Hydrology and Earth System Sciences Discussions

Identification of hydrological model parameters for flood forecasting using data depth measures

T. Krauße[1] and J. Cullmann[2]

[1]Institute of Hydrology and Meteorology, Department of Hydrology, Dresden University of Technology, Bergstr. 66, 01062 Dresden, Germany
[2]2IHP/HWRP Secretariat, Federal Institute of Hydrology, Am Mainzer Tor 1, 56068 Koblenz, Germany

Correspondence to: T. Krauße (thomas.krausse@tu-dresden.de)

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Identification of model parameters by data depth measures

T. Krauße and
J. Cullmann

Abstract

The development of methods for estimating the parameters of hydrological models considering uncertainties has been of high interest in hydrological research over the last years. Besides the very popular Markov Chain Monte Carlo (MCMC) methods which estimate the uncertainty of model parameters in the settings of a Bayesian framework, the development of depth based sampling methods, also entitled robust parameter estimation (ROPE), have attracted an increasing research interest. These methods understand the estimation of model parameters as a geometric search of a set of robust performing parameter vectors by application of the concept of data depth. Recent studies showed that the parameter vectors estimated by depth based sampling perform more robust in validation. One major advantage of this kind of approach over the MCMC methods is that the formulation of a likelihood function within a Bayesian uncertainty framework gets obsolete and arbitrary purpose-oriented performance criteria defined by the user can be integrated without any further complications. In this paper we present an advanced ROPE method entitled the Advanced Robust Parameter Estimation by Monte Carlo algorithm (AROPE$_{MC}$). The AROPE$_{MC}$ algorithm is a modified version of the original robust parameter estimation algorithm ROPE$_{MC}$ developed by Bárdossy and Singh (2008). AROPE$_{MC}$ performs by merging iterative Monte Carlo simulations, identifying well performing parameter vectors, the sampling of robust parameter vectors according to the principle of data depth and the application of a well-founded stopping criterion applied in supervised machine learning. The principals of the algorithm are illustrated by means of the Rosenbrock's and Rastrigin's function, two well known performance benchmarks for optimisation algorithms. Two case studies demonstrate the advantage of AROPE$_{MC}$ compared to state of the art global optimisation algorithms. A distributed process-oriented hydrological model is calibrated and validated for flood forecasting in a small catchment characterised by extreme process dynamics.

1 Introduction

Conceptual hydrological models are designed to approximate the general physical mechanism which govern the rainfall-runoff process within a specific catchment. Many of the currently available models serve engineers in practice and hydrologists in research. Most of these models require calibration because available input data is rarely sufficient to directly parameterise the models to the desired accuracy. The success of a model application is thus strongly dependent on a good calibration.

Traditionally models are calibrated manually. This is very labour-intensive and the success of the calibration is highly dependent on the experience and hydrological knowledge of the modeller. Therefore, recently automatic methods for model calibration have evolved significantly (e.g. Duan et al., 1992; Gupta et al., 1998; Kobold et al., 2003; Theiner and Wieczorek, 2006) and have found a common acceptance and broad use in the hydrological community (e.g. Hogue et al., 2000; Cullmann, 2006; Kunstmann et al., 2006; Marx, 2007). These approaches are based on solving a mathematical optimisation problem which is formulated by the help of a purpose-oriented objective function which evaluates the model performance. The result is traditionally a single best performing parameter vector. However, regardless of the model and optimisation algorithm used, many studies applying such methods have reported problems in estimating unique best performing parameter vectors (Duan et al., 1992; Gupta et al., 1998; Wagener et al., 2004). The probability to estimate the same model performance for different estimated parameter vectors was described by Beven and Binley (1992) as the equifinality problem. In terms of the optimisation problem, the reason for this is the existence of multiple local optima with both small and large regions of attraction, discontinuities in the first derivatives, and curving ridges in the multidimensional parameter space. The difficulties to estimate single best performing parameter vectors are even stronger, when the focus is put on specific aspects of the catchment behaviour, e.g. flood events (e.g. Cullmann, 2006; Cullmann and Wriedt, 2008; Fenicia et al., 2007). In this context, hydrological models are not yet able to equally well describe the

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

full range of processes that drive the runoff generation. This holds both for simple concept models and detailed process models with physically based components. One of the main reasons for this lack of "process fidelity" are the highly dynamic characteristics of such events. Besides the discussed issues in model calibration, the uncertainties in the used observations is often completely neglected.

To overcome this problem a whole variety of Markov Chain Monte Carlo (MCMC) methods have recently been developed in order to get a well-founded estimate of the the uncertainty of model parameters in the settings of a Bayesian framework. The approaches developed by Vrugt et al. (2003) and Kuczera et al. (2006) have recently attracted much scientific interest and enjoyed rising popularity in the hydrological community (e.g. Bos and de Vreng, 2006; Thyer et al., 2006; Frost et al., 2007; Grundmann, 2010). One major advantage of a Bayesian framework is the possibility to describe all relevant sources of uncertainty in a closed form and consequently the estimation of mathematically well founded results. However, also for those kind of approaches a modeller has to make assumptions of all sources of uncertainty to be considered. Often these assumptions are quite arbitrary because the information for a well founded decision is not available. Subsequently these decisions have a non-neglect-able influence on the results. Thus, the uncertainty estimates might get a rather subjective touch – a fact that contradicts the original intention of the application of a Bayesian framework. Furthermore in many real-world applications modellers call for purpose-specific objectives in calibration, this is difficult to integrate in the likelihood function Bayesian uncertainty framework. For example the formulation and implementation of a likelihood function considering both peak flow difference and the Nash-Suttcliffe efficiency is not straightforward.

A completely new kind of approach to address this problem was presented by Bárdossy and Singh (2008) who applied the concept of data depth in order to sample robust model parameter vectors. In a previous study (Bárdossy, 2007) it was shown that the set of parameter vectors with good model performance (according to Bárdossy and Singh, 2008, they are from now on called the *good parameter vectors*) is

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

|◄ | ►|
◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

geometrically well structured for commonly used hydrological models. This result was also found by other studies dealing with model calibration (e.g. Kuczera and Parent, 1998) and by preliminary studies with WaSiM in the Rietholzbach catchment (Pompe, 2009). However, the actual goal of a successful model calibration is not just a good model performance in calibration, but the estimation of *robust parameter vectors*. A good definition of robustness of hydrological model parameters is given by Bárdossy and Singh (2008). We call all parameter vectors robust if they:

1. lead to good model performance over the selected time period

2. lead to a hydrologically reasonable representation of the corresponding processes

3. are not sensitive: small changes of the parameters should not lead to very different results

4. are transferable: they perform well for other time periods and might also perform well on other catchments (i.e. they can be regionalised)

Studies of computational geometry and multivariate statistics (e.g. Liu et al., 2006; Bremner et al., 2008) showed that members geometrically deep within a set, are more robust in order to represent the whole set. These points can be estimated by the concept of data depth, which has recently attracted a lot of research interest in multivariate statistics and robust modelling (e.g. Cramer, 2003; Liu et al., 2006). The estimation of a set of parameter vectors can be done in an evolutionary process, as presented by Bárdossy and Singh (2008) in a very first sketch. The estimated results are very promising. Therefor we reviewed the presented methods, implemented improved methods to sample deep parameter vectors and well founded stopping criteria, and applied a further developed version of the ROPE method, called $AROPE_{MC}$ in order to calibrate a hydrological model. Compared to the study of Bárdossy and Singh (2008) we applied the method to a distributed process-oriented model with a higher temporal resolution (hourly instead of daily time-step) in a catchment where the dominant processes have high dynamics.

2   Case study area and the hydrological model

The concept of this paper will be illustrated with examples from the Rietholzbach catchment. Out of a time series from 1981–2008, 24 significant flood events were selected for model calibration and validation. The hydrological model chosen is the model WaSiM-ETH. A short description of the catchment, the used data and the model is provided in this section.

2.1   Study area

The real-world case studies presented within this paper were carried out on the Rietholzbach catchment. This basin has been observed as a research catchment by the ETH Zurich since 1975. The outlet drains a $3.18\,\mathrm{km^2}$ hilly pre-alpine watershed with an average precipitation of $1600\,\mathrm{mm}$ per year, generating a mean annual runoff of $1046\,\mathrm{mm}$. As a sub-basin of the Thur catchment it is located in the northeast of Switzerland. Its geographical location and basic land-use characteristics are listed in Table 1. A significant number of studies have been conducted in this basin. For further information refer to Gurtz et al. (1999); Zappa (2002) and the website http://www.iac.ethz.ch/research/rietholzbach.

2.2   Hydrological model

The used hydrological model is WaSiM-ETH/6.4 (in the further referred to as WaSiM). It is a spatial distributed process-oriented rainfall-runoff model and was developed by Schulla (1997) at the ETH Zurich. WaSiM has been used successfully for modeling the rainfall-runoff processes in several studies in catchments located within mid mountain ranges (e.g. Grundmann, 2010) and especially also in the pre-alpine Rietholzbach catchment (Gurtz et al., 1999, 2003a,b). Additionally WaSiM-ETH has been used for extrapolation of extreme flood events by Cullmann (2006). For this study we used the version with the Richards approach for the simulation of the unsaturated zone.

In the case studies presented discussed in this work, WaSiM will be calibrated for the simulation of extreme discharges. Therefore the main focus of attention in this short model presentation will be on the model part representing the unsaturated zone. WaSiM transforms rainfall into runoff according to the scheme shown in Fig. 3. Here, three exemplary soil water compartments receive infiltration which is computed by a modified approach according to Green and Ampt (1911). This module is also used to determine the direct runoff Q_d in the model. Q_d is then routed via a flow-time grid and finally projected cell-wise to the catchment outlet by means of a simple bucket type function (Eq. 1). The recession coefficient of this function is the model parameter k_d.

$$Q_d(t) = Q_d(t-1) \, e^{\frac{-\partial t}{k_d}} \tag{1}$$

where $Q_d(t)$ is the direct runoff at timestep t with timestep length ∂t.

The soil water movement through the different soil layers is modeled by means of the discrete form of the Richards-equation which can be written as:

$$\frac{\partial \Theta}{\partial t} = \frac{\partial q}{\partial z} = q_{in} - q_{out} \tag{2}$$

where $\partial \Theta$ denotes the change in soil water content, ∂t defines the time step and ∂q is the change in specific flux. The fluxes q_{in} and q_{out} characterise the influx and efflux from the specific neighboring soil layer respectively. The thickness of the soil layers is defined by ∂z.

In the model each soil layer produces interflow (Q_{ifl}) according to Eq. (3), which is cell-wise scaled with the scalar model parameter dr.

$$Q_{ifl} = k_s(\Theta_l) \, \partial z \cdot dr \cdot \tan \beta \tag{3}$$

where k_s denotes the hydraulic conductivity at the water content Θ_l in the considered soil layer l, dr is a conceptual model parameter to be estimated and β characterises the local slope in the grid cell.

Corresponding to the direct runoff, the interflow is again projected to the catchment outlet by means of a flow-time grid and a second bucket type function. Here the model parameter k_i represents the recession coefficient in analogy to Eq. (1). For further details about WaSiM refer to Schulla (1997) and the official website of the model http://www.wasim.ch. Table 2 gives the model parameters considered for calibration. These are the storage coefficients of direct runoff and interflow, k_d and k_i, and the drainage density dr which is a scaling parameter of interflow generation. In previous studies (Cullmann, 2006; Pompe, 2009; Grundmann, 2010) these three parameters have been proven to be sensitive with respect to modelling flood events. Besides the specified upper and lower boundaries of the model parameters, the additional constraint $k_i \geq 1.05 \, k_d$ was introduced in order to account for the basic consideration that the direct runoff from a cell has a shorter travel time to the catchment outlet than the generated interflow in the unsaturated zone.

2.3 Data

Due to its longterm observation as a research catchment and its limited size, the Rietholzbach catchment has a long record of hourly data sets and the perturbing impact of data heterogeneity is relatively small in this catchment. The data we based our study upon is a time series consisting 27 years of meteorological and discharge measurements. Out of this time series we selected a set of 24 flood events with a peak flow of at least $1 \, \mathrm{mm \, h^{-1}}$. All events are in the time from May until October to avoid the problem of modeling snow accumulation and melting processes. An overview of all selected 24 flood events with their specific characteristics and pre-conditions is given in Table 3.

3 Objective criteria

Within this study commonly used objective functions are applied. There are global criteria which try to assess the general quality of the fit of model and the catchment

Identification of model parameters by data depth measures

T. Krauße and
J. Cullmann

behavior whereas local criteria just focus on a specific attribute, e.g. peak flow values or snow melt periods. The efficiency criterion according to Nash and Sutcliffe (1970) (NS) has been widely used to quantify the global performance of hydrological models. The relative deviation in peak flow (rPD) is a simple local criterion to assess the model's skill to represent the catchment behavior for flood events. In order to obtain both a good estimate of the peak flow values and a minimum reasonable representation of the catchment behavior we aggregated the global criterion NS and the local criterion rPD in a performance criterion, we call flood skill (FloodSkill). An overview of all criteria used as an objective within the case studies of this paper is given in Table 4.

4 Data depth

Data depth is a statistical method used for multivariate data analysis which assigns a numeric value to a point with respect to a set of points based on its centrality. This approach provides center-outward orderings of points in Euclidean space of any dimension and provides the possibility of a new non-parametric multivariate statistical analysis in which no distributional assumptions are needed. Tukey (1975) introduced this concept first in order to estimate the center of a multivariate dataset. A formal definition of an arbitrary depth function D for the d-dimensional space $\mathbb{R}_d$ is given as follows:

$$D : \mathbb{R}^d \times \left(\mathbb{R}^d \times \mathbb{R} \right) \rightarrow \mathbb{R} \tag{4}$$

To be called a depth function, D has to fulfill specific properties (Zuo and Serfling, 2000). The concept of data depth is illustrated in Fig. 4 by a small 2-dimensional example. For a random point set in $\mathbb{R}^2$ the data depth was computed for each point of the set with respect to the point set itself. The used data depth function was the halfspace depth. It is one of the best known among the data depth measures in nonparametric statistics, and in discrete and computational geometry. According to Tukey (1975)

and Donoho and Gasko (1992) the halfspace depth of an arbitrary point $\boldsymbol{\theta} \in \mathbb{R}^d$ with respect to a d-dimensional data set $Z = \{\boldsymbol{z}_i = (z_{i1}, \cdots, z_{id}); i = 1, \cdots, n\}$ is defined as the smallest number of data points in any closed halfspace with boundary through $\boldsymbol{\theta}$. This is also called the Tukey or location depth, and it can be written as:

$$h\text{depth}\,(\boldsymbol{\theta}|Z) := \min_{\|\boldsymbol{u}\| = 1} \{i, \boldsymbol{u}^\top \boldsymbol{z}_i \geq \boldsymbol{u}^\top \boldsymbol{\theta}\} \tag{5}$$

where $\boldsymbol{u}$ ranges over all vectors in $\mathbb{R}_d$ with $\|\boldsymbol{u}\| = 1$.

Very often the halfspace depth is normalized by division with the number of points in the set Z:

$$h\text{depth}^*(\theta|Z) := \frac{h\text{depth}\,(\boldsymbol{\theta}|Z)}{\#\,\{Z\}} \tag{6}$$

The first publication of Tukey (1975) was then followed by many generalizations and other definitions of this concept (e.g. Oja, 1983; Donoho and Gasko, 1992; Rousseeuw and Struyf, 1998; Rousseeuw and Hubert, 1999; Vencálek, 2008). A good overview of a broad range of different definitions of the concept of data depth and its application for multivariate data analysis is given by Hugg et al. (2006) and Liu et al. (2006). In the following the symbol D is used for an arbitrary depth function.

For a given data set Z, the set D_k of all points $\in \mathbb{R}_d$ with depth at least k is called the contour of depth k in statistics (s. Donoho and Gasko, 1992). The application of that concept in the sampling of parameter vectors with a least depth with respect to a set of good parameter vectors is the underlying approach of the algorithm presented in this work. We use the definition of data depth introduced by Tukey (1975) with an implementation according to Rousseeuw and Struyf (1998). Furthermore it has proved to be a very robust measure in order to identify the center of a multivariate dataset (e.g. Rousseeuw and Struyf, 1998; Cramer, 2003; Serfling, 2006). For a study of further data depth functions with the algorithm presented in this paper, refer to Krauße (2011).

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

5 Implementing advanced methods for robust parameter estimation: the AROPE$_{MC}$ algorithm

Bárdossy and Singh (2008) developed the ROPE method with the goal to estimate a set of robust parameter vectors in the feasible space by the application of the principle of data depth in order to generate parameter vectors deep with respect to an identified set of good parameter vectors. The first presented ROPE algorithm (Bárdossy and Singh, 2008) is given as follows[1].

Algorithm 5.1 ROPE$_{MC}$

1: Select d model parameters, to be considered for calibration and identify prior boundaries $[\boldsymbol{x}_{lb}, \boldsymbol{x}_{ub}]$ for all selected parameters
2: n random parameter vectors forming the set X_n are generated in the d-dimensional rectangle bounded by the defined boundaries.
3: **repeat**
4:     The hydrological model is run for each parameter vector in X_n and the corresponding model performances are calculated
5:     The subset X_n^* of the best performing parameters is identified. This might be for example the best 10% of Θ_n
6:     m random parameter vectors forming the set Y_m are generated, such that $\forall \boldsymbol{\theta} \in Y_m : D(\boldsymbol{\theta}|X_n^*) \geq L$ where $L \geq 1$
7:     The set Y_m is relabeled as X_n and steps 3–6 are repeated until
8: **until** the performance corresponding to X_n and Y_m does not differ more than what one would expect from the observation errors
9: **return** Y_{m_i}

In principle the general proceeding of this algorithm, can be divided into three important parts. After the input and a pre-processing a set of good parameter vectors is identified (lines 4 and 5). Afterwards a set of deep parameter vectors (w. r. t. the good

---

[1] Note that the notation was marginally changed from Bárdossy and Singh (2008) in order to have a consistent syntax with other publications in the field of data depth.

ones) is generated (line 6). These two operations are evolutionary repeated and after each iteration a stopping criterion is checked (line 8). The general approach of the presented algorithm is well-founded and a case study with an application to the calibration of the conceptual hydrological model HBV for a catchment in south-west Germany on a daily time step resulted in parsimonious and reasonable results. However, in first studies we experienced some problems particularly with the application of the latter two parts of the procedure, the generation of deep parameter vectors and the exact definition of the stopping criterion. In the following we will give a brief overview of the problems and explain how the new AROPE$_{MC}$ algorithm addresses these shortcomings.

One of the major premises of the application of the concept of data depth is the assumption that the set of good parameter vectors is geometrically well-structured. In concrete terms we rely on the assumption that the depth contours will be indicative of the shape of the cloud of good parameter vectors, while generating deep parameters. However, for most depth functions, this does not hold on point sets that are distributed in non-convex position (Hugg et al., 2006). Unfortunately the parameter space of most hydrological models is dominated by distinct regions of attraction and non-convex multi-dimensional ridges (e.g. Duan et al., 1992; Sorooshian et al., 1993; Grundmann, 2010). To overcome this conflict we propose to substitute the generation of deep parameter vectors with the strategy, entitled GenDeep, as given below in Algorithm 5.2.

Additionally we implemented alternative sampling strategies for the sampling of candidate points for the sets of deep parameter vectors. A simple sampling strategy of candidates is a uniform sampling within the bounding box for the considered set of good parameter vectors. This strategy gets ineffective and computationally intensive for higher dimensions. That is due to the fact that the volume ratio of the bounding box to the set of parameter vectors itself decreases with rising dimension. This issue is illustrated by Fig. 5 where the ratio between the volume of the unit sphere and the unit cube is plotted. Additionally the computational complexity of most depth functions increases tremendously for higher dimensions. To address this problem we suggest

Algorithm 5.2 GenDeep

1: Perform a cluster analysis on the set of good parameter vectors X_n^*, e.g. with the expectation maximization (EM) algorithm according to Dempster et al. (1977), which identifies the most probable number of clusters k in X_n^* and assigns all members of the set X_n^* to one (in case of ambiguity also to more than one) of the clusters c_i, where $i \in \{1, \ldots, k\}$.

2: $Y_m \leftarrow \varnothing$

3: **for all** $c_i \in \{c_1, \ldots, c_k\}$ **do**

4: $\quad m_i \leftarrow m\left(\frac{\#\{C_i\}}{n}\right)$

5: $\quad m_i$ random parameter vectors forming the set Y_{m_i} are generated, such that $\forall \boldsymbol{\theta} \in Y_{m_i} : \{D(\boldsymbol{\theta}|c_i) \geq L \text{ where } L > 0\}$

6: $\quad Y_m \leftarrow Y_m \cup Y_{m_i}$

7: **end for**

8: **return** Y_{m_i}

to approximate the considered set of good parameter vectors by a Gaussian mixture model (GMM) whose parameters can be estimated by the EM algorithm which is called anyway in order to do the cluster analysis in the presented strategy. For further details of the proposed strategy for the generation of deep parameter vectors refer to Krauße

5 (2011).

Another issue of the ROPE$_{\text{MC}}$ algorithm is the loosely defined stopping criterion: "until the performance corresponding to X_n and Y_m does not differ more than what one would expect from the observation errors" (Bárdossy and Singh, 2008, p. 1280). The problem is that there are countless possibilities in the prior estimation of the tolerance

10 in the model performance due to uncertainty in the observation data and it can hardly be determined exactly. A broad definition of this tolerance can lead to sets with inferior model performance, whereas a tighter tolerance can easily result in overfitting. This is a severe shortcoming because it undermines the actual goals of the algorithm. Overfitting in the context of robust parameter estimation means that the model perfor-

15 mance on the calibration data still can be increased by further shrinking the estimated

set of the deep model parameter vectors, whereas the model performance on (reasonably similar) control data decreases by further shrinking. Figure 6 illustrates this fact with the results of the calibration of WaSiM in the Rietholzbach catchment w. r. t. to flood events. The FloodSkill criterion was used as objective and the flood events no. 4 and no. 14 were used as calibration and control data, respectively. It is evident that the model performance on the control data considerably decreases from iteration 3 whereas the model performance on the calibration data could be increased by further iterations.

To address this problem we implemented two changes to the algorithm. First, we slightly changed the evolutionary shrinking of the generated deep parameter vectors. To avoid the unintended exclusion of possibly robust parameter vectors close to the boundary of the initial set of good model parameters, we suggest merging the set of generated deep parameter vectors and the identified good parameter vectors as initial set for the next iteration, as follows:

$$X_n \leftarrow Y_m \cup X_n^* \tag{7}$$

Furthermore we introduced a new stopping criterion in order to avoid overfitting. We suggest the splitting of the data used for model calibration in a calibration and a control set. Just the calibration set is used for the actual model calibration, whereas the control set is just used to supervise the control process in order to avoid overfitting. In each iteration of the algorithm the model performance is estimated both on the calibration and control set. The moment the performance does not improve anymore for the control set, the algorithm is stopped. This kind of approach is a state of the art method in the supervised training of artificial neural networks in order to avoid overfitting (Tetko et al., 1995). The new algorithm, entitled the Advanced Robust Parameter Estimation by Monte Carlo ($AROPE_{MC}$) is given in a brief form in pseudocode in Algorithm 5.3. A more detailed illustration of the approach is given in Fig. 7. The algorithm was implemented in the MATLAB programming language. The implementation is open source and available from the author.

2436

Algorithm 5.3 AROPE$_{MC}$

1: Select d model parameters, to be considered for calibration and identify prior boundaries $[\boldsymbol{x}_{lb}, \boldsymbol{x}_{ub}]$ for all selected parameters

2: n random parameter vectors forming the set X_n are generated in the d-dimensional rectangle bounded by the defined boundaries.

3: **repeat**

4:     **if** #iterations > 1 **then**

5:         $X_n \leftarrow Y_m \cup X_n^*$

6:     **end if**

7:     The hydrological model is run for each parameter vector in X_n and the corresponding model performances on the calibration data are calculated

8:     The subset X_n^* of the best performing parameters is identified. This might be for example the best 10% of Θ_n

9:     The hydrological model is run for each parameter vector in X_n^* and the corresponding model performances on the control data are calculated

10:     $Y_m \leftarrow \text{GenDeep}(X_n^*)$

11: **until** the performance corresponding to X_n and Y_m does not differ more than what one would expect from the observation errors or the performance on the control data gets worse

12: **return** Y_m

6 Case studies

6.1 Case study I: estimating the minimum of the Rosenbrock and Rastrigin function

The goal of the AROPE$_{MC}$ algorithm is not the estimate of the global optimum of a problem. However, the estimated set of robust points should encompass a region close to the global optimum. We investigate the performance of the AROPE$_{MC}$ algorithm for the estimation of the minimum of two simple test functions of the form $f : \mathbb{R}^n \to \mathbb{R}$, often used as a performance test problem for optimization algorithms: the *Rosenbrock's*

function and the *Rastrigin's function*. The former is a non-convex function with a unique minimum value of 0 attained at the point **1**. Finding the minimum is a challenge since it has a shallow minimum inside a deeply curved valley. The Rastrigin function is a typical example of non-linear multimodal function. It was first proposed by Rastrigin as a 2-dimensional function and has been generalised for multiple dimensions by Mühlenbein et al. (1991). This function is a fairly difficult problem due to its large search space and its large number of local minima. The formal definition of both functions is given in Eqs. (8) and (9) respectively. Figure 8 shows the plot for both functions for two variables, to give the reader a better impression of the nature of the problem.

$$f_a(\boldsymbol{x}) = \sum_{i=1}^{N-1} \left[\left(1 - x_i^2\right)^2 + 100 \left(x_{i+1} - x_i^2\right)^2 \right] \tag{8}$$

$$f_b(\boldsymbol{x}) = A \, n + \sum_{i=1}^{N} \left[x_i^2 - A \cos\left(2 \, \pi \, x_i\right) \right] \tag{9}$$

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \, A = 10$$

We applied the original ROPE_{MC} algorithm and the AROPE_{MC} algorithm for estimating the minimum of both test functions for the dimension 2–4. As boundary for the parameters $\boldsymbol{x}$ we chose [**−10**, **10**]. In order to have a reference performance, we applied the genetic algorithm (GA) according to Conn et al. (1997) in the same feasible space with a comparable number of maximum function evaluations and the same tolerance[2]. To get a fair mean best result for comparison, we ran the GA each iteratively until the value for the overall mean of the best estimates got stable. For each estimate we computed its fitness value as the absolute of the difference between its function value and the known global optimum.

---

[2]For all problems and all algorithms we adjusted their settings to ensure at the most 10 000 function evaluations and a tolerance in the performance of the estimated parameter sets of 0.1.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Figure 9 illustrates the principle of operation of the algorithm by the scatter plots of the estimated parameter vectors after several iteration steps for the the 2-dimensional case. It is evident that new parameter vectors are sampled deep with respect to the estimated set in the previous iteration. For the Rastrigin's function several clusters can be clearly identified which improves the performance of the sampled parameter vectors tremendously.

Table 5 presents the comparison of the fitness values for all used algorithms. It is evident that the AROPE$_{MC}$ algorithm can achieve a reasonable performance and estimate a set of model parameters in the region with the highest possible performance for all test cases. In most of the cases the deepest estimated parameter vectors have a better performance than the mean of the estimated set. For the Rosenbrock's function both the ROPE$_{MC}$ and the AROPE$_{MC}$ algorithm perform comparable well for all dimensions. The results for the Rastrigin's function show an improvement of AROPE$_{MC}$ with respect to ROPE$_{MC}$ algorithm for higher dimensions due to the cluster based sampling with the DepGeen strategy. However, the results for the Rastrigin's function also show that the proposed algorithm still suffers from the general shortcomings of a Monte-Carlo type approach for high dimensional problems with a very large number of areas of attraction. A too small sample size can result in an inaccurate clustering and consequently undermine the improvements of this strategy. Note that this problem can easily be compensated by a higher sample size on the cost of more computation time. As another possible solution for high-dimensional problems with multiple regions of attraction we propose the previous use of an approved evolutionary search strategy for high-dimensional parameter spaces, e.g. the particle swarm concept.

6.2 Case study II: calibration of the hydrological model WaSiM with focus on flood events

In a second case study we studied the influence of observation errors on the calibration results with AROPE$_{MC}$ in comparison with state of the art optimisation algorithms of the process-oriented hydrological model WaSiM for flood events. For classical optimisation

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

CC BY

algorithms we used the interior-point method (IPM) according to Waltz et al. (2006) which is a gradient based method and the GA already used in the previous case study.

We assume that the influence of observation errors in temperature measurements is negligible for the simulation of flood events whereas the uncertainty of the measured precipitation can be expressed by an ensemble. To keep the problem still computationally feasible we do not consider the influence on the estimated parameter sets due to the uncertainties in the observed precipitation and just use the ensemble mean for the model calibration, but just focus on the influence on the observation errors of the measured discharge. Following the assumptions of Bárdossy and Singh (2008) we assume an accuracy of the measured discharge $q_{obs}(t)$ of 5%. Thus, the real but unknown discharge $q(t)$ can be written as:

$$q(t) = q_{obs}(t)\,(1 + e(t)) \tag{10}$$

with $e(t)$ being a random error. This random error is due to uncertainties of the rating curve, non-uniqueness of the stage discharge relationship, changes of the cross section etc. (Bárdossy and Singh, 2008). As many other authors (e.g. Kuczera et al., 2006; Bárdossy and Singh, 2008) we assume that this error obeys a normal distribution with a standard deviation of the measurement accuracy: $\mathcal{N}(0, 0.05)$. For each observed discharge time series we used this model and produced an ensemble with 100 members. With both the IPM and the GA for different each of the 24 flood events with respect to every single ensemble member each and validated the set of 100 estimated best parameter vectors. The same was done for the AROPE$_{MC}$ algorithm. In order to reduce the computation time, we previously checked the stability of the estimated robust parameter sets with AROPE$_{MC}$ for a subset of the discharge ensemble members and subsequently just used the ensemble mean for calibration. The used objective in all of the following case studies was the proposed FloodSkill criterion.

6.2.1 Model calibration with limited observation data

At first we calibrated WaSiM with limited observation data. Therefore the hydrological model was calibrated for each of the 24 flood events and subsequently validated on all 24 flood events. Flood event 14 was used as control set, because WaSiM can portray this event very well, so serious outliers in the measurements and massive occurrence of non considered runoff problems should not be a problem for this event. Furthermore its peak flow value is close to the center with respect to all observed flood events considered in this study. For the calibration of event 14 itself we chose event 12 as control event. Table 6 shows the calibration performance for the estimated parameter vectors for each algorithm. It is evident that all calibration algorithms achieve a reasonable calibration performance. For most of the cases the $AROPE_{MC}$ algorithm and the IPM are outperformed by the and GA. That is not disappointing, because the goal of the $AROPE_{MC}$ algorithm is primarily not to achieve a better calibration performance but to estimate a more robust and preferably better validation performance with respect to the parameter vectors estimated by classical optimisation. For the events marked with "∗" overfitting was an issue and was limited by the control event. An example[3] of the effectiveness of the overfitting stopping criterion for the calibration with the flood event 4 is given in Fig. 6.

The results of the mean model performance over all validation events for the specific flood events used for calibration is shown in Fig. 11. Detailed statistics of the overall validation performance averaged over all calibration events are given in Table 7. From the plots in Fig. 11 it is obvious that regardless from the used parameter estimation algorithm the validation performance for the single event calibration estimates are very volatile and are strongly dependent on the used calibration event. Referring to the FloodSkill criterion and the NS, the validation performance of the $AROPE_{MC}$ estimates averaged over the results of all 24 single event calibrations is slightly better than those

---

[3]Consider that this example was already discussed within the presentation of the principles of the $AROPE_{MC}$ algorithm in the previous section.

estimated by GA and IPM. Surprisingly the parameter vectors estimated by the GA have by far the worst validation performance. These results show that one flood event is not sufficient to estimate a stable solution and therefore also cannot confirm the supposed advantages of the approach.

5 A deeper investigation of the geometrical structure of the estimated sets of parameter vectors reveals possible explanations for this results and can also be an indicator for the stability of the achieved solution. Consider again that the underlying assumption of the sampling of parameter vectors by data depth is that the model performance both on the calibration and validation data of parameter vectors deep within the set

10 of good ones is within the upper range of the set and has a smaller spread than the ones at the boundary. The scatter plots in Fig. 10 show the correlation of the validation performance of parameter vectors estimated by $AROPE_{MC}$ with respect to their data depth. Referring to the NS, parameter vectors with higher depth have a consistently good model performance, whereas for the results calculated by the rPD criterion such

15 a relationship can not be shown at all. These results are strongly confirmed by the validation results calculated by the NS and the rPD (cf. Table 7). For the NS criterion the validation results of the $AROPE_{MC}$ estimates are significantly better in mean than the ones estimated by pure optimisation and have a smaller standard deviation. However, referring to the rPD criterion the $AROPE_{MC}$ estimates perform even worse than the

20 ones estimated by IPM. The problems for the rPD criterion are due to the fact that for the calibration with one single flood event, the rPD is calculated by the comparison of just two values. Random errors in the observed peak value and problems of the model structure to simulate that value can result in a spiky and not well-defined structure of the set of parameter vectors with good model performance. It is obvious that the

25 application of data depth to such kind of problems does not make sense. This might be avoided by both the use of more flood events for calibration and the the use of "smooth" performance criteria which do not account for just a very small number of observations measurements.

6.2.2 Calibration with multiple flood events

After the analysis of the results for the calibration with one single flood event, we decided to study the improvement of the validation results of $AROPE_{MC}$ in comparison with IPM and GA using more observation data for calibration and control. In particular we studied the geometrical shape and the stability of the estimated solutions. Therefore we divided the set of 24 flood events into three subsets, a set of calibration events, a set of control events and a set of validation events as given in Table 8. The assignment of the events to one of the three was rather arbitrary, but we tried to keep the proportion of events with high and low peak flow values balanced within the subsets. The same holds true for the type of its precipitation (convective vs. stratiform). Again we used the FloodSkill criterion as objective, in particular to be able to compare the results with those of the calibration with single flood events.

Figure 12 illustrates how the sets of candidate parameter vectors evolve after each iteration of the $AROPE_{MC}$ algorithm and form a geometrically well-structured cloud. A scatter plot of the estimated parameter vectors of all three compared parameter estimation algorithms is given in Fig. 13. The sets of parameter vectors estimated by IPM and GA form geometrically less-defined clouds. Furthermore the central region of the set estimated by GA is roughly approximated by the set estimated by $AROPE_{MC}$, whereas the set estimated by the gradient based IPM algorithm has another form. We validated all estimated results for all flood events in the validation set (see Table 8). For better comparison we also computed the model performance for the parameter vectors estimated by the single event calibration in the previous case study on the validation set in this case study. A boxplot of the overall validation results referring to the FloodSkill criterion is given in Fig. 14. Detailed statistics for all referred performance criteria are given in Table 9. It is evident that the use of more data for model calibration improves the model performance in validation tremendously for all three approaches. $AROPE_{MC}$ outperforms IPM and GA for the used objective, the FloodSkill criterion and the NS.

Furthermore for the calibration objective also the standard deviation of the validation results is significantly smaller for the parameter vectors estimated by AROPE$_{MC}$ than those estimated by IPM and GA which indicates that the transfer of these parameters is more reasonable. Referring to the rPD the parameter vectors estimated by AROPE$_{MC}$ have approximately the same validation performance as those estimated by IPM and GA. However, the best parameter vectors of the cloud in AROPE$_{MC}$ perform slightly better than the best ones estimated by IPM and GA.

Additionally we checked whether the set of parameter vectors estimated by AROPE$_{MC}$ form a stable solution, that means whether vectors with higher data depth not just have less standard deviation in their corresponding model performance but also tend to have a better model performance on the validation data. Therefore we calculated the correlation between the data depth of each parameter in the final set with respect to the complete set and its model performance on the validation events. The results are given in Table 10. Besides on event 16 the parameter vectors with higher data depth tend to have a better model performance on each single validation event and on the overall set of validation events. The correlation is much stronger for the NS than for the rPD. Thus, the estimated set seems to be more robust with respect to the NS than to the rPD. This might be due to still too less calibration events or due to problems of the model structure to represent the global system behaviour and the peak flow values equally well. Possibly the good parameter sets with respect to the rPD criterion for the given calibration events still do not form a well-structured geometrical set. Consider that the rPD for three calibration events is just computed by comparison with three observations. This might by tackled by more calibration data or by the use of other approaches than iterative Monte-Carlo simulations for the identification of parameter vectors with good model performance.

Figure 15 shows the simulated discharge for four validation events computed by the estimated parameter sets of all three compared parameter estimation algorithms. The catchment characteristics are in general better represented by the model runs with the estimates of AROPE$_{MC}$ than the estimates in mean and has a lower standard

deviation. However the results for flood event 4 also show the slight disadvantage of the AROPE$_{MC}$ estimates compared to the GA estimates with respect to the rPD criterion.

Due to the high process dynamics and the small catchment size (uncertainties and errors are not shrinked to the same degree by averaging of many values) the confidence bounds of the model performance are higher than those presented by Bárdossy and Singh (2008). Nevertheless, the results of the multiple events calibration study confirm the outcome of Bárdossy and Singh (2008): the application of the principle of data depth can be very useful to estimate a set of robust parameter vectors considering uncertainties. However, considering the results of the single event calibration it is also evident that in cases where no geometrically well-structured parameter set of model parameter vectors with good model performance can be identified, the application of the principle of data depth is not suitable. This problem might be avoided considering the selection of appropriate performance criteria and the required amount of observation data to be used for calibration. Furthermore the previous application of approved population based algorithms might be useful in order to identify sets of good parameter vectors with a more complicated geometrical structure, before deep parameter vectors are selected.

7 Discussion and conclusions

– This paper presents a depth based parameter estimation method, which is well suited for the robust calibration of hydrological models considering uncertainties. The Advanced Robust Parameter Estimation by Monte Carlo (AROPE$_{MC}$), is a modified version of the depth based parameter estimation procedure presented by Bárdossy and Singh (2008). There are two differences between the AROPE$_{MC}$ algorithm and the original ROPE$_{MC}$ algorithm. The further development enables us sampling from different non-convex regions of attraction and at the same time preventing AROPE$_{MC}$ from overfitting calibration data. We compare the effectiveness of the newly developed algorithm for estimating robust model parameter

vectors with the original ROPE$_{MC}$ algorithm and the GA and IPM algorithms in three case studies.

- The first case study reveals that the AROPE$_{MC}$ algorithm estimates a set of parameter vectors in the region with the highest possible performance for 2 benchmark functions. The new sampling strategy leads to a better performance of AROPE$_{MC}$ if compared to the original ROPE$_{MC}$ algorithm for problems with multiple regions of attraction.

- The results of the second case study in this paper show that just the small observational uncertainty of the discharge leads to a high variability of the model performance in validation. Parameter vectors with equal model performance on the calibration data can lead to very different results in validation. The proposed method of an evolutionary sampling of model parameter vectors by the help of data depth functions can help to identify sets of robust parameter vectors. Parameters with low data depth are near the boundary and are sensitive to small changes and do transfer to other time periods less well as high depth ones.

- Especially for processes with high dynamics (short time steps in the models), the selection of appropriate performance criteria and the required amount of observation data have to be considered to estimate robust model parameter vectors.

- In this paper, model performance was expressed by just one aggregated objective function. The presented algorithm can be easily altered to a general multi-objective parameter estimation procedure.

We propose further research on the application of data depth functions for parameter estimation. We suggest merging the concept of depth based sampling with the strength of approved search strategies for high-dimensional parameter spaces, e.g. the particle swarm concept in order to overcome the shortcomings of the Monte Carlo based approach in order to generate sets of good parameter vectors. In this study we compared

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

the data depth based approach with other state of the art optimisation algorithms coping with uncertainty in the output data. A comparison with other multi-objective model uncertainty methods, e.g. MCMC or multi-objective optimisation algorithms might be very useful and is in the scope of future research.

References

Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrol. Process., 6, 279–298, doi:10.1002/hyp.3360060305, 1992. 2425

Bos, A. and de Vreng, A.: Parameter optimization of the HYMOD model using SCEM-UA and MOSCEM-UA, Tech. rep., University of Amsterdam, The Netherlands, Faculty of Science, 2006. 2426

Bremner, D., Chen, D., Iacono, J., Langerman, S., and Morin, P.: Output-sensitive algorithms for Tukey depth and related problems, Stat. Comput., 18, 259–266, 2008. 2427

Bárdossy, A.: Calibration of hydrological model parameters for ungauged catchments, Hydrol. Earth Syst. Sci., 11, 703–710, doi:10.5194/hess-11-703-2007, 2007. 2426

Bárdossy, A. and Singh, S. K.: Robust estimation of hydrological model parameters, Hydrol. Earth Syst. Sci., 12, 1273–1283, doi:10.5194/hess-12-1273-2008, 2008. 2426, 2427, 2433, 2435, 2440, 2445, 2467

Conn, A. R., Gould, N. I. M., and Toint, P. L.: A Globally Convergent Augmented Lagrangian Barrier Algorithm for Optimization with General Inequality Constraints and Simple Bounds, Math. Comput., 66, 261–288, 1997. 2438

Cramer, K.: Multivariate Ausreißer und Datentiefe, Ph.D. thesis, Wirtschafts- und Sozialwissenschaftliche Fakultät, Universität zu Köln, Köln, Germany, 2003. 2427, 2432

Cullmann, J.: Online flood forecasting in fast responding catchments on the basis of a synthesis of artificial neural networks and process models, vol. 4 of Dresdner Schriften zur Hydrologie, Institut für Hydrologie und Meteorologie, Technische Universität Dresden, Dresden, Germany, 2006. 2425, 2428, 2430

5 Cullmann, J. and Wriedt, G.: Joint application of event-based calibration and dynamic identifiability analysis in rainfall-runoff modelling: implications for model parametrisation, J. Hydroinform., 10, 301–316, 2008. 2425

Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum likelihood from incomplete data via the EM algorithm, J. Roy. Stat. Soc. B, 39, 1–38, 1977. 2435

10 Donoho, D. L. and Gasko, M.: Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness, Ann. Stat., 20, 1803–1827, 1992. 2432

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models, Water Resour. Res., 28, 1015–1031, 1992. 2425, 2434

Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: A comparison of alternative multi-
15 objective calibration strategies for hydrological modeling, Water Resour. Res., 43, W03434, doi:10.1029/2006WR005098, 2007. 2425

Frost, A. J., Thyer, M. A., Srikanthan, R., and Kuczera, G.: A general Bayesian framework for calibrating and evaluating stochastic models of annual multi-site hydrological data, http://www.sciencedirect.com/science/article/B6V6C-4NHM5BR-1/2/
20 e33fcfea8d4bf5fa4a5b938c8cd07660, J. Hydrol., 340, 129–148, 2007. 2426

Green, W. and Ampt, G.: Studies on Soil Physics: I. The flow of air and water through soils, J. Agr. Sci., 4, 1–24, 1911. 2429

Grundmann, J.: Analyse und Simulation von Unsicherheiten in der fláchendifferenzierten Niederschlags-Abfluss-Modellierung, vol. 8 of Dresdner Schriften zur Hydrologie, Institut für
25 Hydrologie und Meteorologie, Technische Universität Dresden, Dresden, Germany, 2010. 2426, 2428, 2430, 2434

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resour. Res., 34, 751–763, 1998. 2425

30 Gurtz, J., Baltensweiler, A., and Lang, H.: Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins, Hydrol. Process., 13, 2751–2768, 1999. 2428

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Gurtz, J., Verbunt, M., Zappa, M., Moesch, M., Pos, F., and Moser, U.: Long-term hydrometeo-rological measurements and model-based analyses in the hydrological research catchment Rietholzbach, J. Hydrol. Hydromech., 51, 1–13, 2003a. 2428

Gurtz, J., Zappa, M., Jasper, K., Lang, H., Verbunt, M., Badoux, A., and Vitvar, T.: A compar-ative study in modelling runoff and its components in two mountainous catchments, Hydrol. Process., 17, 297–311, doi:10.1002/hyp.1125, 2003b. 2428

Hogue, T. S., Sorooshian, S., and Gupta, H.: A Multistep Automatic Calibration Scheme for River Forecasting Models, J. Hydrometeorol., 1, 524–542, 2000. 2425

Hugg, J., Rafalin, E., Seyboth, K., and Souvaine, D.: An Experimental Study of Old and New Depth Measures, in: Workshop on Algorithm Engineering and Experiments (ALENEX06), 51–64, 2006. 2432, 2434

Kobold, M., Sušelj, K., Polajnar, J., and Pogačnik, N.: Calibration Techniques Used For HBV Hydrological Model in Savinja Catchment, in: XIV-th Conference of the Danubian Countries On The Hydrological Forecasting And Hydrological Bases of Water Management, edited by: Brilly, M. and Šraj, M., Slovenian National Committee for the IHP UNESCO, 2003. 2425

Krauße, T.: Data depth – parsimonious sampling of robust parameters in hydrological models, Appl. Math. Modell., submitted, 2011. 2432, 2435

Kuczera, G. and Parent, E.: Monte Carlo assessment of parameter uncertainty in concep-tual catchment models: the Metropolis algorithm, http://www.sciencedirect.com/science/article/B6V6C-3VKCC3M-6/2/d6e9d6d6db0a01eeb1af268e11b7d545, J. Hydrol., 211, 69–85, 1998. 2427

Kuczera, G., Kavetski, D., Franks, S., and Thyer, M.: Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, http://www.sciencedirect.com/science/article/B6V6C-4KBVX50-2/2/b066c6156e2b79860d5e07dd57111ff2, J. Hydrol., 331, 161–177, 2006. 2426, 2440

Kunstmann, H., Krause, J., and Mayr, S.: Inverse distributed hydrological modelling of Alpine catchments, Hydrol. Earth Syst. Sci., 10, 395–412, doi:10.5194/hess-10-395-2006, 2006. 2425

Liu, R. Y., Serfling, R., and Souvaine, D. L.: Data Depth: Robust Multivariate Analysis, Compu-tational Geometry and Applications, vol. 72 of Series in Discrete Mathematics and Theoreti-cal Computer Science, American Mathematical Society, USA, 2006. 2427, 2432

Marx, A.: Einsatz gekoppelter Modelle und Wetterradar zur Abschätzung von Niederschlagsin-tensitäten und zur Abflussvorhersage, vol. 160 of Mitteilungen des Instituts für Wasserbau,

Identification of model parameters by data depth measures

T. Krauße and J. Cullmann

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄◄ | ►►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Universität Stuttgart, Institut für Wasserbau, Universität Stuttgart, Stuttart, Germany, 2007. 2425

Mühlenbein, H., Schomisch, D., and Born, J.: The Parallel Genetic Algorithm as Function Optimizer, Parallel Comput., 17, 619–632, 1991. 2438

5 Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I – A discussion of principles, http://www.sciencedirect.com/science/article/B6V6C-487FF7C-1XH/1/75ac51a8910cad95dda46f4756e7a800, J. Hydrol., 10, 282–290, 1970. 2431

Oja, H.: Descriptive Statistics for Multivariate Distributions, Stat. Prob. Lett., 1, 327–332, 1983. 2432

10 Pompe, K.: Untersuchungen zur optimalen Parametrisierung von Prozessmodellen in der Hydrologie, Diplomarbeit, Institut für Hydrologie und Meteorologie, Technische Universität Dresden, Dresden, Germany, 2009. 2427, 2430

Rousseeuw, P. and Hubert, M.: Regression depth, J. Am. Stat. Assoc., 94, 388–402, 1999. 2432

15 Rousseeuw, P. J. and Struyf, A.: Computing location depth and regression depth in higher dimensions, Stat. Comput., 8, 193–203, 1998. 2432

Schulla, J.: Hydrologische Modellierung von Flussgebieten zur Abschätzung der Folgen von Klimaänderungen, Ph.D. thesis, Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland, diss. ETH Nr. 12018, 1997. 2428, 2430

20 Serfling, R.: Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications, vol. 72 of DIMACS Series in Discrete Mathematics and Theoretical Computer Science, chap. Depth functions in nonparametric multivariate inference, American Mathematical Society, 1–16, 2006. 2432

Sorooshian, S., Duan, Q., and Gupta, V. K.: Calibration of rainfall-runoff models: Application of
25 global optimization to the Sacramento Soil Moisture Accounting Model, Water Resour. Res., 29, 1185–1194, doi:10.1029/92WR02617, 1993. 2434

Tetko, I., Livingstone, D., and Luik, A.: I. Neural network studies, 1. Comparison of overfitting and overtraining, J. Chem. Inform. Comput. Sci., 35, 826–833, 1995. 2436

Theiner, D. and Wieczorek, M.: Reduction of Calibration Time of Distributed Hydrological Mod-
30 els by use of Grid Computing and Nonlinear Optimisation Algorithms, in: Proceedings of the 7th International Conference on Hydroinformatics HIC 2006, Nice, 2006. 2425

Thyer, M., Frost, A. J., and Kuczera, G.: Parameter estimation and model identification for stochastic models of annual hydrological data: Is the observed

Identification of model parameters by data depth measures

T. Krauße and
J. Cullmann

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

|◀ | ▶|

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

record long enough?, http://www.sciencedirect.com/science/article/B6V6C-4K428WH-2/2/899aa9b003ba9b596314c0e934b79334, J. Hydrol., 330, 313–328, 2006. 2426

Tukey, J. W.: Mathematics and the picturing of data, in: Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), vol. 2, Canad. Math. Congress, Montreal, Que., 523–531, 1975. 2431, 2432

Vencálek, O.: Weighted data depth and its properties, in: Robust'2008, Charles University, Prague, Czech Republic, 2008. 2432

Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, Water Resources Research, 39(8), 1201, doi:10.1029/2002Wr001642, 2003. 2426

Wagener, T., Wheater, H. S., and Gupta, H. V.: Rainfall-Runoff Modelling in Gauged and Ungauged Catchments, Imperial College Press, London, UK, 2004. 2425

Waltz, R., Morales, J., Nocedal, J., and Orban, D.: An interior algorithm for nonlinear optimization that combines line search and trust region steps, Math. Program., 107, 391–408, 2006. 2440

Zappa, M.: Multiple-Response Verification of a Distributed Hydrological Model at Different Spatial Scales, Ph.D. thesis, Swiss Federal Institute of Technology (ETH) Zürich, Zürich, Switzerland, 2002. 2428

Zuo, Y. and Serfling, R.: General notions of statistical depth function, Ann. Stat., 28, 461–482, 2000. 2431

Table 1. Overview of the most important basin characteristics.

area	3.18 km^2
maximum altitude	938 m a.s.l.
altitude at basin outlet	681 m a.s.l.
mean altitude	796 m a.s.l.
mean slope	12.5˚
pasture land	67%
forest	25%
wetland	4%
bushes	2%
roads	2%

Table 2. Overview of the used model parameters considered for calibration; the reference parameter vector $\boldsymbol{\theta}_{wb}$ was estimated in order to use WaSiM for water-balance simulations in the Rietholzbach catchment.

parameter	reference ($\boldsymbol{\theta}_{wb}$)	upper and lower boundary		description
k_d [h]	7	0.01	25	storage coefficient of direct runoff
k_i [h]	20	0.01	40	storage coefficient of interflow
dr [−]	2.1	0.01	80	drainage density

Table 3. Overview of the database of the 24 flood events used for calibration and validation sorted by peak flow value.

	Event		Discharge		Rainfall		Pre Soil Moisture
No.	Time (CET)	Hydrograph	Peak [mm h^{-1}]	Volume [mm]	Type	Volume [mm]	$\Theta_{0-0.5m}$ [vol %]
1	6 Jul 1994 24:00	---	7.61	33.2	convective	75.0	65
2	23 Jun 1986 22:00	---	7.52	20.2	convective	39.7	75
3	6 Jul 1994 21:00	---	6.23	33.2	convective	75.0	65
4	8 Aug 2007 20:00	---	5.69	51.2	convective	64.1	90
5	7 Jun 2007 23:00	---	5.48	16.0	convective	36.1	79
6	6 Aug 1982 19:00	---	5.35	21.2	convective	28.3	87
7	25 Jul 1989 20:00	---	4.64	11.7	convective	13.7	80
8	30 May 1995 22:00	---	4.01	16.9	convective	39.4	84
9	15 Aug 1982 20:00	---	4.01	18.3	convective	43.6	80
10	6 Aug 2000 13:00	---	3.61	28.9	stratiform	48.8	83
11	17 Sep 2006 09:00	---	3.13	36.1	convective	101.8	71
12	24 Sep 2002 10:00	---	3.04	120.1	stratiform	125.2	87
13	9 Sep 2001 01:00	---	2.96	18.4	convective	49.4	78
14	9 Jun 1994 07:00	---	2.52	27.9	stratiform	40.7	86
15	11 May 1991 24:00	---	2.50	44.6	stratiform	75.3	80
16	1 Jul 1987 16:00	---	2.37	12.1	convective	30.7	78
17	2 Aug 2005 22:00	---	2.03	26.0	convective	54.2	74
18	25 May 1990 11:00	---	2.02	14.3	stratiform	54.9	75
19	11 Jun 1995 07:00	---	1.98	28.8	stratiform	36.8	88
20	31 May 2000 09:00	---	1.76	19.4	stratiform	57.7	78
21	9 Sep 2005 21:00	---	1.62	6.9	convective	37.1	67
22	28 Sep 1995 01:00	---	1.47	17.3	stratiform	38.2	81
23	13 May 2002 01:00	---	1.02	4.6	convective	21.4	81
24	14 Sep 1993 01:00	---	1.02	10.9	stratiform	26.8	84

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄◄ | ►►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 4. Objective functions used in this study, where x_i and $y_i(\boldsymbol{\theta})$ are the observed and simulated discharge (by the parameter vector $\boldsymbol{\theta}$) at time-step i respectively and n is the number of observation points.

Criterion	Formula
Nash-Suttcliffe efficiency (NS)	$1 - \dfrac{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i(\boldsymbol{\theta}))^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
rel. peak flow deviation (rPD)	$\dfrac{|x_{\max} - y_{\max}(\boldsymbol{\theta})|}{x_{\max}}$
FloodSkill	$0.5\,(-(NS - 1)) + 0.5\,(rPD)$

Table 5. Comparison of the fitness values for the estimates computed by the three used algorithms.

Function	Dimension	GA	ROPE$_{MC}$		AROPE$_{MC}$	
			mean	10% deepest	mean	10% deepest
f_a	2	$6.86\,e-3$	$5.04\,e-2$	$2.41\,e-2$	$4.03\,e-2$	$3.33\,e-2$
	3	$1.78\,e-1$	$6.00\,e-2$	$7.61\,e-2$	$1.12\,e-2$	$9.42\,e-3$
	4	$5.34\,e-1$	$4.64\,e-2$	$2.63\,e-2$	$4.71\,e-2$	$1.07\,e-2$
f_b	2	$1.49\,e-2$	$7.87\,e-1$	$1.96\,e-1$	$2.67\,e-3$	$1.47\,e-3$
	3	$5.15\,e-2$	$2.44\,e+0$	$1.87\,e+0$	$7.34\,e-1$	$7.25\,e-2$
	4	$1.79\,e-1$	$7.65\,e+0$	$7.81\,e+0$	$2.40\,e+0$	$2.56\,e+0$

Table 6. Mean calibration results (FloodSkill) for the three compared algorithms.

Calib Event	$AROPE_{MC}$	IPM	GA
1	0.28*	0.05	0.04
2	0.59	0.59	0.43
3	0.49	0.43	0.42
4	0.23*	0.14	0.14
5	0.55	0.51	0.41
6	0.51*	0.45	0.44
7	0.92*	0.84	0.72
8	0.26	0.22	0.14
9	0.28	0.35	0.25
10	0.10	0.49	0.34
11	0.36	0.14	0.10
12	0.41*	0.20	0.08
13	0.28	0.25	0.23
14	0.09	0.22	0.16
15	0.30	0.21	0.13
16	0.46	0.42	0.39
17	0.11	0.40	0.19
18	0.19	0.41	0.34
19	0.13*	0.09	0.10
20	0.38	0.25	0.19
21	0.12	0.06	0.06
22	0.33	0.22	0.13
23	0.15	0.37	0.31
24	0.24*	0.53	0.23
Overall	0.32	0.33	0.24

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀ | ▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 7. Mean overall validation results of the single event calibration for the three compared algorithms.

	FloodSkill				NS				rPD			
	μ	σ	worst	best	μ	σ	worst	best	μ	σ	worst	best
AROPE$_{MC}$	0.49	0.009	0.52	0.47	0.41	0.018	0.37	0.45	0.40	0.005	0.41	0.39
IPM	0.52	0.025	0.58	0.47	0.32	0.043	0.22	0.43	0.36	0.015	0.40	0.33
GA	0.68	0.037	0.78	0.60	0.03	0.063	−0.13	0.18	0.40	0.014	0.43	0.36

Table 8. Sub-division of all flood events in a calibration, control and validation set.

Set	Used flood events
Calibration	{6, 11, 20}
Control	{14, 18, 24}
Validation	{1, 2, 3, 4, 5, 7, 8, 9, 10, 12, ... 13, 15, 16, 17, 19, 21, 22, 23}

Table 9. Mean overall validation performance for the parameter vectors estimated by the three compared algorithms; the model performance was calculated on the validation set according to Table 8 for both the calibration with single flood events (above) and for the multiple event calibration (below).

	FloodSkill				NS				rPD			
	μ	σ	worst	best	μ	σ	worst	best	μ	σ	worst	best
single event calibration												
AROPE$_{MC}$	0.49	0.010	0.51	0.46	0.41	0.018	0.37	0.45	0.40	0.006	0.41	0.39
IPM	0.51	0.021	0.56	0.47	0.32	0.043	0.22	0.43	0.36	0.015	0.40	0.33
GA	0.59	0.026	0.65	0.53	0.03	0.063	−0.13	0.18	0.40	0.014	0.43	0.36
multiple event calibration												
AROPE$_{MC}$	0.41	0.015	0.44	0.37	0.55	0.036	0.47	0.61	0.38	0.030	0.44	0.30
IPM	0.43	0.044	0.58	0.38	0.52	0.049	0.33	0.58	0.37	0.048	0.52	0.32
GA	0.42	0.024	0.46	0.38	0.53	0.033	0.43	0.61	0.38	0.030	0.46	0.32

Table 10. Correlation between data depth and validation performance of all parameter vectors estimated with $AROPE_{MC}$ for the multiple event calibration.

Event	FloodSkill	NS	rPD
1	−0.16	0.18	−0.14
2	−0.05	0.10	−0.01
3	−0.25	0.32	−0.16
4	−0.13	0.16	−0.09
5	−0.19	0.30	−0.11
7	−0.09	0.12	−0.04
8	−0.12	0.17	−0.06
9	−0.17	0.29	−0.01
10	−0.22	0.27	−0.14
12	−0.27	0.29	−0.24
13	−0.23	0.27	−0.16
15	−0.29	0.35	−0.16
16	0.16	−0.15	0.16
17	−0.20	0.23	−0.06
19	−0.32	0.40	−0.11
21	−0.19	0.34	−0.10
22	−0.32	0.36	−0.05
23	−0.44	0.14	−0.33
Overall	−0.25	0.33	−0.13

Fig. 1. The Rietholzbach catchments with the main measurement site "Büel" on 7 May 2005 at the beginning of the summer season. You see the typical pastures which cover two third of the catchment area and some of the sporadic small patches of forest in the background.
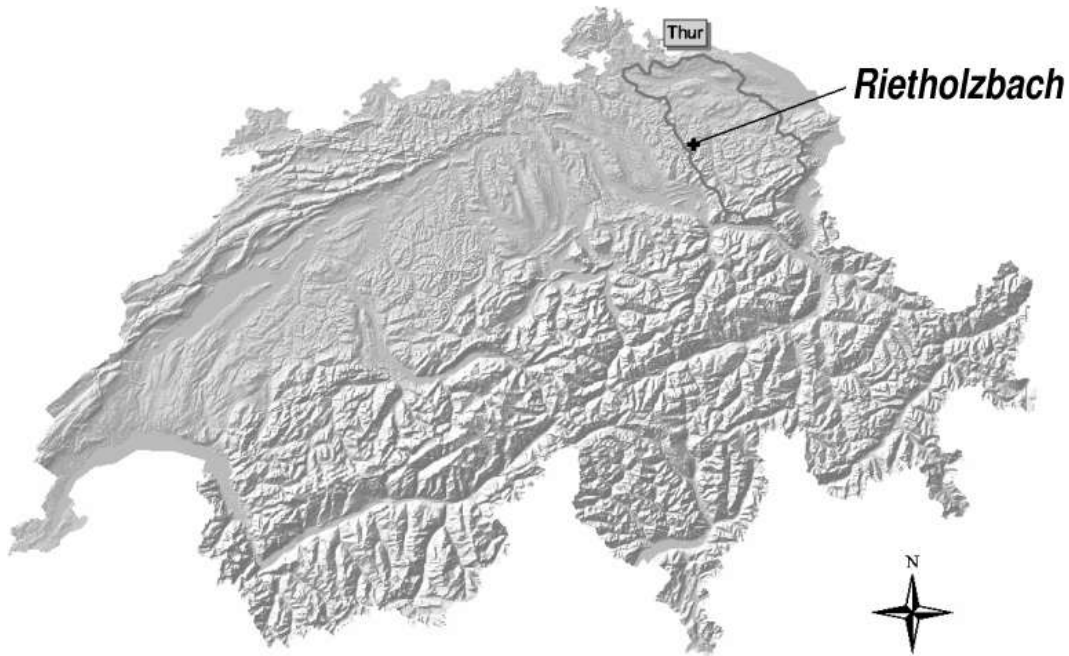
Fig. 2. The Rietholzbach catchment is located in the north-east of Switzerland.
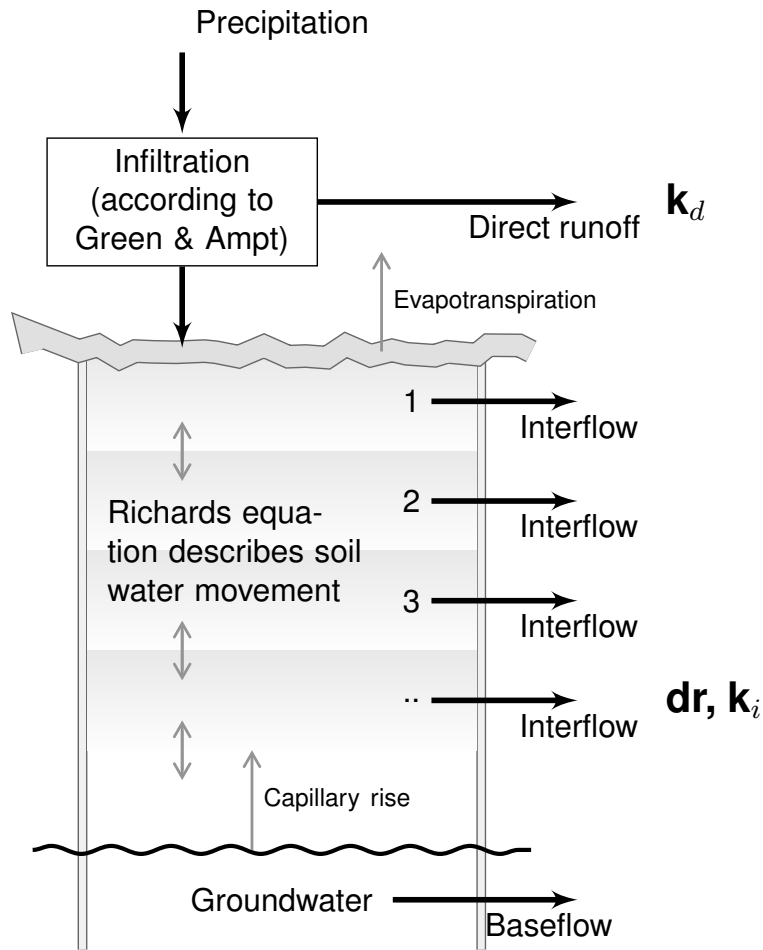
Fig. 3. Scheme of the WaSiM soil module with location of impact of conceptual model parameters (bold).

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
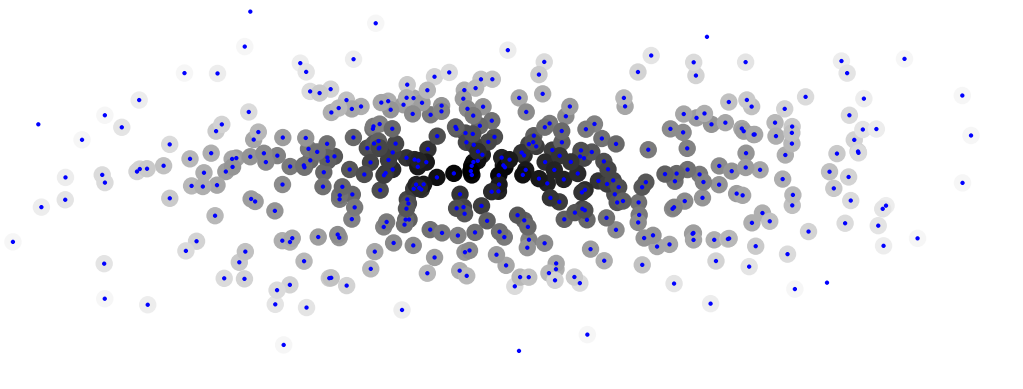
Fig. 4. 2-dimensional point set shaded according to assigned depth. A darker point represents higher depth. The used depth function was halfspace depth.
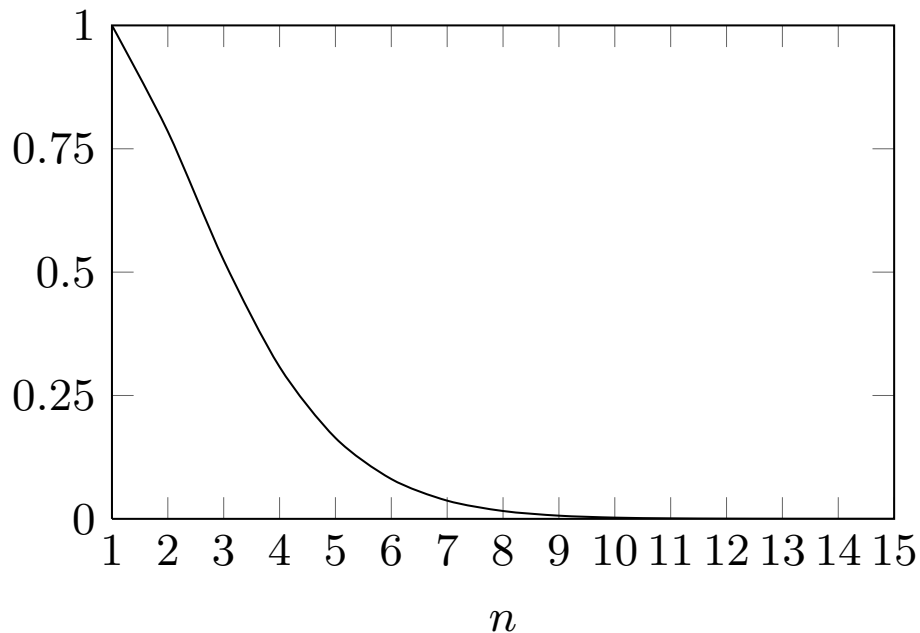
Fig. 5. Volume ratio of the unit sphere to the unit cube in n dimensions as a continuous function of n.
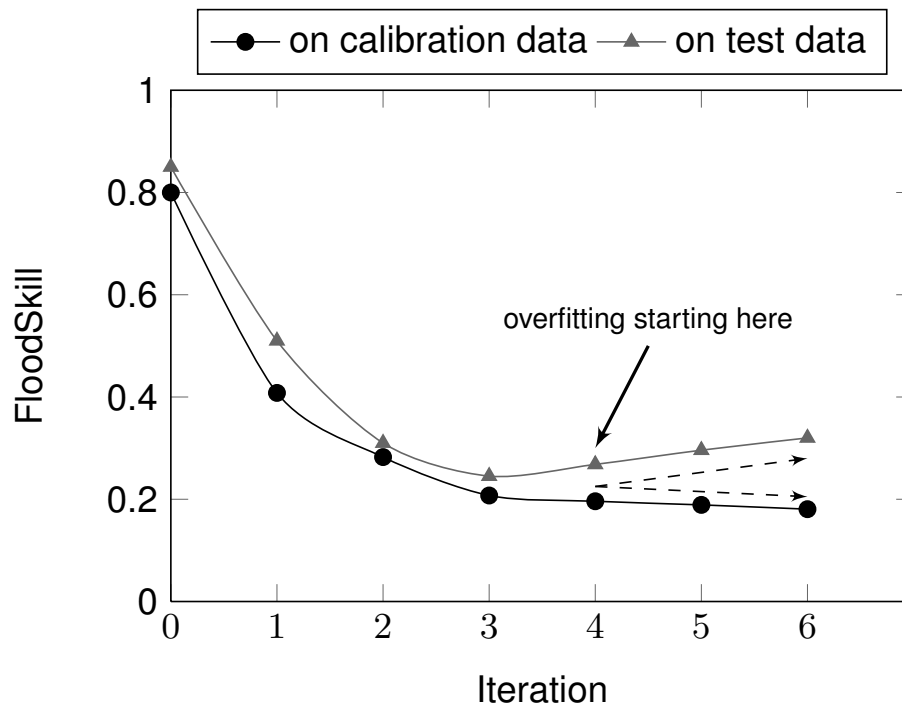
Fig. 6. Overfitting for the calibration of the model WaSiM while calibrated with the method according to Bárdossy and Singh (2008); Flood event no. 4 was used for calibration and event no. 14 was used as control set.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

START

Input: d=dimension, n=population size, m=final size
constraints=CHR_d, priors=$\{p_1(\boldsymbol{\theta}),\ldots,p_n(\boldsymbol{\theta})\}$
maxIter=maximum no. of iterations

if possible, split the used observation data $\{\eta,\boldsymbol{x}\}$ into a calibration and control set

Sample n parameter vectors in the feasible space, Θ, according to the given priors and constraints:

$\mathbf{X_n} := sample(\Theta, \{p_1(\boldsymbol{\theta}),\ldots,p_d(\boldsymbol{\theta})\}, CHR_d);$

Identify good

The model η is run for the calibration and control set; the corresponding model performances are calculated by a purpose-specific objective function U:

$\forall \boldsymbol{\theta} \in \mathbf{X_n}$ **do** :

$\hat{\mathbf{y}} := \eta(\xi,\boldsymbol{\theta});$
$perf(\boldsymbol{\theta}) := U(\mathbf{x},\hat{\mathbf{y}});$

endfor

A subset $X_n^* \subset X_n$ of the good performing parameter vectors in X_n is identified, e.g. such that X_n^* comprises the best performing 10% of all parameters vectors in X_n:

$\mathbf{X_n^*} := identify(X_n, perf)$

Generate robust

Generate a set of deep parameters $\mathbf{Y_m}$ w.r.t. X_n^* ⟷ **GenDeep** strategy

(1) no. of iterations $\geq$ *maxIter*?

(2) Improvement on calibration data gets smaller than tolerance?

(3) Improvement on control data decreases w.r.t previous iteration?

Stopping criterion satisfied?

No → $\mathbf{X_n} := X_n^* \cup Y_m;$

Yes

Output: Y_m=set of robust parameter vectors
perf=calibration performance

STOP

Fig. 7. Flowchart of the Advanced Robust Parameter Estimation by Monte Carlo Simulation (AROPE$_{MC}$) algorithm.
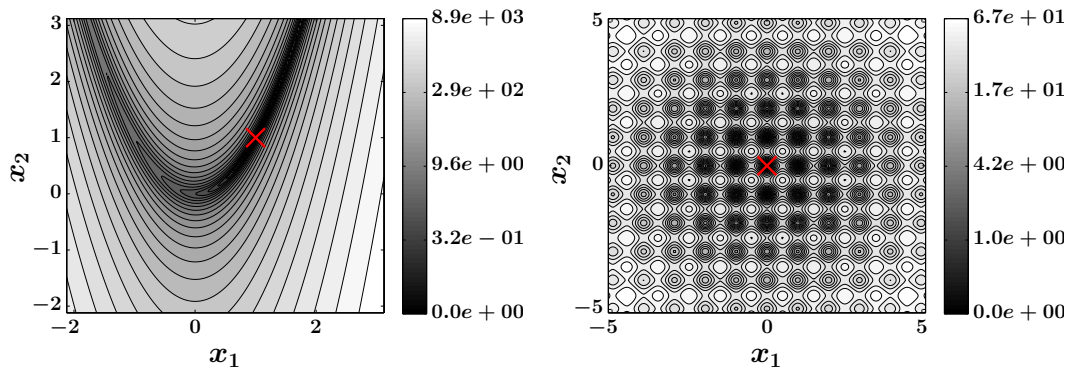
Fig. 8. Contour plot of the Rosenbrock (upper) and Rastrigin (lower) function for the 2-dimensional case; the known global optima are marked by a red cross.
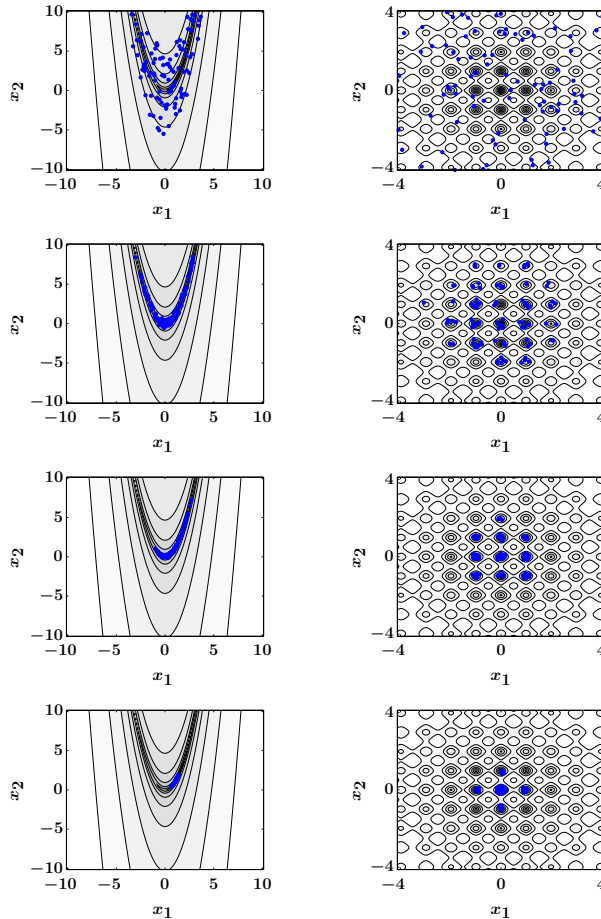
Fig. 9. Scatter plot of the results of the AROPE$_{MC}$ algorithm of the Rosenbrock (left) and Rastrigin (right) function for the 2-dimensional case after the first, second, fourth and final iteration from top to bottom.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀◀ | ▶▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
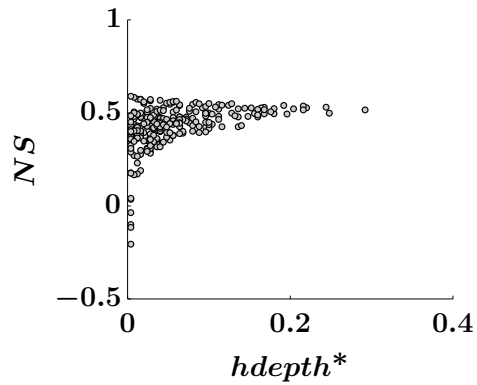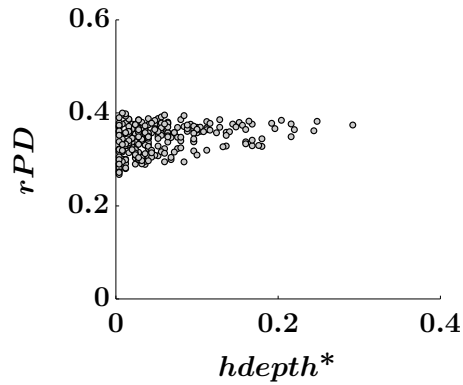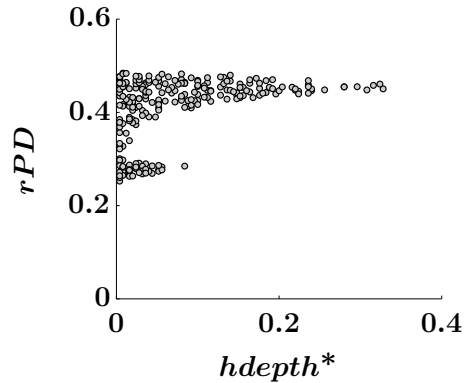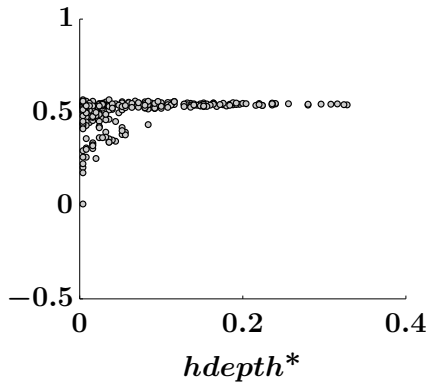
Fig. 10. Relationship between data depth and validation performance (NS and rPD) for the parameter vectors estimated for flood events 6 **(a)** and 14 **(b)**.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
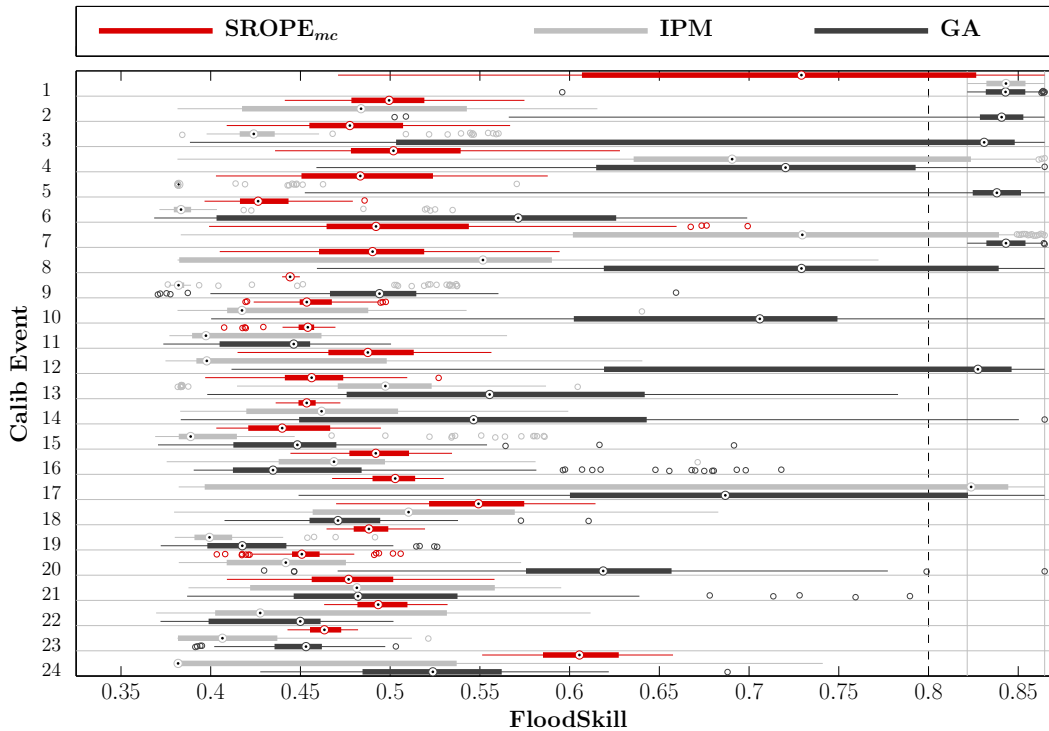
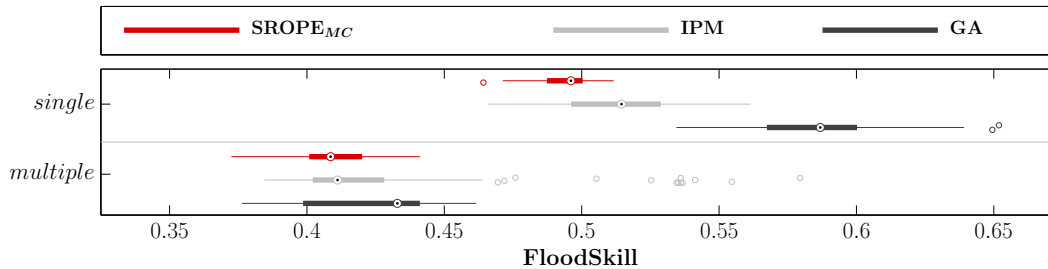**Fig. 11.** Statistics of the mean FloodSkill on the validation data for the estimates of the calibration with single flood events each.

Fig. 12. Evolution of the candidate model parameter vectors before each iteration in the AROPE$_{MC}$ algorithm for the multiple event calibration.
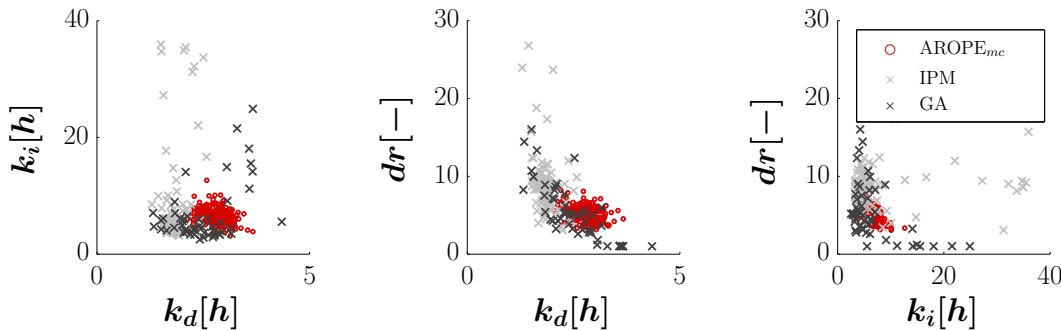
Fig. 13. Scatter plot of the final estimated parameter vectors of the three compared parameter estimation algorithms.

Fig. 14. Statistics of the mean FloodSkill on the validation data for the estimates of the calibration with single flood events each.
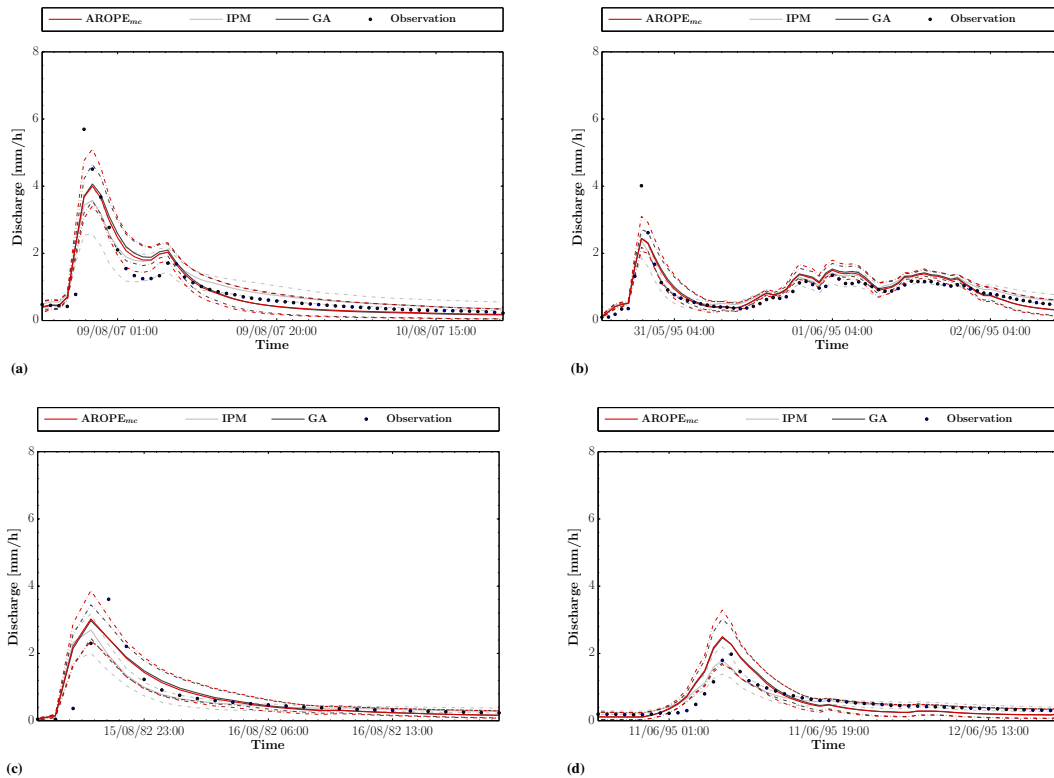
Fig. 15. Simulated hydrographs for the flood events 4 **(a)**, 8 **(b)**, 9 **(c)** and 19 **(d)** computed by the parameter vectors estimated of all three compared algorithms; the mean value is plotted as thick solid line and the confidence interval of the parameter uncertainty ($Q_{.95} - Q_{.05}$) plotted as thin dash-dot line.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

⏮ | ⏭

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion