Calibration Methodologies in Hydrological Modeling: State of the Art

P. Droogers
W.W. Immerzeel

NIVR
Nederlands Instituut voor Vliegtuigontwikkeling en Ruimtevaart
Netherlands Agency for Aerospace Programmes

Research paper no. 2

IWMI
International
Water Management
Institute

Future Water
Science for Solutions

P.O. Box 35
2600 AA Delft
The Netherlands
Kluyverweg 1
2629 HS Delft
The Netherlands
Tel. +31 (0)15 278 80 25
Fax +31 (0)15 262 30 96
Email info@nivr.nl
www.nivr.nl

NATIONAL USER SUPPORT PROGRAMME (NUSP) 2001-2005

http://www.ao-go.nivr.nl

The National User Support Programme 2001-2005 (NUSP) is executed by the Netherlands Agency for Aerospace Programmes (NIVR) and the Space Research Organization of the Netherlands (SRON). The NUSP is financed from the national space budget. The NUSP subsidy arrangement contributes to the development of new applications and policy-supporting research, institutional use and use by private companies.

The objectives of the NUSP are:

- To support those in the Netherlands, who are users of information from existing and future European and non-European earth observation systems in the development of new applications for scientific research, industrial and policy research and operational use;

- To stimulate the (inter)national service market based on space-based derived operational geo-information products by means of strengthening the position of the Dutch private service sector;

- To assist in the development of a national Geo-spatial data and information infrastructure, in association with European and non-European infrastructures, based on Dutch user needs;

- To supply information to the general public on national and international space-based geo-information applications, new developments and scientific research results.

Calibration Methodologies in Hydrological Modeling: State of the Art

Authors: P. Droogers
W. Immerzeel

This research was funded by: NIVR/SRON/NWO User Support Programme 2

Generaal Foulkesweg 28 ● 6703 BS Wageningen ● Tel: 0317-460050
E-mail: p.droogers@futurewater.nl ● Website: www.futurewater.nl

Preface

This report contributes to a research project undertaken by FutureWater entitled: "Remotely Sensed based hydrological model calibration for basin scale water resources planning: embedding case for Krishna Basin, India" (GO-2005/025). This project is financially supported by NIVR (Nederlands Instituut voor Vliegtuigontwikkeling en Ruimtevaart) in the context of "Tijdelijke subsidieregeling Nationaal Programma Gebruikers Ondersteuning (GO).

The report describes the state of the art in hydrological model calibration and will serve as starting point for the research to be undertaken.

FutureWater
Generaal Foulkesweg 28
6703 BS Wageningen
The Netherlands
tel: + 31 (0)317 460050
email: info@futurewater.nl

Table of contents

1 INTRODUCTION 5

2 HYDROLOGICAL MODELLING 7

2.1 Introduction 7
2.2 Concepts of modeling 10
2.3 Model classification 10
2.4 Existing model overviews 11
2.5 Model reviews 12
2.6 SWAT model 14

3 CALIBRATION IN HYDROLOGICAL MODELLING 17

3.1 Introduction 17
3.2 Objective function 18
3.3 Optimization algorithm 20
3.4 Termination criteria 22
3.5 Calibration data 23
3.6 Parameters to be optimized 23
3.7 Calibration tools 25
3.7.1 PEST 25
3.7.2 UCODE 26
3.7.3 MATLAB 26
3.7.4 GLOBE 27

4 CONCLUSIONS 29

5 REFERENCES 31

1 Introduction

The number of major water related disasters such as droughts and floods is on the rise, as well as the number of people affected, total loss in lives, and economic damage. Improved water management is highly required and it is evident that the current focus on day-to-day management should shift to more strategic planning. A key issue in this strategic planning of water resources is the ability to have proper planning and management tools available. These tools can be divided in Operational and Strategic Decision Support Systems: O-DSS and S-DSS, both relying on simulation models that can mimic reality. The O-DSS are mainly hydraulic oriented models able to predict on a time scale from hours to days in high detail how water will flow in river and canal systems, relying on accurate flow measurements upstream to predict timing and quantity of water downstream. These systems have been very effective in reducing the number of fatalities by so-called early-warning systems, but are not very helpful in a more strategic planning of water resources management.

The Strategic Decision Support Systems (S-DSS) have a much more hydrological focus, describing the entire water cycle including natural as well as human induced processes. The objective of these tools is not to predict as accurate as possible the time a flood or drought might occur, but the probability of exceedance of these events and what long-term options might be feasible to reduce these risks. Besides this capability to estimate these extremes, these S-DSS are extremely powerful in evaluating the impact of changes in water management such as reservoir building, changes in water allocation between and within sectors, and impact of climate change.

Despite substantial progress in the development of these S-DSS techniques, the weakest part is currently the lack of data to apply and calibrate these S-DSSs. Traditionally, these S-DSSs are fine-tuned by a calibration process where observed hydrographs are compared to simulated ones. By adjusting the most sensitive and most unreliable input parameters the S-DSS can be calibrated and performs better in describing the current situation and is therefore also more reliable to explore water management options for the future. It must be emphasized here that this is the standard practice for almost every hydrological modeling study. It is clear that in data scarce areas such an approach, which requires observed streamflow data, is impossible and analyses are therefore often based on non-calibrated models, resulting in erroneous output.

The focus of the GO research project "Remotely Sensed based hydrological model calibration for basin scale water resources planning, India" is to use an innovative model calibration procedure based on Remotely Sensed evapotranspiration data. This report discusses the use of hydrological simulation models in general and specifically focuses on calibration procedures.

2 Hydrological modelling

2.1 Introduction

The increasing water scarcity, the growing demand for food, and the need to link those two in a sustainable way is the challenge for the next decades. Seckler et al. (1999) estimated that by 2025 cereal production will have to increase by 38% to meet world food demands. The World Water Vision, as outcome from the Second World Water Forum in The Hague in 2000, estimated a similar figure of 40% based on various projections and modeling exercises (Cosgrove and Rijsberman, 2000). These figures were more or less confirmed by projections based on an econometric model which showed that the rate of increase of grain production will be about 2% per year for the 2000-2020 period (Koyama, 1998).

To produce this increasing amount of food substantial amounts of water are required. Global estimates of water consumption per sector indicate that irrigated agriculture consumes 85% from all the withdrawals and that this consumptive use will increase by 20% in 2025 (Shiklomanov, 1998). Gleick (2000) presented estimates on the amount of water required to produce daily food diets per region. According to his figures large differences can be found between regions ranging from 1,760 liters per day per person for Sub-Saharan Africa to 5,020 for North America. Differences come from the larger number of calories consumed and the higher fraction of water-intensive meat in the diet of a North American.

This increase in food, and therefore water, requirements coincide with a growing water scarcity at an alarming rate. Recently, a study by the United Nations (UN, 1997) revealed that one-third of the world's total population of 5.7 billion lives under conditions of relative water scarcity and 450 million people are under severe water stress. This relative water scarcity and severe water scarcity are defined using the Relative Water Demand (RWD) expressed as the fraction water demand over water supply. A RWD greater then 0.2 is classified as relative water scarce, while a RWD greater then 0.4 as severe water stress. However, these values as mentioned by the UN are based on national-level totals, ignoring the fact that especially in bigger countries, huge spatial differences can occur. Vörösmarty et al. (2000) showed that including these in-country differences 1.8 billion people live in areas with sever water stress. Using their global water model and some projections for climate change, population growth and economic growth, they concluded that the number of people living in sever water stress will have grown to 2.2 billion by the year 2025.

A study published by the International Water Management Institute (Seckler et al., 1999), based on country analysis, indicated that by the year 2025 8 percent of the population of countries studied (India and China where treated separately, because of their extreme variations within the country) will have major water scarcity problems. Most countries, which contain 80% of the study population, need to increase withdrawals to meet future requirements, and only for 12% of the population no actions are required.

Although the exact numbers on how sever water stress actually is, or will be in the near future, and how much more food we should produce, differ to a certain extent, the main trend is unambiguous: more water for food and water will be scarcer.

References given before are related to the global scale, but it is very clear that at smaller scales, such as basins, extreme variations will occur and many basins with tremendous water problems can be found. This, in combination with the "think globally, act locally" principle, makes the basin the most appropriate scale to focus on.

Data is essential to assess the current conditions of water resources and to explore trends in the past. However, to explore options for the future tools are required that are able to see the impact of future trends and how we can adapt to these in the most sustainable way. Simulation models are the appropriate tools to do these analyses. R.K Linsley, a pioneer in the development of hydrologic simulation at Stanford University wrote already in 1976:
"In summary then it can be said that the answer to 'Why simulate?' is given by the following points:

1. Simulation is generally more adequate because it involves fewer approximations than conventional methods.
2. Simulation gives a more useful answer because it gives a more complete answer.
3. Simulation allows adjustment for change which conventional methods cannot do effectively.
4. Simulation costs no more than the use of reliable conventional methods (excluding empirical formulae which should not be used in any case).
5. Data for simulation is easily obtained on magnetic tape from the Climatic Data Service or the Geological Survey.
6. No more work or time is required to complete a simulation study than for a thorough hydrologic analysis with conventional methods. Often the time and cost requirements are less.
7. In any case, if the time and cost are measured against the quality and completeness of the results, simulation is far ahead of the conventional techniques.
8. Even though the available data are limited, simulation can still be useful because the data are used in a physically rational computational program."

These points are still valid nowadays and can be more or less summarized by the two main objectives where models can be used for: (i) understanding processes and (ii) scenarios analyses. Understanding processes is something that starts right from the beginning during model development. In order to build our models we must have a clear picture on how processes in the real world function and how we can mimic these in our models. The main challenge is not in trying to build in all processes we understand, which is in fact impossible, but lies in our capabilities to simplify things and concentrate on the most relevant processes of the model under construction.

The main reason for the success of models in understanding processes is that models can provide output over an unlimited times-scale, in an unlimited spatial resolution, and for difficult to observe sub-processes. These three items are the weak point in experiments, but are at the same time exactly the components in the concept of sustainable water resources management. A typical example of the application of models to understand processes is shown in Figure 1. Soil moisture profiles are shown on a daily base (high temporal resolution), at every centimeter depth (high spatial resolution) and for relatively difficult measurable processes (soil moisture movement).

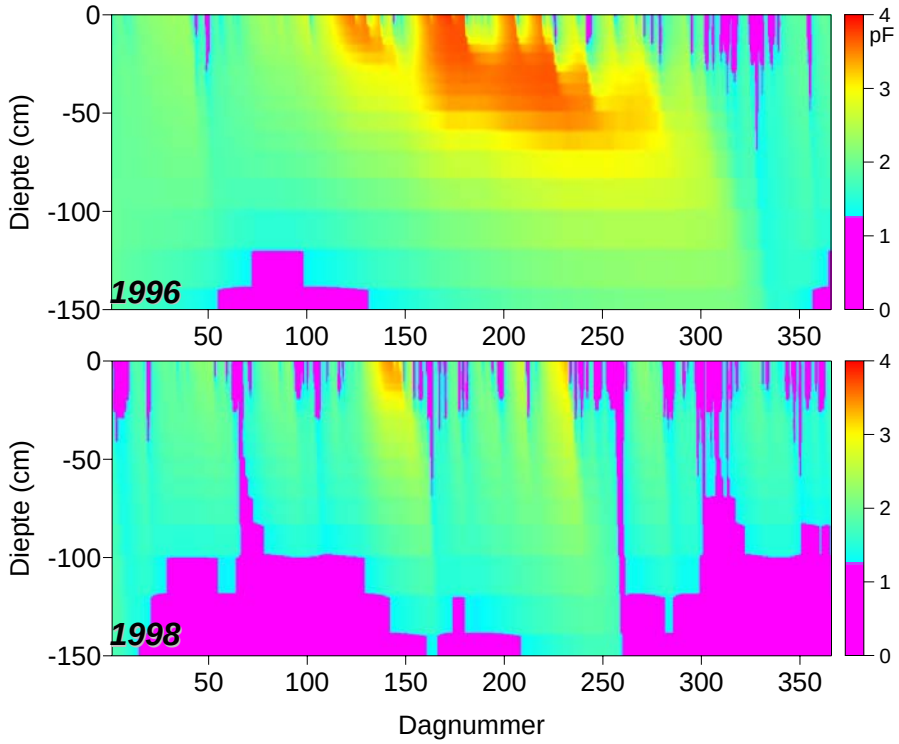

Figure 1. Daily soil moisture profiles for a typical Dutch polder soil. (Meijer et al., 2004)

The most important aspect of applying models however is in the use to explore different scenarios. These scenarios can refer to aspects that cannot directly be influenced, such as population growth and climate change. These are often referred to as projections. On the contrary to this are the so-called management scenarios where water managers and policy makers can make decision that will have a direct impact. Examples are changes in reservoir operations, water allocation and agricultural/irrigation practices. In other words: models enable to change focus from a re-active towards a pro-active approach. (Figure 2).

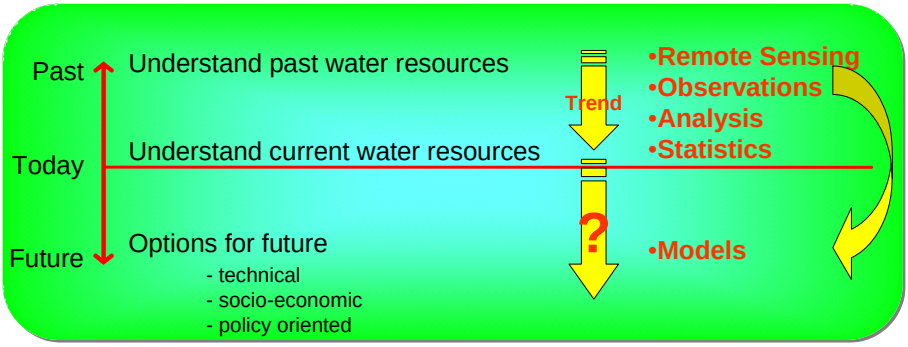

Figure 2. The concept of using simulation models in scenario analysis.

2.2 Concepts of modeling

The term modeling is very broad and includes everything where reality is imitated. The Webster dictionary distinguishes 13 different meanings for the word model where the following definition is most close to the one this study is focusing on: "a system of postulates, data, and inferences presented as a mathematical description of an entity or state of affairs". However we will restrict our definition here to computer models and that a model should have a certain degree of process oriented approach, excluding statistical, regression oriented models. This leads to the following definition: "a model is a computer based mathematical representation of dynamic processes".

The history of hydrological and agro-hydrological models, based on this somewhat restricted definition, is relatively short. One of the first catchment models is the so-called Stanford Watershed Model (SWM) developed by Crawford and Linsley in 1966, but the main principles are still used in nowadays catchment models to convert rainfall in runoff. SWM did not have much physics included as the catchment was just represented by a set of storage reservoirs linked to each other. The value of parameters describing the interaction between these different reservoirs was obtained by trying to optimize the simulated with the observed streamflows. At the other end of the spectrum are the field scale models describing unsaturated flow processes in the soil and root water uptake. One of the first models to be developed was the SWATR model by Feddes et al (1978) based on Richards' equation. Since, these models are based on points and use the concept that unsaturated flow is highly dominated by only vertical transport of water, much more physics could be built in from the beginning.

A huge number of hydrological models exits and applications are growing rapidly. The number of pages on the Internet including "hydrological model" is over 2.7 million (using Google on March 2006). A relevant question for hydrological model studies is therefore related to appropriate model selection. One of the most important issues to consider is the spatial scale to be incorporated in the study and how much physical detail to be included. Figure 3 illustrates the general relation between an increasing amount of physical detail and a decreasing amount of spatial detail.. The figure shows the position of commonly used models in this continuum. .

2.3 Model classification

The number of hydrological simulation models is unknown, but must be in the order of thousands. Even if we exclude the one-time models developed for a specific study and count only the more generic and more applied models it must exceed thousands. Some existing model overviews, as described later in more detail, include numerous models: IRRISOFT: 105, USBR: 100, CAMASE: 211, and REM: 675, amongst others. Interesting is that there seems to be no standard model or models emerging, as can be seen for example in groundwater modeling where ModFlow is the de-facto standard. Two hypotheses for this lack of standard can be brought forward. The first one is that model development is still in its initial phase, despite the about 25 years of history, and is therefore easy to start developing one's own model in a reasonable amount of time and effort that can compete with similar existing ones. A stimulating factor related to this is that a serious scientist is considered to have his/her own model or has at least developed one during his or her PhD studies. A second reason for the large number of models is a more fundamental one saying that hydrological processes are so complex and diverse that each case requires its specific model or set of models.

It is therefore interesting to see how models can be classified and see whether such a classification might be helpful in selecting the appropriate model given a certain question or problem to be solved. Probably the most generally used classification is the spatial scale the model deals with and the amount of physics included (Figure 3). These two characteristics determine other model behavior as data need, expected accuracy, required expertise, user-friendliness amongst others.

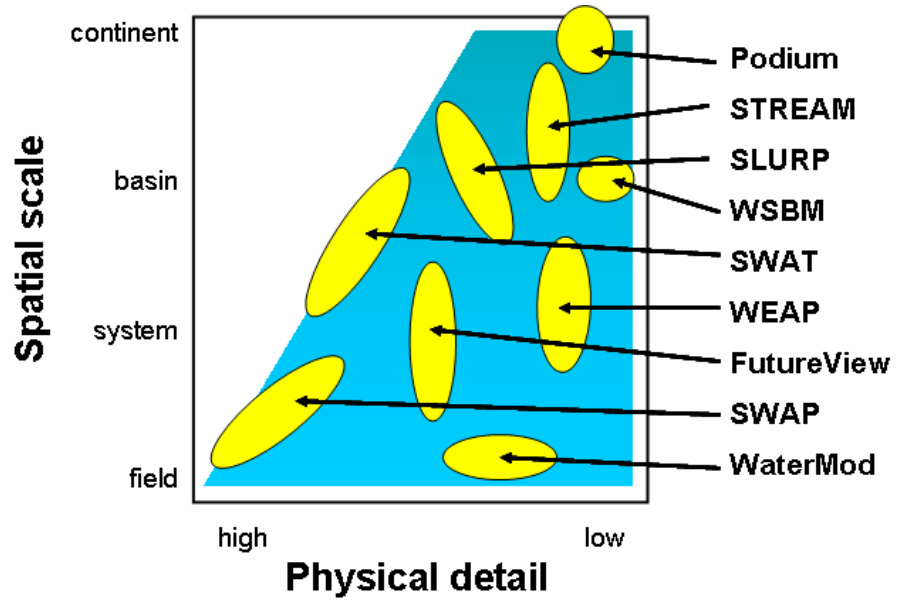

Figure 3. Spatial and physical detail of hydrological models.

2.4 Existing model overviews

A substantial number of overviews exist listing available models and a short summary. Most of this information is provided by the developers of the model themselves and tends therefore biased towards the capacities of the model. The most commonly used model overviews are discussed briefly here, keeping in mind that these overviews are changing rapidly, in size and number, since the Internet provides almost unlimited options to start and update such an overview in a automatic or semi-automatic way. A clear example is the Hydrologic Modeling Inventory project from the United States Bureau of Reclamation, where about 100 mainly river basin models are registered by model developers (USBR, 2002).

An overview of agro-ecosystems models is provided by a consortium named CAMASE (Concerted Action for the development and testing of quantitative Methods for research on Agricultural Systems and the Environment; CAMASE, 2005). The following types of models are distinguished: crop science, soil science, crop protection, forestry, farming systems, and land use studies, environmental science, and agricultural economics. A total of 211 models are included and for each model a nice general overview is provided. Unfortunately the last update of the register was in 1996 and advancements in model development over the last six years are not taken into account.

The United States Geological Survey (USGS, 2006) provides an overview of all their own models, about 50, divided in four categories: geochemical, ground water, surface water, water quality, and general. Some of the models are somewhat outdated, but some commonly used ones are included too. All the models are in the public domain and can be used without restrictions. For most of the models source code is provided as well.

The United States Department of Agriculture provides also models to be used in crop-water related issues. The National Water and Climate Center of the USDA has an irrigation page (NWCC, 2006) with some water management tools related to field scale irrigation.

United States Environmental Protection Agency is very active in supporting model development. The SWAT model, originating from their research programs, might have the potential to become the de-facto standard in basin scale modeling, and has been included in the BASINS package (BASINS, 2006). More linkages to models and other model overviews are provided too (EPA, 2006).

Modeling efforts of USGS, USDA, USACE, and EPA, combined with some other models, are brought together by the USGS Surface water quality and flow Modeling Interest Group (SMIG, 2006a). SMIG has setup the most complete link to models archives nowadays including links to 40 archives (SMIG, 2006b).

The most up-to-date overview of models used crop growth modeling is the Register of Ecological Models (REM, 2006), with 675 models as per 12-Dec-2005. Besides this overview of models the same website provides general concepts and links to modeling.

2.5 *Model reviews*

In the previous section an overview of existing model inventories has been given. Although useful as a catalog it does not provide any independent judgment of model quality. As argued before, the best model does not exist and is a function of the application and questions to be answered. Few studies have been undertaken where a limited amount of models have been tested and reviewed. The majority of these studies focuses on two or three models that are almost similar in nature and conclusions are that models are reasonable comparable.

Texas Natural Resource Conservation Commission evaluated 19 river basin models, referred to as Water Availability Models, in order to select the most suitable model used for management of water resources, including issuing new water right permits (TNRCC, 1998). A total of 26 evaluative criteria were identified as important functions and characteristics for selecting a model that fits the need for the 23 river basins in Texas. Most importantly was the ability of the model to supports water rights simulation. During the evaluation process, each model was assessed and ranked in order of its ability to meet each criterion. The 19 models were in the first phase narrowed down to five: WRAP, MODSIM, STATEMOD, MIKE BASIN, OASIS. Models not selected included WEAP (no appropriation doctrine) and SWAT (not intuitive and user-friendly). The final conclusion was to use the WRAP model with the HEC-PREPRO GUI. As mentioned, the study focused only on models able to assist in water rights questions.

A similar study was performed to select an appropriate river basin model to be used by the Mekong River Commission (MRC, 2000). In fact, it was already decided that considering the requirements of the MRC not one single model could fulfill the needs, but three different types of model were necessary: hydrological (rainfall-runoff), basin water resources, hydrodynamic. Three main criteria were used to select the most appropriate model: technical capability, user friendliness, and sustainability. Considering the hydrological models 11 were evaluated and the SWAT model was considered as the most suitable one. Since water quality and sediment processes were required models like SLURP were not selected. Interesting is that grid based models were not recommended as they were considered as relatively new. The selected basin simulation model was IQQM. ISIS was reviewed as the best model to be used to simulate the hydrodynamic processes.

An actual model comparison, where models are really tested using existing data, is initiated by the Hydrology Laboratory (HL) of the National Weather Service (NWS), USA. The comparison is limited to hydrological models and their ability to reproduce hydrographs, based on detailed radar rainfall data. This model comparison, referred to as DMIP (Distributed Model Intercomparison Project) has the intention to invite the academic community and other researchers to help guide the NWS's distributed modeling research by participating in a comparison of distributed models applied to test data sets. Results have been published recently, but no distinct conclusions were drawn (Reed et al., 2004).

Sing et al. (2005) evaluated the performance of two popular watershed scale simulation models HSPF and SWAT. Both models were calibrated for a nine-year period and verified using an independent 15-year period by comparing simulated and observed daily, monthly, and annual streamflow. The characteristics of simulated flows from both models were mostly similar to each other and to observed flows, particularly for the calibration results. The final conclusion was SWAT predicts flows slightly better than HSPF for the verification period, with the primary advantage being better simulation of low flows.

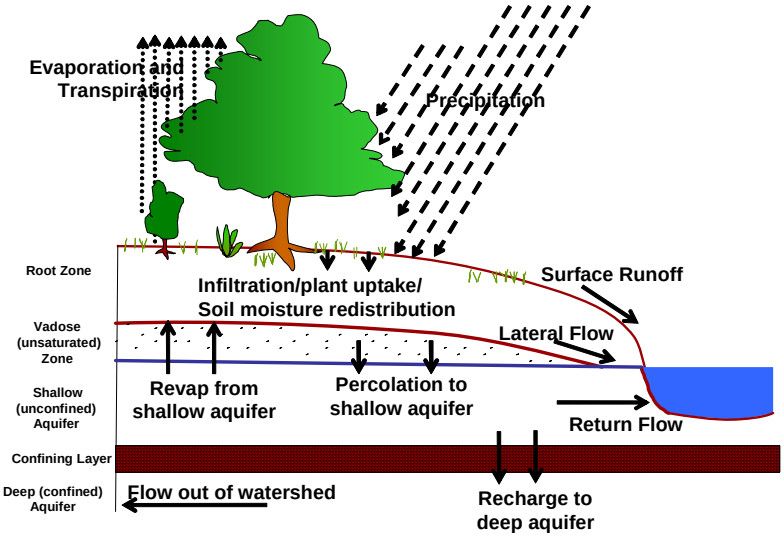

Figure 4. Main land phase processes as implemented within SWAT

2.6 SWAT model

SWAT is the acronym for Soil and Water Assessment Tool, a river basin model developed originally by the USDA Agricultural Research Service (ARS) and Texas A&M University that is currently one of the worlds leading spatially distributed hydrological models.

A distributed rainfall-runoff model – such as SWAT – divides a catchment into smaller discrete calculation units for which the spatial variation of the major physical properties are limited, and hydrological processes can be treated as being homogeneous. The total catchment behavior is a net result of manifold small sub-basins. The soil map and land cover map within sub-basin boundaries are used to generate unique combinations, and each combination will be considered as a homogeneous physical property, i.e. Hydrological Response Unit (HRU). The water balance for HRU's is computed on a daily time step. Hence, SWAT will distribute the Rio Bravo into units that have similar characteristics in soil, land cover and that are located in the same sub-basin.

Irrigation in SWAT can be scheduled by the user or automatically determined by the model depending on a set of criteria. In addition to specifying the timing and application amount, the source of irrigation water must be specified, which can be: canal water, reservoir, shallow aquifer, deep aquifer, or a source outside the basin.

SWAT can deal with standard groundwater processes. Water enters groundwater storage primarily by infiltration/percolation, although recharge by seepage from surface water bodies is also included. Water leaves groundwater storage primarily by discharge into rivers or lakes, but it is also possible for water to move upward from the water table into the capillary fringe, i.e. capillary rise. As mentioned before, water can also be extracted by mankind for irrigation purposes. SWAT distinguishes recharge and discharge zones.

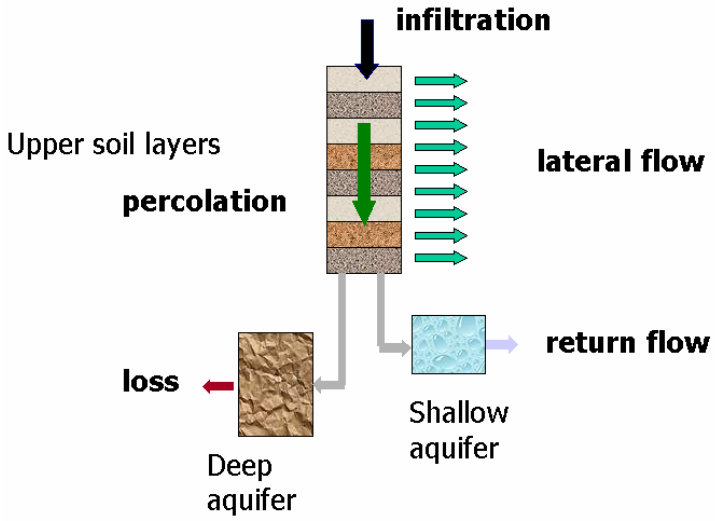

Figure 5. Schematic diagram of the sub-surface water fluxes

Recharge to unconfined aquifers occurs via percolation of excessively wet root zones. Recharge to confined aquifers by percolation from the surface occurs only at the upstream end of the confined

aquifer, where the geologic formation containing the aquifer is exposed at the earth's surface, flow is not confined, and a water table is present. Irrigation and Link canals can be connected to the groundwater system; this can be an effluent as well as an influent stream.

After water is infiltrated into the soil, it can basically leave again the ground as lateral flow from the upper soil layer – which mimics a 2D flow domain in the unsaturated zone – or from return flow that leaves the shallow aquifer and drains into a nearby river. The remaining part of the soil moisture can feed into the deep aquifer, from it can be pumped back. The total return flow thus consists of surface runoff, lateral outflow from root zone and aquifer drainage to river.

For each day of simulation, potential plant growth, i.e. plant growth under ideal growing conditions is calculated. Ideal growing conditions consist of adequate water and nutrient supply and a favorable climate. The biomass production functions are to a large extend similar to SEBAL. First the Absorbed Photosynthetical Radiation (APAR) is computed from intercepted solar radiation, followed by a Light Use Efficiency (LUE) that is in SWAT essentially a function of carbon dioxide concentrations and vapor pressure deficits. The crop yield is computed as the harvestable fraction of the accumulated biomass production across the growing season.

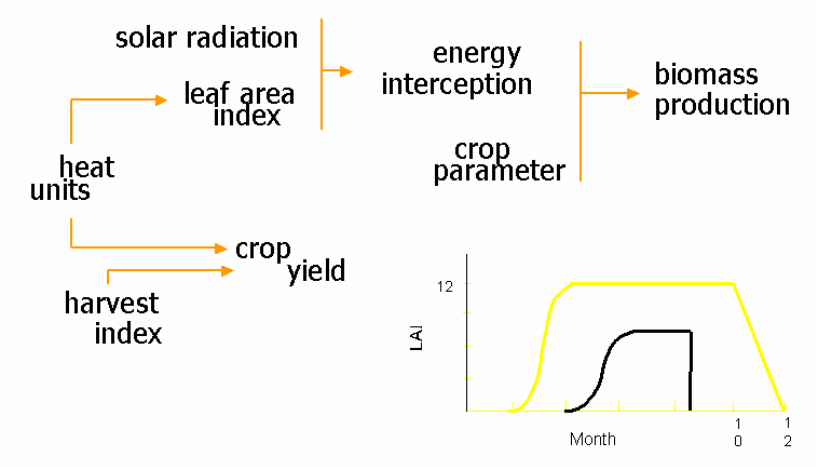

Figure 6. Parameterization of crop production

3 Calibration in hydrological modelling

3.1 Introduction

Hydrological models use input data that have, by definition, inaccuracies. These input data or parameters must be estimated for a given catchment and for each computational segment of the model. They must be estimated either by some relationship with physical characteristics or by tuning the parameters so that model response approximates observed response, a process known as calibration.

The process of model calibration is quite complex because of limitations of the models, limitations of the input and output data, imperfect knowledge of basin characteristics, mathematical structure of the models and limitations in our ability to express quantitatively our preferences for how best to fit the models to the data. As a result of these limitations, it is even not clear that a unique set of values exists for the model parameters for a given watershed. When comparing model outputs to observations, a basic question is what causes the differences. (Duan et al., 2003)

The first paper published in the Journal of Hydrology mentioning "calibration" and "model" was published in 1973 by McCuen, although the emphasis was more on sensitivity analysis. Douglas et al. (1973) were the first to publish in the same Journal about a real calibration approach of models. Attention drawn on calibration procedures has been growing over the last 10 years. In the 2005 issues of the Journal of Hydrology for example 41 articles were published having the word "calibration" mentioned in the abstract. One decade ago (1995) the number of articles was limited to seven.

Recently interest in using simulation models in ungauged or sparsely gauged basins has increased leading to some concerted actions. The most relevant ones are PUB (Prediction in Ungauged Basins) and MOPEX (Model Parameter Estimation Experiment)

Prediction in Ungauged Basins (PUB) refers to the prediction of streamflow, sediment and water quality variables at multiple scales, which is not based on the availability of measured data of these variables, and thus precludes "local tuning" or "calibration". On the contrary, PUB requires the development of new predictive approaches that are based on a deep "understanding" of hydrological functioning at multiple space-time scales. Indeed, PUB will herald a major change of paradigm in surface hydrology from the present one based on "calibration" to one based on "understanding".

The Model Parameter Estimation Experiment (MOPEX) project was established in 1996 with the primary goal to develop techniques for the a priori estimation of the parameters used in land surface parameterization schemes of atmospheric and hydrological models. MOPEX is an open international collaborative endeavor and has a loose group of contributors and participants. The major sponsor is the International Association of Hydrological Sciences (IAHS). So far, the major outcomes are a couple of active working meetings and publications in various journals and working papers. No real secretariat or organizational structure exists.

Calibration of hydrological models can be considered as parameter estimation or more general as optimisation. Calibration requires four dominant elements: (i) objective function, (ii) optimisation algorithm, (iii) termination criteria, (iv) calibration data (Singh and Woolhiser, 2002). Besides these four elements a fifth one can be added: (v) parameters to be optimized. These five elements will be discussed here in more detail.

3.2 *Objective function*

The objective function describes the difference between the observed and the simulated value. In conventional hydrological model calibration these observed and simulated values are discharges. Only few studies included other observations, such as groundwater levels and soil moisture contents, in the calibration process. Objective functions come at different flavors and the most frequently used ones will be discussed here.

DeSmedt et al. (2005) performed a flood control modeling study where four different objective functions were evaluated criteria as proposed by Hoffmann et al. (2004). The objective functions are presented in Table 1. CR_1 is the model bias, for which the value 0 represents a perfect simulation of the flow volume. CR_2 represents the model variance. CR_3 is the Nash-Sutcliffe coefficient for evaluating the ability of reproducing the time evolution of flows with a best value of 1. CR_4 is a logarithmic transformed Nash-Sutcliffe criterion, giving emphasize for evaluating the quality of low-flow simulations. CR_5 is an adapted version of the Nash-Sutcliffe criterion giving more weight to high discharges, and used for evaluating model efficiency for high flows.

Table 1. Evaluation criteria for the assessment of model performance.

Code	Criterion	Description
CR_1	$\sum_{i=1}^{N} Qs_i \Big/ \sum_{i=1}^{N} Qo_i - 1$	Model bias for evaluating the ability of reproducing the water balance
CR_2	$\sum_{i=1}^{N} \left(Qs_i - \overline{Qo}\right)^2 \Big/ \sum_{i=1}^{N} \left(Qo_i - \overline{Qo}\right)^2$	Determination coefficient representing the model variance
CR_3	$1 - \sum_{i=1}^{N} \left(Qs_i - Qo_i\right)^2 \Big/ \sum_{i=1}^{N} \left(Qo_i - \overline{Qo}\right)^2$	Model efficiency for evaluating the ability of reproducing all stream-flows
CR_4	$1 - \sum_{i=1}^{N} \left[\ln(Qs_i) - \ln(Qo_i)\right]^2 \Big/ \sum_{i=1}^{N} \left[\ln(Qo_i) - \ln(\overline{Qo})\right]^2$	Model efficiency for evaluating the ability of reproducing low flows
CR_5	$1 - \sum_{i=1}^{N} \left(Qo_i + \overline{Qo}\right)\left(Qs_i - Qo_i\right)^2 \Big/ \sum_{i=1}^{N} \left(Qo_i + \overline{Qo}\right)\left(Qo_i - \overline{Qo}\right)^2$	Model efficiency for evaluating the ability of reproducing of high flows

with
Q : flow ($m^3 s^{-1}$)
Qs : simulated flow ($m^3 s^{-1}$)
Qo : observed flow ($m^3 s^{-1}$)
N : number of observations

In practical cases, however, only one objective function is used, where the RMSE and the Nash-Sutcliffe criterion are the most commonly used ones (Lipiwattanakarn et al., 2006). The Root mean square error (RMSE) is expressed as:

$$RMSE = \left(\frac{\sum_{i}^{N} (Q_{obs,i} - Q_{sim,i})^2}{N} \right)^{1/2} \quad (1)$$

RMSE measures the average error between the observed and the simulated discharges, where Q_{obs} is the observed discharge, Q_{sim} is the simulated discharge and N is the number of observations. The closer the RMSE value is to zero, the better the performance of the model. The most frequently used objective function in hydrology is the efficiency index or Nash-Sutcliffe criterion:

$$EI = 1 - \left(\frac{\sum_{i=1}^{N} (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^{N} (Q_{obs,i} - \overline{Q}_{obs})^2} \right) \quad (2)$$

The Nash-Sutcliffe criterion value is in the range of [-∞, 1]. The zero value means the model performs equal to a naive prediction, that is, a prediction using an average observed value. The value less than zero means the model performs worse than the average observed value. A value between 0.6-0.8 is moderate to good. A value more than 0.8 is a good fit. A value of one is a perfect fit.

Most of these objective functions consider only discharge data as calibration factor. If more than one data source will be used, a combined objective function has to be used. Madsen and Jacobsen (2001) calibrated the MIKE-SHE model using groundwater as well as flow data, based on the following combined RMSE approach:

$$F_h(\theta) = \frac{1}{M} \sum_{j=1}^{M} \sqrt{\frac{1}{n_i} \sum_{i=1}^{n_i} [h_{obs,i,j} - h_{i,j}(\theta)]^2} \quad (3)$$

$$F_q(\theta) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} [q_{obs,i,j} - q_{i,j}(\theta)]^2} \quad (4)$$

$$F_{agg}(\theta) = w_h F_h(\theta) + w_q F_q(\theta) \quad (5)$$

where:
$F(\theta)$ an objective function that measures the goodness-of-fit of the simulated model with respect to the parameter set θ.
h: groundwater levels
M: total number of groundwater locations
n: total number of observations at one groundwater location
q: runoff at the catchment outlet
w_h : weight assigned to groundwater level data
w_q : weight assigned to runoff data

The first equation is the average root mean squared error (RMSE) of the groundwater levels (h) at M locations, and the second equation is the RMSE of the discharge (q). The weighted average measure is given by the last equation.

Selection of the appropriate objective function is critical for a successful optimization, but the best objective function is problem dependent. Many research papers have been published to compare different objective functions all with different conclusions. Some examples:

- Rao and Han (1987) analysed several objective functions in calibrating the urban watershed runoff model and found the least-squares criterion to be the best.
- Servat and Dezetter(1991) employed five different objective functions for calibrating a rainfall-runoff model on a Sudanese savannah area in the Ivory Coast and found the Nash-Sutcliffe efficiency to be the best.

3.3 *Optimization algorithm*

Although optimization is in principle quite straightforward, e.g. minimizing the objective function, in real life the problem is very complex as many objective functions have multiple extremes (minima).

In general, optimization algorithms can be categorized as "local" and "global" search methods (Sorooshian and Gupta, 1995). Depending on the hill climbing strategy employed, local search algorithms may be further divided into "direct" and "gradient-based" methods. Direct search methods use only information on the objective function value, whereas gradient-based methods also use information about the gradient of the objective function. Local search methods are efficient for locating the optimum of a uni-modal function since in this case the hill climbing search will eventually reach the global optimum, irrespective of the starting point. Interesting is that in groundwater modelling, parameter estimation has mainly been based on local gradient-based search techniques (e.g. McLaughlin and Townley, 1996).

In practice, hydrological simulation models have numerous local optima on the objective function surface, and in such cases local search methods are less effective because the estimated optimum will depend on the starting point of the search. For such multi-modal objective functions global search methods are more effective, where the term "global" refers to algorithms that are especially designed for locating the global optimum and not being trapped in local optima. Popular global search methods are the so-called population-evolution-based search strategies such as the shuffled complex evolution (SCE) algorithm (Duan et al., 1992) and genetic algorithms (GA) (Wang, 1991).

In summary five classes of optimization algorithms can be distinguished: (i) direct search methods, (ii) gradient search methods, (iii) random search methods, (iv) multistart algorithms, and (v) shuffled complex algorithms (Sorooshian and Gupta, 1995). The first two can be considered as local search methods and the latter three as global search methods.

In terms of global optimisation techniques Solomatine (1998) distinguished the following five groups:

- set (space) covering techniques;
- random search methods;
- evolutionary and genetic algorithms (can be attributed to random search methods);

- methods based on multiple local searches (multistart) using clustering;
- other methods (simulated annealing, trajectory techniques, tunneling approach, analysis methods based on a stochastic model of the objective function).

For calibration of lumped, conceptual hydrological catchment models a large number of studies have been conducted that compare different automatic algorithms (e.g. Duan et al., 1992; Gan and Biftu, 1996; Cooper et al., 1997; Kuzcera, 1997; Franchini et al., 1998; Thyer et al., 1999). The main conclusion from these studies is that the global population-evolution based algorithms are more effective than multi-start local search procedures, which in turn perform better than pure local search methods.

Global search procedures has been applied in steady state groundwater modelling (Heidari and Ranjithan, 1998; Solomatine et al., 1999), but to the authors' knowledge no attempt has yet been made to apply these techniques for calibration of integrated and distributed catchment models.

Singh and Woolhiser (2002) stated that the shuffled complex evolution global optimisation algorithm has been found to be consistent, effective, and efficient in locating the globally optimum hydrologic model parameters.

Recently SWAT was used to evaluate several the optimization algorithms: Shuffled Complex Evolution (SCE), real-valued simple Genetic Algorithm (GA), multi-start Simplex and Monte Carlo Sampling (MCS) and a new algorithm called the Global Greedy Search (GGS) algorithm (Tolson and Schoemaker, 2006). All algorithms used in this study were coded in Matlab and compared at default or recommended algorithm parameter settings. For the two case studies a maximum of 2500 (6 parameters) respectively 6000 (14 parameters) SWAT model evaluations were used. Results indicated, similar as in the majority of previous studies, that the SCE algorithm outperformed the Simplex, GA and MCS algorithms.

One of the major conclusions of research on optimization is somewhat disappointing: there is no best algorithm. One algorithm can be very good on one problem and show poor performance on another problem and/or vice versa. Solomatine (1998; 1999) defined three performance indicators that can be used to evaluate the algorithm for a specific problem:

- effectiveness: how close the algorithm gets to the global minimum;
- efficiency: running time of an algorithm measured by the number of function evaluations needed (the running time of the algorithm itself is negligible compared with the former);
- reliability: robustness of the algorithms can be measured by the number of successes in finding the global minimum, or at least approaching it sufficiently closely.

The book published by Duan et al. (2003) on calibration of watershed models came to the clear conclusion that the Shuffled Complex Evolution (SCE) can be considered as the de-facto standard optimization algorithm.

A detailed description of the method appears in Duan et al. (1992). In summary the algorithm has five distinct steps:

1. A "population" of points is selected randomly from the feasible parameter space.

2. The population is partitioned into several complexes, each with $2n + 1$ points, where n is the number of parameters to be optimized.
3. Each complex is evolved independently based on the downhill simplex method.
4. The population is periodically shuffled to share information, and new complexes are formed.
5. Evolution and shuffling are repeated until the specified convergence criteria are satisfied.

Conclusions from most recent studies that the SCE algorithm outperforms other ones should be taken into consideration that these conclusions are mainly based on optimization on flows, while the focus of our study is on calibration using evapotranspiration. Also Skahill and Doherty (2006) recently showed the advantages of the more traditional Gauss–Marquardt–Levenberg (GML) algorithm over the global search ones.

3.4 *Termination criteria*

Termination criteria are needed to determine when to stop the iterative search. Methods that have been used include (i) function convergence, (ii) parameter convergence, or (iii) a maximum number of iterations (Hogue et al., 2000) When an algorithm is unable to appreciably change parameter values and improve the objective function value, parameter convergence is achieved. Function convergence occurs when the algorithm is unable to improve the objective function beyond a predefined increment in one or more iterations. A calibrator also may set a maximum number of iterations to stop the search procedure, ensuring that the algorithm does not enter an endless loop.

One simple way to terminate the search is to stop when the algorithm is unable to appreciably improve the value of the function over one or more iterations (**function convergence**). While this can indicate arrival at the location of an optimum, it could also mean only that a very flat region of the response surface has been reached. If precise detection of an optimum is not considered as important, then function convergence can be a very useful termination criterion. One typical implementation of this criterion is to stop when:

$$\frac{f_i - f_{i-1}}{f_i} \leq Err_f \tag{6}$$

where f_{i-1} and f_i are the function values at the $(i-1)^{th}$ and i^{th} steps, respectively, and Err_f is the function convergence criterion (for example $Err_f = 10^{-3}$).

Another way to terminate the search is to stop when the algorithm is unable to appreciably change the parameter values and simultaneously improve the function value over one or more iterations (**parameter convergence**). While this can indicate arrival at an optimum, it could also mean only that a region of high parameter interaction (long narrow valley) on the response surface has been reached. One typical implementation of this criterion is to stop when:

$$\frac{\theta(j)_{i-1} - \theta(j)_i}{\theta(j)_{\max} - \theta(j)_{\min}} \leq Err_\theta \tag{7}$$

where $\theta(j)_{i-1}$ and $\theta(j)_i$ are the values of the j^{th} parameter at the $(i-1)^{th}$ and i^{th} steps, respectively, and Err_θ is the parameter convergence criterion (for example $Err_\theta = 10^{-3}$).

If computer time is limited, and to ensure that the algorithm does not somehow enter an infinite loop, it is normal to terminate the search if a prespecified maximum number of iterations is exceeded, unless the parameter or function convergence criteria are met first (maximum number of iterations). For random search methods, this is the normal way to terminate the search. It is not really possible to give guidelines on the value for this criterion, because it is both algorithm- and problem – dependent. The maximum iterations criterion is used as a backup to prevent waste of computer time.

None of these termination criteria guaranty that the search arrival at the global optimum, except in the most trivial cases where the function is convex and well behaved. These criteria can be used in the same program, so the search will terminate when one of the three criterion is reached (Xu, 2002).

3.5 *Calibration data*

It is common practice to use observed discharge data to calibrate hydrological models. In some cases, especially if groundwater models are considered, hydraulic heads in aquifer systems are used. As discussed previously, this may lead to models able to generate runoff accurately even if processes are not well described by model and model parameters. It is therefore erroneous to use models calibrated on discharge data to evaluate other (land) processes, or undertake scenario analysis. The objective of this study is therefore to explore the use of evapotranspiration data as alternative for calibration purposes.

The number of data to apply in a calibration procedure is somewhat less studied. It has been a common practice to use as much data as were available for the calibration, after setting aside part of the data set for verification of the results. However, studies by Sorooshian et al (1983) and Xu and Vandewiele (1994) indicated that the use of longer data sets than what is necessary served only to marginally improve the parameter estimates. In general, from a statistical point of view, the data set used should be at least 20 times the number of parameters to be estimated (for example, if there are 10 parameters, then at least 200 observed data points should be used for computing the function). This is of course an approximate rule of thumb. Gupta and Sorooshian (1985) showed that the standard error (σ) of the estimate of parameter (j) decreases with sample size n approximately according to the formula:

$$\sigma(j) \approx \frac{1}{\sqrt{n}} \tag{8}$$

Because the marginal improvement in $1/\sqrt{n}$ becomes small after 500 to 1000 data points, this suggests that two to three years of calibration data should be sufficient for a daily model with not more than 10 parameters, provided the data are of good quality.

3.6 *Parameters to be optimized*

The number of parameter to be estimated and which parameters to include in the calibration processes depends on many factors, such as: model considered, observation parameters and number,

optimization algorithm, and objectives of the study. Appropriate parameter sensitivity should be done prior to any calibration. Some suggestions specific to SWAT are provided in the SWAT manual:

- ESCO: Soil evaporation compensation factor
- CN2: Curve number that controls runoff according to the American SCS method
- REVAP: Groundwater revap coefficient.
- SHALLST: Threshold depth of water in the shallow aquifer.
- ALPHA_BF: Baseflow alpha factor.

Tolson and Schoemaker (2006) calibrated SWAT using several optimization algorithms: Shuffled Complex Evolution (SCE), real-valued simple Genetic Algorithm (GA), multi-start Simplex and Monte Carlo Sampling (MCS) and a new algorithm called the Global Greedy Search (GGS) algorithm. For two case studies 6 parameters respectively 14 parameters were used (Table 2).

Table 2. Parameters used to calibrate the SWAT model (in brackets name of the input file).

Case 1 (6 parameters)
SMTMP (.bsn), Snow melt base temperature (ºC)
SURLAG (.bsn), Surface runoff lag coefficient
GW_DELAY (.gw), Groundwater delay time (days)
ALPHA_BF (.gw), Baseflow alpha factor (days)
BIO_E (crop.dat), Radiation-use efficiency ((kg/ha)/(MJ/m2))
BLAI (crop.dat), Maximum potential leaf area index
Case 2 (14 parameters)
TIMP (.bsn), Snow pack temperature lag factor
SURLAG (.bsn), Surface runoff lag coefficient
APM (.bsn), Peak rate factor for subbasin sediment routing
PRF (.bsn), Peak rate factor for main channel sediment routing
SPCON (.bsn), Linear channel sediment reentrainment factor
SPEXP (.bsn), Exponent channel sediment reentrainment factor
GW_DELAY (.gw), Groundwater delay time (days)
ALPHA_BF (.gw), Baseflow alpha factor (days)
BIOMIX (.mgt) A, Biological mixing efficiency
CN2 (.mgt) A, SCS runoff curve number for moisture condition II
AWC_f (.sol) A & B, Available water capacity factor
SOL_K_f (.sol) A & C, Saturated hydraulic conductivity
T_OPT (crop.dat), Optimal temperature for plant growth (ºC)
T_BASE (crop.dat), Minimum temperature for plant growth (ºC)

Bastiaanssen et al. (2006) applied the SWAT model for the Rio Bravo basin in Mexico and performed a simple calibration using the following parameters as shown in Table 3.

Table 3. Parameters used to calibrate the SWAT model (in brackets name of the input file).

SOL_Z (.sol), Depth from soil surface to bottom of layer (mm)
SOL_AWC (.sol), Available water capacity of the soil layer (mm H_2O/mm soil)
GW_REVAP (.gw), Groundwater "revap" coefficient
RFINC(.sub), Rainfall adjustment (% change)

3.7 Calibration tools

3.7.1 PEST

PEST is a nonlinear parameter estimation package that can be used to estimate parameters for just about any existing computer model, whether or not a user has access to the model's source code. PEST is able to "take control" of a model, running it as many times as it needs to while adjusting its parameters until the discrepancies between selected model outputs and a complementary set of field or laboratory measurements is reduced to a minimum in the weighted least squares sense.

The Gauss-Marquardt-Levenberg algorithm is used by PEST to optimize the model. For linear models (ie. models for which observations are calculated from parameters through a matrix equation with constant parameter coefficients), optimisation can be achieved in one step. However for nonlinear problems (most models fall into this category), parameter estimation is an iterative process. At the beginning of each iteration the relationship between model parameters and model-generated observations is linearised by formulating it as a Taylor expansion about the currently best parameter set; hence the derivatives of all observations with respect to all parameters must be calculated. This linearised problem is then solved for a better parameter set, and the new parameters tested by running the model again. By comparing parameter changes and objective function improvement achieved through the current iteration with those achieved in previous iterations.

Derivatives of observations with respect to parameters are calculated using finite differences. During every optimisation iteration the model is run once for each adjustable parameter, a small user-supplied increment being added to the parameter value prior to the run. The resulting observation changes are divided by this increment in order to calculate their derivatives with respect to the parameter. This is repeated for each parameter. This technique of derivatives calculation is referred to as the method of "forward differences".

Derivatives calculated in this way are only approximate. If the increment is too large the approximation will be poor; if the increment is too small roundoff errors will detract from derivatives accuracy. Both of these effects will degrade optimisation performance. To combat the problem of derivatives inaccuracy, PEST allows derivatives to be calculated using the method of "central differences". Using this method, two model runs are required to calculate a set of observation derivatives with respect to any parameter. For the first run an increment is added to the current parameter value, while for the second run the increment is subtracted. Hence three observation-parameter pairs are used in the calculation of any derivative (the third pair being the current parameter value and corresponding observation value). The derivative is calculated either by (i) fitting a parabola to all three points, (ii) constructing a best-fit straight line for the three points or (iii) by simply using finite differences on the outer two points.

PEST is very flexible and is becoming the de-facto standard in groundwater modeling. However, one of the most relevant restrictions is the use of the Gauss-Marquardt-Levenberg algorithm, a gradient based method, sensitive to local minima.

3.7.2 UCODE

UCODE_2005 can be used with existing process models to perform sensitivity analysis, data needs assessment, calibration, prediction, and uncertainty analysis. Any process model or set of models can be used; the only requirements are that models have numerical (ASCII or text only) input and output files, that the numbers in these files have sufficient significant digits, that all required models can be run from a single batch file or script, and that simulated values are continuous functions of the parameter values. An estimated parameter can be a quantity that appears in the input files of the process model(s), or a quantity used in an equation that produces a value that appears in the input files. In the latter situation, the equation is user-defined.

UCODE_2005 can compare observations and simulated equivalents. The simulated equivalents can be any simulated value written in the process-model output files or can be calculated from simulated values with user-defined equations. The quantities can be model results, or dependent variables. For example, for ground-water models they can be heads, flows, concentrations, and so on. Prior, or direct, information on estimated parameters also can be considered. Statistics are calculated to quantify the comparison of observations and simulated equivalents, including a weighted least-squares objective function. In addition, data-exchange files are produced that facilitate graphical analysis.

UCODE_2005 can be used in model calibration through its sensitivity analysis capabilities and its ability to estimate parameter values that result in the best possible fit to the observations. Parameters are estimated using nonlinear regression: a weighted least-squares objective function is minimized with respect to the parameter values using a modified Gauss-Newton method or a double-dogleg technique. Sensitivities needed for the method can be read from files produced by process models that can calculate sensitivities, such as MODFLOW-2000, or can be calculated by UCODE_2005 using a more general, but less accurate, forward- or central-difference perturbation technique. Statistics are calculated and printed for use in (1) diagnosing inadequate data and identifying parameters that probably cannot be estimated; (2) evaluating estimated parameter values; and (3) evaluating how well the model represents the simulated processes.

One of the disadvantages of UCODE_2005 is the use of Gauss-Newton as optimization algorithm, which is a gradient based one. For non-linear problems this might lead to finding local minima rather than global ones. However, Skahill and Doherty (2006) argued the advantages of the more traditional Gauss–Marquardt–Levenberg (GML) algorithm over the global search ones. In summary: model-run efficiency, report useful information on parameter sensitivities and covariances, easily adaptable.

More information on UCODE can be found at: http://www.mines.edu/igwmc/freeware/ucode/

3.7.3 MATLAB

Matlab can be considered as the de-facto standard for scientific high-level technical computing and interactive environment for algorithm development, data visualization, data analysis, and numeric computation. MATLAB is used in a wide range of applications, including signal and image processing, communications, control design, test and measurement, financial modeling and analysis, and computational biology. Add-on toolboxes (collections of special-purpose MATLAB functions, available separately) extend the MATLAB environment to solve particular classes of problems in these application areas.

One of the Toolboxes relevant to this study is the Model-Based Calibration Toolbox. Models are available that can be incorperated directly within this toolbox. One of the examples of this is the Shuffled Complex Evolution (SCE-UA) global optimization algorithm, that is provided by the authors and can be downloaded free of charge (http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=7671&objectType=file).

3.7.4 GLOBE

GLOBE is an optimization tool that can find the minimum a function of multiple variables which value is given by an external program or a dynamic-link library (DLL). It is possible to impose box constraints (bounds) on the variables' values. No special properties of the function are assumed. GLOBE implements the "global" minimization. There are seven (with variations – nine) algorithms implemented that the user can tune to the problem and that can be run in a batch for the same function. GLOBE includes advanced visualization features.

The following algorithms are currently implemented in GLOBE:

- Controlled random search (CRS) (Price, 1983)
- Genetic Algorithm (GA)
- Adaptive cluster covering (ACCO/ACCOL) (Solomatine, 1995)
- Multis (a version of Powell-Brent non-derivative algorithm, with multiple randomized starts)
- M-Simplex (a version of the simplex decent algorithm of Nelder and Mead, with randomized multiple starts)
- Improved Controlled random search (CRS4a) (based on Ali, Storey, 1994)
- Adaptive cluster descent (ACD) (Solomatine, 1999)

Details about the GLOBE program can be found at: http://www.unesco-ihe.org/hi/sol/global.htm

4 Conclusions

The key features of the GO research project "Remotely Sensed based hydrological model calibration for basin scale water resources planning, India" is that calibration will be based on remote sensed based ET observations rather than on discharges. In this context the following conclusions can be drawn:

- Selection of the appropriate **model** is essential where a balance has to be made between issues as: scale, physical detail, support, availability of source code, expertise, and option of linking the model with remote sensed based ET. The SWAT model offers the best opportunities in this regard.
- There appears to be a general consensus that the appropriate **objective function** for discharge based calibration is the Nash-Sutcliffe one. However, whether this is also the case for ET calibration is questionable and should be evaluated during this study.
- For most calibration studies the global **optimization algorithms** outperformed the local ones. Most studies indicated that the Shuffled Complex Evolution (SCE) algorithm performed best. Since these conclusions are based on discharge calibration it is not clear whether this holds for this study as well. It is therefore required to test several optimization algorithms.
- The **tool** to be used should therefore include more algorithms that can be all applied in one generic step. The selection of GLOBE or MATLAB might be therefore preferable. However, given the complexity of the problem and the innovative aspects of using ET as calibration data, and the uncertainaty in parameter selection, a start with a more direct search algorithm is preferable. The PEST program is therefore the best selection.

5 References

Allen, R., Pereira, L.A., Raes, D., Smith, M., 1998, Crop evapotranspiration; guidelines for computing crop water requirements, FAO Irrigation and Drainage Paper No. 56, FAO, Rome

BASINS, 2006. BASINS, Better Assessment Science Integrating Point and Nonpoint Sources. http://www.epa.gov/waterscience/basins/index.html

Bastiaanssen, W., Klaasse, A., Zwart, S., Immerzeel, W. and Droogers, P., 2006, The hydrological flow path and options for sustainable water resources management in the overexploited Rio Bravo Basin; A preliminary analysis from remote sensing and hydrological modeling, WaterWatch report, Wageningen

Burnash, R.J.E., Ferral. R.L. & McGuire, R.A. (1973)..A Generalised Streamflow Simulation System, Joint Federal-State River Forecast Center, Sacramento, California.

CAMASE. 2005. Agro-ecosystems models. http://library.wur.nl/camase/

Cosgrove, W.J.; and F.R. Rijsberman. 2000. *World Water Vision: Making water everybody's business*. London, UK: Earthscan..

Courault, D., Seguin, B. and Olioso, A., 2005, Review on estimation of evapotranspiration from Remote Sensing data: From empirical to numerical modelling approaches, Irrigation and Drainage Systems, 19, 223-249

Crawford, N.H. and R.K. Linsley. 1966. Digital Simulation in Hydrology: Stanford Watershed Model IV, Stanford Univ., Dept. Civ. Eng. Tech. Rep. 39, 1966.

De Bruin, H.A.R., van den Hurk, B.J.J.M., Kohsiek, W., 1995. The scintillation method tested over a dry vineyard area. Boundary-Layer Meteorol. 76, 25–40.

DeSmedt, F. Y.B. Liu and S. Gebremeskel. 2005. Integrated modelling of hydrological processes on basin scale. Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel.

Douglas, J.R., R.T. Clarke, S.G. Newton. 1975..The use of likelihood functions to fit conceptual models with more than one dependent variable. Journal of Hydrology, Vol. 29: 181-198

Duan, Q., Sorooshian, S. and Gupta, V., 1992, Effective and Efficient Global Optimisation for Conceptual Rainfall-Runoff Models, Water Resources Research, Vol. 28, No. 4, pages 1015-1031.

Duan, Q., Sorooshian, S. and Gupta, V., 1993, a Shuffled Complex Evolution Approach for Effective and Efficient Optimisation, Journal of Optimisation Theory Application, Vol. 76, No. 3, pages 501-521.

Duan, Q., Sorooshian, S. and Gupta, V., 1994, Optimal Use of the SCE-UA Global Optimisation Method for Calibrating Watershed Models, Journal of Hydrology, No. 158, pages 265-284.

Duan, Q. H. V. Gupta, S. Sorooshian, A.N. Rousseau, and R. Turcotte. 2003. Calibration of watershed models. Water Science and Applications Series Volume 6. American Geophysical Union.

DMIP. 2002. Distributed Model Intercomparison Project. http://www.nws.noaa.gov/oh/hrl/dmip/

EPA. 2006. Environmental Protection Agency, Information Sources. http://www.epa.gov/epahome/models.htm

GLOBE. 2006. Global optimization tool GLOBE. http://www.unesco-ihe.org/hi/sol/global.htm

Hoffmann L, El Idrissi A, Pfister L, Hingray B, Guex F, Musy A, Humbert J, Drogue G, Leviandier T., Development of regionalized hydrological models in an area with short hydrological observation series, River Research and Applications 20(3), 243-254, 2004.

IRRISOFT. 2000. Database on IRRIGATION & HYDROLOGY SOFTWARE. http://www.wiz.uni-kassel.de/kww/irrisoft/irrisoft_i.html#index

IWMC. 2002. Water quality model comparison chart. Illinois Watershed Management Clearinghouse. http://web.aces.uiuc.edu/watershed/model/COMPARE.HTM

Hogue, T.S., S. Sorooshian, H. Gupta, A. Holz, D. Braatz. 2000. A Multistep Automatic Calibration Scheme for River Forecasting Models. Journal of Hydrometeorology: Vol. 1, No. 6, pp. 524–542.

Jackson, R.B., Carpenter, S.R., Dahm, C.N.,.McKnight, D.M.,.Naiman, R.J.,.Postel, S.L., and Running, S.W., 2001. Water in a changing world. Ecological Applications 11, 1027-1045.

Kite, G.W. and Droogers, P., 2000, Comparing evapotranspiration estimates from satellites, hydrological models and field data, Journal of Hydrology, 229, 3-18

Koyama, O. 1998. *Projecting the future world food situation*. Japan International Research Center for Agricultural Sciences Newsletter 15. http://ss.jircas.affrc.go.jp/kanko/newsletter/nl1998/no.15/04koyamc.htm.

Lagouarde, J.-P. 1991. Use of NOAA-AVHRR data combined with an agrometeorological model for evaporation mapping. International Journal of Remote Sensing 12: 1853–1864.

Lipiwattanakarn, S., N. Sriwongsitanon and S. Saengsawang. 2006. Performance Comparison of a Conceptual Hydrological Model and a Back-propagation Neural Network Model in Rainfall-Runoff Modeling. Department of Water Resources Engineering, Faculty of Engineering, Kasetsart University, Bangkok, Thailand.

Linsley, R.K. 1976. Why Simulation? Hydrocomp Simulation Network Newsletter, Vol: 8-5. http://www.hydrocomp.com/whysim.html

Lyall and Macoun Consulting Engineers (1986)..*WARAS Reference Manual, Version 1.0*, prepared for the former Department of Water Resources, NSW..

Madsen, H. and T. Jacobsen. 2001. Automatic calibration of the MIKE SHE integrated hydrological modelling system. *In*: 4th DHI Software Conference, 6-8 June, 2001, Scanticon Conference Centre, Helsingør, Denmark.

McCuen, R.H. 1973. The role of sensitivity analysis in hydrologic modeling. Journal of Hydrology, Volume 18: 37-53.

Meijer, F., M. Jaarsma, R. Loeve, P. Droogers. 2004. Flexible weirs to retaining water. (In Dutch). H2O 12:24-27.

Monin, A.S., and A.M. Obukov, 1954. Basic laws of turbulent mixing in the atmosphere near the ground. Tr. Akad. Nauk SSSR Geoph. Inst., No. 24(151), 1963-1987.

Monteith, J.L., 1965, Evaporation and environment, Sym. Soc. Exp. Biol., 19, 205-234

MRC, Mekong River Commission. 2000. Review of Available Models. Water Utilisation Project Component

Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models, 1, a discussion of principles. J. Hydrol., 10(3), 282-290.

Neitsch, S.L., J.G. Arnold, J.R. Kini , J.R. W Lliams, K.W. King. 2002. Soil And Water Assessment Tool, Theoretical Documentation,Version 2000. Texas Water Resources Institute, College Station, Report TR-191

NWCC. 2006. Water Management Models. http://www.wcc.nrcs.usda.gov/nrcsirrig/irrig-mgt-models.html

Penman, H.L., 1948, Natural evaporation from open water, bare soil and grass, Proc. Roy. Soc., A, 193, 120-146.

Rao, A. R., and Han, J. 1987. Analysis of objective functions used in urban runoff models. Adv. Water Resour., 10, 205–211.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.J. 2004. Overall distributed model intercomparison project results. Journal of Hydrology, Volume 298, Issue 1-4: 27-60.

REM. 2006. Register of Ecological Models (REM). http://www.wiz.uni-kassel.de/ecobas.html

Riou, C., Itier, B. & Seguin, B. 1988. The influence of surface roughness on the simplified relationship between daily evaporation and surface temperature. International Journal of Remote Sensing 9(9): 1529–1533.

Seckler, D.; R. Barker; and U. Amarasinghe. 1999. Water scarcity in the twenty-first century. *Water Resources Development* 15: 29-42.

Seguin, B. & Itier, B. 1983. Using midday surface temperature to estimate daily evaporation from satellite thermal IR data. International Journal of Remote Sensing 4: 371–383.

Servat, E., and Dezetter, A. 1991. Selection of calibration objective functions in the context of rainfall-runoff modeling in a Sudanese savannah area. Hydrol. Sci. J., 6: 307–330.

Singh, J., H.V. Knapp, J.G. Arnold, M. Demissie. 2005. Hydrological modeling of the Iroquois river watershed using HSPF and SWAT. Journal of the American Water Resources Association. Volume 41, Number 2, April 2005, pages 343-360.

SMIG. 2006a. Surface-water quality and flow modeling interest group. http://smig.usgs.gov/SMIG/SMIG.html

Skahill, B.E., J. Doherty. 2006. Efficient accommodation of local minima in watershed model calibration. Journal of Hydrology (in press).

SMIG. 2006b. Archives of Models and Modeling Tools. http://smig.usgs.gov/SMIG/model_archives.html

Solomatine. D.P. 1998. Genetic and other global optimization algorithms - comparison and use in calibration problems. Proc. 3rd Intern. Conference on Hydroinformatics, Copenhagen, Denmark, 1998. Balkema Publishers. pp.1021-1028

Solomatine, D.P. (1999). Two strategies of adaptive cluster covering with descent and their comparison to other algorithms. Journal of Global Optimization, 1999, vol. 14, No. 1, pp. 55-78.

Sorooshian, S., and Gupta V. K. 1995. Chapter 2: Model calibration. Computer models of watershed hydrology, V. P. Singh, ed., Water Resources Publications, Littleton, Colorado: 23–68.

TNRCC, Texas Natural Resource Conservation Commission. 1998. An Evaluation of Existing Water Availability Models. Technical Paper #2. http://www.tnrcc.state.tx.us/permitting/waterperm/wrpa/wam.html

Tolson, B.A., Schoemaker, C.A. 2005. Comparison of optimization algorithms for the automatic calibration of SWAT2000. In:.Karim Abbaspour and Raghavan Srinivasan, (ed.) 3rd International SWAT Conference July 11-15 2005, Zürich, Switzerland.

United Nations. 1997. Comprehensive Assessment of the Freshwater Resources of the World (overview document) World Meteorological Organization, Geneva.

USBR. 2002. Hydrologic Modeling Inventory. http://www.usbr.gov/pmts/rivers/hmi/2002hmi/index.html

USGS. 2006. Water Resources Applications Software. http://water.usgs.gov/software/

Vijay P. Singh, V.P., and D.A. Woolhiser. 2002. Mathematical Modeling of Watershed Hydrology. Journal of Hydrologic Engineering, Vol. 7, No. 4, 270-292.

Xu, Chong-yu. 2002. Hydrological Models. Uppsala University Department of Earth Sciences Hydrology.