Component based environmental modelling using the JAMS framework

S. Kralisch, P. Krause, M. Fink, C. Fischer and W.-A. Flügel

Department of Geoinformatics, Hydrology and Modelling, School of Chemical and Earth Sciences
Friedrich-Schiller-University, Jena, Germany. E-Mail: sven.kralisch@uni-jena.de

Keywords: *Modelling Frameworks, Integrated Water Resources Management, Environmental Modelling*

EXTENDED ABSTRACT

Current challenges in sustainable water resources management have created demand for integrated, flexible and easy to use environmental simulation models which are able to simulate the quantitative and qualitative aspects of the hydrological cycle with a sufficient degree of certainty. Along with these models various accompanying applications are needed in order to support the setup, application and evaluation of models or the search for optimized management alternatives. Among them are tools for automated parameterization and sensitivity analyses of models and graphical user interfaces that support model setup and visualization of simulation results.

Existing applications which have been developed to cover these demands are often constrained to specific scales or purposes and can not easily be adapted to meet different challenges. Reasons for this can be a monolithic software architecture that hinders implementation of extensions, constraining software licenses that permit modifications to the system or simply a closed code base. As a consequence resulting from this shortage, modelling frameworks that allow an easy implementation and extension of component based simulation models have gained increasing attention from both model developers and users during the last years.

This paper presents various applications of the Jena Adaptable Modelling System (JAMS) that cover the abovementioned subjects while focusing especially on the reuse of existing modelling components. First we show how spatial and temporal domains typical for environmental models can be represented in JAMS (figure 1). Then we describe the implementation of a distributed hydrological model and its application in more detail. Following, we point out how the hydrological model can be extended by new components in order to allow the simulation of nutrient dynamics. After a brief description of the single components, we present results of their application to a meso-scale catchment. The last part of the paper focuses on parameter optimization procedures in JAMS models. After briefly showing how optimization algorithms can be represented as JAMS components, results of an application of an evolutionary optimization procedure to the hydrological model discussed earlier are presented.

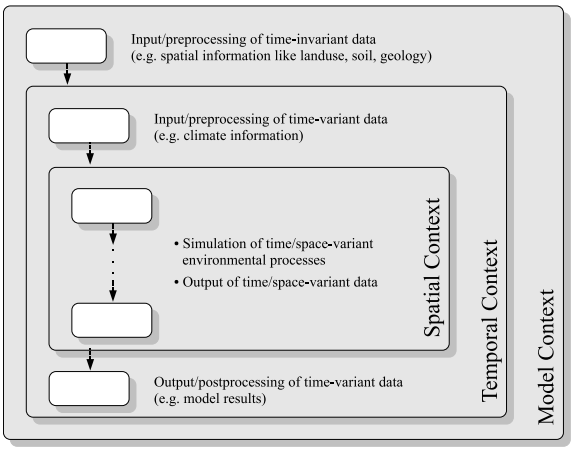

Figure 1. Common structure of distributed environmental simulation models in JAMS

1 INTRODUCTION

Problems of integrated water resources management (IWRM) like e.g. the implementation of the European Water Framework Directive (European Union, 2000) demand flexible software systems that can be used for both oriented development of environmental simulation models as well as their application. Such requirements are met by modelling frameworks, which can be used *(i)* to define a software architecture (e.g. by means of interfaces and abstract classes) that specifies how modelling components inside the framework can interact with each other and *(ii)* to control the creation, linkage and execution of such components.

Environmental modelling frameworks which are suited to cover the special demands of environmental simulation models are usually characterized by some of the following additional properties:

- They provide special data types that can be used to describe the spatial and temporal domains.
- The control flow inside these frameworks is not fixed but can be configured from outside, e.g. by means of special components.
- They provide functions for managing and manipulating environmental data, e.g. for reading and writing time series data or for unit conversion.

The Jena Adaptable Modelling System (JAMS) is one of those frameworks that features these properties (Kralisch and Krause, 2006). During the last two years a number of JAMS modelling components have been developed that cover various aspects of IWRM, particularly hydrological modelling, nutrient modelling, parameter optimization and visualization of model results. The following sections will provide an overview of these components and examples of their application to IWRM problems.

2 DISTRIBUTED CATCHMENT MODELS

2.1 Intended model structure

JAMS has been developed with the main objective to create models that can simulate environmental processes at discrete points in time and/or space. This approach is widely-used by many distributed hydrological models applied in current practice.

JAMS provides two specific type of building blocks, named *components* and *context components*. Components are used to implement specific knowledge and/or process algorithms whereas the model structure is defined by the context components. An environmental process (e.g. potential evapotranspiration) can be implemented as a JAMS component without any knowledge about it's later execution context, e.g. the temporal resolution or the type of spatial discretisation of the modelled area. The only precondition for the later application of a component in conjunction with others is the proper declaration of it's desired input data and provided output data by means of metadata in the component's source code.

Context components in JAMS define the model structure that is needed to represent the necessary temporal and spatial iteration loops. These software components can be used to manage the repeated execution of other components while varying their sets of input data. Each set of input data may then represent one point in time (*temporal context*) or space (*spatial context)*. A context component in JAMS can thus be seen as a scope that defines an environment for the execution of other software components. Together with another context (*model context*) that manages the execution of the real simulation, the above-mentioned environmental models can easily be produced.

Figure 1 shows the different contexts and the workflow usually executed within them. Other than shown in this illustration, a JAMS model can include more than only one temporal or spatial context. As an example, a simulation model considering hydrological processes in the river channel could additionally include a second spatial context managing the iteration over river reaches and the execution of appropriate process components at a time.

Figure 2 shows an example of the interaction of components and context components. Here a component is depicted which is executed during model simulation *(i)* only once (left), *(ii)* for each time step (middle), and *(iii)* for each time step and each spatial object (e.g. raster cell or polygon) (right).

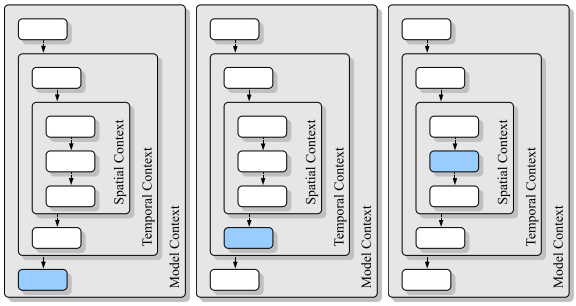

Figure 2. Different execution contexts for the same component

The JAMS contexts allow a flexible structuring of environmental models while reusing existing components or whole submodels, which will be

shown by examples in the following sections.

2.2 Hydrological modelling

Water is one of the most important environmental factors. Because of this paramount importance hydrological modelling has a long history. Computer models for hydrological purposes exist since the end of the 60s of the last century and up today an discountable number of models for various hydrological purposed had been developed. Lumped, semi-distributed or distributed process-oriented conceptual models are the most widely used model types in practice. Despite the large number of such models only a limited number of different concepts and methods for the interpretation of the hydrological cycle are used inside them, with a sometimes tremendous overlap between different models.

JAMS was used in the last two years to implement a number of existent hydrological process concepts as well as for the development of entire new models and model components. Hydrological models which had been implemented are the Thornthwaite water balance model (Thornthwaite, 1948), HYMOD based on Moore (1985), J2000 (Krause, 2001), parts of WASIM-ETH (Schulla and Jasper, 1999) and PRMS (Leavesley et al., 1983), and the SNOW17 processes of Anderson (1973). Hereby a solid library of hydrological process components was developed which can be used as a basis for a problem tailored model development.

Such a development is the model J2000g which was recently implemented as a JAMS model. The reason for the development was the need of an easy to use, robust but simple model applicable for the entire German state of Thuringia with an area of 16 172 km^2. The constraints for the model were: *(i)* continuous modelling in daily or monthly time steps, *(ii)* applicable for the entire state but also for single catchments, *(iii)* process oriented model concept, *(iv)* robust with only a small number of calibration parameters, *(v)* applicable with historical time series but also with climate scenario data for the next 100 years, and *(vi)* flexible distribution concept for the area or interest.

As the name implies J2000g has been developed as a derivative of the original J2000 model which was considered as too complex for the envisaged goal. Of the 33 components in the new model, only 5 needed to be programmed from scratch whereas the remaining 28 components could be directly taken from the existing component library, which made the development and implementation a very rapid process. The model itself can be classified as a simplified distributed conceptual hydrological water balance model. To provide the model with spatial distributed knowledge, information about topography, landuse, soil types and hydrogeology of the area of interest is needed, which is used to describe the physio-geographical properties of each modelling unit. These units can be of various types, e.g. entire catchments (i.e. lumped version), subbasins, response units but also raster cells or even a mixture of different types. To drive the model, measured climate data (precipitation, minimum, average, maximum air temperature, sunshine duration, wind speed and relative humidity) from at least one climate station is needed. This point information is transferred to each model unit with the regionalisation approach of J2000, which takes horizontal and vertical variability of the climate data into account. The regionalised information is then used for simulating the hydrological processes as shown in figure 3.

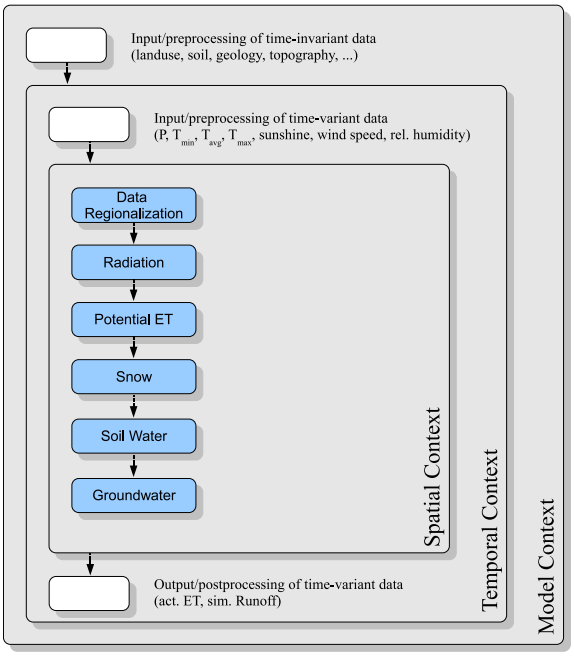

Figure 3. Simplified illustration of the J2000g layout

The regionalised data is used for the radiation calculation following the FAO guideline of Allen et al. (1998) which is providing net radiation for the subsequent potential evapotranspiration module. This module implements the approach of Penman-Monteith also following the guideline of Allen et al. (1998) and computes the potential ET for each unit and each time step with the actual weather conditions and the unit's vegetation type.

Snow accumulation and melt processes are considered by a simple approach which calculates snow accumulation whenever the arithmetic mean of the minimum and average air temperature is lower than a user definable temperature threshold (`Tbase`). Snow melt is computed with a time-degree-factor (`TMF`) and the arithmetic mean of average and maximum air temperature.

The soil water module comprises a simple water tank which is filled by precipitation and snow melt and emptied by evapotranspiration. As this behaviour corresponds to the pore volume of the field capacity, this parameter is used to determine the storage's maximum capacity. For calibration purposes the entire distribution of storage capacity values of all modelling units can be modified by a constant multiplier (`FAC`). The retrieval of water from the soil tank by ET is governed by the actual saturation, the potential ET and a calibration coefficient `ETR`. This coefficient defines a specific threshold of the actual soil water saturation which triggers a linear reduction of the maximum ET rate whenever soil water saturation falls below `ETR`. Runoff occurs when the soil water capacity is saturated. From this point on any surplus of water is distributed into two runoff components: direct runoff and percolation. The relative part for each component is computed by the slope of the model unit and a calibration factor `LVD`. Both runoff components are transferred to catchment wide storages. Originating from these tanks direct runoff and baseflow is calculated with linear outflow functions and one calibration coefficient each, to account for retention from runoff generation. If the model is applied in monthly time steps the storage of the direct runoff tank is not held back at all.

The model is applied operationally for the Thuringian Environmental Agency for quantifying the state wide historical and future hydrological balance and dynamics. Model calibration and validation was carried out in a number of selected typical basins with areas between 13 and 5940 km^2. The model performance was quantified by the comparison of the model result with measured streamflow values and produced Nash-Sutcliffe efficiencies between 0.6 and 0.85 which can be considered as reasonably good when taking the model's simplicity into account. An example of the simulated monthly runoff for the 150 km^2 large Bode basin is given in figure 4.

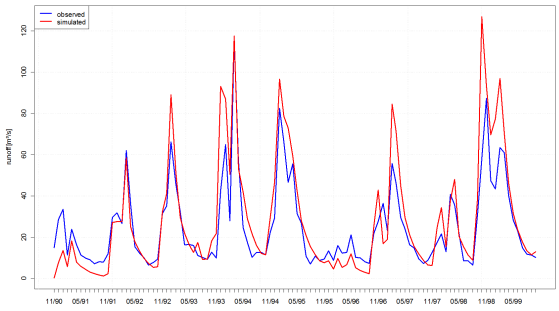

Figure 4. Observed and simulated monthly runoff achieved with J2000g in the Bode basin.

2.3 Nutrient modelling

Nitrogen is an important nutrient in landscapes and also shows a highly complex turnover dynamic. Due to these facts this element is the subject of intense research. The nitrogen cycle is a dynamic system that includes the water, atmosphere and soil. In order to simulate these processes in the context of a JAMS model, descriptions from the SWAT 2000 model (Arnold et al., 1998; Neitsch et al., 2002) have been utilized. For the description of the hydrological dynamics the components of J2000 (Krause, 2001) were used. The calculation of the nutrient dynamics is separated into four main components (i.e. landuse management, plant growth, soil temperature and soil nitrogen balance), which will be described briefly in the following paragraphs.

The component for the landuse management has been implemented according to the description in the SWAT model. Basically, the management operations that control the plant growth cycle, the time of planting, fertilizer application and the removal of plant biomass are controlled by this component. This component is also able to describe the management of complex crop rotations.

The description of the plant growth and plant nutrient uptake is based on the SWAT model as well. The phenological development of the plant is controlled by the sum of the difference of the mean air temperature and a plant specific base temperature. The development of the leaf area index (LAI) and the root depth is a function of this calculated temperature sum. The production of biomass is simulated by the LAI dependent light interception and the conversion of intercepted light into biomass, given the plant species-specific radiation use efficiency. Moreover, the calculated total biomass is differentiated into the root development and above ground biomass, each simulating the N-uptake, residues and yield. This potential biomass production is limited by three different stress factors: temperature, water and nitrogen stress. To gather the information for the plant growth simulation, interactions with the land use management, soil water and soil nitrogen components are required.

The bio-chemical processes in the soil are strongly influenced by the soil's temperature. Soil temperature is calculated depending on air temperature and global radiation with empirical equations (Neitsch et al., 2002; Williams et al., 1984) and is damped compared to air temperature because of aboveground biomass, snow cover, soil matrix and soil water. The damping increases with soil depth down to the lower border of the simulation domain for which the average annual air temperature is assumed to be representative. For the calculation various informations of other

components are necessary, i.e. the soil water content from the soil water module, the snow depth from the snow module and aboveground biomass from the plant growth module. For the simulation of nitrogen balance in the soil horizons, five different nitrogen pools are considered: nitrate, ammonium, stable organic, active organic and plant residue (Neitsch et al., 2002). The structure of the resulting J2000-S nutrient simulation model is shown in figure 5.

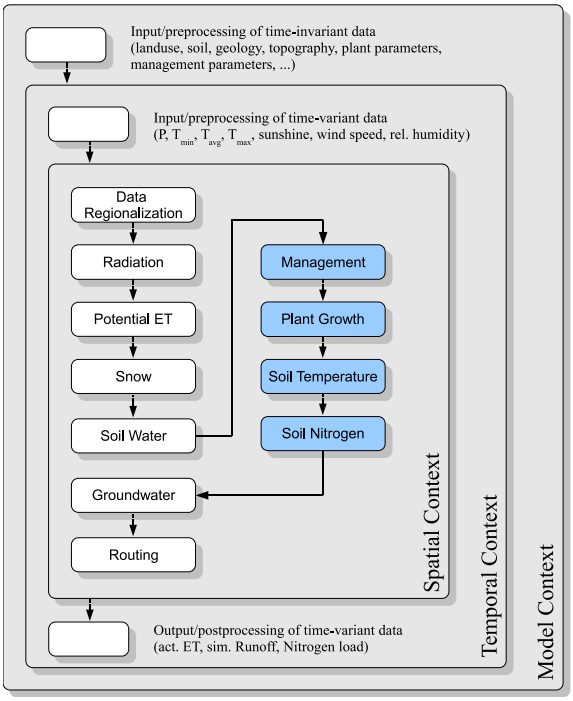

Figure 5. Simplified illustration of the J2000-S layout with nutrient simulation components

For testing and validation of the J2000-S, the 844 km^2 large upper Gera catchment located in Thuringia, Germany was selected. The runoff modelled with the J2000-S (figure 6) showed a visual good fit which is also confirmed by the Nash-Sutcliffe efficiency of 0.71 and the nearly perfect match of the simulated long-term mean runoff (5.9 m^3/s) with the observed one (5.8 m^3/s). The figure shows that the model tends to overpredict single peak-flow events. This can be explained partly by the influence of the various water reservoirs in the catchments which are not considered by the model. The representation of the simulated nitrogen-load dynamics (figure 7) shows a good fit as well, with a coefficient of determination of 0.64. The long-term modelled average nitrate concentration of 23.9 mg/l is close to to the observed value of 22.2 mg NO_3/l. These results show that J2000-S is able to describe the water and nitrogen dynamics in meso-scale catchments with a satisfying degree of determination. More details of this study are given in Fink et al. (2007).

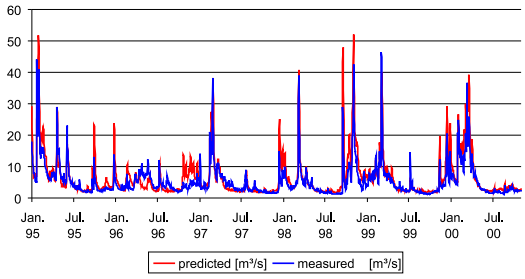

Figure 6. Predicted and measured discharge at the catchment outlet (Fink et al., 2007)

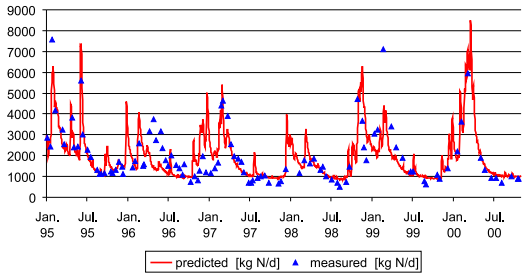

Figure 7. Predicted and measured nitrogen load at the catchment outlet (Fink et al., 2007)

3 PARAMETER OPTIMIZATION

The parameters of conceptional hydrological models often can not be measured or directly determined a priori. This makes it necessary to calibrate a model by choosing the parameters in such a way that the behaviour of a catchment is simulated as best as possible. Because manual calibration can be a time consuming task, the use of optimization algorithms which can search automatically for such best parameter values is more efficient.

Applications of traditional optimization methods (e.g. uniform random search, gradient descent methods or the newton method) showed that the search for a global best solution with these algorithms is hard if not impossible (Sorooshian and Gupta, 1983). The most common problems are the existence of multiple local optima, incontinuities, and the often large number of parameters to optimize. This failure of traditional approaches led to the development of new methods that where capable to handle the occuring difficulties.

In 1992 the SCE-UA (Shuffle Complex Evolution - University of Arizona) algorithm was introduced by Duan et al. (1992). This evolutionary optimization method was developed especially for the application of parameter optimization in hydrological models and designed with special focus on efficiency, effectivity and modern standards.

The core idea of evolutionary optimization methods is to treat the optimization problem as a natural evolutionary process. Main subject of SCE-UA is therefore a population of samples – each of them representing one solution candidate. This population is divided into complexes that evolve independently from each other. In order to create new samples, subcomplexes are formed which act as parents. While generally every subcomplex is able to produce a new sample, especially promising subcomplexes are preferred in this reproduction step. The produced children must fulfill some minimum requirements before they are added to the population and supersede the currently worst sample. Their fitness can be determined with various objective measures (e.g. Nash-Sutcliffe efficiency), demanding the model to be evaluated every time. After some iterations of reproduction, the complexes are joined. This guarantees that the information of each initial population is beeing used most efficiently. The process of complex segmentation and reproduction is repeated until no further improvement of the samples fitness can be accomplished.

The SCE-UA algorithm shows good convergence for a vast variety of problems, meaning that with a sufficiently high (fixed) number of model evaluations the method has a fairly high probability to find the global optimum. In order to make it available for JAMS-based models, SCE-UA was implemented as a JAMS component based on the MATLAB sources provided by the SARAH hydroarchive (http://www.sahra.arizona.edu). Since the method must be able to execute the underlying model whenever it is necessary (i.e. for each evaluation of a sample's fitness), SCE-UA was implemented as a context component. Due to the flexibility of the JAMS framework and the generality of the SCE-UA method, nearly no knowledge about the model is needed. Only the relevant model parameters, the respective objective function value and the number of complexes must be made available to the SCE-UA component. According to Duan et al. (1992), the population size is determined based on the number of complexes and parameters.

The resulting JAMS model structure is shown in figure 8. The SCE-UA context component encloses the model to optimize and can access its parameter values and controls its execution. Since every search based optimization procedure comprises the two tasks search and evaluation, this setup depicts a general structure for any search based parameter optimization component in JAMS. It has also been utilized for the implemenation of other optimization procedures, namely a modified gradient descent method for testing purposes, and the algorithms MOCOM (Yapo et al., 1998) as well as MOSCEM-UA (Vrugt et al., 2003) which allows handling of multi-objective global optimization problems.

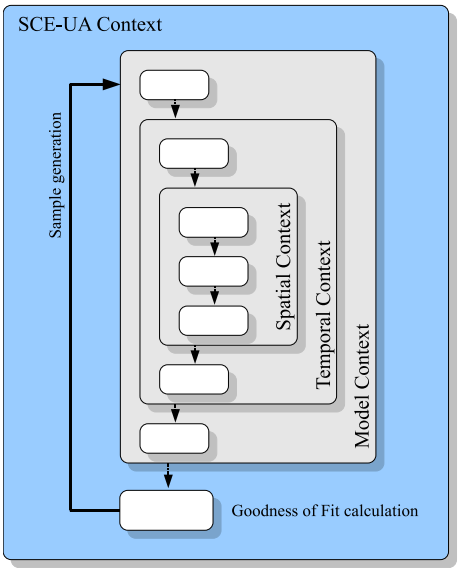

Figure 8. Model structure for search based parameter optimization

The SCE-UA method has been applied to calibrate the six parameters of the above mentioned J2000g model. The component was configured to use 5 complexes, the Nash-Sutcliffe efficiency was chosen as objective function. A satisfying parameter set was found after only 76 evaluations of the objective function. After a further slight improvement the optimization finally terminated after 1096 model executions because convergence had been achieved. The finally obtained Nash-Sutcliffe efficiency value was 0.84. Figure 9 shows the efficiency of the best sample found depending on the iteration step and demonstrates the fast movement of the objective function's value towards the optimum.

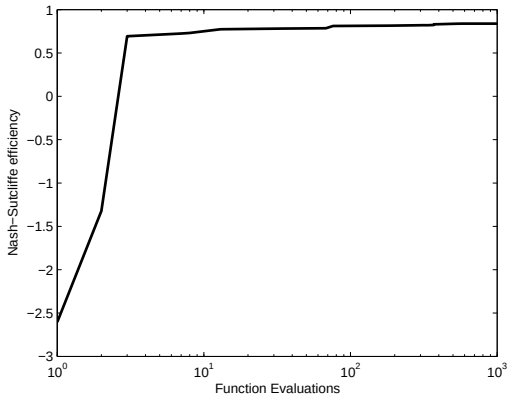

Figure 9. Development of Nash-Sutcliffe efficiency during SCE-UA application

4 CONCLUSIONS

The JAMS modelling framework described in this paper can be used to build complex environmental simulation models from single, easy manageable components. We have shown how the temporal and spatial domains can be represented in this system and applied it to different environmental modelling problems in order to prove its flexibility. First we created a distributed hydrological model, then we added a number of components that allow the simulation of the nutrient dynamics based on the original model. Finally, we pointed out how optimization procedures can be implemented as a JAMS component in order to calibrate the original hydrological model. These applications of the JAMS framework show that this system can be used to easily create problem tailored models for a broad range of environmental problems while reusing existing solutions.

5 ACKNOWLEDGEMENT

We would like to thank the German Federal Ministry of Education and Research for funding the JAMS development in the course of the program "Entrepreneurial Regions – the BMBF Innovation Initiative for the New German Länder".

References

Allen, R. G., Pereira, L., Raes, D., and Smith, M. (1998). Crop evapotranspiration: Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56, FAO, Rome.

Anderson, E. A. (1973). National Weather Service River Forecast System – Snow Accumulation and Ablation model. Tech. mem. nws-hydro-17, National Oceanographic and Atmospheric Administration (NOAA).

Arnold, J., Srinivasan, R., Muttiah, R., and Williams, J. (1998). Large Area Hydrologic Modelling and Assessment. Part I, Model Development. *Journal of the American Water Resource Association*, 34(1):7389.

Duan, Q., Sorooshian, S., and Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28(4):1015–1031.

European Union (2000). Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 – Establishing a framework for community action in the field of water policy. *Official Journal L 327*, pages 0001–0073.

Fink, M., Krause, P., Kralisch, S., Bende-Michl, U., and Flügel, W.-A. (2007). Development and Application of the Modelling System J2000-S for the EU-Water Framework directive. *Advances in Geosciences*, 11:123130.

Kralisch, S. and Krause, P. (2006). JAMS – A Framework for Natural Resource Model Development and Application. In Voinov, A., Jakeman, A., and Rizzoli, A., editors, *Proceedings of the iEMSs Third Biannual Meeting "Summit on Environmental Modelling and Software"*, Burlington, USA.

Krause, P. (2001). *Das hydrologische Modellsystem J2000: Beschreibung und Anwendung in großen Flußeinzugsgebieten*. Schriften des Forschungszentrum Jülich : Reihe Umwelt/Environment; Band 29.

Leavesley, G. H., Lichty, R. W., Troutman, B. M., and Saindon, L. G. (1983). The Precipitation Runoff Modeling System – PRMS: User's manual. Water Resources Investigation Report 83-4238, USGS, Denver, Colorado.

Moore, R. J. (1985). The Probability-Distributed Principle and Runoff Production at Point and Basin Scale. *Hydrological Sciences Journal*, 30(2):273–297.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., Williams, J. R., and King, K. W. (2002). Soil and Water Assessment Tool - Theoretical Documentation - Version 2000. TWRI Report TR-191, Texas Water Resources Institute, Temple.

Schulla, J. and Jasper, K. (1999). Model Description WaSiM-ETH (Water balance Simulation Model ETH). Technical report, ETH, Zurich, CH.

Sorooshian, S. and Gupta, V. (1983). Automatic Calibration of Conceptual Rainfall-Runoff Models: The Question of Parameter Observability and Uniqueness. *Water Resources Research*, 19(1).

Thornthwaite, C. W. (1948). An Approach toward a Rational Classification of Climate. *Geographical Reviews*, 38:55–94.

Vrugt, J., Gupta, H., Bastidas, L., Bouten, W., and Sorooshian, S. (2003). Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resources Research*, 39(8):1214.

Williams, J. R., Jones, C. A., and Dyke, P. T. (1984). EPIC, The Erosion-Productivity Impact Calculator, Volume I., Model documentation. Report, United States Department of Agriculture, Agricultural Research Service, Temple.

Yapo, P., Gupta, H., and Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. *Journal of Hydrology*, 204(1):83–97.