

Multi-objective calibration of a distributed hydrological model (WetSpa) using a genetic algorithm

M. Shafii and F. De Smedt

Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Belgium

Received: 24 October 2008 – Published in Hydrol. Earth Syst. Sci. Discuss.: 12 January 2009

Revised: 26 October 2009 – Accepted: 26 October 2009 – Published: 12 November 2009

Abstract. A multi-objective genetic algorithm, NSGA-II, is applied to calibrate a distributed hydrological model (WetSpa) for prediction of river discharges. The goals of this study include (i) analysis of the applicability of multi-objective approach for WetSpa calibration instead of the traditional approach, i.e. the Parameter ESTimator software (PEST), and (ii) identifiability assessment of model parameters. The objective functions considered are model efficiency (Nash-Sutcliffe criterion) known to be biased for high flows, and model efficiency for logarithmic transformed discharges to emphasize low-flow values. For the multi-objective approach, Pareto-optimal parameter sets are derived, whereas for the single-objective formulation, PEST is applied to give optimal parameter sets. The two approaches are evaluated by applying the WetSpa model to predict daily discharges in the Hornad River (Slovakia) for a 10 year period (1991–2000). The results reveal that NSGA-II performs favourably well to locate Pareto optimal solutions in the parameters search space. Furthermore, identifiability analysis of the WetSpa model parameters shows that most parameters are well-identifiable. However, in order to perform an appropriate model evaluation, more efforts should be focused on improving calibration concepts and to define robust methods to quantify different sources of uncertainties involved in the calibration procedure.

and Uresti-Charre, 1997). After the first pioneering studies on evolutionary multi-objective optimization in the mid-1980s (Schaffer, 1984; Fourman, 1985), these algorithms were successfully applied to various multi-objective optimization problems (e.g. Ishibuchi and Murata, 1996; Cunha et al., 1997; Valenzuela-Rend'on and Uresti-Charre, 1997; Fonseca and Fleming, 1995). There have also been significant contributions on application of GAs for multi-objective optimization in water resources research (Ritzel et al., 1994; Cieniawski et al., 1995; Reed et al., 2001; Reed and Minsker, 2004).

Conceptual rainfall-runoff (RR) models, aiming at predicting stream flow from the knowledge of precipitation over a catchment, have become basic tools for flood and drought forecasting, catchment basin management, spillway design, and flood protection. Calibration of RR models is a process in which parameter adjustment are made so as to match (as closely as possible) the dynamic behaviour of the RR model to the observed behaviour of the catchment. Because of the multi-objective nature of RR calibration processes, automatic calibration methodologies have been shifted from single-objective towards multi-objective formulation in recent years. Gupta et al. (1998) discussed for the first time the advantages of multiple-objective model calibration and showed that such schemes are applicable and desirable. Subsequently, more research has been focused on multi-objective approaches for calibration of RR models (Yapo et al., 1998; Seibert, 2000; Cheng et al., 2002; Boyle et al., 2000; Madsen, 2000; Vrugt et al., 2003).

Over past recent years, population-based search algorithms have shown to be powerful search methods for multi-objective optimization problems and have been applied for multi-objective RR calibration, especially when there are a large number of calibration parameters (Boyle et al., 2000; Madsen, 2000; Vrugt et al., 2003; Khu et al., 2005). Tang et al. (2006) comprehensively assessed the efficiency, effectiveness, reliability, and ease-of-use of three multi-objective

1 Introduction

Genetic algorithms (GA) have become increasingly popular for solving complex multi-objective optimization problems because of their better performance compared to other search strategies (Fonseca and Fleming, 1995; Valenzuela-Rend'on



Correspondence to: M. Shafii
(mshafiih@vub.ac.be)

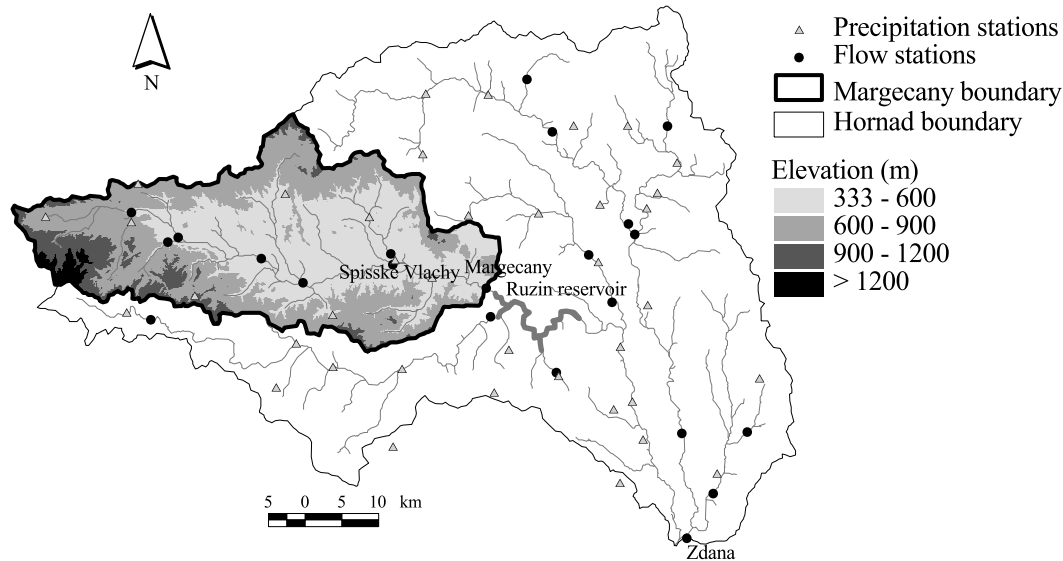


Fig. 1. Hydrologic network of the Hornad catchment with topography of Margecany sub-catchment and location of gauging and meteorological stations.

evolutionary optimization algorithms (MOEAs) for hydrologic model calibration. Another comprehensive comparison between other optimization algorithms was dealt with by Wöhling et al. (2008). Moreover, some researchers have applied MOEAs to develop automatic multi-objective calibration strategies for distributed hydrological models (Madsen, 2003; Ajami et al., 2004; Muleta and Nicklow, 2005a, b; Vrugt et al., 2005; Bekele and Nicklow, 2007).

This paper presents an application of a MOEA, Non-dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al., 2002), for multi-objective RR calibration of a distributed hydrological model (WetSpa; Wang et al., 1997). In the past, calibration of this model has been performed by classical least squares minimization with the Parameter ESTimator software (PEST; Doherty and Johnson, 2003), e.g. Liu et al. (2005); Bahremand et al. (2007). Usually, the single-objective to be minimized is the sum of squared differences between observed and estimated river discharges. However, this criterion is known to be biased for high-flows. An alternative approach is to use log-transformed discharges to emphasize low-flows, but this can lead to quite different optimal parameter values, creating a dilemma for the user which parameter set to prefer. In this paper, we apply a multi-objective approach to calibrate the WetSpa model using both criteria in order to find out whether a compromise is possible with equal attention to both high- and low-flows. We aim at (i) investigating the difference between single- and multiple-objective model calibration approaches in terms of how the optimal regions of model parameters vary over the search space, and also (ii) assessment of the identifiability of these parameters.

The paper is organized as follows. Section 2 provides material and methods used in this paper, i.e. the study area,

WetSpa model, representation of the multi-objective optimization algorithm (NSGA-II), description of the single-objective optimization routine (PEST), and the framework of the WetSpa model calibration within these two approaches. Section 3 describes the models application results, corresponding analyses, and discussions. Finally, conclusions and recommendations for further research are presented in Sect. 4.

2 Material and methods

2.1 Study area

The WetSpa hydrological model is applied to the Hornad River, located in Slovakia. The drainage area of the river up to the Margecany gauging station is 1.131 km². Figure 1 shows the Hornad catchment, the topography until Margecany, and the location of gauging and meteorological stations. The basin is mountainous with elevations ranging from 339 to 1556 m. The basin has a northern temperate climate with distinct seasons. The highest amount of precipitation occurs in the summer period from May to August while in winter there is usually only snow. The mean annual precipitation is about 680 mm, ranging from 640 mm in the valley to more than 1000 mm in the mountains. The mean temperature of the catchment is about 6°C and the annual potential evapotranspiration about 520 mm. About half of the basin is covered by forest, while the other half consists mainly of grassland, pasture, and agriculture areas. The dominant soil texture is loam, which covers about 42% of the basin, and sandy loam and silt loam about 24% and 23% respectively. Detailed information about the study area along with the methodology

Table 1. Global WetSpa model parameters to be calibrated: description, symbols, preset feasible range, range of Pareto optimal values obtained with NSGA-II, and optimal values obtained with PEST (Solutions No. 1 and No. 2).

Description	Parameter	Units	Feasible range	NSGA-II		PEST solutions	
				Min	Max	No. 1	No. 2
Interflow scaling factor	K_i	–	0–10	1.83	1.88	2.082	2.105
Groundwater recession coefficient	K_g	d^{-1}	0–0.05	0.0059	0.0087	0.0085	0.0064
Initial soil moisture factor	K_s	–	0–2	1	1	1.008	1.158
Correction factor for PET	K_e	–	0–2	1.16	1.33	1.16	1.48
Initial groundwater storage	K_{gi}	mm	0–500	44	46	43	20
Groundwater storage scaling factor	K_{gm}	mm	0–2000	133	545	133	1188
Base temperature for snowmelt	K_t	$^{\circ}C$	–1–1	–0.13	0.61	0.25	0.89
Temperature degree-day coefficient	K_{td}	$mm\ ^{\circ}C^{-1}\ d^{-1}$	0–10	0.87	0.97	0.89	1.34
Rainfall degree-day coefficient	K_{rd}	$^{\circ}C^{-1}\ d^{-1}$	0–0.05	0.028	0.036	0.02	0.047
Surface runoff coefficient	K_m	–	0–5	2.8	3.12	2.76	4.26
Rainfall scaling factor	K_p	mm	0–500	433	497	500	79

to extract required data for the WetSpa model has been provided by Bahremand et al. (2007). Observations of daily precipitation, temperature, potential evaporation, and discharge are available for the period 1991–2000. The first 5 years of the 10-year period is chosen for model calibration and the second 5 years for model validation.

2.2 WetSpa hydrological model

WetSpa is a grid-based distributed hydrologic model for water and energy transfer between soil, plants and atmosphere, which was originally developed by Wang et al. (1997) and adapted for flood prediction on hourly time step by De Smedt et al. (2000, 2004), and Liu et al. (2003, 2004, 2005). For each grid cell, four layers are considered in the vertical direction, i.e. the plant canopy, the soil surface, the root zone, and the groundwater zone (Fig. 2). The hydrologic processes considered in the model are precipitation, interception, depression storage, surface runoff, snowmelt, infiltration, evapotranspiration, interflow, percolation, and groundwater drainage. The model predicts peak discharges and hydrographs, which can be defined for any numbers and locations in the channel network, and can simulate the spatial distribution of basin hydrological variables. Interested readers may refer to Liu et al. (2003) and De Smedt et al. (2004) for detailed information about WetSpa and its methodology to predict stream flow.

The WetSpa distributed model potentially involves a large number of model parameters to be specified during the model setup. Most of these parameters can be assessed from field data, e.g. hydrometeorological observations, maps of topography, soil types, land use, etc. However, comprehensive field data are seldom available to fully support specification of all model parameters. In addition, some model parameters are of a more conceptual nature and cannot be directly assessed. Hence, some parameters have to be determined

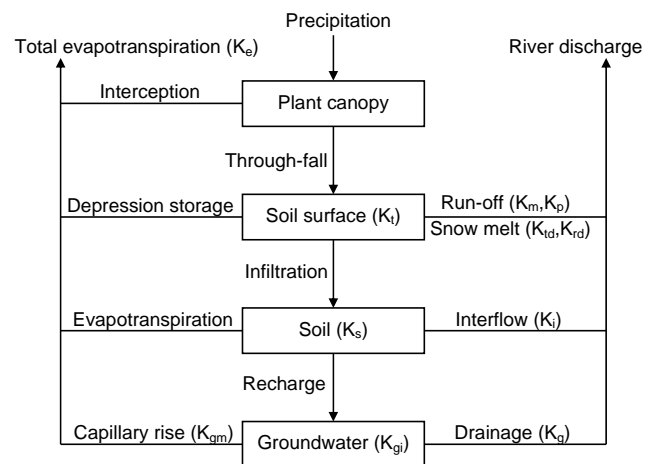


Fig. 2. Schematic representation of the general model structure of WetSpa: arrows represent hydrological processes, boxes represent storage zones, symbols between brackets refer to WetSpa global model parameters to be calibrated as explained in Table 1.

through a calibration process. The choice of parameters to calibrate is based on earlier studies of the WetSpa model (Liu et al., 2003; Liu and De Smedt, 2005; Bahremand et al., 2007). The model parameters that have to be determined through calibration (i.e. eleven parameters) are listed in Table 1 and their impact on the different model components of WetSpa is schematically depicted in Fig. 2. All other model parameters, i.e. spatial hydrological properties related to soil type, land-use, and topography, are automatically derived using GIS tools and need not to be adjusted through calibration.

2.3 Multi-objective optimization algorithm (NSGA-II)

Multi-objective genetic algorithms (MOGAs) and the Pareto optimality concept (Pareto, 1896) have been widely applied

in water resources studies. MOGAs are search algorithms based upon the mechanics of natural selection, derived from the theory of natural evolution. They represent the solutions using strings (also referred to as chromosomes) of variables, which are comprised of a number of genes (decision variables). The fitness of each chromosome is an expression of the objective function value. A MOGA starts with a population of initial chromosomes, which through genetic operators such as selection, crossover, and mutation produce successively better chromosomes.

NSGA-II is one of the MOGAs, proposed by Deb et al. (2002) as a significant improvement to the original NSGA (Srinivas et al., 1993) by using a more efficient ranking scheme and improved selection to capture the Pareto front. Zitzler et al. (2000) and Deb et al. (2002) have shown that NSGA-II performs as well as or better than other algorithms on difficult multi-objective problems. In NSGA-II, the selection process at various stages of the algorithm toward a uniformly spread-out Pareto optimal front is guided by assigning fitness to chromosomes based on domination and diversity. Domination is determined by ranking all chromosomes in the population, where chromosomes with higher rank are considered to have better fitness. Chromosomes with the same rank are compared based on their diversity which is defined based on a crowding measure for each chromosome. Chromosomes with larger values of crowding distance are preferred more to be selected for next generations. Interested readers may refer to Deb et al. (2002) for a detailed description of the algorithm. A brief step-by-step description of NSGA-II, with specific application to the calibration problem of this paper, is as follows:

1. Start with a random generation of a parent population (i.e. a set of parameter sets), followed by sorting based on domination and crowding distance.
2. Create an offspring population of the same size as the parent population through tournament selection with tournament size of 2 (Goldberg and Deb, 1991).
3. Apply a single-point-cut crossover operator (Michalewicz, 1994) to replace parts of designated offspring parameter sets with values from parent solutions. The crossover probability (i.e. the percentage of entire offspring population which is affected by crossover operator) is 90%.
4. Perform a uniform mutation (Michalewicz, 1994) by altering the value of one variable per parameter set, i.e. the mutation probability is $1/s$, where s is the number of parameters.
5. Combine parent and offspring populations, and rank the parameter sets based on domination and diversity.
6. Transfer the top half best parameter sets to the next generation.

7. Repeat steps 2 to 6 till termination criteria are met.

The C-function, proposed by Zitzler and Thiele (1999), is applied as stopping criterion. Let X and X' be two sets of Pareto parameter sets, of which the latter belongs to one generation after the former. The C-function maps the ordered pair (X, X') to the interval $[0,1]$ based on how much X is better than X' , as follows:

$$C(X, X') := \frac{|\{a' \in X'; \exists a \in X : a \prec a'\}|}{|X'|}, \quad (1)$$

where a and a' are respectively individual components of X and X' , and \prec is the sign of domination. The nominator in Eq. (1) indicates the number of parameter sets of X' which are dominated at least by one of the elements of X , and the denominator is the total number of elements in X . The C-function is a measure of the improvement over the iterations expressed as a value between zero and one. Value $C(X, X') = 1$ means that all solutions in X' are dominated by or are equal to all solutions in X . The opposite $C(X, X') = 0$ indicates that none of the solutions in X' are dominated by or are equal to solutions in X . The C-function measure has been used by Zitzler and Thiele (1999) to compare the performance of multiple methods. However, in this study, we apply the C-function to see how much improvement is achieved in the Pareto front of a particular generation compared to the previous one. If the value remains equal to 1 for a number of consecutive iterations (10 in this study), the search algorithm has converged and can be terminated.

2.4 Single-objective optimization routine (PEST)

PEST is a non-linear parameter estimation and optimization package, offering model independent optimization routines (Doherty and Johnston, 2003). Unlike evolutionary algorithms such as NSGA-II, PEST uses a gradient-based methodology (i.e. Levenberg-Marquardt algorithm) to search for the optimal solution. The best set of parameters is selected from within reasonable ranges by adjusting the values until the discrepancies between the model generated values and observations is reduced to a minimum in the weighted least squares sense. Since its development, PEST has gained extensive use in many different fields, as for instance automated calibration of surface runoff and water quality models (e.g. Baginska et al. 2003; Syvoloski et al., 2003). Liu et al. (2005) and Bahremand et al. (2007) applied PEST for calibration of the WetSpa model.

PEST minimizes the sum of squared residuals, i.e. normally the differences between observed and predicted discharges, but PEST can also be applied on log-transformed discharges to put more emphasis on low-flows. PEST also provides useful information for parameter sensitivity analysis and uncertainty assessment. In addition to the best parameter estimates, m , PEST also estimates the standard deviation, s , of the parameter estimates, so that confidence intervals for each parameter are obtained as $m \pm t_{\alpha, n} s$, where

$t_{\alpha,n}$ is student's t-distribution with probability α and n degrees of freedom (usually $\alpha=0.025$, so that each parameter is contained within the predicted confidence interval with a probability of $1-2\alpha$, i.e. 95%).

2.5 Framework of the calibration problem

The main goal of the study is to evaluate the applicability of a multi-objective calibration framework and to analyze the impact of multiple objectives on the optimal regions of the model parameters. The main components for the calibration framework include objective functions and optimization procedures. There are different objective functions which can be applied for calibration. As Madsen et al. (2000) have pointed out, good agreement between simulated and observed peak flows, as well as good agreement for low flows, are among these objective functions. In general, trade-offs exist between different criteria used for calibration. For instance, one may find a set of parameters that provide a very good simulation of peak flows but a poor simulation of low flows, and vice versa. Hence, in order to obtain a successful calibration, it is necessary to formulate performance measures in a multi-objective framework. The following objective functions are used in the present study:

$$CR_1 = 1 - \frac{\sum_{i=1}^N (Q_{s_i} - Q_{o_i})^2}{\sum_{i=1}^N (Q_{o_i} - \overline{Q_0})^2} \quad (2)$$

$$CR_2 = 1 - \frac{\sum_{i=1}^N [\ln(Q_{s_i}) - \ln(Q_{o_i})]^2}{\sum_{i=1}^N [\ln(Q_{o_i}) - \overline{\ln(Q_0)}]^2} \quad (3)$$

where, Q_{o_i} is the observed discharge at time i , Q_{s_i} the simulated discharge at time i , the bar stands for average, and N is the total number of time steps in the calibration period. The first criterion, CR_1 , is the model efficiency (Nash and Sutcliffe, 1970) which evaluates the ability of reproducing all stream flows, but is known biased for peak flows. The second criterion, CR_2 , is the model efficiency for reproducing log-transformed discharges, giving more emphasis to low-flow values. Therefore, the goal of the multi-objective calibration (i.e. objective functions addressed in this study) is to maximize both CR_1 and CR_2 . However, the result of the optimization will not be a single unique set of parameters but will consist of Pareto front solutions. For the single-objective procedure, the two criteria will be assessed separately to derive the best parameter sets with different emphasis on high- and low-flows (i.e. working on normal and log-transformed discharges, respectively).

The multi-objective calibration will be performed with NSGA-II resulting in the optimal Pareto front, while the single-objective optimization will be performed with PEST. Because NSGA-II and PEST belong to different groups of optimization techniques, we do not intend to compare them

Table 2. Objective functions values of the Pareto optimal solutions (No. 1–15) obtained with NSGA-II and PEST solutions 1 and 2, for the calibration and validation periods.

Description	No.	Calibration period		Validation period	
		CR_1	CR_2	CR_1	CR_2
NSGA-II Pareto front solutions	1	0.752	0.685	0.682	0.757
	2	0.759	0.637	0.673	0.750
	3	0.714	0.722	0.668	0.736
	4	0.744	0.708	0.665	0.745
	5	0.705	0.724	0.655	0.741
	6	0.748	0.698	0.667	0.743
	7	0.758	0.651	0.669	0.757
	8	0.756	0.664	0.670	0.743
	9	0.735	0.714	0.664	0.746
	10	0.760	0.616	0.674	0.752
	11	0.760	0.558	0.675	0.741
	12	0.690	0.725	0.620	0.736
	13	0.760	0.585	0.675	0.748
	14	0.726	0.719	0.673	0.741
	15	0.753	0.675	0.688	0.754
PEST solution 1		0.746	0.568	0.703	0.747
PEST solution 2		0.671	0.682	0.584	0.699

in terms of efficiency or technical aspects. We only want to compare the obtained parameter values and explore their identifiability.

The optimization procedure starts with identifying feasible parameter values. Model parameters ranges are chosen according to the basin characteristics, as discussed in the documentation and user manual of the WetSpa model (Liu and De Smedt, 2004) and a previous study on the same area by Bahremand et al. (2007). The preset feasible parameter ranges are given in Table 1. Next, initial values of the parameters for multi- and single-objective algorithms need to be selected. To generate the initial population of NSGA-II, a Latin Hypercube Sampling (LHS) (Iman and Conover, 1980) technique is used to explore the full range of all feasible parameter values. Thousand parameter sets are generated using the LHS technique and WetSpa is run to evaluate the objective criteria. The solutions are subsequently ranked based on the concept of Pareto dominance and the top 50 parameter sets are selected to be the initial population of NSGA-II. For PEST, as it only needs a single solution as the starting point of the search process, we first rank the previously obtained 1000 LHS samples based on the first criterion, CR_1 , and the best parameter set is considered as starting values for optimization with PEST using criteria CR_1 ; we will term the resulting optimal parameter set “PEST solution 1”. Next, the 1000 LHS solutions are ranked according to the second criterion, CR_2 , and likewise, the best parameter set is used as starting values for optimisation with PEST using criteria CR_2 ; the result will be termed “PEST solution 2”.

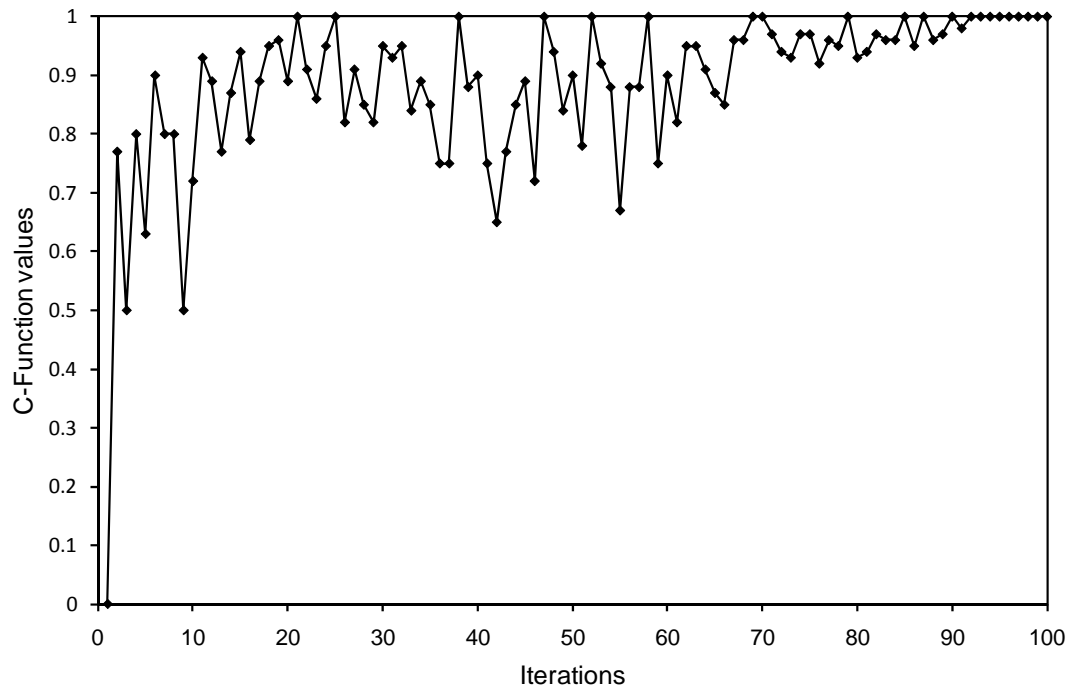


Fig. 3. Convergence measure (C-Function) values versus number of NSGA-II iterations.

3 Results and discussion

3.1 Objective functions values

NSGA-II and PEST (i.e. multi- and single-objective routines) are used according to aforementioned objective functions and application framework to calibrate the WetSpa model. In NSGA-II, CR_1 and CR_2 are maximized, while in PEST, the goal is to maximize CR_2 only. An important issue in NSGA-II application is the termination criterion. In this study, the C-function is considered as the convergence measure, and thus as stopping criterion of the NSGA-II algorithm. The variation of this index over the iterations is shown in Fig. 3. The C-function is low for the first iterations of NSGA-II, but becomes larger as the algorithm proceeds approaching the value of one. Often it becomes equal to one, but becomes smaller again in next iterations. However, after 90 iterations it becomes one and remains equal to 1 for 10 more iterations. Hence, this means that there is no more improvement to be found and consequently the algorithm is considered to be converged and terminated after these 100 iterations. Using a population size of 50, the corresponding total number of function evaluations is 5000.

After convergence of the NSGA-II algorithm, 15 Pareto front solutions are obtained, of which the corresponding objective function values for the calibration and validation periods are given in Table 2. For the calibration period, the model efficiency CR_1 ranges between 0.690 and 0.760, and the low flow model efficiency CR_2 between 0.558 and 0.725.

The values obtained for the validation period are lower for CR_1 , but generally better for the low flow efficiency CR_2 . The latter can be explained by the fact that the validation period is generally dryer; accordingly, flows and residuals are smaller leading to a better efficiency measure for low-flows. All solutions listed in Table 2 are Pareto optimal for the calibration period, and therefore, are all worthy candidates for model calibration depending upon the preferences of the user and the goals of the model application.

The corresponding objective function values for the optimization with PEST are shown in the bottom part of Table 2. Notice that PEST only calibrates the model based on simple least squares. Hence, both objective functions have afterwards been manually evaluated from the simulated discharges. The results are similar, i.e. for the calibration period, the better model efficiency CR_1 is about 0.746 for PEST solution 1 and CR_2 about 0.682 for PEST solution 2, while the values obtained for the validation period are lower for CR_1 but better for CR_2 .

In order to visualize the results of Tables 2, a bi-criterion CR_1 - CR_2 plot of Pareto front solutions and PEST solutions 1 and 2 for the calibration period are shown in Fig. 4. It is observed that the spread or trade-off of NSGA-II solutions between criteria CR_1 and CR_2 is quite uniform. Moreover, this trade-off is properly distributed in-between the two PEST solutions. This illustrates the applicability of multi-objective calibration to better explore the optimal region and to obtain more optimal solutions, which provides stakeholders with flexibility to make decisions. Comparison between the

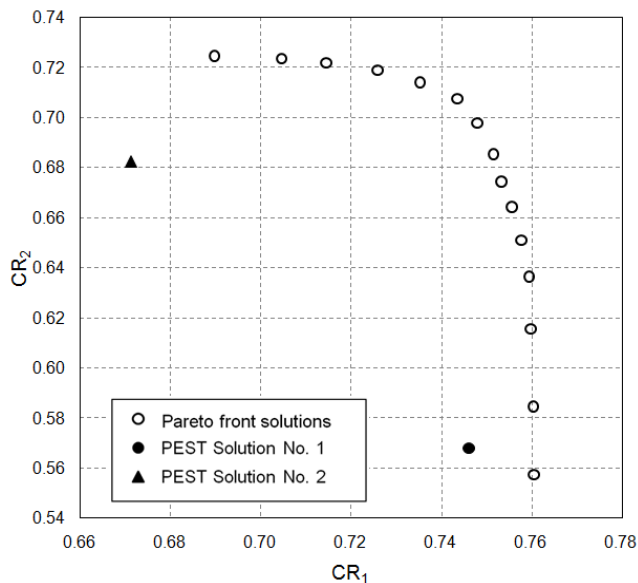


Fig. 4. Bi-criterion CR_1 – CR_2 plot of Pareto optimal solutions obtained by NSGA-II and PEST solutions 1 and 2, as given in Table 2 for the calibration period.

results obtained by NSGA-II and by PEST shows that some Pareto front solutions have resulted in a good performance of the simulation model, in terms of CR_1 and CR_2 values. In particular, some of the Pareto solutions are better than PEST solution 1 for criterion CR_1 and also better than PEST solution 2 for CR_2 . This does not prove that NSGA-II is more efficient than PEST but rather indicates that NSGA-II performs well and is capable of searching the parameter space to obtain optimum results. The less performance of PEST in this respect lies, in essence, in its restriction to locally search the parameter space in the neighbourhood of initial starting points. If PEST had been used with more initial parameter starting values, probably better solutions would have been obtained, though at a higher computations cost. In general, it can be concluded that the multi-objective calibration of the WetSpa model, i.e. using NSGA-II, performs favourably well compared to the traditional single-criterion calibration as with PEST.

3.2 Model parameters values

The optimal parameter values obtained with NSGA-II and PEST are presented in Table 1. For NSGA-II, only the range (i.e. minimum and maximum) of the 15 Pareto optimal values for each parameter is given, whereas for PEST, optimal parameter values are listed for both solutions 1 and 2. These optimum parameter values, along with the 95% confidence intervals of the PEST solutions, are also graphically depicted in Fig. 5, whereby the values are normalized according to the preset initial range of the parameters as given in Table 2. Figure 5 demonstrates that there is a relatively high consist-

tency between the results obtained with both techniques, because the parameters values obtained with NSGA-II either fall within or are close to the range reported by the two PEST solutions. This is generally the case for all parameters except for K_t where the range obtained by NSGA-II partly covers the distance between the two PEST solutions, which is a sign of the high uncertainty and model insensitivity associated with this parameter. If the range between the PEST-obtained solutions is considered as a measure of the distance between two optimal regions of the search space with different attentions to high- and low-flows, it can be concluded that NSGA-II can properly explore this range. This is also in line with the distribution of objective function values shown in the previous section.

It is highly important to point out here that the range of optimum parameters values provided within the multi-objective framework are only a reflection of the population size, as well as the considered objective functions. Consideration of a larger population size and/or other objective functions would definitely help to better explore the search space and obtain more accurate parameter sets. However, there are limitations to population size and objective functions, which are primarily related to computational cost, algorithmic issues, and feasibility or reliability of obtained solutions.

3.3 Uncertainty evaluation

Along with calibration to identify a set of optimal parameter sets giving the best performance of a simulation model, there are other important issues such as uncertainty and identifiability of the parameters which should be taken into account for a proper model evaluation. Among various approaches developed over past years to deal with different sources of uncertainty specifically in RR modelling, multi-objective calibration can be considered as one of these methods, as implied by Gupta et al. (2005), because it takes into account the imperfection of the model structure to reproduce all aspects of hydrograph equally well within a single parameter set. Thus, the outcome is a set of models that are constrained (by the data) to be structurally and functionally consistent with available qualitative and quantitative information and which simulate, in an uncertain way, the observed behaviour of the watershed (Gupta et al., 2005).

Figure 6 shows a graphical comparison between calculated and observed daily flow at Margecany for the year 1991 of the calibration period, and Fig. 7 for the year 2000 of the validation period. Figures 6a and 7a show the model outcome obtained with the optimal parameter sets of NSGA-II, while Figs. 6b and 7b give model results obtained with the optimal parameter sets of PEST. The model results obtained with the 15 Pareto front solutions are shown as a range of simulated discharges as a grey shaded area, and similarly the range of discharges obtained with the two PEST solutions are also shown as a shaded area (generally the upper one is solution 1 and the lower one solution 2). Observed

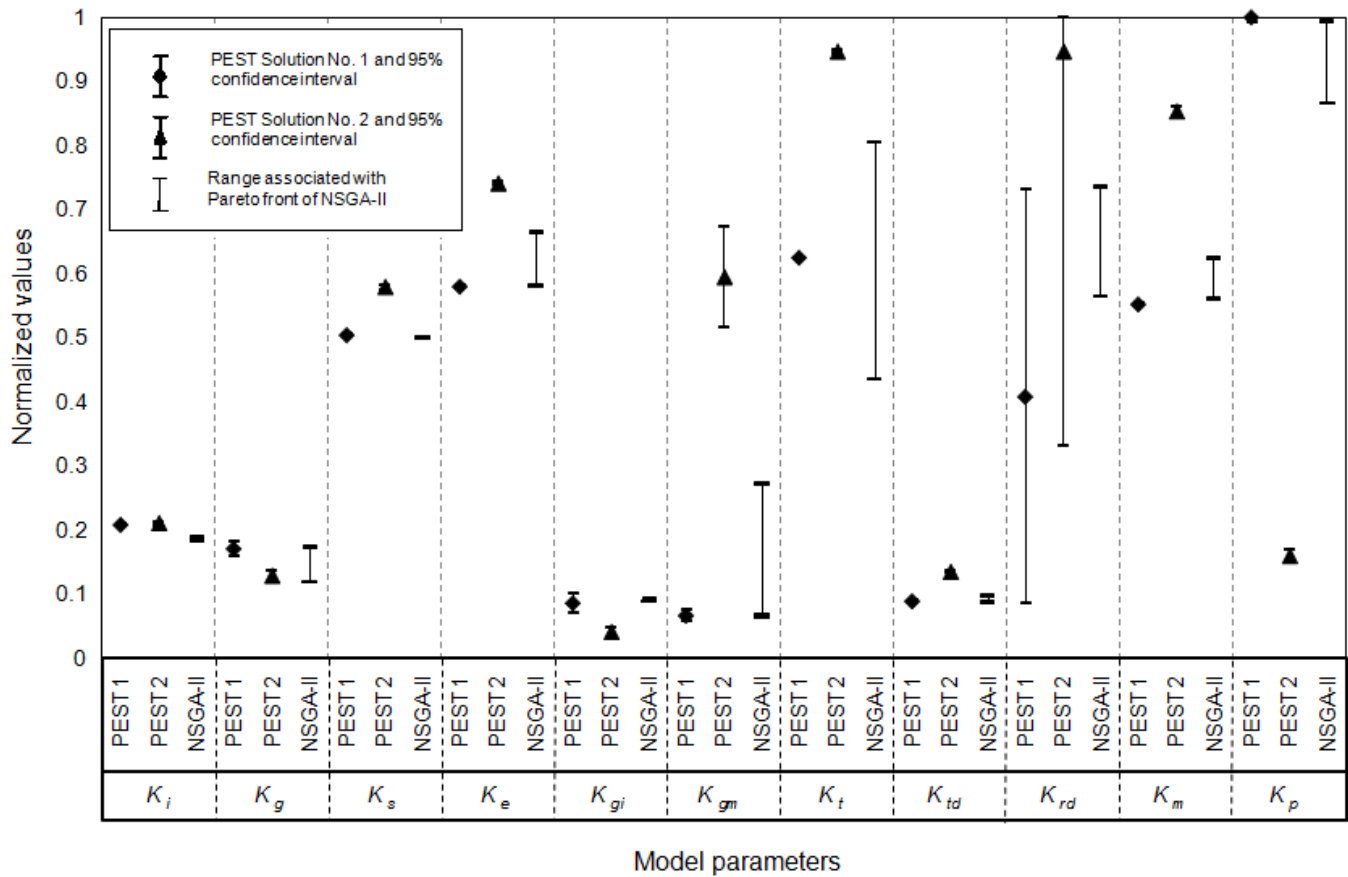


Fig. 5. Normalized optimal parameters values obtained by PEST solution 1 and 2, and their 95% confidence intervals, along with the normalized parameters range of the 15 Pareto front optimal solutions obtained by NSGA-II.

discharges are shown as a dashed line. It is observed that there is consistency between the results of these two approaches and that the ranges are rather narrow. However, both the stream flow estimations in the calibration period and the stream flow predictions in the validation period display systematic errors with respect to the observations. These deviations are likely due to inconsistencies associated with the input data, model structure, and/or inaccuracies in the observed discharges. The limitation of multi-objective calibration approach to take these uncertainties into account is that it does not articulate an identifiable error model, and consequently it becomes difficult to provide uncertainty bounds on the parameter estimates as these are strongly related to specific error models (Kavetski et al., 2002). Hence, as well as improving the calibration routines, it is also required to improve the model structure or to provide suitable methods to appropriately quantify model and parameters uncertainties.

Another interpretation of the model predictions shown in Figs. 6 and 7 might be related to the concept of equifinality introduced by Beven (1993), i.e. the fact that there may be different parameter sets equally suitable to reproduce the observed behaviour of the system. Hence, the hydrograph

ranges obtained by the NSGA-II Pareto front solutions can be a reflection of equifinality in WetSpa calibration. Although it may be argued that this issue is not really a problem for practical models applications, because any of these parameter sets may be applied (Lindstorm, 1997), it is, nevertheless, desirable to address the prediction uncertainty due to these parameter sets (i.e. quantitative analysis of discharge ranges for validation period). Although multi-objective equivalence of parameter sets is different from the probabilistic representation of parameter uncertainty, the Pareto set of solutions defines the minimum uncertainty in the parameters that can be achieved without stating a subjective relative preference for minimizing one specific component of the hydrograph at the expense of another (Vrugt et al., 2003). Combination of deterministic multi-objective calibration (i.e. such as the approach addressed in this study) and probabilistic methods might be a promising approach to analyze different sources of uncertainty. In line with this, research aimed at improving the WetSpa model and development of a methodology to quantify model and parameter uncertainties is ongoing by the authors.

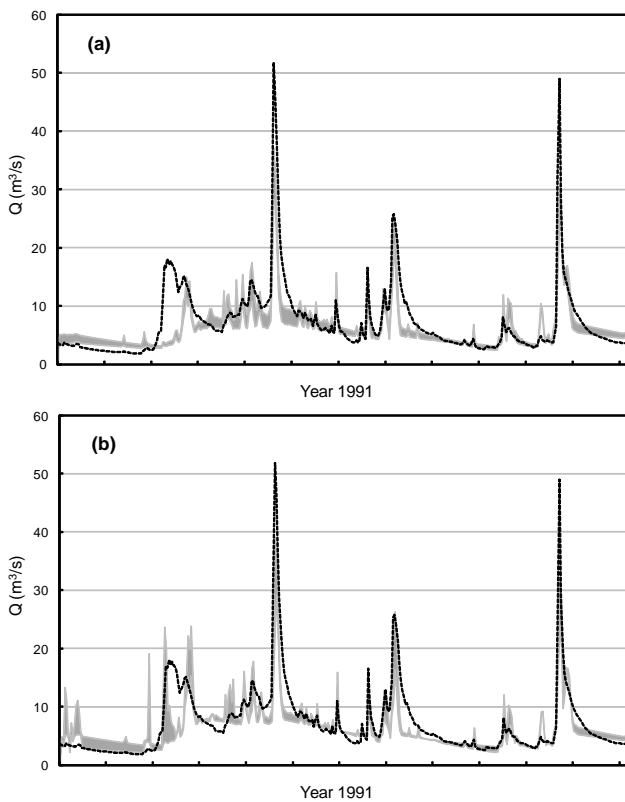


Fig. 6. Observed hydrograph (dashed line), calculated hydrograph with (a) optimal NSGA-II Pareto solutions (shaded area in the upper panel as the range of simulated discharges), and (b) two PEST solutions (shaded area in the lower panel), at Margecany for the year 1991 of the calibration period.

3.4 Identifiability analysis

The purpose of identifiability analysis in RR modelling is the identification of the model structure and a corresponding parameter set that are most representative of the catchment under investigation, while considering aspects such as modelling objectives and available data (Wagener et al., 2001). Assuming a particular model structure, e.g. the WetSpa model addressed in this study, estimation of a suitable parameter set as the result of calibration would complete the model identification process. In order to investigate the identifiability of the WetSpa model's parameters within the multiobjective calibration procedure of this study, the range of parameters values associated with the Pareto front was taken into account. The normalized values of different parameters are depicted in Fig. 8 versus number of NSGA-II iterations, i.e., all parameter values of the solutions contained in Pareto fronts of different iterations are shown. These values are normalized based on their initially preset feasible minimum and maximum values as given in Table 1. Initially, as there is no a priori information about optimal values for each parameter, the values were generated randomly within the feasible pa-

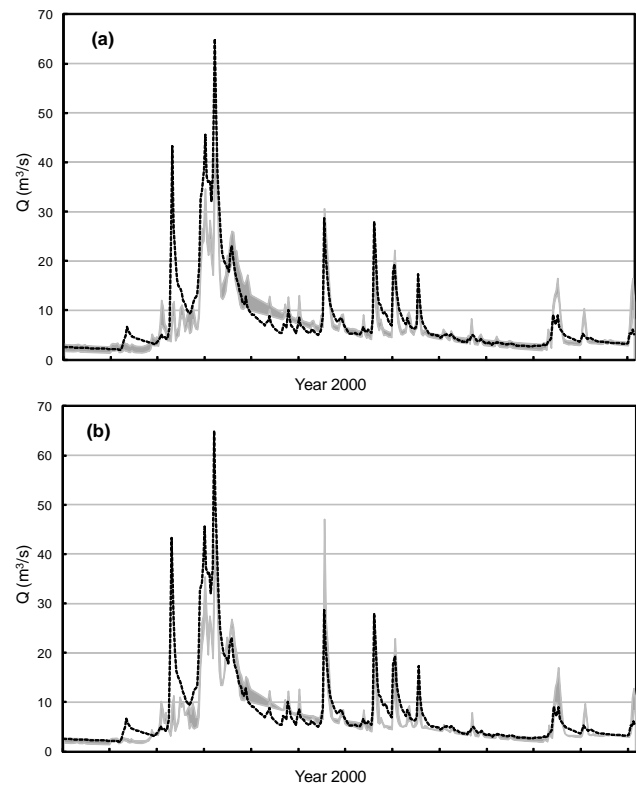


Fig. 7. Observed hydrograph (dashed line), calculated hydrograph with (a) optimal NSGA-II Pareto solutions (shaded area in the upper panel as the range of simulated discharges), and (b) two PEST solutions (shaded area in the lower panel), at Margecany for the year 2000 of the calibration period.

rameters space. But over the iterations, Pareto optimal solutions are obtained with better parameter values, located in optimal regions of the parameter space. As seen in Fig. 8, most WetSpa parameters (i.e. K_i , K_g , K_s , K_e , K_{gi} , K_{td} , and K_m) are well identified because the range of values of the Pareto optimal solutions quickly become much more bounded compared to their initial range. However, some parameters are poorly identifiable (i.e. K_{gm} , K_t , K_{rd} , and K_p) exhibiting ranges that do not converge.

Considering Figs. 6 and 7, the range of Pareto solutions seems to have little impact on the predictive flows. This could be due to the relatively small difference between formulations of objective functions CR_1 and CR_2 translated into the simulations of flow, but also points out that WetSpa is more sensitive to well-identifiable parameters (i.e. these will occupy a relatively small range in the optimal region of the parameter space) than poorly-identifiable parameters.

Table 3 gives the correlation between the different WetSpa model parameters for all Pareto front solutions. The correlation between most of the parameters is typically low, further confirming that most of the WetSpa parameters are well defined. Hence, it can be concluded from the results

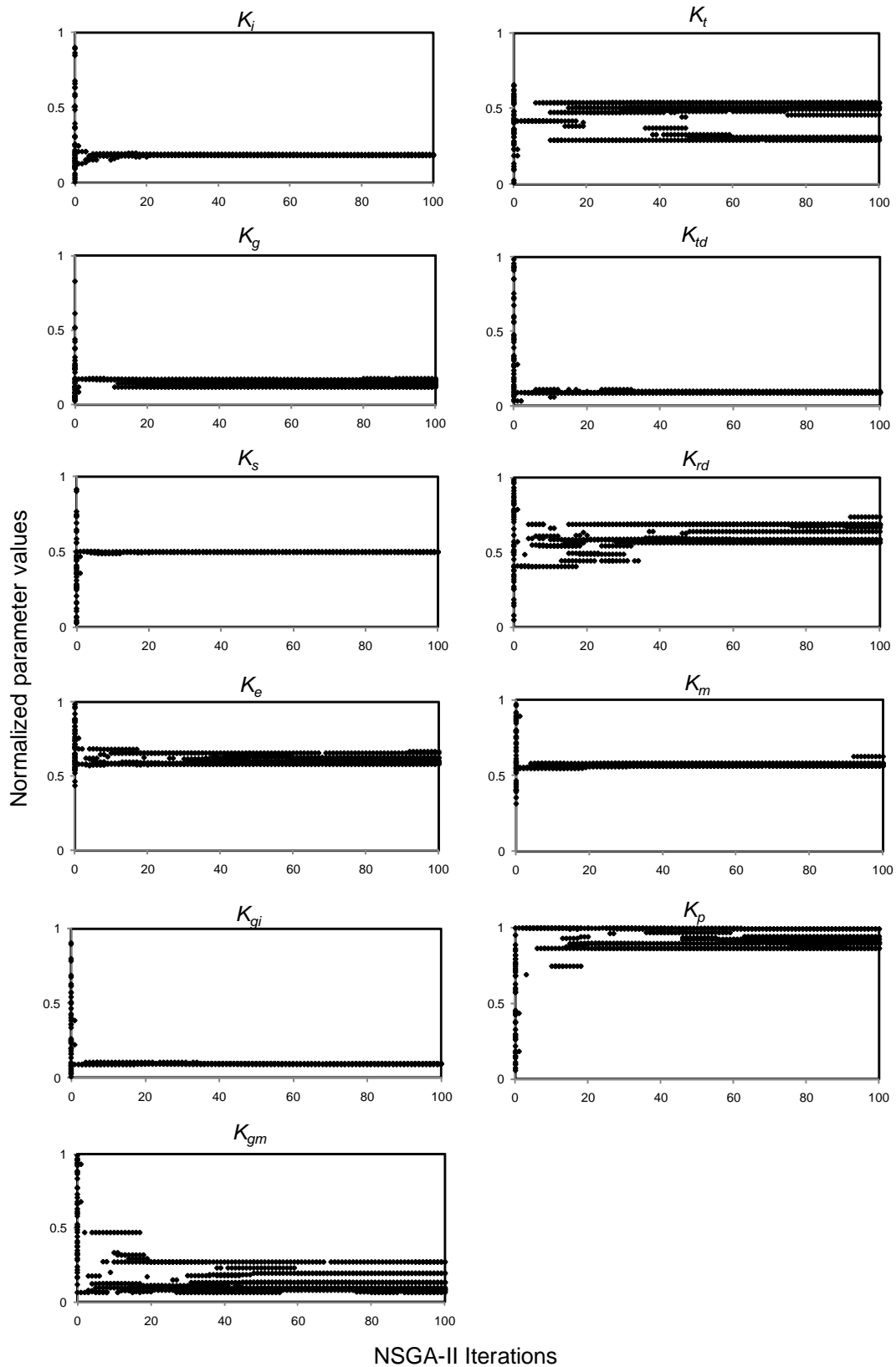


Fig. 8. Plot of normalized values of the WetSpa model parameters versus number of iterations of the NSGA-II search algorithm; shown are all parameter values of all Pareto front solutions through 100 iterations.

Table 3. Correlation between the WetSpa model parameters derived from the Pareto front solutions of all NSGA-II iterations.

Parameter	K_i	K_g	K_s	K_e	K_{gi}	K_{gm}	K_t	K_{td}	K_{rd}	K_m	K_p
K_i	1	0.25	0.20	0.23	0.44	0.12	0.00	0.42	0.12	0.21	0.01
K_g		1	0.34	0.09	0.43	0.05	0.13	0.70	0.05	0.21	0.03
K_s			1	0.02	0.29	0.14	0.01	0.47	0.12	0.08	0.00
K_e				1	0.32	0.57	0.05	0.36	0.08	0.08	0.07
K_{gi}					1	0.17	0.00	0.65	0.09	0.16	0.01
K_{gm}						1	0.04	0.33	0.14	0.00	0.09
K_t							1	0.02	0.00	0.01	0.03
K_{td}								1	0.13	0.14	0.01
K_{rd}									1	0.04	0.00
K_m										1	0.22
K_p											1

presented in Fig. 8 and Table 3 that for this particular watershed and dataset, most WetSpa parameters can be reasonably calibrated using multi-objective formulation. Obviously, this conclusion is also based on particular algorithm, objective functions, and initial solutions used in this paper. Thus, more efforts to define better settings for these items will definitely help to get more insight into parameters identifiability.

As a simple action to extend the identifiability analysis of the WetSpa model parameters, multiple NSGA-II runs were made considering different set of initial solutions, each of which obtained by LHS. The resulting Pareto fronts from different runs are fairly identical in terms of model performance, and furthermore show that most of the optimized parameters are located in the same region of their feasible space. This is in line with the findings of previous paragraph. The parameters which varied the most from one run to the other are the ones that were previously shown to be poorly-definable. Possibly, multi-modality of response surface of the optimization problem and/or the concept of equifinality might have been the reasons of these variations.

4 Conclusions

In this study, a multi-objective genetic algorithm, NSGA-II (Deb et al., 2002), was applied to calibrate a hydrological model (WetSpa). The objective functions were the Nash-Sutcliffe model efficiency (i.e. ability to reproduce all stream-flows), and the model efficiency for log-transformed stream-flows to emphasize low-flow values. The concept of Pareto dominance was used to solve the multi-objective optimization problem. In order to analyze the applicability of the approach and to analyze the impact of multiple objective functions on optimal regions of the parameters space, the single-objective local search technique of PEST (the classical method to calibrate the WetSpa model) was also used to calibrate the model. PEST was applied in two modes: (1) minimizing the sum of squared differences between ob-

served and predicted discharges, and (2) similarly but for log-transformed discharges to enhance the importance of low-flows. Furthermore, we also aimed to assess the identifiability of the model parameters through multi-objective calibration. The two approaches, NSGA-II and PEST, were evaluated through application of the WetSpa model to the Hornad River located in Slovakia.

Based on the objective function values obtained from the NSGA-II and PEST runs, it can be concluded that the multi-objective approach proposed in this paper performs well. Hence, it can be considered as an alternative way to calibrate the model instead of using PEST. Moreover, due to the uniform spread of Pareto front solutions in the objectives space, and also in the parameters space, it is possible for stake-holders to select a particular parameter set based on existing priorities. Hence, multi-objective calibration can provide stake-holders with a proper decision support system.

The obtained results of the identifiability analysis also clearly demonstrate that most of the WetSpa model parameters are well identifiable. For the parameters which are poorly identified, which might be due to multi-modality of the problem, application of more efficient calibration strategies such as multi-population evolutionary algorithms or a combination of these search methods together with mathematical local search procedures might be highly useful, as for instance the AMALGAM multi-objective evolutionary search strategy of Vrugt and Robinson (2007). Research aimed at further improvement of the optimization approach proposed in this study is also ongoing.

According to literature on multi-objective calibration and uncertainty analysis as well as what was shown in this paper, this approach can define a minimum level of uncertainty associated with the model structure. This uncertainty is shown in terms of parameters ranges of the Pareto front solutions, and/or a band of model simulations. Nevertheless, it would be desirable to adopt a more robust methodology to quantify different sources of uncertainty such as input, parameters and

model uncertainties. Possibly, a combination of probabilistic principles and multi-objective evolutionary algorithms might deal with this issue. In this respect, some approaches might be (i) to use Markov Chain Monte Carlo samplers such as MOSCEM-UA (Vrugt et al., 2003) to estimate the Pareto front, or (ii) to adopt the Probabilistic Multi-Objective Genetic Algorithm (PMOGA) proposed by Singh et al. (2008) for rainfall-runoff calibration.

Acknowledgements. The authors thank two anonymous reviewers and the associate editor for providing constructive comments that greatly improved this paper.

Edited by: J. Vrugt

References

- Ajami, N. K., Gupta, H. V., Wagener, T., and Sorooshian, S.: Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system, *J. Hydrol.*, 298, 112–135, 2004.
- Baginska, B., Milne-Home, W., and Cornish, P.: Modelling nutrient transport in Currency Creek, NSW with AnnAGNPS and PEST, *Environ. Model Softw.*, 18, 801–808, 2003.
- Bahreman, A., De Smedt, F., Corluy, J., and Liu, Y. B.: WetSpa Model Application for Assessing Reforestation Impacts on Floods in Margecany-Hornad Watershed, Slovakia, *Water Resour. Manag.*, 21, 1373–1391, 2007.
- Bekele, E. G. and Nicklow, J. W.: Multi-objective automatic calibration of SWAT using NSGA-II, *J. Hydrol.*, 341(3–4), 165–176, 2007.
- Beven, K. J.: Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, 16, 41–51, 1993.
- Beven, K. J. and Binley, A. M.: The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Processes*, 6, 279–298, 1992.
- Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, *Water Resour. Res.*, 36, 3663–3674, 2000.
- Cheng, C. T., Ou, C. P., and Chaw, K. W.: Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration, *J. Hydrol.*, 268(1–4), 72–86, 2002.
- Cieniawski, S. E., Eheart, J. W., and Ranjithan, S. R.: Using genetic algorithms to solve a multiobjective groundwater monitoring problem, *Water Resour. Res.*, 31, 399–409, 1995.
- Coello, C. C.: An updated survey of GA-based multiobjective optimization techniques, *ACM Comput. Surv.*, 32(2), 109–143, 2000.
- Cunha, A. G., Oliviera, P., and Covas, J.: Use of genetic algorithms in multicriteria optimization to solve industrial problems, in: *Proceedings of the Seventh International Conference on Genetic Algorithms*, edited by: Bäck, T., San Francisco, California, 682–688, Morgan Kaufmann, 1997.
- De Smedt, F., Liu, Y. B., Gebremeskel, S., Hoffmann, L., and Pfister, L.: Application of GIS and remote sensing in flood modeling for complex terrain, in: *GIS and Remote Sensing in Hydrology*, edited by: Chen, Y. B., Takara, K., Cluckie, I., and De Smedt, F., *Water Resources and Environment*, IAHS Publ., 289, 23–32, 2004.
- De Smedt, F., Liu, Y. B., and Gebremeskel, S.: Hydrologic modeling on a basin scale using GIS and remote sensed land use information, in: *Risk Analysis II*, edited by: Brebbia, C. A., 295–304, WTI press, Southampton, Boston, 2000.
- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.: A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II, *IEEE Trans. Evol. Computation*, 6(2), 182–197, 2002.
- Doherty, J. and Johnston, J. M.: Methodologies for calibration and predictive analysis of a watershed model, *J. Am. Water Resour. Assoc.*, 39, 251–265, 2003.
- Fonseca, C. and Fleming, P. J.: An overview of evolutionary algorithms in multiobjective optimization, *Evolutionary Comput.*, 3(1), 1–16, 1995.
- Fourman, M. P.: Compaction of symbolic layout using genetic algorithms, in: *Proceedings of an International Conference on Genetic Algorithms and Their Applications*, edited by: Grefenstette, J. J., Pittsburgh, PA, 141–153, sponsored by Texas Instruments and US Navy Center for Applied Research in Artificial Intelligence (NCARAI), 1985.
- Gen, M. and Cheng, R.: *Genetic algorithms and engineering optimization*, New York, John Wiley & Sons, 2000.
- Goldberg, D. E.: *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Co., Inc., Redwood City, CA, USA, 1989.
- Goldberg, D. E. and Deb, K.: A comparative analysis of selection schemes used in genetic algorithms, in: *Foundations of Genetic Algorithms*, edited by: Rawlins, G. J. E., Morgan Kaufmann, San Mateo, CA, 1991.
- Gupta, H. V., Beven, K. J., and Wagener, T.: Model Calibration and Uncertainty Estimation, *Encyclopedia of Hydrological Sciences*, Part 11, Rainfall-Runoff Modeling, John Wiley & Sons, Ltd, 2005.
- Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrological models: multiple and noncommensurable measures of information, *Water Resour. Res.*, 34(4), 751–763, 1998.
- Iman, R. L. and Conover, W. J.: Small sample sensitivity analysis techniques for computer models, with an application to risk assessment, *Communication in Statistics*, A9(17), 1749–1842, 1980.
- Ishibuchi, H. and Murata, T.: Multi-objective genetic local search algorithm, in: *Proceedings of 1996 IEEE International Conference on Evolutionary Computation (ICEC'96)*, Piscataway, NJ, 119–124, 1996.
- Kavetski, D., Franks, S. W., and Kuczera, G.: Confronting input uncertainty in environmental modeling, in: *Calibration of Watershed Models*, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., AGU Water Science and Applications Series, 6, 49–68, 2002.
- Khu, S. T., Savic, D., and Liu, Y.: Evolutionary-based multi-objective meta-model approach for rainfall-runoff model calibration, *Geophys. Res. Abstr.*, 7, 09858, 2005.
- Lindstorm, G.: A simple automatic calibration routine for the HBV model, *Nordic Hydrology*, 28, 153–168, 1997.
- Liu, Y. B., Gebremeskel, S., De Smedt, F., Hoffmann, L., and Pfister, L.: A diffusive transport approach for flow routing in GIS-based flood modelling, *J. Hydrol.*, 283, 91–106, 2003.

- Liu, Y. B., De Smedt, F., Hoffmann, L., and Pfister, L.: Parameterization using ArcView GIS in medium and large watershed modelling, in: *GIS and Remote Sensing in Hydrology*, edited by: Chen, Y. B., Takara, K., Cluckie, I., and De Smedt, F., *Water Resources and Environment, IAHS Publ.*, 289, 50–58, 2004.
- Liu, Y. B., Batelaan, O., De Smedt, F., Poorova, J., and Velcicka, L.: Automated calibration applied to a GIS-based flood simulation model using PEST, in: *Floods, from Defence to Management*, edited by: van Alphen, J., van Beek, E., and Taal, M., Taylor & Francis Group, London, 317–326, 2005.
- Liu, Y. B. and De Smedt, F.: *WetSpa extension, documentation and user manual*, Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Belgium, 2004.
- Liu, Y. B. and De Smedt, F.: Flood modeling for complex terrain using GIS and remote sensed information, *Water Resour. Manag.*, 19(5), 605–624, 2005.
- Madsen, H.: Parameter estimation in distributed hydrological catchment modeling using automatic calibration with multiple objectives, *Adv. Water Resour.*, 26, 205–216, 2003.
- Madsen, H.: Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, *J. Hydrol.*, 235, 276–288, 2000.
- Michalewicz, Z.: *Genetic Algorithms + Data Structures = Evolution Programs*, AI Series, Springer-Verlag, New York, 1994.
- Muleta, M. and Nicklow, J.: Decision Support for Watershed Management Using Evolutionary Algorithms, *J. Water Res. Pl.*, 131, 35–44, 2005a.
- Muleta, M. K. and Nicklow, J. W.: Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model, *J. Hydrol.*, 306, 127–145, 2005b.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models. Part 1: A discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.
- Osyczka, A.: *Multicriteria Optimization for Engineering Design*, Academic Press, 1985.
- Pareto, V.: *The Rise and Fall of the Elites*, Totowa, N. J., Bedminster Press, 1968.
- Reed, P. and Minsker, B. S.: Striking the Balance: Long-Term Groundwater Monitoring Design for Conflicting Objectives, *J. Water Res. Pl.*, 130, 140–149, 2004.
- Reed, P., Minsker, B. S., and Goldberg, D. E.: A multiobjective approach to cost effective long-term groundwater monitoring using an Elitist Nondominated Sorted Genetic Algorithm with historical data, *J. Hydroinform.*, 3, 71–90, 2001.
- Ritzel, B. J., Eheart, J. W., and Ranjithan, S. R.: Using genetic algorithms to solve a multiple objective groundwater pollution containment problem, *Water Resour. Res.*, 30, 1589–1603, 1994.
- Schaffer, J. D.: Multiple objective optimization with vector evaluated genetic algorithms, in: *Proceedings of an International Conference on Genetic Algorithms and Their Applications*, edited by: Grefenstette, J. J., Pittsburgh, PA, 93–100, sponsored by Texas Instruments and US Navy Center for Applied Research in Artificial Intelligence (NCARAI), 1984.
- Seibert, J.: Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, *Hydrol. Earth Syst. Sci.*, 4, 215–224, 2000, <http://www.hydrol-earth-syst-sci.net/4/215/2000/>.
- Singh, A. and Minsker, B. S.: Uncertainty-based multiobjective optimization of groundwater remediation design, *Water Resour. Res.*, 44, W02404, doi:10.1029/2005WR004436, 2008.
- Srinivas, N. and Deb, K.: Multiobjective optimization using nondominated sorting in genetic algorithms, *Tech. Rep.*, Indian Institute of Technology, Delhi, India, 1993.
- Syvoloski, G., Kwicklis, E., Eddebarh, A. A., Arnold, B., Faunt, C., and Robinson, B.: The site-scale saturated zone flow model for Yucca Mountain: calibration of different conceptual models and their impact on flow paths, *J. Contam. Hydrol.*, 62–63, 731–750, 2003.
- Tang, Y., Reed, P., and Wagener, T.: How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration?, *Hydrol. Earth Syst. Sci.*, 10, 289–307, 2006, <http://www.hydrol-earth-syst-sci.net/10/289/2006/>.
- Valenzuela-Rend'on, M. and Uresti-Charre, E.: A nongenerational genetic algorithm for multiobjective optimization, in: *Proc. 7th Int. Conf. Genetic Algorithms*, edited by: Bäck, T., San Francisco, CA, Morgan Kaufmann, 658–665, 1997.
- Vrugt, J., 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., Schoups, G., Hopmans, J. W., Young, C., Wallender, W. W., and Harter, T.: Inverse modeling of large-scale spatially distributed vadose zone properties using global optimization, *Water Resour. Res.*, 41, W06003, doi:10.1029/2004WR003698, 2005.
- Vrugt, J. A. and Robinson, B. A.: Improved evolutionary optimization from genetically adaptive multimethod search, *Proc. Natl. Acad. Sci. USA*, 104, 708–711, doi:10.1073/pnas.0610471104, 2007.
- Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., and Sorooshian, S.: A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, 5, 13–26, 2001, <http://www.hydrol-earth-syst-sci.net/5/13/2001/>.
- Wang, Z., Batelaan, O., and De Smedt, F.: A distributed model for Water and Energy Transfer between Soil, Plants and Atmosphere (WetSpa), *Phys. Chem. Earth* 21(3), 189–193, 1997.
- Wöhling, Th., Barkle, G. F., and Vrugt, J. A.: Comparison of three multiobjective optimization algorithms for inverse modeling of vadose zone hydraulic properties, *Soil Sci. Soc. Am. J.*, 72(2), 305–319, 2008.
- Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-objective global optimization of hydrological models, *J. Hydrol.*, 204, 83–97, 1998.
- Zitzler, E., Deb, K., and Thiele, L.: Comparison of Multiobjective Evolutionary Algorithms: Empirical Results, *Evolutionary Comput.*, 8(2), 173–195, 2000.
- Zitzler, E. and Thiele, L.: Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach, *IEEE Trans. Evol. Computation*, 3(4), 257–271, 1999.