



Robust multi-objective calibration strategies – possibilities for improving flood forecasting

T. Krauß¹, J. Cullmann², P. Saile², and G. H. Schmitz¹

¹Institute of Hydrology and Meteorology, Department of Hydrology, Dresden University of Technology, Bergstr. 66, 01062 Dresden, Germany

²IHP/HWRP Secretariat, Federal Institute of Hydrology, Am Mainzer Tor 1, 56068 Koblenz, Germany

Correspondence to: T. Krauß (thomas@krausze.com)

Received: 24 March 2011 – Published in Hydrol. Earth Syst. Sci. Discuss.: 15 April 2011

Revised: 24 May 2012 – Accepted: 15 June 2012 – Published: 15 October 2012

Abstract. Process-oriented rainfall-runoff models are designed to approximate the complex hydrologic processes within a specific catchment and in particular to simulate the discharge at the catchment outlet. Most of these models exhibit a high degree of complexity and require the determination of various parameters by calibration. Recently, automatic calibration methods became popular in order to identify parameter vectors with high corresponding model performance. The model performance is often assessed by a purpose-oriented objective function. Practical experience suggests that in many situations one single objective function cannot adequately describe the model's ability to represent any aspect of the catchment's behaviour. This is regardless of whether the objective is aggregated of several criteria that measure different (possibly opposite) aspects of the system behaviour. One strategy to circumvent this problem is to define multiple objective functions and to apply a multi-objective optimisation algorithm to identify the set of Pareto optimal or non-dominated solutions. Nonetheless, there is a major disadvantage of automatic calibration procedures that understand the problem of model calibration just as the solution of an optimisation problem: due to the complex-shaped response surface, the estimated solution of the optimisation problem can result in different near-optimum parameter vectors that can lead to a very different performance on the validation data. Bárdossy and Singh (2008) studied this problem for single-objective calibration problems using the example of hydrological models and proposed a geometrical sampling approach called Robust Parameter Estimation (ROPE). This approach applies the concept of data depth in order to overcome the shortcomings of automatic calibration procedures

and find a set of robust parameter vectors. Recent studies confirmed the effectivity of this method. However, all ROPE approaches published so far just identify robust model parameter vectors with respect to one single objective. The consideration of multiple objectives is just possible by aggregation. In this paper, we present an approach that combines the principles of multi-objective optimisation and depth-based sampling, entitled Multi-Objective Robust Parameter Estimation (MOROPE). It applies a multi-objective optimisation algorithm in order to identify non-dominated robust model parameter vectors. Subsequently, it samples parameter vectors with high data depth using a further developed sampling algorithm presented in Krauß and Cullmann (2012a). We study the effectivity of the proposed method using synthetic test functions and for the calibration of a distributed hydrologic model with focus on flood events in a small, pre-alpine, and fast responding catchment in Switzerland.

1 Introduction

Hydrologic models are simplified, conceptual representations of a part of the hydrologic cycle. They relate rainfall to streamflow on a continuous basis. Many of those models are driven by a vector of parameters that cannot be measured directly, but must be determined through indirect methods. This is an important aspect of model calibration. Efficient and effective parameter estimation techniques are a crucial factor for the successful application of these models. In the process of parameter estimation, the values of the

model parameters are adjusted until the catchment behaviour is closely matched.

Traditionally, this calibration is performed manually adjusting the parameters while visually inspecting the agreement between observations and model predictions. For the application of automatic approaches, the calibration is formulated as an optimisation problem. A purpose-specific objective function f quantifies the agreement between observations and simulation results. Many practical studies suggest that one single objective function, no matter how carefully chosen, is often insufficient to represent all characteristics of the system behaviour (Gupta et al., 1998; Cullmann, 2006; Gill et al., 2006; Fenicia et al., 2007). For instance, the mean absolute error of the discharge at the catchment outlet might be a good indicator for the ability to represent the water balance; however, it is likely to be inadequate to measure the model performance for flood forecasts where a correct simulation of the peak flow value and timing is crucial. Consequently, single-objective calibration approaches that provide one unique, global best parameter vector are in many cases not considered acceptable by experienced hydrologists. The most elementary solution to circumvent this problem is to aggregate several objective functions. However, this approach involves a great deal of subjective judgment and neglects the global bests for individual objective functions. Another advanced option is a multi-objective view of the optimisation problem referring to the concept of Pareto optimality (Yapo et al., 1998). A multi-objective optimisation algorithm approximates the set of non-dominated (i.e. Pareto optimal) solutions with respect to a set of given objectives. The set of Pareto optimal solutions is denoted as the Pareto optimal set or Pareto set \tilde{P} , and the image of \tilde{P} under the mapping of all considered objectives F is called Pareto front. Often, both terms are used synonymously. This can however lead to misunderstandings. Consider that the term Pareto optimal set means the set of true or estimated solutions in the parameter space, whereas the term Pareto front means its image in the objective space. This difference is important for the major concept of this paper. The Pareto optimal set or Pareto set reflects the trade-offs among all given objectives. Evolutionary algorithms, i.e. genetic algorithms (GA), and particle swarm optimisation (PSO) or other population-based algorithms can be used to perform this task. There are many different strategies with different strengths and shortcomings. Recently, multi-method algorithms have found favour in order to approximate this set. The approximated Pareto optimal set is indicated by \tilde{A} .

Bárdossy and Singh (2008) studied one of the major shortcomings of automatic calibration procedures that understand the problem of model calibration just as an optimisation problem: the parameters of fitted hydrologic models depend upon the input data. The quality of input data cannot be assured as there may be measurement errors for both input and state variables. That is why the estimated best parameter vectors can lead to a very different performance on the

validation data. The actual goal of a good model calibration should not be to find parameter vectors that perform best for the calibration period but to find parameter vectors that are robust. In this context, the term robust is closely related to the terms transferable and less sensitive. Robust parameter vectors should not just perform well on the calibration data but also provide a sufficient model performance in the validation. Hence, these parameters can be transferred to other time periods. Furthermore, robust parameter vectors should be as insensitive as possible. This means that a small change of the estimated parameter vectors should not lead to a significant change (usually a decrease) of the model performance. Against this background, the term sensitive is always related to a good model performance and to the general sensitivity of the considered model parameters. This means that a sensitive model parameter will always stay sensitive. However, a small change of this parameter within the region of robust model parameter vectors should have a smaller effect on the (good) model performance than in other regions. This requirement is important, because the set of optimal parameters cannot be exactly determined and may even slightly change under slightly different conditions.

The application of advanced parameter estimation approaches that do not understand a model calibration as a pure optimisation task is a necessary though not sufficient condition for the determination of robust model parameters. The success of a robust modelling strategy also strongly depends on a good implementation of the principle of parsimony. The goal of this concept is to obtain a model that represents a desired structure with as few parameters as possible. The essential number of parameters depends on the available information, e.g. the number of gaging stations in a catchment or the temporal resolution of the observations. The selection of a parsimonious model structure was subject of many studies in hydrology (e.g. Beven, 1989; van der Linden and Woo, 2003; Boyle et al., 2006; Wagener and Wheeler, 2006). Another essential prerequisite for the estimation of robust model parameters is the selection of appropriate calibration data. The data used for calibration should be both representative for the considered processes and contain as much information as possible that can be used to identify the parameters that describe the considered processes best. The development of well-founded statistical methods based on information theory that are suitable to perform this task for both data-driven and process-oriented models has attracted rising scientific interest (see Maier and Dandy, 2000; Wagener and Wheeler, 2002; Wagener and Gupta, 2005; Thyer et al., 2006; Singh, 2010). Of course also the selection of appropriate objective functions plays an important role for a successful model calibration. Typically, the formulation of uncorrelated criteria that use several considered model outputs adds new information. A comprehensive overview and a more detailed discussion of these issues within the scope of multi-objective optimisation are given in Efstratiadis and Koutsyiannis (2010).

The approach presented in this paper focusses on the enhancement of the results provided by approved evolutionary multi-objective optimisation methods. The underlying principle to achieve this goal is the application of data depth metrics in order to sample robust parameter vectors with respect to a set of identified parameter vectors with good model performance. Bárdossy and Singh (2008) developed a first method using this approach and called it Robust Parameter Estimation Method (ROPE). Recent studies using further developed versions of this methodology (e.g. Krauß and Cullmann, 2012a) confirmed the potentials of the depth-based parameter sampling for the estimation of robust hydrologic parameter vectors.

However, all ROPE approaches published so far just identify robust model parameter vectors with respect to one single objective. The consideration of multiple objectives is just possible by aggregation. In this paper we present a new method, entitled Multi-Objective Robust Parameter Estimation (MO-ROPE) that synthesizes the advantages of both the multi-objective view and robust parameter estimation. In order to quantify the uncertainty of the parameterisation with respect to the given objectives, the method estimates a set of robust model parameter vectors applying a two-step approach. Within the first step, a suitable multi-objective optimisation algorithm is used to approximate the Pareto optimal set. In a second step, parameter vectors with high data depth (with respect to the Pareto set) are sampled assuming that those parameter vectors are more robust than the complete Pareto optimal set. Notwithstanding the already raised fact that the application of advanced parameter estimation approaches is not sufficient for the estimation of robust model parameters, we will follow the notation of Bárdossy and Singh (2008) and denote the parameter estimation using data depth metrics by the abbreviation ROPE.

In the following section, we introduce a multi-objective parameter estimation technique that applies evolutionary multi-objective optimisation algorithms and the concept of data depth in order to estimate a robust set of parameter vectors for a given multi-objective calibration problem. First, we discuss suitable techniques for the approximation of the Pareto optimal set. We suggest the application of a modified version of the advanced multi-method calibration framework AMALGAM first presented by Vrugt et al. (2009). It applies an additional hybrid search strategy entitled MO-PSO-GA that is a modified version of a search strategy that has proven successful within the frame of another single-objective ROPE strategy. Afterwards, we briefly introduce and discuss the principle of data depth and the depth-based sampling approach. The developed solution is tested on a set of well-known multi-objective test problems. We study the effectivity and efficiency of the suggested multi-objective optimisation strategy and show the advantages of the depth-based sampling strategy for several low- and high-dimensional test problems that are subject to uncertainty. Furthermore, a real world case study shows the success of

the developed methodology calibrating a distributed hydrologic model in a small catchment with high process dynamics focussing on flood events.

2 Combining multi-objective optimisation and depth-based parameter sampling

The proposed MO-ROPE approach synthesizes the concepts of multi-objective model calibration and the depth-based ROPE approach. Therefore, we will thoroughly introduce both concepts in this section.

2.1 Effective and efficient approximation of the Pareto set

An essential prerequisite for a successful application of the proposed MO-ROPE method is an effective and efficient approximation of the Pareto set for a given calibration problem. Recently, several evolutionary search strategies, e.g. genetic algorithms (GA) and particle swarm optimisation (PSO), became popular to fulfill this task. Multi-method approaches even go one step further and apply several search strategies in parallel with the goal to exploit their strengths, and to minimise their weaknesses. A very popular and effective state-of-the-art method in this category is the AMALGAM approach presented by Vrugt and Robinson (2007) that merges the strengths of four well-founded multi-objective optimisation strategies. Benchmark results using a set of well-known multi-objective test problems show that AMALGAM approaches a factor of 10 improvement over current optimization algorithms for more complex, higher-dimensional problems. Following Wolpert and Macready's "no free lunch" theorem, showing that it is impossible to develop a single search algorithm that will always be superior to any other algorithm, an outperformance of AMALGAM by a single newly developed algorithm is thus virtually excluded.

However, the existence of a superior multi-method approach should not completely prevent the development of alternative multi-objective calibration strategies. In case that there is a new method that provides at least in some cases a superior performance to other considered single calibration strategies, development and application within the frame of a multi-method framework might be successful. Furthermore, there are many real world multi-objective calibration problems that are simple enough to be solved by such approaches. That is why we adapted a search strategy called PSO-GA_u that is based on PSO and GA in order to approximate the Pareto set¹. It has proven to be successful in estimating a set of good parameter vectors with a given uncertainty bound within the scope of a ROPE framework that is provided in Krauß and Cullmann (2012a). The developed approach is entitled Multi-Objective optimisation by

¹The subscript "u" indicates that the PSO-GA_u is not just a classical optimisation procedure but also provides an uncertainty range.

Algorithm 1 MO-PSO-GA.

```

initialise the non-dominated front:  $\tilde{\mathcal{A}} \leftarrow \emptyset$ 
for all particles  $i$  do
  initialise position and local best  $\mathbf{x}_i, \hat{\mathbf{x}}_i \in \mathcal{U}[\mathbf{x}_{lb}, \mathbf{x}_{ub}]$ 
  and velocity  $\mathbf{v}_i \leftarrow \mathbf{0}$ 
end for
while stop criteria not met do
  for all particles  $i$  do
    in case that the particle  $\mathbf{x}_i$  is not dominated by all mem-
    bers in  $\tilde{\mathcal{A}}$ , add it to this set  $\tilde{\mathcal{A}} \leftarrow \tilde{\mathcal{A}} \cup \mathbf{x}_i$ 
    remove all members from  $\tilde{\mathcal{A}}$  that are dominated by  $\mathbf{x}_i$ 
    update the local best  $\hat{\mathbf{x}}_i$  as the best position found so far
    from the particle with index  $i$  in case that it weakly domi-
    nates all positions explored by this particle so far
  end for
  assign to each particle a random member of the archive  $\hat{\mathbf{g}}_i \in$ 
   $\tilde{\mathcal{A}}$  as a “personal” global best
  discard the worst  $n_\psi \leftarrow \psi \cdot \#\{\text{particles}\}$  from the population

  initialise genetic offspring  $o_{ga} \leftarrow \emptyset$ 
  for  $i = 1$  to  $\frac{n_\psi}{2}$  do
    select a pair  $\{\mathbf{x}_1, \mathbf{x}_2\}$  from the population by tournament
    selection
    apply the VPAC operator to generate new offspring
     $\{\mathbf{x}_{1o}, \mathbf{x}_{2o}\} \leftarrow \text{vpac}(\{\mathbf{x}_1, \mathbf{x}_2\})$ 
     $o_{ga} \leftarrow o_{ga} \cup \{\mathbf{x}_{1o}, \mathbf{x}_{2o}\}$ 
  end for
  for all particles  $i$  do
    update velocity using equation
     $\mathbf{v}_i \leftarrow \omega \mathbf{v}_i + \phi_1 R_2(\hat{\mathbf{x}}_i - \mathbf{x}_i) + \phi_2 R_1(\hat{\mathbf{g}}_i - \mathbf{x}_i)$ 
    update position using equation
     $\mathbf{x}_i \leftarrow \mathbf{x}_i + \mathbf{v}_i$ 
  end for
  merge new population with genetic offspring
  particles  $\leftarrow$  particles  $\cup o_{ga}$ 
end while

```

Particle Swarm Optimisation and Genetic Algorithm (MO-PSO-GA). A pseudocode listing of the algorithm is given in Algorithm 1. The algorithm was set up according to the ideas provided by Settles and Soule (2005). They introduced a hybrid between a GA and PSO. The algorithm is controlled by a parameter called the breeding ratio ψ . This parameter controls the proportion of the population that is not moved according to the PSO strategy 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 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 different considered calibration problems. 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 technique: from the pool of possible candidates, a subset of parameter vectors is selected by tournament selection. They are recombined by an oper-

Algorithm 2 VPAC operator.

```

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

```

ator that is called Velocity Propelled Averaged Crossover (VPAC). The goal of its application is to create two children whose positions are in between the parent’s positions, but accelerated away from the parent’s current direction (negative velocity) in order to increase diversity in the population. This might be effective because, towards the end of a typical PSO run, the population tends to be highly concentrated in a small portion of the search space which reduces the effective search space. Algorithm 2 shows how the new child position vectors and velocities are calculated using VPAC. The used PSO strategy is an adaption of the strategy used in the PSO-GA_u approach provided in Krauße and Cullmann (2012a). The movement of the particles is controlled by three parameters: the particle inertia weight ω , the cognitive attraction ϕ_1 and the social attraction ϕ_2 . The parameter ω determines the velocity of the proper motion of the particles. ϕ_1 specifies the degree of movement towards the local optimum, and ϕ_2 controls the movement towards the global optimum of the whole particle swarm. We set the algorithm’s parameters according to literature recommendations (see Table 1). For further details and studies regarding the setting of the algorithm’s parameters, refer to Perez and Behdinan (2007) and Settles (2005) who 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).

In contrast to normal single-objective approaches, it stores all so far approximated Pareto optimal parameter vectors in a set $\tilde{\mathcal{A}}$ which is used to direct the search towards the complete region enclosing the Pareto optimal parameter vectors. Differently to other MO-PSO approaches (e.g. Gill et al., 2006), the the so-called “personal” global best for each particle is not the closest particle of the so far approximated Pareto optimal set $\tilde{\mathcal{A}}$, but a random member of $\tilde{\mathcal{A}}$. Furthermore, each particle has a local best that is just updated by the current position in case that it weakly dominates all so far found positions of the particle. After a given number of iterations, the set $\tilde{\mathcal{A}}$ holds a number of parameter vectors that represent the Pareto set of the given multi-optimisation problem.

We assume that the introduction of additional features to a normal PSO search strategy can outperform other simple multi-objective calibration strategies and might be a useful complement to existing advanced multi-method optimisation strategies, e.g. the AMALGAM algorithm. Therefore, we integrated the developed algorithm as an additional search strategy into the AMALGAM framework. In the following, the extended AMALGAM approach will be denoted by AMALGAM*. The AMALGAM framework contains a simple handling of boundary constraints. Infeasible solutions that are out of the bounds are just set to the bounds in order to preserve their feasibility. In order to enable the framework to deal with more complex bounds, we added some features of a constraint handling technique based on adaptive penalty functions and a distance measure proposed by Woldeesenbet et al. (2007). This method uses the number of feasible individuals in the population in order to be able to control whether a modified objective function focusses just on the objective values or the constraint violation. As long as all members of the current population are feasible, the objectives remain unchanged.

2.2 Robust parameter estimation

The application of automatic calibration procedures for model parameter estimation often completely neglects possible uncertainties in the observations used to quantify the matching of simulated values and measurements. Bárdossy and Singh (2008) studied this problem: due to the complex-shaped response surface and the erroneous observations, the solution of the optimisation problem can lead to very different near-optimum parameter vectors that correspond to a much different model performance on the validation data. The actual goal of a good model calibration should not be to find parameter vectors that perform best for the calibration period but to find parameter vectors that

- lead to good model performance over the selected time period;
- are not sensitive: small changes of the parameters should not lead to very different results;
- are transferable: they perform well for other time periods and might also perform well on other catchments.²

Typically, such parameter vectors also lead to a hydrologically reasonable representation of the corresponding processes in the context of the possibilities of the used model. According to Bárdossy and Singh (2008), we call such parameter vectors robust. There are two possibilities to improve existing calibration approaches in order to identify robust parameter vectors. One starting point that recently attracted ris-

²Singh (2010) studied the influence of robust parameter estimation on the transferability of model parameters. Consider however that this aspect is not a focus in this paper.

Table 1. Default values of the parameters controlling the MO-PSO-GA 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

ing scientific interest is a more intelligent selection of the calibration data (see Wagener and Wheeler, 2002; Wagener and Gupta, 2005; Thyer et al., 2006); another one is the development of advanced methods for the identification of robust model parameters.

A methodology that focusses on the latter approach is the robust parameter estimation approach (ROPE). It was first presented by Bárdossy and Singh (2008). The ROPE approach is based on the application of the principle of data depth in order to sample robust parameter vectors. Data depth is a statistical method used for multivariate data analysis which has recently attracted a lot of research interest (e.g. Cramer, 2003; Liu et al., 2006). A specific data depth function assigns a numeric value to a given point which corresponds to its centrality, with respect to a set of points. This approach provides a center-outward ordering of points in Euclidean space of any dimension with respect to a given point set or distribution. This provides the possibility of a new non-parametric multivariate statistical analysis 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 with high depth with respect to a given point set, are more robust in order to represent the underlying distribution than a whole set. The deep points can be identified using a data depth metric implemented in a data depth function. Most proposed metrics used in data depth functions are inherently geometric, with a numeric value assigned to each data point that represents its centrality within the given dataset. The depth median, the point of maximal depth, is the depth-based estimator for the center of the dataset. Depth contours can be used to visualize and quantify the data. For instance they can be used to define central regions enclosing a part of the space with high depth. The concept of data depth is illustrated in Fig. 1 by a simple 2-dimensional example. For a random point set in \mathbb{R}^2 , the data depth was computed for each point of the set with respect to the point set itself. The used depth function is the halfspace depth. According to its originator, it is also known as Tukey depth (Tukey, 1975). It is one of the best known among the data depth measures in nonparametric statistics, and in discrete and computational geometry and has proved to be a very robust measure in order to identify the center of a multivariate dataset (e.g. Rousseeuw and Struyf, 1998; Cramer, 2003; Serfling, 2006). According to Donoho

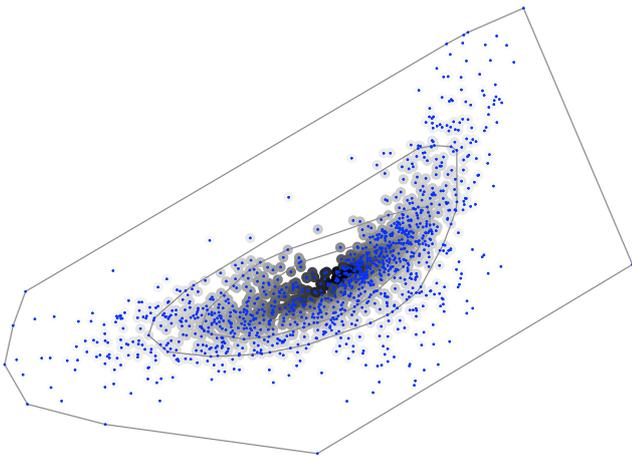


Fig. 1. 2-dimensional point set shaded according to assigned depth. A darker point represents higher depth. The lines indicate convex hulls enclosing the 25 %, 50 %, 75% and 100 % deepest points. The used depth function is halfspace depth.

and Gasko (1992), the halfspace depth of an arbitrary point $\theta = (\theta_1, \dots, \theta_d) \in \mathbb{R}^d$ with respect to a d -dimensional dataset Z is defined as the smallest number of data points in any closed halfspace with boundary through θ . It can be written as

$$\text{hdepth}(\theta | Z) := \min_{\|\mathbf{u}\|=1} \#\{i, \mathbf{u}^\top \cdot \mathbf{x}_i \geq \mathbf{u}^\top \cdot \theta\} \quad (1)$$

where \mathbf{u} ranges over all vectors in \mathbb{R}_d with $\|\mathbf{u}\| = 1$. Within this equation, the symbol \top just indicates that the labelled vector is transposed.

Very often the halfspace depth is normalised by dividing the “hdepth” value by the number of points in the set Z :

$$\text{hdepth}^*(\theta | Z) := \frac{\text{hdepth}(\theta | Z)}{\#Z}. \quad (2)$$

The principle of the halfspace depth of a point θ with respect to a point set Z in \mathbb{R}^2 is illustrated in Fig. 2. The plot shows some lines through θ that split the parameter space into two halfspaces. Obviously, each possible halfspace through θ contains at least two points of the given set. Hence, the non-normalised halfspace depth $\text{hdepth}(\theta | Z) = 2$. For any further details of the halfspace depth, refer to Tukey (1975); Donoho and Gasko (1992); Rousseeuw and Struyf (1998).

As mentioned above, Bárdossy and Singh (2008) developed a method entitled ROPE that applies the principle of data depth in order to identify a set of robust model parameters. The approach was further developed by Krauße and Cullmann (2012a). In general all ROPE approaches consist of two steps:

1. In a first step, a set of model parameter vectors with a reasonable good model performance is identified. The

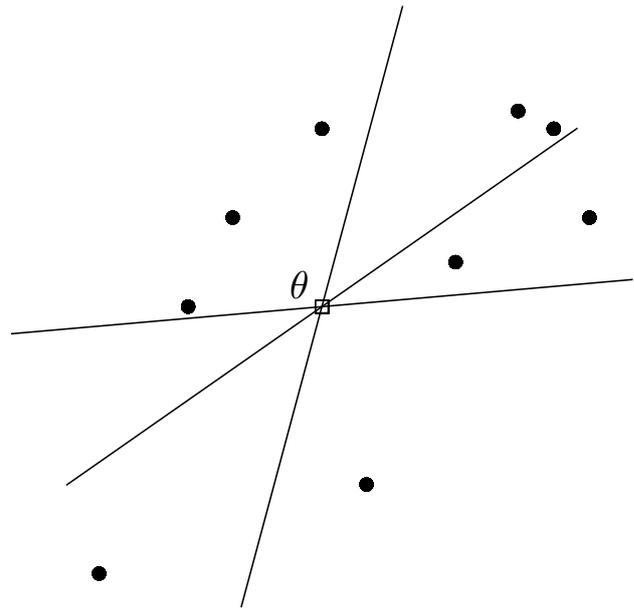


Fig. 2. Illustration of the halfspace depth in the two-dimensional space. The point θ has depth 2 with respect to the given point set.

estimated parameter vectors achieve the best possible performance with respect to the given objectives and calibration data. Thus, they are from now on called the good parameter vectors. For single-objective problems, the good parameter vectors can be expressed by a set comprising the global optimum plus an uncertainty bound that depends on the assumed uncertainties. Considering multi-objective calibration, we suggest to express the good parameter vectors by a set of non-dominated parameter vectors, i.e. the Pareto optimal set.

2. Afterwards, a set of parameter vectors deep with respect to the previously identified set of good parameter vectors is generated under the assumption that those parameter vectors are more robust than the complete set of good parameter vectors.³

The depth-based sampling with ROPE can be very useful for the estimation of robust hydrologic model parameters. Different studies with the semi-distributed HBV model focussing on the simulation of the water balance (Bárdossy and Singh, 2008) and the modelling of flood events with a distributed hydrologic model (Krauße and Cullmann, 2012a) illustrate the advantages and potential of this approach.

³“The reason for this is that one assumes that the low depth points can be regarded as an iso-hypersurface corresponding to the selected level. If one assumes continuity of the objective function then higher values of the function are expected in the interior of the set” (Bárdossy and Singh, 2008, p. 1278).

Algorithm 3 MO-ROPE.

- 1: Execute the multi-objective PSO-based MO-PSO-GA procedure to estimate an approximation of the Pareto optimal set of the problem. This set is denoted by \tilde{A} .
- 2: Sample parameter vectors $\mathcal{D}_{\tilde{A}}$ with high data depth w.r.t. \tilde{A} by the *GenDeep* strategy provided in Krauß and Cullmann (2011a). Consider that the data depth is computed within the parameter space.
- 3: **return** $\mathcal{D}_{\tilde{A}}$

2.3 Synthesizing multi-objective optimisation and robust parameter estimation

We propose to synthesize the strengths of multi-objective optimisation and robust parameter estimation with data depth functions in a new algorithm, entitled multi-objective robust parameter estimation (MO-ROPE). It synthesizes the advantages of a multi-objective view and the ROPE approach. The approach firstly applies a suitable multi-objective optimisation algorithm in order to estimate a set of non-dominated solutions \tilde{A} , thus the Pareto optimal set. A suitable method to do this task could be one of the methods discussed above, e.g. the AMALGAM approach. In a second step, the algorithm samples a set of parameter vectors that are deep with respect to the previously identified set \tilde{A} . The result is a set of parameter vectors $\mathcal{D}_{\tilde{A}}$ that has high data depth with respect to the set \tilde{A} . We propose to do the sampling using a stratified approach called *GenDeep*. It provides advantages for the sampling from non-elliptic, banana-shaped and multimodal distributions. That might be a crucial in case of Pareto optimal sets that are distributed non-linearly or even over different distinct regions in the parameter space. The *GenDeep* strategy is thoroughly discussed in Krauß and Cullmann (2012b). It was already successfully applied for the sampling of deep parameter vectors within the frame of a further developed single-objective ROPE strategy (Krauß and Cullmann, 2012a). Consider that MO-ROPE samples the deep parameters in the parameter space and not in the objective space. Hence, the estimated robust solutions are just dependent on the geometrical structure of the previously identified Pareto optimal set. Any further assumptions regarding realistic ranges of the considered objectives are not required. This makes it an in principle different approach in comparison to other published approaches that distinguish robust and non-behavioural solutions by subjective cut-off thresholds in the objective space (e.g. Efstratiadis and Koutsyiannis, 2010). Although this solution can easily select robust parameter vectors, it requires additional assumptions on the optimal range of the considered objectives. Furthermore, the non-behavioural solutions might not always exclusively correspond to the extreme tails of the Pareto front. Hence, the depth-based sampling completely forgets about the location of the Pareto optimal solutions in the objective space.

It just tries to identify parameter vectors that are located in the central regions of the Pareto optimal set in the parameter space. A pseudocode listing of the proposed MO-ROPE approach is provided in Algorithm 3. The developed solution enables the possibility to apply the ROPE approach on multi-objective calibration problems. The algorithm was implemented in MATLAB and C++ and is embedded in a robust parameter estimation framework which comprises other published robust parameter estimation approaches. The framework is open source and available from the author.

3 Case studies

3.1 Preliminary case study: investigating the effectivity and efficiency of different multi-method multi-objective algorithms using common test problems

The published results of single-objective robust parameter estimation approaches have shown that the quality of the set of good parameter vectors plays an important role for the robustness of the estimated model parameter vectors (e.g. Krauß and Cullmann, 2012a). This is why an effective estimation of the non-dominated solutions in the first algorithmic step is an important prerequisite for a successful application of a multi-objective robust parameter estimation approach. Thus, in a preliminary case study, we applied the multi-objective optimisation algorithms discussed in the previous section in order to approximate the Pareto set for several benchmark problems with analytical solutions. We compare the performance of the stand-alone MO-PSO-GA approach with the results obtained by other simple evolutionary multi-objective optimisation strategies. Furthermore, we discuss the advantages of the integration of this hybrid approach into the advanced multi-method algorithm AMALGAM. We compare the original version of AMALGAM that already applies four approved multi-objective optimisation algorithms⁴ and an extended version called AMALGAM* that comprises the MO-PSO-GA approach as well.

For the first problem, entitled COELLO function, we followed the settings in Gill et al. (2006). We ran the the algorithms with a population size of 30 and set the maximum number of function evaluations to 5000. The true front of and the non-dominated front estimated by MO-PSO-GA are shown in Fig. 3. It is obvious that the true front is approximated well and all sections of the front are uniformly covered. The same holds true for the estimates of both considered AMALGAM modifications. Thus, for an objective comparison of the approximated front against the true front, we calculated two performance metrics and compared them

⁴The original AMALGAM contains the genetic NSGA-II algorithm (Deb et al., 2002), a simple PSO approach (Kennedy et al., 2001), the adaptive metropolis search (AMS) (Haario et al., 2001), and differential evolution (DE) (Storn and Price, 1997).

Table 2. Description of two simple multi-objective test problems used in this study

Problem	N	Parameter range	Objective functions	Characteristics
COELLO	2	[0, 7]	$f_1(\mathbf{x}) = x_1^2 - x_2$	Nonconvex
			$f_2(\mathbf{x}) = -\frac{x_1}{2} - x_2 + 1$	
			subject to	
			$0 \geq \frac{x_1}{6} + x_2 - 6.5, \quad 0 \geq \frac{x_1}{2} + x_2 - 7.5,$ $0 \geq 5x_1 + x_2 - 30, \quad x, x_2 \geq 0$	
VRUGT $_N$	{2,10,30}	[0, 1]	$f_1(\mathbf{x}) = \sum_{i=1}^N x_i^2$	Convex
			$f_j(\mathbf{x}) = \sum_{i=1}^{j-1} x_i^2 + (x_j - 1)^2 \sum_{i=j+1}^N x_i^2$ where $2 \leq j \leq (N + 1)$	

Table 3. Performance metrics for the non-dominated fronts estimated by several evolutionary multi-objective optimisation algorithms edited according to Gill et al. (2006).

Algorithm	GD	SP
AMALGAM*	0.0005	0.0570
AMALGAM	0.0005	0.0581
MO-PSO-GA	0.0006	0.0621
MOPSO (Gill et al., 2006)	0.0122	0.1415
MOPSO (Coello et al., 2004)	0.0365	0.1095
NSGA-II	0.0842	0.0985
Micro-GA	0.1508	0.315
PAES	0.1932	0.1101

with a published overview according to Gill et al. (2006). The results are provided in Table 3. The metrics considered in this comparison are the generational distance (GD) metric introduced by van Veldhuizen and Lamont (1998) that measures the distance between the elements of the estimated non-dominated set and the known true front, and the spacing metric (SP) according to Schott (1995) that measures the mutual distance between the elements of the estimated non-dominated front. Lower values of GD and SP denote a better approximation and a more uniform spread, respectively, with zero being the optimum. We abstain from a more detailed introduction of these measures, as they are just used for this small comparison. The results show that all three suggested algorithms provide excellent results for the COELLO problem. The MO-PSO-GA approach outperforms existing PSO approaches and obtains results that are almost equivalent to the results of AMALGAM and AMALGAM*. In a second test, we consider a problem that was proposed and used as a benchmark by Vrugt et al. (2003). That is why we refer to this benchmark as VRUGT. It is originally defined for two dimensions with three objectives. The Pareto solution set consists

of a triangular-shaped area in the parameter space, having the corner points (0, 0), (0, 1) and (1, 0) for x_1 and x_2 , respectively. We defined an extended version that is defined for any dimension greater than or equal to two. It is defined by $N + 1$ objectives. In a first simple experiment, we estimated the Pareto set for the 2-dimensional case. A higher dimensional version with 30 dimensions is considered in a following case study that studies the advantages of deep parameter vectors. According to Vrugt et al. (2003), we set the population size to values of 10, 20, 50, and 100, but limited the number of iterations to 50. Thus, the maximum number of function evaluations is limited to 500, 1000, 2500 and 5000. The estimated parameters for the MO-PSO-GA and AMALGAM are given in Figs. 4 and 5. The results of AMALGAM* are just about equal and are thus neglected here. The results estimated by MO-PSO-GA and AMALGAM have a similar quality. A detailed visual comparison with published results estimated by MOSCEM (Vrugt et al., 2003) and two other multi-objective PSO implementations (Coello et al., 2004; Gill et al., 2006) confirms that the non-dominated front estimated by MO-PSO-GA is less clustered and provides a fairly better approximation of the true Pareto front. Already the run with population size 20 and a corresponding number of function evaluations of just 1000 provides better results. This underlines the efficiency of the MO-PSO-GA algorithm in comparison with other single-method search strategies.

The previous benchmarks underlined the efficiency of the MO-PSO-GA algorithm in comparison with other simple multi-objective optimisation. However, the test problems were still far too easy to notice any differences between the different suggested algorithms. That is why we tested the suggested solutions on another set of more complex and well-known multi-objective benchmark problems provided by Zitzler et al. (2000), Deb et al. (2002) and Fonseca and Fleming (1993) that are commonly used in literature (e.g. Vrugt and Robinson, 2007). An overview

Table 4. Description of complex multi-objective benchmark problems used in this study.

Problem	N	Parameter range	Objective functions	Characteristics
ZDT1	30	[0, 1]	$f_1(\mathbf{x}) = x_1$ $f_2(\mathbf{x}) = g(\mathbf{x})[1 - \sqrt{x_1/g(\mathbf{x})}]$ $g(\mathbf{x}) = 1 + 9[\sum_{i=1}^N x_i/(N - 1)]$	Convex
ZDT2	30	[0, 1]	$f_1(\mathbf{x}) = x_1$ $f_2(\mathbf{x}) = g(\mathbf{x})[1 - (x_1/g(\mathbf{x}))^2]$ $g(\mathbf{x}) = 1 + 9[\sum_{i=1}^N x_i/(N - 1)]$	Nonconvex
ZDT3	30	[0, 1]	$f_1(\mathbf{x}) = x_1$ $f_2(\mathbf{x}) = g(\mathbf{x})[1 - \sqrt{x_1/g(\mathbf{x})} - (x_1/g(\mathbf{x}))\sin(10\pi x_1)]$ $g(\mathbf{x}) = 1 + 9[\sum_{i=1}^N x_i/(N - 1)]$	Convex, disconnected
ZDT4	10	$x_1 \in [0, 1]$ $x_{2...N} \in [-5, -5]$	$f_1(\mathbf{x}) = x_1$ $f_2(\mathbf{x}) = g(\mathbf{x})[1 - (x_1/g(\mathbf{x}))^2]$ $g(\mathbf{x}) = 1 + 10(N - 1) + \sum_{i=1}^N (x_i - 10\cos(4\pi x_i))$	Nonconvex
ZDT6	10	[0, 1]	$f_1(\mathbf{x}) = 1 - e^{-4x_1} \cdot \sin^6(6\pi x_1)$ $f_2(\mathbf{x}) = 1 - \left(\frac{f_1(\mathbf{x})}{g(\mathbf{x})}\right)^2$ $g(\mathbf{x}) = 1 + 9\left[\sum_{i=1}^N x_i/(N - 1)\right]^{0.25}$	Nonconvex, nonuniformly spaced
FON _N	{3,10}	[-4, 4]	$f_1(\mathbf{x}) = 1 - e^{-\sum_{i=1}^N (x_i + 1/\sqrt{3})^2}$ $f_1(\mathbf{x}) = 1 - e^{-\sum_{i=1}^N (x_i - 1/\sqrt{3})^2}$ $f_1(\mathbf{y}) = y_1$ $f_2(\mathbf{y}) = g(\mathbf{y})e^{-y_1/g(\mathbf{y})}$	Nonconvex
ROT	10	[0, 1]	$g(\mathbf{y}) = 1 + 10(N - 1) + \sum_{i=2}^N y_i^2 - 10\cos(4\pi y_i)$ $\mathbf{y} = \mathcal{R}\mathbf{x}$	Rotated, correlated

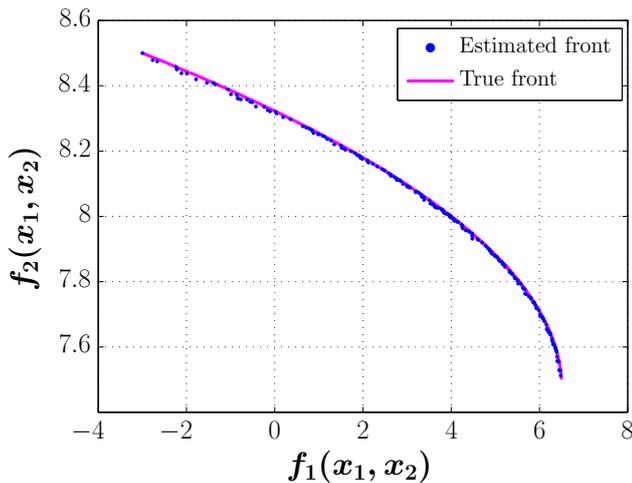


Fig. 3. True Pareto front and the non-dominated front estimated by MO-PSO-GA using Coello’s function.

over the used benchmarks is given in Table 4. Pre-generated, uniformly spaced, reference sets representing the known exact Pareto optimal sets for the given optimisation problems are provided by <http://www.tik.ee.ethz>.

ch/sop/download/supplementary/testproblems. We chose the versions with 1000 members each. According to the studies conducted in Vrugt and Robinson (2007), we applied each considered optimisation algorithm for the approximation of the Pareto optimal sets for the given test problems using a population size of 100 points in combination with 150 generations. Each optimisation run was repeated 30 times in order to avoid an interference of the results by outliers. In addition to the three considered algorithms MO-PSO-GA, AMALGAM and AMALGAM*, we also considered the NSGA-II algorithm according to Deb et al. (2002). It was subject of attention in many scientific publications and demonstrated superiority over many existing multi-objective optimisation methods. Furthermore, it already served as a reference in the first study with AMALGAM presented by Vrugt and Robinson (2007). Thus, it is a good indicator to assess the performance of the hybrid MO-PSO-GA algorithm.

The goal of a multi-objective optimisation algorithm is to approximate the Pareto optimal set of a given optimisation problem. The Pareto set in the parameter space corresponds to a Pareto front in the objective space. Thus, a good algorithm should estimate solutions that correspond to a non-dominated front that approximate the true Pareto front as close as possible. With the evolution of multi-objective

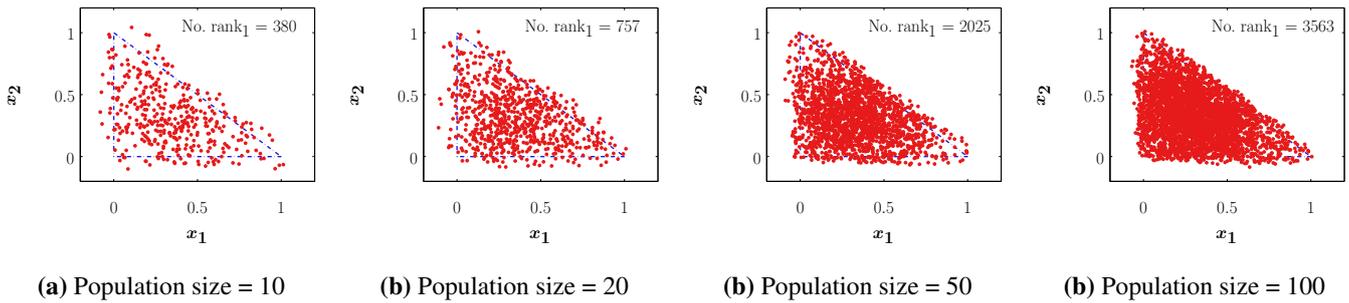


Fig. 4. Scatter plots for the rank 1 points (blue dots) for 50 evaluation generations estimated by MO-PSO-GA with a population size of (a) 10, (b) 20, (c) 50, and (d) 100 respectively. The boundaries of the triangle enclosing the true front are the blue dot-dashed blue lines.

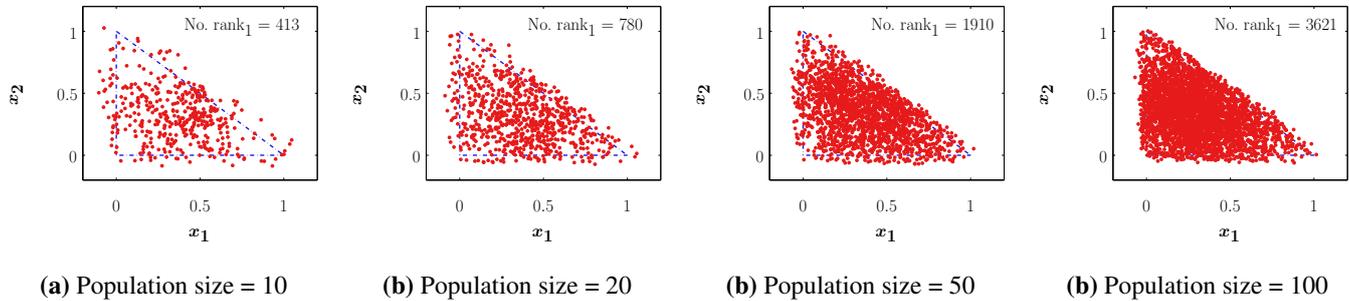


Fig. 5. Scatter plots for the rank 1 points (blue dots) for 50 evaluation generations estimated by AMALGAM with a population size of (a) 10, (b) 20, (c) 50, and (d) 100 respectively. The boundaries of the triangle enclosing the true front are the blue dot-dashed blue lines.

optimisation algorithms, a lot of metrics have emerged in order to quantify the effectivity and efficiency of compared approaches. We follow the suggestions of Vrugt and Robinson (2007) and use the following three metrics proposed in Deb (2001) and Deb et al. (2002) in order to measure the effectivity and efficiency of the compared optimisation algorithms.

The *convergence metric* (Y) measures the extent of convergence of the approximated set of non-dominated solutions \tilde{A} to the known true set of Pareto optimal solutions \tilde{P} . The metric Y computes by the average of the Euclidean distance of each point in \tilde{A} to its closest neighbour in \tilde{P} . The closer the value of Y is to zero, the better is the matching of the estimated solutions to the true Pareto optimal set. Consider that the distances are computed in the objective space.

Another important performance criterion for evolutionary algorithms for multi-objective optimization problems is the *diversity metric* Δ . It measures the extent of the spread or diversity of the estimated solutions along the true Pareto front. In order to calculate this metric, we compute the Euclidean distances between two consecutive solutions in \tilde{A} . These distances are stored in a vector \mathbf{d} with length $L - 1$ where L denotes the cardinality of the set \tilde{A} . In a second step, we calculate the Euclidean distances between the extrapolated extreme solutions of the true Pareto front and the boundary solutions of the obtained non-dominated set, d_f and d_l . Thus, d_f

denotes the distance between the first (leftmost) point in the obtained approximation and the beginning of the true front. The same applies for d_l and the last (rightmost) point of the obtained Pareto approximation. The metric Δ is a measure of the deviation of the individual distance between the consecutive solutions in \tilde{A} from its mean value weighted by the distance of the extreme tails of \tilde{A} from the tails of the true front \tilde{P} . Thus Δ is computed by

$$\Delta = \frac{d_f + d_l + \sum_{i=1}^{L-1} |d_i - \bar{d}|}{d_f + d_l + (L - 1)\bar{d}} \quad (3)$$

A perfect approximation of the true Pareto front would have a Δ value of zero. A perfect distribution, whose tails are ideally identical to those of the true front, would correspond to $d_f = d_l = 0$. Furthermore, such an approximation should be perfectly distributed so that all $d_i = \bar{d}$. This leads to $\Delta = 0$. Hence, smaller values for this metric indicate a better spread and uniform distribution along the front.

The *relative hypervolume indicator* (RHV) quantifies both convergence and diversity of the so far estimated approximation within a single measure, and is therefore one of the best unary measures available to diagnose whether the non-dominated solution set estimated by an optimisation algorithm is a good approximation of the true Pareto set. This metric is computed by one minus the ratio between the

objective space dominated by the estimated non-dominated set, and the objective space dominated by the given true Pareto set.

$$RHV = 1 - \frac{HV(\tilde{A})}{HV(\tilde{P})} \quad (4)$$

Thus, a perfect fit of the true front corresponds to a relative hypervolume indicator of zero. Following Vrugt and Robinson (2007), we use a relative hypervolume indicator with value of less than 0.005 as a criterion indicating that the estimated solution \tilde{A} is sufficiently close to the true Pareto set \tilde{P} , and counts the function evaluations carried out so far in order to measure the efficiency of the optimisation algorithms.

The effectivity of the considered optimisation algorithms in terms of the metrics Y and Δ is provided in Table 5. The obtained results for the existing algorithms AMALGAM and NSGA-II are close to the values published in literature (Vrugt and Robinson, 2007). This indicates that the framework was set up correctly. Of particular relevance are however the results of the proposed approaches MO-PSO-GA and AMALGAM*. Aside from the ROT problem, the MO-PSO-GA approach provides results that are comparable or better to those estimated by NSGA-II. Notwithstanding these convincing results, the achieved performances are one step worse than the results yielded by AMALGAM. However, the already excellent results of the AMALGAM approach can even be slightly improved by an extension with the MO-PSO-GA approach. For all test problems, the AMALGAM* provides the best results for all benchmark problems considered. An exception is again the ROT problem where the additional MO-PSO-GA approach does not provide any additional improvements. These results both confirm the superiority of a well-founded multi-method approach, such as the AMALGAM approach. Furthermore, the benefits of additional single or hybrid search strategies can be used in order to improve the results of advanced multi-method approaches. It also confirms Wolpert and Macready’s “no free lunch” theorem that shows that it is impossible to construct one single (parameter) search algorithm that will always outperform any other algorithm (Wolpert and Macready, 1997). Further research is required to estimate a perfect set of strategies in order to obtain a multi-method approach that yields the best possible results. According to the obtained results, we suggest to use the AMALGAM* approach for the estimation of the Pareto optimal set.

3.2 Studying the effectivity of depth-based sampling for multi-objective calibration problems that are subject to uncertainty

3.2.1 Uncertainty in synthetic test cases

In the previous case study, we studied different strategies that are suitable to approximate the Pareto set of a given multi-objective optimisation problem. This is a nec-

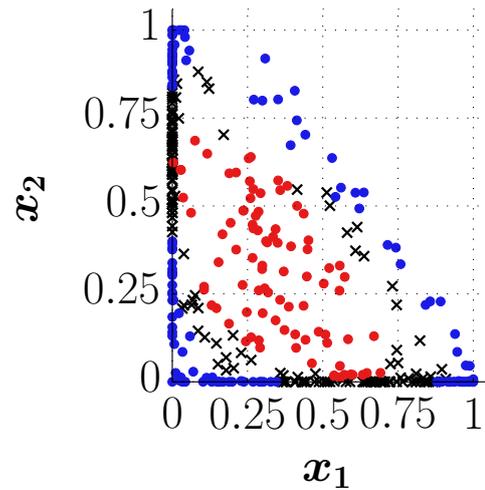


Fig. 6. Scatter plot of the approximated Pareto set for the problem VRUGT₂ in the 2-dimensional parameter space; the deep and the shallow solutions (hull) are indicated by red and blue dots.

essary and important step towards the identification of robust parameter vectors. The main thrust of the MO-ROPE approach is however the sampling of deep parameter vectors that are considered to be robust and thus a more representative solution in environments that are affected by a considerable degree of uncertainty. In such situations that are common for real-world problems, the Pareto set cannot be uniquely identified. The erroneous data can even lead to very different optimal parameter vectors. This issue is related to the requirement that robust parameter vectors should be as insensitive as possible. In this regard the estimated parameter vectors are considered as sensitive if a small change of the whole vector might lead to a big change in the performance of the model. To consider this issue, we just slightly altered the estimated parameter vectors and studied the results for the whole identified set, the deep parameters and the hull. The alteration was done using the following equation:

$$\mathbf{x} = \mathbf{x} + \epsilon_{\mathbf{x}} \quad (5)$$

where $\epsilon_{\mathbf{x}}$ is a multi-dimensional random independent Gaussian error with mean zero and a standard deviation that equals one percent of the range of the parameter in the estimated Pareto set in each considered dimension. Consider that this procedure alters the parameter vectors even less than in a similar experiment for sets of good parameters estimated by single-objective optimisation algorithms conducted by Bárdossy and Singh (2008).

Furthermore, we assumed that the actual values of the objective function cannot be exactly identified. That is due to errors in the observations. We consider this issue as follows. For a given multi-objective calibration problem $F = \{f_1 \dots f_m\}$, we defined a corresponding problem $F^u = \{f_1^u \dots f_m^u\}$, where every objective function $f_i^u \in F^u$ is exactly the same as its corresponding member $f_i \in F$ besides

Table 5. Effectivity of the four presented algorithms in terms of the performance metrics Y and Δ for the given more complex test benchmarks. The statistics represent mean values over 30 optimisation runs.

Problem	Convergence metric: Y			
	MO-PSO-GA	NSGA-II	AMALGAM	AMALGAM*
ZDT1	0.0024	0.0040	0.0009	0.0006
ZDT2	0.0174	0.0053	0.0007	0.0004
ZDT3	0.0016	0.0024	0.0012	0.0011
ZDT4	0.0148	0.0411	0.0028	0.0024
ZDT6	0.0161	0.0502	0.0008	0.0004
FON ₃				
FON ₁₀				
ROT	13.57	3.66	0.46	0.46
Problem	Diversity metric: Δ			
	MO-PSO-GA	NSGA	AMALGAM	AMALGAM*
ZDT1	0.40	0.41	0.35	0.34
ZDT2	0.46	0.45	0.33	0.33
ZDT3	0.60	0.61	0.44	0.43
ZDT4	0.75	0.81	0.41	0.40
ZDT6	0.57	0.56	0.47	0.45
FON ₃				
FON ₁₀				
ROT	1.62	1.10	0.65	0.67

Table 6. Efficiency of the four presented algorithms in terms of the number of model runs that are required to result in a relative hypervolume smaller than 0.005. The statistics represent mean values over 30 optimisation runs.

Problem	Number of function evaluations			
	MO-PSO-GA	NSGA	AMALGAM	AMALGAM*
ZDT1	1.576	6.124	756	704
ZDT2	8.624	7.152	1.148	988
ZDT3	4.404	4.956	1.252	1.040
ZDT4	9.376 (14)*	13.635 (16)*	6.472	5.916
ZDT6	6.736	8.780	960	874
FON ₃	4.710	4.930	1.510	1.380
FON ₁₀	4.710	4.930	1.510	1.380
ROT	N/C**	N/C**	9.527	9.874

* Number of runs that have failed to converge are shown in parentheses. ** None of the 30 optimisation runs have converged after 150 generations.

an additional uncertainty term. The uncertainty is subject to the following error model:

$$f_i^u(\mathbf{x}) = f_i(\mathbf{x}) + \epsilon_{f_i} \quad (6)$$

where ϵ_{f_i} is a random independent Gaussian error with constant variance. We assumed mean and standard deviation of just one percent of the range of the objective f_i in the true Pareto front of the problem original problem. Even though the defined error models are quite subjective, the resulting disturbances of both the objectives and the parameters are still well below the typical changes observed in hydrologic

model calibration and should not affect the performance to a too large extent.

We now conducted the following experiment in order to study the advantages of the deep parameter vectors and the effectivity of the depth-based sampling:

1. In a first step, we applied the MO-ROPE framework in order to estimate robust parameter vectors for a multi-objective calibration problem F . According to the results of the previously presented case study we chose the AMALGAM* for the approximation of the Pareto optimal set \hat{A} . Afterwards, we sampled a set of deep

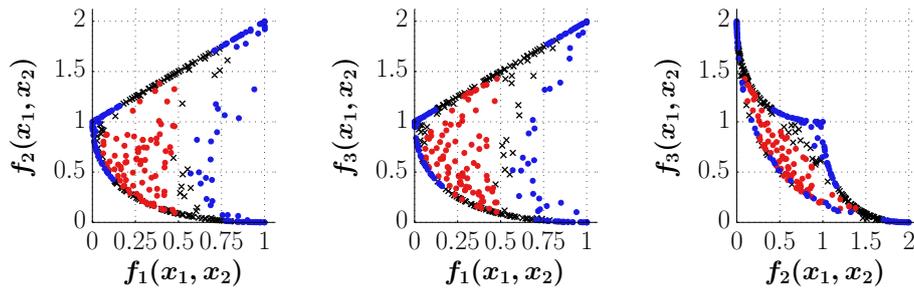


Fig. 7. Scatter plot of the approximated Pareto set for the problem VRUGT₂ in the 3-dimensional objective space; the deep and the shallow solutions (hull) are indicated by red and blue dots.

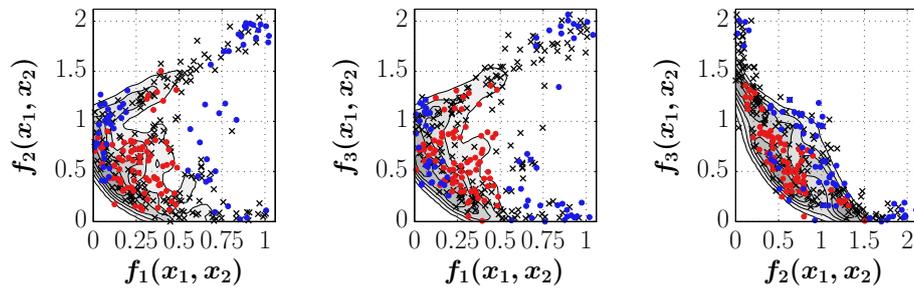


Fig. 8. Scatter plot indicating the distribution of the actual Pareto fronts for the problem VRUGT₂^u considering uncertainty. Scatter plot of the approximated Pareto set for the problem VRUGT₂ in the 3-dimensional objective space; the deep and the shallow solutions (hull) are indicated by red and blue dots.

parameter vectors $\mathcal{D}_{\tilde{A}}$. Additionally, we generated another set that comprises all vectors in \tilde{A} with shallow depth. We call it the hull of \tilde{A} and denote it by $\mathcal{H}_{\tilde{A}}$. Consider once again that the deep parameters are sampled in the parameter space and not in the objective space. This means that the depth-based sampling does not require any a priori knowledge of the used objectives.

- In a second step, we compared the performance of the complete estimated Pareto set and the sampled deep parameter vectors on the problem F_u . We generated 10 000 different realisations of the errors ϵ_{f_i} and ϵ_x . For each realisation, we altered the parameter vectors and identified the corresponding Pareto optimal set as a subset of \tilde{A} and $\mathcal{D}_{\tilde{A}}$. Now we compared the performance of the complete approximated Pareto set \tilde{A} , the sampled deep parameter vectors $\mathcal{D}_{\tilde{A}}$ and its hull $\mathcal{H}_{\tilde{A}}$ regarding their ability to represent the set of true Pareto sets for F_u .

We performed this experiment for the test problems VRUGT, FON and ROT in lower and higher dimensions that were already considered in the previous case study. For the problems VRUGT and FON, a subscript indicates the used number of dimensions. The Pareto set of the problem VRUGT is a multi-dimensional more or less convex region where the sampling of deep parameter vectors might be possible. The same applies, to a lesser extent, for the prob-

lems FON and ROT where the true Pareto set is a multi-dimensional line that is usually approximated by a cigar-shaped set. The same applies for reasonable models that simulate natural or technical processes where the Pareto optimal parameters also form such geometric structures (e.g. Bárdossy, 2007; Vrugt et al., 2003). We did not consider the problems ZDT1-ZDT6 where the corresponding Pareto set is a one-dimensional line with just one free parameter. In such cases, the sampling of deep parameters is not useful.

The principle of the depth-based sampling from the Pareto optimal sets is illustrated using the example of the problem VRUGT₂^u. The estimated non-dominated solutions \tilde{A} , the accordingly sampled deep parameter vectors and the hull in both parameter and objective space are provided by Figs. 6 and 7. The region defined by the deep parameter vectors corresponds to central parts of the Pareto front. However, there is no direct relationship to a simple cutoff value or even a direct one to one relationship that maps just deep parameter vectors to the central part of the Pareto front. This is in particular underlined by the distribution of the hull that corresponds to larger parts of the tails but also contains members in the central region of the Pareto front. The advantage of the depth-based sampling approach is illustrated by Fig. 8. It shows the density of the distribution of all possible Pareto fronts for VRUGT₂^u that were estimated for the different realisations of ϵ_f and ϵ_x . Furthermore, there is an overlay with scatter plots that represent the distribution of the estimated

Table 7. Effectivity of the estimated Pareto optimal set (\tilde{A}) and the corresponding deep ($\mathcal{D}_{\tilde{A}}$) and shallow ($\mathcal{H}_{\tilde{A}}$) parameter vectors in terms of the performance metrics Y and Δ , and the mean sum of all squared objective functions. The metrics were calculated with respect to the individually identified non-dominated front (as a subset of the originally approximation) for each realisation of the perturbances ϵ_x and ϵ_f . The statistics represent the mean values over 10 000 realisations.

Problem	Convergence metric: Y			Diversity metric: Δ			$\frac{1}{m} \sum_{i=1}^m (f_i^u)$		
	\tilde{A}	$\mathcal{D}_{\tilde{A}}$	$\mathcal{H}_{\tilde{A}}$	\tilde{A}	$\mathcal{D}_{\tilde{A}}$	$\mathcal{H}_{\tilde{A}}$	\tilde{A}	$\mathcal{D}_{\tilde{A}}$	$\mathcal{H}_{\tilde{A}}$
VRUGT ₂ ^u	0.12	0.07	0.17	0.73	0.60	0.88	2.10	1.10	2.79
VRUGT ₁₀ ^u	0.02	1E-6	0.05	0.37	0.29	0.30	21.4	12.4	30.6
VRUGT ₃₀ ^u	0.03	≈ 0	0.0514	0.35	0.34	0.34	89.3	40.2	114.0
FON ₃ ^u	0.015	0.010	0.016	0.65	0.59	0.63	0.84	0.81	0.87
FON ₁₀ ^u	0.11	0.09	0.12	1.13	1.04	1.11	1.37	1.34	1.39
ROT ^u	22.1	12.4	28.3	0.94	0.61	1.02	31225	17440	39951

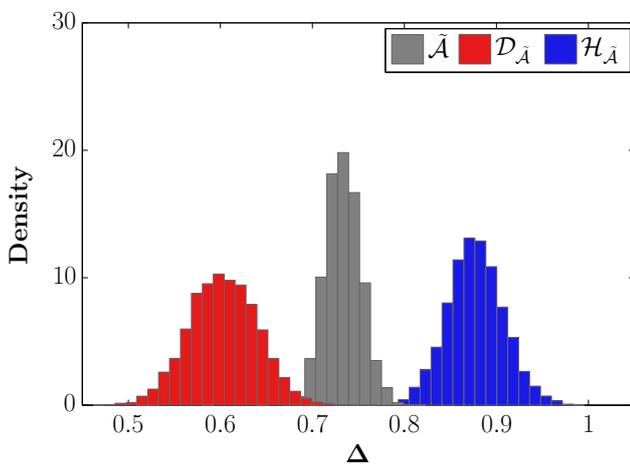


Fig. 9. Distribution of the diversity metric Δ for the approximated Pareto set for the problem VRUGT₂; its shallow solutions (hull) are indicated by red and blue bars.

approximation of the Pareto set of VRUGT (\tilde{A}), the corresponding deep parameter vectors ($\mathcal{D}_{\tilde{A}}$) and the hull ($\mathcal{H}_{\tilde{A}}$) in the objective space of the problem VRUGT₂^u. In comparison with the complete Pareto set, the deep parameter vectors do not just show a better convergence to the set of true Pareto fronts for the problem considering uncertainty. According to the distributions of the distribution metric Δ given in Fig. 9, they also provide a better mean approximation of the spread of the true Pareto fronts for the problem VRUGT₂^u.

The advantages of the depth-based sampling are not limited to simple low-dimensional problems. The results for all considered problems including uncertainty are quantified in Table 7. It compares both performance metrics used in the previous case, i.e. the convergence metric Y that measures the extent of convergence to the actual Pareto sets for F^u and the diversity metric Δ that provides information about the spread of the solutions along the Pareto front. Furthermore, it provides the mean sum of the squares of all individual objectives. For each realisation of the errors ϵ_x and

ϵ_f , the metrics were computed with respect to the individually identified non-dominated front (as a subset of the originally approximation \tilde{A}). For all given test problems, the sampled deep parameter vectors provide, in comparison with the complete estimated Pareto set, a better mean approximation of the actual Pareto sets in the objective space considering uncertainty regarding both the matching and a uniform distribution. Furthermore, the mean sum of the squares of all objectives is less for the deep parameter vectors. The effectivity of the depth-based sampling becomes even more objectionable if one considers the results for the parameter vectors with shallow depth, i.e. the hull. The parameters in the hull show the least suitability regarding its ability to represent a robust solution for the set of potential Pareto solution in an uncertain environment. Consider that the benefit of the depth-based sampling is greater for the problem VRUGT where the Pareto optimal set comprises a larger region. Typically, this is the case for hydrologic models. The results show that the depth-based sampling might be a suitable method to identify robust solutions with respect to a previously identified Pareto optimal set.

3.2.2 Uncertainty in the synthetic application of a hydrologic model

In the previous case study, we showed the advantages of the depth-based sampling in comparison with a pure multi-objective optimisation using synthetic benchmark experiments. One last experiment should establish that the MOROPE method really produces robust parameter vectors in the context of hydrologic model calibration. The experiment is done carrying out the following steps:

1. We chose a simple lumped hydrologic model and used observation data to generate synthetic discharge series.
2. Afterwards, we added noise to the observations and the synthetic discharge. The noisy data were split into a calibration and validation period in three different ways.

Table 8. Overview of the used model parameters considered for the synthetic hydrologic model calibration.

parameter	unit	lower/upper boundary	
a	[-]	0	1
k_1	[h]	0.01	20
k_2	[h]	20	60

- The model was now calibrated and validated for each realisation comparing AMALGAM and the proposed MO-ROPE approach.

The used observations were generated using the measurements in the Rietholzbach catchment. The used hydrologic model is described below. The model is a very simple lumped model consisting of two linear storage units. It takes both precipitation (prec) and potential evapotranspiration (etp) as input values and computes the corresponding discharge with two linear storage units. In a first step the precipitation and potential evapotranspiration are used to compute the effective precipitation as given in Eq. (7).

$$p_{\text{eff}}(t) = \max(0, \text{prec}(t) - \text{etp}(t)) \quad (7)$$

The effective precipitation p_{eff} is distributed to two linear storage units using the parameter a as defined in Eq. (8).

$$q_{\text{in}_1}(t) = a * p_{\text{eff}}(t) \quad (8)$$

$$q_{\text{in}_2}(t) = (1 - a) * p_{\text{eff}}(t) \quad (9)$$

The linear storage units are filled with the computed inflows. Each of both linear storage units generates discharge proportional to the stored water content V_w as defined in Eq. (10). The storage cons k describes how fast the storage unit drains the stored water content to the outlet.

$$q_{\text{out}}(t + 1) = \frac{V_w(t)}{k} \quad (10)$$

The total discharge is the sum of the two outflows q_{out_1} and q_{out_2} . Thus, the calibration problem consists of the three parameters a, k_1 and k_2 with the boundaries given in Table 8. The boundaries for the storage coefficients k_1 and k_2 account for the fact that one linear storage unit should represent the faster runoff components while the other one models the interflow and groundwater discharge. The initial storage content was fixed to $V_{w_1} = 0.015$ for the faster responding storage unit and $V_{w_2} = 5$ for the less responsive one.

The data basis was the meteorological observations for 24 historic flood events in the Rietholzbach catchment (for details, see Sect. 4). We combined the time periods in three different ways for calibration and validation sets as shown in Table 9. The observations were altered by adding Gaussian noise as given in Eq. (11). This was repeated 100 times resulting in 100 different possible observation time series.

$$\text{etp}_\epsilon(t) = \text{etp}(t) + \epsilon \text{etp}(t) \quad (11)$$

Table 9. Overview of the used combination of calibration and validation events.

Experiment	Calibration events	Validation events
S1	{ 12 }	{ 5, 6, 10, 17, 18, 20, 22 }
S2	{ 4, 12 }	{ 5, 6, 10, 17, 18, 20, 22 }
S3	{ 4, 12, 8 }	{ 5, 6, 10, 17, 18, 20, 22 }

with

$$\epsilon \sim \mathcal{N}(0, 0.5 \text{etp}) \quad (12)$$

Additionally, the generated synthetic discharge time series were altered as defined in Eq. (13).

$$q(t) = q(t)(1 + \epsilon q(t)) \quad (13)$$

with

$$\epsilon \sim \mathcal{N}(0, q) \quad (14)$$

The lumped model was now calibrated with respect to two objective functions: the relative deviation of simulated and observed peak runoff (rPD) and the Nash-Sutcliffe efficiency (NS). For details, see Table 13 in the following section. The mean validation results for both a pure multi-objective calibration using the AMALGAM algorithm and the corresponding deep parameter vectors are given in Tables 10 and 11. It is obvious that the deep parameters provide a slightly better model performance in mean and have less outliers on the negative side of the performance scale in terms of both considered performance criteria. This is also indicated by a slightly smaller standard deviation of the model performances. These results show that the advantages of the depth-based sampling approach apply not only for completely synthetic benchmarks but also for the calibration of a simple hydrologic model in a controlled environment. This underlines the possibilities of the depth-based sampling for the estimation of robust hydrologic model parameters.

4 Calibrating WaSiM with MO-ROPE focussing on flood forecasting

4.1 Case study area and the hydrologic model

In a real world application, the MO-ROPE approach is tested on the calibration of the distributed hydrologic model WaSiM-ETH /6.4 model (further referred to as WaSiM). he model was developed by Schulla (1997). WaSiM transforms rainfall into runoff according to the scheme shown in Fig. 10. It exemplarily shows that the soil water compartments receive infiltration which is computed by a modified approach according to Green and Ampt (1911). This module is also used to determine the direct runoff Q_d and the interflow Q_{if} in

Table 10. Validation performance of the solutions estimated by AMALGAM and MO-ROPE in terms of the Nash-Sutcliffe efficiency (NS).

Experiment	AMALGAM			MO-ROPE		
	Mean	Std	Range	Mean	Std	Range
S1	0.95	0.28	0.88–0.99	0.96	0.21	0.91–0.99
S2	0.97	0.21	0.91–0.99	0.98	0.13	0.93–0.99
S3	0.98	0.12	0.93–0.99	0.98	0.08	0.96–0.99

Table 11. Validation performance of the solutions estimated by AMALGAM and MO-ROPE in terms of the relative peak deviation (rPD).

Experiment	AMALGAM			MO-ROPE		
	Mean	Std	Range	Mean	Std	Range
S1	0.04	0.08	0.02–0.16	0.03	0.06	0.02–0.13
S2	0.03	0.06	0.02–0.14	0.02	0.04	0.02–0.11
S3	0.03	0.05	0.02–0.10	0.02	0.03	0.02–0.07

the model. Q_d is then routed via a flow-time grid and finally projected cell-wise to the catchment outlet by means of a simple bucket-type function. The recession coefficient of this function is the model parameter k_d . The soil water movement through the different soil layers is modelled by means of the discrete form of the Richards equation. The WaSiM model was calibrated for the Rietholzbach catchment, a small pre-alpine catchment located in the north-east of Switzerland. A significant number of hydrologic studies have been conducted in this basin. It has been observed as a research catchment by the ETH Zurich since 1975. The outlet drains a 3.18 km² hilly pre-alpine watershed with an average annual precipitation of 1600 mm, generating a mean runoff of 1046 mm per year. For further information, refer to Gurtz et al. (1999); Zappa (2002) and the website <http://www.iac.ethz.ch/research/rietholzbach>. In this case study, WaSiM will be calibrated for the simulation of extreme discharges. Out of a time series of 27 yr (1981–2008) of hourly measurements (precipitation, temperature, wind speed, global radiation, and streamflow), 24 significant flood events were identified, such that the whole range of possible flood characteristics occurring in the catchment is well covered. An exception was made in that we did not use any winter events to avoid a blurring of the results due to the dynamics of snow accumulation and melt. For further details and a more comprehensive overview, refer to Krauß and Cullmann (2011a).

The WaSiM model has been chosen in this study because its physically based unsaturated zone module maintains the characteristic physics of dynamic rainfall-runoff processes even for unobserved events. This is especially important for correctly portraying the pre-event catchment conditions. “WaSiM is therefore – amongst the available models – one of the best suited for extrapolation into the range of extreme flood events” (Cullmann, 2006, p. 20). That is why it has been widely used for the modelling of flood

events (e.g. Marx, 2007; Cullmann et al., 2008; Grundmann, 2010). We performed this case study in the Rietholzbach catchment because it has a long record of hourly datasets and the perturbing impact of data heterogeneity is relatively small in this catchment. Furthermore, the WaSiM model has been thoroughly tested within this catchment (Schulla, 1997). Notwithstanding these advantages, the modelling of flood events in such a small catchment is a challenging task. Typically, the achievable model performance is just moderate and the modelling process is subject to many uncertainties that can hardly be quantified. However, the results of previous studies dealing with flood forecasting suggest that an improved parametrisation of headwater catchments can have a big impact on the model performance for gauging stations in the lower reaches that are usually monitored by operational flood forecasting centres (e.g. Cullmann, 2006; Grundmann, 2010). Many studies dealing with model calibration focussing on flood events showed that there is a tremendous tradeoff between a correct modelling of the peak flow values and a good representation of the catchment behaviour in terms of the streamflow at the outlet (e.g. Moussa and Chahinian, 2009; Grundmann, 2010). The goal of this case study is thus to study the advantages of the developed MO-ROPE strategy for the calibration of a process-oriented hydrologic model focussing on flood events with respect to multiple objectives.

4.2 Calibrating WaSiM for flood events using two objectives

In a first case study, we calibrated WaSiM using two objectives. The relative peak flow deviation (rPD) quantifies the agreement between observed and simulated peak flow value, whereas the coefficient of efficiency according to Nash and Sutcliffe (1970) (NS) assesses the global fit of the observed and simulated hydrograph. According to our experience with

Table 12. Overview of the used model parameters considered for calibration; the reference parameter vector θ_{wb} was estimated in order to use WaSiM for water-balance simulations in the Rietholzbach catchment; the parameterisation of the soil hydraulic parameters is done for each soil according to the pedotransfer functions provided in Wösten et al. (1999) and Brakensiek et al. (1984).

Parameter	Unit	Reference (θ_{wb})	Lower/Upper boundary		Description
k_d	[h]	7	0.01	25	storage coefficient of direct runoff
k_i	[h]	20	1	60	storage coefficient of interflow
dr	[-]	2.1	1	80	drainage density
k_{rec}	[-]	0.1	0.01	1	gradient of k_s with increasing depth
$\beta_{SL} = \left\{ \begin{array}{l} k_s \\ \Theta_s \\ \Theta_r \\ \alpha \\ n \end{array} \right.$	k_s [ms ⁻¹]	$2.22e-5$	$3.19e-6$	$1.32e-4$	saturated hydraulic conductivity
	Θ_s [-]	0.412	0.38	0.46	saturation water content
	Θ_r [-]	0.01	0.01	0.01	residual water content
	α [-]	4.60	1.62	8.97	empirical shape parameter (MVG)
	n [-]	1.29	1.18	1.45	empirical shape parameter (MVG)
$\beta_{SiL} = \left\{ \begin{array}{l} k_s \\ \Theta_s \\ \Theta_r \\ \alpha \\ n \end{array} \right.$	k_s [ms ⁻¹]	$7.12e-7$	$1.07e-7$	$2.95e-6$	saturated hydraulic conductivity
	Θ_s [-]	0.42	0.41	0.46	saturation water content
	Θ_r [-]	0.01	0.01	0.01	residual water content
	α [-]	1.36	0.67	2.17	empirical shape parameter (MVG)
	n [-]	1.26	1.14	1.46	empirical shape parameter (MVG)

Table 13. Objective functions used in this study, where q_{x_i} , $q_{y_i}(\theta)$, and Θ_{x_i} , $\Theta_{y_i}(\theta)$ are the observed and simulated discharge and mean soil moisture at time step i , respectively. The simulated values are computed by the parameter vector θ . n denotes the number of observations.

Name	Description	Formula
NS	Nash-Sutcliffe efficiency	$1 - \frac{\frac{1}{n} \sum_{i=1}^n (q_{x_i} - q_{y_i}(\theta))^2}{\frac{1}{n} \sum_{i=1}^n (q_{x_i} - \bar{q}_x)^2}$
rPD	rel. peak flow deviation	$\frac{ q_{x_{max}} - q_{y_{max}}(\theta) }{q_{x_{max}}}$
FloodSkill	aggregate between NS and rPD	NS-rPD
r_{Θ}	moisture correlation coefficient	$\frac{\sum_{i=1}^n (\Theta_{x_i} - \bar{\Theta}_x) \cdot (\Theta_{y_i} - \bar{\Theta}_y)}{\sqrt{\sum_{i=1}^n (\Theta_{x_i} - \bar{\Theta}_x)^2 \cdot \sum_{i=1}^n (\Theta_{y_i} - \bar{\Theta}_y)^2}}$

WaSiM, both criteria have a small correlation and are most suitable to quantify the model performance focussing on flood events. In a further case study, we additionally considered the correlation between the simulated soil moisture and the observed soil moisture in terms of the lysimeter weight in the catchment. An overview of objective criteria referred to in the following case studies is provided in Table 13. The model parameters considered for calibration are the conceptual model parameters k_d , k_i , “dr” and k_{rec} . k_d and k_i are storage coefficients that control the outflow of linear storage events that collect the generated direct runoff and interflow, whereas “dr” is a scalar value that controls the generation of interflow. The conceptual soil parameter k_{rec} determines the decrease of the saturated soil conductivity with increasing depth. Furthermore, we considered the soil hydraulic parameters of the dominating soils SL and SiL as

additional calibration parameters. A reasonable a priori estimation of the distribution of the soil hydraulic parameters and a scaling of these values to one scaling parameter per soil (β_{SL} and β_{SiL}) were done using an approach presented in Grundmann (2010). Previous studies with WaSiM focussing on flood events (e.g. Cullmann, 2006; Grundmann, 2010) showed that the selected set of model parameters is the most sensitive and suitable calibrating WaSiM for the modelling of flood events in fast responding catchments. These results were confirmed by preliminary sensitivity analyses with WaSiM in the Rietholzbach catchment (Cullmann, 2006; Seifert, 2010). All calibration parameters with their a priori distribution are clearly summarised in Table 12.

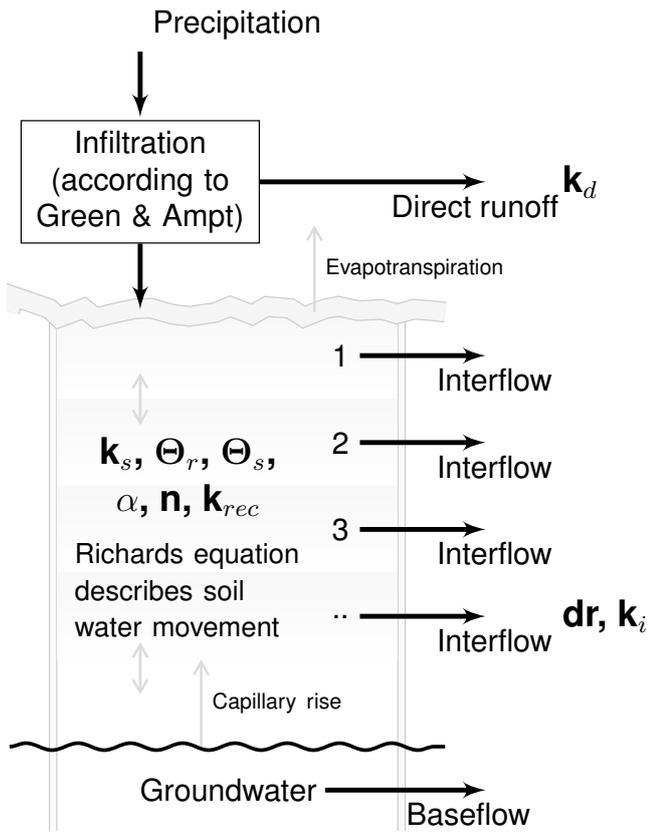


Fig. 10. Scheme of the WaSiM soil module with location of impact of soil hydraulic and conceptual model parameters (bold).

We used five flood events with different characteristics⁵ for calibration, and the estimated parameter vectors were validated on further 19 flood events.

We applied MO-ROPE using the given setup in order to estimate a set of robust model parameter vectors for the modelling of flood events using the two objectives rPD and NS. The Pareto set was estimated using the extended AMALGAM* framework. We checked the convergence of the algorithm using the development of the hypervolume dominated by the so far estimated non-dominated Pareto front in the consecutive generations as given in Fig. 12. It is evident that the hypervolume reaches stationary after approximately 80 generations. With a population size of 50, this corresponds to 4000 function evaluations. We decided to consider an additional buffer and thus set the maximum number of function evaluations to 6000. The trade-off surface for both objectives rPD and NS on the basis of all parameter vectors evaluated by AMALGAM* and the estimated approximation of the Pareto optimal set \tilde{A} is given in Fig. 11. The evaluation of the model performances corresponding to the

⁵We considered flood events caused by convective and advective precipitation, different peak flow values and events with single and multiple flood crests.

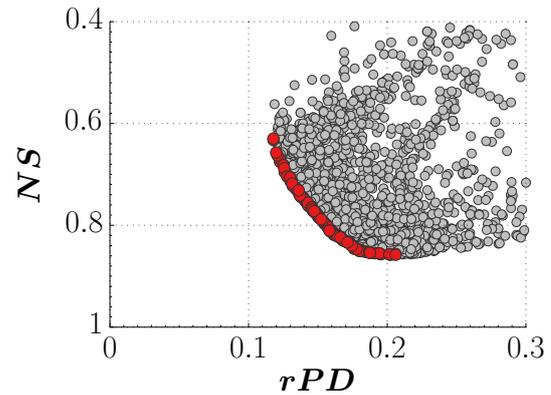


Fig. 11. Evolution of the trade-off surface of the WaSiM parameter vectors in the two-dimensional objective space. The grey dots represent the whole set of parameter vectors that were evaluated during the approximation of the Pareto optimal set \tilde{A} , indicated by red dots.

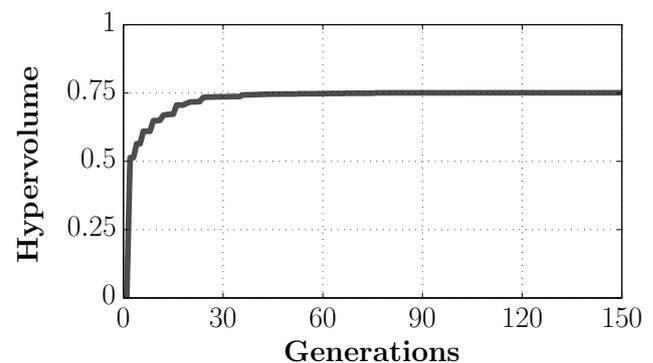
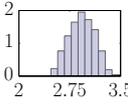
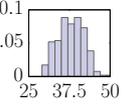
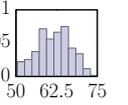
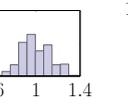
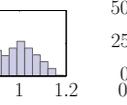
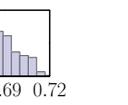
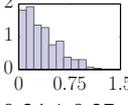
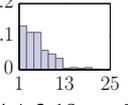
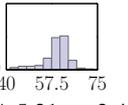
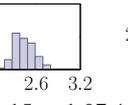
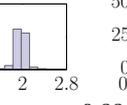
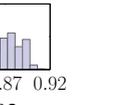
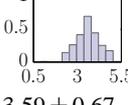
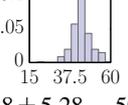
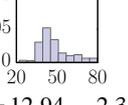
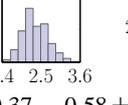
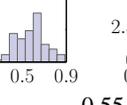
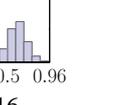
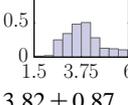
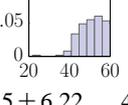
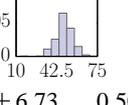
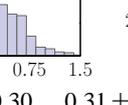
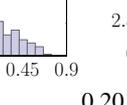
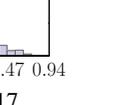


Fig. 12. Illustration of the development of the hypervolume of the estimated non-dominated front as a function of the generation number; the hypervolume was calculated with respect to the point [1, 0] which corresponds to 100 % peak flow deviation and a Nash value of zero.

estimated parameter vectors shows a clear tradeoff between both used objectives. That means that the best parameter vectors with respect to the peak flow value cannot represent the global behaviour of the catchment for flood events equally well. The following final step, the sampling of deep parameter vectors, was done using the GenDeep strategy. We compared the results of the estimated approximation of Pareto set \tilde{A} and the deep parameter vectors $\mathcal{D}_{\tilde{A}}$. Additionally, we compared this set with the parameter vectors determined by the single-objective robust parameter estimation algorithm ROPE-PSO, using both the singular objectives rPD and NS, and an aggregated criterion, called FloodSkill (see Table 13), as objective functions.

The estimated Pareto set and the sampled deep parameter vectors in the parameter space are provided in Fig. 13. The Pareto set is almost convex which is advantageous for the depth-based sampling. The deep parameter vectors cover the central region of the set, and its convex hull is indicative

Table 14. Distribution of the robust parameter vectors of the WaSiM model estimated by MO-ROPE and the three different single-objective ROPE-PSO runs.

	k_d	k_i	dr	β_{SL}	β_{SiL}	k_{rec}
Initial range	0.01–25	1–60	1–80	0.06–16	0.04–9.5	0.01–1
MO-ROPE						
Mean ± Std	2.93 ± 0.19	37.9 ± 3.94	61.7 ± 5.06	1.00 ± 0.14	0.99 ± 0.08	0.69 ± 0.01
Range	[2.47–3.47]	[29.1–49.0]	[50.4–72.9]	[0.61–1.37]	[0.82–1.16]	[0.66–0.72]
ROPE-PSO(rPD)						
Mean ± Std	0.34 ± 0.27	5.4 ± 3.18	59.2 ± 5.21	2.44 ± 0.15	1.97 ± 0.15	0.88 ± 0.02
Range	[0.01–1.33]	[1.1–15.4]	[40.9–71.0]	[1.84–2.93]	[0.69–2.50]	[0.83–0.92]
ROPE-PSO(FloodSkill)						
Mean ± Std	3.59 ± 0.67	43.8 ± 5.28	50.8 ± 12.94	2.34 ± 0.37	0.58 ± 0.12	0.55 ± 0.16
Range	[0.88–5.03]	[19.4–59.9]	[22.0–77.4]	[1.43–3.54]	[0.16–0.89]	[0.05–0.87]
ROPE-PSO(NS)						
Mean ± Std	3.82 ± 0.87	50.5 ± 6.22	48.5 ± 6.73	0.50 ± 0.30	0.31 ± 0.17	0.20 ± 0.17
Range	[2.00–5.96]	[27.9–60.0]	[30.6–73.4]	[0.08–1.43]	[0.01–0.88]	[0.01–0.72]

of the shape of the complete set. A comparison of the estimated sets of robust parameter vectors for both the MO-ROPE run as well as the single-objective ROPE-PSO is given in Table 14. It provides the distributions of the estimated robust parameter vectors and some basic statistical properties, i.e. the mean value, the standard deviation and their range. Obviously, there is an strong dependence of the parameters k_d and k_{rec} on the used performance criterion. The more a correct representation of the observed peak flow value is measured by the used objective criterion, the lower the value of k_d and the higher the value of k_{rec} . In general these results are reasonable and consistent with the model structure and the corresponding understanding of the hydrologic system. Lower values of k_d increase the dynamics of the generated direct runoff; a higher value of k_{rec} decreases the effective saturated conductivity of deeper soil layers. This leads to a faster generation of direct runoff. Within the Rietholzbach catchment, direct runoff on the surface has hardly ever been observed, not even during large and intensive convective storm events. However, there are many underground pipes that drain the slopes. The effect of these drainage system becomes significantly large during times of intensive precipitation. The estimated parameters indicate that the drainage flow can be effectively represented by the direct runoff component in

WaSiM. In terms of the spread of the parameters, there exists an in general strong relationship between the used performance criterion and the spread of the parameters. Typically, the variance of the estimated parameter distributions is smaller the more the used performance criterion quantifies the peak flow deviation. As explained above, it is necessary to force the direct runoff storage coefficient k_d to extremely low values in order to represent the fast runoff components with very high dynamics. As a consequence, most of the remaining fast and intermediate runoff components have to be represented by the interflow component. Hence, the storage coefficient of the interflow k_i is also forced to relatively small values, the value of the drainage “dr” forced to higher values (more interflow generation), and the soils adjusted more conductively (higher values of the scaling parameters β_{SL} and β_{SiL}) in all soil layers (higher values of k_{rec}). Furthermore, the variance of these parameters becomes smaller as several flow components with different temporal dynamics have to be represented by just one component of the interflow. On the other hand, the estimates for the efficiency criterion NS show in general a larger spread. The NS criterion quantifies the agreement of the observed and simulated catchment behaviour in a global way with an overweighting of the time steps with higher streamflow. A calibration with this

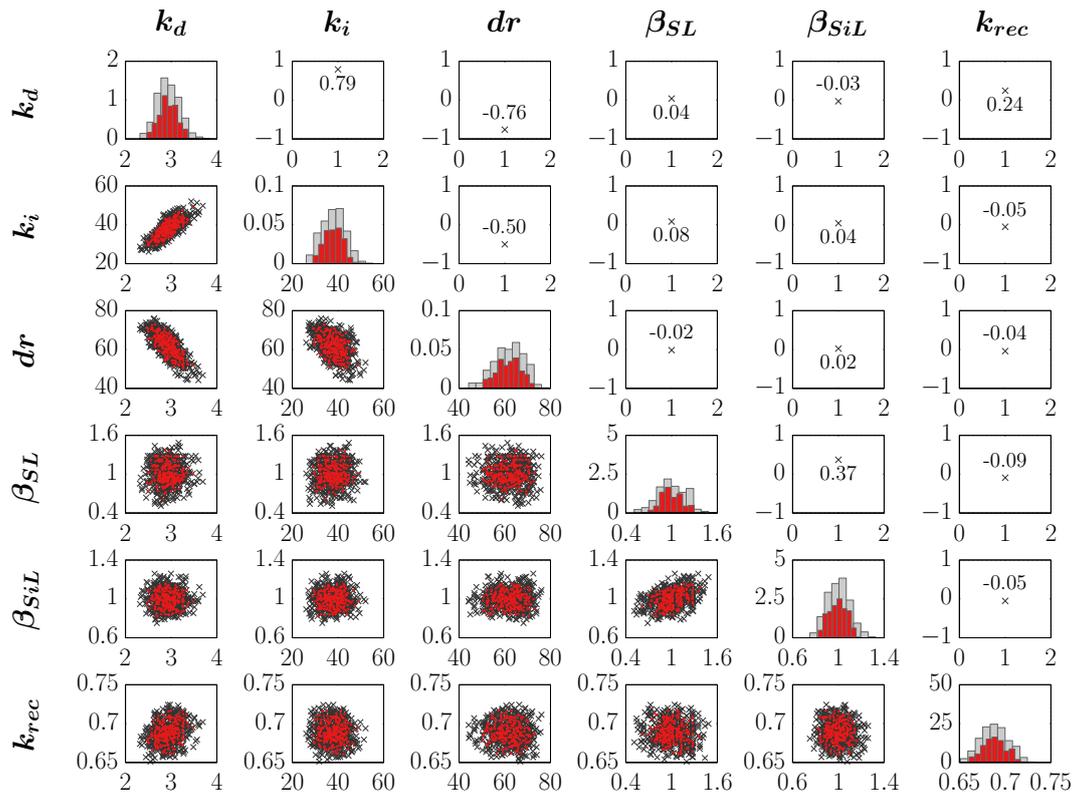


Fig. 13. Distribution of the Pareto optimal set estimated by MO-ROPE (applying the AMALGAM* strategy) and the sampled deep parameter vectors for the WaSiM model using two objectives (rPD and NS).

criterion does not necessarily force the model to its limits and can neglect a part of the fast runoff components in order to achieve a high NS value. The corresponding peak flow deviations are higher, but the flow components can be represented by more model components and as a consequence the variance of the estimated parameter distributions becomes larger. The parameter distributions estimated by the multi-objective MO-ROPE approach have mean values that are in between the mean values of the single objective calibrations using the rPD and the aggregated compromise criterion FloodSkill. In comparison with the single-objective calibration runs, the use of the multi-objective MO-ROPE approach obtains tighter variation intervals. These results confirm the outcome of previous studies dealing with the multi-objective calibration of conceptual hydrologic models focussing on flood events (e.g. Engeland et al., 2006; Moussa and Chahinian, 2009).

A comparison of the complete Pareto set $\tilde{\mathcal{A}}$, its hull $\mathcal{H}_{\tilde{\mathcal{A}}}$ (thus the points with shallow depth) and the sampled deep parameter vectors $\mathcal{D}_{\tilde{\mathcal{A}}}$ in the objective space for both the calibration and the validation data is provided in Fig. 14a and b. These figures illustrate the advantage of the depth-based sampling. In the objective space for the calibration data, the sampled deep parameter vectors are obviously concentrated in a central part of the Pareto front, whereas the parameter vectors in the hull are distributed over all parts of the Pareto

front with high density in the area of the tails and low density in the central part of the front. The sampled deep parameter vectors show a better performance on the validation data than the complete Pareto set. The deep parameter vectors have less outliers with a worse performance and are a good approximation of the (theoretical) Pareto front in the objective space based on the validation data. The distribution of the deep parameter vectors suggests that the tails of the Pareto front estimated in the calibration are not required for a robust set of model parameter vectors. For example, the best parameter vectors with respect to rPD on the calibration data do not have a better rPD in the validation than the sampled deep parameter vectors. However, these vectors correspond to clearly worse NS values. This shows that the deep parameter vectors are better transferable to other periods and events and thus more robust. Notwithstanding the fact that several shallow parameter vectors also correspond to central parts of the Pareto front, we compared the deep parameter vectors with the results of an approach using subjective cutoff thresholds to select “behavioural” and reject “non-behavioural” solutions as proposed by Efstratiadis and Koutsoyiannis (2010). We simply used cutoff thresholds defined by the boundaries of the model performances of the deep parameter vectors on the calibration data. The result is given in Fig. 14c. It is evident that the behavioural parameter

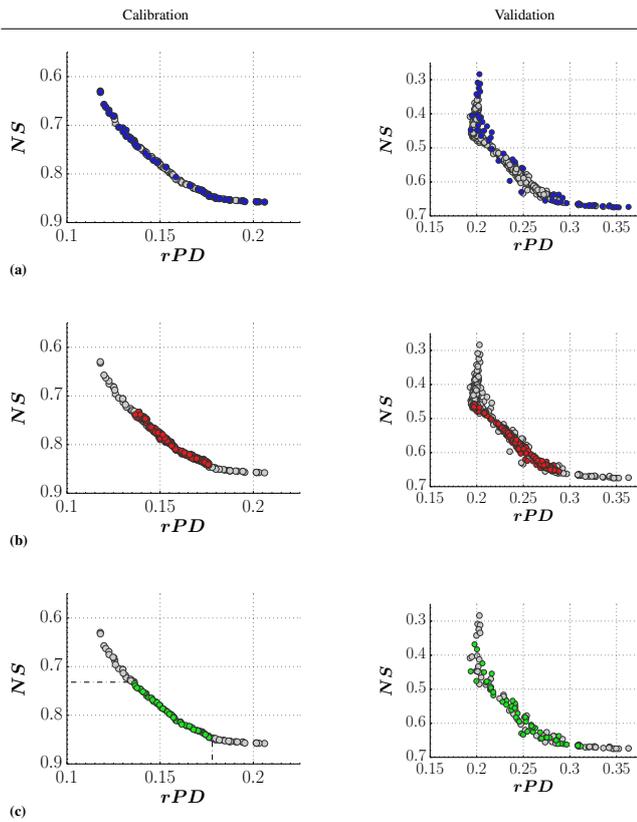


Fig. 14. Comparison of the identified approximation of the Pareto optimal set $\tilde{\mathcal{A}}$ (grey dots) in the objective space spanned by the two objectives rPD and NS. The plots in the left column are computed on the calibration events, whereas the right column shows the results on the validation data. Subplot (a) additionally emphasises the parameter vectors in the hull $\mathcal{H}_{\tilde{\mathcal{A}}}$ (indicated by blue dots); subplot (b) shows the sampled deep parameter vectors (indicated by red dots), and subplot (c) provides the results for a simple cut-off procedure with threshold values defined by the performance of the deep parameter vectors on the calibration data. The selected parameter vectors are indicated by green dots.

vectors identified by the cut-off thresholds are better suitable to approximate the (theoretical) Pareto front in the objective space based on the validation data than the complete Pareto set identified for the calibration events. Nonetheless, the deep parameter vectors provide a closer, denser and more uniform approximation of the Pareto front on the validation data. Moreover, the depth-based sampling does not require any assumptions regarding the threshold values⁶ and additionally provide the possibility to assign likelihood values to the parameter vectors according to their data depth. The advantage of the depth-based sampling is illustrated again in Fig. 15. It visualises the dependency between the model performance of a sampled parameter vector on the validation data in terms

⁶In this study the threshold values were only defined by the help of the deep parameter vectors.

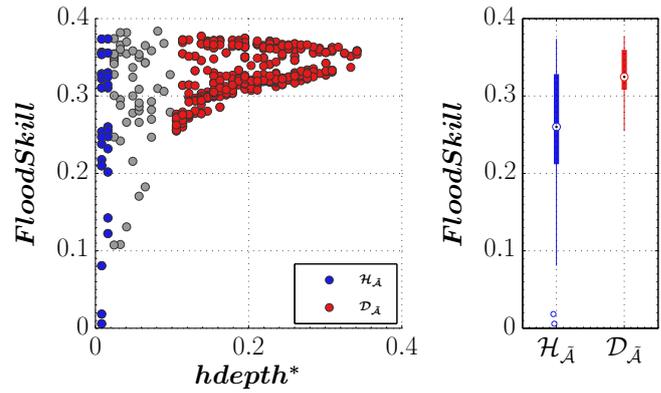


Fig. 15. Correlation between data depth and overall validation performance in terms of the FloodSkill criterion of all parameter vectors in the approximated Pareto set $\tilde{\mathcal{A}}$ estimated by the MO-ROPE calibration using the objectives rPD and NS. The blue dots indicate the hull $\mathcal{H}_{\tilde{\mathcal{A}}}$, i.e. the parameter vectors with shallow data depth; the red dots indicate the ones with high data depth $\mathcal{D}_{\tilde{\mathcal{A}}}$, and the grey ones are members of the complete Pareto set that have a depth less than the deep ones and are not members of the hull.

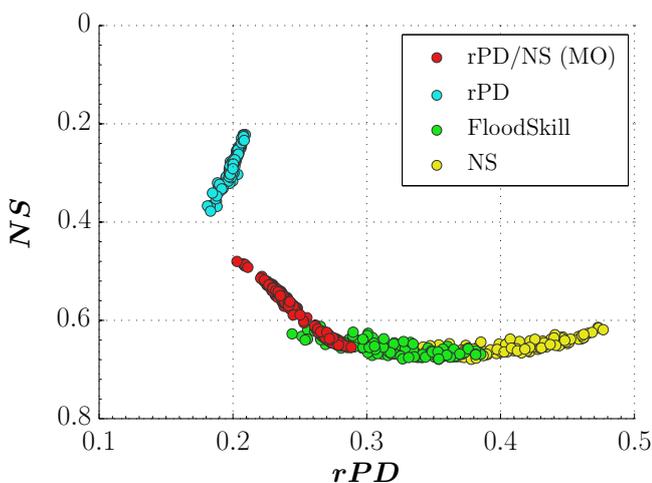
of the FloodSkill criterion⁷ and its data depth with respect to the approximated Pareto set $\tilde{\mathcal{A}}$. The parameter vectors with high data depth show a on average better model performance (higher FloodSkill values) with less variance than the parameter vectors with low data depth. This implies that parameter vectors with low data depth are more likely to be an outlier with bad model performance. The estimated results confirm the underlying assumption of the ROPE approach for multi-objective calibration tasks.

The model performances of the parameter vectors on the validation events estimated by both the multi-objective MO-ROPE and the single-objective robust parameter estimation runs are provided in Fig. 16. These results once again illustrate the evident tradeoff between the two criteria rPD and NS. The parameter vectors estimated by the single-objective ROPE-PSO considering just one criterion (rPD or NS) show indeed a good validation performance on those criteria. However, the validation performance in terms of the complementary performance criterion that is not considered for calibration is not sufficient for a robust modelling of flood events. For instance, the NS values for the ROPE-PSO estimates using the rPD as calibration objective are in the range of approximately 0.2–0.4, which is not acceptable, whereas the NS estimates correspond to peak flow deviations of 30–50 %, which is not sufficient either. Also the aggregated FloodSkill criterion just slightly improves the results. It still overweighs the NS criteria, and the parameter vectors estimated using the FloodSkill criterion do not fully exploit the model’s abilities

⁷The FloodSkill is a criterion aggregated by rPD and NS that represents a compromise between these two criteria (cf. Table 13 at the beginning of this section).

Table 15. Distribution of the robust parameter vectors of the WaSiM model estimated by MO-ROPE using the 2-objective (a) and 3-objective calibration (b).

(a)									
	Mean	Std	CV	k_d	k_i	dr	β_{SL}	β_{SiL}	k_{rec}
k_d	2.93	0.19	0.07	1.00	0.78	-0.72	-0.04	-0.06	0.29
k_i	37.9	3.94	0.10	...	1.00	-0.43	0.06	0.04	0.01
dr	61.7	5.06	0.08	1.00	0.06	0.01	-0.07
β_{SL}	1.00	0.14	0.14	1.00	0.38	-0.09
β_{SiL}	0.99	0.08	0.08	1.00	-0.04
k_{rec}	0.69	0.01	0.02	1.00
(b)									
	Mean	Std	CV	k_d	k_i	dr	β_{SL}	β_{SiL}	k_{rec}
k_d	2.88	0.20	0.07	1.00	0.83	-0.68	-0.02	0.11	-0.38
k_i	42.5	5.33	0.13	...	1.00	-0.52	0.01	0.13	-0.38
dr	60.6	3.43	0.06	1.00	0.08	-0.07	0.39
β_{SL}	1.01	0.08	0.08	1.00	0.39	-0.01
β_{SiL}	1.01	0.05	0.05	1.00	0.00
k_{rec}	0.23	0.04	0.16	1.00

**Fig. 16.** Comparison of the robust model parameter vectors estimated by the multi-objective MO-ROPE using two objectives (rPD/NS) and the single-objective robust parameter estimation algorithm ROPE-PSO using the objectives rPD, FloodSkill and NS, respectively. The plot shows the estimated solution in the objective space based on the validation events.

in terms of the peak flow deviation. The MO-ROPE results however provide a good model performance on the validation data with respect to a good global representation of the catchment behaviour and a good modelling of the peak flow values. This confirms the assumption that the use of an advanced multi-objective calibration technique can improve the performance of hydrologic models used for the simulation of flood events. The corresponding hydrographs

of three validation events and the corresponding parameter and model uncertainties for both the robust multi-objective estimates and the single-objective estimates, using the aggregated FloodSkill criterion as a kind of compromise solution, are given in Fig. 17. The complete model uncertainty was computed by two normal distributions 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). Contrary to Engeland et al. (2010), we considered observations with a discharge greater than or equal to 0.33 mm h^{-1} to account for the focus on the correct simulation of higher flow values. The hydrographs confirm the results of the previous analysis of the model performances. The peak flow values are better represented by the parameter vectors estimated by the multi-objective calibration. The confidence band of the parameter uncertainty is slightly smaller for the multi-objective calibration than for the single-objective calibration. The variation intervals of the resulting complete model uncertainty are approximately the same for both approaches. The remaining uncertainty is due to the discussed shortcomings of the model structure and other neglected uncertainties in the observations.

4.3 Calibrating WaSiM for flood events using three objectives describing the fit with both observed hydrograph and soil moisture measurements

In another case study, we calibrated WaSiM again. However, this time we accounted for the provided information of the soil moisture in the catchment for model calibration as an additional calibration objective. Recent studies showed that

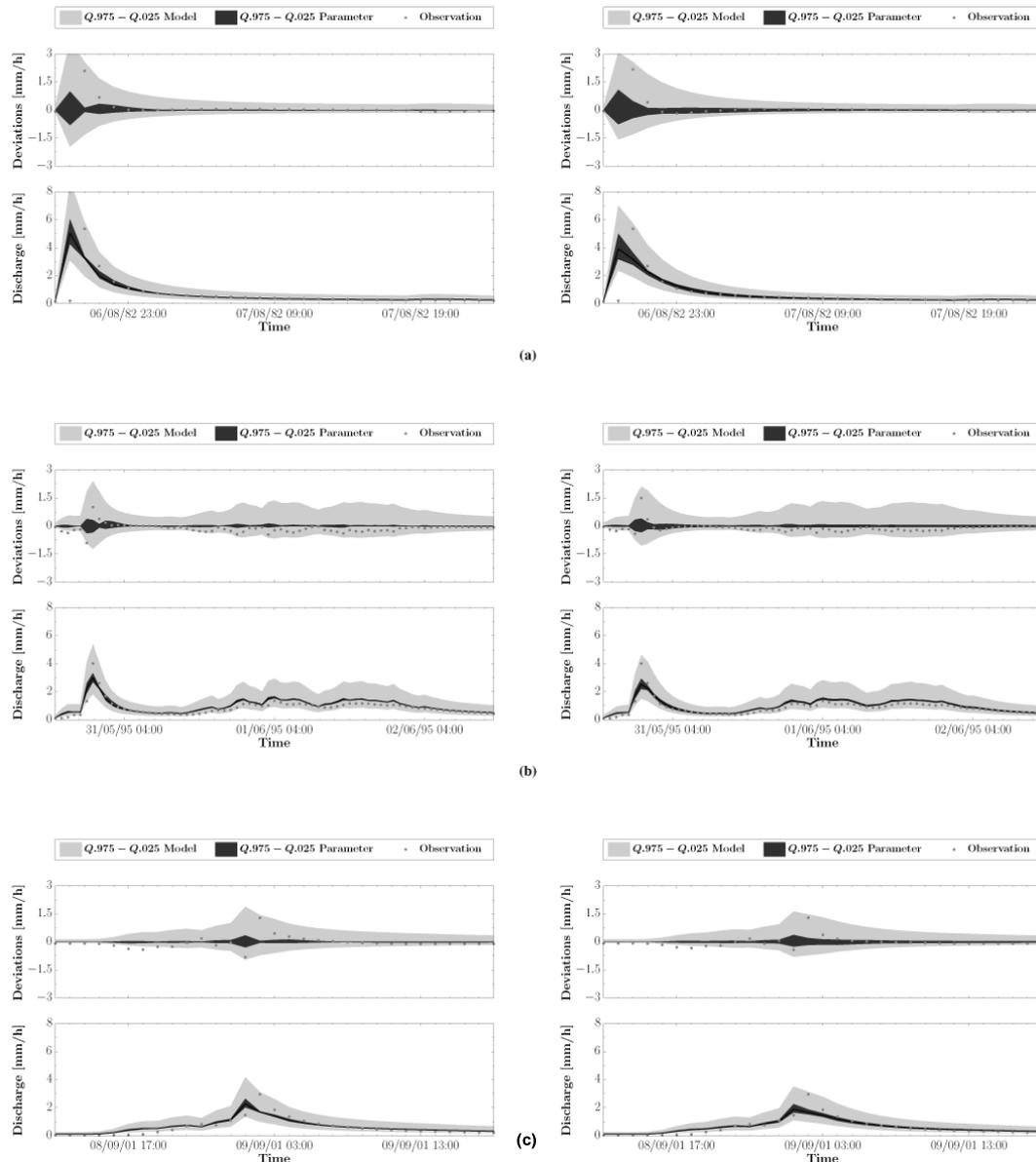


Fig. 17. Hydrograph prediction uncertainty associated with the uncertainty in the model (lighter shading) and parameter estimates (darker shading) for the flood events 6 (a), 8 (b) and 13 (c), estimated by MO-ROPE (left column) and ROPE-PSO using the FloodSkill criterion (right column). The dots correspond to the observed streamflow data. The shaded areas of uncertainty correspond to the 95 % confidence intervals.

spatially distributed soil moisture measurements can help to improve the performance of hydrologic models used for flood forecasting (e.g. Oudin et al., 2003). The Rietholzbach catchment has been intensively monitored as a scientific research catchment since 1975. Lysimeter measurements deliver the necessary soil moisture information on hourly basis. Due to the small size of the basin and the location of the lysimeter station in the centre of the basin situated on grassland, which is the major land use in the catchment, the measurements are on average representative for the whole catchment for the long term (Gurtz et al., 2003). Unfortu-

nately, there are no other soil moisture measurements that have a substantial length and no other significant studies for the catchment that examine the soil moisture dynamics for storm events in a more distributed way. Therefore, we studied the temporal dynamics of the soil moisture measurements at the lysimeter station and found a relatively strong correlation between the lysimeter measurements and the simulated soil moisture in the upper layers up to a soil depth of 0.5 m also during periods with high streamflow values, i.e. flood events (Müller, 2009). Hence, we chose the correlation between observed and simulated soil moisture as an

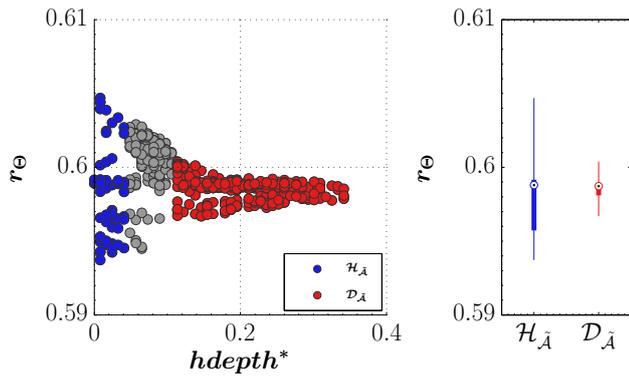


Fig. 18. Correlation between data depth and overall validation performance in terms of the soil moisture correlation r_{Θ} of the Pareto set $\tilde{\mathcal{A}}$ estimated during the 2-objective calibration. The blue dots indicate the hull $\mathcal{H}_{\tilde{\mathcal{A}}}$, i.e. the parameter vectors with shallow data depth and the red dots indicate the sampled parameter vectors with high data depth $\mathcal{D}_{\tilde{\mathcal{A}}}$.

additional objective. We call this performance criterion soil moisture correlation coefficient and denote it by r_{Θ} . Its definition is given in the already referenced Table 13. The underlying idea considering the soil moisture correlation coefficient as an additional objective is that parameter vectors do not only correspond to a good representation of catchment's behaviour in terms of the measured discharge, but also the mean soil moisture dynamics are potentially better transferable to other flood events and can be used for extrapolation, i.e. they are potentially more robust. The r_{Θ} has a relatively low correlation with both objectives that are already taken into account and should thus provide additional information for the parameter estimation. Nonetheless, this is no guarantee that this will really improve the calibration results as “the possibility of operational use of the variational method depends heavily on the availability and quality of the datasets, especially regarding soil moisture data” (Oudin et al., 2003, p. 685). Doubts about the effectivity of the available soil moisture information are for instance reinforced by a further evaluation of the results estimated in the previous case study. Figure 18 provides information about the dependance between r_{Θ} and the data depth with respect to the estimated Pareto set. Obviously, the deep parameter vectors in $\mathcal{D}_{\tilde{\mathcal{A}}}$ that are robust in terms of the objectives rPD and NS provide no better correlation between observed and simulated soil moisture.

We repeated the calibration of WaSiM using the MO-ROPE algorithm using the three introduced objectives rPD, NS and r_{Θ} . All settings of the model and calibration setup remain the same as in the previous case study. In two runs robust parameter vectors were estimated in terms of all three objectives and just the rPD and NS. From now on we refer to both calibration runs as 2-objective and 3-objective calibration. The distributions of the estimated parameter vec-

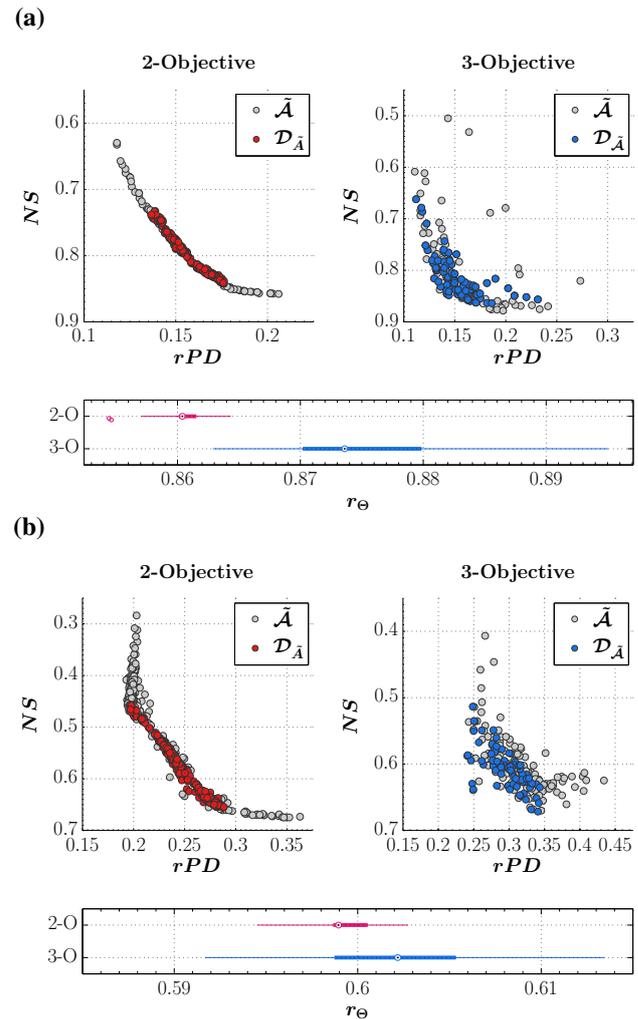


Fig. 19. Comparison of the Pareto optimal sets $\tilde{\mathcal{A}}$ and the subsequently sampled set of deep parameter vectors $\mathcal{D}_{\tilde{\mathcal{A}}}$ for the 2-objective (left) and 3-objective (right) calibration run in the objective space based on the calibration events (a) and validation events (b). Additionally, each plot contains a boxplot providing information about the distribution of the corresponding soil moisture correlation for the sampled deep parameter vectors each (below).

tors for both runs are provided in Table 4.2. Obviously, the results for the 2-objective calibration are the same as those of the previous case study. For the calibration run using all three objectives the distribution of the most sensitive parameter k_d tends to result in smaller values than for the calibration run considering just the objectives rPD and NS. The distribution of the parameter k_i has approximately the same mean and variance but a significantly larger negative skewness. The mean value of the parameter “dr” is approximately the same with a significantly lower variance. The same holds for the soil parameters β_{SL} and β_{SiL} . An obvious difference can be observed in the distribution of the parameter k_{rec} . The mean value is significantly lower with a higher variance. That

means that the effective saturated conductivity for deeper soil layers decreases faster. This means that the lower soil layers in the model have a lower saturated conductivity. The correlations between the most important conceptual model parameters are approximately the same for both calibration runs. However, the correlation between the parameter k_{rec} and the other model parameters is fundamentally different for the 3-objective calibration run.

The corresponding model performances on the calibration and validation data for both the 2-objective calibration and the 3-objective calibration considering the soil moisture dynamics are provided in Fig. 19. The scatter plots show the rPD and the NS for the approximated Pareto optimal set \tilde{A} and the subsequently sampled deep parameter vectors $\mathcal{D}_{\tilde{A}}$. Evaluating the Pareto front of the 3-objective problem, consider that the Pareto front for this problem is actually a curved 3-dimensional face. Hence, the 2-dimensional plot of this 3-dimensional front is not a non-dominated front with respect to just 2 objectives. The calibration results for the 3-objective calibration are the same in terms of the best values for the individual objectives. However, the variation intervals for the two objectives rPD and NS are clearly larger. The NS reaches values down to 0.5 instead of 0.63 and the rPD goes up to values of 0.28 instead of 0.22. The differences between the two calibration runs in terms of the criterion r_{\ominus} are just 0.015. Referring to the validation results, the improvements of the 3-objective calibration in terms of the correlation between the observed and simulated soil moisture are again negligibly small. An increase of the correlation coefficient of just one hundredths is no substantial improvement. However, the validation performance in terms of the criteria assessing a good representation of the observed runoff is slightly worse, particularly for the rPD. For the 3-objective calibration, the rPD values on the validation events are approximately in the range of 0.24 up to 0.34, whereas they reach from 0.19 up to 0.29 for the 2-objective calibration. In terms of the NS values, the results are comparable. The 3-objective estimates correspond to NS values in the range of 0.51–0.67 instead of 0.47–0.67 for the 2-objective calibration. In brief, the soil moisture measurements available in the Rietholzbach catchment are no suitable additional criterion for the calibration of the hydrologic model WaSiM focussing on flood events. The negligibly small improvement in terms of the representation of the soil moisture dynamics and the global NS criterion brings a slight drop of the model performances focussing on an exact representation of the streamflow values. One possible explanation for the disappointingly small improvement might be the shortcomings in the model structure, already discussed in the previous case study. The soils of the Rietholzbach catchment are characterised by many macropores and some smaller drainages. This induces the generation of preferential flow which is a very fast runoff component. However, the water movement in the unsaturated zone is described by the model in terms of the Richards equation that accounts just for the matrix flow in the soil but cannot

represent the preferential flow component natively. Thus, the fast runoff components have to be represented by the direct runoff component. The identified parameter vectors provide a good representation of the catchment's behaviour in terms of the discharge at the outlet. In such a situation an additional criterion that requires a reasonable representation of the observed soil moisture dynamics can even be counterproductive. Another possible explanation for the expected improvements that have not been achieved is the question of whether the soil moisture measurements are representative for the soil moisture dynamics in the whole catchment. Recent studies using soil moisture data for the improvement of hydrologic modelling emphasise the importance of a good quality of these datasets (e.g. Oudin et al., 2003; Norbiato et al., 2008). Focussing on flood events, this requires not only a good accuracy and temporal resolution of the soil moisture information but also many spatially distributed measurements. Unfortunately, these measurements are not available in the Rietholzbach catchment. We suggest further studies in operational flood forecasting studies for fast responding medium-scale catchments where such information is available. Although this case study did not fully succeed, it confirms the advantages of the depth-based sampling. The sampled deep parameters provide a robust approximation of the Pareto front with tighter variation intervals in terms of all considered objectives. Furthermore, it underlines that a successful robust modelling does not require just an advanced parameter estimation procedure but also the selection of a as parsimonious as possible model structure, representative calibration data and appropriate calibration objectives. The combination of multi-objective optimisation and depth-based parameter sampling can be a good tool to obtain robust model parameters. On its own, the depth-based sampling is however not sufficient to achieve robustness.

5 Discussion and conclusions

- This paper presents a hybrid parameter estimation approach, entitled multi-objective robust parameter estimation (MO-ROPE), that merges the strength of evolutionary multi-objective optimisation algorithms and depth-based parameter sampling. In a first step, the algorithm employs a suitable multi-objective optimisation algorithm in order to approximate the Pareto optimal set of model parameter vectors for the given calibration problem. Within our framework, we apply an extended version of the advanced multi-method framework AMALGAM. We extended the original version of AMALGAM with a hybrid evolutionary optimisation algorithm (MO-PSO-GA) that employs the concepts of particle swarm optimisation and genetic programming in order to effectively estimate the Pareto optimal set. An adapted version of this search strategy has proven successful within the frame of single-objective ROPE

algorithms. In a second step, we apply the concept of data depth for the sampling of robust model parameter vectors with respect of the identified Pareto optimal set. We study the efficiency and effectiveness of the developed solution by means of a set of synthetic benchmarks and the calibration of a process-oriented hydrologic model focussing on flood events.

- A set of synthetic benchmarks studies the efficiency and effectiveness of the AMALGAM framework extended by the newly developed MO-PSO-GA algorithm in order to estimate the Pareto optimal set for a given (constrained) multi-objective optimisation problem. For the given test problems, the developed framework proved its reliability and efficiency in comparison with other established approaches. The new MO-PSO-GA strategy is a useful search strategy that is able to outperform approved single-strategy multi-objective optimisation algorithms. As a consequence within the frame of the AMALGAM framework, it improves the effectiveness and efficiency of this approach using four approved search strategies.
- Another case study shows the effectiveness of the depth-based sampling approach for several multi-objective test problems that are subject to uncertainty. The sampled deep parameter vectors are less sensitive to small changes with respect to their ability to provide a robust approximation of the set of the possible Pareto fronts in the objective space considering uncertainties.
- In a real world case study, we compared the multi-objective MO-ROPE approach with the single-objective robust parameter estimation approach ROPE-PSO estimating three conceptual model parameters and three soil parameters of the hydrologic model WaSiM in the small Swiss research catchment Rietholzbach focussing on flood events. Previous studies have already shown that the model has problems to represent the global catchment behaviour and the peak flow values equally well. This applies in particular to small and fast responding catchments. The tradeoff between the representation of the peak flow values and the global catchment behaviour suggests the application of a multi-objective calibration strategy. The results of this study show that the application of MO-ROPE is a preferable option in order to identify robust solutions for such calibration problems. We showed that the parameter vectors in the approximated Pareto optimal set on the calibration data can lead to very different results in validation. The sampled parameters with high depth show a robust performance with less negative outliers. They represent a robust approximation of the (theoretical) Pareto set on the validation data.
- In the scope of a second application, we studied the calibration of WaSiM considering three objectives. Besides

the peak flow deviation and the Nash-Sutcliffe efficiency criterion, we used the agreement of the observed and simulated soil moisture dynamics for the estimation of robust parameter vectors for the modelling of flood events. Although the deep parameter vectors show once again the advantages of the depth-based sampling, the improvements in comparison with the calibration using just two objectives are negligibly small. This is due to shortcomings in the model structure and a limited significance of the soil moisture measurements at only one single spot. We strongly propose similar applications at larger scale with a sufficient set of spatially distributed measurements. This case study underlines that the depth-based parameter sampling can be a very useful technique for the identification of robust parameter vectors for multi-objective calibration problems.

- This paper introduced a new multi-objective method for the estimation of robust parameter vectors that can improve the results of classical multi-objective optimisation techniques. Despite all of the presented benefits, it must not be forgotten that the estimation of robust parameter vectors cannot be reduced to the development of robust parameter estimation methods. The shortcomings of all kind of calibration techniques in the last case study clearly show necessity that more effort should also go into finding better process descriptions resulting in hydrologic models that are able to represent catchments in a better way. Furthermore, better methods have to be developed to improve the identification of representative calibration data, i.e. the basis of a subsequent model calibration. The achievement of robustness requires the combination of improvements in all of these areas.

In this paper, the estimation of the Pareto optimal sets was done using an extended version of AMALGAM approach. Consider that one might substitute this component with any suitable multi-objective optimisation technique. The application of the technique of depth-based sampling is a relatively new method which was applied to a limited number of case studies. We strongly propose its application to further models, catchments and also other fields of study where measurement errors with unknown distribution and model structures that cannot be easily identified are present.

Acknowledgements. We would like to thank the German Research Foundation and the Cusanuswerk for the funding of this work. Furthermore we thank Irene Lehner from the land-climate interactions group at the ETH Zurich for the provision of all the measurement data in the Rietholzbach catchment and the Center for Information Services and High Performance Computing (ZIH) at the University of Technology, Dresden for the supply with the required computing power to carry out all the optimisation runs.

Edited by: H. Cloke

References

- Beven, K.: Changing ideas in hydrolog – the case of physically-based models, *J. Hydrol.*, 105, 157–172, 1989.
- Boyle, D., Lamorey, G., Bassett, S., Pohl, G., Krause, P., and Kralisch, S.: Investigating the value of hydrologic model complexity and spatial information using multicriteria methods, in: 7th International Conference on Hydroinformatics, Nice, France, 2006.
- Brakensiek, D., Rawls, W., and Stephenson, G.: Modifying SCS hydrological soil groups and curve numbers for rangeland soils, ASAE Paper, 203, 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.: Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth Syst. Sci.*, 11, 703–710, doi:10.5194/hess-11-703-2007, 2007.
- 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.
- Coello, C., Pulido, G., and Lechuga, M.: Handling multiple objectives with particle swarm optimization, *IEEE T. Evolut. Comput.*, 8, 256–279, 2004.
- 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.
- Cullmann, J., Krauß, T., and Philipp, A.: Enhancing flood forecasting with the help of processed based calibration, *Phys. Chem. Earth*, 33, 1111–1116, 2008.
- Deb, K.: *Multi-Objective Optimization Using Evolutionary Algorithms*, Wiley, New York, 2001.
- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.: A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II, *IEEE T. Evolut. Comput.*, 6, 182–197, 2002.
- 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, 1992.
- Efstratiadis, A. and Koutsoyiannis, D.: One decade of multiobjective calibration approaches in hydrological modelling: a review, *Hydrolog. Sci. J.*, 55, 58–78, 2010.
- Engeland, K., Braud, I., Gottschalk, L., and Leblois, E.: Multi-objective regional modelling, *Hydrol. Process.*, 327, 339–351, 2006.
- 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, 2010.
- Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: A comparison of alternative multiobjective calibration strategies for hydrological modeling, *Water Resour. Res.*, 43, W03434, doi:10.1029/2006WR005098, 2007.
- Fonseca, C. and Fleming, P.: Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization, in: *Genetic Algorithms: Proceedings of the Fifth International Conference*, 416–423, San Mateo, CA, 1993.
- 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.
- Green, W. and Ampt, G.: Studies on Soil Physics: I. The flow of air and water through soils, *J. Agr. Sci.*, 4, 1–24, 1911.
- 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, 2003.
- Haario, H., Saksman, E., and Tamminen, J.: An adaptive Metropolis algorithm, *Bernoulli*, 7, 223–242, 2001.
- Kennedy, J., Eberhart, R. C., and Shi, Y.: *Swarm Intelligence*, Morgan Kaufmann Series in Evolutionary Computation, Morgan Kaufman, 2001.
- Krauß, 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.
- Krauß, T. and Cullmann, J.: Towards a more representative parametrisation of hydrological models via synthesizing particle swarm optimisation and robust parameter estimation, *Hydrol. Earth Syst. Sci.*, 16, 603–629, doi:10.5194/hess-16-603-2012, 2012a.
- Krauß, T. and Cullmann, J.: Data depth – parsimonious sampling of robust parameters in hydrological models, *Environ. Modell. Softw.*, submitted, 2012b.
- Krzysztofowicz, R.: Transformation and normalization of variates with specified distributions, *J. Hydrol.*, 197, 286–292, doi:10.1016/S0022-1694(96)03276-3, 1997.
- 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 Theoretical Computer Science*, American Mathematical Society, 2006.
- Maier, H. R. and Dandy, G. C.: Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, *Environ. Modell. Softw.*, 15, 101–124, doi:10.1016/S1364-8152(99)00007-9, 2000.
- 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/>, 2007.

- Moussa, R. and Chahinian, N.: Comparison of different multi-objective calibration criteria using a conceptual rainfall-runoff model of flood events, *Hydrol. Earth Syst. Sci.*, 13, 519–535, doi:10.5194/hess-13-519-2009, 2009.
- Müller, F.: Untersuchung der ROPE-Methodik zur robusten Kalibrierung von hydrologischen Modellen zur Hochwasservorhersage am Beispiel des Schweizer Testeinzugsgebietes Rietholzbach, Master's thesis, Institut für Hydrologie und Meteorologie, Technische Universität Dresden, 2009.
- Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I – A discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.
- Norbiato, D., Borga, M., Esposti, S. D., Gaume, E., and Anquetin, S.: Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins, *J. Hydrol.*, 362, 274–290, doi:10.1016/j.jhydrol.2008.08.023, 2008.
- Oudin, L., Weisse, A., Loumagne, C., and Hégarat-Masclé, S. L.: Assimilation of soil moisture into hydrological models for flood forecasting: a variational approach, *Can. J. Remote Sensing*, 29, 679–686, 2003.
- Perez, R. E. and Behdinan, K.: Particle swarm approach for structural design optimization, *Computers and Structures*, 85, 1579–1588, 2007.
- Rousseeuw, P. J. and Struyf, A.: Computing location depth and regression depth in higher dimensions, *Stat. Comput.*, 8, 193–203, 1998.
- Schott, J.: Fault tolerant design using single and multicriteria genetic algorithm optimization, Master's thesis, Dep. of Aeronaut. and Astronaut., Mass. Inst. of Technol., Cambridge, Mass., 1995.
- Schulla, J.: Hydrologische Modellierung von Flussgebieten zur Abschätzung der Folgen von Klimaänderungen, Ph.D. thesis, Eidgenössische Technische Hochschule Zürich, diss. ETH Nr. 12018, 1997.
- Seifert, M.: Untersuchungen zum Einfluss unsicherer Bodeninformationen im Schweizer Testeinzugsgebiet Rietholzbach, Master's thesis, Institut für Hydrologie und Meteorologie, Technische Universität Dresden, 2010.
- Serfling, R.: Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications, vol. 72 of DIMACS Series in Discrete Mathematics and Theoretical Computer Science, chap. Depth functions in nonparametric multivariate inference, 1–16, American Mathematical Society, 2006.
- 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.
- Singh, S.: Robust Parameter Estimation in Gauged and Ungauged Basins, Vol. 198 of Mitteilungen des Instituts für Wasserbau, Universität Stuttgart, Institut für Wasserbau, Universität Stuttgart, 2010.
- Storn, R. and Price, K.: Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, *J. Global Optim.*, 11, 341–359, doi:10.1023/A:1008202821328, 1997.
- Thyer, M., Frost, A. J., and Kuczera, G.: Parameter estimation and model identification for stochastic models of annual hydrological data: Is the observed record long enough?, *J. Hydrol.*, 330, 313–328, 2006.
- 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.
- van der Linden, S. and Woo, M.-K.: Application of hydrological models with increasing complexity to subarctic catchments, *Journal of Hydrology*, 270, 145–157, 2003.
- van Veldhuizen, D. A. and Lamont, G.: Multiobjective evolutionary algorithm research: A history and analysis, Tech. Rep. TR-98-03, Dep. of Electr. Comput. Eng., Grad. Sch. of Eng., Air Force Inst. of Technol., Wright-Patterson Air Force Base, Ohio, 1998.
- Vrugt, J. A. and Robinson, B. A.: Improved evolutionary optimization from genetically adaptive multimethod search, in: Proceedings of the National Academy of Sciences of the United States of America, 104, 708–711, doi:10.1073/pnas.0610471104, available at: www.pnas.org/cgi/content/full/0610471104/DC1, 2007.
- 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, 2003.
- Vrugt, J. A., Robinson, B. A., and Hyman, J. M.: Self-Adaptive Multimethod Search for Global Optimization in Real-Parameter Spaces, *IEEE Transaction On Evolutionary Computation*, 13, 243–259, 2009.
- Wagner, T. and Gupta, H.: Model identification for hydrological forecasting under uncertainty, *Stoch. Env. Res. Risk A.*, 19, 378–387, doi:10.1007/s00477-005-0006-5, 2005.
- Wagner, T. and Wheeler, H. S.: A Generic Framework for the Identification of Parsimonious Rainfall-Runoff Models, in: Integrated Assessment and Decision Support (Proc. First Biennial Meeting of the International Environmental Modelling and Software Society), edited by: Rizzoli, A. E. and Jakeman, A. J., Vol. 1, 434–439, 2002.
- Wagner, T. and Wheeler, H. S.: Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, *J. Hydrol.*, 320, 132–154, doi:10.1016/j.jhydrol.2005.07.015, 2006.
- Woldesenbet, Y. G., Tessema, B. G., and Yen, G. G.: Constraint handling in multi-objective evolutionary optimization, in: IEEE Congress on Evolutionary Computation, Piscataway, NJ, USA, 2007.
- Wolpert, D. and Macready, W.: No free lunch theorems for optimization, *IEEE Transactions on Evolutionary Computation*, 1, 67–82, doi:10.1109/4235.585893, 1997.
- 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, 1999.
- Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-objective global optimization for hydrologic models, *J. Hydrol.*, 204, 83–97, doi:10.1016/S0022-1694(97)00107-8, 1998.
- 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.
- Zitzler, E., Deb, K., and Thiele, L.: Comparison of Multiobjective Evolutionary Algorithms: Empirical Results, *Evol. Comput.*, 8, 173–195, 2000.