Hydrology and Earth System Sciences

Comparison of different multi-objective calibration criteria using a conceptual rainfall-runoff model of flood events

R. Moussa[1] and N. Chahinian[2]

[1]INRA, Laboratoire d'étude des Interactions entre Sol – Agrosystème – Hydrosystème, UMR LISAH SupAgro-INRA-IRD,
2 Place Pierre Viala, 34060 Montpellier, France
[2]HydroSciences Montpellier, Université Montpellier 2, Case courrier MSE, Place Eugène Bataillon,
34095 Montpellier Cedex 5, France

Abstract. A conceptual lumped rainfall-runoff flood event model was developed and applied on the Gardon catchment located in Southern France and various single-objective and multi-objective functions were used for its calibration. The model was calibrated on 15 events and validated on 14 others. The results of both the calibration and validation phases are compared on the basis of their performance with regards to six criteria, three global criteria and three relative criteria representing volume, peakflow, and the root mean square error. The first type of criteria gives more weight to large events whereas the second considers all events to be of equal weight. The results show that the calibrated parameter values are dependent on the type of criteria used. Significant trade-offs are observed between the different objectives: no unique set of parameters is able to satisfy all objectives simultaneously. Instead, the solution to the calibration problem is given by a set of Pareto optimal solutions. From this set of optimal solutions, a balanced aggregated objective function is proposed, as a compromise between up to three objective functions. The single-objective and multi-objective calibration strategies are compared both in terms of parameter variation bounds and simulation quality. The results of this study indicate that two well chosen and non-redundant objective functions are sufficient to calibrate the model and that the use of three objective functions does not necessarily yield different results. The problems of non-uniqueness in model calibration, and the choice of the adequate objective functions for flood event models, emphasise the importance

of the modeller's intervention. The recent advances in automatic optimisation techniques do not minimise the user's responsibility, who has to choose multiple criteria based on the aims of the study, his appreciation on the errors induced by data and model structure and his knowledge of the catchment's hydrology.

1 Introduction

It is common for hydrologists to model individual storm events at the catchment scale (e.g. Bates and Ganeshanandam 1990; Zarriello, 1998; Moussa et al., 2002; Jain and Indurthy, 2003), for flood forecasting, spillway design or flood protection schemes. The first important challenge that awaits the modeller in this task is to choose a rainfall-runoff model, and to calibrate a set of parameters, that can accurately simulate a number of flood events and related hydrographs shapes. Most of the models used currently for flood forecasting in France are lumped conceptual models (Garçon, 1996; Yang and Michel, 2000; Paquet, 2004) i.e. they have parameters which cannot, in general, be obtained directly from measurable catchment characteristics, and hence model calibration is needed. Various calibration algorithms and procedures have been presented in the literature extensively (Rosenbrock, 1960; Nelder and Mead 1965; Duan et al., 1992; Gan and Biftu, 1996; Yapo et al., 1998; Vrugt et al. 2003a and b; see a review in Gupta et al., 2003). Although they differ in the ways they seek the optimal value, they all aim at minimising or maximising an objective function. It is important to note that, in general, trade-offs exist between the different objective functions. For instance, when using the

Correspondence to: R. Moussa
(moussa@supagro.inra.fr)

bias as an objective function, one may find a set of parameters that provides a very good simulation of volume, but a poor simulation of the hydrograph shape or peak flow, and vice versa. Conventional objective functions such as the root mean square error, the Nash and Sutcliffe (1970) efficiency coefficient, or the index of agreement (Willmott, 1981) tend to emphasize the high flows, and consequently, are oversensitive to extreme values and outliers (Legates and McCabe, 1999). On the opposite, the mean absolute percent error tends to emphasize the low flows (Yu and Yang, 2000).

However, in most real world applications, models are used to reproduce the entire shape of the hydrograph, sometimes even to simulate more than one flow component in the catchment i.e. groundwater and surface water, water and nutrient fluxes etc. In these occurrences, the use of a single objective function may be questionable and it would be advisable to take into account various objective functions by considering the calibration problem in a multi-objective framework (Yapo et al., 1998; Madsen, 2000). After Pareto (1906) introduced the concept of non-inferior solutions in the context of economics, Zadeh (1963) and Stadler (1979) began to apply the notion of Pareto optimiality to the fields of engineering. Over the last three decades tremendous amount of research effort has been expanded on multi-objective decision making leading to the publication of many interesting results in the literature (Changkong and Haimes, 1983; Miettinen, 1999; Ehrgott, 2005). One of the most widely used methods for solving multi-objective optimisation problems is to transform a multi-objective problem into a series of single-objective problems. When an appropriate set of solutions is obtained by the single-objective optimisations, the solutions can approximate a Pareto front (case of a two-objective function) or Pareto surface (case of three-objective functions). In hydrology, most of the studies related to multi-objective calibration have investigated the use of two-objective functions and few ones have looked into the use of three or more functions (Madsen et al., 2002; Schoops et al., 2005; Parajka et al., 2007). In this study we will use the multi-objective calibration approach with three objective functions as suggested by Madsen (2000) and Madsen et al. (2002). In addition, while most of these studies have dealt with continuous simulations, we will investigate multi-objective calibration issues related to event based modelling.

Hence, the objective of this paper is to develop an event based lumped conceptual rainfall-runoff model and then to compare single-objective and multi-objective calibration approaches. The emphasis is put on the impact of the selected objective functions on the actual hydrograph shapes and simulation errors rather than on the calibration algorithm as several studies have already investigated this issue (Yu and Yang, 2000; Johnsen et al., 2005; Tang et al., 2006). The Gardon catchment located in southern France is used in the applications because of the recurrent flooding problems its inhabitants encounter yearly. The objective functions used in this study can be divided in two broad categories: global
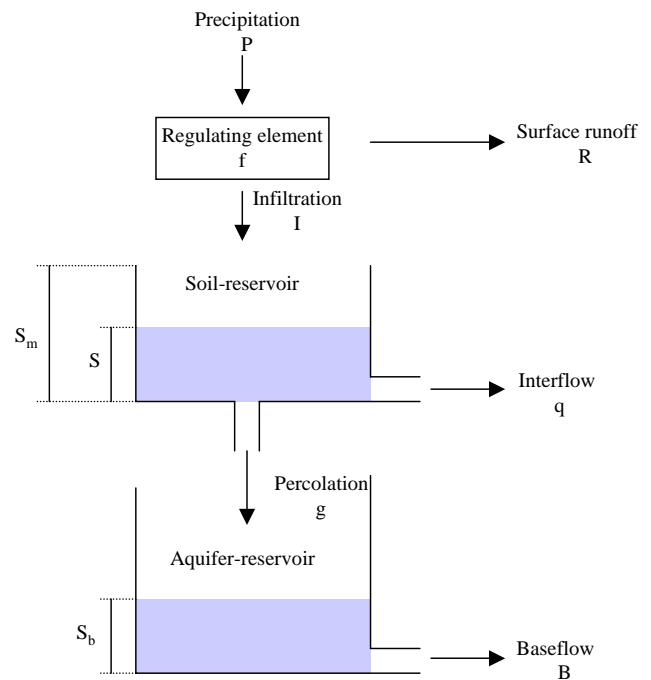

Fig. 1. Schematic of the model production function.

and relative. Given the diversity of the flood events to be modelled such an approach was deemed necessary as the first type of criteria gives more weight to large events whereas the second considers all events to be of equal weight. For each category, three different objective functions are considered: volume conservation which is important for gauging problems, peakflow reproduction which is essential for flood applications and the root mean square error as a measure of the global agreement between the simulated and observed curves for high flows. The paper is organised in four sections: i) model presentation; ii) formulation of the calibration criteria; iii) presentation of the study zone, and iv) model performance analysis based on single-objective and multi-objective calibration.

2 The lumped conceptual rainfall-runoff model

Since the 1960s, lumped, conceptual rainfall-runoff models have been used in hydrology (e.g. Crawford and Linsley, 1966; Cormary and Guilbot, 1969; Duan et al., 1992; Bergström, 1995; Donigan et al., 1995; Havnø et al., 1995). These models consider the catchment as an undivided entity, and use lumped values of input variables and parameters. For the most part (for a review, see Fleming, 1975; Singh, 1995), they have a conceptual structure based on the interaction between storage elements representing the different processes, with mathematical functions to describe the fluxes between the stores.

The modelling approach followed herein will be lumped and the catchment will be considered as a single entity. A two-reservoir-layer model has been developed to represent the catchment on the basis of the CREC model (Cormary and Guilbot, 1969) and the Diskin and Nazimov (1995) production function (Fig. 1). Evaporation is not represented since the purpose of the model is to simulate individual flood events during which evaportanspiration is negligible. The first layer, denoted "soil-reservoir", represents the upper soil layer and controls surface runoff, infiltration, interflow and percolation. The second layer, denoted "aquifer-reservoir", represents the aquifer where mainly base flow occurs. A unit hydrograph transfer function is used to route flows to the outlet. The output of the model will be a simulated hydrograph which will be compared to the original measured hydrograph to assess model performance. A general description of each procedure is given below.

2.1 The production function

A regulating element f separates the precipitation P into surface runoff R, and infiltration I. The soil-reservoir element has one input, the infiltration I, and two outputs, the lateral interflow q, and the vertical flow g which represents the percolation from the upper soil layer to deeper layers. The state variable of the regulating element, denoted f, is determined by the magnitude of the reservoir state variable S according to the Diskin and Nazimov (1995) relationship

$$f = f_0 + (f_c - f_0) \frac{S}{S_m} \qquad (1)$$

where f_0 [ms^{-1}] is the maximum infiltration capacity (corresponding to S=0), f_c [ms^{-1}] the minimum infiltration capacity (corresponding to $S=S_m$), and S_m [m] the maximum storage in the soil-reservoir layer. The value of f_c characterises the soil's infiltration capacity at saturation, and the term (S/S_m) characterises the relative soil moisture. The two outputs I and R of the regulating element depend on the value of the state variable f and on the value of the input P, at the same instant according to the following equations

If $P < f$ then $I=P$ and $R=0$ \qquad (2)

If $P > f$ then $I=f$ and $R=P-f$ \qquad (3)

The two outputs of the soil-reservoir, q and g, are calculated as function of a parameter b (with $0 \le b \le 1$) and the term (S/S$_m$)

$$q=f_c b \frac{S}{S_m} \text{ and } g=f_c (1-b) \frac{S}{S_m} \qquad (4)$$

It should be noted that the sum $q+g=f_c \frac{S}{S_m}$ is independent of the parameter b and verifies the soil-reservoir output of the Diskin and Nazimov (1995) model. As the storage S approaches the threshold value S$_m$, both the infiltration capacity f and the sum q+g tend to the same value f_c.

The aquifer-reservoir has one input, the percolation g, and one output, the base flow B [ms^{-1}], which is calculated as function of the aquifer-reservoir level S$_b$ using a linear relation

$$B=kS_b \qquad (5)$$

where k [s^{-1}] is a constant characterizing the recession curve of the aquifer. In order to reduce the number of parameters, the aquifer-reservoir does not have a maximum storage depth. The value of the state variable S_b of the aquifer-reservoir is obtained using the continuity equation

$$\frac{dS_b}{dt} = g(t) - B(t) \qquad (6)$$

2.2 The transfer function

A transfer function is used to route the rainfall excess to the catchment outlet. A unit hydrograph linear model, based on a Hayami (1951) kernel function, which is an approximation of the diffusive wave equation, was used to simulate the transfer of the sum of (R + q + B) to the outlet (Moussa and Bocquillon, 1996). Let I(t) [m^{3}s^{-1}] be the input hydrograph

$$I(t) = (R + q + B) A \qquad (7)$$

where A [m^2] is the catchment area. Let $O(t)$ be the routed hydrograph at the outlet

$$O(t) = \int_0^t I(\tau) H(t - \tau) d\tau \qquad (8)$$

with H(t) the Hayami kernel function defined as

$$H(t) = \left(\frac{wz}{\pi}\right)^{\frac{1}{2}} \frac{\exp^{z\left(2 - \frac{t}{w} - \frac{w}{t}\right)}}{(t)^{3/2}} \text{ with } \int_0^\infty H(t)dt = 1 \qquad (9)$$

where w[s] is a time parameter that represents the centre of gravity of the unit hydrograph, z [dimensionless] a form parameter, π=3.1416 and t the time [s]. For low values of z (i.e. z=1, 2 or 5), the unit hydrograph represents both translation and diffusivity (approximation of the diffusive wave equation), while for high values of z (i.e. z=20, 50 or 100), the unit hydrograph tends to represent only a translation equal to w (approximation of the kinematic wave equation).

2.3 Model properties and parameters

The input rainfall P is usually given as a function of time in the form of a histogram using a fixed time interval. Consequently, the other variables are also presented as functions of time, and the computations are carried out for the same fixed time interval. The regulating element f and the soil-reservoir element are linked by a feedback path transmitting information about the state of the storage element to the regulating element. The regulating element is related to the soil-reservoir element by the fact that one of its outputs is the

input of the soil-reservoir element. It is also related to the transfer function by the fact that its output R is one input of the transfer function.

Computations start at an instant adopted as zero time $t=0$, with a known, or an assumed, initial value of the soil-reservoir S_0 and the aquifer-reservoir S_{b0} at the start of that time interval. Initial conditions can considerably influence modelling results. However, our model is more sensitive to the ratio between the initial and maximum storage capacities than to the value of the initial condition. The 5-day antecedent rainfall index is frequently used in hydrological modelling (i.e. Chevallier, 1983; Fedora and Beschta, 1989; Heggen, 2001; Peugeot et al., 2003; Brocca et al., 2008). Although some studies have used wider time spans to calculate this index (Anctil et al., 2004), the 5-day rainfall is considered as a standard. For each flood event, the value of S_0 is defined according to the 5-day antecedent rainfall R_{5d} corresponding to three classes representing dry, normal and wet soil moisture initial conditions as suggested by the Soil Conservation Service-USDA (1972) method.

$$\text{If } R_{5d} < 10 \text{ mm then } S_0 = 0.25 \, S_m \tag{10}$$

$$\text{If } 10 < R_{5d} < 30 \text{ mm then } S_0 = 0.50 \, S_m \tag{11}$$

$$\text{If } R_{5d} > 30 \text{ mm then } S_0 = 0.75 \, S_m \tag{12}$$

At the beginning of the rainfall event, the measured discharge Qo(0) at $t=0$ is the sum of the lateral interflow q and the base flow B. Using Eq. (4) and (5), and substituting S by S_0, the initial value of the aquifer-reservoir S_{b0} is calculated as

$$S_{b0} = \frac{1}{k} \left(\frac{Qo(0)}{A} - f_c b \frac{S_0}{S_m} \right) \tag{13}$$

For each time interval, the three state variables $f(t)$ which separates rainfall into surface runoff and infiltration, the level $S(t)$ in the soil-reservoir and the level $S_b(t)$ in the aquifer-reservoir are calculated from the known values of the variables at the beginning of the time interval and the rainfall input to the model during the interval. The values of the other variables at the end of the computation time interval are derived from the value of the three state variables by using the equations above.

The model needs : i) five parameters for the production function, the minimum value of the infiltration capacity f_c, a coefficient "a" that relates the maximum and minimum infiltration capacities i.e. $f_0 = a f_c$ (with $a > 1$), the maximum level of the soil-reservoir S_m, the parameter b of the lateral interflow, and the parameter k of the aquifer-reservoir's recession curve, ii) two parameters for the transfer function, the lag time w and the shape parameter z and iii) two initial conditions S_0 and S_{b0} calculated as function of the 5-day antecedent rainfall R_{5d} and the measured discharge Qo(0) at $t=0$.

3 Formulation of calibration criteria

The objective of model calibration is to select parameter values so that the model simulates the measured hydrograph as closely as possible. Our aim is to consider multiple objectives that measure different aspects of the hydrological response (Madsen, 2000): i) a good agreement between the average simulated and observed runoff volume (i.e. a good water balance); ii) a good overall agreement of the hydrograph shape; iii) a good agreement of the peak flows.

When a calibration procedure is used, the quality of the final model parameters will depend on the structure of the model, the power of the optimisation algorithm, the quality of the input data, and the estimation criteria or objective functions used in the optimisation procedure (Gan and Biftu, 1996). It is beyond the scope of this paper to address all the above factors. We will focus on the last one only and discuss the definition of objective functions and multi-objective calibration procedures.

3.1 The objective functions

The objective functions used in this study include both relative and absolute error measures as suggested by Legates and McCabe (1999). The selected criteria can be divided in two broad categories: "global" which gives more weight to larger flood events and "relative" which considers all events to be of equal weight. For each category, three different objective functions were considered: volume conservation, the root mean square error (RMSE) (which is related by a one-to-one relationship to the widely used Nash and Sutcliffe (1970) efficiency measure) and the peakflow prediction:

1. The global volume error V_g and the relative volume error V_r

$$V_g = \left| \frac{\sum_{i=1}^{N} (Ls_i - Lo_i)}{\sum_{i=1}^{N} n_i} \right| \text{ and } V_r = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Ls_i - Lo_i}{Lo_i} \right| \tag{14}$$

where N is the total number of flood events used for calibration, i an index representing a flood event ($1 \leq i \leq N$), and for each flood event i: n_i the number of time steps, Lo_i the observed runoff depth and Ls_i the simulated runoff depth.

2. The global root mean square error RMSE_g and the relative root mean square error RMSE_r

$$\text{RMSE}_g = \left[\frac{\sum_{i=1}^{N} \sum_{j=1}^{n_i} (Qo_{ij} - Qs_{ij})^2}{\sum_{i=1}^{N} n_i} \right]^{1/2} \text{ and} \tag{15}$$

$$\text{RMSE}_r = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{1}{n_i} \sum_{j=1}^{n_i} (Qo_{ij} - Qs_{ij})^2 \right]^{\frac{1}{2}}$$

where n_i is the number of time steps in the flood event i, j is an index representing the time step in a flood event i ($1 \leq j \leq n_i$), Qo_{ij} the observed discharge at time j in the flood event i and Qs_{ij} the simulated discharge at time j on the flood event i.

3. The global peakflow P_g and the relative peakflow P_r

$$P_g = \frac{1}{N} \sum_{i=1}^{N} \left| (Qxs_i - Qxo_i) \right| \text{ and } P_r = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Qxs_i - Qxo_i}{Qxo_i} \right| \quad (16)$$

where Qxo_i is the observed peak flow of discharge in the flood event i and Qxs_i is the simulated peak flow of discharge in the flood event i.

The six objective functions, V_g, V_r, RMSE_g, RMSE_r, P_g and P_r, are positive functions, and the optimum value of the parameters corresponds to the minimum value of each "0". A single-objective calibration procedure was undertaken separately with each of the six criteria. Most model calibration procedures suffer from the same problems, namely the existence of multiple optima and the presence of high interaction or correlation between subsets of fitted model parameters. A coupled manual and automatic calibration procedure, similar to adaptive cluster based methods (Solomatine, 1999) was used. In the first phase, the manual method is used on the basis of expert knowledge of the hydrological model structure and the catchment characteristics (rainfall, discharge, events type). The manual calibration method is based on the results of previous sensitivity analysis, and uses a trial and error method to obtain a set of parameters for each flood event separately (Moussa, 1991; Moussa et al., 2007). The manual calibration is followed by a collective calibration using all the flood events simultaneously (Chahinian et al., 2005, 2006). Through this first phase, the lower and higher limits of each parameter range are obtained. The second phase consists in an automatic calibration procedure based on model simulations using a progressively finer grid in space. The second phase aims at exploring the whole space of parameters, especially around the optimum obtained during the first phase. The initial grid is obtained by subdividing the interval of variation of each parameter into steps, not necessarily regular, on the basis of the results of the first phase. Different local minima can appear; around which, smaller grid steps are defined. This approach enables the exploration of a larger space, but is limited by the difficulties due to non-linearities and thresholds in the model which can produce local minima in the objective functions space. The procedure is repeated iteratively and guides the grid adaptation to focus on more promising regions in the space of parameters. The number of simulations was limited to 50000 runs. In the applications, this procedure was first used for single-objective calibration runs.

3.2 Multi-objective calibration procedures

When using multiple objectives, the calibration problem can be stated as follows (Madsen, 2000)

$$\min \{ F_1(\theta), F_2(\theta), ..., F_m(\theta) \} \quad \text{with} \quad \theta \epsilon \Theta \quad (17)$$

where $F_i(\theta)$ (i=1, 2 ..., m) are the different objective functions. The optimisation problem is constrained because θ is restricted to the feasible parameter space Θ. The parameter space is usually defined as a hypercube by specifying lower and upper limits on each parameter. These limits are chosen according to physical and mathematical constraints, information about physical characteristics of the system, and from modelling experiences (Kuczera, 1997; Madsen, 2000, 2003; Madsen et al., 2002).

The solution of Eq. (17) will not, in general, be a single unique set of parameters but will consist of the so-called Pareto set of solutions according to different trade-offs between the different objectives (Gupta et al., 2003). The parameter space can be divided into "good" (Pareto optimal) and "bad" solutions, and none of the "good" solutions can be said to be "better" than any of the other "good" solutions (Madsen, 2000). A member of the Pareto set will be better than any other member with respect to some of the objectives, but because of the trade-off between the different objectives it will not be better with respect to other objectives. In practical applications, the entire Pareto set may be too expensive to calculate, and one is only interested in part of the Pareto optimal solutions.

Generally, when dealing with the multi-objective calibration, the problem is usually transformed into a single-objective optimisation problem by defining a scalar that aggregates the various objective functions (Madsen, 2000; Parakja et al., 2007). However herein, the balanced optimum Pareto criterion was calculated using the simulation results obtained by the single-objective calibration. The procedure was first undertaken for each of the couples within the same function type i.e. (V_g and V_r), (RMSE_g and RMSE_r) and (P_g and P_r). Then, we crossed two "global" criteria i.e. (V_g and RMSE_g), (V_g and P_g), (RMSE_g and P_g), and two "relative" criteria (V_r and RMSE_r), (V_r and P_r) and (V_r and RMSE_r). In the last step, the calibration was carried out using all functions of the same type (V_g, RMSE_g and P_g) and (V_r, RMSE_r and P_r).

4 The study site

4.1 Catchment description

The Gardon d'Anduze is a 543 km^2 Mediterranean catchment located in Southern France. It has a highly marked topography consisting of high mountain peaks, narrow valleys, steep hillslopes and a herring-bone shaped channel network. The highest point is the Mont Aigoual at 1567 m a.s.l.

and the outlet is located at Anduze at 123 m a.s.l. The catchment's soils developed essentially on metamorphic (64% of the catchment area) and granitic terrains. The substrate is made of shale and crystalline rocks overlain by silty clay loams (83% of the catchment area) and sandy loam top soil. The vegetation is dense and composed mainly of beech and chestnut trees, holm oaks and garrigue, conifers, moor, pasture and cultivated lands. These vegetation classes are typical of Mediterranean forests.

Rainfall data for the 1977–1984 period were obtained on paper medium from the "Direction Départementale de l'Equipement du Gard" of the French Ministry of Equipment on seven rain gauge stations. The "Direction Départementale de l'Equipement du Gard" provided also analogue streamflow hydrographs at the outlet at Anduze (Moussa, 1991). The Gardon region is characterized by the highest rainfall intensities recorded in France e.g. a maximum daily rainfall of 608 mm was recorded on the Mont Aigoual during 24 h on 30–31 October 1963. The analysis of long rainfall time series shows that a daily rainfall of 70 mm has a return period of 1 year and daily rainfall of 170 mm has a return period of 100 years. The conjunction of high intensity rainfall, shallow soils and steep slopes produce devastating floods in autumn.

Mean rainfall was calculated as the arithmetic mean of the seven rain gauges. It is true that a mean value of rainfall cannot fully express rainfall heterogeneity. However, in lumped hydrologic models, one cannot distribute rainfall without altering the model structure into a semi-distributed scheme. Herein the number of rain gauge is seven, and the mean precipitation can be calculated using either simple graphical methods or more elaborate techniques based on spline functions or kriging methods. However, Lebel et al. (1987) showed that on this catchment, for a small number of rainfall gauges, the arithmetic mean and the Thiessen models yield comparable results to more complicated methods such as kriging and spline.

4.2 Characteristics of the studied flood events

Flood events from the 1977–1984 period were selected based on a continuous rainfall greater than 50 mm. All corresponding hyetographs and hydrographs were digitised at an hourly time step (Moussa, 1991). In total, 29 events were retained: the event durations range between 24 h and 108 h, the total rainfall between 50 mm and 300 mm, the total runoff between 9 and 166 mm, the runoff coefficients between 15% and 67%, and the initial discharge between 3 and 92 $m^3 s^{-1}$. Figure 2 shows the relations between the total rainfall, the total runoff depth, runoff coefficient, peakflow, and the initial discharge. No clear correlation can be seen between them i.e. the most important rainfall events in terms of precipitation volume are not necessarily those that have the highest runoff coefficients or peakflow. The initial discharge value which represents the catchment's moisture condition does not seem to yield linear trends either. This finding is typical
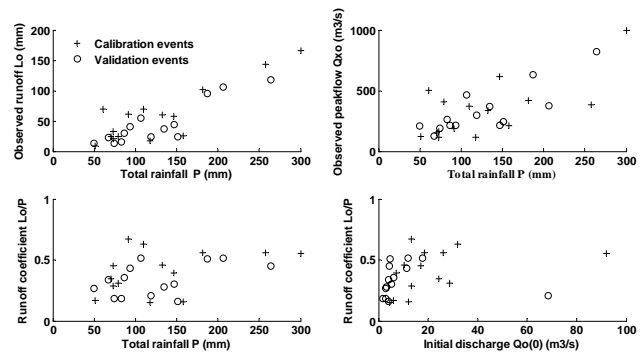

Fig. 2. Relationships between total rainfall, total runoff, peak flow, initial disccharge and runoff coefficient for the calibration (+) and validation (o) events.

of Mediterranean climatic conditions where short duration and high intensity rainfall events are often the cause of the most important runoff events in terms of both runoff depth and peakflow.

Fifteen events, corresponding to the 1977–1979 period, were chosen for calibration and the remaining fourteen, corresponding to the 1980–1984 period, were used for validation (Fig. 2). Both data sets are representative of the various hydrological behaviours observed on the catchment. They cover all climatic seasons and display a large spectrum of rainfall intensity, peakflow and runoff coefficient values.

5 Single- and mutli-objective calibration and validation results

For the fifteen events of the calibration period, a number of tests were carried out in order to optimise the parameters first using single-objective functions, and then to estimate the Pareto front and analyse the trade-offs between the different objectives when using two- or three-objective approaches.

5.1 A priori ranges for the model parameters

We defined a lower and an upper variation bound for each parameter. The hypercube search space shown in Table 1 was used for all the numerical tests:

- The soil-reservoir maximum capacity S_m represents the storage in the root zone for a soil depth of approximately 1–2 m, and was estimated equal to 365 mm on the Gardon d'Anduze by Moussa (1991) and Moussa et al. (2007). S_m's variation range was set between 0 and 1000 mm.

- "f_c" represents the soil's infiltration capacity at natural saturation. As a first approximation, this parameter was considered equal to the hydraulic conductivity at natural saturation using Rawls and Brakenseik's pedotransfer functions (1989). Based on soil and geology maps,

Table 1. Parameter ranges applied for the different automatic calibration together with the single-objective calibration and the balanced multi-objective Pareto optimum solution.

Objective Function	S_m (m)	a	f_c ($\times 10^{-5}$) (ms^{-1})	b	k ($\times 10^{-5}$) (s^{-1})	w (h)	z
			a. Parameter ranges				
	[0–1.0]	[1–70]	[0–30]	[0–1]	[0–30]	[0–72]	[0.01–100]
			b. Single-objective calibration				
V_g	0.015	1.39	2.634	0.809	0.361	23.03	13.71
V_r	0.009	19.92	0.112	0.024	0.249	21.29	11.03
RMSE$_g$	0.051	12.94	0.289	0.260	0.103	5.68	3.48
RMSE$_r$	0.053	12.52	0.295	0.267	0.111	5.79	3.24
P_g	0.029	1.70	1.313	0.462	0.789	3.06	17.16
P_r	0.071	17.84	0.737	0.064	0.342	4.32	4.20
			c. Pareto optimum with two objective functions				
			c.1. Crossing one "global" and one "relative" objective function				
(V_g)–(V_r)	0.076	1.44	5.695	0.010	0.278	20.83	4.93
(RMSE$_g$)–(RMSE$_r$)	0.051	12.94	0.289	0.260	0.103	5.68	3.48
(P_g)–(P_r)	0.158	5.89	5.206	0.010	3.592	3.16	15.78
			c.2. Crossing two "global" objective functions				
(V_g)–(RMSE$_g$)	0.315	5.10	0.051	0.474	9.867	7.82	11.43
(V_g)–(P_g)	0.120	24.25	0.011	0.700	5.620	5.23	0.22
(RMSE$_g$)–(P_g)	0.040	9.68	0.129	0.634	0.112	11.65	0.42
			c.3. Crossing two "relative" objective functions				
(V_r)–(RMSE$_r$)	0.212	27.07	0.036	0.955	0.820	6.08	15.01
(V_r)–(P_r)	0.223	13.69	0.041	0.792	0.602	8.53	13.38
(RMSE$_r$)–(P_r)	0.075	16.80	0.637	0.065	0.382	4.62	5.10
			d. Pareto optimum with three objective functions				
(V_g)–(RMSE$_g$)–(P_g)	0.172	10.549	5.547	0.220	0.2375	4.36	28.2
(V_r)–(RMSE$_r$)–(P_r)	0.225	4.289	7.372	0.055	0.367	3.03	14.1

we distinguish two soil classes (Moussa et al., 2007). The first one corresponds to silty clay loam soils with low permeability (estimated $f_c=2.8\times10^{-7}$ ms^{-1}) while the second one to sandy loam soils, with higher permeability values (estimated $f_c=1.4\times10^{-6}$ ms^{-1}). f_c's variation range was set between 0 and 300×10^{-6} ms^{-1}.

– "a" is an empirical parameter that relates the maximum f_0 and minimum f_c infiltration capacities such as $f_0=af_c$ (with a>1). f_0 depends also on soil hydrodynamic properties as stated above for f_c. Its variation range was set between 1 and 70.

– "b" is an empirical parameter used in Eq. (4) to separate the infiltration from the soil-reservoir into interflow q and deep percolation g (Fig. 1). "b" can vary between 0 and 1.

– "k" is an empirical parameter characterizing the exponential recession curve of the aquifer. k's variation range was set between 0 and 30×10^{-5} s^{-1}.

– "w" is a time parameter that represents the centre of gravity of the unit hydrograph. For the Gardon d'Anduze, the lag time for the studied flood events

ranges generally between 3 and 9 h. The range of variation of w was set between 0 and 72 h.

– "z" is a form parameter of the unit hydrograph. For low values of z, the unit hydrograph represents an approximation of the diffusive wave equation, while for high values, the unit hydrograph tends to represent an approximation of the kinematic wave equation. z ranges between 0.01 and 100 as suggested by Moussa and Bocquillon (1996).

Note that "z" has a feasible range that covers several decades (0.01–100); that is why in the calibration procedure, we used a small interval for lower values of z (z<1) and a larger interval for higher values (z>1). This procedure is equivalent to a Log transformation of parameter z which allows a better distinction between small and very large values. For all remaining parameters, the initial grid is obtained by subdividing the interval of variation of each parameter into regular steps.

5.2 Results of the single-objective calibration procedure

The calibrated parameter values for each of the six objective functions V_g, V_r, RMSE$_g$, RMSE$_r$, P_g and P_r, are presented

in Table 1. Results show that the parameters vary considerably depending on the objective function used.

For the production function, the soil-reservoir maximum capacity S_m ranges between 9 and 71 mm; this appears to be a small value if it is supposed to represent the storage in the root zone. f_c, which represents the soil's infiltration capacity at natural saturation, varies between 0.29×10^{-5} and 2.6×10^{-5} ms^{-1} and compares well with the values estimated when using Rawls and Brakenseik's (1989) pedotransfer function. It is interesting to note that the two empirical parameters a and b have the widest span, probably because these parameters are used to compensate for the errors made on the remaining parameters of the production function. Indeed, even when using conceptual models, modellers tend to be less permissive with parameters that can evoke physical characteristics or be assimilated to such. As a direct consequence, such parameters, even in an un-constrained calibration procedure, will be allowed to vary in a tighter interval.

The transfer function has two parameters w and z. The travel time values of w obtained through the use of V_g and V_r (w=21–23 h) are clearly overestimated both in comparison with the observation data (the basin response time ranges between 3 and 9 h) and the values obtained through RMSE$_g$, RMSE$_r$, P_g and P_r (w=3–4 h). This is because the two volume criteria V_g and V_r are less sensitive to the hydrograph's global shape and consequently to the parameters of the transfer function. The second parameter of the transfer function z ranges between 3 and 17. These values highlight the importance of the diffusive factor.

For most parameters the use of RMSE$_g$ or RMSE$_r$ yields close results. However, when using P_g or P_r, the calibrated parameters differ from those obtained with V_g, V_r, RMSE$_g$ or RMSE$_r$. These results highlight the differences between the various criteria: V_g or V_r describe the mean value of the discharge, RMSE$_g$ or RMSE$_r$ describe the whole shape of the hydrograph, while P_g and P_r refer to a single point representing peakflow.

When overlaying the simulated hydrographs for the same event using the calibrated parameters for each objective function, one can note major differences between the shapes of simulated hydrographs. Figure 3 shows the simulated hydrographs for the event of 22 October 1977 using the calibrated parameters obtained from the single-objective calibration (Table 1) for respectively V_g, RMSE$_g$ and P_g. When using the parameters calibrated using the V_g-objective function (Fig. 3, normal dotted line), we observe that the simulated hydrograph reproduces the total volume of the hydrograph accurately but fails to represent the global shape of the hydrograph; neither the peakflow amplitude, nor the time of occurrence of peakflow are well reproduced. When using the parameters calibrated using the RMSE$_g$-objective function (Fig. 3, bold dotted line), the global shape of the hydrograph is well represented; as is the total volume and the peakflow is better represented than in the previous case. Finally, when using the parameters calibrated using the P_g-objective func-
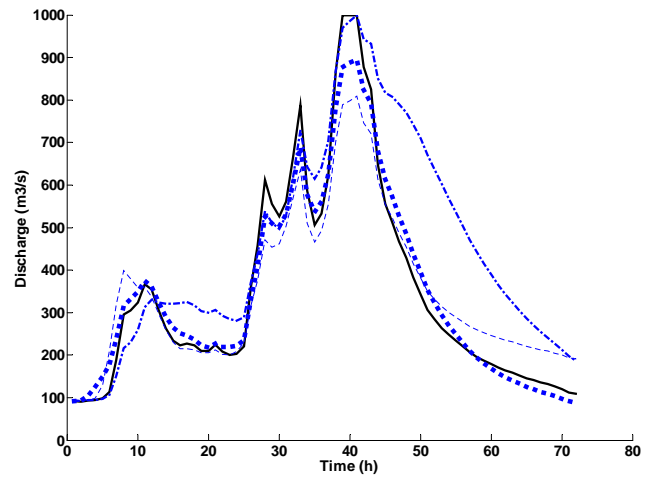

Fig. 3. Simulated hydrographs for the event of 22 October 1977 using the calibrated parameters obtained from the single-objective calibration (Table 1) for respectively V_g (normal dotted line), RMSE$_g$ (bold dotted line) and P_g (dashdot line) objective functions; the solid bold line shows the observed hydrograph.

tion (Fig. 3, dashdot line), the amplitude and the time of occurrence of the peakflow are very well simulated, while the total volume and the global shape of the hydrograph are not. We observe that the use of each single-objective function improves the simulation of some characteristics of the shape of the hydrograph, but never the three studied criteria simultaneously.

Also, no unique solution to the single-objective problem is found and an "equifinality" of parameter sets can be seen as stated by Beven and Binely (1992) and Beven (1993) i.e. many different parameter combinations give acceptable solutions. However, it is important to distinguish equifinality as defined by Beven and Binley (1992), and which forms the basis for the generalised likelihood uncertainty estimation (GLUE) procedure, and trade-offs between different objectives in a multi-objective optimisation framework. The former is related to the fact that different parameter sets may give very similar model performance measured with respect to some likelihood measure (or single objective function), whereas the latter is referred to as multi-objective equivalence of parameter sets and defined in terms of the Pareto criterion (see e.g. Madsen (2000) for a discussion on this).

5.3 Results of the multi-objective calibration procedure when crossing two objective functions

The results of the multi-objective calibration obtained when crossing two objective functions are shown in Figs. 4, 5 and 6. The dotted plots represent the objective function values during the calibration process; the "*" indicates the Pareto front and the bold point indicates the balanced aggregated objective function. In Fig. 4, the calibration is based on two objective functions, the global volume error V_g and the relative
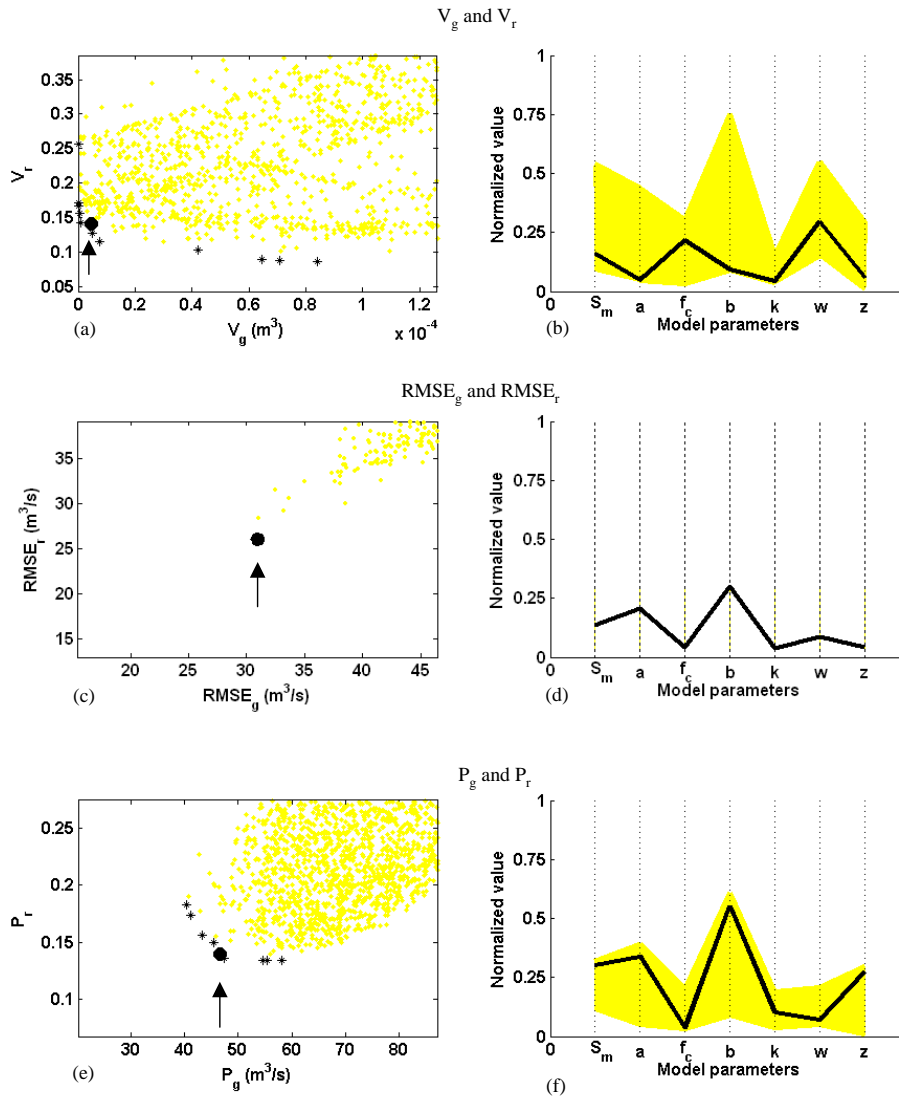
Fig. 4. Multi-objective calibration using one "global" and one "relative" objective function. **(a, c, e)** Objective function values during the calibration process; the "*" indicates the Pareto front and the bold point indicates the balanced aggregated objective function. **(b, d, f)** Normalized range of parameter values along the Pareto front shown in (a), (c) and (e) respectively; full bold line indicate the parameter set corresponding to the Pareto balanced aggregated objective function.

volume error V_r (Fig. 4a), the global RMSE$_g$ and the relative RMSE$_r$ (Fig. 4c) and the global peakflow P_g and the relative peakflow P_r (Fig. 4e). Figure 4a, c and e show the objective function values corresponding to the evaluated parameter sets for two different objective functions. The grey dots indicate the results of the calibration runs. The Pareto optimal front (indicated by a star "*" on Fig. 4a, c and e) is identified, and finally the balanced aggregated objective function (indicated by an arrow on Fig. 4a, c and e) is calculated. Note that in Fig. 4c, the Pareto front collapses totally because conflicts do not exist between the subsets of the analysed objective functions (i.e. RMSE$_g$ and RMSE$_r$). The single-objective optimisation provides the tails of the Pareto

front, and the optimisation based on the balanced aggregated measure approximates the balanced central part of the Pareto front.

The estimated Pareto front for the calibration of V_g and V_r (Fig. 4a) presents a trade-off. A very good calibration of V_g (corresponding to V_g=0) provides a bad calibration of V_r (V_r=16.7%), and vice-versa (V_g=0.166×10^{-4} m^3 for V_r=8.7%). The same comment can be made about P_g and P_r: P_g=40.3 m^3 s^{-1} when P_r=18.3% and P_g=58.0 m^3 s^{-1} when P_r=13.3%. This result is not surprising as peakflow refers to instantaneous values that are both difficult to determine and simulate. The Pareto front of RMSE$_g$ and RMSE$_r$ shows a high correlation. This can be explained by the fact
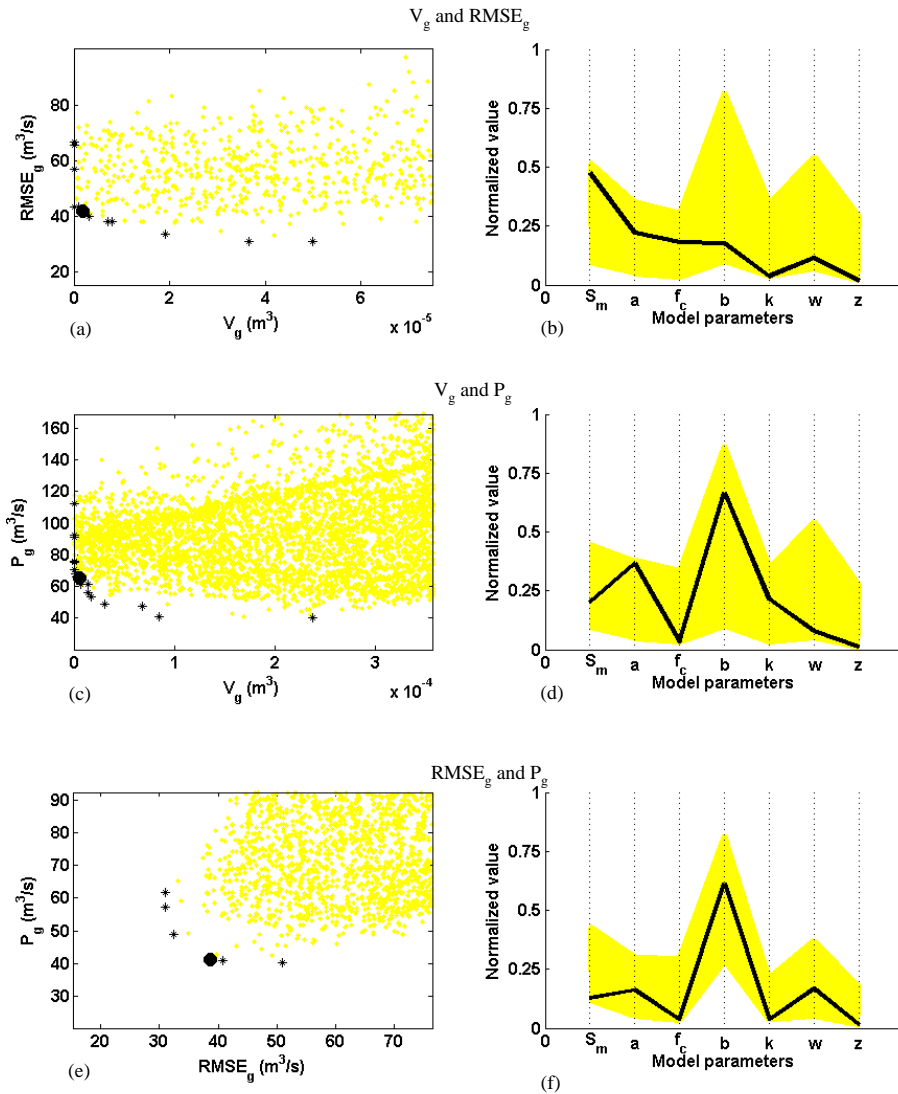
Fig. 5. Multi-objective calibration using two "global" objective functions. **(a, c, e)** Objective function values during the calibration process; the "*" indicates the Pareto front and the bold point indicates the balanced aggregated objective function. **(b, d, f)** Normalized range of parameter values along the Pareto front shown in (a), (c) and (e) respectively; full bold line indicate the parameter set corresponding to the Pareto balanced aggregated objective function.

that the majority of flood events have similar duration (48 to 96 h) and consequently the two objective functions presented in Eq. (15) tend to be similar.

The variation of the optimum model parameter sets along the Pareto front is shown in grey in Fig. 4b, d and e for the multi-objective calibration of (V_g and V_r), ($RMSE_g$ and $RMSE_r$) and (P_g and P_r) respectively. The parameter values are normalised with respect to the upper and lower limits given in Table 1 so that the range of all normalised parameters is between 0 and 1. For the calibration of V_g and V_r (Fig. 4b), a remarkably large span is observed in the parameter values when moving along the Pareto front. The range is larger than 50% for the main parameters S_m, a, b, w and z.

The compromised solution using the Pareto front is shown in bold on Fig. 4b, d and e, and the corresponding values of the calibrated parameters are given in Table 1. The calibrated parameters of the compromised solution of (V_g and V_r) in Table 1 are within the interval delimited by the calibrated parameters of V_g and V_r separately (Table 1). Similar results are observed for ($RMSE_g$ and $RMSE_r$) and for (P_g and P_r).

Figure 5 shows results when crossing two "global" criteria (V_g and $RMSE_g$; Fig. 5a and b), (V_g and P_g; Fig. 5c and d) and ($RMSE_g$ and P_g; Fig. 5e and f) and Fig. 6 shows results when crossing two "relative" criteria (V_r and $RMSE_r$; Fig. 6a and b), (V_r and P_r; Fig. 6c and d) and ($RMSE_r$ and P_r; Fig. 6e and f). Note that when we calibrate the model
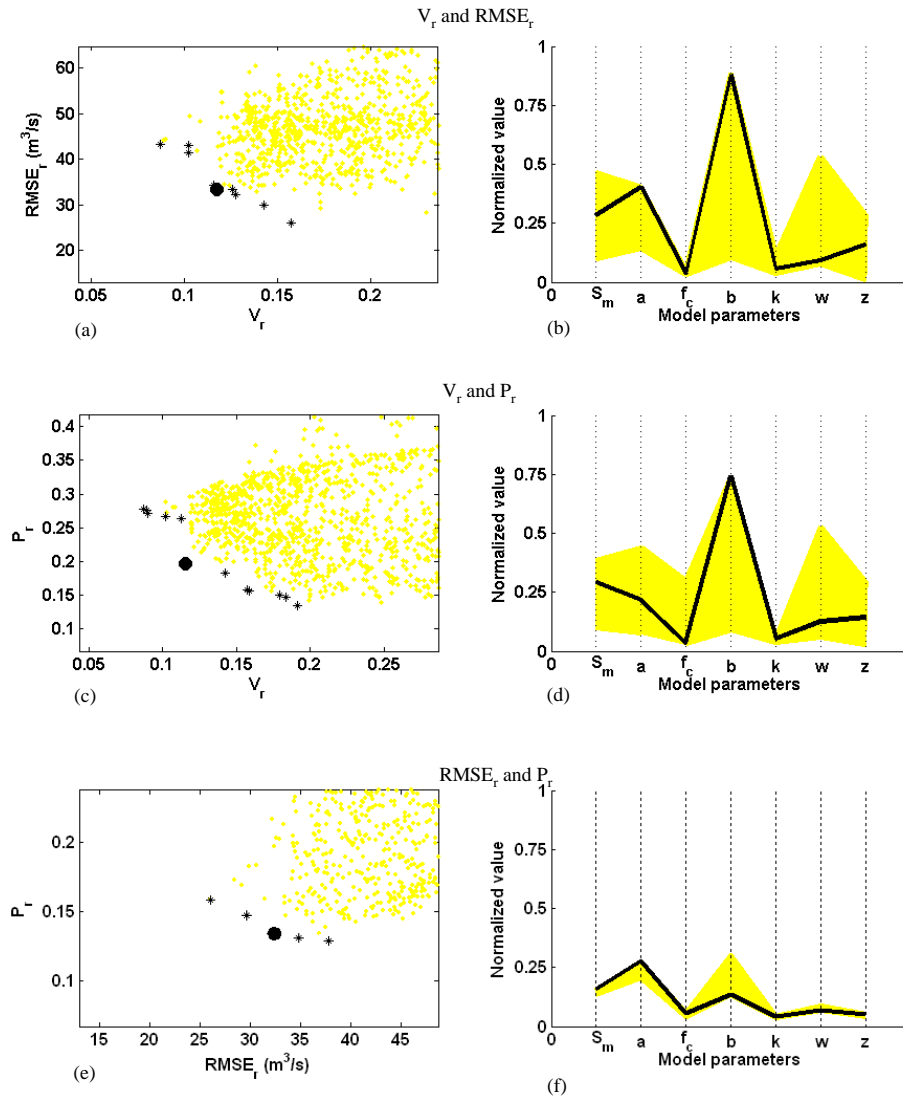
Fig. 6. Multi-objective calibration using two "relative" objective functions. **(a, c, e)** Objective function values during the calibration process; the "*" indicates the Pareto front and the bold point indicates the balanced aggregated objective function. **(b, d, f)** Normalized range of parameter values along the Pareto front shown in (a), (c) and (e) respectively; full bold line indicate the parameter set corresponding to the Pareto balanced aggregated objective function.

using the global volume objective function V_g, the criterion tends to zero because the calibration procedure allows the simulation of the exact total volume; this is not the case for the other criteria. Note also that the volume based objective functions V_g and V_r are calculated in volume per time step. Consequently, the corresponding objective function (i.e. less than 0.0001 m³/time step in Fig. 5) is small in comparison with other objective functions (i.e. the relative volume error ranges between 0.05 and 0.25 in Fig. 6). Again, we observe significant trade-offs for the three cases of Fig. 5 :

- V_g=0 when RMSE$_g$=65.5 m³ s⁻¹ and V_g=5.0×10⁻⁵ m³ when RMSE$_g$=31.0 m³ s⁻¹ (Fig. 5a);

- V_g=0 when P_g=92.4 m³ s⁻¹ and V_g=23.8×10⁻⁵ m³ when P_g=40.3 m³ s⁻¹ (Fig. 5c);

- RMSE$_g$=31.0 m³ s⁻¹ when P_g=61.6 m³ s⁻¹ and RMSE$_g$=50.9 m³ s⁻¹ when P_g=40.3 m³ s⁻¹ (Fig. 5e).

Comparable results are obtained with the relative criteria in Fig. 6a, c and e. The variation of the optimum model parameter sets along the Pareto front for the multi-objective calibration of (V_g and RMSE$_g$; Fig. 5b), (V_g and P_g; Fig. 5d), (V_r and RMSE$_r$; Fig. 6b), (V_r and P_r; Fig. 6d) are similar and sometimes higher than those observed for (RMSE$_g$ and P_g; Fig. 5f) and (RMSE$_r$ and P_r; Fig. 6f). Figures 4 to 6 indicate that the global volume criteria (V_g) is the most insensitive i.e.

volume conservation can be easily respected with a number of parameter combinations. This translates into a somehow flat Pareto front (Figs. 4a, 5a and b). In comparison RMSE and peakflow produce sharper fronts; RMSE seems to be the most restrictive of the tested criteria. The relative criteria seem to yield sharper fronts than the global criteria which seem to be controlled by the extreme events.

In comparison with the single-objective calibration, the use of a multi-objective calibration technique seems to yield tighter variation intervals. These findings are in accordance with those of Engeland et al. (2006). The variation ranges we obtained for the multi-objective calibration are smaller than those reported in Schoops et al. (2005) and Madsen (2000) who had respectively 3 and 2 additional parameters to calibrate. However, the number of free parameters cannot be the only explanation behind the wider variation spans as Gupta et al. (2003) obtained tighter intervals when calibrating the 13 parameters of the SAC-SMA model.

5.4 Results of the multi-objective calibration procedure when crossing three objective functions

The same methodology was extended in order to combine three objective functions. Figure 7 shows a volumetric view for the three-objective calibration space, when crossing the three global objective functions (V_g, $RMSE_g$ and P_g; Fig. 7a), and the three relative objective functions (V_r, $RMSE_r$ and P_r; Fig. 7c). Results show comparable parameter variation ranges for both (V_g, $RMSE_g$ and P_g; Fig. 7b) and (V_r, $RMSE_r$ and P_r; Fig. 7d). Note that Fig. 5a, c and e give the projection on two-objective planes, respectively (V_g, $RMSE_g$), (V_g, P_g) and ($RMSE_g$, P_g) of the three global objective functions (V_g, $RMSE_g$, P_g) of Fig. 7a. Similarly, Fig. 6a, c and e give the projection on two-objective planes, respectively (V_r, $RMSE_r$), (V_r, P_r) and ($RMSE_r$, P_r) of the three relative objective functions (V_r, $RMSE_r$, P_r) of Fig. 7c. For the optimal solution (Table 1), the soil-reservoir maximum capacity S_m is equal to 172 mm when crossing (V_g, $RMSE_g$ and P_g) and 225 mm when crossing (V_r, $RMSE_r$ and P_r). This value is more representative of the storage in the root zone. For both combinations (V_g, $RMSE_g$ and P_g) and (V_r, $RMSE_r$ and P_r), the optimal parameter f_c ranges between 5.5×10^{-5} and 7.4×10^{-5} ms^{-1}, and compares well with pedotransfer functions (Rawls and Brakenseik, 1989). It is interesting to note that the two empirical parameters a and b still have the widest span. The travel time values of w obtained range between 3 and 4 h and correspond to the approximate lag time of the basin during intense flood events, while the dimensionless shape parameter z, ranges between 14 and 28. The parameter variation ranges are comparable to those obtained when using only two functions. The parameters that vary most between the two combinations are b, k and z. Parameters b and k are clearly linked to each other as they both control the "outward" fluxes and can be used to correct possible flow over-estimation by increasing percolation and

interflow. Data on water levels both in the unsaturated and saturated zones could have been useful in constraining these parameters further. Unfortunately such data were unavailable for the studied period.

A similar approach can be used for other sets of three-, four-, five- or six-objective functions. However, we limit our analysis to the three global objective functions (V_g, $RMSE_g$ and P_g) and the three relative objective functions (V_r, $RMSE_r$ and P_r), because the aim of this study is to conduct a sensitivity analysis of the calibrated parameters when moving from single- to two- or three-objective functions, and not to explore the whole space of the objective functions. Also, it is not simple to analyse simultaneously all possible combinations of the objective functions due to their large number, and because three or more objective functions could yield very complex geometries for the Pareto front. Moreover, mathematically, the Pareto front will "collapse dimension" in subspaces if conflicts do not exist between the subsets of the analysed objectives (see Fig. 4c for an example of a dimensional collapse). Herein, the analysis was conducted in a six-objective space, and a single analysis using all single objective functions implicitly yields all the presented results. It was also possible to present results moving from higher to lower dimensional analysis (i.e. from six-, to five-, four-, three-, two- and single-objective functions). However, it is not simple to analyse data in a six-D space and even in three-D space due to the complex shape of functions. Consequently, starting the analysis with single-, than two- and finally three- objective functions allows the analysis of the shape of each objective function separately and then two by two, then three by three, etc.

The methodology presented in this paper has also its limits, especially in hydrologic modelling where non-linearities and thresholds yield non-convex Pareto fronts such as in Fig. 6a and c. As discussed in a number of studies by Changkong and Haimes (1983), Koski (1985), Das and Dennis (1997) and Messac and Mattson (2002), the traditional approaches used in the calculation of the Pareto front have drawbacks on non-convex parts of the Pareto front. This is due to the fact that the weighted-sum method is often implemented as a convex combination of objectives, where the sum of all weights is constant and negative weights are not allowed. Generally, multi-objective optimisation algorithms perform well for bivariate problems, but scale poorly to multiple objectives (Changkong and Haimes, 1983). More recently, new methods developed to generate the Pareto optimal solutions in non-convex regions (i.e. Kim and Weck, 2005) give promising results in the case of two-objective functions, but further theoretical work is still needed to overcome the limitations for multi-objective functions.

5.5 Validation and uncertainty analysis

As the "global" and "relative" criteria gave comparable results and interpretation, we choose to validate the results of
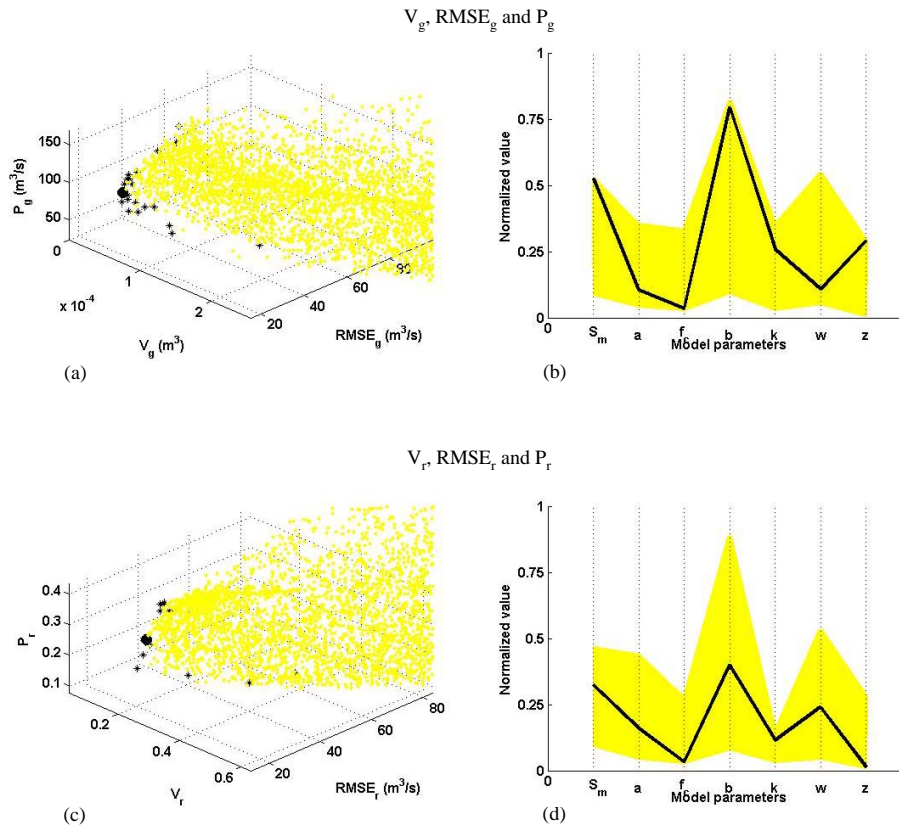
V_g, RMSE$_g$ and P$_g$

(a)

(b)

V_r, RMSE$_r$ and P$_r$

(c)

(d)

Fig. 7. Multi-objective calibration using three "global" **(a** and **b)** and three "relative" **(c** and **d)** objective functions. a and c) Objective function values during the calibration process; the "∗" indicates the Pareto front and the bold point indicates the balanced aggregated objective function. b and d) Normalized range of parameter values along the Pareto front shown in (a) and (c) respectively; full bold line indicate the parameter set corresponding to the Pareto balanced aggregated objective function.
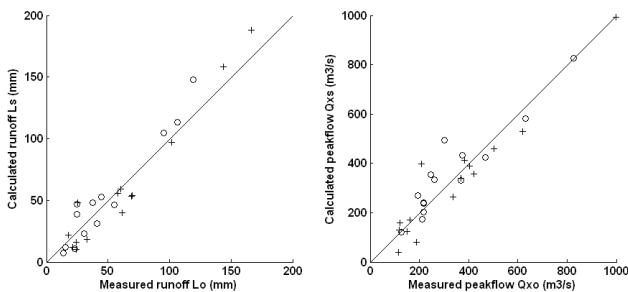

Fig. 8. Comparison between measured and simulated runoff depths **(a)** and peakflows **(b)** using the set of parameters from the multi-objective V_r–RMSE$_r$–P_r procedure.

the "relative" criteria (V_r, RMSE$_r$ and P_r) which are less sensitive to extreme events.

To compare the calibration procedures, we computed the value of each objective function (Table 2). It is not surprising that the values of V_r, RMSE$_r$ and P_r are minimal when single-objective calibration are conducted function of each of these criteria : V_r=8.7%, RMSE$_r$=26.0 m^3 s^{-1} and

P_r=13.4%. The maximum value of V_r=19.1% is obtained when using a single-objective calibration minimising P_r; the maximum value of RMSE$_r$=43.2 m^3 s^{-1} is obtained when using a single-objective calibration minimising V_r; and the maximum value of P_r=27.8% is also obtained when using a single-objective calibration minimising V_r. Once again the use of a volume based criterion is restrictive while the use of RMSE can yield relatively acceptable results for peakflow and vice versa. The sensitivity of the RMSE criterion to peakflow has already been reported by many authors (Parada et al., 2003; Yapo et al., 1998) and our findings are similar to theirs.

When using two- or three-objective functions, the error values fall within the intervals indicated above. However, it is interesting to note that for all three criteria, the maximum errors values obtained with the multi-objective methods are always lower than those obtained when using a single-objective calibration in which the given criterion is not considered. This is essentially a consequence of the trade-offs between objective functions. Finally, the combination of (V_r, RMSE$_r$ and P_r) gives a reasonable compromise between the three criteria with V_r=12.6%, RMSE$_r$=33.3 m^3 s^{-1}
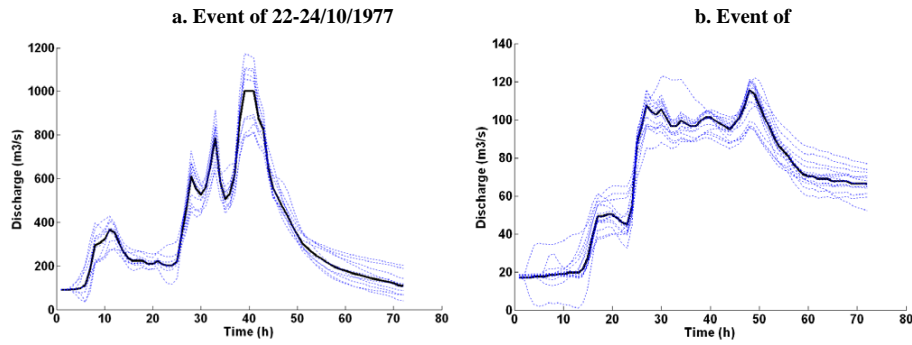
Fig. 9. Example of simulated hydrographs of high intensity (**a**) and low intensity events (**b**) using the set of parameters from the multi-objective V_r–RMSE$_r$–P_r procedure. The full bold line shows the measured hydrograph, and the dotted lines the simulated hydrographs.

Table 2. Values of the three objective functions V_r, RMSE$_r$ and P_r when using the calibrated parameters with the single-objective calibration and the balanced two-objective or three-objective Pareto optimum solution.

Objective function used for calibration	Criteria value		
	V_r [-]	RMSE$_r$ (m^3 s^{-1})	P_r [-]
Single-objective (V_r)	0.087	43.2	0.278
Single-objective (RMSE$_r$)	0.158	26.0	0.159
Single-objective (P_r)	0.191	32.4	0.134
Two-objective (V_r)–(RMSE$_r$)	0.117	33.4	0.216
Two-objective (V_r)–(P_r)	0.116	34.4	0.197
Two-objective (RMSE$_r$)–(P_r)	0.180	29.4	0.145
Three-objective (V_r)–(RMSE$_r$)–(P_r)	0.126	33.3	0.214

Table 3. Means $\overline{\varepsilon_V}$ and $\overline{\varepsilon_Q}$ of the relative prediction error on runoff depth and peakflow of the calibration and validation events when using the calibrated parameters with the single-objective calibration and the balanced two-objective or three-objective Pareto optimum solution.

Objective function used for calibration	Runoff depth $\overline{\varepsilon_V}$	Peakflow $\overline{\varepsilon_Q}$
Single-objective (V_r)	−0.049	−0.269
Single-objective (RMSE$_r$)	0.047	0.129
Single-objective (P_r)	0.172	−0.107
Two-objective (V_r)–(RMSE$_r$)	0.021	−0.161
Two-objective (V_r)–(P_r)	0.026	−0.190
Two-objective (RMSE$_r$)–(P_r)	0.169	−0.112
Three-objective (V_r)–(RMSE$_r$)–(P_r)	0.014	−0.183

and P_r=21.4%. These values are quite comparable to those obtained when using just two objective functions i.e. V_r, RMSE$_r$.

To further compare the performance of the calibration procedures, we computed the relative error on both runoff depth and peakflow: for a given event i, the error on runoff depth and peakflow are defined respectively by ε_{vi}=(Ls$_i$−Lo$_i$)/Lo$_i$ and ε_{Qi}=(Qxs$_i$−Qxo$_i$)/Qxo$_i$. Let $\overline{\varepsilon_V}$ and $\overline{\varepsilon_Q}$ be the mean of ε_{vi} and ε_{Qi} respectively. The quantities $\overline{\varepsilon_V}$ and $\overline{\varepsilon_Q}$ represent

the bias of runoff depth and peakflow predictions. Table 3 illustrates the findings when using either the single or multi-objective method (i.e. two or three objective functions) and shows that $\overline{\varepsilon_V}$ and $\overline{\varepsilon_Q}$ fall within similar ranges for two- or three-objective functions. It is worthy to note that the greatest errors on runoff depth are obtained through the use of a peakflow criterion while the greatest errors on peakflow are caused by the use of a volume criterion. Apart from two cases, the error on runoff depth is acceptable (<5%) while, not surprisingly, the error on peakflow is far higher (>10%) but less than 27%. In addition, the model seems to have a tendency to overestimate runoff depth and to underestimate peakflow.

The best compromise between both errors is reached by using the parameters obtained through multi-objective calibration with three functions (V_r, RMSE$_r$ and P_r) but the combination of V_r and RMSE$_r$ gives once again comparably good results. In this instance the use of two "well" chosen and complementary objective functions seems to be sufficient for runoff simulation. However, the lack of soil-moisture and groundwater data prevents us from extending our results to the other vertical components of the model.

Figure 8 compares the measured and simulated runoff depths (Fig. 8a) and the measured and simulated peakflows (Fig. 8b) obtained when using the parameters of the balanced aggregated objective function (V_r, RMSE$_r$ and P_r) given in Table 1. It can be seen that the model gives similar results both for the calibration and validation periods. An illustration of the equifinality problem is shown in Fig. 9 where hydrographs are simulated using the parameters corresponding to the entire Pareto front of (V_r, RMSE$_r$ and P_r); the bold line shows the measured hydrograph and the dotted lines show the various model simulations using the multi-objective V_r-RMSE$_r$-P_r procedure's parameters. The results indicate that the hydrographs are also well simulated for high (Fig. 9a) and low intensity events (Fig. 9b). The figure clearly shows that many different parameter sets, may produce equally good simulations according to the three objective functions.

6 Discussion and conclusion

A conceptual lumped event based rainfall-runoff model, coupling a production and a transfer function, was developed and applied on the Gardon catchment in southern France. The model has seven free parameters which need to be calibrated. The model was calibrated on 15 events and validated on 14 others. The results of both the calibration and validation phases were compared on the basis of their performance with regards to six objective functions, three global and three relative, representing volume, the root mean square error, and peakflow.

The results showed that the calibrated parameter values were dependent on the type of the objective function used. Trade-offs are observed between the different objectives and no single set of parameter was able to optimise all objectives simultaneously. Thus a set of Pareto optimal solutions and a balanced aggregated objective function were calculated with two and three objective functions.

The comparison of the single and multi-objective calibration results, using two or three functions, illustrates the "non-uniqueness problem" (Beven and Binley, 1992) since many different parameter combinations gave acceptable solutions according to a given objective. However, the volume conservation criterion seemed to be the most permissive whereas the RMSE and the peakflow prediction criteria yielded sharper Pareto fronts. The model was then validated with the set of optimised parameters obtained using the "combined relative" criteria (V_r, RMSE_r and P_r). The use of a triple objective function does not seem to be justified in our case. Indeed given the impact peakflow values have on the RMSE, there seems to be a redundancy in their use, hence a combination of either (V_r and P_r) or (V_r and RMSE_r) can yield equally acceptable results.

Differences between measured and simulated hydrographs were assessed by calculating the bias of the simulated runoff depth and peakflow. These errors can be due to the use of non-optimal parameter values but also to errors inherent to the model structure and the meteorological input data. In the model calibration herein, only the error due to the parameter values is minimised. However, the calibration of model parameters can also compensate the other error sources. In our case the best results in terms of bias were obtained through multiple calibration with a volume and an RMSE criterion

The choice of an adequate objective function when modelling separate flood events, emphasise the importance of the modeller's intervention for tailoring the model calibration to a specific application. Attempts have already been made to include this knowledge objectively in the model calibration procedure (Boyle et al., 2000). Our results highlight the importance of the modeller's professional judgement as often the criteria values and error estimates are within close bounds and may not be significantly different from a statistical point of view. It is therefore important to plot the hydrographs and assess the graphical differences in the simulated hydrograph's shape.

A sound hydrological knowledge is required to evaluate data and model errors. In most real world application, especially in an operational framework and for real-time predictions, data quality checks could be too time consuming and hence difficult to carry out. Thus a robust calibration procedure becomes even more essential.

Edited by: H. Bormann

References

Anctil, F., Michel, C., Perrin, C., and Andréassian, V.: A soil moisture index as an auxiliary ANN input for stream flow forecasting, J. Hydrol., 286(1–4), 155–167, 2004.

Bates, B. and Ganeshanandam, S.: Bootstrapping non-linear storm event models, In: A.S.O.C. Engineering (Editor), National conference of Hydraulic engineering, San Diego, CA, USA, 330–335, 1990.

Bergström, S.: The HBV model, in: Computer Models of Watershed Hydrology, edited by: Singh, V. P., Water Resources Publications, Colorado, USA, 443–476, 1995.

Beven, K. J.: Prophesy, 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. Process., 6, 279–298, 1992.

Boyle, D., Gupta, H., and Sorooshian, S.: Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods, Water Resour. Res., 36, 3663–3674, 2000.

Brocca, L., Melone, F., and Moramarco, T.: On the estimation of antecedent wetness conditions in rainfall-runoff modelling, Hydrol. Process., 22, 629–642, 2008.

Chahinian, N., Moussa, R., Andrieux, P., and Voltz, M.: Comparison of infiltration models to simulate flood events at the field scale, J. Hydrol., 306(1–4), 191–214, 2005.

Chahinian, N., Moussa, R., Andrieux, P., and Voltz, M.: Accounting for temporal variation in soil hydrological properties when simulating surface runoff on tilled plots, J. Hydrol., 326(1–4), 135–152, 2006.

Changkong, N., and Haimes, Y. Y.: Multiobjective decision making: Theory and methodology, North-Holland Publishing, New York, Amsterdam, 406 pp., 1983.

Chevallier, P.: L'indice des précipitations antérieures, évaluation de l'humectation des sols des bassins versants représentatifs, Cah. ORSTOM, Série Hydrologie, 20(3–4), 179–189, 1983.

Crawford, N. H. and Linsley, R. K.: Digital simulation in hydrology: Stanford watershed model IV, Technical Report 39, Department of Civil Engineering, Stanford University, Stanford, California, USA, 210 pp., 1966.

Cormary, Y. and Guilbot, A.: Relations pluie-débit sur le bassin de la Sioule, Rapport D.G.R.S.T. No. 30, Université des Sciences et Techniques, Montpellier, France, 1969.

Das, I. and Dennis, J. E.: A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multi-criteria optimization problems, Struct. Optim., 14, 63–69, 1997.

Diskin, M. and Nazimov, N.: Linear reservoir with feedback regulated inlet as a model for the infiltration process, J. Hydrol., 172, 313–330, 1995.

Donigan, A., Bicknell, B., and Imhoff, J. C.: Hydrological simulation program – Fortran (HSPF), in: Computer Models of Watershed Hydrology, edited by: Singh, V. P., Water Resource Publications, Colorado, USA, 395–442, 1995.

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimisation for conceptual rainfall-runoff models, Water Resour. Res., 28, 1015–1031, 1992.

Ehrgott, M.: Multicriteria optimization, Springer, Heidelberg, Berlin, Germany, 323 pp., 2005.

Engeland, K., Braud, I., Gottschalk, L., and Leblois, E.: Multiobjective regional modelling, Hydrol. Process., 327, 339–351, 2006.

Fedora, M. and Beschta, M.: Storm runoff simulation using an antecedent precipitation index (API) model, J. Hydrol., 112(1–2), 121–133, 1989.

Fleming, G.: Computer simulation techniques in hydrology, Environmental Science Series, Elsevier, 333 pp., 1975.

Gan, T. and Biftu, G.: Automatic calibration of conceptual-runoff models: Optimization algorithms, catchment conditions, and model structure, Water Resour. Res., 32(12), 3513–3524, 1996.

Garçon, R.: Prévision opérationnelle des apports de la Durance à Serre-Ponçon à l'aide du modèle MORDOR : Bilan de l'année 1994–1995, La Houille Blanche, 51(2), 71–79, 1996.

Gupta, H., Sorooshian, S., Hogue, T., and Boyle D.: Advances in the automatic calibration of watershed models, in: Calibration of watershed models, Water Science and Application American Geophysical Union, Washington DC, USA, 6, 9–28,, 2003.

Havnø, K., Madsen, M. N., and Dørge J.: MIKE 11 – a generalized river modelling package, in: Computer Models of Watershed Hydrology, edited by: Singh, V. P., Water Resources Publications, Colorado, 733–782, 1995.

Hayami, S.: On the propagation of flood waves, Disaster Prev. Res. Inst. Bull., 1, 1–16, 1951.

Heggen, R. J.: Normalized antecedent precipitation index, J. Hydrol. Eng., 6(2), 377–381, 2001.

Jain, A. and Indurthy, P.: Comparative analysis of event based rainfall-runoff modelling techniques-Deterministic, statistical and artificial neural networks, J. Hydrol. Eng., 8(2), 93–98, 2003.

Johnsen, K., Mengelkamp, H., and Huneke, S.: Multi-objective calibration of the land scheme TERRA/LM using LIFTASS-2003 data, Hydrol. Earth Syst. Sci., 9(3), 586–595, 2005.

Kim, I. Y. and de Weck, O. L.: Adaptive weighted-sum method for bi-objective optimisation: Pareto front generation, Struct. Multidisc. Optim., 29, 149–158, 2005.

Koski, J.: Defectiveness of weighting method in multicriterion optimization of structures, Commun. Appl. Numer. Methods, 1, 333–337, 1985.

Kuczera, G.: Efficient subspace probabilistic parameter optimisation for catchment models, Water Resour. Res., 33, 177–186, 1997.

Lebel, T., Bastin, G., Obled, C., and Creutin, J. D.: On the accuracy of areal rainfall estimation: a case study, Water Resour. Res., 23(11), 2123–2134, 1987.

Legates, D. and McCabe, G.: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35(1), 233–241, 1999.

Madsen, H.: Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, J. Hydrol., 235, 276–288, 2000.

Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, Adv. Water Resour., 26, 205–216, 2003.

Madsen, H., Wilson, G., and Ammentrop, H.: Comparison of different automated strategies for calibration of rainfall-runoff models, J. Hydrol., 261, 48–59, 2002.

Messac, A., and Mattson, C.A.: Generating well-distributed sets of Pareto points for engineering design using physical programming, Optim. Eng., 3, 431–450, 2002.

Miettinen, K. M.: Non-linear Multi-objective optimization, Kluwer's Academic Publishing, 1999.

Moussa, R.: Variabilité spatio-temporelle et modélisation hydrologique. Application au bassin du Gardon d'Anduze, PhD dissertation, University of Montpellier II, France, 314 pp., 1991.

Moussa, R. and Bocquillon, C.: Algorithms for solving the diffusive wave flood routing equation, Hydrol. Process., 10, 105–123, 1996.

Moussa, R., Voltz, M., and Andrieux, P.: Effects of spatial organization of agricultural management on the hydrological behaviour of farmed catchment during flood events, Hydrol. Process., 16(2), 393–412, 2002.

Moussa, R., Chahinian, N., and Bocquillon, C.: Distributed hydrological modelling of a Mediterranean mountainous catchment - model construction and multi-site validation, J. Hydrol., 337, 35–51, doi:10.1016/j.jhydrol.2007.01.028, 2007.

Nash, I. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models. Part I: a discussion of principles, J. Hydrol., 10, 282–290, 1970.

Nelder, J. and Mead, R.: A simplex method for function minimization, Comput. J., 7, 308–313, 1965.

Paquet, E.: Évolution du modèle hydrologique MORDOR: modélisation du stock nival à différentes altitudes, La Houille Blanche, 2(2), 75–82, 2004.

Parada, L., Fram, J., and Liang, X.: Multi-resolution calibration of hydrologic models, in: Calibration of watershed models, Water Science and Application, American Geophysical Union, Washington DC, USA, 6, 197–212, 2003.

Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, Hydrol. Process., 21, 435–446, 2007.

Pareto, V.: Manuale di Economia Politica, Societa Editrice Libraria, Milano, Italy, (translated into English by: Schwier, A. S., 1971): Manual of Political Economy, New York, USA, Macmillan, 1906.

Peugeot, C., Cappelaere, B., Vieux, B., Séguis, L., and Maia, A.: Hydrologic process simulation of a semiarid, endoreic catchment in Sahelian West Niger, 1. Model-aided data analysis and screening, J. Hydrol., 279(1–4), 224–243, 2003.

Rawls, W. J. and Brakensiek, D. L.: Estimation of soil hydraulic properties. Unsaturated flow in hydrologic modeling: Theory and practice. NATO ASI series, Series C, Mathematical and physical sciences, edited by: Morel-Seytoux, H. J., 275, Kluwer Academic, Boston, USA, 275–300, 1989.

Rosenbrock, H.: An automatic method for fitting the greatest or least value of a function, Comput. J., 3, 175–184, 1960.

Hydrol. Earth Syst. Sci., 13, 519–535, 2009

www.hydrol-earth-syst-sci.net/13/519/2009/

Schoops, G., Hopmans, J., Young, C., Vrugt, J., and Wallender, W.: Multi-criteria optimization of a regional spatially-distributed subsurface water flow model, J. Hydrol., 311, 20–48, 2005.

Singh, V. P.: Computer Models of Watershed Hydrology, Water Resources Publications, Colorado, USA, 1130 pp., 1995.

Soil Conservation Service-USDA: Estimation of direct runoff from storm rainfall, National Engineering Handbook, Section 4-Hydrology, 10.1–10.24, 1972.

Solomatine, D. P.: Two strategies of adaptive cluster covering with descent and their comparison to other algorithms, J. Global Optimization, 14(1), 55–78, 1999.

Stadler, W.: A survey of multicriteria optimization, or the vector maximum problem, JOTA, 29, 1–52, 1979.

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/.

Vrugt J., Gupta, H., Bastidas, L., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water Resour. Res., 39(5), 1214, doi:10.1029/2002WR001746, 2003a.

Vrugt J., Gupta, H., Bouten, W., and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, Water Resour. Res., 39(5), 1201, 2003b.

Willmott, C.: On the validation of models, Phys. Geogr., 2, 184–194, 1981.

Yang, X. and Michel, C.: Flood forecasting with a watershed model: a new method of parameter updating, Hydrol. Sci. J., 45(4), 537–546, 2000.

Yapo, P., Gupta, H., and Sorooshian, S.: Multi-objective global optimization for hydrologic models, J. Hydrol., 204, 83–97, 1998.

Yu, P. S. and Yang, T. C.: Fuzzy multi-objective function for rainfall-runoff model calibration, J. Hydrol., 238, 1–14, 2000.

Zadeh, L.: Optimality and Non-Scalar-Valued Performance Criteria, IEEE T. Autom. Control 8, 59–60, 1963.

Zarriello, P.: Comparison of Nine Uncalibrated Runoff Models to Observed Flows in Two Small Urban Watersheds, In: Subcommittee on Hydrology of the Interagency Advisory Committee on Water Data (Editor) First Federal Interagency Hydrologic Modeling Conference, Las Vegas, NV, USA, 7-163–7-170, 1998.