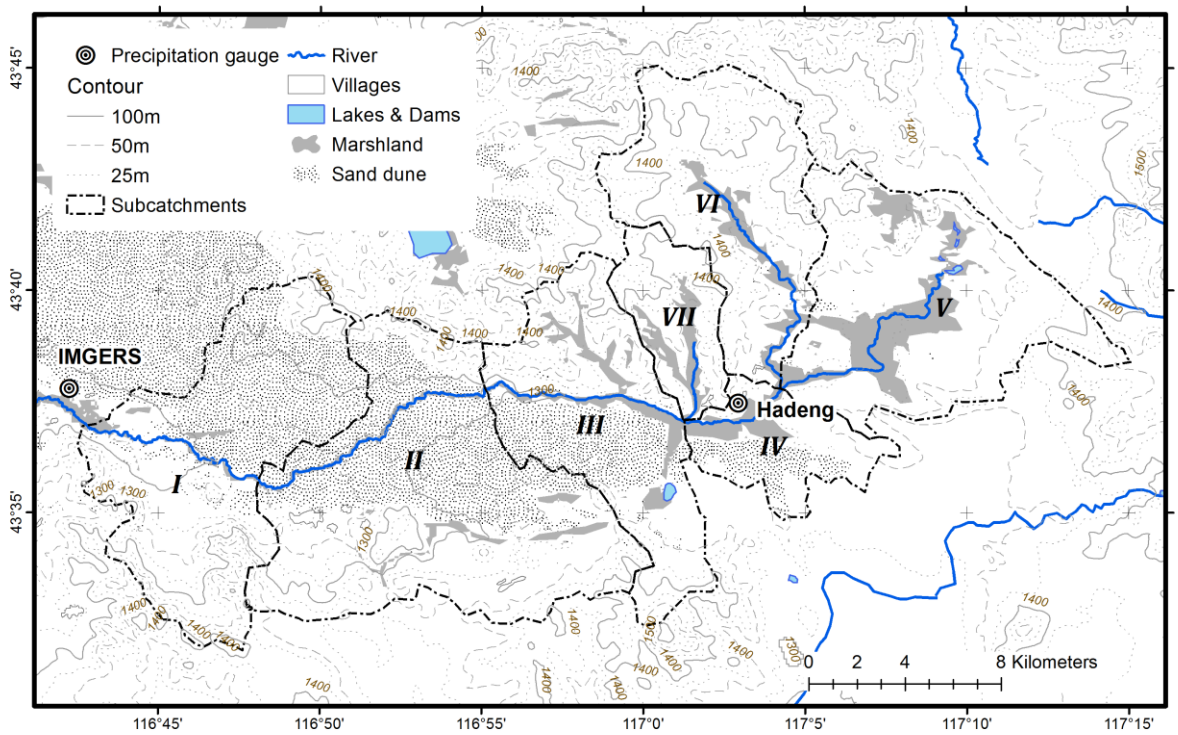

Fig. III-1: Map of the upper Xilin river catchment. Subcatchments are labeled I-VII. The two rivers in the north east of the study area entering the map boundaries are seeping into the ground.

The data set that serves as input into our models was collected in a subcatchment of the Xilin River Basin (43˚24' to 44˚40' N and 115˚20' to 117˚13' E) which is an endorheic river system located in the steppe regions of the Inner Mongolia Autonomous Province, China. The subcatchment, subsequently referred to as the Xilin catchment, drains an area of about 475 km^2 and has been described in detail by Barthold et al. (2010) which we will summarize in the following. The Xilin catchment has a difference in elevation of 325 m and ranges between 1175 m and 1500 m a.s.l. between its outlet and the highest point in the catchment. The semi-arid continental climate generates a maximum monthly averages of 18˚C in July and minimum monthly averages of -23˚C in January which average to -2˚C over the course of the year (Chen, 1988). Precipitation is highly variable in space and time due to prevailing convective weather conditions. A mean annual range of precipitation between 150 and 500 mm has been reported by Chen (1988) of which 60 to 80% fall between

June and August. Mean actual ET during the summer months is more than 90% of incoming precipitation (Schneider et al., 2007), but lower throughout the rest of the year.

The dominant vegetation in the area is steppe which covers about 40.5% of the catchment (Fig. III-1). Steppe turns into mountain meadows in higher elevated areas in the north and the east of the catchment. Marshland is developed in the vicinity of the river and its tributaries. Striking features of the catchment are paleo sand dunes that stretch through the center of the catchment and cover about 19.4% of the area (Fig. III-1). A substantial and considerably increasing part of the area is also used for cultivation of crops such as maize, wheat and rapeseed (Guo et al., 2004). Typical steppe soils such as Phaeozems, Chernozems and Kastanozems (IUSS Working GROUP, 2007), dominate in the Xilin catchment. Arenosols cover the sand dune area and Gleysols are developed in wet areas near the river and its tributaries. Calcisols, Cryosols and Regosols occur mainly in the vicinity of the Gleysols.

Material and Methods

Available data

The challenge of understanding the dominant hydrological processes in this catchment is data scarcity. Although many different data types were collected within the framework of the MAGIM research project (DFG Research Unit 536), the produced time series are often short, site specific and incomplete. The measured meteorological data depicts the general data situation: One station exists with a complete time series from 1999 – 2009 with precipitation, air temperature and humidity measurements.

Table III-1: Available observations for the upper Xilin river catchment

Measured data within the catchment (A)	Derived processes, qualitative data (B)	Qualitative and proxy data, expert knowledge (C)
A1 - Discharge during vegetation period from three consecutive years (2006, 2007, 2008)	B1 - Land cover map derived from LandSat TM [*Wiesmeier et al.*, 2010]	C1 - The riparian zone is water saturated all year (visual observation by F.K. Barthold, K. Schneider)
A2 - Near stream ground water table at 3 locations of a transect in summer 2007	B2 - The sand dune is a major contributor to discharge (soft data, from tracer analysis (Barthold et al., 2010) in some years, while deep groundwater and an unidentified source are mainly contributing in dry years, qualitative data)	C2 - Daily discharge 60 km downstream of the catchment under investigation
A3 - 20 locations of soil property measurements	B3 - Groundwater table at few locations in the year 1973 from a hydrogeological map (Geological survey, Inner Mongolia, China, K50-3)	C3 - Geochemical tracer composition and information from a DEM and a hydrogeological map suggest contribution from a GW aquifer located in the east of the catchment, outside the topographic catchment area
A4 - Meteorological data measured at IMGERS (1999-2009); occasional precipitation measurements at the headwater area, near Hadeng		
A5 - Snapshot discharge profile measurement along the river in 2008		
A6 - ET from eddy covariance measurements [*Schneider et al.*, 2007]		

Incomplete time series of a few months in the summer periods of 2004-2008 were measured also (Schneider et al., 2007; Barthold et al., 2008, 2010). While these additional measurements are not suitable for longer model runs, they show a high spatial variability of the rainfall. This observation is supported by reports of local farmers. However, a long term gauging station located 60 km downstream of the study area delivered discharge data for a catchment that drains 3600 km² of the area. The time series consists of daily discharge data for the period between 1954 and 2004 (Barthold et al., 2008). A continuous time series of meteorological data for the same time period was delivered from the Inner Mongolian Grassland Ecosystem Research Station (IMGERS) which is located in the centre of the catchment. Different models with varying complexity, including SWAT, HBV and a tank model based on geochemical tracers, were applied and calibrated against the discharge time series. Each application has failed to predict runoff well (Barthold et al., 2008; Schneider et al., 2007; Schäfer, 2009). Andréassian et al. (2010) introduced for such catchments the term “monster”, and encourage hydrologists not to hide their monsters from the public, but learn from the unsuccessful applications. Silberstein (2006) explains why runoff is not a good indicator for hydrological processes in semi-arid landscapes: Since only a tiny fraction of precipitation runs off, and the major part of precipitation is evaporated, a 10% error in evaporation leads to a 100% error in discharge. As a consequence, we used discharge measurements as one indicator of the processes only, but focus on other observations from the catchment, trying to make “clever use of data” (Clark et al., 2011). Namely the magnitude and recession of discharge during low flow conditions is more stable and depending on aquifer size and conductivity, than the peak discharges. A summary of observations used in the model application presented here is given in Table III-1. Discharge is the only integrative measurement of the catchment; however, its importance for the water cycle in this study area is limited. Hence quantitative data are scarce, the final model of the catchment has rather to be structural correct and in accordance with qualitative properties of the catchment, than show a high, directly measurable efficiency. This restricts the use of automated calibration of specific parameter sets and

favors changes to the model structure over changes of parameters to increase the plausibility of predictions.

Model approach

To facilitate structural changes, models of the upper Xilin catchment were created using the "Catchment Modeling Framework" (CMF) (Kraft, Vache, et al., 2011). CMF extends the Python programming language with a framework to design hydrological models based on a network of finite volumes. Fluxes between finite volumes are calculated using built in governing equations. The resulting system is integrated over time with an implicit solver for ordinary differential equation systems, in our case the CVODE solver by Hindmarsh et al. (2005) to avoid the numerical problems outlined by Kavetski and Clark (2011). Setting up a customized model with CMF involves the following steps: (1) Horizontal discretization of the study area surface into cells and of the surface water bodies which are treated as homogenous units. (2) Create discrete soil and surface water storages based on step (1) and a defined vertical model resolution. (3) Define the boundary conditions of the system. (4) Define the rules for water flux between storages and boundaries using either gradient or single state based equations.

The catchment size of 475 km² is not ideal for fully distributed modeling due to calculation time restrictions. Semi-distributed or lumped approaches restrict the hydrological fluxes to run only from a specific upslope area directly to the river and not to any other soil water storage. We therefore decided to combine the best from the two worlds. The saturated riparian zone (see Table III-1, C1) is considered to be an important indicator for the hydrological processes. Hence, bidirectional water exchange between the riparian zone and the river, as well as connectivity of the riparian zone to other water storages needs to be implemented into the model. The study area is discretized in a semi-distributed way, but bidirectional, potential driven connectivity between the discretized storages is implemented, in difference to most models of this class. The study area is discretized into seven subcatchments (I to VII, see Fig III-1.) following river junctions and geologic changes along the river. Each subcatchment is further divided into land cover classes: marshland, steppe, sand dune, bare soil, mountain meadow and

agricultural land (Table III-1, B1). Since not all land use units appear in every subcatchment, our discretization results in 28 different landscape units, called cells. The spatial extents of land cover units vary between the subcatchments, i.e. between the headwater region (subcatchments IV, V, VI, VII) and the downstream region (subcatchment I, II, III) (Fig. III-2). In the headwater region, extensive areas (9-20% of the subcatchment) of all year saturated marshlands surround the river. The upslope area of the headwater region is covered by steppe and mountain meadow. The downstream region is characterized by large sand dune areas with sparse vegetation cover and no wetlands in the direct vicinity of the river. In the outer areas of this region, steppe and mountain meadow ecosystem types are found. The valleys of the upslope area are partially covered by marshlands. In all our model approaches, only those land use units that are adjacent to the river contribute directly to stream discharge, i.e. the marshland cells in the headwater region and sand dune and bare soil areas in the downstream region. Fig. III-2 presents the conceptual models.

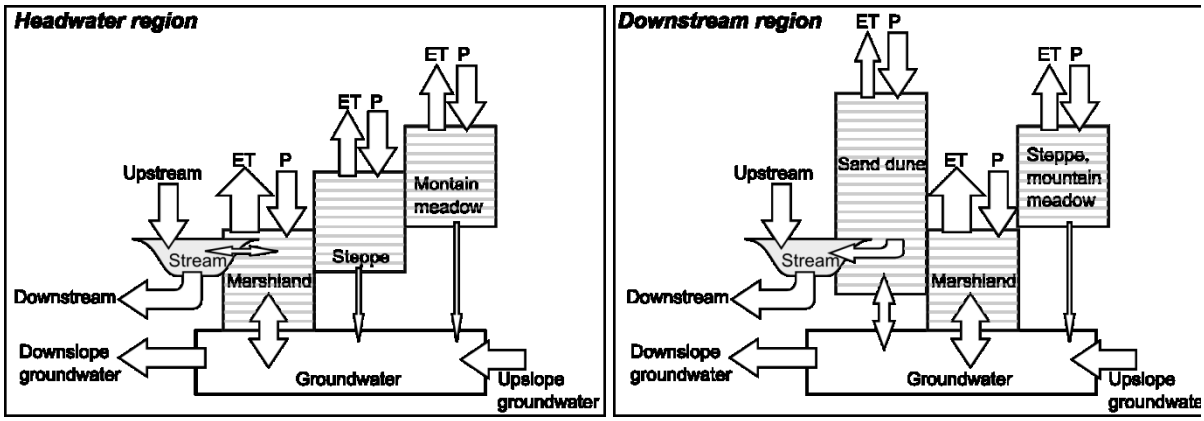

Fig. III-2: Conceptual model of water fluxes in different types of subcatchments of the upper Xilin river, Inner Mongolia. Groundwater does not contribute directly to discharge but indirectly through regions covered by marshland and sand dune.

Process representation

Vertical water transport is calculated using a one dimensional Richards equation approach. The retention curve for steppe, mountain meadow, agriculture and marshland is defined by a mean curve derived from various measured retention curves by Krümmelbein et al. (2006) at differently intensive used steppe sites near the outlet of the study area. The parameters of the retention curve for the sandy soils of the sand dune are estimated and were varied for several runtime realizations and conductivity is taken from Table III-1, A3. Actual transpiration and

evaporation is calculated using Shuttleworth-Wallace method (Shuttleworth and Wallace, 1985) following the implementation of the BROOK 90 model (Federer, 1995) according to land use (Table III-1, B1). Snow melt is modeled using a simple degree-day method. The routing of seepage and surface water is varied in course of the model adaption process. The routing involves gradient based, linear storage based and demand based approaches.

The different types and number of connections between water storages as well as observations (Fig. III-2) have been fed into different model structures by connectivity matrices and rated against each other. The four model structures we set up were the following:

(M1) Linear storage driven ground water system, with distinct land cover units contributing to the groundwater. Groundwater is redirected to the river. No transport between the land cover / subcatchment units.

(M2) Soil water from upland areas (steppe and mountain meadows) is routed by a variable saturated gradient based equation either to groundwater or the riparian zone. Surface water is routed using a kinematic wave approach. Groundwater is divided into an upper aquifer and a lower aquifer. For both aquifers, water is routed with a gradient based approach between subcatchments. Groundwater can feed the riparian zone, and the riparian zone is laterally connected with reaches using a gradient based approach.

(M3) Soil water from upland areas is routed only to the groundwater, surface water and groundwater storages are routed as in approach 2, although the groundwater system is simplified by omitting the upper groundwater aquifer.

(M4) The observation of a constantly wet riparian zone in the headwater area was used as prerequisite of the model. The water flow from the groundwater to the riparian zone is controlled by the model in order to hold a constant head of the soil water in the riparian zone.

The model structures were parameterized differently to describe the magnitude of hydrological fluxes between the modeled landscape elements. All model structures were tested using a wide range of

parameter sets. The type and number of the varied parameters differ between the approaches how subsurface flow between the landscape elements of the model is being realized. To test the response of the model structures, the parameters were varied in steps on a logarithmic scale. As a first step, the sensitive parameters of each model were identified. In a second step, the sensitive parameters were varied using a finer logarithmic grid, until the behavior of the system was known.

Due to the large uncertainty in the spatial distribution of rainfall and the low rainfall-runoff ratio, classic efficiency measures like Nash-Suttcliff efficiency or the correlation coefficient of modeled versus observed discharge were not used. Instead, results were evaluated against the following six observations (see Table III-1 for more details):

(1) the modeled ET in the steppe region was compared with eddy covariance estimations by Schneider et al. (2007) (Observation A6);

(2) continuous saturated conditions of the riparian zone over the modeling period (A2, C1);

(3) the sum of summer discharge should be in the same magnitude as observed (A1);

(4) the magnitude of the flood event in 2008 is met (A1)

(5) the observed recession of the base flow during the vegetation period of 2005 is an indicator of the groundwater storage size (A1); and

(6) the contribution of the sand dune to discharge, as described by the geochemical analyses in Barthold et al. (2010) (A5, B2).

Results

In a first step of model evaluation we checked whether the different model structures M1 to M4 were capable of matching the valuation criteria listed above. Criteria (1) and (6), i.e. the magnitude of modeled versus estimated ET in the steppe and the contribution of the sand dune area to discharge in wet years, were fulfilled by all model setups. However, the other observations were not met by the model set ups, e.g. the riparian zone dries out in model setups 1, 2 and 3, while model setup 4 was not converging to a balance of input and outputs. Hence, an

additional water source to the system from the catchment water balance needed to be assumed, which is supported by results from Barthold et al. (2010), who also suggest an unknown end member that significantly contributes to discharge. This missing end member could be groundwater influx from outside the topographic catchment area, e.g. the mountain ranges in the east of the catchment (Fig. III-1). The topographic setting in the headwater area also supports the idea of a strong groundwater influx from the eastern mountain range (see also Table III-1, observations B3, C3). Accordingly, model setups were extended to test the effect of an external subsurface influx to the catchment from the eastern mountain range. Since model setups M1 and M2 showed already numerical instable behavior for many parameter sets without the external input, they were discarded for this follow up experiment. In case of model 3, the influx is implemented as a Dirichlet boundary condition using a constant groundwater head in the headwater area (Table III-1, B3). In model 4, the groundwater head is set to a constant value, and external recharge ensured the head constant. The performances of the additional four model setups are compared in Table III-2.

The parameterization of model setup M3 with additional influx produced the best matches between model outputs and observations (Table III-2). Hence, M3 was chosen to demonstrate the overall model behavior in the following. The reaction of the modeled versus observed discharge for different parameterizations of M3 is shown in Fig. III-3. The model fails to predict short term reactions of the system, notably the extreme discharge event in 2008. However, the model satisfyingly reproduces the discharge

Table III-2: Criteria met by the different model structures. 'X' indicates that the model structure was in agreement with the relevant observation while '-' denotes a failure to match the observation. ET = evapotranspiraiton. See text for a description of the differences in model structures M3 and M4.

	Observation					
Model structures	Summer ET in steppe as observed	Riparian zone stays wet	Total summer discharge in range of observations	High flow event in 2008	Recession 2005 matches observation	Sand dune is contributing
M3, no influx	X	-	-	-	-	X
M3, influx	X	X	X	-	X	X
M4, no influx	X	X	-	-	-	X
M4, influx	X	X	-	-	-	X

recession after snowmelt in 2005, the general trend of the hydrograph in 2006 and the average of the low flow period in 2007. Parameters sets that have a stronger emphasis on surface runoff produce partially better fits regarding the discharge peaks, but fail to calculate the observed base flow. However, even these set ups do not reproduce the extreme event of 2008 as observed. This latter failure may be attributed to the fact that rainfall during the vegetation period is highly variable in the study area and rainfall events in the upper Xilin basin are local. A correlation between rainfall recorded at a station located in the headwater area (Fig. III-1, subcatchment I) and the IMGERS station located near the catchment outlet (Fig. III-1, subcatchment V), with a horizontal distance of 34 km between both, gives an idea about the extent of these convective weather conditions on spatial variability of rainfall (Fig. III-4). Note that the rainfall data from the upstream site was not included into the models due to subsequent instrument failure at this site.

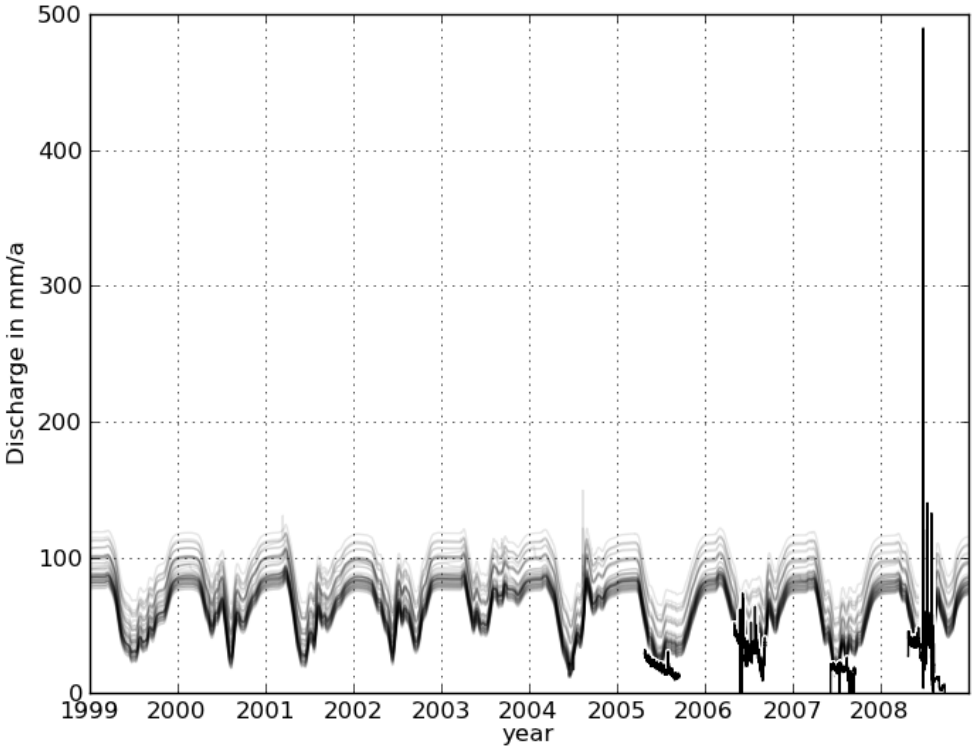

Fig. III-3: Modeled discharge for model structure M3 with influx (grey, 60 model realizations) and observed discharge (black).

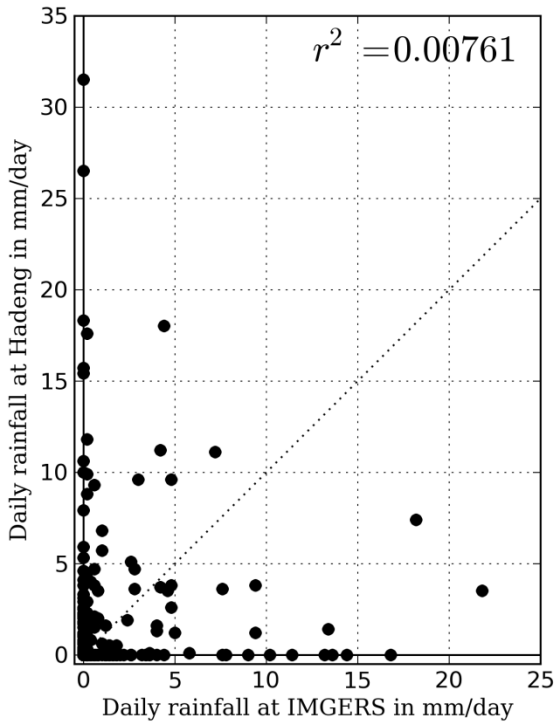

Fig. III-4: Correlation of rainfall at Hadeng (in the headwater region, subcatchment V) and IMGERS station (near to subcatchment I). The dotted line shows the 1:1 line. The measurements were performed in the summer years 2004 to 2006 [Schneider et al., 2007].

Due to this fact, one cannot expect the modeled hydrograph to fit the measured discharge at a time scale below seasonality. To compare model and reality, there is a need to utilize other similarities between the modelled and observed results. Silberstein (2006) emphasizes the importance of ET compared to discharge in semiarid landscapes. We therefore compared observed versus simulated ET. The modelled results of the steppe cell in subcatchment I agree in magnitude with results from eddy covariance measurements made by B. Ketzer and C. Bernhofer (Schneider et al., 2007) (Fig. III-5). The mismatch of timing between observed and simulated ET can be explained by the fact the model was not calibrated against this data set.

Fig. III-5: Comparison of modeled evapotranspiration (ET) from steppe land use in subcatchment I and eddy covariance measurements by B. Ketzer and C. Bernhofer [Schneider et al., 2007].

These high ET rates result in a nearly complete consumption of incoming rainfall in the steppe. However, different land cover and soil types are likely to influence the results. In respect to their parameterizations three different groups of land cover types exist: (1) The dry and silty upland areas, with steppe, agricultural land and mountain meadow as dominant land cover types, (2) the sandy areas, covered by the sand dune vegetation with its vegetation free areas (bare soil) and (3) the water saturated marshlands (Table III-3).

In our model the land cover types steppe, mountain meadow and agricultural land share a similar parameterization concerning land cover and soil type, thus show the same characteristic of ET. The sandy areas (sand dune and bare soil) differ from (1) in their soil physical properties, and were, due to the lack of measured retention curves, tested with different soil physical parameters. The vegetation parameters were chosen to fit dry open forests at the sand dune; no vegetation was used for bare soil. The marshland area was parameterized as a dense shrub land, while the soil properties, due to constant saturation do not influence the actual ET and thus have no influence on modeled ET. In all land cover types, except bare soil, the complete rooting zone (assumed to

be 1 m in depth) is utilized for water uptake by the vegetation. These assumptions lead to the differing results for ET presented in Table III-4.

Noteworthy amounts of precipitation seep to groundwater only on the land cover type bare soil. With vegetation, most of the percolating water is taken up by roots. Another special case is the marshland, where upwelling groundwater is feeding the water demand of the shrub vegetation in most of the subcatchments. In subcatchment V, the riparian zone loses 850 mm/a by ET, or 14.1 Million m³/a, according to our model with 4.8 Mill. m³/a rainfall. The difference of 9.3 Million m³/a has to be contributed by groundwater (note: The change of units is to avoid confusion between differing reference areas if area normalized units (mm) were used).

As we do not know to what extent groundwater is also feeding the Xilin river, an additional amount groundwater resources are needed to close the water balance of the Xilin catchment. Our model predicts a surplus in the water balance of only 1.5 m³/a in the headwater area and similar behaviors occur in other subcatchments. This results in a mismatch of water input and output. This water gap is the reason for failing to predict wet riparian zones by model setups without external water input. By including an external water input to M3 we are able to close the total, long term water balance of the whole study area as $P + Q_{in} = ET + Q_{out}$.(Table III-5). Although large uncertainties remain for all given water balance terms, all model results for M3 indicate that the Xilin River discharge is driven by groundwater resources outside of the catchment, except for extreme events. Based on our model results we conclude that 84% of the catchment area contributes to runoff only during extreme rainfall events by surface runoff and groundwater recharge is nearly negligible. We further reason that base flow originates from the mountain range outside the topographic catchment area and the bare soil areas of the sand dune within the topographic catchment boundaries.

Table III-3: Parameters for different land use types for the Shuttleworth-Wallace ET model.

Landuse	LAI	Veg. height [m]	Root depth [m]	Stomatal resistance [s/m]	Albedo [-]	Canopy closure [-]
Sand dune	3.5	0.7	0.5	150	0.23	0.5
Bare soil	0	0	0	1000	0.20	0
Riparian zone	5	2.5	0.5	100	0.17	1
Steppe	2	0.3	0.3	100	0.23	0.8
Mountain meadow	2	0.3	0.3	100	0.23	0.8
Agricultural land	1	0.5	0.5	100	0.27	0.8

Table III-4: Simulated mean actual evapotranspiration (ET) of the various land cover types in the Xilin river catchment.

Land cover	ET [mm/a]	Fraction of evaporated precipitation [%]	Fraction of ET from land cover type [%]	Area fraction of land cover type [%]
Mountain meadow	268.3	92.0	19.7	20.6
Agricultural land	268.7	92.2	1.7	1.8
Steppe	268.9	92.3	40.6	42.6
Marshland	489.0	167.9*	16.0	9.2
Bare soil	165.3	56.7	3.9	6.8
Sand dune	267.1	91.7	18.0	19.0

* a value for ET above 100% indicates ET from upwelling groundwater resources

Table III-5: Simulated water balance terms for model structure M3 with external inflow for the simulation period 2000-2009.

Water balance term		Mean [mm/a]	Sd* [mm/a]
Precipitation	P	291	
Subsurface inflow	Q_{in}	78	± 11
Evapotranspiration	ET	292	± 2
Discharge	Q_{out}	68	± 10

* standard deviation from parameter variance in model M3 with influx

Discussion and conclusion

Silberstein (2006) has shown the challenges for modeling semi-arid catchments, while Sivapalan et al. (2003) called for a whole decade to focus on the prediction in ungauged basins. In this study, we have faced both challenges. Hence, failing of classical runoff based modeling approaches should be rather the norm than the outlier. Establishing a model of the dominant hydrological processes is especially problematic, due to the ill-defined boundary conditions in our catchment: not only the properties of the subsurface are unknown, but also the distribution of precipitation and ET over space and the major contribution of snow melt to discharge which has not been measured directly. Kuczera et al. (2010) conclude from failed model applications caused by standard errors of 25% in precipitation: "It is one of the major shortcomings of catchment hydrology that errors in the inputs are typically ignored or treated unduly simplistically." Although we cannot calculate the standard error of precipitation from only one station and some additional rainfall measurements in Xilin catchment, Fig. III-4 indicates that precipitation is highly variable in space and that using only a single rain gauge stations introduces *per se* an error in prediction in this catchment.

Does the poor data quality explain the poor runoff modeling results completely? The model structure approach shows, in difference to the classical modeling approach of fitting only one fixed model to observations, a better conformance with the overall water balance of the system, and in particular a better representation of low flow conditions. The credibility of the model is even substantially increased, if inter basin groundwater flow is considered as one of the major source for stream discharge and riparian zone ET. Le Moine et al. (2007) have concluded from applying rainfall runoff models in 1040 catchments of France that interbasin flow is not an outlier, but rather the norm. But Kuczera et al. (2010) claim to explain the water balance for one of the 1040 catchments reviewed by Le Moine et al. (2007) with uncertainties of rainfall, which are certainly lower than in the Xilin catchment. The importance of groundwater inflow is also shown by Barthold et al. (2011) through a geogenic tracer analyses, despite the fact that contribution of this influx is heavily depending on the choice of tracers. Even in regions that have

been well studied, e.g. for 50 years like the Great Basin in Nevada and California, the importance of interbasin flow is still debated (Belcher et al., 2009). Intensive tracer studies and detailed mapping of hydrogeologic features, e. g. faults, does not result in a scientific consensus if water sources in the death valley are fed from interbasin groundwater flow (Belcher et al., 2009) or from ancient groundwater reservoirs (Anderson et al., 2006). With this background, the conclusion of this study to explain the area of wetlands and the relative high base flow of the Xilin by interbasin groundwater inflow might seem bold, but plausible.

However, to scrutinize this explanation, additional measurements are needed to constrain the boundary conditions of the system. Fan et al. (2006) used remote sensing data to model the regional distribution of ET from the ground energy balance for the study area. Since only one Landsat scene has been used, shortly after a rainfall event, the modeled values could not be used as an indicator for cumulated water losses to the atmosphere. Applying this approach to a longer series of Landsat scenes could provide an important data base to correct the water fluxes back to the atmosphere. However, most crucial is the better knowledge of the regional distribution of rainfall, e.g. by using radar technology. To gain more knowledge concerning the groundwater fluxes a complete inventory of groundwater wells in the region including the Baicha Shan Mountains in the vicinity of the catchment and the lake Da Li Nuo'er together with a questionnaire of groundwater usage from each well is needed. From this data, a steady state groundwater flow model could be derived to support the interbasin flow conclusion. To exclude ancient water from within the catchment as the major source like in the Death Valley debate, the age of in the groundwater of the headwater region need to be identified in the range of centuries, using ^{14}C measurements (McDonnell et al., 2010), although the geomorphology of the headwater area indicates the interbasin flow hypothesis.

Despite all uncertainties and chances to complete the measurements listed above, the following conclusions can be drawn: It is likely (in a Baysian sense), that land use change in the topographic catchment of the Xilin river will not influence groundwater recharge and the low flow conditions of the river, since this study showed a range of indications for

dominant external inflow. Flood events however, are driven by surface water runoff, both in the model and indicated in the field by extensive gully systems.

Usage of a modular framework in this study has been useful, not only to follow roughly the multiple hypothesis approach (Clark et al., 2011), but also to model the distinct features of the catchment, like the large wetlands in the arid environment. Classical semi-distributed approaches, lacking the ability to route water between terrestrial elements, are not suitable to model such systems. Failing with the initial modeling approaches for an ungauged catchment is natural. To adopt the measurement strategy to system insight gained from these failings, the modeling strategy needs to start early in a project and include a systematic method to reveal the normal anomalies of the study area.

Acknowledgments

The authors like to thank Katrin Schneider for the kind provision of raw data, the Deutsche Forschungs Gemeinschaft for funding the field work (DFG Research Unit 536, MAGIM) and the European Union for funding the development of the modeling tools by NitroEurope IP, component C4.

IV. Using Python as a coupling platform for integrated catchment models

P. Kraft, S. Multsch, K. B. Vaché, H.-G. Frede, L. Breuer

This article is published in Advances in Geosciences 27, 51-56.

Keywords: Model coupling, integrated modelling, Python, framework, CMF, PMF

Abstract

Interdisciplinary sharing of knowledge is a key for understanding matter fluxes in landscapes. However, models of transport and reactive fluxes from different disciplines need to work seamlessly together, to capture the tight feedback loops between different compartments and process domains of a landscape. Techniques to facilitate the integration of model codes for integrated catchment modelling exist, but are still scarcely used. In this paper, we are testing a scripting language, Python as a model coupling platform, and demonstrates effects of feedback loops on a virtual agriculturally used hillslope.

Introduction

The movement and storage of water in the environment is defined by a series of complex relationships involving atmosphere, biosphere, pedosphere, and hydrosphere. These relationships vary in time and space, and while capturing them within a simulation requires simplification, the process of simplifying the relationships may lead to useful predictions and insights into system function. Solute-based models, which add an additional layer of complexity, are often based upon theory developed in various disciplines. In a watershed model, including nutrient dynamics, the theories from a wide range of disciplines are needed for the model formulation. Disciplines involved range at least from soil and hydrologic sciences, to biogeochemistry and agronomy, not to mention physics and mathematics.

We argue that the degree of success any model has in capturing the key features of such a wide variety of fields, depends on buy-in from

disciplinary experts, and further, that this vetting process is facilitated through transparency associated with the model development strategy.

In this paper, we outline the strategies that have been developed to produce such integrated models of environmental process. These include soft coupling (e. g. Cui et al., 2005), redevelopment (e. g. Band, Tague, Brun, et al., 2001), as well as the use of an explicit coupling platform (e. g. Gregersen et al., 2007). The paper outlines some advantages and disadvantages of each of these strategies, and suggests a fourth, less explored but potentially fruitful alternative. Specifically, this alternative involves the development of a high-level, object oriented programming language, accessible through standard scripting tools, and targeted to the development of integrated process models. Ousterhout promoted this use of a scripting language as "glue" between models written in compiled, high performance languages over a decade ago (Ousterhout, 1998). However, the potential of scripting languages to design coupled, yet independent model suites is up to now scarcely used. We suggest that a targeted language possesses a number of significant advantages that have yet to be explored adequately. The system outlined here is proposed as an initial step in the development of an open source standard based code, which focuses on accessibility and portability, as called for by Buytaert et al. (2008).

The processes to be modelled in an integrated matter flux catchment model fall in one of two categories: transport processes and local turn over processes. Transport processes, usually by water, air or management, need to be modelled with spatially explicit models. Turn over models, like plant growth models, biogeochemical process models or local energy budget models are rather plot models without a definite spatial domain. In an integrated catchment approach, a coupling needs to be established between one instance of each type of transport model and many instances of the local models.

This paper focuses on the description of such a strategy, designed to facilitate the integrated simulation of watershed scale hydrology and solute transport.

Method

To illustrate the simplicity of coupling independent models using Python as a "glue" language, three different models were coupled and applied on a virtual hillslope. A water and solute transport model, a plant growth model and a model of organic matter decomposition.

Turnover of dissolved inorganic nitrogen (DIN) and water into biomass is calculated by the plant growth model, while the decomposition model calculates the turnover from dead biomass to the components DIN dissolved organic carbon (DOC) and gaseous carbon losses (CO2). Relocation of the two dissolved compounds DOC and DIN is governed by the water and solute transport model. The models used in this exercise were not chosen to fit a specific theory, but for accessibility and simplicity.

Transport model

The hydrological model framework CMF (Catchment Model Framework) by Kraft et al. (2008), which is based on the rejectionist framework by Vaché and McDonnell (2006), is an extension to the Python programming language designed to desgin water transport models. A model in CMF is set up as a network of storages and boundary conditions, connected by flux calculating submodels. CMF allows for the development of detailed mechanistic models as well as lumped large scale linear storage based models. In this study, a two dimensional Richards-based hill slope model was setup, whereas in an ungauged artificial catchment study (Holländer et al., 2009) a 2.5 dimensional Green-Ampt / Darcy approach was chosen.

The framework was designed to be connectible with other models through the implementation of a clear application programming interface (API). To demonstrate this functionality, a virtual hillslope using a variable saturated, continuous model was set up, using a discretized form of the two dimensional Richards equation. The hillslope is divided laterally into cells and each cell is divided vertically into layers. The flux between the layers of one cell (percolation and capillary rise) and between the layers of adjacent cells (lateral flow) is calculated using the wetness-based form of the Richards equation, spatially discretized using a finite volume approach. Surface flow is routed to the bottom of the hillslope

using a kinematic wave approach. At the bottom of the slope, a constant head boundary condition of 25 cm below ground is imposed to simulate a downslope ditch with constant head. A soil depth of 3m with impervious bedrock is assumed, whereas evapotranspiration is simulated by the plant growth model, through the API. However, for applications not involving a plant growth model, methods of calculating evapotranspiration are implemented in CMF. Solute transport is modelled using a simple advective transport scheme.

Turnover model 1: plant growth

Plant growth is determined with the Plant growth Modeling Framework (PMF)(publication in preparation). The model divides the plant into its physical components root, shoot, leaf, stem and storage organs. In the physical structure the growth processes are calculated on an abstract level. These components are related to process modules, which hold numerical solutions for the growth processes. The model can be adjusted to agriculture crops without changing the fundamental structure. Two interfaces handle the data transfer between PMF and other models or databases.

In this study, PMF is parameterized to represent summer wheat. Daily biomass accumulation is calculated with the radiation use efficiency and solar radiation (Acevedo et al., 2002). The biomass is allocated at the plant organs in relation to the development phase, which is determined using the thermal time concept (Monteith and Moss, 1977; Miller et al., 2001). Drought and nitrogen stress limit growth, but the plant can adapt to these stresses by varying the root biomass distribution. Stress is defined in PMF as 1-α, where α is the ratio between actual uptake of water or nutrients and potential uptake.

Root water uptake is calculated from potential transpiration and a crop specific response function relating uptake and soil matrix potential. The water uptake is represented as sink term in the water flux equation. This concept is similar to the macroscopic water uptake approach type II (Hopmans and Bristow, 2002; Feddes et al., 2001). Nitrogen uptake is divided into an active and a passive component following Simunek and Hopmans (2009). Passive uptake is the product from the dissolved

nitrogen concentration and water uptake. Active uptake is determined from the residual nitrogen demand after passive uptake assuming Michaelis-Menten type kinetics (Simunek and Hopmans, 2009).

Turnover model 2: decomposition of organic matter

The DECOMP model (Wallman et al., 2006) is a semi-deterministic model of decomposition of organic matter. It was developed as a part of the integrated plot scale forest biogeochemical model ForSAFE. It includes four carbon pools representing decomposable components, cellulose-like material, lignin-like material and recalcitrant material. Each pool has a potential transformation rate, and parameters to describe the reaction of the decomposition rate to environmental conditions, like wetness, soil temperature and soil acidity. The products of decomposition are distributed between dissolved organic carbon (DOC) and carbon dioxide (CO2).

For simplicity, only one Nitrogen pool exists for the four carbon pools. The gross mineralisation rate of Nitrogen is calculated from the mass of carbon decomposed to DOC and CO2. The mineralised N is partly released to the soil solution and partly immobilized, depending on the N content in the organic matter. To extract the model from the integrated ForSAFE code, a minimal model version was reimplemented using C++ as a Python extension, similar to CMF. The slightly different curve shape to calculate the immobilization / mineralization ratio does not affect the usability of the model to be used as a demonstration code example in this model coupling exercise. The parameterisation of the model was taken from Wallman et al. (2005). The parameters were determined to represent the behaviour of forest soils. Agricultural soils may behave differently, but as stated earlier, the focus of this paper is rather the feasibility of integrated models and the coupling approach, than on the parameterization of the models.

Integration

The three models were coupled with each other using a common setup script in Python. Since CMF and DECOMP are Python extensions, written in C++ and PMF is entirely written in Python, the integration is straightforward. Existing legacy model codes can be wrapped as a Python

extension using tools like SWIG (Simple Wrapper Interface Generator) (Beazley and Lomdahl, 1996) for model codes written in C or C++ or F2PY (Peterson, 2009) for model codes written in FORTRAN. The level required to wrap an existing model code depends primarily on the modularity of the existing code.

For our study, CMF was set up as a fully connected two dimensional hillslope model. For each lateral unit (cell in CMF) an instance of the plant growth model was created, and for each vertical unit (layer) in each cell an instance of DECOMP setup.

Figure IV-1 depicts the communication between the models. To keep the figure clear, only matter fluxes between model domains are shown.

Two different interfacing strategies has been chosen. Exchange of data between DECOMP and CMF is implemented using a method to be called by the main time loop, copying explicitly the data between the models. This simple strategy of data exchange between independent models is facilitated using a scripting language as interface, since necessary conversion or interpolation of data can be accomplished using the built-in features or existing mathematical libraries of the scripting language. To illustrate the implementation of this strategy, the data exchange between CMF and DECOMP at the layer scale is shown in the appendix A.

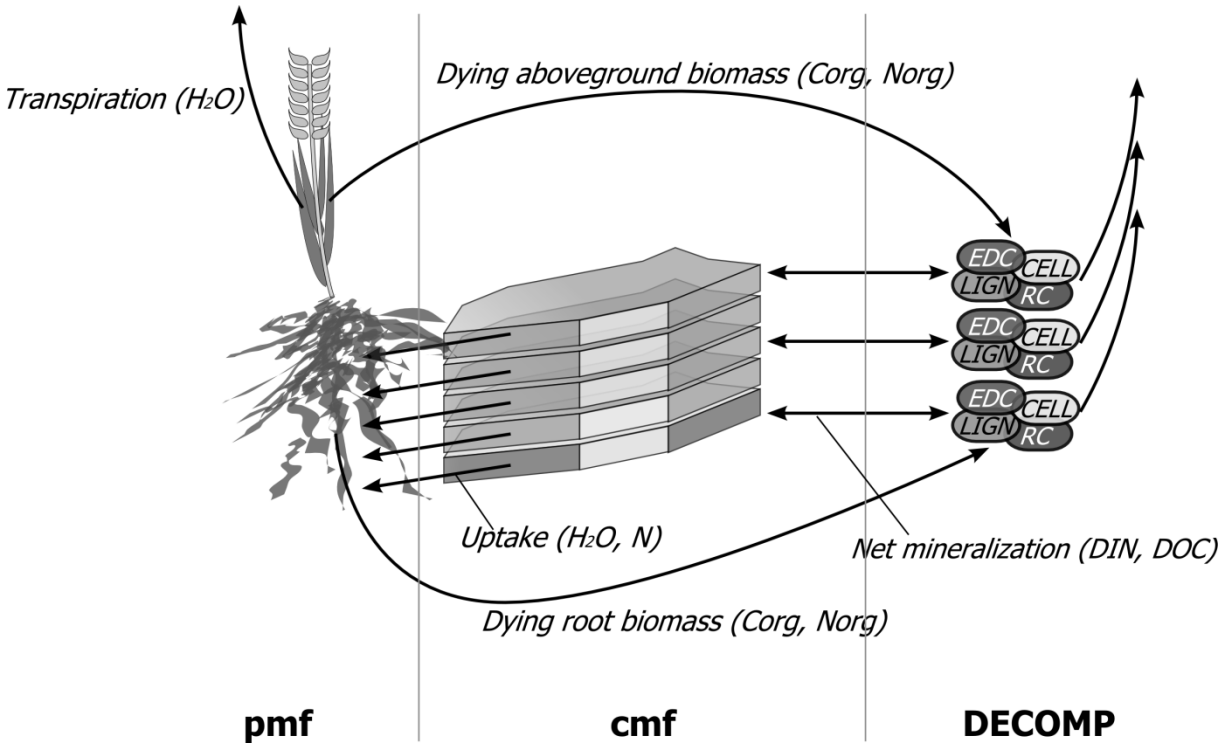

Figure IV-1: Matter fluxes in one cell of the integrated model. Matter fluxes across model domain boundaries are addressed integrated setup script, while matter fluxes between cells are calculated by the transport model CMF. EDC (easily decomposing compounds), CELL (cellulose like compounds), LIGN (lignin like compounds) and RC (recalcitrant compounds) denote the four organic matter compounds in DECOMP.

PMF, on the other hand, is specifically designed for different types of input of boundary conditions, namely nitrate concentration and soil moisture. PMF expects at setup time an interface providing the required data. These interfaces can route the data queries to a database with measured data, or to a model providing the requested data. A class wrapping a single cell of the CMF based transport model can be constructed, implementing the interface of PMF. Using this strategy, no direct copying of data between the models is required. The synchronization of water and solute fluxes between PMF, CMF and DECOMP, as shown in Figure IV-1 is rather implemented using the first strategy.

Input data, like meteorological time series, are imported from the setup script, using the advanced syntax of Python for text analyzes and partly hard coded into the setup script. The main time loop of the integrated model is part of the common setup script. The turnover models in this application run with a daily time step. However, the transport model uses a variable time step, which might be in the range of seconds during rainfall events. Therefore, the setup runs each of the models for one day,

taking one model time step for the turnover models while CMF is taking internally many time steps.

Results

To demonstrate the feedback loops, the model suite is run for 12 years using climatic data from the German meteorological service station "Giessen". Apart from the climatic data, the application of fertilizer is considered as an external driver. The slope is fertilized three times a year. During sowing in early spring (March 1st), 20 kg N/ha manure is applied, and shortly before germination (mid April) mineral fertilizer with 80 kg N/ha is given. A third application of 80kgN/ha mineral fertilizer is carried out during the shooting phase of the crops (end of May). The soil properties are assumed to be constant throughout the hillslope profile, and resemble a sandy soil.

Despite the simple model setup, the data produced leaves wide space for different interpretations. Due to the limitations of the chosen approach, as discussed below, it is not the objective of this study to show realistic model behaviour, but rather the potential how lateral transport influences model results. A single model time step, June 28, 1992 is chosen out of the simulated time series to explain possible effects of lateral nutrient transport on simulated crop growth in different parts of a virtual hill slope. At this time step, the shooting phase of the summer wheat is finished and nutrient storages in the rooting zone depleted. The date is located at the end of a four week period with less than 20 mm rainfall. Figure IV-2 shows the state of the hillslope near to the end of the simulation. The upper figure shows the distribution of soil moisture and plant water stress, while the lower figure displays the distribution of DIN and resulting plant stress. Bar length in both figures indicates the produced biomass in kg/m² and bar colours indicate the stress state of the plant: a green bar shows a plant where the growth is not hampered by lack of water (upper figure) or nitrogen in the soil solution, while dark red shows drastic reduced growth.

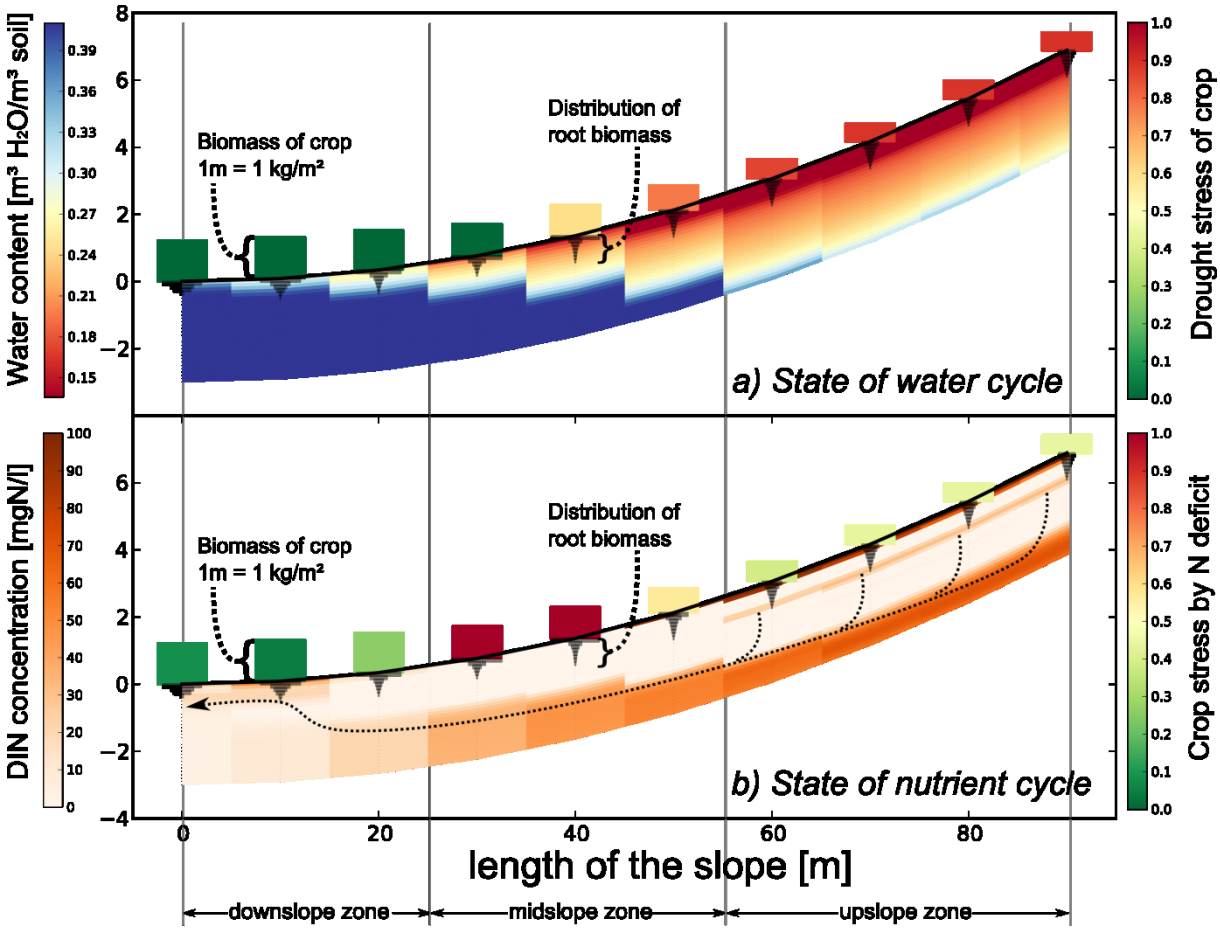

Figure IV-2: State of crops and soil water modelled by the integrated PMF-CMF-DECOMP system at 28.6.1992 after 12.5 years of integrated model run time. Each model cell has a size of 10 m x 10 m. Coloured rectangular bars indicate the crop state. The upper figure shows the spatially distribution of soil water content θ (ranging from 0.15 to 0.42 m^3/m^3, red to blue), and crop stress (colour of rectangles), ranging from 0 (no stress, green) to 1 (no growth due to draught, dark red). The soil cross-section in the lower figure is coloured according to the calculated DIN concentration in soil solution, ranging from 0(white) to 100 (brown) mgN/l. The dashed arrow shows the long term transport path for excessive DIN from the upslope zone to the ditch.

DIN is added to the transport system by application of fluid fertilizer and by mineralization of plant residuals (DECOMP) in the first layer and removed by plant uptake in the layers containing fine roots (PMF). The rooting depth changes over the vegetation period (PMF) from germination (uptake from first layer only) to a maximum of ca. 1 m (uptake from the upper 10 layers).

As shown in Figure IV-2, the production rate in the upslope zone of the hillslope (50 to 90 m from ditch) is limited due to drought. As a result, the fertilizer applied to this area is only partly taken up by the growing crops, while excessive DIN is leached from the rooting zone. The excessive DIN is then transported at the soil bedrock interface by lateral flows towards the saturated zone in the midslope zone of the hillslope

(30 to 50 m from ditch). Here, the higher DIN concentrations in the saturated zone are not accessible for the plant roots. Therefore plant growth is limited by DIN deficit, despite the high concentrations below the rooting zone. In the downslope zone (0 to 30 m from ditch), the saturated zone is accessible for the plants and the excessive DIN from the upslope zone can be taken up. The dashed arrow shows the long term transport of DIN through the hillslope. Since water from the saturated zone is ascending by capillary rise and accessible for the crop roots in the downslope zone of the hillslope, excessive DIN transported from the upper region becomes available to the crops of the lower region.

Discussion

The exemplary model setup shown above is not designed to be competitive for "real" problems. For operational use, the parameterization of DECOMP has to be validated for agricultural soils and a spin-up time of at least 100 years should be used. However, the goal of the application presented here is to show the effect of connecting typical plot scale models like plant growth models or biogeochemical soil models with a transport scheme. For "real" applications, sensitivity and uncertainty measures need to be taken into account.

In the exemplary setup shown above, the nutritional demand of a crop under water limited conditions affects the nutritional supply of crops in the riparian zone. This kind of spatial relationship cannot be captured with a classical one dimensional plant growth and nutrient turnover model, including a percolation model. Examples for this approach are PnET (Aber et al., 1997), EPIC (Williams et al., 1984), CENTURY (Parton et al., 1983) and WOFOST (Diepen et al., 2007). On the other hand, two dimensional Richards equation based models, such as CATFLOW (Maurer, 1997; Zehe et al., 2001) or HYDRUS 2D (Simunek et al., 1999) do neither include submodels to calculate nutritional uptake by plants, nor models for decomposition of the crop residuals after harvest. Although the combination of an existing crop growth model with an existing transport model in a single code base is feasible, the resulting integrated model might suffer a specific functionality for the next arisen use case. The most prominent example of a statically coupled distributed nutrient flux model is RHESSys (Band, Tague, Brun, et al., 2001) and influenced the

framework approach presented. However, since the models integrated into RHESSys were completely redeveloped, using alternative models is a time consuming task.

The second option for a tight coupling, the use of standardized model coupling interfaces (Lagarde et al., 2001; e. g. Gregersen et al., 2007) avoids this problem. Model codes to be used in this kind of coupling environment need to implement the interfaces of the coupling platforms, and can be coupled with each other. However, building an interface for a coupling platform is not trivial and therefore better suited for model codes unlikely to be changed. With a scripting language as coupling infrastructure, as promoted by Ousterhout (1998) and implemented in this study, a lightweight, flexible and less formalized interface can be used for exchanging data between models. Converting an existing model, like one of the mentioned well known plant growth models into a module of a scripting language is relatively simple, given automation tools such as SWIG (Beazley and Lomdahl, 1996) for model codes in C or C++ or F2PY (Peterson, 2009) for model codes in FORTRAN. In difference to coupling platforms, the user of a "pure" submodel integrated into a scripting language yields benefits, such as simplified testing or model run batching. However, the effort to implement an existing model as an extension to the Python language depends mainly on the quality and modularity of the code. Well structured and documented model codes might be wrapped in a few days. Other codes might not be suitable for integration into the Python language, due to their unstructured design or by inaccessibility of the source code. The main challenge for wrapping is the often limited possibility to disable specific process descriptions in models, covered by another coupled model already. For example most plant growth models include a more or less sophisticated model of percolation, which has to be disabled when coupled with a transport model like CMF.

However, since benefits exist in having a model as a Python extension, even for stand-alone applications and development, the authors wish to encourage model developers to create Python wrappers for their own models, and release the model codes to the public using an open source license. A multitude of models with a built-in facility of coupling even

during model development might establish a fast knowledge exchange path between different disciplines.

Appendix A: Data exchange between CMF and DECOMP

```
def RunDECOMP(dt, cmf_layer, DECOMP_SOM):
    # parameter exchange, DECOMP queries the wetness
    # and DOC concentration from cmf
    DECOMP_SOM.wetness = cmf_layer.wetness
    DECOMP_SOM[DOC] = cmf_layer.conc(DOC)
    # run DECOMP for timestep and return N and DOC release rate
    Nmin, newDOC, newCO2 = DECOMP_SOM.run(dt)
    # Convert kg/(ha day) to g/day
    Nmin_g_day = Nmin * cmf_layer.cell.Area / 1e4
    newDOC_g_day = newDOC * cmf_layer.cell.Area / 1e4
    # Add N and DOC release as tracer source in cmf
    cmf_layer.Solute(N).source += Nmin_g_day
    cmf_layer.Solute(N).source += newDOC_g_day
```

Acknowledgements

This study was funded by the European Union through the NitroEurope IP (www.nitroeurope.eu).

V. Towards a new approach to simulating regional N_2O emissions - the LandscapeDNDC Model

Edwin Haas, Steffen Klatt, Alexander Fröhlich, Philipp Kraft, Christian Werner, Ralf Kiese, Rüdiger Grote, Lutz Breuer, Klaus Butterbach-Bahl

This paper is submitted to the Journal Landscape Ecology. If accepted, the copyright will be transferred to *Springer Science + Business Media*

Philipp Kraft mainly contributed to the coupling issues, chapter 4

[1] Karlsruhe Institute of Technology, Institute for Meteorology and Climate Research, Kreuzeckbahnstr. 19, 82467 Garmisch-Partenkirchen, Germany

[2] Institute for Landscape Ecology and Resources Management (ILR), Justus-Liebig-University Giessen, Heinrich-Buff-Ring 26, 35392 Giessen, Germany

[3] Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, 60325 Frankfurt, Germany

Abstract

We present a new model system, which facilitates scaling of ecosystem processes from the plot to regional simulation domains. The new framework LandscapeDNDC is partly based on the widely used biogeochemical plot scale model DNDC, thereby inheriting a series of new features with regard to process descriptions, model structure and data I/O functionality. LandscapeDNDC incorporates different vegetation types and management systems for simulating carbon, nitrogen and water related biosphere-atmosphere-hydrosphere fluxes in forest, arable and grassland ecosystems. The modeling concept divides ecosystems into six substates (canopy air chemistry, microclimate, physiology, water cycle, vegetation structure and soil biogeochemistry) and provides alternative modules dealing with these substates. These modules account for model application across various ecosystem types, thus also allow simulation of dynamic land use transition.

The model can be applied on the plot scale, as well as for three-dimensional regional simulations operating either on raster (structured) or polygonal (unstructured) grids. The advantage of the regional LandscapeDNDC is that it integrates all grid cells synchronously forward

in time. This allows easy coupling to other spatial distributed models (e.g. for hydrology or atmospheric chemistry) and efficient two-way exchange of states The model can be coupled to any other model as a library or by direct communication with other models via parallel computer techniques.

This paper describes the fundamental design concept of the model and its object-oriented software implementation. An application demonstrates the data preprocessing of the regional model input data derived from GIS holding all relevant spatial information on soil properties, climate and agricultural management to build the input files for regionalization of nitrous oxide emissions from agricultural soils for the state of Saxony (Germany) in the year 2000. The computational effort for the LandscapeDNDC preprocessing and simulation could be speed up by a factor of almost 100 compared to the approach using the original DNDC version 9.3. The emission inventory simulated by LandscapeDNDNC was compared to results obtained with the original DNDC model, the IPCC methodology and the German National Inventory Report (NIR). While the IPCC Tier I methodology resulted in 1107 t N_2O–N/ha/a and the German NIR (IPCC Tier II) in 2100 t N_2O–N/ha/a, the process based simulations with LandscapeDNDC resulted in a regional source strength of agricultural soils of 2693 t N_2O–N/ha/a, which is very close to the value obtained with the original DNDC model of 2725 t N_2O–N/ha/a. By variation of input data a regional sensitivity study based on 32 different realizations of the regionalization revealed soil organic carbon and bulk density as well as the fertilization management to have highest impacts on the magnitude of N_2O emissions. To illustrate the capabilities of LandscapeDNDC for building a fully coupled model system on the landscape scale a first application of a coupled hydrology-biogeochemistry model for a virtual hillslope is presented.

Keywords: regionalization, greenhouse gas emissions, inventory, DNDC, LandscapeDNDC, model coupling

Introduction

Understanding human impacts on greenhouse gas (GHG) emissions from terrestrial ecosystems is vital for understanding climate change, mitigating emissions and developing adaptation strategies (Meinshausen et al. 2009). GHG emissions from soils such as CH_4 emissions from wetlands and rice paddies or N_2O emissions from arable soils and grasslands are mainly the result of microbial processes such as methanogenesis and methanotrophy (Cicerone and Shetter 1981; Wassmann et al. 1993), nitrification and/ or denitrification (Firestone and Davidson 1989). Due to the strong dependency of these microbial processes on environmental factors such as temperature, moisture, soil and vegetation properties and anthropogenic land management, emissions of GHG from soils exhibit a high degree of temporal and spatial variability (Blagodatsky et al. 2011; Butterbach-Bahl et al. 2004b; Del Grosso et al. 2010; Del Grosso et al. 2005a; Li et al. 2005). A reliable estimate of the regional source strength of soils for GHGs based on measurements would thus require a combination of different bottom-up (chamber measurements, eddy covariance flux measurements) and top-down approaches (remote sensing, tall-tower measurements of GHG fluxes) (Schulze et al. 2010) which is currently not available or still associated with large uncertainties. Therefore, many countries still apply the IPCC (Tier 1) methodology for calculating and reporting the national GHG sink and source strength of soils. In case of nitrous oxide this method simply assumes that a fixed proportion (emission factor, EF) of the applied nitrogen fertilizer is emitted as N_2O. In general EF for GHG emissions are addressed with high uncertainties since they are still based on a limited number of field observations (IPCC 2007) and represent more a global average rather than being representative for countries or even smaller regions.

The EF approach has obvious short-comes since temporal and spatial variations of emissions cannot be delineated, so that regional emission hotspots due to soil or climatic conditions or hot moments – e.g. due to nitrogen fertilizer applications – can hardly be covered. Moreover, IPCC Tier 1 methodology does not allow the development of region and site-specific mitigation strategies resulting in a need for more detailed

approaches to characterize spatial and temporal patterns of ecosystem GHG exchange (Butterbach-Bahl et al. 2004a; Butterbach-Bahl et al. 2004c; Del Grosso et al. 2005a; Li et al. 2005; Smith et al. 2008). For the given reasons the application of more mechanistic approaches as forseen in IPCC Tier 3 such as application of process-based models is highly desirable.

Recent studies on N_2O emissions from soils with process-based biogeochemical models like DNDC (Beheydt et al. 2007; Blagodatsky et al. 2011; Li et al. 1992; Werner et al. 2007), FASSET (Chatskikh et al. 2005; Chirinda et al. 2010), DAYCENT (Del Grosso et al. 2010; Del Grosso et al. 2005b), CERES-EGC (Gabrielle et al. 2006; Lehuger et al. 2009), Expert-N (Kaharabata et al. 2003) or COUP (Norman et al. 2008) have shown that these models allow simulating of soil based GHG emissions for a range of terrestrial ecosystem types at site scale as well as on regional scale.

Even though the mentioned models have been used on regional and national levels for calculating N-trace gas emission inventories there are still technical challenges related to generating, handling and assessing of input/ output data. In principal the models are designed for site-specific applications such that regional applications are still based on single runs for each individual spatial unit. Moreover, all existing models used for calculating regional/ national GHG inventories are of one-dimensional character, thereby neglecting lateral matter exchange with adjacent simulation units e.g. driven by topographical differences. However, lateral fluxes may significantly affect carbon nitrogen and water cycles of the simulated ecosystem, e.g. biosphere-atmosphere exchange of riparian zones are largely depending on water and nutrient input from the surrounding landscape. However, most of the currently available models used for simulating biosphere-atmosphere exchange of GHG emissions are structurally not capable to consider landscape fluxes and the importance of regional hydrology or nutrient dispersion for site specific GHG fluxes.

So far consideration of lateral fluxes relies generally on makeshift solutions. For example Cui et al. (2005) described a linkage of the Mike She watershed model to the Wetland-DNDC model for simulating C and N dynamics and GHG emissions in forested wetlands. Due to the

shortcomings in the software structure of Wetland-DNDC they could only link water table dynamics generated by Mike She to one/ several site scale simulations of Wetland-DNDC. A technology of synchronizing the two models including feedbacks like e.g. nutrient transport – and therefore a fully two-way coupling – between Wetland-DNDC and Mike She could not be realized.

Other approaches do not couple different models but are complementing existing models with new features. For example Pohlert et al. (2007b) incorporated biogeochemical processes into the hydro-biogeochemistry model SWAT and successfully applied it to predict nitrate leaching and transport on catchment scale (Pohlert et al. 2007a). Such approaches use a simpler description of biogeochemical processes as compared to process-based models like DNDC and they lack the capability of computing and compiling regional soil GHG emission inventories.

The far most advanced integrated ecosystem model for the assessment of landscape scale fluxes is the ECOSYS model (Grant et al. 1993; Grant and Pattey 2003). The model incorporates soil hydrology, crop growth and biogeochemical C and N cycling. However, the complexity of the model results in highly demanding parameterizations and initializations and therefore limits its applicability, in particular on regional/ national scales.

In this study, we present the design concept of a recently developed model system for the simulation of ecosystems processes from leaf to plot and regional scale. It follows the design concepts outlined in Boyer et al. (2006). The model framework takes advantage of a modular design allowing the use of modules to simulate different ecosystem processes and to facilitate the exchange of data with regional models (e.g. hydrological or atmospheric airchemistry models) in order to consider the importance of landscape fluxes for actual site simulations. The model framework of LandscapeDNDC is suitable for:

I. Applications on site and regional scales to facilitate an efficient calculation and compilation of regional/ national GHG inventories.

II. Coupling it to other regional models to allow the consideration of lateral nutrient transport and feedbacks between different ecosystem compartments.

III. Consistent simulation of soil- and vegetation processes during land-use change (LUC).

In the following section we will outline our objectives driving the development of LandscapeDNDC and some of the challenges in regional ecosystem modelling from a more theoretical point of view. This will be followed by an introduction to the model concept concentrating on software engineering issues as well as its implementation in an object oriented way. The model performance will be discussed with a focus on computational efficiency. We will present a regional application of LandscapeDNDC demonstrating its capability to efficiently compile regional greenhouse gas emission inventories including a regional sensitivity analysis. Finally, we show results of a fully coupled model application of LandscapeDNDC and a physically based hydrological model.

Demands for landscape modelling of biosphere-atmosphere-hydrosphere processes

Land Use Change (LUC)

The motivation of designing a new ecosystem model framework is strongly driven by the challenges in simulating land use change (LUC) (IPCC 2000) as a major driver of increased atmospheric CO_2 concentrations (Houghton and Hackler 1999; IPCC 2000; Koomen et al. 2008) and as potential strategy to sequester atmospheric CO_2 in terrestrial ecosystems (Conant 2011; Smith et al. 2008). A suitable biogeochemical model for assessing the effects of LUC on soil and ecosystem C and N stocks and turnover needs to consider biogeochemical processes that are applicable to different ecosystem types such as forests, grassland and arable land. This request is fundamental for simulating the transition of different ecosystems due to LUC and will result in a unique feature as no process-based ecosystem model of higher

complexity is so far designed to simulate multi-ecosystems with an identical parameterization of the biogeochemistry.

Within LandscapeDNDC we have unified the soil biogeochemical processes of the agricultural DNDC (Li et al. 1992) and of the Forest-DNDC (Kesik et al. 2005; Kiese et al. 2011; Li et al. 2000; Stange et al. 2000) into a general soil biogeochemistry module. This module is capable to simulate ecosystem C and N turnover and changes in soil C and N stocks for various land use (LU) types and allows simulating the dynamic C and N cycling during periods of land use change (LUC).

The new approach groups the physiological functioning of the different LU types into exchangeable modules resulting in a DNDC-based physiology module for agricultural crop growth including grassland (Li et al. 1992; Wolf et al. 2010) and a PnET-based forest growth module (Aber and Federer 1992; Li et al. 2000; Miehle et al. 2010). This allows describing LUC in a transient way since only respective physiology modules are switched and no re-initialization of soil properties is necessary. The design also allows an easy incorporation of future/ new physiological modules e.g. for wetlands, paddy rice systems or tropical savannas.

Regional applications

Recent studies (Butterbach-Bahl et al. 2009; Gabrielle et al. 2006; Kesik et al. 2005; Rolland et al. 2010; Werner et al. 2007) in which biogeochemical models have been used to assess the regional or global source and sink strengths of soils for GHG are based on ecosystem simulations using large domains with high spatial disaggregation. For example Werner et al. (2007) based their global N_2O emission inventory for tropical rainforest soils on individual calculations for approximately 40,000 grid cells, which were simulated several times for assessing sensitivity or the effect of input uncertainties on simulated fluxes. All individual model runs were associated with I/O operations and the creation of individual files. Another example can be found in Kiese et al. (2011) operating more than 10,000 grid cells for regionalisation of nitrate leaching rates of forest soils in Germany. Following such an approach, pre- and post processing

of data and simulation results are rather complex, unefficient and are introducing a high risk for errors. Moreover, such approaches are computational unefficient that even modern high performance computing (HPC) systems can get to their limits due to the large amount of single files and I/O operations involved.

To overcome this computational as well as conceptual problem we defined that the new model framework should incorporate the following features:

Site- and regional scale model

The model system has to be grid based (structured or unstructured) representing many ecosystem units (grid cells) within one software instance allowing an efficient application on regional/ national scale. In order to implement a professional computational infrastructure for regional model simulations, alternatives to the use of single files have to be invented. The input data has to be handled in a condensed form in regional (gridded) input files (i.e. XML or HDF5 files) or will be supplied via a link to a database system. Input generation is handled by preprocessor programs that access GIS databases holding the spatially distributed input data. The new model framework can also be used for site scale simulations when running on one single grid cell.

Parallelization

The model framework needs to allow the efficient use of hybrid and heterogeneous computing environments such as multicore workstations via OpenMP and distributed HPC systems utilizing MPI parallelization.

Model core concept

The underlying process descriptions should apply to site and regional model runs alike.

Library concept

The design has to enable the compilation of the complete model into a library, which allows creating a model instance from any other software system loading this library. It features functionalities, e.g. building a test

bed for model development when integrated into a high-level, interpreter-based programming environment such as Python.

Coupling the model framework to other regional models

So far most regional modelling studies dealing with the emission of GHG from soils (Butterbach-Bahl et al. 2009; Gabrielle et al. 2006; Kesik et al. 2005; Rolland et al. 2010; Werner et al. 2007) have traditionally focused on the simulation of site specific ecosystem fluxes, thereby neglecting the potential importance of landscape fluxes of water and nutrients for GHG emissions e.g. downwind or downstream the site of nitrogen fertilizer application.

Open key questions with regards to C and N cycling and feedbacks concern a) the dominant processes (microbial/plant versus physico-chemical) determining residence times, transport and emissions/ deposition of nitrogen at different temporal and spatial scales (Skiba et al. 2009), b) the spatial distribution and magnitude of hotspots and hot moments of C and N losses at the landscape scale and their importance for annual and/ or regional nutrient balances (Butterbach-Bahl and Dannenmann 2011) and c) the pathways of C and N losses to the atmosphere (VOCs, CO_2, NH_3, N_2O / NO or N_2) and hydrosphere (DOC, DON, NO_3, NH_4, particle bound C an N) and the redistribution of nutrients in the landscape (Sutton et al. 2007).

Addressing these questions in a thorough manner requires the implementation of a coupled biosphere-hydrosphere modelling system capable to simulate processes and fluxes on the catchment / landscape scale. Such a system would allow to improve the understanding of the fate of C and N within the landscape and to develop strategies to enhance C and N use efficiency e.g. in agriculture. Approaches in this direction are currently stuck on a very small scale that is not appropriate for comprehensive landscape analyses (Roering et al. 2004; Schmitter et al. 2010).

The desired environmental modelling tool should be suitable to assess environmental impacts of global change (climate, land us and management) not only on site, but also on the even more relevant landscape scale. How can this coupling be done? Modern model coupling

tools like the PalmMP model coupler of CERFACS (Fouilloux et al. 1999) enable to couple different model codes (even written in different programming languages) on basis of their respective source code to form one new coupled model. Using the multiple programs and multiple data paradigm of MPI parallel programming, PalmMP allows compiling different model codes together with the PalmMP coupler into one logical MPI application consisting of multiple coupled parts i.e. the coupled models and additional process control counterparts. The exchange of data between the coupled parts will utilize PalmMP *put* and *get* functions (corresponding to MPI send and receive), which have to be implemented into the source code of the coupled models among other functionalities for process control and synchronization.

To finally meet the requirements for the construction of a coupled biosphere-hydrosphere-atmosphere modelling system we designed our model framework in such a way that it allows:

- Synchronous time integration of all grid cells

  The time integration for all grid cells has to be performed synchronously forward in time. This enables the exchange of the system's state for all grid cells (the regional state) for a distinct model time (e.g. daily or hourly data exchange) between the component models of the coupled modelling system.
- Library concept

  PalmMP will replace the coupled models main routine by a function, which will be called by the PalmMP master application when running in coupled mode. Therefore the models have to be compiled into a library, which will be included in the PalmMP coupled application.
- Parallelization

  Modern coupling tools – such as the PalmMP Coupler – take advantage of parallelized models when building a coupled system.

The Library concept also allows an alternative approach when dynamically coupling models to test the coupled system, for example by importing the code from an interpreter system like Python. The python shell allows interactive observation and manipulation of all public

variables (states) of any of the model instances, especially when biogeochemical and hydrologic models are coupled. With this functionality it forms an efficient tool to debug the model system and investigate its behaviour. Building a time integration loop and integrating the models forward in time accompanied by a suitable data exchange will form a coupled modelling system. The advantage of this straightforward approach is its simplicity with regard to the implementation of a coupled prototype. However, this approach lacks computational efficiency and it cannot take advantage of the parallelization of the individual models within Python. Therefore, in order to establish a landscape/regional scale coupled system the use of a parallel model coupler like PalmMP is without any alternative.

Model concept and software design

Model concept and implementation

The regional model system constitutes a framework that performs simultaneous single site simulations by iterating over a set of single site submodels (cells) and integrating them forward in time. Such a submodel is a one-dimensional vertical column model for simulating ecosystem processes on the plot scale. All submodels in a regional simulation mode are integrated synchronously forward in time such that all grid cells will have the same simulation time. This is a fundamental difference to regional application of DNDC and CERES, as these models compute the full model timespan for one grid cell and then proceed to the next cell. However, a synchronized timestep for the full simulation domain as implemented in LandscapeDNDC is essential to allow two-way coupling of LandscapeDNDC to other regional (hydrological or atmospheric) models.

The presented approach of LandscapeDNDC enables the validation of the regional model by evaluating it in site mode at sites for which observations are available.

At this stage there is no communication in LandscapeDNDC between the cells when running in regional mode. We emphasize that this is perfectly suited for parallelization.

In principle, *modules* (mathematical description of biological, physical and biogeochemical processes) act on subsets of the representation of the ecosystem state. The behaviour of the modules as well as the initial state of the system are controlled by site specific parameters (soil texture, carbon content etc.) and boundary conditions (air temperature, precipitation etc.).

LandscapeDNDC is implemented in Ansi C++, designed to run on Unix platforms and MS Windows. It also compiles with recent versions of the most common compilers (gnu gcc, intel icc, portland pgcc, mircosoft vsc++). All components are implemented as classes (see UML diagram in Fig V-1). The software architecture is given in Fig. V-2. The size of the simulation, i.e. the number of cells is determined by the input data. After the I/O subsystem successfully processed the input data a *model setup* is created by instantiating an object of the LandscapeDNDC class by handing it over a *set of input*. Each input set (drivers and parameters of different type) corresponds to a cell and the information is used to initialize the LandscapeDNDC class object. Modules are set up after state initialization hence the state is available to modules as well.

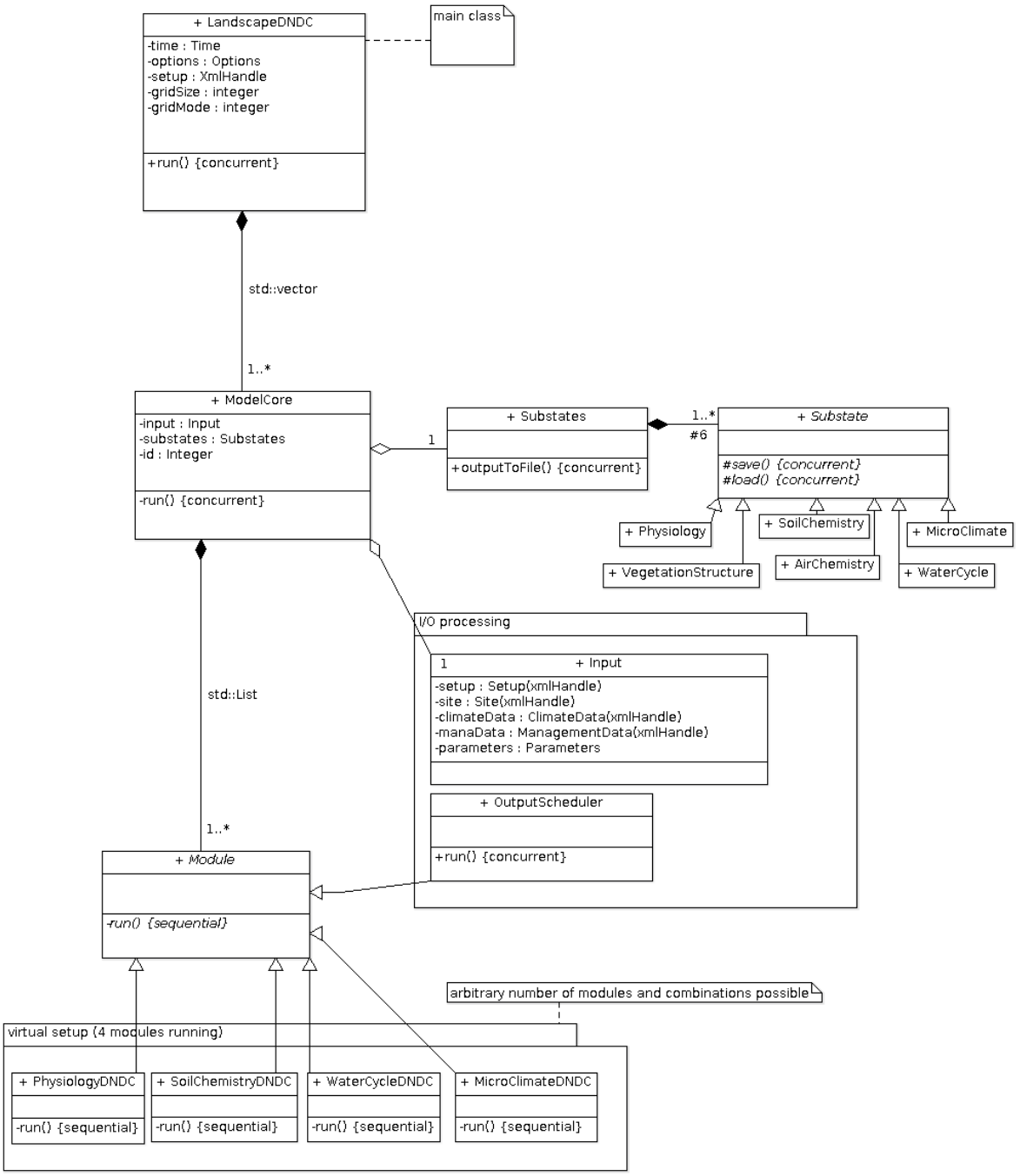

Figure V-1: UML diagram of the object structure of LandscapeDNDC.

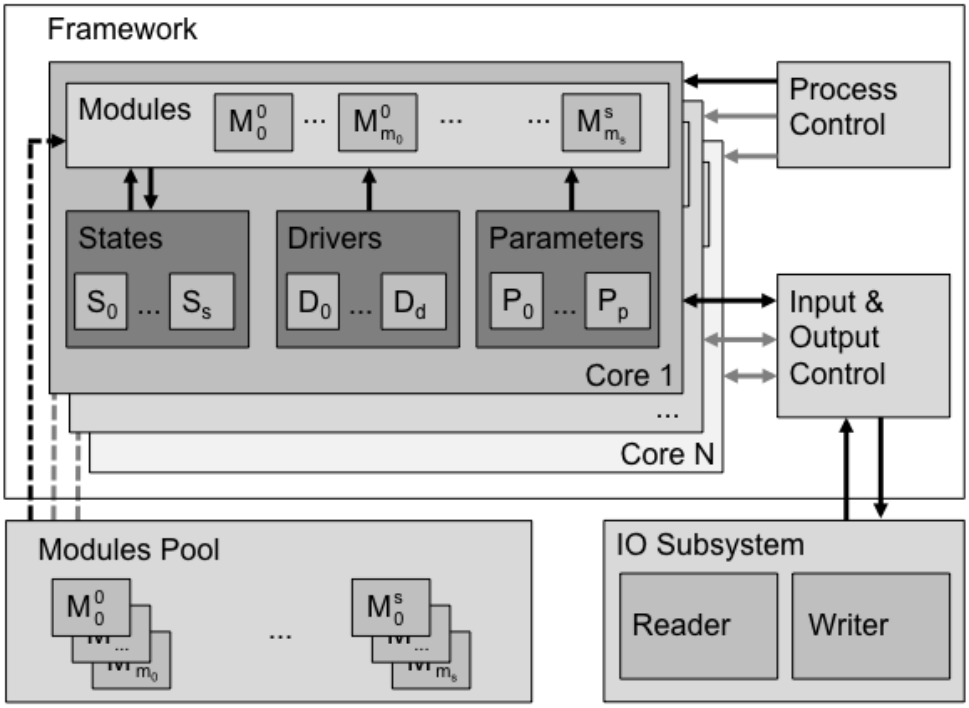

Figure V-2: Schematic sketch of the LandscapeDNDC model framework.

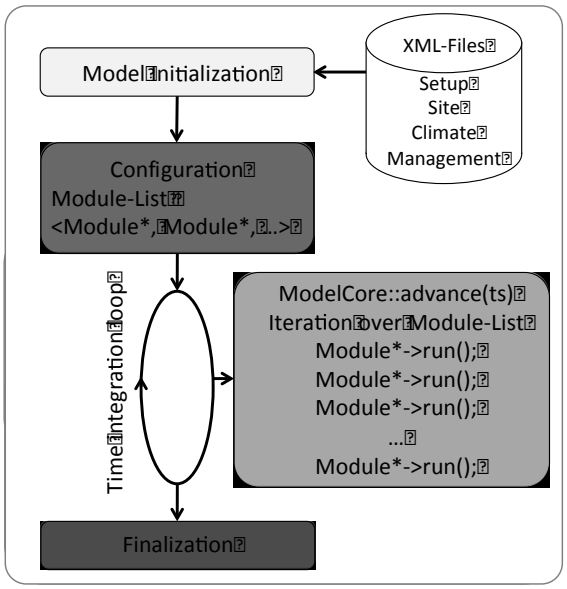

Figure V-3: Data flow scheme of LandscapeDNDC in site mode.

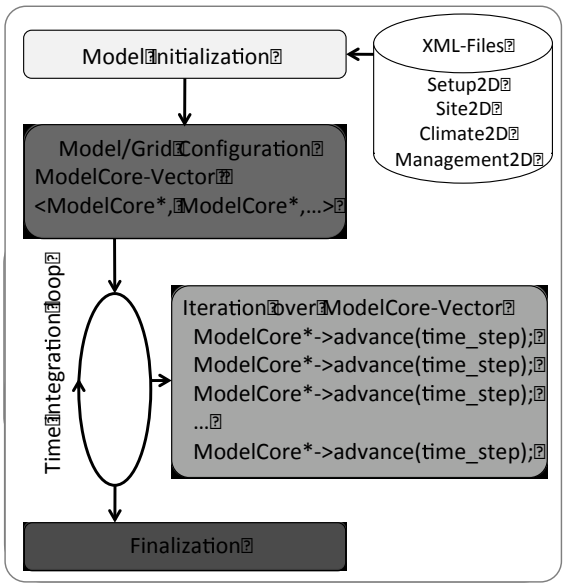

Figure V-4: Data flow scheme of LandscapeDNDC in regional mode.

After successful initialization the framework will perform the time integration loop (see Fig. V-4). Within the time loop the framework calls each cell in turn to advance by one time step. During execution of a cell, its modules are performed once in a sequential fashion (see Fig. V-3) by iterating over the module list. The module list will be created automatically during the initialization of the model core and allows an efficient execution of the modules. If the setup does not contain LUC actions, the module list of each core will not change during the simulation.

At occurance of LUC events the module list may be altered accordingly to represent the new land use. As the ecosystem state is being kept in the substates, the change in modules (by deleting or creating module instances) will not influence any state variable. As the LUC transition will be performed when continuing with the time integration, the new module configuration will act on the substates and therefore transform the ecosystem state – represented in the substates - towards the new LU.

State and substates

The state is an abstract representation of the ecosystem of a single cell. Because of lack of communication between cells we omit to define a combined regional state. A substate describes a subset of the state. Currently LandscapeDNDC defines six distinct substates that emerge from a natural decomposition of an ecosystem: canopy air chemistry, canopy and soil microclimate, crop/ vegetation physiology, vegetation structure (forests only), water cycle and finally soil biogeochemistry which all have been derived from previous work (Grote et al. 2011; Grote et al. 2009; Kiese et al. 2011; Li 2000; Li et al. 1992; Stange et al. 2000).

The states can be saved to, and restored from, a XML file. This can also be used to provide initial state conditions from a prerun to the model in order to decrease the overall simulation effort e.g., scenario or sensitivity analysis with common spinup conditions.

Drivers

External data used to set the initial state and driving the time integration are called drivers. Typical model drivers are climatic and chemical

boundary conditions, e.g.air temperature, precipitation, solar radiation, gas concentrations in ambient air (e.g. CO, CO_2, CH_4, N_2O, O_2, NO, NO_2), but also agricultural and forest management actions (e.g., fertilizing, manuring, tilling, irrigating, flooding, cutting and grazing).

Input and Output functionality

Input (regional and site scale) is read prior to instantiating the simulation framework. This is accomplished by the I/O subsystem and motivated by the demand to feed input not only to LandscapeDNDC but also to converters, data synthesizers, or validators. Each driver and parameter set is read from a separate file while each file holds data for all cores. The top-level input file (*setup*) contains global configuration settings such as the number of cells, general time parameters (e.g. start date, duration) and names of additional input files (e.g. site description, climate, agricultural/ forest management). For each grid cell the module setup will be defined in the setup file as well. To assist input preparation, especially for large regional simulations, a Python preprocessing toolbox is provided. Theoretically, any input format can be used with the toolbox. LandscapeDNDC supports currently, XML, HDF5, netCDF and plain text files.

The model uses data structures to collect output during the time integration in order to avoid excessive I/O operations and to facilitate different output formats. The data is stored temporarily and flushed to the output interface on an adjustable time-step. The standard output interface will write data according to an output template (controlling the output variable selection) into text files (site mode). In regional mode the physical output supports plain text or HDF5 file formats in order solve access problems originating from parallelization.

Parameters

Process-based ecosystem modelling relies on detailed process description and thus on a large number of process parameters (e.g., diffusion coefficients, biological process rate constants, kinetic velocities, plant and microbial growth parameters etc.). Parameters are considered constant during a simulation run. However, some parameters are specific to a given land-use and soil type, and thus can vary in the domain such that

each core might employ its own set of *local* parameters. This excludes *global* parameters that apply to all cores. An example of such a parameter is e.g. the water use efficiency of a given vegetation type or the maximum yield which can be expected for a given crop variety.

The model framework provides local parameters as default values but they can be overwritten by reading them in from an external parameter file.

Modules

Modules describe the biological, physical and geochemical processes in an ecosystem. By definition a module is assigned to only one substate that it can modify while having read-only access to the other states.

In this respect, a module is a specific realization of given substate process. The ability to exchange and combine modules offers a way of comparing different substate descriptions that might provide different levels of complexity, concepts or numerical implementation under consistent boundary conditions. This concept has been introduced in previsous work by (Boyer et al. 2006) and (Holst et al. 2010). While technically not being part of the model framework they are managed and controlled (e.g. invoked) by it (see Fig. 2). From a pool of modules the user can make a module selection that best fits the environment/ ecosystem on a per cell basis.

The available modules within LandscapeDNDC are based on the DNDC model which was first published by (Li et al. 1992) (DNDC agricultural plant growth, soil biogeochemistry & water cycle functionalities) and Forest-DNDC (Li et al. 2000; Stange et al. 2000). For forest biomass development, i.e. in structured forests, the Physiological Simulation Model PSIM (Grote 2007) is available as an alternative to the PNET-N physiology module of the Forest-DNDC and forest dimensional growth can be calculated based on wood carbon growth in a separate module (Grote et al. 2011).

In general, the integration of functionality from an existing model into LandscapeDNDC as one single module (such as the PSIM physiology) or multiple modules (such as the DNDC-plant growth, DNDC-soil-

biogeochemistry & DNDC-watercycle) generally requires code adaptation. Porting a model into LandscapeDNDC that is implemented in C/C++ will in the most simple case only require adaptations to the frameworks variable names, while it is likely that a recoding is necessary when it is coded in a different programming language.

Management action

The model system features the simulation of i) agricultural management as well as it enables the study of ii) LUC scenarios resulting from changes in LU and/ or LU management on C, N and water fluxes at site and landscape scale.

The most common management actions in LandscapeDNDC are DNDC-like agricultural management events like cropping, harvesting, tilling, fertilizing, manuring, flooding, irrigating, grazing and cutting (Li et al. 1992). All management actions or events are defined in the management input as xml data structures.

For management actions that involve LUC, the vegetation (physiology) module in LandscapeDNDC may be exchanged during the simulation to account for example for a re-forestation of former grassland or a deforestation of a site with a follow-up use as grassland or arable land. LUC was previously simulated (for site-scale applications) by writing the complete system state into a xml file before the LU transition and restarting the model using this dump together with a changed configuration representing an alternative ecosystem type. For simulations of LUC on the regional scale, with incorporating of a huge number of grid cells, this approach is not adequate. Therefore LUC simulations in LandscapeDNDC are based on a dynamic change of the used physiology module during the simulation to account for the ecosystem change. All bio-geochemical states such as soil moisture, temperature, C & N stocks or other soil characteristics are not changed. In the current version soil properties like bulk density or the saturated hydraulic conductivity are fixed and therefore not adapted during the LUC simulation whereas in reality they might adapt due to the LUC.

Regional and Landscape Scale Applications

To prove the ability of LandscapeDNDC for the compilation of regional GHG emission inventories the model framework was applied to simulate the N_2O source strength of agricultural soils for the state of Saxony, Germany. The second test case is an coupled application of LandscapeDNDC and the hydrological CMF model on landscape scale which shows simulation results of C & N cycling and transport on a virtual hillslope.

Regional N_2O inventory and sensitivity analysis

Compiling regional GHG emission inventories from agricultural land requires the decomposition of the agricultural land into simulation units (grid cells) representing homogeneous soil, crop and climate conditions. In this study we used a polygonal grid with 661 polygones. Since several different crops are grown in a polygone, multiple simulation runs were performed. With respect to computational performance we have limited the maximum number of possible crops to 10 major types as described below. Each cropping system was simulated in regional mode. The first regional run simulates crop type 1 (e.g. corn) on all polygones, the second run crop type 2 (e.g wheat) and so on. The results of the 10 regional simulations were aggregated onto the original polygones according to the crop cover fraction given by statistics (data source: LfULG Saxony), with the weighted average flux being represented in the final inventory.

Input data

Region specific input data was supplied by the environmental and geological services of the state of Saxony (LfULG). The LU for the simulation domain was derived from the CORINE database (published by the Commission of the European Communities). The study presented accounts for classic arable and grassland systems. Soil data was used from the German soil survey map BUEK1000 (resolution 1:1.000.000) of the “Federal Institute for Geosciences and Natural Resources (BGR)”, Braunschweig. The BUEK1000 dataset delivers region specific soil attributes such as organic carbon content, soil pH, clay / sand / silt content and bulk density in different soil strata (upper and lower soil horizon). Ten major crop types were taken into account for modelling the

agricultural systems of the simulation domain: winter wheat, spring barley, winter barley, rye, oat, silage maize, potatoes, sugar beets, rape seeds and non legume perennial grass, as proposed by the LfULG Saxony. The climate data was obtained from DWD station observations provided by LfULG. The dataset includes daily temperature (max, min, mean) and precipitation. In the GIS preprocessing each polygon was assigned to the nearest station and this information was used during preprocessing to build the connection of the grid cell and the climate input.

The fertilization and manuring input data was derived from district statistics supplied by LfULG as no high resolution data was available. Therefore all grid cells within the districts have received the same amounts of fertilizers. The amount and distribution of the manure application was derived from statistics on life stock per district supplied by LfULG. The manure distribution was based on the crop N-demand. The mineral fertilizer application rate was derived from statistics on sold fertilizer per district per year and the distribution was again based on the crop N-demand. The timing of fertilizer and manure applications was based on recommendations of local agricultural advisers. The distribution of the total fertilizer applied is shown in Fig. V-7.

Table V-1: Optimum yields and residues left on the field after harvest for simulated crops.

DNDC Crop	Maximum yield kg C / ha	Residuals %
Cereals		15
Winter wheat	6 857	15.0
Rye	4 940	15.0
Winter barley	5 143	15.0
Spring barly	4 000	15.0
Oats	3 333	15.0
Corn	10 667	30.0
Root crops		30.0
Potato	5 444	30.0
Beets	8 888	30.0
Animal feed		
Non-legume hay	7 500	100.0
Commercial crops		
Sun flowers	5867	15.0
Rapeseed	2 400	15.0

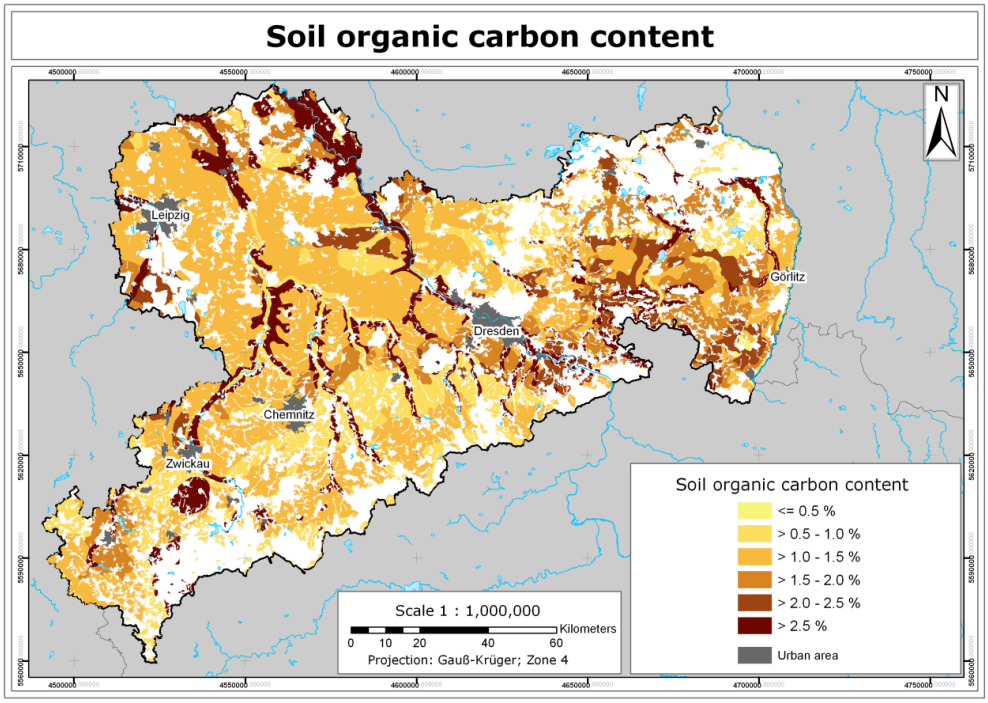

Figure V-5: Soil organic carbon (SOC) in agricultural soils of Saxony according to the BUEK1000 soil map.

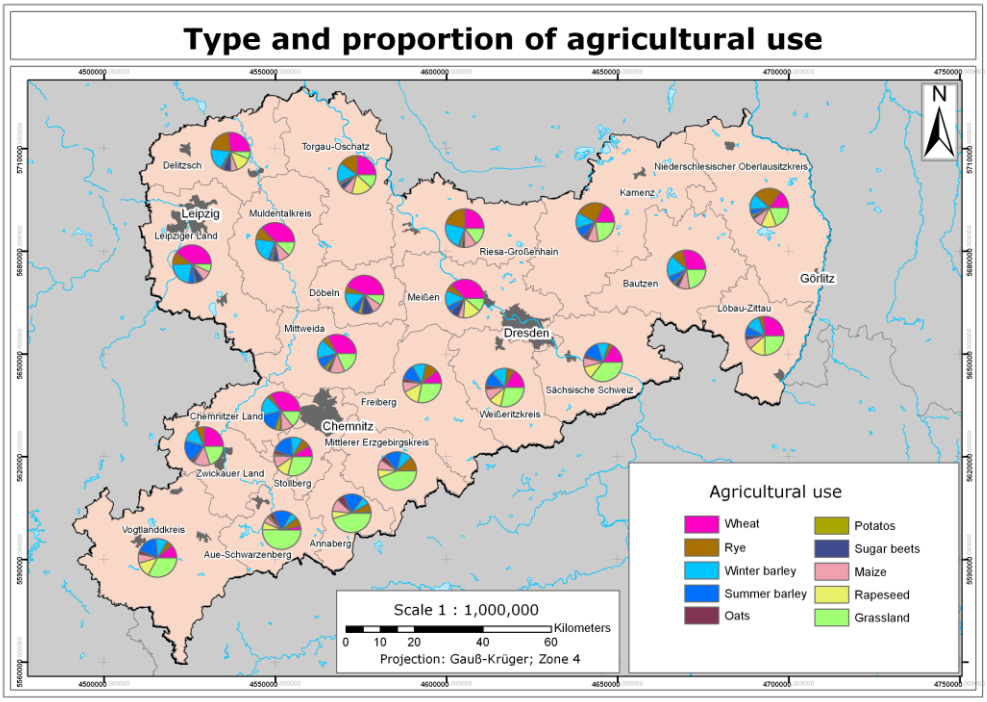

Figure V-6: Main crops cultivated in the administrative districts of Saxony in the year 2000 (data source: LfULG Saxony).

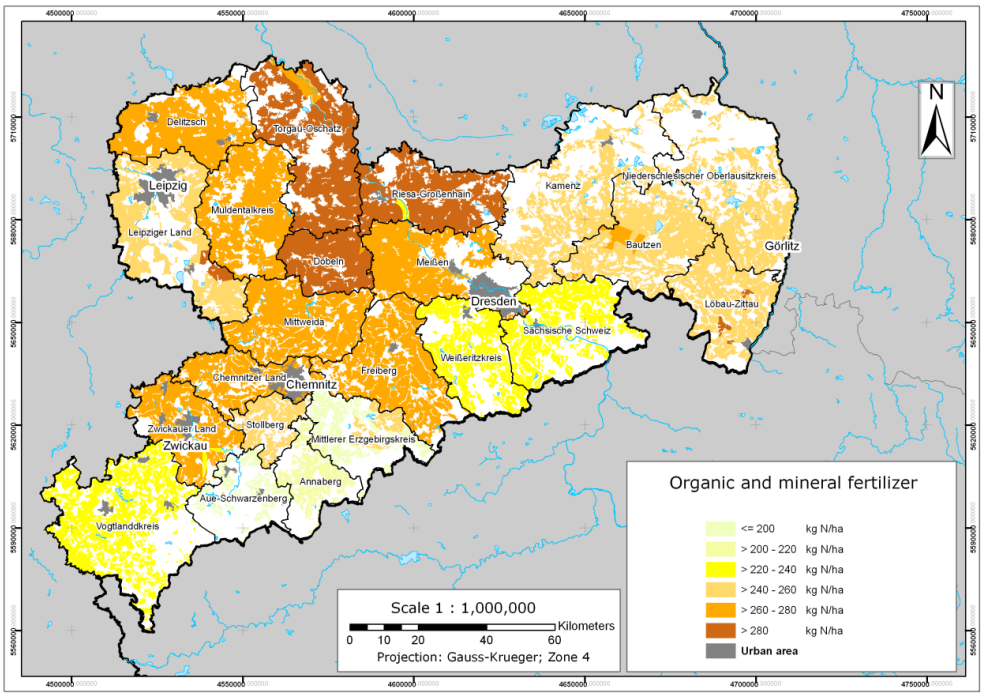

Figure V-7: Total inorganic and organic nitrogen fertilizer use for different districts in Saxony (data source: LfULG Saxony)

The generation of model input data for agricultural management and practice is an essential step when creating the input database for a regional inventory. Besides the lack of detailed data (Dämmgen and Grünhage 2002) a further drawback is the poor spatial resolution of available datasets, which mostly originated from farming statistics (crop distribution and yield, number of farms, head numbers of life stock) on district level. Important parameters affecting ecosystem N cycling and associated N trace gas emissions are rates and timing of fertilizer applications (mineral and organic), crop specific parameters such as optimal biomass yield and residues left on the field after harvest (see Table V-1), vegetation phenology, nitrogen efficiency, and other information on agricultural practices such as timing of seeding, harvest or tilling. Since agricultural practices may change from farm to farm we mostly relied on expert knowledge and statistics at district level. Since a regional distribution of the crops cultivated was not available on high resolution it was needed to use data on district basis, thereby assuming that crops were distributed homogeneous across a given district.

From Fig. V-6 it is obvious that management practice differs from north to south. grassland dominates in the hilly southern areas, whereas wheat is the pre-dominant crop in the flat terrain of northern Saxony. This pattern reflects the climatological and geomorphological differences of Saxony, as the southern region is mountainous (the Ore Mountains) and therefore more privileged for grassland than for arable cropping systems.

The crop distribution was taken from the land use statistics for 2000 (Bodennutzungshaupterhebung supplied by the Federal Statistical Office of Germany). Differences between the available model crops and the cultivated crops of the LU statistics have been overcome by lumping the different cultivated crops onto the 10 main crops realized in LandscapeDNDC.

For our test case we assumed a simplified crop rotation. It consists of a one year pre-run followed by the same crop for the second year (the year of interest) to avoid effects of fallow soil. For example, winter wheat will be initialized at the first of January, harvested in summer and resown in autumn in the prerun year such that on the first of January in the second

year the plant is already growing on the field and has not been initialized. It will be harvested in the second year and resown in autumn.

N_2O emissions based on IPCC methodology (Tier 1)

The IPCC2 defined in their last report "IPCC Guidelines for National Greenhouse Gas Inventories Volume 4 Agriculture, Forestry and Other Land Use" the most recent procedure to estimate the direct N_2O emission strength from agricultural soils for signatory states. The methodology estimates soil N_2O emissions at a site of N-fertilizer application (direct emissions) as a percentage loss of the applied N-fertilizer. The so called EF is assumed to be 0.01 kg N_2O-N per kg N applied as mineral fertilizer or manure with an uncertainty range from 0.002 – 0.018 kg N_2O-N per kg N applied.

Total N-fertilzer use in Saxony is 110 709 t-N (19 526 t-N manure and 91 183 t-N mineral fertilizer). Therefore, direct soil N_2O emissions due to N-fertilization in Saxony amount to 1107 t N_2O-N if the IPCC Tier 1 approach is used. The uncertainty range is from 221 up to 1993 t N_2O-N. This is equivalent to an average emission rate of 1.23 kg N_2O–N/ha/a for the 902 114 ha of agricultural land with the uncertainty of 0.25 to 2.21 kg N_2O–N/ha/a.

N_2O emissions based on the German national emission methodology NIR (IPCC Tier 2)

The German national trace gas emission methodology (NIR) was recently outlined by (Dämmgen and Grünhage 2002). Among others, it supplies detailed information about trace gas emissions from agricultural soils for all federal states of Germany. Even though the NIR methodology is also empirical it is far more detailed and uses national based emission factors in comparison to the much corser IPCC methodology (see Table V-2).

Following the German National Emission Inventory (Dämmgen et al. 2007) agricultural used soils in Saxony are emitting 3050 t N_2O–N in the year 2000 which results in an average emission rate of 3.38 kg N_2O–N/ha/a. The numbers include 2100 t N_2O–N direct emissions from fertilization resulting in 2.33 kg N_2O–N/ha/a. Table V-2 includes indirect emissions from grazing, residues and deposition of reactive nitrogen. These emissions have been accounted for, as LandscapeDNDC will incorporate these processes as well.

N_2O inventory calculated by LandscapeDNDC

According to LandscapeDNDC simulations 2693 t N_2O-N were emitted in 2000 in Saxony from agricultural soils. Highest averaged N_2O emission rates were simulated for rapeseed (3.7 kg N_2O–N/ha/a), closely followed by grassland (3.6 kg N_2O–N/ha/a) and winter wheat (3.3 kg N_2O–N/ha/a). Rapeseed and winter wheat received the highest amount of (mineral and organic) fertilizer being 335 kg N/ha/a and 353 kg N/ha/a respectively. Averaged fertilizer amount applied on grassland was 201 kg N/ha/a with a maximum value of 373 kg N/ha/a. The highest simulated N_2O emission rate across all grids was found for a polygon with a repeseed cultivation with 14.8 kg N_2O–N/ha/a. This high emission strength is based on the local site conditions, as the soil is an organic soil containing a high amount of organic carbon (> 6%) whereas the lowest emissions of 0.2 kg N/ha/a occurred for potatoes cultivated on a mineral soil poor in organic

Table V-2 Source strength of N_2O emissions from agricultural soils in Saxony for the year 2000 according to the German National Emission Inventory (Dämmgen et al. 2007).

Name	Description	N_2O emissions [t N2O-N]
EM1001.02	Emissions due to mineral fertilizer	1 500
EM1001.03	Emissions due to manure	600
EM1001.05	Emissions due to cultivation on organic soils	-
EM1002.04	Emissions resulting from legumes	-
EM1002.05	Emissions resulting from grazing	180
EM1002.06	Emissions resulting from residues	470
EM1002.07	Emissions due to deposition of reactive N	300
Total		*3 050*

carbon (< 0.2%). For details see Fig. V-5, V-7 and V-8. However, the total amount of N_2O released per a grid cell depends on the area fraction of the respective crops.

The average EF calculated as the relative portion of N_2O emitted from applied fertilizer range from 0.96% for winter wheat to 1.8% for grassland sites. The average EF across all agricultural systems was found to be 1.19%. The average emission strength for all agricultural soils equalled 3.02 kg N_2O-N/ha/a.

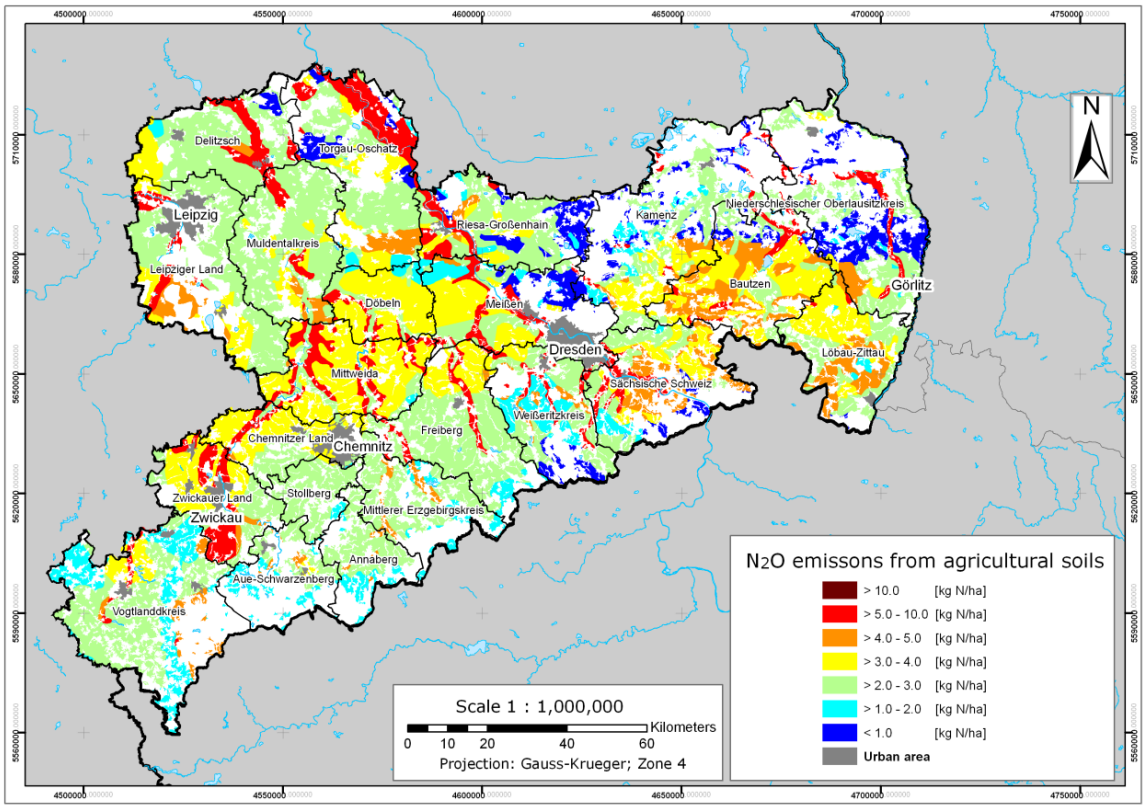

Figure V-8: Simulated N_2O emissions from agricultural soils in Saxony for the year 2000 using LandscapeDNDC.

Fig. V-8 shows the regional distribution of N_2O emissions from agricultural soils for Saxony for the year 2000 based on LandscapeDNDC simulations. Hotspots with highest N_2O fluxes (> 5 kg N_2O-N/ha/a) were generally found in regions with soil organic carbon contents > 2.5 %.

Simulating the inventory with the original DNDC model (version 9.3) resulted in a slightly higher total emission of 2 725 t N_2O-N (see Table V-3).

Table 1 N_2O emission strength of agricultural soils in the State of Saxony for the year 2000. Total agricultural land equals to 902 114 ha, with 714 014 ha (79.15 %) arable land and 181 252 ha (20.09%) grassland.

Method	N_2O emissions	
	Cumulative emission (t N_2O-N)	Average emission rate (t N_2O-N/ha/a)
IPCC Tier 1	1 107	1.227
NIR (IPCC Tier 2)	3 050	3.380
LandscapeDNDC	2 693	2.985
DNDC (version 9.3)	2 725	3.018

Summarizing these results LandscapeDNDC has proven its capabilities to adequately simuate spatially distributed regional trace gas emission inventories. Compared to previous approaches by Butterbach-Bahl et al. (2004a) for the same region the presented study did result in comparable results and differences are mostly due to different N fertilization rates based on different sources of statistics. Butterbach-Bahl et al. (2004a) already conducted a sensitivity analysis, however only for one of the 23 districts of the inventory, due to restrictions by computational capacities. Conclusions from this reduced sensitivity analysis to the overall sensitivity of the inventory may be uncertain. The postulation for a detailed sensitivity analysis accompanying any compilation of a model based GHG emission inventory requires to take into account the full simulation domain.

LandscapeDNDC can fulfil this requirement due to its advanced design with respect to file handling, and parallelization enabling improved calculation of model sensitivity.

Regional sensitivity analysis

The regional GIS database uses explicit input values for each polygone, i.e. assumes homogeneity of a given parameter, thereby neglecting local variations of e.g. soil pH, SOC content, soil texture or small scale differences in climate and agricultural management (fertilization and manure application). To estimate the importance of sub-grid variability on calculated values of the regional emission strength a detailed sensitivity analysis assessing effects of input data uncertainty was performed. To address the sensitivity of the resulting N_2O flux to variability in local conditions and model drivers, selected parameters were varied by fixed percentages. In our sensitivity study we varied the amount of fertilizer N, precipitation, bulk density, SOC and clay content

as well as the soil parameter "field capacity" by ± 10% and ± 25%. Daily mean temperature was changed for ± 1 and ± 2.5˚C and soil pH was changed for ± 4% and ± 10%.

The sensitivity of the simulated N_2O emissions at regional scale to variations in input parameters was calculated using the approach of Friend et al. (1993). The variation was evaluated for each grid cell and expressed by the index of variation β:

$$\beta = \frac{\frac{N_2O_- - N_2O_+}{N_2O_-}}{\frac{p_- - p_+}{p_-}} \tag{6}$$

where p denotes the parameter examined and p_+ and p_- represent the increased/decreased parameter. N_2O_+ in equ. (1) will be the N_2O emissions resulting from a positive change, N_2O_- will be the N_2O emissions resulting from a negative change of the underlying parameter p. A histogram of all values of β explains the sensitivity on the regional scale. The sign of the sensitivity index β determines the direction of the correlation for each grid cell. Positive β values indicate that an increase in the parameter p resulted in an increase in N_2O emissions, while negative β values indicate a decrease in N_2O emissions. The absolute value of β determines the proportion of the sensitivity of the underlying parameter for the N_2O emission.

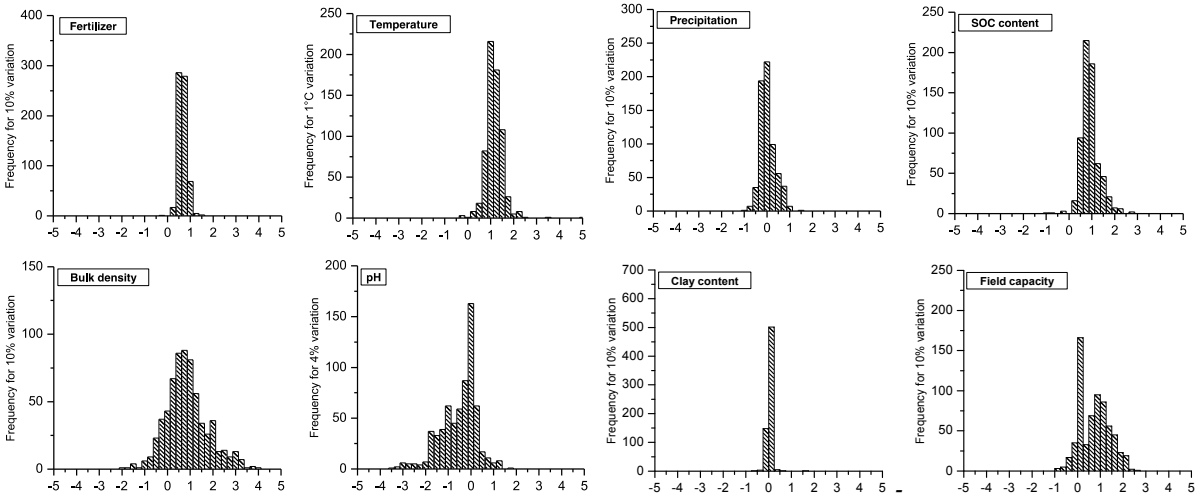

Figure V-9: Sensitivity analysis of the regionalization for N fertilizer input, mean temperature, precipitation, SOC content, buld density, soil pH value, clay content and field capacity by means of the β index.

The frequency distribution of all indexes of variation β (calculated for each grid cell) determines the regional sensitivity of the inventory against the examined parameter which are shown in Fig. V-9. For all grid cells the β values for fertilizer application are positive, i.e. increasing N-fertilization results in elevated N_2O emissions. However, they are ranging between zero and one indicating that for both, a 10% and a 25% variation of fertilizer N inputs, the resulting N_2O emission changes are smaller than 10% or 25% respectively. Highest frequency is observable for values around 0.5. In the 25% variation of fertilizer N inputs, the distribution of the β values is very similar with a higher frequency around 0.5. The nonlinear response is due to the fact that parts of the added fertilizer nitrogen may get leached as nitrate to deeper soil layers or volatilized as NH_3 following fertilizer application rather than resulting in a proportional increase in N_2O emission. In general, nitrification and denitrification activity depend on soil moisture and therefore one would assume in general a positive correlation. However, at least in regions with sandy soils having a high water drainage capacity, increases in precipitation resulted in increases of nitrate leaching, but not necessarily in increased N_2O emissions.

The sensitivity of modelled N_2O emissions is higher for variations in air temperature as for fertilizer input or amount of precipitation. Highest values are up to 3.0 for variations of daily mean air temperatures for 2.5˚C. For most grid cells values between 0.75 and 1.25 were calculated. This result was expected as nitrification and denitrification is modelled in LandscapeDNDC with explicit temperature dependence.

Fig. V-9 shows that sensitivity index values resulting from variations in bulk density are higher than for any other of the investigated parameters.

Table V-4: Scenarios for the regional sensitivity analysis

No.	scenario	desription
1	baseline	No changes
2a	N-fertilization	+25%, +10%, -10%, -25%
2b	temperature	+1˚C, +2.5˚C, -1˚C, -2.5˚C
2c	percipitation	+25%, +10%, -10%,-25%
2d	Soil organic carbon initialization	+25%, +10%, -10%,-25%
2e	Soil bulk density	+25%, +10%, -10%,-25%
2f	Soil texture (clay content)	+25%, +10%, -10%,-25%
2g	Soil pH value	+10%, +4%, -4%,-10%
2h	Field capacity	+25%, +10%, -10%,-25%

The index values go even up to 4, with highest frequency observed between 0.5 and 1 for the 10% variation and between 0.0 and 0.5 for the 25% variation. The high sensitivity of simulated regional N_2O emissions to variations in bulk density is due to an increase in SOC as well as effects of soil aeration (increased bulk density → decreased soil aeration → increased production of N_2O via denitrification) and soil water flow (increased bulk density → decreased saturated hydraulic conductivity → decreased soil aeration at higher soil water contents results in increased production of N_2O via denitrification). Comparable results were also obtained for variations in field capacity.

As already outlined earlier, for sensitivity tests with the original DNDC model (Butterbach-Bahl et al. 2004a) regional simulations of soil N_2O emissions are very sensitive to variations in SOC content. An increase in SOC will result in increased N_2O emissions (see V-5 and V-9). The histograms for a variation of soil C content for 10% and for 25% are almost identical, with highest frequency observed for values around 0.75. The DNDC biogeochemistry concept uses a fixed C/N ratio for organic matter pools of different quality, so that under conditions of increased SOC it is also assumed that the labile fractions are proportionally increasing which might be not realsitic. This could be changed by assuming a different split of the C pools of different quality when increasing SOC. However, in the frame of this study we did not consider differences in the split of C pools for different initial SOC contents, so that increases in simulated initial SOC contents will directly result in increases in organic N, too. Due to increased mineralization activity at high SOC content, more organic N is mineralized and thus also available for microbial N trace gas production.

The regional emission strength varied in the frame of the sensitivity study from 2 030 t N_2O-N/a (for 25% decreased bulk density) to 3 397 t N_2O–N/ha/a (for 25% increased SOC content).

Parallel performance

The measures to evaluate parallel performance are speedup and efficiency. Speedup is defined as

$$S_p = \frac{T_1}{T_p} \tag{7}$$

where p is the number of processors, T_1 is the execution time of the sequential algorithm and T_p is the execution time of the parallel algorithm with p processors.

Linear speedup or ideal speedup is obtained when $S_p = p$, such that running an algorithm with linear speedup, doubling the number of processors doubles the speed. The parallelization is considered to have a good scalability on a target platform, if the speedup values are high (close to the number of processors). Parallel efficiency is defined as

$$E_p = \frac{s_p}{p} \tag{8}$$

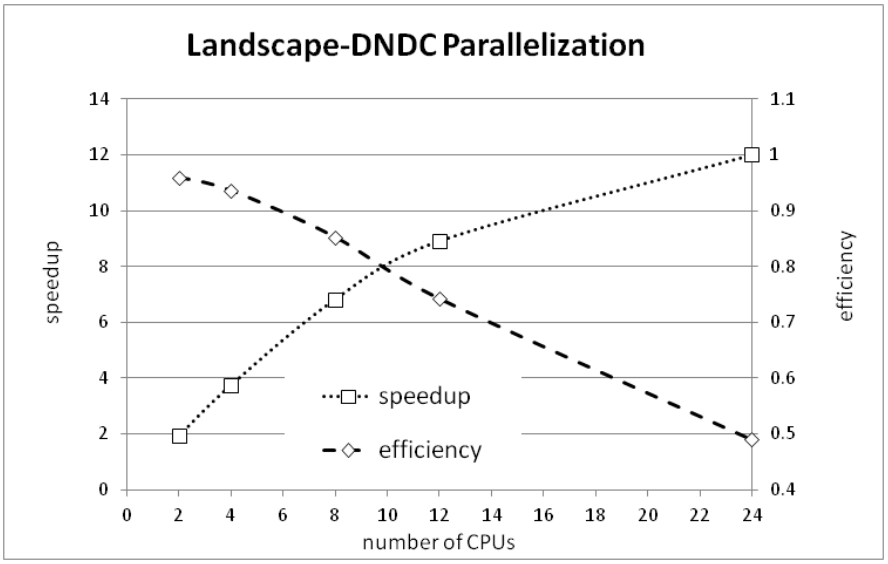

Figure V-10: LandscapeDNDC parallel performance: Speedup and efficiency

Table V-5:Parallelization speedup and efficiency; the measures compare the parallel implementation against the best non-parallel implementation for different CPU configurations from 1 to 24 CPUs using OpenMP parallelization with the gcc 4.4 C++ compiler on a 24 core AMD Opteron workstation.

CPU	1	2	4	8	12	24
Speedup	1	1.92	3.74	6.82	8.91	11.9
Efficiency	1	0.96	0.94	0.85	0.74	0.49

Its value is typically between zero and one estimating how well-utilized the processors are in solving the problem, compared to how much effort is wasted in communication, synchronization, creating threads and other

parallel overhead. Algorithms with linear speedup and algorithms running on a single processor have an efficiency of 1.

The OpenMP parallelization of LandscapeDNDC shows good speedup and efficiency characteristics (see Table V-5 and Fig. V-4). The performance benchmarks were carried out on a 24 core AMD server (2 x AMD 12 Core Opteron). The measures are mean values of 5 simulations to reduce measurement artefacts on the benchmarking systems.

Coupled simulations of N cycling along a virtual hillslope

To prove the advantages of LandscapeDNDC with respect to the option of simulating coupled water and nutrient fluxes at landscape level we coupled LandscapeDNDC to a hydrological model and applied this coupled ecosystem-hydrology model to a virtual hillslope in order to investigate effects of simulated lateral water and nutrient transport on C and N cycling and emission of N2O.. The hydrological model used is the Catchment Modeling Framework CMF (Kraft et al. 2011). CMF is a flexible toolbox that can be used to build water transport models form the plot to the catchment scale. The Python programming language is used to compose the coupled system by loading CMF and LandscapeDNDC as libraries and constructing congruent configurations in CMF and LandscapeDNDC. As the Python coupling program has full access to both models, it can easily perform the data exchange between the CMF and LandscapeDNDC during the time integration.

The virtual hillslope experiment

A virtual 2-dimensional hillslope was designed for a proof of concept experiment to test the general applicability of the coupled biogeochemical-hydrological model. The experiment consists of two hillslopes forming a valley, each 130 m long with an altidudinal difference of 4 m (see Fig. V-11). The right hillslope is formed by an intensively used agricultural upland system cultivated with a barley-maize-wheat rotation with inorganic N input of 300 kg-N/ha/a split in several applications. The left hillslope and the riparian zone represent an extensively used grassland. The riparian zone acts as buffer zone for surface runoff and interflow transported nutrients before entering the stream established at the bottom of the slopes, by a constant head

boundary condition of 0.1 m below ground. An impervious bedrock is set at 3 m belowground, resulting in rapid interflow at the soil-bedrock interface. Cambisol soils cover the shoulder and middle section of the hillslope while gleyic soil conditions are found at the foothill. Evapotranspiration is predicted by CMF following the Penam-Monteith method. Synthetic climate data (mean annual precipitation 1000mm, mean annual temperature 9.6°C) in daily resolution was taken to drive the simulations representing conditions of temperate oceanic climate. The hillslope is discretized into 26 grid cells of 10 m horizontal length each using 14 soil layers with a vertical layer thickness from 10 (topsoil) to 30 cm. The cells are allocated to each grid cell with lateral flux connections between the layers of neighbouring cells. Lateral fluxes are simulated using the Richards equation at unsaturated conditions and the Darcy concept for saturated flow. A kinematic wave approach is used to route surface flow to the bottom of the hillslope. Temporal resolution of the data exchange of the coupled system is 1 day, for which following sequence is executed: (1) run Landscape-DNDC for one day for all grid cells, (2) CMF update of states, parameters and boundary conditions with data supplied by LandscapeDNDC, (3) run CMF for one day and (4) Landscape-DNDC update of states, parameters and boundary conditions with data supplied by CMF. Water transport calucations in Landscape-DNDC are switched off to avoid model interferences. As solute storages are driven by both models (reactive fluxes by Landscape-DNDC and adjective fluxes by CMF), execution steps (2) and (4) overwrite state variables at each time step. The coupled simulation was performed for a period of 3 years.

Effects of lateral flow of nutrients on soil N_2O fluxes

The study consists of two simulations: (i) a baseline scenario in which the biogeochemical model simulates the C and N cycling of each grid cell stand alone neglecting lateral nutrient transport; (ii) a modelling scenario including lateral nutrient transport of NO_3, NH_4, DON and DOC by use of LandscapeDNDC coupled to CMF at daily time scale. Fig. V-11a) shows discharge formation (blue arrow) resulting from rainfall events and corresponding lateral water fluxes along both hillslopes as simulated by the CMF model. Fig. V-11b) illustrates the soil nitrate content (scaled and

indicated in colour transition from blue to yellow) together with the above ground biomass (green bars) and the yearly cummulated N_2O emissions (yellow bars) for each grid cell along both hillslopes in the third simulation year just before harvest. The illustration visualizes the impact of the high N fertilizer input at the upper part of the right hillslope causing a much higher plant productivity (green bars) as well as higher nitrous oxide emissions (yellow bars) compared to the low input system at the left hillslope. Furthermore, the lateral nutrient transport along the intensively used hillslope towards the riparian zone is causing a higher nutrient availability and thus, an increase in biomass productivity and also N_2O emissions further down the hillslope where no N fertilizer was applied.

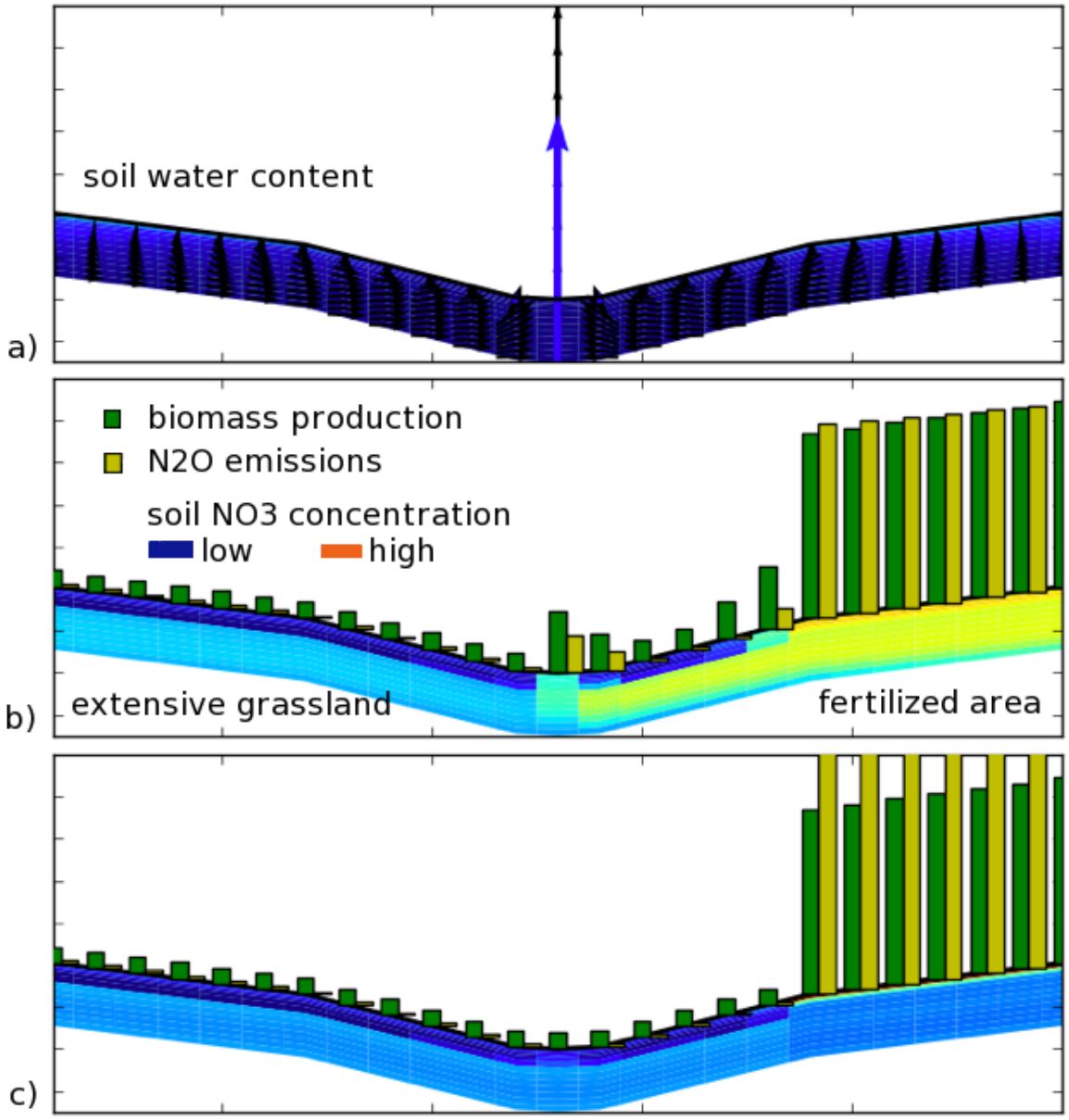

Figure V-11: Coupled LandscapeDNDC/ CMF and baseline (LandscapeDNDC stand alone) simulations of water and nitrogen fluxes along two hillslopes (left extensive grassland; right intensive fertilized rotation of winter barley, maiz, winterwheat: a) Soil water content (indicated by the blue countour plot) and the water fluxes along the hillslope as well as discharge formation (indicated by arrows) as simulated by CMF. b) Coupled simulation and c) stand alone simulaton of soil nitrate content (indicated by the contour plot of the soil layer) and accumulated plant biomass production (green bars) and N_2O emissions (golden bars) before harvest in the third year of simulation.

Fig. V-11c) clearly demonstrates that the impact of the topography causing lateral water and nutrient transport can not be captured by the simple stand alone application of LandscapeDNDC, since elevated plant growth and formation of N_2O emissions is restricted to the area of N fertilizer application only.

Conclusions

With LandscapeDNDC we present a new model system, which facilitates scaling of ecosystem processes from the plot to regional simulation domains and which can be coupled to hydrological models in order to better describe impacts of landscape topography on matter fluxes of water, carbon and nitrogen.

Comparing the LandscapeDNDC approach for regionalization of N_2O emissions of Saxony on a 24 CPU workstation to the standard approach of simulating all grid cells by duplication of the site model and running 24 sites individually at the same time, revealed a acceleration of the overall computation time by a factor of more than 10 when working on a local file systems. Using a NFS mounted file server even increased the factor to 25. This is mostly based on the enormous file input / output operations running concurrently and blocking each other when 24 instances of the site model are running individually at the same time whereas LandscapeDNDC uses only four regional input files and writes gridded output efficiently to the regional output files. This has to be compared to approx. 90000 files being created while using the original DNDC coupled to a GIS. Utilizing expensive ram disk based file systems might be a way to reduce this slow down in overall performance by the previous approach, but the problem of handling the large amount of single files remains and the handling of such a huge number of single files was rather inefficient not only from the computational point of view, but also with regard to data evaluation and visualisation.

The new realized strategy of handling I/O routines also allows to apply more sophisticated sensitivity analyses at regional levels. In our study the amount of input files used for the sensitivity analysis was reduced to 132 regional input files and 231 regional output files (4 input files and to 7 output files for each regional simulation) compared to approx. 2.1 million files using the previous appoach. The preprocessing to create the whole set of input data for the regional simulation was speed up by a factor of more than 100 compared to the previous process of creating four files for each grid cell. The same applies to the post-processing. For these reasons, LandscapeDNDC has proven its advantages in compiling regional GHG inventories compared to the common model approaches, in particular

with respect to the feasibility of regional or national sensitivity/ uncertainty analysis which are of uppermost importance in the framework of UNFCC GHG reporting using TIER 3 approaches.

The potential of the LandscapeDNDC model addressing future challenges like C & N cycling as affected by landscape topography was successfully demonstrated by construction and application of the coupled biogeochemical ecosystem with a spatially distributed hydrological model for a virtual hillslope. The application shows that for a better understanding and simulation of plant productivity and N_2O fluxes at landscape levels it is a prerequisite to consider lateral transport and dispersion processes of nutrients. The application could show the formation of indirect N_2O emissions and the increase in plant productivity at the lower hillsplope due to water driven nutrient transport from areas of fertilization at the upper part of the hillslope. In a next step the coupled model system as well as the capability of LandscapeDNDC for simulation of transient LUC will be tested against "real" field data.

Acknowledgements

The authors thank the European Commission funded research project NitroEurope IP (contract 017841) and the German Helmholtz Society for supporting this work within the ENTRANCE project as well as the Robert Bosch Foundation Germany for support.

The authors would like to acknowledge the help and assistance of the bureau of environment, agriculture and geology (LfULG) of the state of Saxony for providing input data for the regionalization.

References

Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E., Rasmussen, J., 1986. An introduction to the European Hydrological System - Systéme Hydrologique Européen, "SHE", 1: History and philosophy of a physically based distributed modelling system. J.Hydrol. 87, 45-59.

Aber JD, Federer CA (1992) A generalized, lumped-parameter model of photosynthesis, evapotranspiration and net primary production in temperate and boreal forest ecosystems. Oecologia

Aber, J.D., Ollinger, S.V., Driscoll, C.T., 1997. Modeling nitrogen saturation in forest ecosystems in response to land use and atmospheric deposition. Ecol.Model. 101, 61-78.

Acevedo, E., Silva, P., Silva, H., 2002. Wheat growth and physiology. Bread wheat: improvement and production.Rome: Food and Agriculture Organization (FAO) 39-70.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements. FAO Irrigation and drainage paper 56.

Anderson, K., Nelson, S., Mayo, A., Tingey, D., 2006. Interbasin flow revisited: The contribution of local recharge to high-discharge springs, Death Valley, CA. J.Hydrol. 323, 276-302.

Andréassian, V., Perrin, C., Parent, E., Bárdossy, A., 2010. The Court of Miracles of Hydrology: can failure stories contribute to hydrological science? Hydrol.Sci.J. 55, 849-856.

Arnold, J.G., Fohrer, N., USDA, A.R.S., 2005. SWAT2000: current capabilities and research opportunities in applied watershed modelling.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment. Part I: Model development. J.Am.Water Res.Assoc. 34, 73-88.

Bakker, M., Kelson, V.A., 2009. Writing Analytic Element Programs in Python. Ground Water 47, 828-834.

Band, L.E., Tague, C.L., Brun, S.E., Tenenbaum, D.E., Fernandes, R.A., 2001. Modeling watersheds as spatial object hierarchies: structure and dynamics. Transactions GIS 4_ _, 181-196.

Band, L.E., Tague, C.L., Groffman, P., Belt, K., 2001. Forest ecosystem processes at the watershed scale: hydrological and ecological controls of nitrogen export. Hydrol.Proc. 15, 2013-2028.

Barthold, F.K., Sayama, T., Schneider, K., Breuer, L., Vaché, K.B., Frede, H., McDonnell, J.J., 2008. Gauging the ungauged basin: a top-down approach in a large semiarid watershed in China. Adv.Geosci. 18, 3-8.

Barthold, F.K., Tyralla, C., Schneider, K., Vaché, K.B., Frede, H.G., Breuer, L., 2011. How many tracers do we need for end member mixing analysis (EMMA)? A sensitivity analysis. Water Resour.Res. 47.

Barthold, F.K., Wu, J., Vaché, K.B., Schneider, K., Frede, H.G., Breuer, L., 2010. Identification of geographic runoff sources in a data sparse region: hydrological processes and the limitations of tracer based approaches. Hydrol. Process. 24, 2313-2327.

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

Beazley, D.M., Lomdahl, P.S., 1996. Lightweight computational steering of very large scale molecular dynamics simulations, in: Proceedings of the 1996 ACM/IEEE conference on Supercomputing. pp. 50-61.

Beheydt D, Boeckx P, Sleutel S, et al. (2007) Validation of DNDC for 22 long-term N2O field emission measurements. Atmos Environ 41:6196–6211. doi: 10.1016/j.atmosenv.2007.04.003

Belcher, W.R., Bedinger, M.S., Back, J.T., Sweetkind, D.S., 2009. Interbasin flow in the Great Basin with special reference to the southern Funeral Mountains and the source of Furnace Creek springs, Death Valley, California, US. J.Hydrol. 369, 30-43.

Bergström, S., 1995. The HBV Model, in: Computer models of watershed hydrology. Water Resources Publications, Highland Ranch, Colorado, USA, p. 443_ _-476.

Beven, K., 2006. Searching for the Holy Grail of scientific hydrology. Hydrol.Earth Syst.Sci 10, 609-618.

Beven, K.J., 2002. Towards an alternative blueprint for a physically-based digitally simulated hydrologic response modelling system. Hydrol.Proc. 16, 189-206.

Beven, K.J., Calver, A., Morris, E.M., 1987. The Institute of Hydrology distributed model (No. Report 98). _Institute of Hydrology; Wallingford, Oxon, England_.

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin 24, 43-69.

Blagodatsky SA, Grote R, Kiese R, et al. (2011) Modelling of microbial carbon and nitrogen turnover in soil with special emphasis on N-trace gases emission. Plant Soil 346:297–330. doi: 10.1007/s11104-011-0821-z

Boyer EW, Alexander RB, Parton WJ, et al. (2006) Modeling denitrification in terrestrial and aquatic ecosystems at regional scales. Ecol Appl 16:2123–2142.

Bronstert, A., 1999. Capabilities and limitations of detailed hillslope hydrological modelling. Hydrol.Proc. 13.

Butterbach-Bahl K, Dannenmann M (2011) ScienceDirect - Current Opinion in Environmental Sustainability : Denitrification and associated soil N2O emissions due to agricultural activities in a changing climate. Current Opinion in Environmental …

Butterbach-Bahl K, Kahl M, Mykhayliv L, et al. (2009) A European-wide inventory of soil NO emissions using the biogeochemical models DNDC/Forest-DNDC. Atmos Environ 43:1392–1402. doi: DOI: 10.1016/j.atmosenv.2008.02.008

Butterbach-Bahl K, Kesik M, Miehle P, et al. (2004a) Quantifying the regional source strength of N-trace gases across agricultural and forest ecosystems with process based models. Plant Soil 260:311–329.

Butterbach-Bahl K, Kock M, Willibald G, et al. (2004b) Temporal variations of fluxes of NO, NO2, N2O, CO2, and CH4 in a tropical rain forest ecosystem. Global Biogeochem Cy 18:GB3012. doi: 10.1029/2004GB002243

Butterbach-Bahl K, Kock M, Willibald G, et al. (2004c) Temporal variations of fluxes of NO, NO2, N2O, CO2, and CH4 in a tropical rain forest ecosystem. Global Biogeochem Cycles. doi: 10.1029/2004GB002243

Buytaert, W., Reusser, D., Krause, S., Renaud, J.P., 2008. Why can't we do better than Topmodel? Hydrol.Proc. 22, 4175-4179.

Chatskikh D, Olesen JE, Berntsen J, et al. (2005) Simulation of Effects of Soils, Climate and Management on N2O Emission from Grasslands. Biogeochemistry 76:395–419. doi: 10.1007/s10533-005-6996-8

Chen, Z., 1988. Topography and climate of the Xilin River basin, in: Inner Mongolia Research Station of Academia Sinica, Research on Grassland Ecosystem. Science Press, Beijing, pp. 13-22.

Chirinda N, Kracher D, Lægdsmand M, et al. (2010) Simulating soil N2O emissions and heterotrophic CO2 respiration in arable systems using FASSET and MoBiLE-DNDC. Plant Soil 343:139–160. doi: 10.1007/s11104-010-0596-7

Ciarapica, L., Todini, E., 2002. TOPKAPI: a model for the representation of the rainfall-runoff process at different scales. Hydrol.Proc. 16.

Cicerone RJ, Shetter JD (1981) Sources Of Atmospheric Methane: Measurements in Rice Paddies and a Discussion. J Geophys Res 86:7203–7209. doi: 10.1029/JC086iC08p07203

Clark, M.P., Kavetski, D., 2010. Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes. Water Resources Research 46, W10510.

Clark, M.P., Kavetski, D., Fenicia, F., 2011. Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resour.Res 47.

Clark, M.P., Slater, A.G., Rupp, D.E., Woods, R.A., Vrugt, J.A., Gupta, H.V., Wagener, T., Hay, L.E., 2008. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resour.Res. W00B02, doi:10.1029/2007WR006735.

Conant RT (2011) Sequestration through forestry and agriculture. Wires Clim Change 2:238–254. doi: 10.1002/wcc.101

Cui J, Li C, Sun G, Trettin C (2005) Linkage of MIKE SHE to Wetland-DNDC for carbon budgeting and anaerobic biogeochemistry simulation. Biogeochemistry 72:147–167. doi: 10.1007/s10533-004-0367-8

Cui, J., Li, C., Sun, G., Trettin, C., 2005. Linkage of MIKE SHE to Wetland-DNDC for carbon budgeting and anaerobic biogeochemistry simulation. Biogeochemistry 72, 147-167.

Dämmgen U, Grünhage L (2002) Trace gas emissions from German agriculture as obtained from the application of simpler or default methodologies. Environ Pollut 117:23–34.

Dämmgen U, Lüttich M, Haenel H, Döhler H (2007) Dämmgen: Berechnungen der Emissionen aus der deutschen Landbauforschung Völkenrode, Institut für Agrarrelevante Klimaforschung Johann Heinrich von Thünen Institut (vTI) Bundesforschungsinstitut für Ländliche Räume, Wald und Fischerei

Del Grosso SJ, Mosier AR, Parton WJ, Ojima DS (2005a) DAYCENT model analysis of past and contemporary soil N2O and net greenhouse gas flux for major crops in the USA. Soil Till Res 83:9–24. doi: 10.1016/j.still.2005.02.007

Del Grosso SJ, Ogle SM, Parton WJ, Breidt FJ (2010) Estimating uncertainty in N2O emissions from US cropland soils. Global Biogeochem Cy 24:GB1009. doi: 10.1029/2009GB003544

Del Grosso SJ, Parton WJ, Mosier AR, et al. (2005b) Modeling soil CO2 emissions from ecosystems. Biogeochemistry 73:71–91. doi: 10.1007/s10533-004-0898-z

Diepen, C.A., Wolf, J., Keulen, H., Rappoldt, C., 2007. WOFOST: a simulation model of crop production. Soil Use Managem. 5, 16-24.

Dragosits, U., Theobald, M.R., Place, C.J., Lord, E., Webb, J., Hill, J., ApSimon, H.M., Sutton, M.A., 2002. Ammonia emission, deposition and impact assessment at the field scale: a case study of sub-grid spatial variability. Environ.Poll. 117, 147-158.

Durner, W., 1994. Hydraulic conductivity estimation for soils with heterogeneous pore structure. Water Resour.Res. 30, 211-224.

Eckert, S., Baaser, H., Gross, D., Scherf, O., 2004. A BDF2 integration method with step size control for elasto-plasticity. Computational Mechanics 34, 377-386.

Exbrayat, J.-F., Viney, N.R., Seibert, J., Wrede, S., Frede, H.-G., Breuer, L., 2010. Ensemble modelling of nitrogen fluxes: data fusion for a Swedish meso-scale catchment. Hydrology and Earth System Sciences 14, 2383-2397.

Fan, L., Liu, S., Bernhofer, C., Liu, H., Berger, F.H., 2006. Regional land surface energy fluxes by satellite remote sensing in the Upper Xilin river watershed (Inner Mongolia, China). Theor.Appl.Clim. 88, 231-245.

Feddes, R.A., Hoff, H., Bruen, M., Dawson, T., de Rosnay, P., Dirmeyer, P., Jackson, R.B., Kabat, P., Kleidon, A., Lilly, A., 2001. Modeling root water uptake in hydrological and climate models. Bulletin of the American Meteorological Society 82, 2797-2809.

Federer, C.A., 1995. BROOK90: a simulation model for evaporation, soil water and streamflow. Computer freeware and documentation.USDA Forest Service, Durham, New Hampshire, USA.

Feldman, A.D., (US), H.E.C., 2000. Hydrologic Modeling System HEC-HMS: Technical Reference Manual. US Army Corps of Engineers, Hydrologic Engineering Center.

Fenicia, F., McDonnell, J.J., Savenije, H.H.G., 2008. Learning from model improvement: On the contribution of complementary data to process understanding. Water Resour. Res. 44.

Fenicia, F., Savenije, H.H.G., Matgen, P., Pfister, L., 2006. Is the groundwater reservoir linear? Learning from data in hydrological modelling. Hydrol.Earth Sys.Sci. 10, 139-151.

Firestone M, Davidson E (1989) Microbial basis of NO and N2O production and consumption. Dahlem Koferenzen

Fouilloux A, Fouilloux A, Fouilloux A, et al. (1999) The PALM Project: MPMD Paradigm for an Oceanic Data Assimilation Software. In: Lecture Notes In Computer Science. Springer, pp 1423–1430

Freeze, R.A., Harlan, R.L., 1969. Blueprint for a physically-based, digitally-simulated hydrologic response model. J.Hydrol. 9, 237-258.

Friend A, Schugart H, Running S (1993) A Physiology-Based Gap Model of Forest Dynamics. Ecology 74:792–797.

Gabrielle B, Laville P, Henault C, et al. (2006) Simulation of nitrous oxide emissions from wheat-cropped soils using CERES. Nutr Cycl Agroecosys 74:133–146. doi: 10.1007/s10705-005-5771-5

Gayle, G.A., Skaggs, R.W., 1978. Surface storage on bedded cultivated lands. Trans.Am.Soc.Agricult.Eng. 21.

Grant RF, Nyborg M, Laidlaw JW (1993) Evolution of nitrous oxide from soil. Soil Science 156:259–265. doi: 10.1097/00010694-199310000-00006

Grant RF, Pattey E (2003) Modelling variability in N2O emissions from fertilized agricultural fields. Soil Biol Biochem 35:225–243. doi: 10.1016/S0038-0717(02)00256-0

Gregersen, J.B., Gijsbers, P.J.A., Westen, S.J.P., 2007. OpenMI: Open Modelling Interface. J.Hydroinformatics 9, 175-191.

Grote R., 2007. Sensitivity of volatile monoterpene emission to changes in canopy structure: a model-based exercise with a process-based emission model. New Phytol 173:550–561. doi: 10.1111/j.1469-8137.2006.01946.x

Grote R, Kiese R, Gruenwald T, et al. (2011) Modelling forest carbon balances considering tree mortality and removal. Agr Forest Meteorol 151:179–190. doi: 10.1016/j.agrformet.2010.10.002

Grote R, Lehmann E, Brummer C (2009) Modelling and observation of biosphere–atmosphere interactions in natural savannah in Burkina Faso, West Africa. Physics and Chemistry of the Earth 34:

Guo, J.K., Liu, J.Y., Huang, G.M., Zhuang, D.F., Yan, H.M., 2004. Cropland identification in Inner Mongolia, China with SPOT-4 VEGETATION imagery, in: IEEE International Geoscience and Remote Sensing Symposium. IEEE, Anchorage, AK, pp. 3988-3991.

Haas, E., Klatt, S., Fröhlich, A., Kraft, P., Werner, C., Kiese, R., Grote, R., Breuer, L., Butterbach-Bahl, K., 2011. Towards a new approach to simulating regional N2O emissions - the LandscapeDNDC Model. Landscape Ecol. submitted.

Harbaugh, A.W., Banta, E.R., Hill, M.C., McDonald, M.G., 2000. MODFLOW-2000, The U. S. Geological Survey Modular Ground-Water Model-User Guide to Modularization Concepts and the Ground-Water Flow Process. United States Geological Survey.

Hindmarsh, A.C., Brown, P.N., Grant, K.E., Lee, S.L., Serban, R., Shumaker, D.E., Woodward, C.S., 2005. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. ACM TOMS 31, 363-396.

Holländer, H.M., Blume, T., Bormann, H., Buytaert, W., Chirico, G.B., Exbrayat, J.F., Gustafsson, D., Hölzel, H., Kraft, P., Stamm, C., Stoll, S., Bl"schl, G., Flühler, H., 2009. Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data. Hydrol.Earth Syst.Sci. 13, 2069-2094.

Holst J, Grote R, Offermann C, et al. (2010) Water fluxes within beech stands in complex terrain. Int J Biometeorol 54:23–36. doi: 10.1007/s00484-009-0248-x

Hopmans, J.W., Bristow, K.L., 2002. Current capabilities and future needs of root water and nutrient uptake modeling. Adv.Agron. 77, 103-183.

Houghton RA, Hackler J (1999) Emissions of carbon from forestry and land-use change in tropical Asia. Global Change Biol 5:481–492.

IPCC (2000) Land use, land-use change, and forestry. Cambridge: IPCC

IPCC (2007) Climate change 2007: The physical science basis. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

IUSS Working GROUP, W., 2007. World Reference Base for Soil Resources 2006, first update 2007. FAO, Rome.

Jarvis, N.J., Hollis, J.M., Nicholls, P.H., Mayr, T., Evans, S.P., 1997. MACRO--DB: a decision-support tool for assessing pesticide fate and mobility in soils. Environmental Modelling and Software 12, 251-265.

Kaharabata S, Drury C, Priesack E, et al. (2003) Comparing measured and Expert-N predicted N2O emissions from conventional till and no till corn treatments. Nutr Cycl Agroecosys 66:107–118.

Karssenberg, D., 2002. The value of environmental modelling languages for building distributed hydrological models. Hydrol.Proc. 16, 2751-2766.

Karssenberg, D., de Jong, K., van der Kwast, J., 2007. Modelling landscape dynamics with Python. International Journal of Geographical Information Science 21, 483-495.

Kavetski, D., Clark, M.P., 2011. Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing. Hydrol. Process. 25, 661-670.

Kesik M, Ambus P, Baritz R, et al. (2005) Inventories of N2O and NO emissions from European forest soils. Biogeosciences 2:353–375.

Kiese R, Heinzeller C, Werner C, et al. (2011) Quantification of nitrate leaching from German forest ecosystems by use of a process oriented biogeochemical model. Environ Pollut 159:3204–3214. doi: 10.1016/j.envpol.2011.05.004

Kirchner, J., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resour.Res. 42, doi:10.1029/2005WR004362.

Koomen E, Stillwell J, Bakema A, Scholten HJ 2008 Modelling Land-Use Change, Progress and Applications. Springer The GeoJournal Library 89, Dorrecht, The Netherlands

Kraft, P., Barthold, F.K., Frede, H.-G., Breuer, L., 2011. Refining model structures in a poorly gauged catchment based on soft data, educated guesses and model failures. Water Resour.Res., submitted, Chapter III

Kraft, P., Vache, K.B., Breuer, L., Frede, H.-G., 2008. A solute and water flux library for catchment models, in: iEMSs 2008: International Congress on Environmental Modelling and Software. Integrating Sciences and Information Technology for Environmental Assessment and Decision Making. 4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings. International Environmental Modelling and Software Society (iEMSs), pp. 2053-2056.

Kraft, P., Vache, K.B., Frede, H.-G., Breuer, L., 2011. A hydrological programming language extension for integrated catchment models. Environ.Modell.Softw 26, 828-830., Chapter II

Krümmelbein, J., Wang, Z., Zhao, Y., Peth, S., Horn, R., 2006. Influence of various grazing intensities on soil stability, soil structure and water balance of grassland soils in Inner Mongolia, PR China. Adv.Geoecol. 38, 93-102.

Kuczera, G., Renard, B., Thyer, M., Kavetski, D., 2010. There are no hydrological monsters, just models and observations with large uncertainties! Hydrol.Sci.J. 55, 980-991.

Lagarde, T., Piacentini, A., Thual, O., 2001. A new representation of data-assimilation methods: The PALM flow-charting approach. Quart.J.R.Met.Soc 127, 189-207.

Le Moine, N., Andréassian, V., Perrin, C., Michel, C., 2007. How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments. Water Resour.Res 43.

Lehuger S, Gabrielle B, van Oijen M, et al. (2009) Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model. Agr Ecosyst Environ 133:208–222. doi: 10.1016/j.agee.2009.04.022

Li C (2000) Modeling trace gas emissions from agricultural ecosystems. Nutr Cycl Agroecosys 58:259–276.

Li C, Aber JD, Stange F, et al. (2000) A process-oriented model of N2O and NO emissions from forest soils: 1. Model development. J Geophys Res 105:4369–4384.

Li C, Frolking S, Butterbach-Bahl K (2005) Carbon sequestration in arable soils is likely to increase nitrous oxide emissions, offsetting reductions in climate radiative forcing. Climatic Change 72:321–338. doi: 10.1007/s10584-005-6791-5

Li, C., Frolking, S., Frolking, T.A., 1992. A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity. J.Geophys.Res. 97, 9759-9776.

Lindgren, G.A., Wrede, S., Seibert, J., Wallin, M., 2007. Nitrogen source apportionment modeling and the effect of land-use class related runoff contributions. Nordic Hydrol. submitted_ _.

Maurer, T., 1997. Physikalisch begründete, zeitkontinuierliche Modellierung des Wassertransports in kleinen ländlichen Einzugsgebieten.

McDonnell, J.J., McGuire, K., Aggarwal, P., Beven, K.J., Biondi, D., Destouni, G., Dunn, S., James, A., Kirchner, J., Kraft, P., Lyon, S., Maloszewski, P., Newman, B., Pfister, L., Rinaldo, A., Rodhe, A., Sayama, T., Seibert, J., Solomon, K., Soulsby, C., Stewart, M., Tetzlaff, D., Tobin, C., Troch, P., Weiler, M., Western, A., Worman, A., Wrede, S., 2010. How old is streamwater? Open questions in catchment transit time conceptualization, modelling and analysis. Hydrol.Proc. 24, 1745-1754.

Meinshausen M, Meinshausen N, Hare W, et al. (2009) Greenhouse-gas emission targets for limiting global warming to 2 ˚C. Nature 458:1158–1162. doi: 10.1038/nature08017

Miehle P, Grote R, Battaglia M (2010) Evaluation of a process-based ecosystem model for long-term biomass and stand development of Eucalyptus globulus plantations - European Journal of Forest Research, Volume 129, Number 3.

Miller, P., Lanier, W., Brandt, S., 2001. Using Growing Degree Days to Predict Plant Stages. Montguide. http://www.ipm.montana.edu/Training/PMT../2006/mt200103.pdf

Monteith, J.L., Moss, C.J., 1977. Climate and the Efficiency of Crop Production in Britain. Philosophical Transactions of the Royal Society of London.Series B, Biological Sciences 281, 277-294.

Norman J, Jansson P-E, Farahbakhshazad N, et al. (2008) Simulation of NO and N2O emissions from a spruce forest during a freeze/thaw event using an N-flux submodel from the PnET-N-DNDC model integrated to CoupModel. Ecol Model 216:18–30. doi: 10.1016/j.ecolmodel.2008.04.012

Ousterhout, J.K., 1998. Scripting: Higher-level programming for the 21st century. IEEE Computer 31, 23-30.

Parton, W.J., Anderson, D.W., Cole, C.V., Stewart, J.W.B., 1983. Simulation of soil organic matter formation and mineralization in semiarid agroecosystems, Special Publ. No. 23. The Univ. of Georgia, College of Agriculture Experiment Stations, Athens, Georgia.

Peterson, P., 2009. F2PY: a tool for connecting Fortran and Python programs. Int.J.Computational Science and Engineering 4.

Plesca, I., Timbe, E., Exbrayat, J.F., Windhorst, D., Kraft, P., Crespo, P., Vaché, K.B., Frede, H.G., Breuer, L., 2011. Model intercomparison to explore catchment functioning: Results from a remote montane tropical rainforest. Ecol.Model.

Pohlert T, Breuer L, Huisman JA, Frede HG (2007a) Assessing the model performance of an integrated hydrological and biogeochemical model for discharge and nitrate load predictions. Hydrol Earth Syst Sc 11:997–1011. doi: 10.5194/hess-11-997-2007

Pohlert T, Huisman JA, Breuer L, Frede HG (2007b) Integration of a detailed biogeochemical model into SWAT for improved nitrogen predictions—Model development, sensitivity, and GLUE analysis. Ecol Model 203:215–228. doi: 10.1016/j.ecolmodel.2006.11.019

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical recipes in FORTRAN: the art of scientific computing. Cambridge Univ Pr.

Qu, Y.Z., Duffy, C.J., 2007. A semidiscrete finite volume formulation for multiprocess watershed simulation. Water Resour.Res. 43, W08419, doi:10.1029/2006WR005752.

Refsgaard, J.C., Storm, B., 1995. MIKE SHE, in: Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado, pp. 809-846.

Roering J, Almond P, Tonkin P (2004): Constraining climatic controls on hillslope dynamics using a coupled model for the transport of soil and tracers: Application to loess-mantled hillslopes, South Island. J Geophys Res, 109

Rolland M, Gabrielle B, Laville P, Cellier P (2010) High-resolution inventory of NO emissions from agricultural soils over the Ile-de-France region. Environmental Pollution 158(3), 711-722

Roussel, C.J., Roussel, M.R., 2003. Generic object-oriented differential equation integrators. C/C++ Users Journal,(11) 18-23.

Sanner, M.F., 1999. Python: a programming language for software integration and development. J.Mol.Graphics Mod 17, 57-61.

Schäfer, J., 2009. Auswirkungen unterschiedlicher Landnutzungsszenarien auf den Wasserhaushalt im Einzugsgebiet des Xilin, Innere Mongolei, China (Diploma Thesis). Justus-Liebig-Universität Gießen.

Schmitter P, Dercon G, Hilger T, Hertel M (2010) ScienceDirect - Agriculture, Ecosystems & Environment : Linking spatio-temporal variation of crop response with sediment deposition along paddy rice terraces. Agriculture

Schmitz, O., Karssenberg, D., Van Deursen, W.P.A., Wesseling, C.G., 2009. Linking external components to a spatio-temporal modelling framework: Coupling MODFLOW and PCRaster. Environmental Modelling and Software 24, 1088-1099.

Schneider, K., 2008. Water fluxes on different spatial and temporal scales in a semi-arid steppe environment: experimental and modelling approaches (Dissertation, Justus-Liebig-Universität, Gießen).

Schneider, K., Ketzer, B., Breuer, L., Vache, K.B., Bernhofer, C., Frede, H.-G., 2007. Evaluation of evapotranspiration methods for model validation in a semi-arid watershed in northern China. Adv.Geosci. 11, 37-42.

Schulze ED, Ciais P, Luyssaert S, et al. (2010) The European carbon balance. Part 4: integration of carbon and other trace-gas fluxes. Global Change Biol 16:1451–1469. doi: 10.1111/j.1365-2486.2010.02215.x

Seibert, J., 1997. Estimation of parameter uncertainty in the HBV model. Nordic Hydrol. 28, 247-262.

Seibert, J., McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. Water Resour.Res. 38, doi: 10.1029/2001WR000978.

Shuttleworth, W.J., Wallace, J.S., 1985. Evaporation from sparse crops-an energy combination theory. Quart.J.R.Met.Soc 111, 839–855.

Silberstein, R.P., 2006. Hydrological models are so good, do we still need data? Environ.Modell.Softw 21, 1340-1352.

Silberstein, R.P., Sivapalan, M., Wyllie, A., 1999. On the validation of a coupled water and energy balance model at small catchment scales. J.Hydrol. 220, 149-168.

Simunek, J., Hopmans, J., 2009. Modeling compensated root water and nutrient uptake. Ecol.Model. 220, 505-521.

Simunek, J., Sejna, M., van Genuchten, M.T., 1999. The HYDRUS-2D software package for simulating two-dimensional movement of water, heat, and multiple solutes in variably saturated media. Version 2.0,IGWMC - TPS - 53, International Ground Water Modeling Center, Colorado School of Mines, Golden, Colorado.

Simunek, J., Sejna, M., van Genuchten, M.T., 2005. The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. University of California, Riverside, Research reports 240.

Simunek, J., van Genuchten, M.T., Sejna, M., 2008. Development and applications of the HYDRUS and STANMOD software packages and related codes. Vadose Zone Journal 7, 587.

Singh, V.P., Frevert, D.K., 2002. Mathematical models of small watershed hydrology and applications. Water Resources Publications, Highlands Ranch.

Sivapalan, M., Blöschl, G., Zhang, L., Vertessy, R., 2003. Downward approach to hydrological prediction. Hydrol.Proc. 17, 2101-2111.

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'connell, P.E., 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. Hydrol.Sci.J. 48, 857-880.

Sivapalan, M., Viney, N.R., Ruprecht, J.K., 1996. Water and salt balance modelling to predict the effects of land-use changes in forested catchments .2. Coupled model of water and salt balances. Hydrol.Proc. 10, 413-428.

Skiba U, Drewer J, Tang Y, Van Dijk N (2009) Biosphere–atmosphere exchange of reactive nitrogen and greenhouse gases at the NitroEurope core flux measurement sites: Measurement strategy and first data sets. Agriculture

Smith P, Martino D, Cai Z, et al. (2008) Greenhouse gas mitigation in agriculture. Philos Trans R Soc Lond, B, Biol Sci 363:789–813. doi: 10.1098/rstb.2007.2184

Stange F, Butterbach-Bahl K, Papen H, et al. (2000) A process-oriented model of N2O and NO emissions from forest soils 2. Sensitivity analysis and validation. J Geophys Res 105:4385–4398.

Sutton MA, Nemitz E, Erisman JW, et al. (2007) Challenges in quantifying biosphere-atmosphere exchange of nitrogen species. Environ Pollut 150:125–139. doi: 10.1016/j.envpol.2007.04.014

Tetzlaff, D., Carey, S.K., Laudon, H., McGuire, K., 2010. Catchment processes and heterogeneity at multiple scales—benchmarking observations, conceptualization and prediction. Hydrol. Process. 24, 2203-2208.

Tetzlaff, D., McDonnell, J.J., Uhlenbrook, S., McGuire, K.J., Bogaart, P.W., Naef, F., Baird, A.J., Dunn, S.M., Soulsby, C., 2008. Conceptualizing catchment processes: simply too complex? Hydrol.Proc. 22, 1727-1730.

Tocci, M.D., Kelley, C.T., Miller, C.T., 1997. Accurate and economical solution of the pressure-head form of Richards' equation by the method of lines. Adv.Water Resour. 20, 1-14.

Uhlenbrook, S., Roser, S., Tilch, N., 2004. Hydrological process representation at the meso-scale: the potential of a distributed, conceptual catchment model. J.Hydrol. 291, 278-296.

Vache, K.B., McDonnell, J.J., 2006. A process-based rejectionist framework for evaluating catchment runoff model structure. Water Resour.Res. W02409, doi:10.1029/2005WR004247.

Wallman, P., Belyazid, S., Svensson, M.G.E., Sverdrup, H., 2006. DECOMP - a semi-mechanistic model of litter decomposition. Environmental Modelling and Software 21, 33-44.

Wallman, P., Svensson, M.G.E., Sverdrup, H., Belyazid, S., 2005. FORSAFE - an integrated process-oriented forest model for long-term sustainability assessments. For.Ecol.Managem. 207, 19-36.

Wassmann R, Papen H, Rennenberg H (1993) Methane emission from rice paddies and possible mitigation strategies. Chemosphere 26:201–217. doi: 10.1016/0045-6535(93)90422-2

Werner C, Butterbach-Bahl K, Haas E, et al. (2007) A global inventory of N2O emissions from tropical rainforest soils using a detailed biogeochemical model. Global Biogeochem Cy 21:–. doi: 10.1029/2006GB002909

Whitehead, P.G., Wilson, E.J., Butterfield, D., 1998. A semi-distributed integrated flow and nitrogen model for multiple source assessment in catchments (INCA): Part I - model structure and process equations. Sci.Total Environ. 210, 547-558.

Wigmosta, M.S., Vail, L.W., Lettenmaier, D.P., 1994. A distributed hydrology-vegetation model for complex terrain. Water Resourc.Res. 30, 1665-1679.

Williams, J.R., Jones, C.A., Dyke, P.T., 1984. A modeling approach to determining the relationship between erosion and soil productivity. Trans.Am.Soc.Agricult.Eng. 27, 129-144.

Wolf B, Chen W, Brüggemann N, et al. (2010) Applicability of the soil gradient method for estimating soil-atmosphere CO2, CH4, and N2O fluxes for steppe soils in Inner Mongolia. J Plant Nutr Soil Sci 174:359–372. doi: 10.1002/jpln.201000150

Zehe, E., Maurer, T., Ihringer, J., Plate, E., 2001. Modeling water flow and mass transport in a loess catchment. Physics and Chemistry of the Earth Part B-Hydrology Oceans and Atmosphere 26, 487-507.

Danksagung

Zahlreichen Menschen bin ich für die Unterstützung meiner Promotion dankbar. Zuallererst möchte ich Prof. Hans-Georg Frede und Dr. Lutz Breuer für die wunderbare Betreuung danken. Durch ihr Vertrauen in meine Arbeit konnte sie werden, was sie jetzt ist. Herrn Prof. Luterbacher danke ich für die Übernahme des zweiten Gutachtens.

Dem EU Projekt NitroEurope IP, der Deutschen Forschungsgemeinschaft und dem Institut für Landschaftsökologie und Ressourcenmanagement danke ich für die Finanzierung meiner Arbeit.

Viele Kollegen haben mir mit Rat und einem offenen Ohr zur Seite gestanden: Besonders möchte Prof. Kellie Vaché danken für die vielen Diskussionen über die Ausrichtung der Arbeit. Über ihn konnte ein Forschungsaufenthalt an der Oregon State University realisiert werden (finanziert durch die DFG) bei dem ich durch Prof. J.J. McDonnell und seiner Arbeitsgruppe wertvolle Erfahrungen machen konnte.

Aus dem Institut für Meteorologie und Klimaforschung, des Karlsruhe Institute of Technology in Garmisch-Partenkirchen, insbesondere von Dr. Edwin Haas und Prof. Klaus Butterbach-Bahl, aber auch von Alexander Fröhlich, Steffen Klatt, Dr. Ralf Kiese, Dr. Christian Werner und Dr. Rüdiger Grote bekam ich wichtige Anregungen und aufmunterndes Interesse, vor allem zu der Modell-Kopplung und zu mathematischen Problemen.

Am Institut für Landschaftsökologie und Ressourcenmanagement muss ich zuallererst Ruth Strittmatter dafür danken, dass sie alle administrativen Problemstellungen herzlich und direkt löste. Die fachliche und persönliche Unterstützung meiner ehemaligen Kollegen Dr. Frauke Barthold, Dr. Katrin Schneider und Dr. Jean-François Exbrayat bleibt unvergessen. David Windhorst, Sebastian Multsch und Ina Plesca haben sich vom „Modelling Framework“-Fieber anstecken lassen, und mich dadurch in meinen Ideen zur Doktorarbeit bestärkt. Insgesamt erlebe ich die fachlichen Diskussionen, das entspannte Arbeitsklima und den soziale Zusammenhalt am Institut als etwas besonderes, dafür möchte ich allen Kollegen danken.

Meinem früheren Arbeitgeber, der ÖKO-DATA GmbH in Strausberg möchte ich für die Erfahrungen, den Raum den ich für meine Weiterentwicklung hatte, und natürlich auch den Freundschaften die entstanden sind, danken. Die Zusammenarbeit mit Hans-Dieter Nagel, Angela Schlutow, Gudrun Schütze und Regine Weigelt-Kirchner prägt mich bis heute.

Klaus und Irmela und besonders Kathrin und meinen Kindern Helen und Paul: Danke.

Erklärung

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.

Gießen, im Januar 2012

(Philipp Kraft)