Hydrology and Earth System Sciences

Towards a more representative parametrisation of hydrologic models via synthesizing the strengths of Particle Swarm Optimisation and Robust Parameter Estimation

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@krausze.com)

Abstract. The development of methods for estimating the parameters of hydrologic models considering uncertainties has been of high interest in hydrologic research over the last years. In particular methods which understand the estimation of hydrologic model parameters as a geometric search of a set of robust performing parameter vectors by application of the concept of data depth found growing research interest. Bárdossy and Singh (2008) presented a first Robust Parameter Estimation Method (ROPE) and applied it for the calibration of a conceptual rainfall-runoff model with daily time step. The basic idea of this algorithm is to identify a set of model parameter vectors with high model performance called good parameters and subsequently generate a set of parameter vectors with high data depth with respect to the first set. Both steps are repeated iteratively until a stopping criterion is met. The results estimated in this case study show the high potential of the principle of data depth to be used for the estimation of hydrologic model parameters. In this paper we present some further developments that address the most important shortcomings of the original ROPE approach. We developed a stratified depth based sampling approach that improves the sampling from non-elliptic and multi-modal distributions. It provides a higher efficiency for the sampling of deep points in parameter spaces with higher dimensionality. Another modification addresses the problem of a too strong shrinking of the estimated set of robust parameter vectors that might lead to overfitting for model calibration with a small amount of calibration data. This contradicts the principle of robustness. Therefore, we suggest to split the available calibration data into two sets and use one set to control the overfitting. All modifications were implemented into a further developed ROPE approach that

is called Advanced Robust Parameter Estimation (AROPE). However, in this approach the estimation of the good parameters is still based on an ineffective Monte Carlo approach. Therefore we developed another approach called ROPE with Particle Swarm Optimisation (ROPE-PSO) that substitutes the Monte Carlo approach with a more effective and efficient approach based on Particle Swarm Optimisation. Two case studies demonstrate the improvements of the developed algorithms when compared with the first ROPE approach and two other classical optimisation approaches calibrating a process oriented hydrologic model with hourly time step. The focus of both case studies is on modelling flood events in a small catchment characterised by extreme process dynamics. The calibration problem was repeated with higher dimensionality considering the uncertainty in the soil hydraulic parameters and another conceptual parameter of the soil module. We discuss the estimated results and propose further possibilities in order to apply ROPE as a well-founded parameter estimation and uncertainty analysis tool.

1 Introduction

Hydrologic models are designed to approximate the general physical mechanism which govern the rainfall-runoff process within a specific catchment. This is why these models have found favour with many hydrologists and engineers in practice and research. Most of the hydrologic models are driven by a vector of model parameters. These parameters are supposed to be estimated in order to approximate the general system behaviour which governs the rainfall-runoff process within a specific catchment. In most cases the model

parameters cannot be related to measurements in a direct way, but are supposed to be estimated through indirect methods such as calibration. In the process of calibration, the modeller adjusts the values of the model parameters such that the model is able to closely match the behaviour of the real system it is intended to represent. Hence the success of a model application is strongly dependent on a good estimation of the model parameters.

In the past models were calibrated by hand. This is very labour-intensive and requires an experienced modeller with profound hydrologic knowledge. Thus recently automatic methods for model calibration have evolved significantly (e.g. Duan et al., 1992; Gupta et al., 1998) and have found a common acceptance and broad use in the hydrologic community (e.g. Hogue et al., 2000; Cullmann, 2006; Kunstmann et al., 2006; Marx, 2007; Grundmann, 2010). The parameter estimation of hydrologic models is affected by numerous uncertainties. Beven and Binley (1992) described the probability to estimate the same model performance for different estimated parameter vectors as the equifinality problem. Recently developed approaches address this problem by estimating the uncertainty of the model parameter vectors considering uncertainties in the observations and the model structure. The uncertainty is often expressed by providing a set of optimal parameter vectors. One well established approach for a parameter estimation including uncertainty are the Markov Chain Monte Carlo (MCMC) methods (e.g. Vrugt et al., 2003b, 2009a; Kuczera et al., 2006). These methods require the setup of a complete Bayesian uncertainty framework. One major advantage of such a 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 negligible 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. That is why also alternative approaches, e.g. the previously mentioned generalized likelihood uncertainty estimation (GLUE) or the robust parameter estimation approach (ROPE) (Bárdossy and Singh, 2008) attracted scientific interest in the hydrologic community. The development of these methods was accompanied by a strong debate in the hydrologic community regarding the requirements of an appropriate framework for uncertainty estimation. The results of recently published studies comparing the newly developed MCMC method DREAM with the GLUE framework suggest that "formal and informal Bayesian approaches have more common ground than the hydrologic literature and ongoing debate might suggest" (e.g. Vrugt et al., 2009b). Hence, the use of both Bayesian and non-Bayesian approaches might be reasonable regarding the requirements of a specific application.

Within this paper we will focus on a further development of the ROPE approach proposed by Bárdossy and Singh (2008). ROPE is a non-Bayesian approach that addresses the parameter and uncertainty estimation problem using the concept of 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 provides the possibility of a center-outward orderings of points in Euclidean space of any dimension and opens up a new non-parametric multivariate statistical analysis method in which no distributional assumptions are needed. Recent studies of computational geometry and multivariate statistics (e.g. Liu et al., 2006; Bremner et al., 2008) showed that members that are in a geometrical central position with respect to a given point set or distribution, are more robust in order to represent the whole set. These points can be estimated applying 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). This concept was basically adapted by an evolutionary parameter estimation method presented by Bárdossy and Singh (2008). They showed that it can be very useful for the estimation of robust hydrologic model parameters. This result was also found by preliminary studies with WaSiM in the Rietholzbach catchment (Pompe, 2009). In a simplified form the ROPE approach consists of two steps. In a first step a set of model parameters with good model performance is identified. According to Bárdossy and Singh (2008) these parameter vectors are from now on called the good parameter vectors. Thereafter a set of parameter vectors with high data depth with respect to the set of good parameter vectors is generated under the assumption that these parameter vectors are more likely to represent a robust solution than the complete set of good parameter vectors. Thus, they called this approach robust parameter estimation method. Within this scope the concept of robustness is related to the term of transferability. Thus, we call parameter vectors robust which not just lead to good model performance over a selected calibration time period but are transferable: they perform well for other time periods and might also perform well on other catchments. Such parameter vectors are more likely to lead to a hydrologically reasonable representation of the corresponding processes and are less sensitive. This means that small changes of the parameters should not lead to very different results on time periods.

The first results of the method provided by Bárdossy and Singh (2008) are very promising. Therefore we reviewed the presented methods and addressed some shortcomings of the

first approach by the implementation of more efficient and effective methods to sample deep parameter vectors and the introduction of well founded stopping criteria. All these ideas were assembled into a further developed version of the ROPE method, called Advanced Robust Parameter Estimation (A-ROPE). Nonetheless the A-ROPE method is still based upon the Monte Carlo method and consequently suffers from its main disadvantage, i.e. that it is slow and requires a large number of model runs to sample the feasible space with reasonable accuracy. The required number of samples increases exponentially with the number of considered parameters. That is in particular a problem for computationally intensive process-oriented models where the number of samples is strictly limited by the available computational capacity. However, the effectiveness of the depth based sampling of parameter vectors is highly dependent on the quality of the identified set of good parameter vectors. To overcome the shortcomings of the Monte Carlo method for the estimation of the good parameter vectors, we suggest to substitute it by an approved evolutionary search strategy for high-dimensional parameter spaces. We are convinced that Particle Swarm Optimisation (PSO) is a suitable candidate for this task. PSO is a search strategy that bases on the concept of swarm intelligence in order to estimate the global optimum for a given single-objective optimisation problem. We modified the search strategy used in the normal PSO algorithm in order to identify a set of good parameter vectors with given tolerance. The modification adapts ideas coming from multi-objective PSO algorithms that also estimate a set of optimal parameters instead of just one global optimum. Afterwards the second step of the ROPE procedure, the depth based parameter sampling can be applied. The new approach merges the strength of PSO and depth based sampling. It is entitled ROPE with Particle Swarm Optimisation (ROPE-PSO).

The remainder of this paper is organised as follows: After the introduction, we will introduce the concept of data depth and subsequently the ROPE method provided by Bárdossy and Singh (2008). This is followed by a presentation of the newly developed approaches A-ROPE and ROPE-PSO and the underlying concepts. The presentation of the algorithms is completed by a brief introduction of an approach provided by Grundmann (2010) that allows considering the uncertainty in soil hydraulic parameters for the calibration of hydrologic models. The general idea of this approach is important to understand the following case studies. We studied both algorithms calibrating a process-oriented hydrologic model with a high temporal resolution (hourly instead of daily time-step) in a catchment where the dominant processes have high dynamics. The focus of the model calibration is the modeling of flood events. The estimated results are discussed comparing them to estimates obtained by the first ROPE algorithm presented by Bárdossy and Singh (2008) and other automatic parameter estimation approaches.
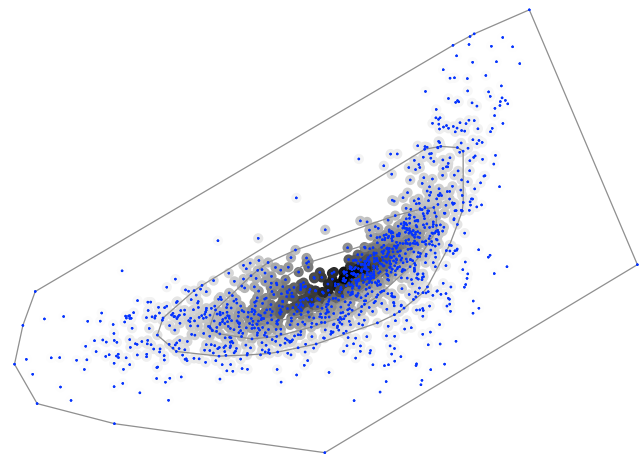

Fig. 1. 2-dimensional point set shaded according to assigned depth. A darker point represents higher depth. The lines indicate convex hulls enclosing the 25 %, 50 %, 75 % and 100 % deepest points. The used depth function was halfspace depth.

2 Parameter estimation using data depth metrics

2.1 Data depth

The algorithm applies the technique of data depth, a new approach used for multivariate data analysis that provides the possibility to analyse, quantify and visualise data sets. Most proposed metrics used in data depth function are inherently geometric, with a numeric value assigned to each data point that represents its centrality within the given data set. 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. Tukey (1975) introduced this concept first in order to estimate the center of a multivariate dataset. The possibilities of the concept of data depth is illustrated in Fig. 1 using a small 2-dimensional synthetical data set.

A formal definition of an arbitrary depth function D for the d-dimensional space $\mathbb{R}_d$ is given as follows:

$$D \colon \mathbb{R}^d \times (\mathbb{R}^d \times \mathbb{R}) \to \mathbb{R} \tag{1}$$

The following concepts apply to the data depth methodology and distinguish it from other statistical methods.

– Non-parametric methodology: scientific measurements can be viewed as sample points drawn from some unknown probability distribution, where the analysis of the measurements involves computation of quantitative characteristics of the probability distribution (estimators), based on the data set. If the underlying distribution is known (for example normal distribution, lognormal distribution, Cauchy, etc.), the characteristics of the data can be computed using methods from classical statistics. However, in most real life experiments

the underlying distribution is not known. The concept of data depth requires no assumption about the underlying distribution and data is analysed according to the relative position of the data points.

– Center-outward ordering of points: the data depth concept allows the creation of a multivariate analog to the univariate statistical analysis tool of *rank statistics*. Rank statistics is based on the ordering of one-dimensional observations, where the order reflects extremeness, contiguity, variability or the effect of external contamination. In higher dimensions the order of multivariate data is not well defined, and several ordering methods were suggested. The data depth concept provides a method of extending order statistics to any dimension by ordering the points according to their depth values.

– Application to multivariate (high-dimensional) data sets: the concept of data depth is defined with respect to points in Euclidean space in any dimension, thus enabling the derivation of multivariate distributional characteristics of a data set.

– Robustness: in the statistical analysis of data sets, observations that deviate from the main part of the data (outliers) can have an undesirable influence on the analysis of the data. Many depth functions are robust against the possibility of several outliers that may occur in the data and yield nevertheless reasonable results.

Tukey (1975) introduced this concept first with the definition of the halfspace depth. According to Donoho and Gasko (1992) the halfspace depth of an arbitrary point $\theta \in \mathbb{R}^d$ with respect to a d-dimensional data set $Z = \{z_i = (z_{i1}, \cdots, z_{id}); i = 1, \cdots, n\}$ is defined as the smallest number of data points in any closed halfspace with boundary through θ. This is also called the Tukey or location depth, and it can be written as

$$\text{hdepth}(\theta|Z) := \min_{||\boldsymbol{u}||=1} \#\{i, \boldsymbol{u}^\top z_i \geq \boldsymbol{u}^\top \theta\} \qquad (2)$$

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

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

$$\text{hdepth}^*(\theta|Z) := \frac{\text{hdepth}(\theta|Z)}{\#\{Z\}} \qquad (3)$$

The first publication of Tukey (1975) was then followed by many generalizations and other definitions of this concept, e.g. convex-hull peeling depth, simplicial depth, regression depth and L1 depth. 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). A comprehensive study of different data depth measures in robust parameter estimation is provided in Krauße and Cullmann (2011b).

2.2 The ROPE algorithm

Algorithm 1 ROPE

1: Select d model parameters, to be considered for calibration and identify prior boundaries $[\boldsymbol{x}_{\text{lb}}, \boldsymbol{x}_{\text{ub}}]$ 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 hydrologic 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 X_n
6:     m random parameter vectors forming the set Y_m are generated, such that $\forall \theta \in \boldsymbol{Y_m}: \boldsymbol{D}(\theta | \boldsymbol{X_n^*}) \geq \boldsymbol{L}$ 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

Bárdossy and Singh (2008) applied the principle of data depth in order to generate parameter vectors that are deep with respect to a previously identified set of good parameter vectors. The algorithm is called ROPE. All details are provided in the pseudocode listing in Algorithm 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. In principle the general proceeding of this algorithm, can be divided into three important parts. After handling and pre-processing the input, a set of good parameter vectors is identified (line 4 and 5). Afterwards a set of deep parameter vectors (w.r.t. the good ones) is generated (line 6). These two operations are evolutionary repeated and after each iteration a stopping criterion is checked (line 8). The iterated Monte Carlo sampling is used in order to circumvent the shortcoming of a normal Monte Carlo simulation in order to improve the sampling quality with a limited number of samples and simultaneously applying the principle of data depth. The general approach of the presented ROPE algorithm is well-founded. Bárdossy and Singh (2008) showed that ROPE might be very useful for the estimation of robust hydrologic model parameters.

2.3 A-ROPE

Bárdossy and Singh (2008) demonstrated the performance of the ROPE algorithm calibrating the conceptual hydrologic model HBV for a catchment in south-west Germany on a daily time step. The estimated results support the concept of depth based sampling. Parameter vectors with high data depth corresponded to a better transferability to other time periods. We could reproduce these results calibrating the process-oriented model WaSiM on a daily time step

focussing on modeling of the water-balance (Pompe, 2009). However, in further studies focussing on flood events we experienced problems in particular with the application of the latter two parts of the ROPE algorithm, 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 A-ROPE 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, most existing depth measures are unable to cope with non-elliptic, non-convex or multimodal data sets (Hugg et al., 2006). This can affect the robustness of the estimated deep parameter vectors because the parameter space of most hydrologic models is dominated by distinct regions of attraction and non-convex multidimensional ridges (e.g. Duan et al., 1992; Sorooshian et al., 1993; Grundmann, 2010). Another issue is the efficiency of the depth based sampling. A very 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. 2 where the ratio between the volume of the unit sphere and the unit cube is plotted. In addition, the computational complexity of most depth functions increases tremendously for higher dimensions. An approximation of the halfspace depth used in this paper can for instance be computed in polynomial time: $O(md^3 + mdn)$. In this equation d denotes the number of dimensions, n is the number of points in the reference point set and m is the number of iterations that determine the accuracy of the results.

To overcome these problems we propose to substitute the uniform sampling of deep parameter vectors with an alternative stratified sampling strategy that samples candidate points that are more likely to have a high data depth. In order to do this, a Gaussian mixture model is fit to the underlying set of good parameter vectors. This is done using an expectation maximization (EM) algorithm that assigns posterior probabilities to each component density with respect to each point in the set. The number of components can either be set due to prior knowledge or estimated by the EM algorithm. Afterwards the candidate points can be sampled from a Gaussian mixture model that is an approximation of the set of the reference point set. Consequently good parameter vectors are more likely to be deep with respect to the reference set. In addition this offers advantages to depth-based sampling from non-elliptic or multimodal distributions. Most existing depth measures are unable to cope with non-convex or
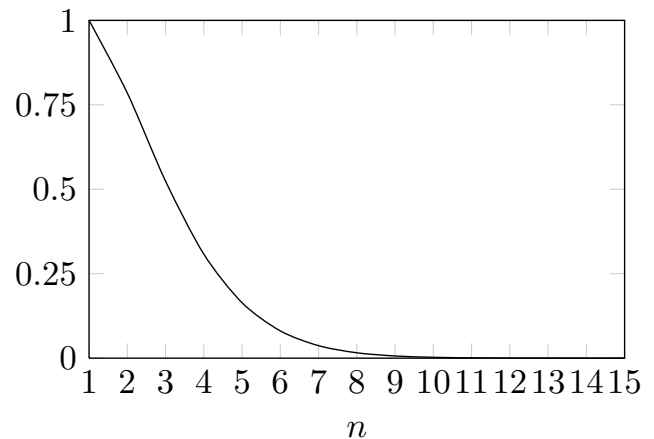

Fig. 2. Volume ratio of the unit sphere to the unit cube in n dimensions as a continuous function of n.

multimodal data sets (Hugg et al., 2006). A clustering of such data sets and a subsequent sampling from single clusters provides the possibility to filter points that are enclosed by the convex hull of the data set, but are actually located outside. The new strategy, entitled GenDeep, is provided in pseudocode in Algorithm 2. Further details of this approach and a number of case studies are provided in Krauße and Cullmann (2011b). Just for a small insight we present the result of a case study discussed in that paper where we performed the depth-based sampling with respect to a previously sampled data set following a banana-shaped and multi-modal distribution (cf. Fig. 3).

Algorithm 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 $i \in \{1, \dots, k\}$.

2: $Y_m \leftarrow \varnothing$
3: **for all** $c_i \in \{c_1, \dots, c_k\}$ **do**
4:     $m_i \leftarrow m\left(\frac{\#\{C_i\}}{n}\right)$
5:     m_i random parameter vectors forming the set Y_{m_i} are generated, such that $\forall \theta \in Y_{m_i} : \{D(\theta | c_i) \geq L$ where $L > 0$
6:     $Y_m \leftarrow Y_m \cup Y_{m_i}$
7: **end for**
8: **return** Y_m

A third issue of the ROPE 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 in the model performance due to uncertainty in the observation data and it
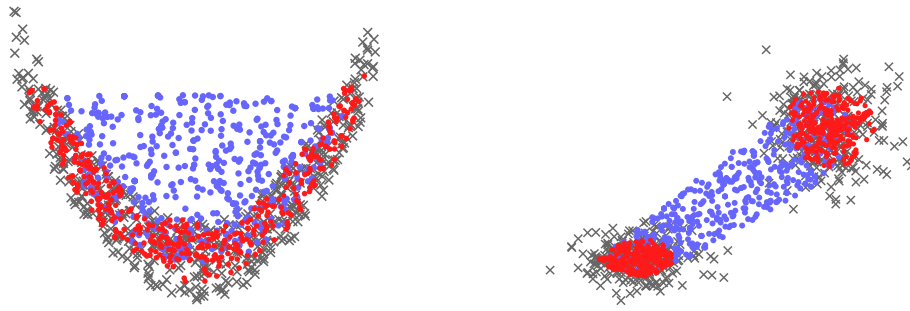
Fig. 3. Two-dimensional data sets (crosses) and corresponding deep points estimated by uniform sampling (blue dots) and stratified sampling with GenDeep (red dots) using the halfspace depth. The datasets were sampled from a non-elliptic banana-shaped distribution (left) and a Gaussian mixture model (right).

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 performance 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 4 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 the problem of overfitting, 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:

$$X_n \leftarrow Y_m \cup X_n^*$$ (4)

Second, we suggest the splitting of the data used for model calibration in a calibration and a control set. 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
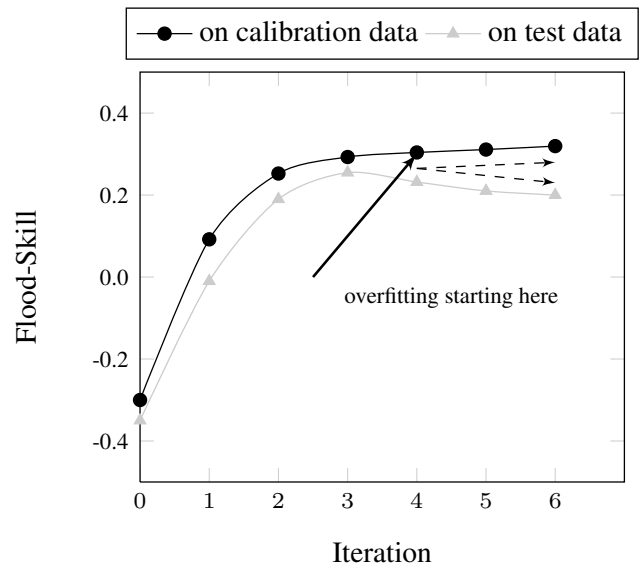

Fig. 4. Overfitting for the calibration of the model WaSiM while calibrated with the method according to Bárdossy and Singh (2008); flood event no. 4 was used for calibration and event no. 14 was used as control set.

(Tetko et al., 1995). The splitting of the calibration data can be done according to an approach based on self-organising maps (SOM) (May et al., 2010) or any other possible strategy. In the case of limited calibration data as for the calibration of hydrological models for flood forecasting we suggest a subjective balanced split, e.g. such that the both calibration and control set contain small and large flood events and all types of characteristics.

All these improvements were integrated into a new approach, we call Advanced Robust Parameter Estimation (A-ROPE). A flowchart of this advanced Monte Carlo-Depth based sampling approach is provided in Fig. 5.
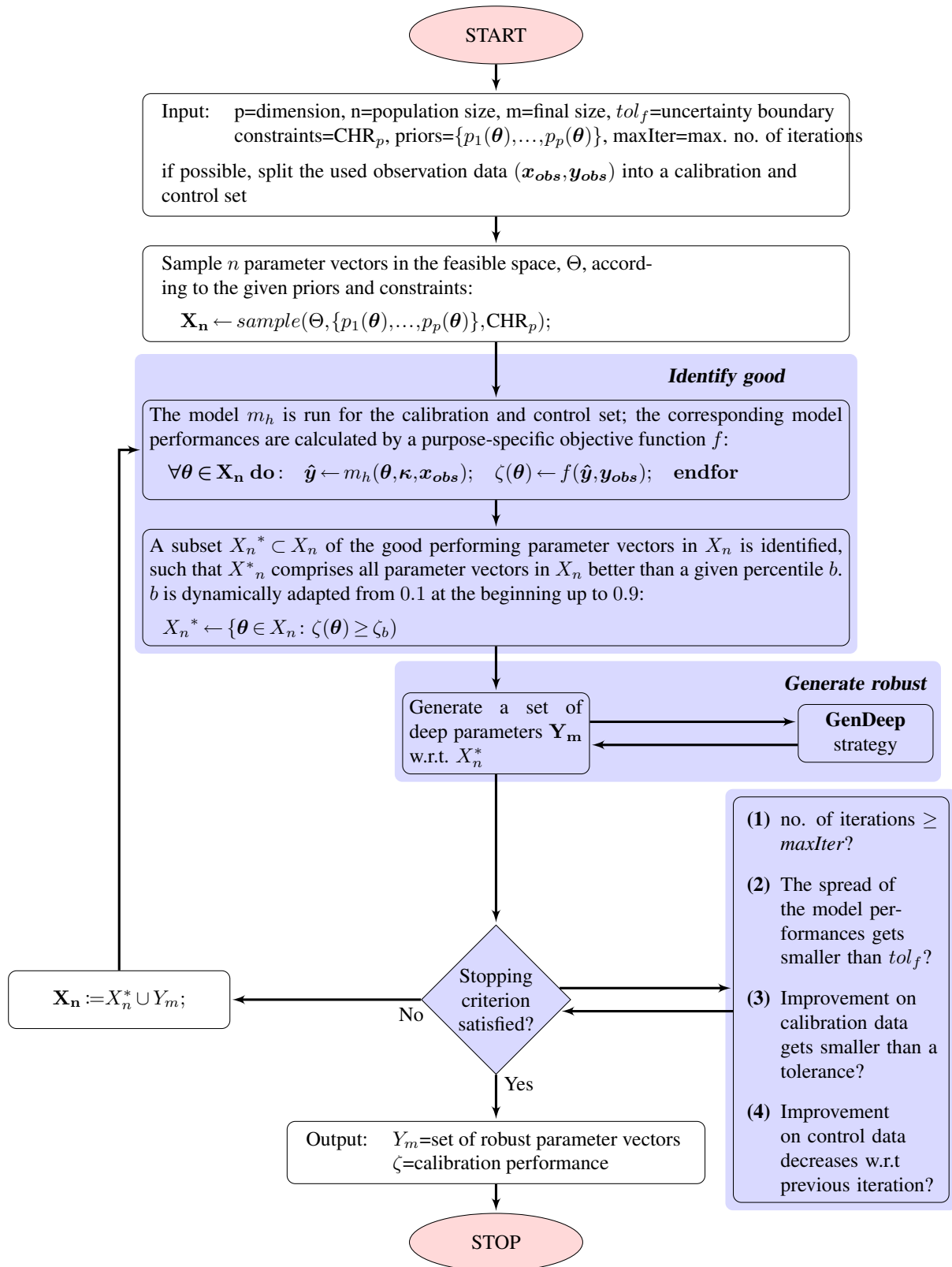
START

Input: p=dimension, n=population size, m=final size, tol_f=uncertainty boundary constraints=CHR$_p$, priors=$\{p_1(\boldsymbol{\theta}),\ldots,p_p(\boldsymbol{\theta})\}$, maxIter=max. no. of iterations

if possible, split the used observation data $(\boldsymbol{x_{obs}},\boldsymbol{y_{obs}})$ into a calibration and control set

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

$\mathbf{X_n} \leftarrow sample(\Theta,\{p_1(\boldsymbol{\theta}),\ldots,p_p(\boldsymbol{\theta})\},\text{CHR}_p);$

Identify good

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

$\forall \boldsymbol{\theta} \in \mathbf{X_n} \text{ do}: \quad \hat{\boldsymbol{y}} \leftarrow m_h(\boldsymbol{\theta},\boldsymbol{\kappa},\boldsymbol{x_{obs}}); \quad \zeta(\boldsymbol{\theta}) \leftarrow f(\hat{\boldsymbol{y}},\boldsymbol{y_{obs}}); \quad \textbf{endfor}$

A subset $X_n^* \subset X_n$ of the good performing parameter vectors in X_n is identified, such that X^*_n comprises all parameter vectors in X_n better than a given percentile b. b is dynamically adapted from 0.1 at the beginning up to 0.9:

$X_n^* \leftarrow \{\boldsymbol{\theta} \in X_n : \zeta(\boldsymbol{\theta}) \geq \zeta_b)$

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) The spread of the model performances gets smaller than tol_f?

(3) Improvement on calibration data gets smaller than a tolerance?

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

Stopping criterion satisfied?

$\mathbf{X_n} := X_n^* \cup Y_m;$

No

Yes

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

STOP

Fig. 5. Flowchart of the Advanced Robust Parameter Estimation by Monte Carlo Simulation (A-ROPE) algorithm.

Algorithm 3 Standard PSO algorithm

1: initialise global best position $\check{\boldsymbol{g}} \in \mathcal{U}[\boldsymbol{x}_{\text{lb}}, \boldsymbol{x}_{\text{ub}}]$
   and global best $g \leftarrow \infty$
2: **for all** particles i **do**
3:     initialise position and local best $\boldsymbol{x}_i, \check{\boldsymbol{x}}_i \in \mathcal{U}[\boldsymbol{x}_{\text{lb}}, \boldsymbol{x}_{\text{ub}}]$
4:     and velocity $\boldsymbol{v}_i = 0$
5: **end for**
6: **while** stop criteria not met **do**
7:     **for all** particles i **do**
8:         update local best position $\check{\boldsymbol{x}}_i$ as best position found so far
           from the particle with index i
9:         update global best position $\check{\boldsymbol{g}}$ and corresponding fitness
           value g as best position found so far from the whole
           swarm
10:    **end for**
11:    **for all** particles i **do**
12:        update velocity using equation
           $\boldsymbol{v}_i \leftarrow \omega \boldsymbol{v}_i + \phi_1 R_2(\check{\boldsymbol{x}}_i - \boldsymbol{x}_i) + \phi_2 R_1(\check{\boldsymbol{g}} - \boldsymbol{x}_i)$
13:        update position using equation
           $\boldsymbol{x}_i \leftarrow \boldsymbol{x}_i + \boldsymbol{v}_i$
14:    **end for**
15: **end while**

3 Merging the strengths of swarm intelligence and depth based parameter sampling

3.1 Particle Swarm Optimisation

An approved search and optimisation strategy for high-dimensional parameter spaces is the Particle Swarm Optimisation which was first presented by Kennedy and Eberhart (1995). It is a population based search algorithm which tries to solve an optimisation problem with arbitrary dimensionality by having a population (swarm) of candidate solutions, here called particles. The performance of each particle is computed and afterwards these particles are moved around in the search-space. The movement is guided by the best found positions of each of the particles and the currently found global best solution. Often those optimisation problems are formulated as the problem of finding the minimum of a function. Algorithm 3 gives a simple version of a PSO algorithm for the minimisation of a function f with upper and lower boundaries $\boldsymbol{x}_{\text{lb}}$ and $\boldsymbol{x}_{\text{ub}}$ respectively. A good introduction into the ideas of swarm intelligence and further reading is given in Kennedy et al. (2001).

3.2 Robust parameter estimation applying Particle Swarm Optimisation: ROPE-PSO

The so far presented depth based parameter estimation algorithms rely on the Monte Carlo method in order to identify a set of good parameter vectors with a given tolerance. Hence the proposed method suffers from the shortcomings of the Monte Carlo method, namely a slow convergence and therefore a large number of samples are needed to estimate a stable solution. This is a major disadvantage for the calibration of

Table 1. Default values of the parameters that control the PSO-GA$_u$ algorithm.

Symbol	Description	Default value
ψ	Breeding ratio	0.5
ω	Particle inertia weight	0.9 decreasing to 0.4
ϕ_1	Cognitive attraction	0.5
ϕ_2	Social attraction	1.25

computationally intensive process-oriented models. Thus in real-world application the maximum number of model runs has to be limited to a computationally feasible maximum. We try to overcome this problem by substituting the Monte Carlo based estimation of a set of good model parameter vectors with a PSO based approach. Although PSO and other evolutionary optimisation approaches were not designed as uncertainty analysis methods they can be adapted in order to be used for a uncertainty quantification with a given tolerance (Mohamed et al., 2010). One possibility to do this is to integrate a Markov Chain element and use them within the scope of a Bayesian framework. Vrugt et al. spent a lot of effort on the development of effective and efficient algorithms following this approach, e.g. SCEM-UA (Vrugt et al., 2003b) and DREAM (Vrugt et al., 2009a). Another approach is to store all so far found parameter vectors within a given uncertainty limit to an archive and direct the search accordingly. One well-known example is the use of PSO algorithms for the approximation of a Pareto optimal set of parameter vectors as the solution of a multi-objective calibration problem, e.g. as provided by Gill et al. (2006).

Algorithm 4 VPAC operator

1: pick random numbers $\phi_1, \phi_2 \sim U(0,1)$
2: update positions using equations
   $\boldsymbol{x}_1 \leftarrow \frac{\boldsymbol{x}_1 + \boldsymbol{x}_2}{2} - \phi_1 \boldsymbol{v}_1$
   $\boldsymbol{x}_2 \leftarrow \frac{\boldsymbol{x}_1 + \boldsymbol{x}_2}{2} - \phi_2 \boldsymbol{v}_2$
3: reset the particles memory $\boldsymbol{p}_1 \leftarrow \boldsymbol{x}_1$ and $\boldsymbol{p}_2 \leftarrow \boldsymbol{x}_2$
4: update the velocities: $\boldsymbol{v}_1 \leftarrow \boldsymbol{v}_1$ and $\boldsymbol{v}_2 \leftarrow \boldsymbol{v}_2$

We will focus on the latter kind of approach. Thus, we developed an algorithm that thoroughly but economically explores the space and stores all solutions within a given uncertainty tolerance to an archive. The basis of our algorithm is a modified version of the PSO presented by Settles and Soule (2005) which is actually a hybrid between a genetic algorithm (GA) and PSO. Thus, we call this algorithm PSO-GA$_u$. It can be controlled by a parameter ψ that is called breeding ratio. This parameter determines the proportion of the population that is not moved according to PSO but is transformed using the GA. Thus, values for the breeding ratio parameter range from $[0 - 1]$. Settles and Soule (2005) propose a default value of 0.5, with the expectation that the best results

Algorithm 5 PSO-GA$_u$

Require: $f =$ objective function, $n =$ population size, $m =$ final size, tol$_f =$ uncertainty boundary, prior boundaries $[\theta_{lb}, \theta_{ub}]$, $t_\alpha =$ minimum depth (corresponds to a desired depth contour)

```
     // INIT
 1:  initialise global best position ǧ ∈ U[θ_lb, θ_ub]
 2:  initialise archive of good solutions with X* ← ∅
 3:  for all particles θ_i ∈ X_n do
 4:      initialise position θ_i, local best θ̌_i and velocity v_i ∈ U[θ_lb, θ_ub]
 5:  end for

     // ITERATE
 6:  while stop criteria not met do

 7:      // UPDATE FITNESS AND BEST
 8:      for all particles θ_i ∈ X_n do
 9:          evaluate model performance f(θ_i)
10:          update the local best θ̌_i as the best position found so far from the particle with index i in case that f(θ_i) better than f(ǧ + tol_f)
11:      end for
12:      update global best position ǧ as best position found so far from the whole swarm
13:      add all current positions with a performance better than f(ǧ) + tol_f to the archive X*
14:      remove all solutions with a performance worse than f(ǧ) + tol_f from the archive X*

         // GA
15:      n_GA ← ψ · n; discard n_GA particles from the population while preserving the 10% best (elitism)
16:      init empty genetic offspring X_GA ← ∅
17:      for i = 1 to (n_GA)/2 do
18:          select a pair {θ_1, θ_2} from the population by tournament selection
19:          apply the VPAC operator to generate new offspring: {θ'_1, θ'_2} ← VPAC({θ_1, θ_2}); consider that this notation includes an update of the velocities and personal best according to the VPAC operator
20:          X_GA ← X_GA ∪ {θ'_1, θ'_2}
21:      end for

         // PSO
22:      assign to each particle a random member ǧ_i ∈ X* as a "personal" global best
23:      for all particles θ_i ∈ X_n do
24:          update velocity using equation v_i ← ωv_i + φ_1 R_2(θ̌_i − θ_i) + φ_2 R_1(ǧ_i − θ_i)
25:          update position using equation θ_i ← θ_i + v_i
26:      end for

         // MERGE
27:      merge the population with genetic offspring
             X_n ← X_n ∪ X_GA
28:  end while

29:  return the set X* as an approximation of the distribution of good parameter vectors within the uncertainty bounds defined by tol_f
```

would be with an even mix of both GA and PSO. However, other values for the breeding ratio may provide better results depending on the characteristics of the considered calibration problem. For further details we refer to Settles and Soule (2005) and referred literature. The evolution of the particles by the GA is done using the following approach: from the pool of possible breeding particles candidates are nominated by tournament selection and recombined. In order to do this they introduced the Velocity Propelled Averaged Crossover (VPAC) operator. The goal is to create two child particles whose position is between the parent's position, but accelerated away from the parent's current direction (negative velocity) in order to increase diversity in the popula-

tion. Algorithm 4 shows how the new child position vectors and velocities are calculated using VPAC. The child particles retain their parent's velocity vector. The previous best vector is set to the new position vector, restarting the child's memory. Towards the end of a typical PSO run, the population tends to be highly concentrated in a small portion of the search space, effectively reducing the search space. With the addition of the VPAC crossover operator, a portion of the population is always pushed away from the group, increasing the diversity of the population and the effective search space. The general movement of the PSO part is a standard implementation that can be controlled by the following parameters: the particle inertia weight ω determines the velocity of

the proper motion of the particles, the cognitive attraction ϕ_1 that controls the degree of movement towards the local optimum and the social attraction ϕ_2 doing the same with respect to the global optimum of the swarm. We set the algorithm's parameters according to literature recommendations (cf. Table 1). For further details and studies regarding the setting of the algorithm's parameters refer to Perez and Behdinan (2007), Settles (2005) that also provide references to additional literature and materials. The used stopping criterion is either a fixed number of iteration steps that has to be set according to the given problem or a maximum number of members in the set of good parameter vectors. Another option might be a check that assesses the stability of the estimated set. We suggest to carry out some test runs with different limits and check the stability of the estimated parameter vectors. For further details regarding this issue we refer to Gill et al. (2006) and Cabrera and Coello (2010).

Algorithm 6 ROPE-PSO

1: Execute the PSO based PSO-GA$_u$ procedure to estimate a set of good model parameter vectors X^* with a model performance within a given tolerance tol$_f$

2: Apply the GenDeep algorithm to sample a set of parameter vectors Y with high data depth w.r.t. X^*

3: **return** Y

An important difference is done considering the local and global optimum of the swarm. Contrary to normal PSO algorithms the PSO-GA$_u$ algorithm does not just account for one global optimum, but for a set of good parameter vectors. Good parameter vectors are all points evaluated so far that correspond to a model performance better than the global optimum found so far plus a given uncertainty tolerance tol$_f$ which must be set with regards to the specific problem. For environmental model calibration tol$_f$ should be set according to the accuracy of the used observations and other sources of uncertainty to be considered. To ensure a sufficient but nonetheless economical cover of the feasible space not just towards the global optimum the PSO-GA$_u$ algorithm follows an idea that is used in multi-objective PSO algorithms: all so far found good parameter vectors are stored in an archive X^*. An important difference between a normal PSO algorithm considers the movement at the end of each iteration. Instead of moving the whole swarm towards the so far found global best position, the algorithm assigns to each particle one random member of the archive $\check{g}_i \in X^*$. We call this position "personal" global best (cf. Algorithm 5, line 14). Another issue affects the update of the local best position $\check{x}_i$. Unlike in a normal PSO, it will only be changed in case that the old local optimum corresponds to a model performance worse than the global optimum plus the uncertainty bound. This prevents a too small shrinkage of the swarm and ensures that the algorithm not just searches into the direction of the so far found global optimum but samples from the whole region within the given tolerance. Consider that this algorithm

can also prevent overfitting just by adding another objective function that assesses the performance on a control set. In this case the position of a particle is considered to be better in case that it shows an improvement on both the calibration and the control data. This basic idea can be used to advance this method into a full multi-objective calibration procedure. A pseudocode listing of the complete proposed PSO-GA$_u$ algorithm is provided in Algorithm 5.

The set of good parameter vectors estimated by PSO-GA$_u$ should not only comprise the global optimum but cover the complete region within the given uncertainty bounds. That is an issue in so far that PSO, i.e. the underlying algorithm of ROPE-PSO, was unlike Monte Carlo not designed to be used as an uncertainty analysis method. Therefore we studied the sets of good parameter vectors estimated by PSO-GA$_u$ and checked whether they follow the same distribution as those estimated by iterative Monte Carlo used in the ROPE and A-ROPE algorithm. We demonstrate this issue using the Rosenbrock function as defined in Eq. (5) as an example. It is a smooth single extremum test function that represents the existence of large flat regions on the error surface. This is quite typical for hydrologic models (Duan et al., 1992). We set a (subjective) tolerance value of 0.2 and applied the iterative Monte Carlo method from in both ROPE and A-ROPE, and the PSO-GA$_u$ algorithm in order to estimate a set of good parameter vectors within the given tolerance.

$$f(\boldsymbol{x}) = \sum_{i=1}^{d-1} \left[(1 - x_i^2)^2 + 100(x_{i+1} - x_i^2)^2 \right]$$

with $\min f(\boldsymbol{x}^*) = 0$ and $x_i^* = 1$ \hfill (5)

The estimated scatter plots are provided in Fig. 6. The dark grey regions represent the true distribution of parameter vectors with a function value less than the given tolerance. It was estimated by a Monte Carlo simulation using Latin hypercube sampling with a number of 1000.000 samples in the target region. Considering the fact that the global optimum of the Rosenbrock function is zero at the position $\theta_{\text{opt}} = \{1, 1\}$ this region matches exactly the targeted distribution. It is evident that the parameters provided by PSO-GA$_u$ cover the complete region of good parameter vectors, suggesting that the modified PSO approach provides a correct estimation of the target distribution and hence does not collapse to a small region comprising the estimated global minimum.

The developed PSO-GA$_u$ approach can be easily used to substitute the Monte Carlo based approach in a robust parameter estimation algorithm. The new approach called Robust Parameter Estimation using PSO (ROPE-PSO) applies PSO-GA$_u$ in order to obtain a set of good parameter vectors X^* with a given uncertainty. Afterwards a set of deep parameter vectors with respect to X^* is sampled using the previously introduced by the GenDeep sampling strategy. A pseudocode listing of the developed ROPE-PSO approach is given in Algorithm 6. The proposed approach methodology can be extended to be used for an uncertainty analysis
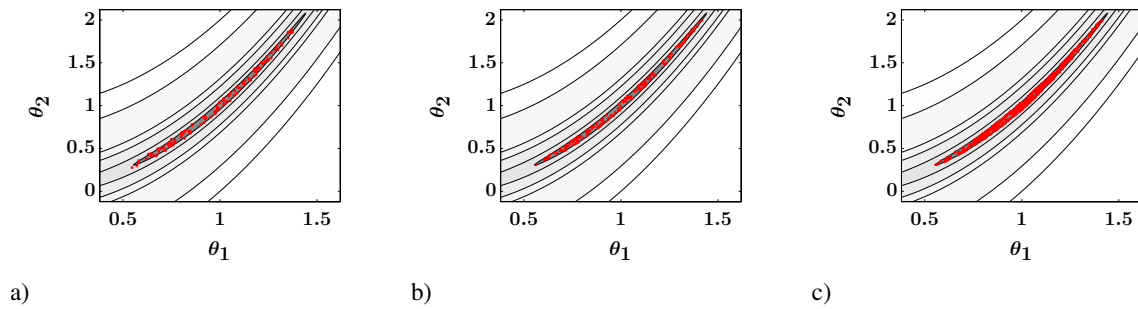
Fig. 6. A scatterplot of the (θ_1, θ_2) samples of good parameter vectors generated using the iterative Monte Carlo approach in the original ROPE **(a)**, the modified version A-ROPE **(b)** and the PSO based PSO-GA$_u$ algorithm **(c)**. The dark grey region represent the true distribution of the good parameter vectors considering a tolerance of 0.2.

within a proper statistical context, by relating the likelihood of the parameter vectors to their depth. However further research is required to complete this task (e.g. Bárdossy and Singh, 2008). The algorithm was implemented in a robust parameter estimation framework which comprises other published ROPE approaches. The implementation was done in the MATLAB programming language. It is open source and available from the author.

4 Accounting for uncertain soil information on hydrologic parameter estimation

The soil hydraulic parameters determine the water retention and conductivity curves and thus govern the process of water movement in the unsaturated zone. For this reason they also influence the generation of direct runoff and interflow in a hydrologic model. In many studies the soil hydraulic parameters are considered as physically based parameters and are used as fixed values. Often those values are simply estimated by applying a pedotransfer function to physical soil properties, e.g. the distribution of the grain-size fractions, humus content and bulk density. Typically the soil information is given in a classified form which provides a possible range of the physical soil properties referring to the used classification system. This information is often visualised in a soil texture triangle. However, in most cases the pedotransfer function is just applied to the mean value for the considered soil type. The uncertainty due to classified soil information is typically not considered. However, neglecting this uncertainty can influence the accuracy and uncertainty of the estimation of other conceptual model parameters.

Grundmann (2010) studied this problem and presented a new method to account for those uncertainties. The proposed approach has a high degree of general applicability because it is independent of the used soil classification system and the used pedotransfer function. Their approach can be summarized by the following algorithmic steps:

1. Identify the lower and upper boundaries of the grain-size fractions for each pre-dominant soil type in the catchment, according to the given soil information and classification system.

2. For each considered soil type, draw a set of possible samples of the grain-size fractions by uniformed sampling (uniform distribution) over the identified range.

3. Apply a suitable pedotransfer function to each sample in order to estimate a set of soil hydraulic parameters describing their prior distribution.

4. The estimated parameters can be scaled to a scaling parameter β using a similar media concept in order to reduce their dimensionality. A suitable approach is presented in Warrick et al. (1977). The scaling parameter β can now be considered for calibration.

The estimated distribution of the scaling parameters is a well-founded a priori estimate of the uncertainty in the soil hydraulic parameters and can be used to study the influence of the uncertain soil information on the simulation results of hydrologic models. Furthermore this information can be used as a well-founded prior distribution for a subsequent model calibration considering the soil hydraulic parameters as calibration parameters. This approach contradicts the physically based philosophy of the Richards equation model. On the other hand, this opens the possibility to adapt this model to the processes in the catchment. We are convinced that this is a reasonable approach because the soil hydraulic parameters can hardly be determined with the necessary precision on a catchment scale just be using soil information maps. During calibration the scaling parameter β is adjusted to suitable values. The model however cannot directly be driven with β but with a vector of soil hydraulic parameters. They can be derived from β using a mapping or rescaling procedure as follows. First, the mapping assigns a soil hydraulic parameter vector from the previously generated set to every
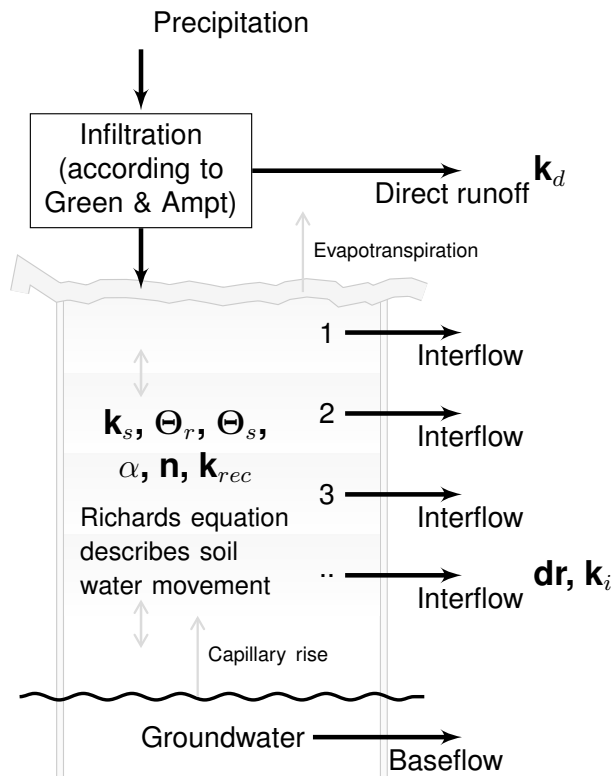
Fig. 7. Scheme of the WaSiM soil module with location of impact of soil hydraulic and conceptual model parameters (bold).
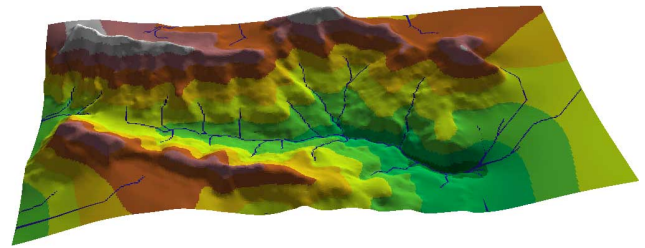

Fig. 8. A 3-D-view of the catchment with the potential rivers flow-paths. This illustration is taken from the official Rietholzbach website (http://www.iac.ethz.ch/en/research/riet/overview.html).

5 Case studies

5.1 Calibrating of the hydrologic model WaSiM focussing on flood events

In two real world case studies the developed approach is tested on the calibration of the hydrologic model WaSiM-ETH/6.4 (in the further referred to as WaSiM). WaSiM 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, 2003b,a); Krauße and Cullmann (2011b). Furthermore WaSiM-ETH has been used for extrapolation of extreme flood events by Cullmann (2006). For this study we used the version with the approach according to Richards for the simulation of the unsaturated zone. An overview of the model structure is given in Fig. 7. For further details of the model, we refer to Schulla and Jasper (2007) and the official website of the model http://www.wasim.ch.

The model was calibrated focussing on flood events in the small prealpine Rietholzbach catchment (3.18 km^2). As a sub-basin of the Thur catchment it is located in the north-east of Switzerland. A 3-D-view of the catchment area is given in Fig. 8. This basin has been observed as a research catchment by the ETH Zurich since 1975. Continuous hourly measurements have taken place since 1981. For the case studies presented in this paper we used a time series consisting 27 yr of meteorological (temperature, precipitation, global radiation, and wind speed) and discharge measurements. 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 yr 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 mm h^{-1}. All events are in the time from May until October to avoid the problem of modeling snow accumulation

estimated scaling parameter and stores this information into a lookup table. For a new β value that is not stored in the database the rescaling selects the vectors of soil hydraulic parameters from the previously generated database with the closest corresponding scaling value. Grundmann (2010) used the estimated prior distribution and estimated the posteriori distribution of the soil hydraulic parameters in the context of a Bayesian framework. The parameter estimation framework used in this paper is no closed Bayesian uncertainty framework. The priori estimates serve however as a well-founded starting population of the evolutionary approach estimating the good model parameter vectors. The mapping used in the rescaling ensures that the soil hydraulic parameters cannot leave the variation intervals given by the prior distributions. Consider once again that we want to adapt the soil hydraulic parameters to the catchment characteristics. Therefore we do not use a probability-possibility transformation and propagate the prior uncertainty described by the possibility distribution through the model. For further information on this approach we refer to Grundmann (2010) and Warrick et al. (1977).

Table 2. 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 [$\mathrm{mm\,h^{-1}}$]	Volume [mm]	Type	Volume [mm]	$\Theta_{0--0.5\,\mathrm{m}}$ [vol %]
1	06/07/1994 24:00		**7.61**	33.2	convective	75.0	65
2	23/06/1986 22:00		**7.52**	20.2	convective	39.7	75
3	06/07/1994 21:00		**6.23**	33.2	convective	75.0	65
4	08/08/2007 20:00		**5.69**	51.2	convective	64.1	90
5	07/06/2007 23:00		**5.48**	16.0	convective	36.1	79
6	06/08/1982 19:00		**5.35**	21.2	convective	28.3	87
7	25/07/1989 20:00		**4.64**	11.7	convective	13.7	80
8	30/05/1995 22:00		**4.01**	16.9	convective	39.4	84
9	15/08/1982 20:00		**4.01**	18.3	convective	43.6	80
10	06/08/2000 13:00		**3.61**	28.9	stratiform	48.8	83
11	17/09/2006 09:00		**3.13**	36.1	convective	101.8	71
12	24/09/2002 10:00		**3.04**	120.1	stratiform	125.2	87
13	09/09/2001 01:00		**2.96**	18.4	convective	49.4	78
14	09/06/1994 07:00		**2.52**	27.9	stratiform	40.7	86
15	11/05/1991 24:00		**2.50**	44.6	stratiform	75.3	80
16	01/07/1987 16:00		**2.37**	12.1	convective	30.7	78
17	02/08/2005 22:00		**2.03**	26.0	convective	54.2	74
18	25/05/1990 11:00		**2.02**	14.3	stratiform	54.9	75
19	11/06/1995 07:00		**1.98**	28.8	stratiform	36.8	88
20	31/05/2000 09:00		**1.76**	19.4	stratiform	57.7	78
21	09/09/2005 21:00		**1.62**	6.9	convective	37.1	67
22	28/09/1995 01:00		**1.47**	17.3	stratiform	38.2	81
23	13/05/2002 01:00		**1.02**	4.6	convective	21.4	81
24	14/09/1993 01:00		**1.02**	10.9	stratiform	26.8	84

Table 3. Overview of the three used conceptual model parameters considered for calibration; the reference parameter vector $\theta_{\mathbf{wb}}$ was estimated in order to use WaSiM for water-balance simulations in the Rietholzbach catchment; additionally we provided the conceptual soil parameter k_{rec} that plays a role in another case study

parameter	unit	reference ($\theta_{\mathbf{wb}}$)	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

and melting processes. An overview of all selected 24 flood events with their specific characteristics and pre-conditions is given in Table 2. A significant number of studies have been conducted in this basin. For further information we refer to Gurtz et al. (1999); Zappa (2002) and the website http://www.iac.ethz.ch/research/rietholzbach.

Table 3 gives the model parameters considered for calibration. Those 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. The reference parameter vector $\theta_{\mathbf{wb}}$ estimated for water-balance simulations was used to estimate reasonable pre-event model states.

Within our case studies 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 $q_{\mathrm{obs}}(t)$ of 5%. Thus, the real but unknown discharge $q(t)$ can be written as:

$$q(t) = q_{\mathrm{obs}}(t)(1 + \epsilon(t)) \tag{6}$$

with $\epsilon(t)$ 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. This ensemble can be used in order to assess the uncertainty due to observation errors. It also opens the possibility to assess the uncertainty of the results of classical optimisation algorithms as follows. For such an algorithm we calibrate the model with respect to every single ensemble member of the set of possible observations. All estimated best parameter vectors are merged into a set that expresses the uncertainty due to the observation errors. For further details refer to Bárdossy and Singh (2008). To ensure comparability the same should be done for the different ROPE algorithms either. However, in order to reduce the computation time, we previously checked the stability of

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

Criterion	Formula
NS	$1 - \dfrac{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i(\theta))^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
rPD	$\dfrac{\lvert x_{\max} - y_{\max}(\theta)\rvert}{x_{\max}}$
FloodSkill	$NS - rPD$

the estimates of the considered ROPE algorithms for a subset of the discharge ensemble members and subsequently just used the ensemble mean for calibration.

A successful calibration requires the definition of a performance criterion that quantifies the matching with the intended objective. Focussing on flood events a suitable performance criterion should quantify both the model's ability to provide a good estimate of the peak flow values and to provide a behaviour that is at least in fundamentals similar to the catchment behaviour. A good measure to asses the the first point is the relative peak flow deviation (rPD) which is defined as the absolute value of the relative deviation of the simulated with respect to the observed peak flow value. The general quality of the fit of the model and the catchment behaviour can be assessed by a global performance criterion, e.g. the efficiency criterion according to Nash and Sutcliffe (1970) (NS). It has been widely used to quantify the global performance of hydrologic models. We combined both the Nash efficiency and the relative peak flow deviation into an aggregated performance criterion, we call Flood-Skill. The higher the FloodSkill the better is the model's ability to represent the catchment's behaviour focussing on flood events. A perfect fit corresponds to a FloodSkill of one. A least acceptable model with Nash value of 0.5 and a relative peak flow deviation of 0.5 correspond to a Flood-Skill of zero[1]. A formal definition of the FloodSkill and the previously introduced criteria is given in Table 4.

5.1.1 Case study I: Comparison of ROPE, A-ROPE and ROPE-PSO considering just the conceptual model parameters

In a first case study we studied the developed ROPE approaches on the basis of a first real world case study. The original ROPE approach and the two classical optimisation algorithms, the interior-point method (IPM) according to

---

[1]Consider that such a model performance is rather realistic for flood events in a very small catchment with less than 5 km^2 using just hourly data.

Table 5. 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	$\{1, 2, 3, 4, 5, 7, 8, 9, 10, 12, ...$ $13, 15, 16, 17, 19, 21, 22, 23\}$

Waltz et al. (2006) which is a gradient based method and a genetic algorithm (GA) according to Conn et al. (1997) served as additional benchmarks. Nonetheless the focus always remains on A-ROPE and ROPE-PSO. The conceptual model parameters of the model WaSiM were calibrated focussing on flood events. Out of the mentioned set of 24 flood events, three events were used for calibration, three for overfitting control and the remaining events were used for validation (cf. Table 5). Both the calibration and control set contain three events covering the possible range of peak flow values in our database of 24 flood events. The ROPE and A-ROPE were set with a population size of 2500 [2] and a maximum of five iterations. For PSO-GA$_u$ we set the population size to 50. These settings were set after a small number of test runs with different population sizes. The used objective was the proposed FloodSkill criterion.

Before deep parameters were generated we validated all good parameter vectors and studied the relationship between their corresponding data depth and model performance on the validation data. An overview of the results is provided in Fig. 9. Especially for the results estimated by ROPE-PSO the correlation between the calibration objective, i.e. the FloodSkill criterion, and the data depth is clearly positive. The correlation for the results estimated by the Monte Carlo based ROPE algorithms is less intense. This a first hint that the set estimated by ROPE-PSO is a stable solution and the application of depth based sampling makes sense in order to improve the calibration results. The scatterplot given in Fig. 10 where each parameter vector of the estimated set is shaded according to its validation performance shows that the parameter vectors with worse model performance are particularly located at the boundary of the estimated set. However, this conclusion does not hold true for the single performance criteria the FloodSkill consists of[3]. As shown in Fig. 11, the optimal regions for this criteria are even located on opposite sides of the estimated set. This is an indication of a clear tradeoff between the two objective criteria rPD and problems of the model to represent the global system behaviour of the catchment and a good representation of the peak flow values. This may also be due to the relatively coarse time step considering the small catchment

size. In this case a multi-objective calibration might be useful, but this is in the scope of future research. In a second step deep parameter vectors were sampled with respect to the estimated set of good parameter vectors. The statistics of the estimated samples in comparison with those estimated by ROPE and A-ROPE are given in Table 6. Although both sets overlap each other, it is evident that the mean value of the storage coefficients k_d and k_i of the ROPE-PSO estimates are considerably higher than those estimated by A-ROPE. However, the correlation between the two most sensitive model parameters focussing on flood events, k_d and dr is approximately the same for both algorithms. The standard deviation for the ROPE-PSO estimates is smaller than the estimates of the Monte Carlo based estimates or all considered model parameters except for the less sensitive parameter k_i. We assume that the iterative Monte Carlo based sampling cannot identify the region with the best parameters as exact as the PSO based ROPE-PSO due to its limited sample size (2500 samples per iteration) and the subjective fixed boundary[4] that is used for the determination of the best parameter in the individual iteration steps. Due to a coarse resolution in the first iterations, this might exclude regions at the outer boundary of the good parameters that might contain better parameters. An adaptive selection of the boundary might solve this problem for the Monte Carlo based search. The validation results of the estimated parameter vectors are shown in Table 7. The ROPE-PSO estimates provide a slightly better validation performance than those estimated by A-ROPE. ROPE-PSO also outperforms all other compared algorithms compared in the case study of Krauße and Cullmann (2011a). In comparison with the classic optimisation algorithms IPM and GA the model performances achieved by the depth based estimates in validation are not just better in mean but show a clearly smaller standard deviation. Among themselves the variances of ROPE, A-ROPE and ROPE-PSO are approximately in the same range.

These results indicate that the estimated solution is robust and transferable. The proposed algorithm also converges faster with less parameter vectors to be evaluated. We studied the stability of the solution estimated by ROPE-PSO for a maximum number of model evaluations of 3000, 1000 and 500. According to this test we set the limit to 1000 whereas the Monte Carlo based A-ROPE made full use of the 10 000 parameter evaluations. This advantage in terms of computational efficiency gets even more weight considering that one parameter evaluation takes approximately three minutes on a standard CPU[5].

---

[2]This corresponds to the parameters n and m in Algorithm 1

[3]These criteria are NS and rPD.

[4] In the current setup the algorithm strictly selects the 10 % best parameter vectors in each iteration.

[5]That means that 3 model runs have to be carried out, one for each calibration event.

Table 6. Mean value (Mean), standard deviation (Std), coefficient of variation (CV), minimum, maximum and correlation coefficients between the generated samples for the conceptual model parameters estimated by ROPE **(a)**, A-ROPE **(b)** and ROPE-PSO **(c)**.

a)

Parameter	Mean	Std	CV	Min	Max	k_d	k_i	dr
k_d	2.34	0.62	0.24	1.56	3.37	1.00	0.11	−0.45
k_i	4.86	2.22	0.46	2.12	9.95	...	1.00	−0.10
dr	5.61	1.44	0.31	2.14	10.02	...	...	1.00

b)

Parameter	Mean	Std	CV	Min	Max	k_d	k_i	dr
k_d	2.30	0.41	0.18	1.47	3.54	1.00	0.05	−0.39
k_i	4.60	1.24	0.27	2.22	8.06	...	1.00	−0.30
dr	5.13	1.52	0.30	2.04	9.91	...	...	1.00

c)

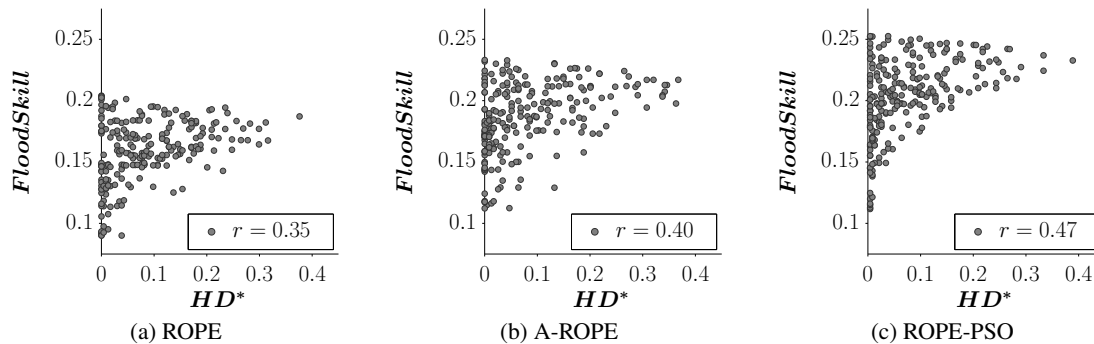| Parameter	Mean	Std	CV	Min	Max	k_d	k_i	dr
k_d	2.78	0.30	0.11	2.04	3.67	1.00	−0.20	−0.38
k_i	6.54	1.36	0.21	3.12	12.65	...	1.00	−0.64
dr	5.23	0.96	0.18	3.08	8.20	...	...	1.00

Fig. 9. Correlation between data depth and overall validation performance for the solutions estimated by ROPE, A-ROPE and ROPE-PSO.

5.1.2 Case study II: Calibrating WaSiM for flood forecasting considering the uncertainty of the soil hydraulic parameters

In a second case study we calibrated WaSiM again, however this time we additionally considered the uncertainty in the soil hydraulic parameters in the model calibration. As already introduced in Sect. 4 the uncertainty due to coarse soil information and the resulting uncertainty in the soil hydraulic parameters can have a tremendous influence on the model uncertainty in the case of flood events. In a preliminary study with WaSiM in the Rietholzbach catchment (e.g. Seifert, 2010) we could prove this conclusion. From the five predominant soil types in the basin (Table 8), we found the soil hydraulic parameters of SL and SiL and the soil parameter k_{rec} to be sensitive referring to the simulated discharge for

flood events. k_{rec} defines the gradient of the saturated conductivity k_s with increasing soil depth and has a valid range between 0.01 and 1. Its default value estimated for water balance calibration runs was set to 0.1. This value of k_{rec} is used to estimate reasonable pre-event model states. We considered both k_{rec} and the introduced soil hydraulic parameters together with the conceptual model parameters for model calibration.

In a first step prior to the actual model calibration we estimated the prior uncertainty in the soil hydraulic parameters of the soils SL and SiL according to the approach developed in Grundmann (2010). As a results we obtained a set of 10 000 soil hydraulic parameter vectors for each soil. Following the ideas of this approach we mapped the computed set of soil hydraulic parameter vectors to two scaling parameters β_{LS} and β_{SiL}, one for each soil. The scaling was
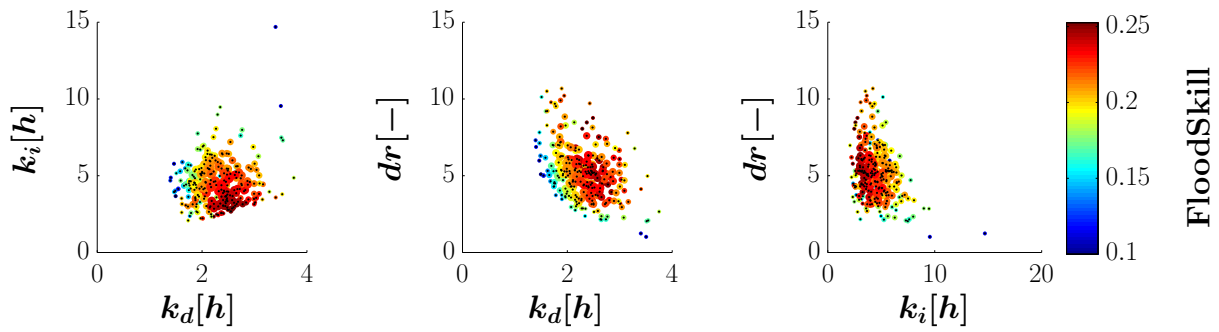
Fig. 10. Scatter plot of the good parameter vectors shaded according to their validation performance. Red points have a good validation performance, blue points are worse (cf. colorbar). The size of the shades is proportional to the data depth of each point with respect to the whole cloud.

Fig. 11. Scatter plot of the good parameter vectors shaded according to their validation performance for the criteria NS and rPD. Red points have a good validation performance, blue points are worse (cf. colorbar). The size of the shades is proportional to the data depth of each point with respect to the whole cloud.
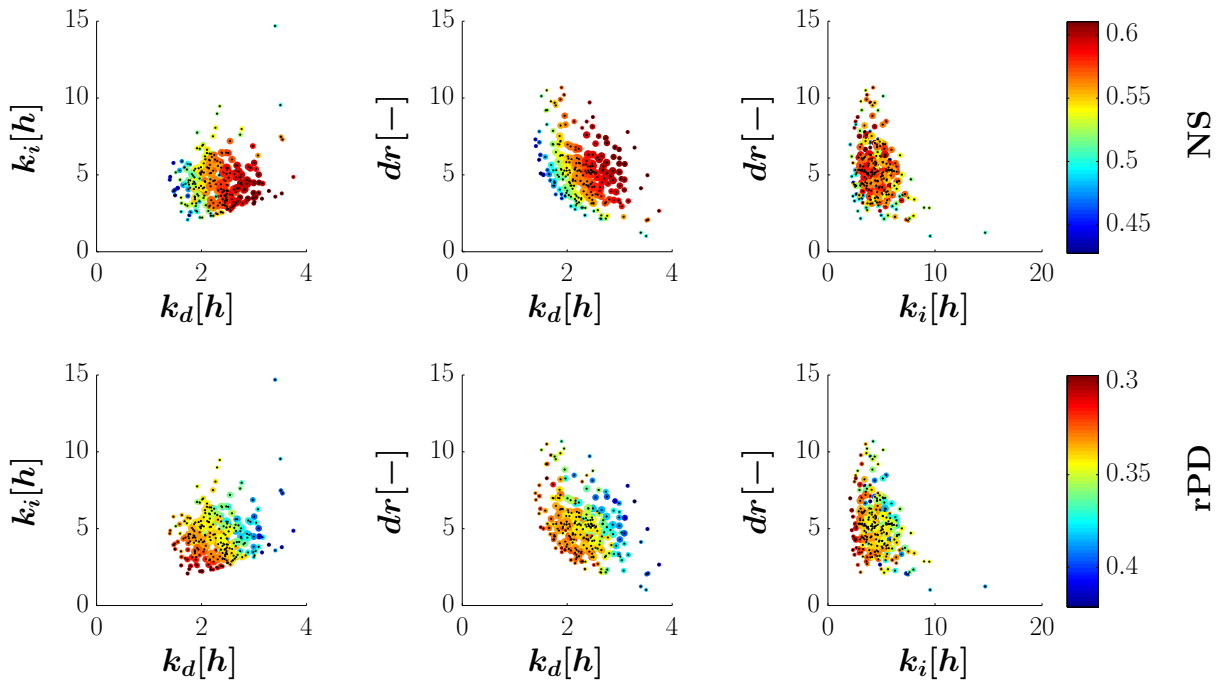
Table 7. Mean overall validation performance for the parameter vectors estimated by all compared algorithms; the model performance was calculated on the validation set as given in Table 5.

	FloodSkill				NS				rPD			
	Mean	Std	Worst	Best	Mean	Std	Worst	Best	Mean	Std	Worst	Best
IPM	0.14	0.088	−0.16	0.23	0.52	0.049	0.33	0.58	0.37	0.048	0.52	0.32
GA	0.15	0.048	0.08	0.24	0.53	0.043	0.43	0.61	0.38	0.040	0.46	0.32
ROPE	0.18	0.029	0.11	0.25	0.56	0.039	0.46	0.61	0.39	0.031	0.46	0.30
A-ROPE	0.15	0.030	0.09	0.23	0.55	0.036	0.47	0.61	0.38	0.030	0.44	0.30
ROPE-PSO	0.20	0.031	0.11	0.25	0.55	0.036	0.44	0.61	0.35	0.021	0.41	0.30

Table 8. Expectation values of the physical properties of the prevailing soil types in the Rietholzbach catchment, classified according to USDA, and corresponding soil hydraulic parameters; the parameterisation of the soil hydraulic parameters is done for each soil according to the approach provided in Grundmann (2010) by the help of the pedotransfer functions provided in Wösten et al. (1999) and Brakensiek et al. (1984); the expectation values are the mean over 10 000 realisations

|	L loam	SL sandy loam	SiCL silty clay loam	SiL silt loam	LS loamy sand
catchment area [%]	15	20	3	51	11
clay [%]	20	10	33.5	13.5	7.5
silt [%]	39	25	56.5	69	15
sand [%]	41	65	10	17.5	77.5
humus content [%]			2.5		
k_s [m s^{-1}]	$1.81 \cdot 10^{-6}$	$1.45 \cdot 10^{-5}$	$8.61 \cdot 10^{-8}$	$2.85 \cdot 10^{-7}$	$4.26 \cdot 10^{-5}$
α [1/m]	3.49	4.48	2.00	1.35	5.52
θ_r			0.01		
θ_s	0.42	0.41	0.43	0.42	0.41
n	1.18	1.27	1.13	1.24	1.32

done using the algorithm provided in Warrick et al. (1977). The prior uncertainty of the soil hydraulic parameters with best fits for Gaussian ($\mathcal{N}$), logarithmic Gaussian (log$\mathcal{N}$), Gamma (Γ) and bimodal Gaussian (GM) distribution is given in Fig. 12. The fitted distributions are just an additional information in order to show that the estimated prior distributions could also be described by commonly used distribution functions. Consider that the residual water content θ_r was constantly 0.01 for both SL and SiL with a deviation of less than $10e - 14$ and thus is not shown in the plots. The distribution of the saturated conductivities has the maximum density in the range of the lowest possible values and is characterised by a high skewness. The other soil hydraulic parameters according to the Mualem – Van Genuchten model have distributions which can be approximated by a normal distribution. The distribution of the corresponding scaling parameters is given in Fig. 13. It is evident that the distribution of the scaling parameters is strongly influenced by the distribution of the saturated conductivities. This is due to the relatively high spread of this parameter.

The conceptual and the soil parameters form a six dimensional calibration problem with the model parameters $\{k_d, k_i, dr, \beta_{SL}, \beta_{SiL}, k_{rec}\}$ to be estimated. Considering that the error surface of WaSiM is very bumpy this is already a challenging calibration problem, especially for Monte Carlo based approach. Again we estimated the parameters of WaSiM with ROPE, A-ROPE and ROPE-PSO. The Monte Carlo based algorithms were limited to a maximum of 10 000 model parameter vector evaluations whereas the ROPE-PSO was tested with both a maximum of 2000, 3000, 5000 and 10 000 model runs. There was not much difference between the estimates of the different runs. This is why we set the maximum to 3000 in order to save computing time. This is another confirmation of the computational efficiency of the ROPE-PSO approach.

The distribution of the parameter vectors estimated by A-ROPE and ROPE-PSO is given in Fig. 14. Furthermore the distribution of the soil hydraulic parameters corresponding to the scaling parameters β_{SL} and β_{SiL} estimated by both algorithms are given in Fig. 15. The soil hydraulic parameters corresponding to the identified distribution of the scaling parameter parameters were generated using the rescaling procedure provided in Sect. 4. For further details we refer to Warrick et al. (1977) and Grundmann (2010). These results

	ks	θ_s	α	n
Mean	2.227e-05	0.4120	4.62	1.29
Std	2.185e-05	0.0083	1.85	0.05
CV	0.98	0.02	0.40	0.04

(a)
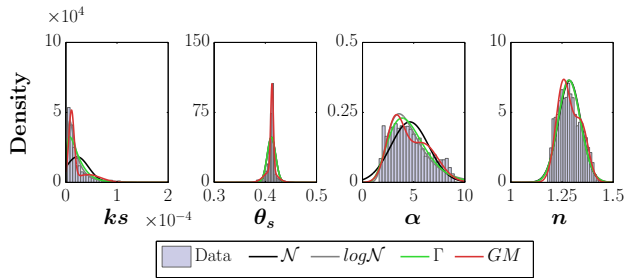
	ks	θ_s	α	n
Mean	7.207e-07	0.4165	1.35	1.26
Std	5.726e-07	0.0048	0.33	0.08
CV	0.79	0.01	0.24	0.06
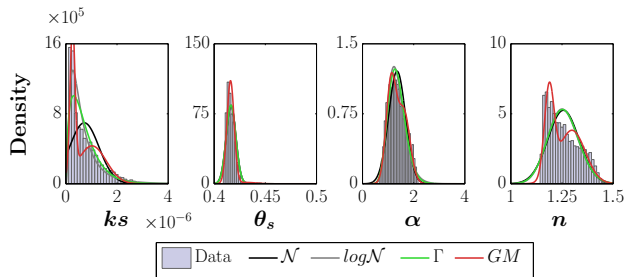
(b)

Fig. 12. Prior distribution of the soil hydraulic parameters for the soils SL **(a)** and SiL **(b)**.

Fig. 13. Prior distribution of the scaling paramters β_{SL} and β_{SiL}

show that the parameter vectors estimated by ROPE-PSO are distributed over a considerably smaller region than those estimated by A-ROPE. The estimated parameter distributions even indicate that the ROPE-PSO estimates form a sub-region of the large region described by the parameter vectors estimated by A-ROPE. This suggests that the PSO based search strategy in ROPE-PSO can more precisely identify the region in the parameter space that corresponds to the highest model performance. Consequently the upper boundary of the least model performance defined by the assumed measurement uncertainty gets higher and the region of good model parameter vectors can be more precisely identified. Another issue concerns the distribution of the soil hydraulic parameters only. By means of comparison of observed and sim-
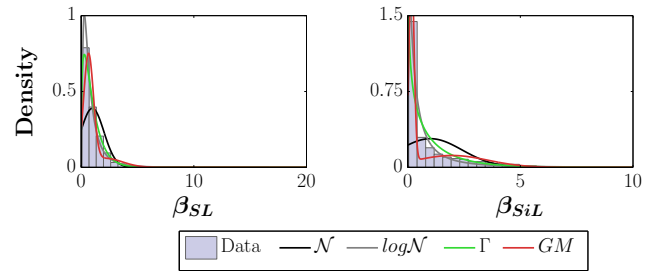
ulated discharge, the parameter estimation algorithms try to reject soil hydraulic parameter vectors that are not suitable to represent the catchment's behaviour and identify a distribution with the most suitable model parameters. It is obvious that the spread of the distributions of the soil hydraulic parameters compared to their prior uncertainty gets smaller for both algorithms. Furthermore it stands out that the mean ks of the ROPE-PSO estimates is smaller than the prior expectancy whereas the mean ks of the A-ROPE values is higher than the prior value. In terms of the model that means that ROPE-PSO identifies parameter vectors that try to simulate just the slow matrix flow in the unsaturated zone whereas faster runoff processes in the unsaturated zone, e.g. preferential flow, are approximated by a fit of the conceptual model parameters controlling direct runoff. Probably that is also the reason why the less sensitive conceptual parameter k_i can be much better identified than in the previous case study. In contrast A-ROPE identifies parameter sets which try to represent the faster components by a higher saturated conductivity in the soil model. Preferential flow in macropores can become a dominant process within the Rietholzbach catchment (Germann, 1981). However, the Richards equation implemented in WaSiM just describes the process of matrix flow in the unsaturated zone and is not able to describe the process of preferential flow. With WaSiM it can either be modelled by adjusting the saturated conductivities in the soil module to higher values or by adapting the conceptual model parameters, in particular the parameters k_d and k_{rec}, in order to represent this phenomena. From a process-oriented point of view it might be better to "blame" the preferential flow on the conceptual model parameters instead of trying to describe a physically completely different process by a physically based model, i.e. trying to fit the Richards equation to represent preferential flow and matrix flow instead of just matrix flow. In general the distribution of the parameters estimated ROPE-PSO has a much smaller spread than the ones estimated by A-ROPE. This suggests that the Monte Carlo based algorithm cannot identify the region with the best model performance as the PSO based ROPE-PSO. Consequently the region comprising the good parameter vectors, i.e. the best plus a given uncertainty tolerance, gets larger

and might be less robust, i.e. less transferable to other flood events, which means the validation set.

The calibration performance of all considered algorithms is given in Fig. 16. The results of PSO-GA$_u$ (before deep samples were drawn) are given for an additional comparison. The calibration performance results are better than the results for the calibration in the previous case study just considering the conceptual model parameters. That result is not surprising. The better fit in the calibration might just be due to a larger number of free model parameters and has to be confirmed on the validation data in order to be considered as improvement. A better model fit is always possible with more free model parameters, but just makes sense if the estimated parameters can be transferred on other time periods or events. Referring to the mean estimated FloodSkill the original ROPE with a value of 0.43 is slightly outperformed by the A-ROPE (0.46) and a bit more by the ROPE-PSO algorithm (0.48). The corresponding uncertainty intervals according to the uncertainty in the observations are nearly the same for all algorithms. They correspond to a bandwidth of the Flood-Skill of approximately 0.1. Hence the PSO provides advantages in finding the region with the best model performance, however the differences are just marginal. Consider however that the slightly worse model performance of the A-ROPE estimates correspond to a much wider distribution of the model parameters as already discussed based on the results provided in Fig. 14. Furthermore it is evident that the deep parameter vectors estimated by ROPE-PSO in comparison with PSO-GA$_u$ do not have a better model performance on the calibration data. The deep parameter vectors just show a small decrease of the standard deviation of the corresponding model performances.

The transferability of the estimated parameters and thus the performance of the algorithms they are generated by can just be shown using the validation data. An overview of the estimated model performances on the validation events[6] is given in Fig. 17. Regardless of the used parameter estimation procedure, the achieved model performance on the validation data is better than the calibration with the conceptual model parameters only. This result confirms the results of Grundmann (2010), that a consideration of the uncertainty in the soil hydraulic parameters for flood events can improve the performance of a process-oriented hydrologic model when focus lies on modeling on flood events. Although the A-ROPE provides a slight advantage over the original ROPE approach, both Monte Carlo based algorithms ROPE and A-ROPE are clearly outperformed by the PSO based PSO-GA$_u$ and related ROPE-PSO estimates. This suggests that the previously discussed advantages of the PSO approach providing a better global optimum and a consequently smaller region of good parameter vectors correspond to a better robustness. That means that the set of estimated parameters contains less outliers with clearly below-average performance. Consid-

ering the limited number of maximum function evaluations used for the ROPE-PSO algorithm these results also confirm the efficiency of the PSO based approach. This is not surprising but should be taken into consideration when choosing between the Monte Carlo based and PSO based ROPE algorithms. Referring to the results of PSO-GA$_u$ and ROPE-PSO it is obvious that the parameters with high data depth do not have just a marginal better model performance on the validation data but also much less outliers on the side of the distribution corresponding to a lower model performance. For instance the worst overall FloodSkill for PSO-GA$_u$ is 0.18 whereas it is 0.26 after the depth based sampling, i.e. for ROPE-PSO. These results show the advantages of the depth based sampling, namely the possibility to filter out parameter vectors corresponding to a more volatile and consequently lower model performance. However a comparison of the Monte Carlo based algorithms (ROPE and A-ROPE) with the PSO based ROPE-PSO also show the limits of this approach. The performance of the sampled deep parameter vectors requires both an optimal sampling of the set of good parameter vectors and an effective sampling of deep parameters with respect to this set. This result is also supported by a clearly lower standard deviation of the ROPE-PSO results referring to all compared performance criteria. The better accuracy of the ROPE-PSO estimates together with less negative outliers is also reflected by a reduced model uncertainty. That means that not just the parameter uncertainty but also the complete model uncertainty can be tremendously reduced. Figure 18 shows the hydrographs and the corresponding parameter and model uncertainties for both algorithms. The model errors were computed by two normal distribution fitted on the positive and negative discharge errors, transformed with the normal quantile transformation (NQT) (Krzysztofowicz, 1997) according to a method presented by Engeland et al. (2010).

6 Discussion and conclusions

– This paper presents two new depth based parameter estimation method, that are well suited for the robust calibration of hydrologic models considering uncertainties. The Advanced Robust Parameter Estimation (A-ROPE), is a modified version of the depth based parameter estimation procedure presented by Bárdossy and Singh (2008). There are two differences between the A-ROPE algorithm and the original ROPE algorithm. The further development enables us sampling from different non-convex regions of attraction and at the same time preventing the algorithm from overfitting. The second algorithm is a PSO based Robust Parameter Estimation algorithm algorithm, entitled Robust Parameter Estimation with Particle Swarm Optimisation (ROPE-PSO). The major difference between ROPE-PSO and the previously presented ROPE algorithm is a substitution of the Monte Carlo based approach for the identification
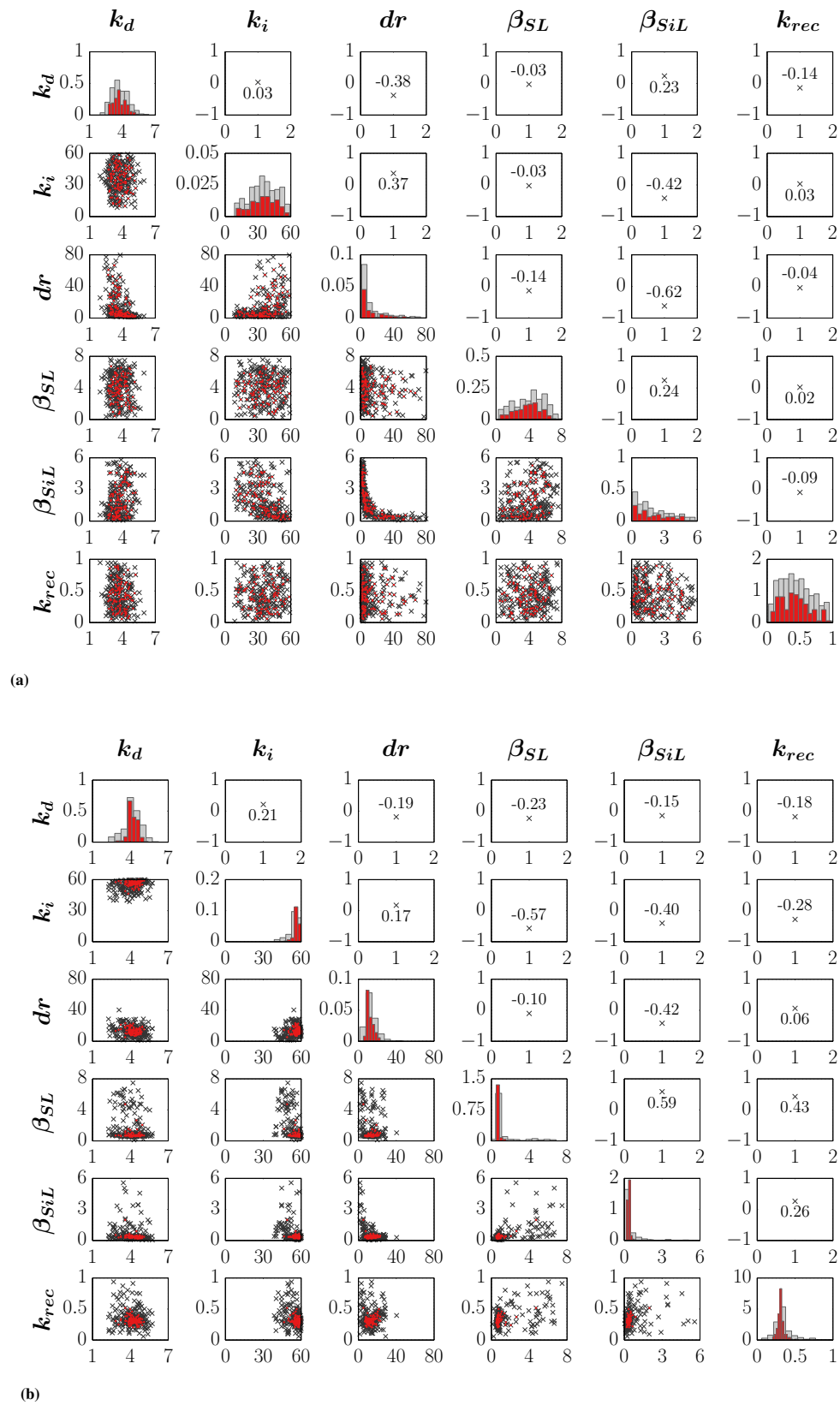
---

[6]For details refer to Table 5

Fig. 14. Distribution of the model parameter vectors estimated by A-ROPE **(a)** and ROPE-PSO **(b)**.

	ks	θ_s	α	n
Mean	7.231e-05	0.400	6.66	1.36
Std	4.224e-05	0.018	2.49	0.03
CV	0.58	0.05	0.37	0.03

	ks	θ_s	α	n
Mean	1.801e-05	0.412	4.24	1.29
Std	1.228e-05	0.003	1.49	0.02
CV	0.68	0.01	0.35	0.02

(a)

(b)

	ks	θ_s	α	n
Mean	1.215e-06	0.416	1.30	1.32
Std	5.893e-07	0.005	0.31	0.06
CV	0.48	0.01	0.24	0.05

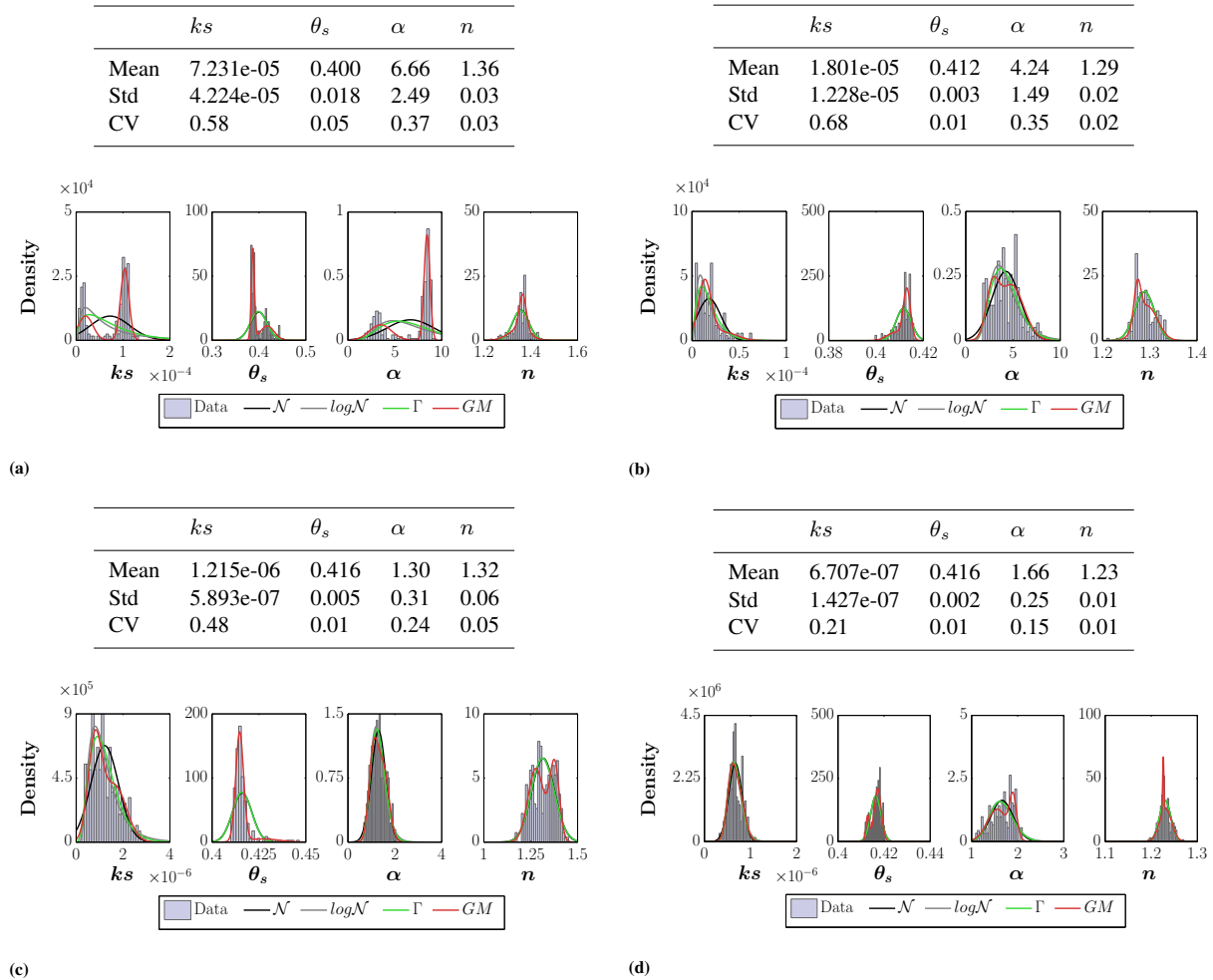|	ks	θ_s	α	n
Mean	6.707e-07	0.416	1.66	1.23
Std	1.427e-07	0.002	0.25	0.01
CV	0.21	0.01	0.15	0.01

(c)

(d)

Fig. 15. Distribution of the soil hydraulic parameters estimated by A-ROPE (left column) and ROPE-PSO (right column) for the soils SL (**a** and **b**) and SiL (**c** and **d**)

of good parameter vectors by a newly developed algorithm, entitled PSO-GA$_u$. We study the effectiveness of the newly developed algorithms in two case studies calibrating a process-oriented hydrologic model focussing on flood events. The results are compared with estimates generated by the original ROPE algorithm and two stats-of-the-art optimisation algorithms.

– In a first case study we compared the original ROPE, and the newly developed A-ROPE and ROPE-PSO approaches estimating three conceptual model parameters of the model WaSiM. We study the effect of observation uncertainty and confirm the results of Bárdossy and Singh (2008): the parameter vectors estimated by classic optimisation algorithms can lead to very different results in the validation and are not robust. Parameter vectors with equal model performance on the calibration data can lead to very different results in validation. Considering a set of identified parameters with

good model performance on the calibration data, members with shallow data depth near the boundary are sensitive to small changes and have a less probability to perform well on other time periods than solutions with high depth. The depth based parameter estimation approaches can identify a set of parameter vectors that shows a clearly better performance in the validation with tighter variation intervals, i.e. less outliers. In comparison with the original ROPE, the modifications of the Monte Carlo based A-ROPE provide slight advantages in terms of the validation performance. These results are even outperformed by the substitution of the Monte Carlo based sampling by ROPE-PSO approach where the good parameters are determined using an adapted PSO strategy.

– In a further case study we increased the number of considered model parameters. The additional parameters allow to account for the uncertainty in the soil hydraulic

Fig. 16. Calibration performance for the parameter vectors estimated by ROPE, A-ROPE, PSO-GA$_u$ and ROPE-PSO.
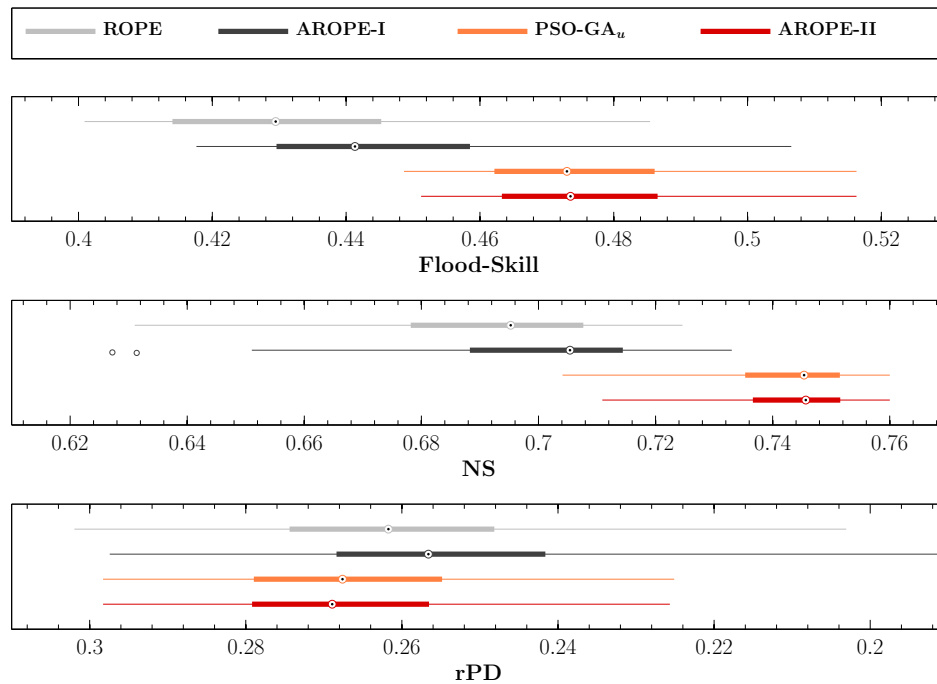
	FloodSkill			NS			rPD		
	Mean	Std	Range	Mean	Std	Range	Mean	Std	Range
ROPE	0.43	0.019	0.40 - 0.49	0.69	0.020	0.63 - 0.72	0.26	0.021	0.20 - 0.30
A-ROPE	0.46	0.019	0.42 - 0.51	0.70	0.019	0.63 - 0.73	0.25	0.021	0.19 - 0.30
PSO-GA$_u$	0.47	0.016	0.44 - 0.52	0.74	0.012	0.70 - 0.76	0.27	0.016	0.23 - 0.30
ROPE-PSO	0.48	0.014	0.45 - 0.52	0.74	0.011	0.71 - 0.76	0.27	0.015	0.23 - 0.30

parameters and an additional conceptual parameter of the soil module in WaSiM. The results of this case study show the limits of the Monte Carlo based ROPE and A-ROPE approaches for problems with a nonsmooth parameter surfaces with large flat areas in higher dimensions. Nonetheless the modifications in A-ROPE help to achieve an improved set of robust solutions. In this case study the effective and efficient PSO based search strategy in ROPE-PSO can show its full potential estimating an optimal approximation of the set of good parameter vectors. This is an important pre-requisite for the effectivity of the depth based sampling. The PSO based strategy can identify a much more concentrated set of good parameter vectors with the same tolerance interval on the calibration data. As a consequence the final deep solutions estimated by ROPE-PSO outperform the ROPE and A-ROPE estimates by lengths.

– The case studies in this paper revealed that the used hydrologic model WaSiM is not able to represent the correct peak flow values and the global catchment behaviour in terms of the streamflow at the catchment outlet with the same parameter vectors equally well. The tradeoff in these two objectives that are important for the modelling of flood events suggests the application of a multi-objective calibration strategy. Consider that the presented algorithm can be easily altered to a general multi-objective parameter estimation procedure.

– The application of data depth metrics can help to identify sets of robust parameter vectors. In general parameters with low data depth are near the boundary of a set of good model performance in the calibration are sensitive to small changes and do transfer to other time periods less well as high depth ones. However the model performance of the sampled deep parameters is also dependent to the quality of the estimated good parameter vectors.

Fig. 17. Validation performance for the parameter vectors estimated by ROPE, A-ROPE, PSO-GA$_u$ and ROPE-PSO.
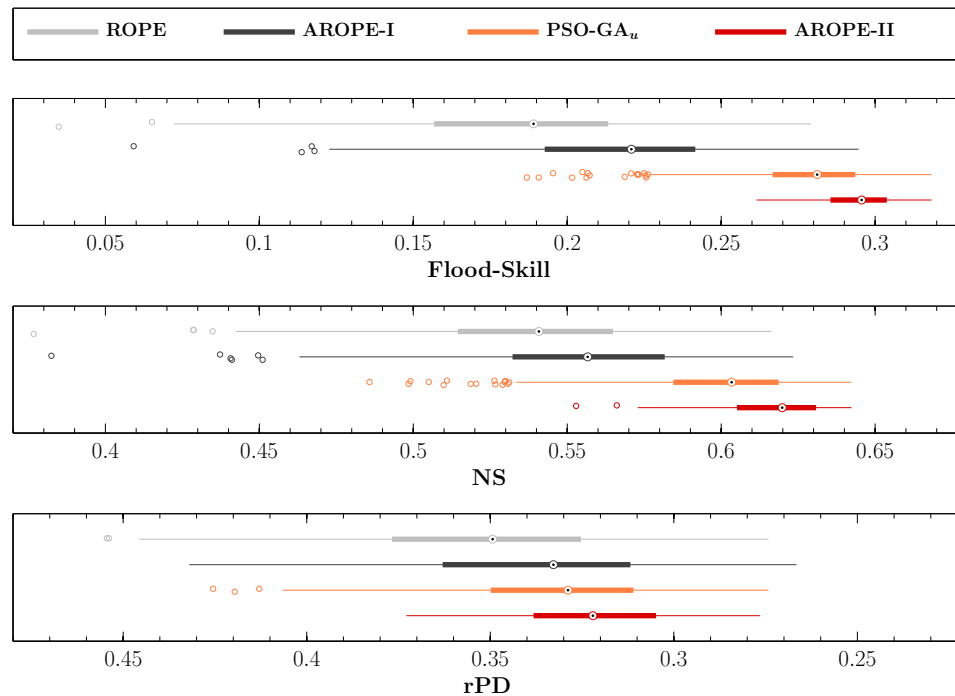
	FloodSkill			NS			rPD		
	Mean	Std	Range	Mean	Std	Range	Mean	Std	Range
ROPE	0.19	0.043	0.03 - 0.28	0.54	0.040	0.38 - 0.62	0.35	0.035	0.27 - 0.45
A-ROPE	0.22	0.038	0.06 - 0.29	0.55	0.039	0.39 - 0.62	0.34	0.033	0.27 - 0.43
PSO-GA$_u$	0.25	0.021	0.18 - 0.32	0.60	0.027	0.48 - 0.64	0.33	0.027	0.27 - 0.42
ROPE-PSO	0.30	0.014	0.26 - 0.32	0.62	0.018	0.55 - 0.64	0.32	0.023	0.28 - 0.37

– The case studies in this paper just consider a limited number of calibration parameters parameters. This is sufficient for the given model setup considering the small amount of observation data for flood events. The suggested method might perform even better for calibration tasks with a higher amount of useful calibration data or lower process dynamics where the set of good parameter vectors is much more clearly defined, e.g. water balance simulations. Bárdossy and Singh (2008) applied the original ROPE method for the estimation of nine parameters in the conceptual model HBV in a much larger catchment on daily basis. According to these results and our experience with other small test problems, we strongly believe that the developed technique might provide good results for higher dimensions as well. Consider however that the application of complex data depth functions, e.g. halfspace depth is limited on lower dimensions due to the computational complexity and the required number of solutions. For a limited number of points and a high dimensionality, all points are in the convex hull and have low depth.

The robust parameter estimation approach is a relatively new method which was applied to a limited number of case studies. We strongly propose a comparison with established uncertainty estimation methods, e.g. MCMC, GLUE or multi-objective calibration, in further research. Furthermore we suggest the further development of the ROPE method to a well-founded calibration tool considering uncertainties, e.g. assigning a likelihood based on the depth of the estimated parameters instead of their model performance as it is done in classical approaches. Due to the probably high tradeoff between the model's ability to represent both the peak flow values and the global system behaviour equally well, we propose the development and application of a multi-objective version of the presented approach.
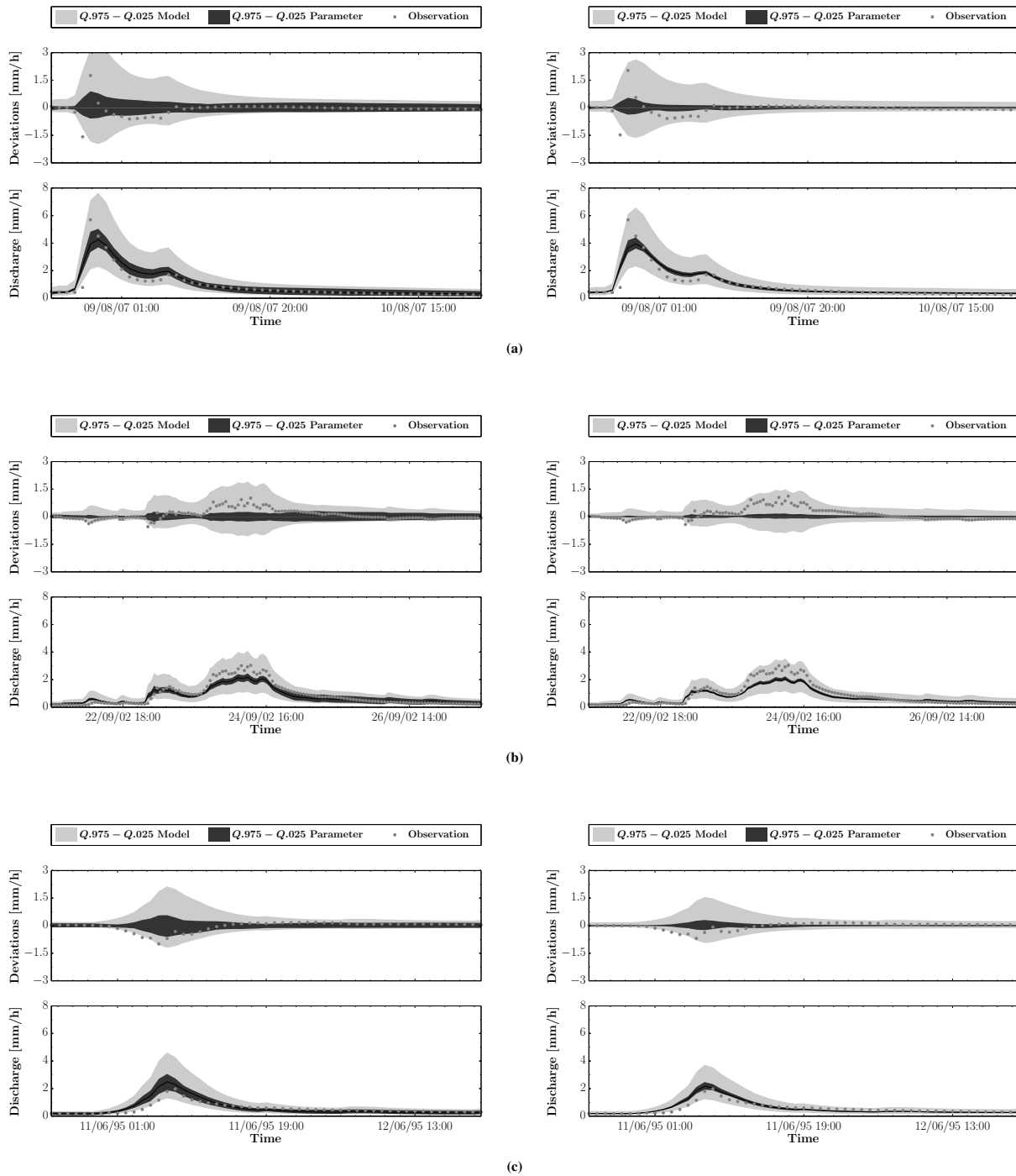
Fig. 18. Hydrograph prediction uncertainty associated with the uncertainty in the model (lighter shading) and parameter estimates (darker shading) for the flood events 4 **(a)**, 12 **(b)** and 19 **(c)**, estimated by A-ROPE (left column) and ROPE-PSO (right column). The dots correspond to the observed streamflow data. The shaded areas of uncertainty correspond to the 95% confidence intervals.

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.

Brakensiek, D., Rawls, W., and Stephenson, G.: Modifying SCS hydrological soil groups and curve numbers for rangeland soils, ASAE Paper, PNR-84-203, St. Joseph/ Michigan, 1984.

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.

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.

Cabrera, J. C. F. and Coello, C. A.: Multi-Objective Swarm Intelligent Systems, vol. 261 of Studies in Computational Intelligence, chap. Micro-MOPSO: A Multi-Objective Particle Swarm Optimizer that Uses a Very Small Population Size, p. 218, Springer Verlag, 2010.

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.

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

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, 2006.

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

Donoho, D. L. and Gasko, M.: Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness, The Annals of Statistics, 20, 1803–1827, http://www.jstor.org/stable/2242368, 1992.

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

Engeland, K., Renard, B., Steinsland, I., and Kolberg, S.: Evaluation of statistical models for forecast errors from the HBV model, J. Hydrol., 384, 142–155, doi:10.1016/j.jhydrol.2010.01.018, 2010.

Germann, P.: Untersuchungen über den Bodenwasserhaushalt im hydrologischen Einzugsgebiet Rietholzbach, no. 51 in Mitteilungen der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie, ETH Zürich, Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie, ETH Zürich, 1981.

Gill, M. K., Kaheil, Y. H., Khalil, A., McKee, M., and Bastidas, L.: Multiobjective particle swarm optimization for parameter estimation in hydrology, Water Resour. Res., 42, W07417, doi:10.1029/2005WR004528, 2006.

Grundmann, J.: Analyse und Simulation von Unsicherheiten in der flächendifferenzierten Niederschlags-Abfluss-Modellierung, vol. 8 of Dresdner Schriften zur Hydrologie, Institut für Hydrologie und Meteorologie, Technische Universität Dresden, 2010.

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.

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.

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

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

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

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.

Kennedy, J. and Eberhart, R.: Particle Swarm Optimization, in: Proceedings of IEEE International Conference on Neural Networks, vol. IV, p. 1942–1948, 1995.

Kennedy, J., Eberhart, R. C., and Shi, Y.: Swarm Intelligence, Morgan Kaufmann Series in Evolutionary Computation, Morgan Kaufman, 2001.

Krauße, T. and Cullmann, J.: Identification of hydrological model parameters for flood forecasting using data depth measures, Hydrol. Earth Syst. Sci. Discuss., 8, 2423–2476, doi:10.5194/hessd-8-2423-2011, 2011a.

Krauße, T. and Cullmann, J.: Data depth - parsimonious sampling of robust parameters in hydrological models, Appl. Math. Modelling, submitted, 2011b.

Krzysztofowicz, R.: Transformation and normalization of variates with specified distributions, J. Hydrol., 197, 286–292, doi:10.1016/S0022-1694(96)03276-3, 1997.

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, J. Hydrol., 331, 161–177, 2006.

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.

Liu, R. Y., Serfling, R., and Souvaine, D. L., eds.: Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications, vol. 72 of Series in Discrete Mathematics and Theoreti-

cal Computer Science, American Mathematical Society, 2006.

Marx, A.: Einsatz gekoppelter Modelle und Wetterradar zur Abschätzung von Niederschlagsintensitäten und zur Abflussvorhersage, vol. 160 of Mitteilungen des Instituts für Wasserbau, Universität Stuttgart, Institut für Wasserbau, Universität Stuttgart, available at: http://elib.uni-stuttgart.de/opus/volltexte/2007/3016/, last access: 6 December 2011, 2007.

May, R., Maier, H., and Dandy, G.: Data splitting for artificial neural networks using SOM-based stratified sampling, Journal of Neural Networks, 23, 283–294, 2010.

Mohamed, L., Christie, M., and Demyanov, V.: Comparison of Stochastic Sampling Algorithms for Uncertainty Quantification, SPE Journal, 15, 31–38, 2010.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I – A discussion of principles, J. Hydrol., 10, 282–290, 1970.

Perez, R. E. and Behdinan, K.: Particle swarm approach for structural design optimization, Comput. Struct., 85, 1579–1588, 2007.

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

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

Schulla, J. and Jasper, K.: Model Description WaSiM-ETH, Eidgenösische Technische Hochschule Zürich, 2007.

Seifert, M.: Untersuchungen zum Einfluss unsicherer Bodeninformationen im Schweizer Testeinzugsgebiet Rietholzbach, Master's thesis, Institut für Hydrologie und Meteorologie, Technische Universität Dresden, 2010.

Settles, M.: Breeding Swarms: A Hybrid GA/PSO for Real Parameter Optimization., Master's thesis, University of Idaho, Moscow, Idaho, 2005.

Settles, M. and Soule, T.: Breeding swarms: a ga/pso hybrid, in: GECCO '05: Proceedings of the 2005 conference on Genetic and evolutionary computation, 161–168, ACM Press, 2005.

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

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

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

Vrugt, J. A., ter Braak, C.J.F., Diks, C., Robinson, B., Hyman, J., and Higdon, D.: Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, Int. J. Nonlin. Sci. Num., 10, 273–290, 2009a.

Vrugt, J. A., ter Braak, C., Gupta, H., and Robinson, B.: Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, Stoch. Env. Res. Risk A., 23, 1011–1026, doi:10.1007/s00477-008-0274-y, 2009b.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water Resour. Res., 39, 1214, doi:10.1029/2002WR001746, 2003a.

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 Resour. Res., 39, 1201, doi:10.1029/2002Wr001642, 2003b.

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.

Warrick, A. W., Mullen, G. J., and Nielsen, D. R.: Scaling Field-Measured Soil Hydraulic Properties Using a Similar Media Concept, Water Resour. Res., 13, 355–362, doi:10.1029/WR013i002p00355, 1977.

Wösten, J. H. M., Lilly, A., Nemes, A., and Le Bas, C.: Development and use of a database of hydraulic properties of European soils, Geoderma, 90, 169–185, http://www.sciencedirect.com/science/article/B6V67-3WWRF2R-C/2/50af110f999083a308c55eedc997ed4c, 1999.

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