Multi-objective calibration of the physically based, spatially distributed SHETRAN hydrological model
Rong Zhang, Madalena Moreira and João Corte-Real

ABSTRACT
Physically based, spatially distributed hydrological models have mostly been calibrated manually; a few were calibrated automatically but without full consideration of conflicting multi-objectives. Here, we successfully applied the non-dominated sorting genetic algorithm II (NSGA-II) and its two variants, namely the reference point-based R-NSGA-II and the extension ER-NSGA-II, to multi-objective, automatic calibration of the SHETRAN hydrological model. Moreover, we demonstrated the possibility of speeding up the calibration process by adjusting the recombination and mutation parameters of the optimization algorithms. The simulated binary crossover and polynomial mutation were used with respective probabilities of 0.9 and 0.1, and crossover and mutation distribution indices ($\eta_c, \eta_m$) with values of (0.5, 0.5), (2.0, 0.5) and (20., 20.). The results indicate that the use of smaller ($\eta_c, \eta_m$) speeded up the optimization process of SHETRAN calibration, especially during the initial stage, for all three algorithms; however, the use of the R-NSGA-II and ER-NSGA-II did not provide a more efficient optimization compared to the NSGA-II. The broad search of the algorithms, enabled by the generation of diversified solutions due to the use of small ($\eta_c, \eta_m$), contributed to the improved efficiency. Finally, we successfully validated the optimal solutions for both the basin outlet and the internal gauging stations.

Key words | multi-objective optimization, NSGA-II, physically based spatially distributed hydrological modelling, reference point approach, SHETRAN model, simulated binary crossover (SBX)

INTRODUCTION
Physically based, spatially distributed hydrological models have been widely used for evaluating the impacts of hypothetic changes resulting from climate (Bathurst et al. 1996; Kilsby et al. 2007; Gül et al. 2010; Bathurst 2011; Goderniaux et al. 2011; Thompson 2012; Mourato et al. 2014, 2015), land-use (McMichael & Hope 2007; Im et al. 2009; Birkinshaw et al. 2011; Kalantari et al. 2014), environment (Birkinshaw & Ewen 2000; Lutz et al. 2013; Hansen et al. 2014; Refsgaard et al. 2014) and management (Adams & Parkin 2002; Mottes et al. 2014). In practice, these models require calibration due to various reasons, namely approximations in model design, unavailability of parameters due to unaffordable cost, experiment constraints, as well as scaling problem (Beven 2012). The calibrations are complicated and expensive due to the sophisticated model structure, heavy computation requirements and a large number of calibration parameters (Refsgaard 1997; Madsen 2005; Blasone et al. 2007; Santos et al. 2012; Zhang et al. 2013). Consequently, calibrations were carried out manually, with consideration of a single objective and few parameters. The efficacy and efficiency of global optimization methods in automatic calibration have been demonstrated by Madsen (2003), Santos et al. (2003, 2012) and Zhang et al. (2013) for the MIKE-SHE, WESP and SHETRAN models. However, such applications are still sparse and few considered the full tradeoff characteristics of conflicted multi-objectives. The models configured with finer spatial resolution represent better the integrated hydrological process of a catchment (Vázquez et al. 2002;
Wildemeersch et al. (2014), which implies an increase of the simulation time for an individual model run. Consequently, automatic calibration of these models with requirements of tens of thousands of simulations to be performed would become unaffordable due to the demands of heavy computations. Therefore, a balance must be found between the desired model performance and manageable computing resources; as well, the calibration objectives should be achieved by a limited number of model simulations.

Multi-objective evolutionary algorithms (MOEAs) are effective and efficient in the calibration of distributed models. Successful examples can be found in Bekele & Nicklow (2007), Shafii & Smedt (2009) and Zhang et al. (2010), by using the non-dominated sorting genetic algorithm II (NSGA-II) (Deb et al. 2002), for the SWAT and WetSpa models. Since modellers are mostly interested in choosing one of a set of solutions instead of the entire non-dominated solutions, their preferred information can also be included to speed up the optimization process of calibration, as Deb et al. (2006) suggested in the reference point-based R-NSGA-II. Furthermore, the extension ER-NSGA-II proposed by Siegmund et al. (2012) directs the optimization straight towards the preference points by using the constrained Pareto-fitness approach and a self-adaptive ε-clearing strategy. These methods suggest great potential in finding a preferred diverse set of non-dominated solutions for automatic calibration of physically based, spatially distributed models in a limited number of model simulations.

The NSGA-II and its variants use the simulated binary crossover (SBX) (Deb & Agrawal 1995) and polynomial mutation (PM) operators (Deb 2001) to generate offspring solutions from the selected parent individuals. The operators involve parameters, called crossover \( \eta_c \) and mutation \( \eta_m \) distribution indices, which determine the spread of offspring solutions relative to the parent solutions. On the one hand, the use of large values for both \( \eta_c \) and \( \eta_m \) results in offspring solutions close to the parent solutions; on the other hand, the use of small values leads to solutions apart from the parents (Deb 2001). Therefore, the selection of appropriate values of \( \eta_c \) and \( \eta_m \) is highly crucial, considering the importance of diversity control of offspring solutions in the optimization processes. However, the proper values may vary with different problems and also during the optimization processes. For example, Deb et al. (2007) developed a self-adaptive procedure for updating the value of \( \eta_c \) and found improvements in solving both single and multiple objective optimization problems; Zeng et al. (2010) proposed a self-adaptive mechanism to adjust the proper value of \( \eta_c \) and got better performances on benchmark problems; Deb & Deb (2014) identified different ranges of \( \eta_m \) for different problems. However, the importance of finding appropriate pairs \( (\eta_c, \eta_m) \) for different problems or even for different stages of the same problem-solving process, has not been paid enough attention to for the hydrological model calibration problems; as one can notice, the use of \( (\eta_c, \eta_m) \) is simply with fixed values in the ranges of \( (15, 20) \), \( (15, 20) \) (Bekele & Nicklow 2007; Tang 2007; Tang et al. 2007; Zhang et al. 2010; Kollat et al. 2012), which may have prevented the algorithms from obtaining better performances.

In the context of anti-desertiﬁcation strategies, this study presents the application of the NSGA-II and its two variants R-NSGA-II and ER-NSGA-II to the multi-objective, automatic calibration of the physically based, spatially distributed SHETRAN hydrological model for a semi-arid middle-sized basin in southern Portugal with high concerns about both water scarcity and ﬂash ﬂoods problems. To complete the above-mentioned limitations, we investigate the effects of the recombination and mutation parameters of the three algorithms on the optimization processes. Since the use of small \( \eta_c \) and \( \eta_m \) may lead to a quicker and more thorough search of the calibration parameters in the physically possible ranges, this study proposes the use of \( (\eta_c, \eta_m) \) with values of \( (0.5, 0.5) \) and \( (2.0, 0.5) \), in comparison to \( (20.0, 20.0) \), which is a typical pair of values used in previous hydrological applications with the three algorithms, namely NSGA-II, R-NSGA-II and ER-NSGA-II, for the automatic calibration of the SHETRAN hydrological model. It is intended to ﬁnd out whether the NSGA-II and its two variants, R-NSGA-II and ER-NSGA-II, can automatically calibrate the SHETRAN model and whether the use of smaller \( (\eta_c, \eta_m) \) can accelerate the optimization processes. The results will be used for evaluating the climate and land-use change impacts on basin hydrology and sediment transport, in areas undergoing desertification.
DATA AND METHODS

Cobres basin

The Cobres basin (37°28′N–37°57′N, 8°10′W–7°51′W, Figure 1) is a sub-basin of the Cobres River, a tributary of the Guadiana River, located in the Alentejo region of southern Portugal. It was selected mainly due to two reasons: first, it is one of the basins that has been well studied concerning hydrological behaviours by using the physically based, spatially distributed hydrological model SHETRAN (Bathurst et al. 1996, 2002; Mourato 2010; Bathurst 2011; Zhang et al. 2013, Zhang 2015; Mourato et al. 2014, 2015), which makes it possible to perform comparisons of the optimized results obtained in this study, with those from the previous studies; second, the basin is a representative area of a Mediterranean region in terms of soil, land-use and climate and has been studied for investigating Mediterranean desertification since the 1990s (Bathurst et al. 1996; Mourato 2010; Zhang 2015). The basin is a middle-sized, slightly sloping agricultural basin with an area of 705 km² and elevation ranging from 103 to 308 m. The soils are characterized by shallow depths of 10 to 50 cm to bedrock, and the main soil types are clay (20%), loam (46%) and sandy loam (34%). The dominant land-use types are crops (70%) and agroforestry (27%). The climate in this region is characteristically Mediterranean and Continental, being classified as Csa by Peel et al. (1988). It has moderate winters and hot and dry summers, high daily temperature range, and a weak and irregular precipitation regime; mean annual precipitation varies between 400 and 900 mm, in general concentrated over 50 to 80 days per year (Ramos & Reis 2002). The mean annual potential evapotranspiration (PET) is around 1,200 mm.

Data used for simulating the Cobres basin include a 10 m resolution digital elevation model, taken from the website of the Portuguese Geographic Information System (http://www.igeoe.pt/), a soil map with a scale of 1:25,000 provided by the Institute of Hydraulics, Rural Engineering and Environment, IHERA, a land-use map with a scale of 1:100,000 from CORINE Land Cover 2006 for Continental Portugal (Caetano et al. 2009), hourly weather and streamflow data provided by the Portuguese Water Resources Information System (SNIRH) for the stations indicated in Figure 1. Hourly PET was calculated by FAO Penman–Monteith method (Allen et al. 1998). Details concerning data can be found in Zhang et al. (2013). In this study, the multi-objective calibration is carried out for the basin outlet from October 1st, 2004 to September 30th, 2006 with the first 10 months as a warm-up period; the calibrated parameters are validated for the basin outlet and the internal gauging stations from October 1st, 2006 to September 30th, 2008. Due to missing data, the period from November 4th, 2006 to November 7th, 2006 is not considered for the basin outlet and for the two internal gauging stations.
SHETRAN modelling system

SHETRAN (http://research.ncl.ac.uk/shetran/) is a physically based, spatially distributed modelling system for water flow and sediment and contaminant transports in river catchments (Ewen et al. 2000; Birkinshaw et al. 2010). The model has a modular design and mainly consists of three components, namely, those concerning water flow, sediment transport and solute transport. These components model the physical processes by finite difference representations of the partial differential equations of mass and energy conservation and by empirical equations derived from independent experimental research. In SHETRAN, a basin is discretized by an orthogonal grid network in the horizontal view and by a column of horizontal layers at each grid square in the vertical view; as well, the river network is simplified by the links running along the edges of the grid squares (Figure 1).

In this paper, the automatic calibration is considered only for the water flow component of SHETRAN (v4.4.0). In the present application, the model represents the physical processes of the hydrological cycle through: (1) the interception calculated from the modified Rutter model; (2) the actual evapotranspiration (AET) calculated from a prescribed ratio of AET/PET as a function of soil water potential; (3) the overland (2D) and channel (1D) flow processes based on the diffusive wave approximation of the Saint–Venant equations; (4) the subsurface flow processes calculated from 3D variably saturated flow equation; (5) the river aquifer interaction calculated from the Darcy equation.

The objective functions of SHETRAN calibration

The objective of SHETRAN calibration is to simultaneously minimize the root mean square error (RMSE) (Moriasi et al. 2007) and the logarithm transformed RMSE (LOGE) (Bennett et al. 2013) between observed and simulated hourly discharges at the basin outlet. RMSE emphasizes fitting of the higher or peak discharges due to the square of errors greater than 1.0 and LOGE is designed to emphasize fitting of the lower discharges through the introduction of logarithms. Both of them range between 0 (perfect match) and $+\infty$. The Nash–Sutcliffe efficiency (NSE) (Nash & Sutcliffe 1970) is also evaluated to facilitate the comparisons of model performances with other studies. NSE is a standardized measure of RMSE, ranging from $-\infty$ to 1 (perfect fit). NSE is linearly related to RMSE$^2$, which is indicated by Equation (4) in this study.

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (O_i - S_i)^2}
\]  
(1)

\[
LOGE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\log \left( \frac{O_i}{S_i} \right))^2}
\]  
(2)

\[
NSE = 1.0 - \frac{\sum_{i=1}^{n} (O_i - S_i)^2}{\sum_{i=1}^{n} (O_i - \bar{O})^2}
\]  
(3)

\[
(1 - NSE) = \frac{nRMSE^2}{\sum_{i=1}^{n} (O_i - \bar{O})^2} = 0.015934RMSE^2
\]  
(4)

where $O_i$ and $S_i$ are, respectively, observed and simulated discharges; $n$ is the total number of data; $\bar{O}$ is the mean value of the observed discharges.

The MOEAs

Similarities and differences between NSGA-II, R-NSGA-II and ER-NSGA-II

The R-NSGA-II (Deb et al. 2006) and ER-NSGA-II (Siegmund et al. 2012) are two variants of the NSGA-II algorithm (Deb et al. 2002). The three algorithms follow the same procedure: (a) generation of an initial population $P$; (b) selection of parent solutions by tournament; (c) generation of offspring solutions by crossover and mutation; (d) evaluation of new offspring solutions $Q$; (e) selection of a new population, out of $P \cup Q$; (f) check of convergence criteria; steps (b)–(f) are repeated until the convergence criteria are satisfied. In this study, 30 sets of the initial population are generated by Latin hypercube sampling to consider the random seed effects. For all three algorithms, the non-dominated sorting
method is used to rank the solutions from high to low non-dominated levels, in terms of all objective functions, so that the solutions from higher levels are absolutely better than those from lower levels but those from the same level are not comparable unless a second fitness measure is specified; SBX and PM are used to generate offspring solutions with respective probabilities of 0.9 and 0.1; convergence is considered being satisfied when the number of SHETRAN simulations reaches 1,500 or all the changes of performance metrics of the algorithm, during the recent five generations, are lower than the limits described in the section ‘The performance metrics of MOEAs’.

The three algorithms are different from each other mainly in the selection criteria and diversity maintenance of steps (b) and (e). As for selection, all the solutions are distinguished first by their non-dominance level, namely the Pareto-fitness; and they are then differentiated by the second fitness measure, which is crowding distance for the NSGA-II, and preference distance for the R-NSGA-II, as indicated in Equations (5) and (6):

\[ d_c = \sum_{k=1}^{m} \frac{|f^k_{i+1} - f^k_{j-1}|}{f^k_{\max} - f^k_{\min}} \] (5)

\[ d_p = \sqrt{\sum_{k=1}^{m} \left( \frac{f^k_i - p^k}{f^k_{\max} - f^k_{\min}} \right)^2} \] (6)

where \( d_c \) is the crowding distance of the \( j^{th} \) solution, \( f \) is the objective function of SHETRAN calibration, \( m \) is the number of objective functions, \( k \) represents a specific objective function, \( f^k_{\max} \) and \( f^k_{\min} \) indicate the maximum and minimum values of the objective function; \( d_p \) is the Euclidean distance between the \( j^{th} \) solution and the \( l^{th} \) reference point \( p \). Large \( d_c \) and small \( d_p \) are preferable for a good calibration.

In terms of diversity maintenance, the NSGA-II prefers a large crowding distance; and the R-NSGA-II and ER-NSGA-II take it into account by keeping solutions with distances larger than epsilon (\( \epsilon \)), among each other. In ER-NSGA-II, the value of \( \epsilon \) is adaptable with the intention of providing the decision-maker with a compact crowded set of solutions near to the reference points, and a more spread set of solutions at the outer ends of the fronts. As indicated in Equations (7) and (8), the value of \( \epsilon \) is proportional to both the distance between the Pareto-front and reference point and the distance between the solutions inside a front. To speed up optimization, the constrained Pareto-fitness approach is used in step (e) of ER-NSGA-II, which selects a predefined number of solutions, from the first fronts, nearest to the reference points.

\[ \epsilon_{ql} = (1 + a\epsilon_l)\epsilon_l \] (7)

\[ \epsilon_l = \frac{d_lK}{2N} \in \left[ \frac{\epsilon}{2}, \frac{\epsilon K}{2N} \right] \] (8)

In these equations, \( \epsilon_{ql} \) is the \( \epsilon \) value assigned to the \( q^{th} \) solution in the cluster of the \( j^{th} \) solution which is nearest to the \( l^{th} \) reference point, \( d_{ql} \) is the Euclidean distance between the \( q^{th} \) and \( j^{th} \) solutions and \( a \) is a user-defined parameter; \( \epsilon_l \) is the \( \epsilon \) value of the \( l^{th} \) reference point, \( d_l \) is the Euclidean distance between the \( l^{th} \) reference point and its closest solution in the Pareto-front, \( K \) and \( N \) are, respectively, the numbers of reference points and population, the minimum and maximum limits of \( \epsilon_l \) being defined as \( \epsilon \) and \( \frac{\epsilon K}{2N} \).

**SBX and PM**

The SBX operator (Deb & Agrawal 1995; Deb 2001) creates two offspring solutions \( x_i^{(1, t+1)} \) and \( x_i^{(2, t+1)} \) from the two parent solutions \( x_i^{(1, t)} \) and \( x_i^{(2, t)} \) by using the polynomial probability distributions indicated in Equation (9). The PM operator (Deb & Goyal 1996; Deb 2001) generates offspring solution \( x_i^{(3, t+1)} \) from the parent solution \( x_i^{(3, t)} \) by using the polynomial probability distribution shown in Equation (10).

\[ P(\beta_t) = \begin{cases} 
0.5(\eta_c + 1)\beta_t^m, & \text{if } \beta_t \leq 1; \\
0.5(\eta_c + 1) \frac{1}{\beta_t^{m+2}}, & \text{otherwise.} 
\end{cases} \] (9)

\[ P(\delta_t) = 0.5(\eta_m + 1)(1 - |\delta_t|)^{n_m} \] (10)

\[ \beta_t = \left| \frac{x_i^{(2, t+1)} - x_i^{(1, t+1)}}{x_i^{(2, t)} - x_i^{(1, t)}} \right| \text{ and } \delta_t = \left| \frac{x_i^{(3, t+1)} - x_i^{(3, t)}}{x_i^{(3, t)} - x_i^{(2, t)}} \right|. \]
where $i$ being the SHETRAN calibration parameter, $P$ is a probability function, $\eta_c$ and $\eta_m$, $\eta_c \in (0, + \infty)$ and $\eta_m \in (0, + \infty)$, are, respectively, crossover and mutation distribution indices. Larger values of $\eta_c$ and $\eta_m$ give higher probabilities in creating solutions near to the parent solution, while smaller values allow distant points to be selected as children. $\beta_i$ and $\delta_i$, $\beta_i \in (0, \infty)$ and $\delta_i \in (-1, +1)$, are, respectively, the spread and perturbation factors and $x_{i}^{(L)} - x_{i}^{(U)}$ is the defined feasible range of the $i^{th}$ parameter.

During the SBX operation, $\beta_i$ can be obtained by equating the area under the probability curve to a random number $u_i \in (0, 1)$, as in Equation (11), and the offspring solutions can be obtained by Equations (12) and (13). For the PM operation, $\delta_i$ can be calculated by Equation (14) and a random number $r_i \in (0, 1)$ and the offspring solution can be derived from Equation (15).

$$
\beta_i = \begin{cases} 
\frac{1}{(2u_i)^{\eta_c} + 1}, & \text{if } u_i \leq 0.5; \\
\frac{1}{2(1 - u_i)}\eta_c + 1, & \text{otherwise} 
\end{cases}
$$  

(11)

$$
x_{i}^{(1,i+1)} = 0.5 \left( (1 + \beta_i) x_{i}^{(1,U)} + (1 - \beta_i) x_{i}^{(2,U)} \right)
$$  

(12)

$$
x_{i}^{(2,i+1)} = 0.5 \left( (1 - \beta_i) x_{i}^{(1,U)} + (1 + \beta_i) x_{i}^{(2,U)} \right)
$$  

(13)

$$
\delta_i = \begin{cases} 
\frac{1}{(2r_i)^{\eta_m} + 1} - 1, & \text{if } r_i \leq 0.5; \\
1 - \left[ 2(1 - r_i)^{\eta_m} + 1 \right], & \text{otherwise} 
\end{cases}
$$  

(14)

$$
x_{i}^{(3,i+1)} = x_{i}^{(3,U)} + (x_{i}^{(U)} - x_{i}^{(L)}) \delta_i
$$  

(15)

Configurations of the optimizations for SHETRAN calibration

The Cobres basin is discretized into 175 grid squares of $2 \times 2$ km$^2$ (Figure 1), to each of which are assigned parameters with unique values. The SHETRAN calibration parameters are the Strickler overland flow resistance coefficient and AET/PET ratio for the grid squares with two main types of land-use and the top soil depth, saturated hydraulic conductivity, saturated water content, residual water content and van Genuchten $n$ and $\alpha$ parameters for those with three main types of soil. As shown in Table 1, the parameter values are constrained within physically realistic ranges according to field measurements of Cardoso (1965) and literature data (Bathurst et al. 1996, 2002; Saxton & Rawls 2006). Details of the SHETRAN model set-up and parameter ranges can be found in Zhang et al. (2013). As for the configuration of MOEAs, the population size is set to 50, the maximum number of generations is 30, and the ($\eta_c$, $\eta_m$) parameters are given three alternative sets $(0.5, 0.5)$, $(2.0, 0.5)$ and $(2.0, 2.0)$, based on the recommended value of $(2.0, 2.0)$ by Deb et al. (2002) and the values previously applied to hydrological applications (Bekele & Nicklow 2007; Tang 2007; Zeng et al. 2010), for all three algorithms NSGA-II, R-NSGA-II and ER-NSGA-II; three reference points with (RMSE, LOGE) of $(1.50, 2.50)$, $(1.25, 2.65)$ and $(1.00, 2.80)$ are set for the R-NSGA-II and ER-NSGA-II, based on the possible ranges of results obtained from a trial run by NSGA-II and as explained in Zhang et al. (2015). For R-NSGA-II, the $\epsilon$ is set to 0.001, according to Deb et al. (2006) and Siegmund et al. (2012). For ER-NSGA-II, the $a$, $\xi$ and $\Zeta$ are set as 0.5, 0.001 and 0.5 to allow for a variation of $\epsilon$ in the range of [0.001, 0.026]; the constraint of the Pareto-fitness approach recommended by Siegmund et al. (2012) is used, and, for step (e) in the section ‘Similarities and differences between NSGA-II, R-NSGA-II and ER-NSGA-II’, the best five non-dominated levels of $P \cup Q$ are supposed to provide 28, 12, 6, 3 and 1 solutions as new populations. A Matlab code was developed for the above-mentioned optimizations based on Seshadri (2009) and Lin (2011).

The performance metrics of MOEAs

Convergence and diversity are two orthogonal goals of a successful MOEA (Deb 2001). This paper evaluates the two aspects of optimization processes of SHETRAN calibration by using the performance metrics hypervolume, $\epsilon$-indicator, generational distance (Reed et al. 2015) and opt-indicator. The normalized objective function values are used for the evaluation, and a trial run optimization is carried out to
Table 1  | The SHETRAN calibration parameters’ description, feasible ranges and values derived from automatic calibrations

<table>
<thead>
<tr>
<th>Parameters (unit)</th>
<th>Description</th>
<th>Range</th>
<th>Sim1a</th>
<th>Sim1b</th>
<th>Sim1</th>
<th>Sim2</th>
<th>Sim3</th>
<th>Sim4</th>
</tr>
</thead>
<tbody>
<tr>
<td>K₁ (m⁰.₃s⁻¹)</td>
<td>Strickler overland flow resistance coefficient for crops</td>
<td>2.5–10.0</td>
<td>9.6</td>
<td>9.7</td>
<td>10.0</td>
<td>10.0</td>
<td>9.1</td>
<td>3.2</td>
</tr>
<tr>
<td>K₂ (m⁰.₃s⁻¹)</td>
<td>Strickler overland flow resistance coefficient for agroforestry</td>
<td>0.5–5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>K₃ (m/day)</td>
<td>Saturated hydraulic conductivity of Vₓ soil</td>
<td>0.110–0.192</td>
<td>0.174</td>
<td>0.190</td>
<td>0.184</td>
<td>0.192</td>
<td>0.192</td>
<td>0.188</td>
</tr>
<tr>
<td>θ₀₁ (m³/m³)</td>
<td>Saturated water content of Vₓ soil</td>
<td>0.506–0.517</td>
<td>0.517</td>
<td>0.517</td>
<td>0.511</td>
<td>0.517</td>
<td>0.506</td>
<td>0.513</td>
</tr>
<tr>
<td>θ₁₁ (m²/m³)</td>
<td>Residual water content of Vₓ soil</td>
<td>0.065–0.073</td>
<td>0.067</td>
<td>0.066</td>
<td>0.068</td>
<td>0.073</td>
<td>0.071</td>
<td>0.069</td>
</tr>
<tr>
<td>n₁ (-)</td>
<td>van Genuchten n of Vₓ soil</td>
<td>1.221–1.403</td>
<td>1.221</td>
<td>1.221</td>
<td>1.221</td>
<td>1.221</td>
<td>1.224</td>
<td>1.226</td>
</tr>
<tr>
<td>α₁ (cm⁻¹)</td>
<td>van Genuchten α of Vₓ soil</td>
<td>0.0055–0.0250</td>
<td>0.0056</td>
<td>0.0055</td>
<td>0.0057</td>
<td>0.0055</td>
<td>0.0055</td>
<td>0.0055</td>
</tr>
<tr>
<td>θ₂₁ (m³/m³)</td>
<td>Saturated water content of Pₓ soil</td>
<td>0.418–0.519</td>
<td>0.418</td>
<td>0.418</td>
<td>0.418</td>
<td>0.418</td>
<td>0.493</td>
<td>0.418</td>
</tr>
<tr>
<td>θ₂₂ (m³/m³)</td>
<td>Residual water content of Pₓ soil</td>
<td>0.041–0.053</td>
<td>0.053</td>
<td>0.053</td>
<td>0.053</td>
<td>0.042</td>
<td>0.051</td>
<td>0.048</td>
</tr>
<tr>
<td>n₂ (-)</td>
<td>van Genuchten n of Pₓ soil</td>
<td>1.345–1.422</td>
<td>1.347</td>
<td>1.347</td>
<td>1.347</td>
<td>1.345</td>
<td>1.345</td>
<td>1.358</td>
</tr>
<tr>
<td>α₂ (cm⁻¹)</td>
<td>van Genuchten α of Pₓ soil</td>
<td>0.0075–0.0225</td>
<td>0.0075</td>
<td>0.0075</td>
<td>0.0075</td>
<td>0.0075</td>
<td>0.0076</td>
<td>0.0088</td>
</tr>
<tr>
<td>h₁ (m)</td>
<td>Top soil depth of Vₓ soil</td>
<td>0.30–0.65</td>
<td>0.30</td>
<td>0.30</td>
<td>0.33</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td>K₄₁ (m/day)</td>
<td>Saturated hydraulic conductivity of Pₓ soil</td>
<td>0.191–0.425</td>
<td>0.424</td>
<td>0.423</td>
<td>0.425</td>
<td>0.423</td>
<td>0.423</td>
<td>0.425</td>
</tr>
<tr>
<td>θ₃₁ (m³/m³)</td>
<td>Saturated water content of Eₓ soil</td>
<td>0.446–0.457</td>
<td>0.446</td>
<td>0.446</td>
<td>0.446</td>
<td>0.446</td>
<td>0.452</td>
<td>0.446</td>
</tr>
<tr>
<td>θ₃₂ (m³/m³)</td>
<td>Residual water content of Eₓ soil</td>
<td>0.051–0.120</td>
<td>0.082</td>
<td>0.065</td>
<td>0.073</td>
<td>0.120</td>
<td>0.083</td>
<td>0.051</td>
</tr>
<tr>
<td>n₃ (-)</td>
<td>van Genuchten n of Eₓ soil</td>
<td>1.511–1.557</td>
<td>1.488</td>
<td>1.381</td>
<td>1.557</td>
<td>1.544</td>
<td>1.321</td>
<td>1.311</td>
</tr>
<tr>
<td>α₃ (cm⁻¹)</td>
<td>van Genuchten α of Eₓ soil</td>
<td>0.0250–0.0690</td>
<td>0.0252</td>
<td>0.0250</td>
<td>0.0275</td>
<td>0.0670</td>
<td>0.028</td>
<td>0.025</td>
</tr>
<tr>
<td>h₃ (m)</td>
<td>Top soil depth of Eₓ soil</td>
<td>0.05–0.10</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.10</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>AETPETFC₁ (-)</td>
<td>The AET/PET ratio at field capacity for crop</td>
<td>0.50–0.90</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.502</td>
<td>0.501</td>
</tr>
<tr>
<td>AETPETFC₂ (-)</td>
<td>The AET/PET ratio at field capacity for agroforestry</td>
<td>0.60–0.80</td>
<td>0.612</td>
<td>0.617</td>
<td>0.622</td>
<td>0.600</td>
<td>0.600</td>
<td>0.607</td>
</tr>
</tbody>
</table>

*`Sim1a`, `Sim1b`, `Sim1`, `Sim2`, `Sim3` and `Sim4` are SHETRAN simulations with (RMSE, LOGE, NSE) of respectively (2.82, 2.73, 0.87), (2.83, 2.73, 0.87), (2.81, 2.74, 0.87), (3.81, 2.53, 0.77), (4.85, 2.49, 0.63) and (5.85, 2.46, 0.46) at basin outlet, for the calibration period (2004–2006). *asim1a`, `Sim1b` and `Sim1` are equivalently good; the hydrograph comparisons are shown in Figure 8 for the simulations `Sim1`, `Sim2`, `Sim3` and `Sim4`. *asim1a`, `Sim1b` and `Sim1` are equivalently good; the hydrograph comparisons are shown in Figure 8 for the simulations `Sim1`, `Sim2`, `Sim3` and `Sim4`. *asim1a`, `Sim1b` and `Sim1` are equivalently good; the hydrograph comparisons are shown in Figure 8 for the simulations `Sim1`, `Sim2`, `Sim3` and `Sim4`. *asim1a`, `Sim1b` and `Sim1` are equivalently good; the hydrograph comparisons are shown in Figure 8 for the simulations `Sim1`, `Sim2`, `Sim3` and `Sim4`. *asim1a`, `Sim1b` and `Sim1` are equivalently good; the hydrograph comparisons are shown in Figure 8 for the simulations `Sim1`, `Sim2`, `Sim3` and `Sim4`.  

To determine the possible maximum and minimum SHETRAN objective values, by specifying a vector of worst objective values, the hypervolume assesses the ratio of volume dominated by the local Pareto approximate set to the volume dominated by the reference Pareto approximate set. The e-indicator estimates the largest distance required to translate a PF solution to dominate its nearest neighbour in the PF. The generational distance measures the average Euclidean distance of points in PF to their nearest corresponding objective vectors in the PF. Details of the definitions are referred to in Nicklow et al. (2010) and Reed et al. (2013). The opt-indicator, defined in this study, indicates the smallest distance required for translating a PF solution to dominate its nearest neighbour in the PF. As it measures the least effort for getting one more PF solution dominating the nearest neighbour in the PF, the opt-indicator is the easiest metric to be improved comparing to the others.

The metrics evaluation is performed with two purposes: first, determination of convergence for each optimization process; second, comparison of performance among different optimization processes. The assessment of convergence by
performance metrics is done every five generations evolution. The convergence is considered satisfied if the changes of hypervolume, $\varepsilon$-indicator, generational distance and opt-indicator are all less than 0.01%. The comparisons among the nine optimizations are implemented only after their completion.

**RESULTS AND DISCUSSION**

The nine optimizations of SHETRAN calibration were carried out by using NSGA-II, R-NSGA-II and ER-NSGA-II algorithms with $(\eta_c, \eta_m)$ of (0.5, 0.5), (2.0, 0.5) and (20., 20.).
All of them were terminated by the time limit established for SHETRAN simulation, before other convergence criteria were satisfied. We compare the optimization processes with two-fold intentions, which are, on the one hand, to find out whether the use of smaller \((\eta_c, \eta_m)\) accelerates the optimization processes of the three algorithms, and, on the other hand, to determine whether the use of R-NSGA-II and ER-NSGA-II improve the performance metrics, in comparison with NSGA-II. As shown in Figures 2–8, the comparisons were made based on the varied non-dominated solutions from each trial run along the evolution of the optimization processes. Figures 2(a)–2(l), 3(a)–3(b), 7(a)–7(f) and 8(a)–8(d) show comparisons of SHETRAN performances for the final results from the nine optimizations. Figures 4(a)–4(l), 5(a)–5(d) and 6(a)–6(d) show comparisons of performance metrics of the three algorithms during the optimization processes.

Comparisons of the use of \((\eta_c, \eta_m)\) with different values

It is shown that, for NSGA-II, R-NSGA-II and ER-NSGA-II, the use of \((\eta_c, \eta_m)\) with value of \((0.5, 0.5)\) gets the best solutions and is most efficient; the use of \((2.0, 0.5)\) gets better solutions and is more efficient than that of \((20., 20.)\). As shown in Figures 2(a)–2(c) and 3(a)–3(b), the use of smaller \((\eta_c, \eta_m)\) makes the three algorithms produce better solutions. Figure 2(a)–2(c) demonstrates the evidence by comparison of ensembles of non-dominated solutions obtained from the 30 trial runs of each optimization after 1,500 SHETRAN simulations; this is displayed in Figure 3(a) through comparisons of the non-dominated solutions of the ensembles, shown in Figure 2(a)–2(c), for all optimizations. On the other hand, it is shown that the use of smaller \((\eta_c, \eta_m)\) may reduce the random seed effects of the three algorithms. As one may notice, this is shown in Figure 2(a)–2(c) from the smaller objective spaces of the ensembles, associated with optimizations with the smaller \((\eta_c, \eta_m)\).

The effects of smaller \((\eta_c, \eta_m)\) in accelerating the optimization are most pronounced in the first, but not in the second, 15 generations of evolutionary processes (or 750 SHETRAN simulations), and the discrepancies vary with different performance metrics. If we define the efficiency of optimization process as the rate of improvement in performance metrics with the number of model runs, the effects of the use of smaller \((\eta_c, \eta_m)\) are clearly noticed by comparisons displayed in Figure 4(a)–4(l). All nine optimizations mostly start...
with their best efficiencies, the smaller the value of $(\eta_c, \eta_m)$ the higher the efficiency; and then, all efficiencies decrease and stabilize in similar values for the second set of evolutionary processes (or 750 SHETRAN simulations). Figures 5(a)–5(d) and 6(a)–6(d) also show the same story, by displaying, respectively, the 50th and 95th percentiles of the dynamic performances. The SBX operator is mainly responsible for this phenomenon, since the probability was set as 0.9 for crossover and 0.1 for mutation operators; and the explanation is probably related to the self-adaptive nature of the SBX operator (Deb et al. 2007). Nevertheless, the use of 0.5 as $\eta_m$ has also promoted the differences of performance metrics between optimizations with $(\eta_c, \eta_m)$ of $(0.5, 0.5)$, $(2.0, 0.5)$ and that of $(20., 20.)$.

The spread of offspring is proportional to that of parent solutions and related to $\eta_c$ as well, as clearly shown in Equations (11), (12) and (13). At the beginning of optimizations, the spread of parent solutions is very large, due to

---

**Figure 4** | Plots of dynamic performance results, namely hypervolume, $\epsilon$-indicator, generational distance and opt-indicator, of the NSGA-II (a), (d), (g) and (j), R-NSGA-II (b), (e), (h) and (k) and ER-NSGA-II (c), (f), (i) and (l) algorithms versus total number of SHETRAN model runs. The shadows with different colours show ranges of performances derived from the use of algorithms with $(\eta_c, \eta_m)$ of $(0.5, 0.5)$, $(2.0, 0.5)$ and $(20., 20.)$ respectively. The bold solid and thin dashed lines represent the respective mean and standard deviation of performances. Each use of the algorithms was repeated 30 times to consider the random seed effects. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/hydro.2015.219.
the random selection of initial solutions, which therefore create offspring solutions far away from the parents and the use of smaller value of $\eta_c$ even further enlarges the spread of offspring. From Figure 4(a)–4(l), it is evident that the broad search of the algorithms in their early generations enables the efficient improvement of the optimizations. However, when the solutions tend to converge due to the action of genetic operators, the spread of parent solutions becomes smaller which diminishes the effect of $\eta_c$ on the spread of offspring solutions; consequently, distance solutions are not likely to be generated and the optimization search focuses on a narrower region. This explains the gradually decreasing efficiency of the nine optimizations during the first 15 generations of evolutionary processes (or 750 SHETRAN simulations). As for $\eta_m$, Equations (14) and (15) demonstrate that the use of smaller $\eta_m$ also favours the creation of distant offspring solutions, which thus facilitates a broader search for the optimizations with $(\eta_c, \eta_m)$ of $(0.5, 0.5)$ and $(2.0, 0.5)$ rather than that of $(20., 20.)$. Furthermore, as shown in Equation (15), the mutation operator introduces diversified solutions with the same efficiency during all the optimization processes. This may explain the nearly constant performance of the nine optimizations during the second 15 generations of evolutionary processes (or 750 SHETRAN simulations). All in all, it is shown in Figure 4(a)–4(l) that a broad search is highly important for all three algorithms.

Figure 5 | Plots for the 50th percentiles of the dynamic performances, namely hypervolume (a), $\varepsilon$-indicator (b), generational distance (c) and opt-indicator (d), of the NSGA-II, R-NSGA-II and ER-NSGA-II algorithms versus total number of SHETRAN model runs, respectively, shown in bold solid, normal solid and normal dashed lines. All the optimizations with $(\eta_c, \eta_m)$ of $(0.5, 0.5), (2.0, 0.5)$ and $(20., 20.)$ are shown, respectively, in different colours. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/hydro.2015.219.
Comparisons among the NSGA-II, R-NSGA-II and ER-NSGA-II

It is shown that, by considering the preference points and maintaining solutions with distances larger than 0.001, the R-NSGA-II does not improve, compared to the NSGA-II; and in the case of the use of \((\eta_c, \eta_m)\) with value of \((20., 20.)\), the R-NSGA-II demonstrates a worse performance than the NSGA-II. This is evident in Figures 2(d)–2(f) and 3(a) by comparing the final results from the 30 trial runs of each optimization, after 1,500 SHETRAN simulations. From Figure 2(d) and 2(e), the ensembles of non-dominated solutions obtained from the R-NSGA-II, by using \((\eta_c, \eta_m)\) of \((0.5, 0.5)\) or \((2.0, 0.5)\), are neither better nor worse than those from the NSGA-II; however, from Figure 2(f), by using \((\eta_c, \eta_m)\) of \((20., 20.)\), R-NSGA-II performs worse than NSGA-II. As well, these conclusions are also true when one compares the performance metrics of the algorithms during the optimization processes. In Figure 4(a)–4(l), the first two columns show that, by using \((\eta_c, \eta_m)\) with values of \((0.5, 0.5)\) and \((2.0, 0.5)\), the average performances of the NSGA-II are always equivalent or slightly better than those from the R-NSGA-II; by using \((\eta_c, \eta_m)\) with values of \((20., 20.)\), the NSGA-II demonstrates much better performances than the R-NSGA-II. Figures 5(a)–5(d) and 6(a)–6(d) also support the same conclusions by comparing the 50th and 95th percentiles of results obtained from the 30 trial runs.

It is also shown that, by considering the preference points, the constrained Pareto-fitness approach and the self-adaptive \(\epsilon\) approach, the ER-NSGA-II does not improve, compared to NSGA-II and R-NSGA-II; and in the case of the use of \((\eta_c, \eta_m)\) with a value of \((20., 20.)\), the ER-NSGA-II displays a
performance similar to that of R-NSGA-II, which is worse than that obtained by the NSGA-II. This is evident in Figures 2(d)–2(f) and 3(a) by comparison of final results; and, as shown in Figures 4(a)–4(l), 5(a)–5(d) and 6(a)–6(d), it is also apparent when it compares the performance metrics of the algorithms during the optimization processes.

From Figures 5(a)–5(d) and 6(a)–6(d), it is clear that using smaller \((\eta_c, \eta_m)\) enables all three algorithms to get much better performances; however, by using the same \((\eta_c, \eta_m)\), the R-NSGA-II and ER-NSGA-II cannot get improvements compared to NSGA-II. As explained in the section ‘Comparisons of the use of \((\eta_c, \eta_m)\) with different values’, the use of smaller \((\eta_c, \eta_m)\) encourages the generations of offsprings further distant from their parent solutions, especially in the early stage of the optimization, which largely increases the diversity of the searched population. For the three algorithms, the diversity of the searched population is determined primarily by the use of \((\eta_c, \eta_m)\) and secondarily by the particular diversity maintenance in the second measure of population selection. Therefore, by the use of smaller \((\eta_c, \eta_m)\), the NSGA-II, R-NSGA-II and ER-NSGA-II show nearly equivalently good performances; and the effect of different diversity maintenance in the second measure of population selection can be seen from the comparison of the three algorithms with the use of larger \((\eta_c, \eta_m)\). Compared to NSGA-II, the R-NSGA-II and ER-NSGA-II abandon the crowding distance as the second measure for population selection, but adopt the preference distance instead, by considering the decision-maker’s preferences and by using the \(\varepsilon\)-clearing strategy to maintain diversity of solutions. However, our application has shown no improvements brought about by the R-NSGA-II and ER-NSGA-II compared to NSGA-II. The direct inclusion of preference points is not necessarily more efficient than the use of crowding distance; this is probably due to their differences in the effects of stimulating

![Figure 7](https://iwaponline.com/jh/article-pdf/18/3/428/478800/jh0180428.pdf)
the generation of diversified solutions. As well, for the ER-NSGA-II, the constrained Pareto-fitness approach may have also contributed to the bad performances by increasing the probability of introducing solutions from worse non-dominated levels into the new population.

**Comparisons of the SHETRAN validation results**

The validation results are satisfactory, which is also in agreement with those obtained from the SHETRAN calibration. As shown in Figure 7(a)–7(f), the non-dominated solutions from the optimization of NSGA-II, R-NSGA-II and ER-NSGA-II with \((\eta_c, \eta_m)\) of (0.5, 0.5) are the best; and those obtained from optimizations with \((\eta_c, \eta_m)\) of (2.0, 0.5) are better than those of (20., 20.). The validation results, shown in Figure 7(d)–7(f), correspond well with the calibration results, shown in Figure 7(a)–7(c). In this study, the SHETRAN model was calibrated and validated at basin outlet, as displayed in Figures 7(a) and 7(d), but the model performances have also been evaluated at two internal gauging stations, as shown in Figure 7(b)–7(c) and 7(e)–7(f). The results show that both the calibration and validation results are satisfactory not only at basin outlet but also at basin internal gauging stations. Figure 7 shows that, in terms of NSE, the best solutions have values of 0.87 and 0.81 at basin outlet, 0.70 and 0.72 at internal gauging station Abernoa and 0.82 and 0.66 at internal gauging station Entradas, respectively, for calibration and validation.
and ‘sim1’, shown in Table 1, have different parameter settings but equivalently good model performances for both the calibration and validation simulations.

Visual comparisons between the observed and simulated hydrographs are made in Figure 8(a)–8(d), for one of the best solutions, for the largest storm events during the respective calibration and validation periods. The best solution ‘sim1’ was calibrated with (RMSE, LOGE, NSE) values of (2.81, 2.74, 0.87); the NSE is 0.89 and 0.75, respectively, for storms No. 1 and No. 4 at basin outlet (Figure 8(a)–8(b)) and it is 0.74 and 0.66 for storm No. 4 respectively at internal gauging stations Alberona and Entredas (Figure 8(c) and 8(d)). The ‘sim1a’ and ‘sim1b’ are not shown because they are indistinguishable from ‘sim1’. To illustrate the differences in SHETRAN objective values, Figure 8(a)–8(d) also display the other three solutions, namely ‘sim2’, ‘sim3’ and ‘sim4’, with calibrated (RMSE, LOGE, NSE) values of (3.81, 2.53, 0.77), (4.85, 2.49, 0.63) and (5.85, 2.46, 0.46). From Table 1, the main differences of the calibrated parameters, between the best solutions and the others are: Strickler overland flow resistance coefficients for both crops and agroforestry and top soil depth of \( V_x \) soil. In ‘sim1a’, ‘sim1b’ and ‘sim1’, the small top soil depth makes more water appear as surface runoff, and the large Strickler overland flow resistance makes surface runoff increase. This explains the larger surface runoff and storm discharges of the best solutions compared to the other ones.

The best solutions have better performances than the previous SHETRAN simulations carried out for the Cobres basin and the optimized parameters are well consistent with the literature data. By using the 12-minute resolution rainfall data, daily PET data and 2 km spatial resolution, Bathurst et al. (1996) obtained NSE of 0.83, 0.81 and 0.61 at basin outlet for the SHETRAN simulations of the Cobres basin, respectively, during the calibration period from 1977 to 1979 and the two validation periods from 1980 to 1982 and from 1983 to 1985. By using the hourly rainfall and PET data, 2 km spatial resolution and the modified shuffled complex evolution (MSCE) optimization algorithm and the same model configuration, Zhang et al. (2013) calibrated and validated the SHETRAN model at Cobres basin for the same periods selected in this study, namely from 2004 to 2006 and from 2006 to 2008, and they got the optimized solution with NSE at basin outlet of 0.86 and 0.74, respectively. The best solutions have similar or close calibration parameters, such as top soil depth, AET/PET ratio and Strickler overland flow resistance coefficients, to the optimized solution obtained by Zhang et al. (2013). The soil depth of \( V_x \) is 0.30 or 0.33 m, which is equal or close to 0.30 m obtained by Zhang et al. (2013) and well inside the range of [0.13, 0.33] m measured by Bathurst et al. (1996). The AET/PET ratio at field capacity is 0.500 and around 0.617, respectively, for crops and agroforestry, which is near their respective lower limits and similar to the values of 0.500 and 0.600 obtained by Zhang et al. (2013). The smaller calibrated AET/PET ratios show the attempt of optimization algorithms in reducing AET to increase the total runoff for a better SHETRAN performance. However, for all these optimized solutions, the surface runoffs are largely underestimated, which may be explained by the reduced soil infiltration resulting from the surface sealing and crust formation (Zhang et al., 2013), physical processes that are not embodied in SHETRAN model and cannot be dealt with by an arbitrarily reduction of the saturated hydraulic conductivity. The Strickler overland flow resistance coefficient is around 10.0 m\(^{1/3}\)s\(^{-1}\) and 5.0 m\(^{1/3}\)s\(^{-1}\), respectively, for crops and agroforestry, which are at their respective large limit and approximate the values of 10.0 and 4.9 m\(^{1/3}\)s\(^{-1}\) obtained by Zhang et al. (2013). By setting a larger Strickler overland flow resistance coefficient, the optimization algorithms try to increase the discharge capacity of river links to compensate for the loss of drainage density resulting from the use of a coarse spatial resolution (Zhang et al., 2015, Zhang 2015). Further studies are required to clarify this point.

**CONCLUSIONS**

This paper successfully applied the NSGA-II optimization algorithm (Deb et al. 2002) and its two variants, namely R-NSGA-II (Deb et al. 2006) and ER-NSGA-II (Siegmund et al. 2012), to the multi-objective calibration of the physically based, spatially distributed hydrological model SHETRAN. SBX and PM were used to generate offspring solutions with respective probabilities of 0.9 and 0.1. The crossover and mutation distribution indices \((\eta_c, \eta_m)\) were configured
as (0.5, 0.5), (2.0, 0.5) and (20., 20.) for all the three algorithms. The nine optimizations were repeated 30 times to consider random seed effects, and all of them terminated due to a limit of 1,500 SHETRAN simulations before the convergence criteria of performance metrics were satisfied. The comparisons among the nine optimizations have shown the distinct effect of using smaller values of ($\eta_c$, $\eta_m$) on getting better efficiency of the three algorithms; however, this effect is clear in the first, but not in the second, half of the optimization processes, due to the self-adaptive property of the SBX operator. On the other hand, the comparisons of the optimization processes, due to the self-adaptive property, might be explained by the lack of diversification of the offspring solutions. The lack of improvement in the optimization of the R-NSGA-II and ER-NSGA-II, compared to the NSGA-II, might be explained by the lack of diversified solutions, which was probably caused by the use of preference distance, rather than crowding distance, as the second measure for population selection.

The validation results of the non-dominated solutions from the nine optimizations are in agreement with those of their calibration. The optimizations with ($\eta_c$, $\eta_m$) of (0.5, 0.5) produce the best solutions and those with ($\eta_c$, $\eta_m$) of (2.0, 0.5) give better solutions than those with (20., 20.). The equifinality problem (Beven & Freer 2001) was identified; the calibration and validation results of the best solutions are satisfactory not only at basin outlet but also at internal gauging stations. In terms of NSE, the best solutions, respectively, for calibration and validation have values of 0.87 and 0.81 at basin outlet, 0.70 and 0.72 at internal gauging station Abernoa and 0.82 and 0.66 at internal gauging station Entradas; as for storm events, the solution has NSE of 0.89 and 0.75, respectively, for storms No. 1 (during the calibration period) and No. 4 (during the validation period) at basin outlet, and 0.74 and 0.66 for storm No. 4 at Albernoa and Entradas, respectively.

Finally, the study has demonstrated the importance of generating diversified offspring solutions for performance improvements of the global optimization methods (Laumanns et al. 2002; Reed et al. 2013) when it comes to automatic calibration of physically based, spatially distributed hydrological models. Although the study has only considered several settings of ($\eta_c$, $\eta_m$) due to the limited computational resources, the conclusions were made based on the mechanism of the optimization processes. Further applications should take into account more parameters of the global optimization methods and the relatively large ranges, as indicated in Reed et al. (2013), for getting better performances.

ACKNOWLEDGEMENTS

This study was made possible by a PhD grant of the Portuguese Foundation for Science and Technology (SFRH/BD/48820/2008). The authors are grateful for the support for the ERLAND (PTDC/AAC-AMB/100520/2008) project. The authors thank SNIRH and SAGRA/COTR for providing the necessary data, the research center CEFAGE (http://www.cefage.uevora.pt/en/) of the University of Évora for offering the required computational resources, and the following researchers for fruitful discussions and for providing access to the SHETRAN model: Prof. Chris Kilby, Prof. James Bathurst, Dr Steve Birkinshaw, Dr Greg O’Donnell and Dr Isabella Bovolo from the University of Newcastle, UK. Also the authors are indebted to Prof. Francisco Santos, Prof. Elsa Sampaio, Prof. Ricardo Serralheiro, Prof. Shokib Shahidian, Dr Julio Lima and Dr Célia Toureiro from the University of Évora and Dr João Pedro Nunes from the University of Aveiro, Portugal. Finally, we very much appreciate the comments and suggestions from the four anonymous reviewers, which considerably improved the manuscript.

REFERENCES

Bathurst, J. C. 2011 Predicting impacts of land use and climate change on erosion and sediment yield in river basins using SHETRAN. In: R. P. C. Morgan & M. A. Nearing (eds),


Cardoso, C. 1965 Solos de Portugal: sua classificação, caracterização e gênesis 1–a sul do rio tejo [The soils of Portugal: its classification, characterization and genesis 1 – the south of the Tejo river]. Lisbon, Portugal.


Im, S., Kim, H., Kim, C. & Jang, C. 2009 Assessing the impacts of land use change on watershed hydrology using MIKE SHE. Environ. Geol. 57 (1), 231–239.


First received 17 November 2012; accepted in revised form 28 August 2015. Available online 23 October 2015