CALIBRATION OF HYDROLOGIC MODELS USING MULTIOBJECTIVE EVOLUTIONARY ALGORITHMS

Francisco Venícius Fernandes Barros, Eduardo Sávio P. R. Martins, Luis Sérgio V. Nascimento & Dirceu Silveira Reis Jr.
FUNCEME, Research Institute for Meteorology and Water Resources

ABSTRACT --- Experience suggests that any single-objective search, no matter how carefully chosen, is not able to identify a solution capable of satisfactorily model a phenomenom of interest. Use of a multiobjective approach can yet be justified by the nature of real world problems, which in general involve multiobjectives, most of the time conflicting objectives. An approach very often used in multi-criteria optimization is the concept of Pareto dominance, which allows us to compare different solutions by using different objectives and to explore different characteristics of the observed data. This paper employs evolutionary algorithms inspired on honey-bee mating for single- (Honey-Bee Mating Optimization - HBMO) and multiobjective (Multiobjective Honey-Bee Mating Optimization - MOHBMO) problems. Both are applied to minimization of test functions and calibration of watershed models. The single-objective version is the one introduced by Haddad et al. (2006), while its multiobjective version is proposed by the present work. As reference of their performance, other well known algorithms were used. The daily hydrological model HYMOD was calibrated for 21 watersheds in Ceará State by using the algorithms aforementioned.
Keywords: calibration, multiobjective, Hydrologic models.

1 - INTRODUCTION

In Engineering, and more specifically in water resources, the need of representation of complexes natural phenomena through models is of crucial importance for water resources planning and management. Through the use of these models, it is possible to understand the natural process and to evaluate the system response to different scenarios, providing support to the decision making process.

Among those models, two classes are of great importance to water resources planning and management: the rainfall-runoff models and the reservoir system operation models. The latter class of models makes use of either systematic records of reservoir's inflows or simulated series obtained by the first class of models, the hydrologic models, which are simple mathematical representations of the natural processes that occur in a watershed.

With respect to rainfall-runoff models, in order to represent satisfactorily the natural process, their parameters should be determined appropriately. In most cases, these parameters cannot be directly determined due to either the impossibility to estimate them in the field or their abstract nature. The indirect determination of the parameters can be made through a calibration study of the model under analysis as long as one has a common period of series that represents both the input and output of this model.

Calibration can be carried out either manually or automatically. The manual method consists of a series of trial-error attempts, in which the parameters are chosen based on the hydrologist experience and knowledge of the region of study. After this choice, the hydrologic model is run and then a comparison is made between the observed and the simulated hydrographs, subjectively, looking after the set of parameters that produces the best result. Solutions of such nature are generally very demanding in terms of work and time, besides requiring the full knowledge of those models, which sometimes are extremely complexes.

The automatic calibration is based upon the use of optimization algorithms that perform a search for the optimal solution with respect to one or more objectives. Several research studies were done in the last decades in this scope, and experience has shown that optimal search based on only one objective, regardless of how carefully it can be made, are not able to determine a solution that models adequately the phenomenom under study. Another factor that favors a multiobjective optimization is that real-world problems frequently require the analysis of conflicting multiobjectives. For such cases, it is possible to use the Pareto Front concept, which makes possible the comparison of solutions with multiple objectives.

Among the algorithms used nowadays, a group in particular has been topic of several research studies due mainly to its widespread use in several areas of science, commerce and engineer, and, also, due to the easiness of their implementation: the evolutionary algorithms. A great advantage of this class of algorithms relative to other approaches is the fact they work with a set of solutions simultaneously, which allows a global perspective of the problem, a greater diversity in the search, and more reliable solutions. Moreover, evolutionary algorithms do not depend on specific characteristics of the objective function to properly work, such as, concavity, convexity and continuity.

2 – LITERATURE REVIEW

The identification of the optimum of uni-modal functions is a problem for which several strategies have been explored in the literature. However, in practice, uni-modal function is barely the case, since most of real-world problems have several local solutions, while only one is the global optimum. This class of problems involves an additional challenge, once one is interested in identifying a set of parameters of a hydrologic model that better represents the behavior of streamflow generation process in a watershed through time (Duan, 2002). Moreover, the involved subjectivity and the required time for manually fit a hydrologic model on a trial-error basis motivated the intense research on automatic calibration of hydrologic models (Vrugt et al., 2003).

However, most studies focused on the uni-objective automatic hydrologic model calibration (Duan *et al.*, 1992, 1993, 1994), while the multiobjective approach has been explored only in the last decade. Within a multiobjective context, a great variety of evolutionary algorithms has been developed based on the concurrent evolution of multiple non-dominate solutions at each iteration of these algorithms (Coello *et al.*, 2004).

Different multiobjective approaches have been employed in the implementation of evolutionary algorithms, such as: objective weighting, lexicographic ordination, use of sub-populations (each one being a uni-objective optimizer), Pareto concept and possible combinations of the previous alternatives. The implementation of multiobjective approaches in evolutionary algorithms has followed this taxonomy, as shown by: 1. Objective weighting (Das and Dennis, 1997; Jin *et al.*, 2001; Baumgartner *et al.*, 2004); 2. Lexicographic ordination (Hu *et al.*, 2003); 3. Use of sub-populations (Schaffer, 1985); 4. Pareto Concept (Knowles and Corne 1999, 2000; Nascimento *et al.*, 2007); and, 5. Possible combinations of the previous alternatives (Xiao-hua *et al.*, 2005).

Within a hydrologic modeling context, two developments can be highlighted: the Multiobjective Complex Evolution (MOCOM-UA; Yapo *et al.*, 1998) and the Multiobjective Shuffled Complex Evolution Metropolis (MOSCEM-UA; Vrugt *et al.*, 2003). The MOCOM-UA resolves the multiobjective calibration problem by the application of the Pareto concept to the global Shuffled Complex Evolution algorithm (SCE-UA; Duan *et al.*, 1992, 1993). MOSCEM was developed in order to better identify the Pareto Front, specially in its extremes, as well as to resolve the MOCOM-UA deficiency regarding the premature convergence in the presence of a large number of parameters and objectives highly correlated (Gupta *et al*., 1998; Yapo *et al*., 1998; Bastidas *et al*., 1999; Boyle *et al*., 2000, 2001; Wagener *et al*., 2001).

This study focus on the application of evolutionary algorithms based upon the mating flights of bees in its uni- (HBMO) and multiobjective (MOHBMO) versions for calibration of hydrologic models. The uni-objective version is that proposed by Haddad *et al.* (2006), while its multiobjective version is proposed by the present study. As a reference, other evolutionary algorithms are used for the uni-objective case, the PSO (Particle Swarm Optimization) and the SCEM (Shuffled Complex Evolution Metropolis), while for the multiobjective case the proposed algorithm will be compared with the multiobjective versions of the previous algorithms, the MOPSO (Multiobjective Particle Swarm Optimization) and the MOSCEM (Multiobjective Shuffled Complex Evolution Metropolis). The PSO and MOPSO algorithms are based on the social behavior of individuals, while the SCEM and the MOSCEM are based upon Markov Chain Monte Carlo methods (MCMC). The PSO and MOPSO algorithms employed here are described in Nascimento *et al.* (2007), while SCEM and MOSCEM are described in Duan *et al.* (1992, 1993) and Vrugt *et al.* (2003), respectively.

3 – METODOLOGY

3.1 – Evolutionary Algorithms

Evolutionary algorithms include search methods that have their insight on natural processes, such as: behavior of social groups, animal reproduction, among others. These algorithms are based on the "survival of the fittest", meaning that the best solutions will prevail over the others. These algorithms have characteristics that make them more robust than other approaches to search optimal solutions, among which can be highlighted:

- The ability to work simultaneously with a population of solutions, which introduces a global perspective and a greater diversity in the search. Such characteristic promotes a great capacity to find a global optimal in problems that have several local optimal solutions;
- Unlike other algorithms based on differential calculus or other specific procedure, evolutionary algorithms work with any objective function and require no specific characteristic, such as continuity, concavity or convexity;
- No previous knowledge of the search space is needed. The search space can be multidimensional, constrained or not, either linear or not.

3.2 – Multiobjective approach using Pareto Dominance Criterion

From a multiobjective context, it is necessary the introduction of a new concept to replace the simple comparison between uni-objective different solutions, i.e., the Pareto dominance concept. This multiobjective approach is described below and the consequent problems due to the use of the Pareto dominance concept are dealt afterwards.

Let a multiobjective minimization problem given by:

$$\min f(\underline{x}) = \left[f_1(\underline{x}) \quad f_2(\underline{x}) \quad \ldots \quad f_M(\underline{x}) \right] \tag{1}$$

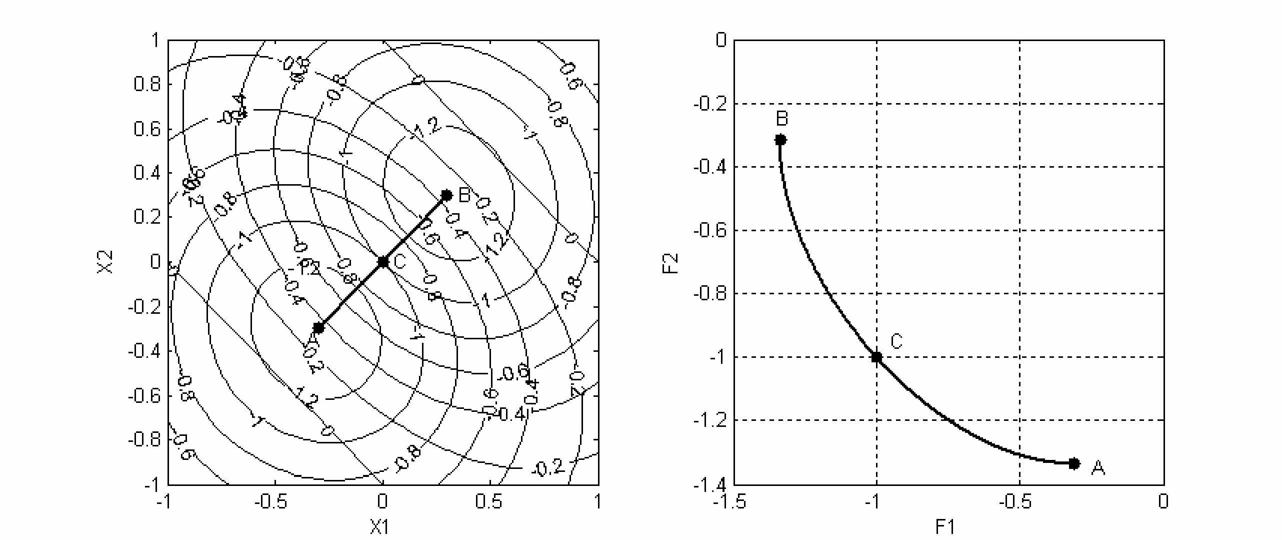
where $f_i(\underline{x})$ is i-th of M objective functions and $\underline{x}$ is a feasible solution.

Looking at equation (1), two distinct solutions (u and v) can be related in the following manner (Alvarez *et al*., 2005):

- If $f_i(u) \le f_i(v) \mid \forall i = 1,\ldots,M$ and $f_i(u) < f_i(v)$ for some i, then v is strictly dominated by u, represented by $u \prec v$;
- Or, if $f_i(u) \le f_i(v) \mid \forall i$, v is said to be weakly dominated by u, represented by $u \preceq v$.

If u is not dominated by v, and v is not dominated by u, u and v are said to be non-dominated solutions. It is clear, then, that multiobjective problems have more than one optimal solution, and this set of solutions is called Optimal Pareto Front or True Pareto Front composed of non-dominated solutions by any other solutions.

Figure 1(a) shows in the parametric space (θ) the location of the minimum of two objective functions represented by A and B, the line that connects these minimum and that are part of the Pareto Front, and the γ point, which represents a possible solution of the optimal front. Figure 1(b) shows the same elements in the objective function space. It should be noticed that the curve that connects A to B is tangent to the function contour lines in the parameter space.



(a) Parameter Space (b) Objective Function Space

Figure 1 – Illustration of the Pareto optimal solution concept for a minimization problem with two objectives (F1, F2) in a search space (Δδ) bi-parametric (θ1, θ2): a. parametric space; b. functions space (Vrugt et al., 2003).

3.3 – Honey bee mating-based Optimization (HBMO and MOHBMO)

The uni-objective version of the algorithm (HBMO) is the one proposed by Haddad *et al.* (2006), in which the honey bee mating has served as inspiration. The relationship between the natural process and the algorithm is established as the algorithm is described mathematically in the sequence.

3.3.1 *HBMO Algorithm (Honey-bees Mating Optimization)*

The algorithm starts with an initial population (hive), composed by a set of solutions randomly sampled from a Uniform Distribution. A fitness value is assigned to each solution of the initial population, equals to the selected objective function. The best solution (queen) is then selected based on the fitness value, the smallest fitness value since it is a minimization problem. All other solutions are discarded and a new iteration is initiated. If the problem at hand is a maximization problem, then the following equation should be used:

$$\max f(\underline{x}) = \min(-f(\underline{x})) \qquad (2)$$

At the beginning of a new iteration, random solutions (drones, D) with certain degree of dependence with the best solution (queen, Q) are generated. Such dependence is a linear or a quadratic function of the number of iterations (see equations 3 and 4), increasing with the number of iterations (maturity of the hive). At the last iteration there is a large dependence among the drones and the queen, which promotes the convergence of the search. In order to guarantee the diversity of the solutions, a minimum randomness factor for the generation of the drone is introduced as an algorithm parameter.

$$D = Q \times \left[(i-1)/nMF\right] + d \times \left[(nMF-(i-1))/nMF\right] \tag{3}$$
$$D = Q \times \left[1-(\delta^2/nMF^2)\right] + d \times \left[\delta^2/nMF^2\right] \tag{4}$$

where nMF is the number of iteration, i is the actual iteration, d is a random solution in the search space. The set of random solutions {d} is centered at the best solution (queen), from which each random solution d is generated. The parameter δ is given by the following expression:

$$\delta = nMF - (i-1) \tag{5}$$

At each iteration, a selective test (mating flight) is performed in order to determine probabilistically whether or not the best solution (queen) will receive information (mating) from the randomly selected solutions (drones). This is done by applying an annealing function, also known as Boltzman function, as suggested by Abbas (2001):

$$p(Q,D) = \exp\left(-\Delta(f)/Sp(t)\right) \tag{6}$$

where $p(Q,D)$ is the probability that the best solution Q receives information from the selected random solution D (crossover between the drone D and the queen Q), $\Delta(f)$ is the absolute difference of the fitness values of solutions Q and D, and $Sp(t)$ is the temperature (speed of the queen) of the annealing function at time t (during flight). Looking at the annealing function, it is evident that the probability is larger when either the temperature (speed of the queen) is high or the differences in fitness are small (the fitness of the drones are close to the fitness of the queen).

The algorithm allows information exchange (mating) between the current solutions (drones and queen) with probability $p(Q,D)$. In case of information exchange (mating), the information of the solution (genetic information of the drone) is selected and stored in a repository (queen's spermatheca), and the temperature of annealing (speed of the queen) decreases as follows:

$$Sp(t+1) = \alpha(t) \times Sp(t) \tag{7}$$

and

$$\alpha(t) = \left[M - m(t)\right]/M \tag{8}$$

where $Sp(t)$ is the temperature (speed of the queen) at time t, $\alpha(t)$ is a value between 0 and 1, M is the size of the repository (queen's spermatheca) and $m(t)$ is the number of randomly selected solutions (drones) for the crossover. The number of information exchange attempts (queen's energy) decreases as follows:

$$E(t+1) = E(t) - \gamma \tag{9}$$

where $E(t)$ is the number of attempts at time t and γ its decay at each time interval. For this study, γ is equals to one.

The best solution (queen) can receive information (mating) as long as both the number of attempts (its energy) is not close to zero and its repository (spermatheca) is not full. From equation (9), it is noticed that the decay value γ will determine how many tests (transitions in the search space, for which exists a probability of the queen meeting a drone) the best solution (queen) can perform at each selection of random solutions (mating flight). Other limiting factors are the temperature of the annealing function (speed of the queen), which should be greater than zero, and the number of random solutions (drones) available for testing, since each random solution only provides information (mating) once and then is discarded (death of the drone).

The generation of new solutions (offspring) occurs after the information exchange by crossover between the information stored in the repository (genes of drones in the spermatheca) and the information of the best solution (genes of the queen). The choice of which information (genetic material in the spermatheca) will be used is randomly determined and can be reused. This generation process is carried out using several crossover operators according to their performance, here evaluated by the percentage of their contribution to the generation of new solutions. It was used two crossover operators as described in section 3.3.3: the Arithmetic Crossover the Blend Crossover.

Once new solutions (new offspring) are generated, an attempt is made to improve both the new solutions and the best solution by use of a mutation procedure, known as Creep mutation operator (See section 3.3.3). The mutation is randomly applied to a pre-specified percentage of the new solutions (new offspring). Also, there is a probability that mutation is applied to the best solution (queen), set herein to 5%.

After mutation, the population is then evaluated based on the objective function. If the best generated solution is better than the best current solution (queen), the best solution is updated, otherwise, the best solution continues the same. At each iteration, all generated solutions are discarded and only the best solution is kept.

The process described previously is repeated until a stop criterion is met, such as the maximum number of iterations. As an attempt to perform a more detailed search, Haddad *et al.* (2006) suggests the use of several queens, selected based on their fitness values. In such case, the process described earlier is applied for each queen, mixing all offspring afterwards. The best solutions are then selected from the sets formed by both queens and their offspring. This approach, used in this study, results in a refinement over the original approach based on only one queen.

3.3.2 *Multiobjective Honey-bees Mating Optimization*

In order to deal with multiobjective problems, some modifications in the HBMO algorithm were made. After the generation of the initial population (hive) and respective evaluation of objective functions, the selection of the "best solutions" (queens) must be made, but no longer based only on the comparison of single objective function values. Under a multiobjective approach, a new concept, such as the Pareto dominance concept, is needed for dealing with different solutions, classifying them as dominated or non-dominated solutions. The "best solutions" (queens) selected from the initial population are the non-dominated solutions.

Once identified the non-dominated solutions (queens), the iterative process is initiated in the same way as in the uni-objective case (mating flights, generation of new queens, improvement of the queens and of the new generation and selection of new queens). Each non-dominated solution (queen) will generate a certain number of solutions (offspring) after each iteration. The criteria for generation and improvement of the solutions (offspring) and of the best solution (queen) are the same employed in the uni-objective version.

With the new generated solutions (new offspring) and the non-dominated solutions from the previous iteration, the new set of non-dominated solutions is identified, which forms the Pareto Front. These new solutions will generate the new solutions in the next iteration. The process is repeated until the stop criterion is satisfied.

Frequently, the number of solutions that belong to the Pareto Front increases as the algorithm evolves, thus each non-dominated solution is a potential generator (queen) of new solutions in the next iteration of the algorithm. This would make the algorithm slower and more inefficient as the number of iteration increases, since each solution would generate a number of new solutions (new offspring and drones), escaping the user control.

In order to increase the algorithm efficiency, it was used a clustering method (Seber, 1984; Spath, 1985) to select the non-dominated solutions (queens) among the front solutions. The clustering technique used here not only promoted a better distribution of the solutions along the front but also improved the performance of the algorithm.

3.3.3 *Crossover and Mutation Operators*

The crossover operators employed here are the Blend Crossover and the Arithmetic Crossover operators. The Blend Crossover performs a linear combination between two solutions as indicated by the following expression (Lacerda and Carvalho, 1999):

$$c = p_1 + \beta \times (p_2 - p_1) \tag{10}$$

where c is the generated offspring, p_1 and p_2 are the parent solutions, and β represents the feasible space for offspring generation with $\beta \sim U(-\varepsilon, 1+\varepsilon)$, where a previously chosen ε allows the generation to occur beyond the interval defined by p_1 and p_2.

The Arithmetic Crossover does a linear combination between two solutions according to the following expressions (Lacerda and Carvalho, 1999):

$$\begin{aligned} c_1 &= \beta \times p_1 + (1-\beta) \times p_2 \\ c_2 &= (1-\beta) \times p_1 + \beta \times p_2 \end{aligned} \tag{11}$$

where $\beta \sim U(0,1)$.

The mutation operator corresponds to the Creep Mutation operator, which performs a small perturbation in one decision variable. This perturbation is carried out according to the following expression:

$$c_n^i = \beta \times c_n^i \tag{12}$$

where c_n^i represents the decision variable i of the solution n and $\beta \sim U(0.95, 1.05)$ is the mutation factor

4 – HYDROLOGIC MODEL

This paper employs the daily rainfall-runoff model HYMOD (Moore, 1985). The used version is a relatively simple model, which uses the probability distribution concept to describe the spatial variation of runoff production process parameters. This allows the integration of the flow response over the whole watershed represented by algebraic expressions.

The idea underneath the model is that the watershed can be viewed as a set of points without interaction among them, while each one has a water storage capacity that, when exceeded, generates runoff. Figure 2 illustrates this representation.

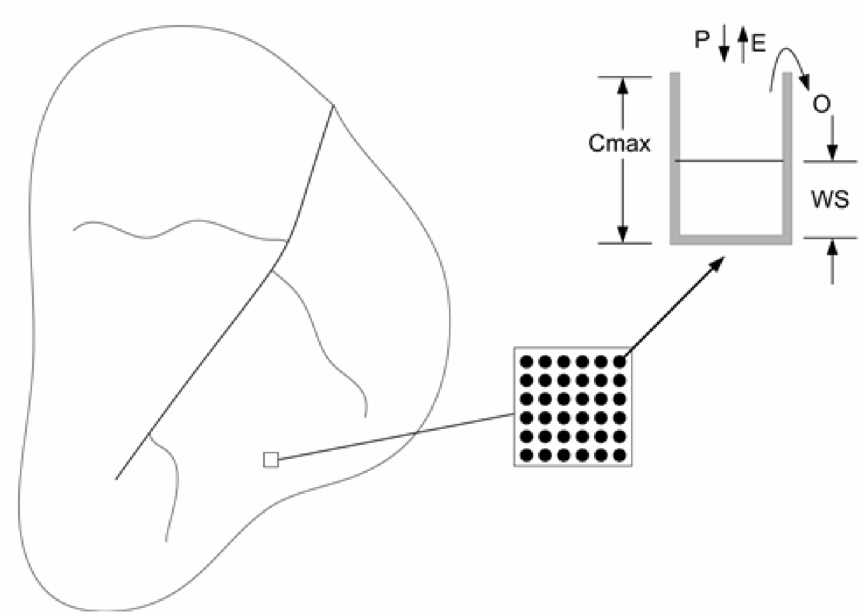


Figure 2 – Watershed representation in the HYMOD model (P: Precipitation; E: Evaporation; O: Outflow; WS: Water Storage; $C_{\max}$ is the largest water storage capacity within the watershed).

The distribution function of the different water storage capacities is defined as:

$$F(C) = 1 - \left(1 - C / C_{\max}\right)^B \tag{13}$$

where F represents the cumulative probability of a certain water storage capacity (C) if a random point is selected; $C_{\max}$ is the largest water storage capacity within the watershed and B is the degree of variability in the storage capacity. Figure 3 shows the schematic representation of the HYMOD model. After a rainfall event, the water can infiltrate up to the soil reaches its water storage capacity, after what runoff will be generated.

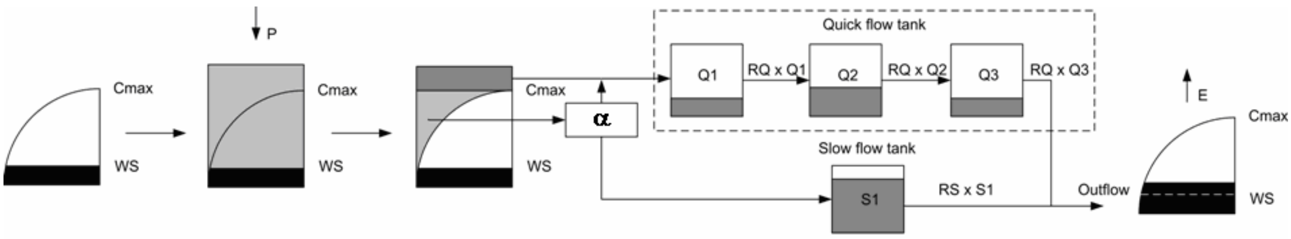


Figure 3 – Schematic representation of the hydrologic model HYMOD.

The fraction that exceeds $C_{\max}$ does not infiltrate and passes through three linear quick flow tanks at a constant flow rate RQ. For those points where the water storage is less than $C_{\max}$, the remaining precipitation which exceeds the water storage is directed to either quick flow tanks or the slow flow tank depending on a constant α. The total outflow of the watershed is obtained by summing the outputs of the quick flow tanks and the slow flow tank.

Finally, the evaporation is taken from the water storage in the watershed. If the available water in storage is greater than the potential evaporation, the real evaporation is equal to the potential evaporation, otherwise all available water evaporates.

This model has five parameters: 1. Largest storage capacity within the watershed ($C_{\max}$); 2. The degree of spatial variability in the water storage capacities (*B*); 3. Factor that divides the amount that exceeds the water storage capacity of points with a capacity lower than $C_{\max}$ between the quick flow tanks and the slow flow tank (α); 4. Residence time for the quick flow tanks (RQ); 5. Residence time for the slow flow tank (RS).

5 – RESULTS AND DISCUSSION

5.1 – Test Functions

The reference algorithms (PSO, MOPSO, SCEM, MOSCEM), the single-objective algorithm HBMO and its multiobjective version proposed here MOHBMO were tested with mathematical functions which represent a challenge for any optimization algorithm, named here simply test functions (Deb, 1999).

Five test functions were used to evaluate the optimization algorithms, noted here as $f_1(x_1,x_2)$, $f_2(x_1,x_2)$, ..., $f_5(x_1,x_2)$.The used functions are described in Table 1, but some of them are used only for the multiobjective case (for example, Function 1). Five combinations of functions in Table 1 were considered for the composition of the multiobjective problems (MO): (a) MO_1: f_1 and f_2; (b) MO_2: f_1 and f_3; (c) MO_3: f_1 and f_4; (d) MO_4: f_1 and f_5; and (e) MO_5: f_3 and f_5.

Table 1 – Test functions used in the uni- and multiobjective optimization.

Function	Range	Observation
$f_1(x_1,x_2) = x_1$	$0 \le x_1, x_2 \le 1$	
$f_2(x_1,x_2) = -20\dfrac{\sin(0.1+\sqrt{(x_1-4)^2+(x_2-4)^2})}{0.1+\sqrt{(x_1-4)^2+(x_2-4)^2}}$	$-10 \le x_1, x_2 \le 20$	Difficulty: The function has several minimum and maximum, with global minimum at $x_1 = 4$ and $x_2 = 4$ equals to $f = -19.6683$.
$f_3(x_1,x_2) = \dfrac{2-\exp\left[-\left(\dfrac{x_2-0.2}{0.004}\right)^2\right]-0.8\exp\left[-\left(\dfrac{x_2-0.6}{0.4}\right)^2\right]}{x_1}$	$0.1 \le x_1 \le 1$ $0 \le x_2 \le 1$	The following expression represents a bimodal function, with a local minimum at $x_1=1$ and $x_2=0.6$ equals to $f=1.2$, and a global minimum at $x_1=1$ and $x_2=0.2$ equals to $f=0.7057$: Difficulty: Singularity at the global minimum.
$f_4(x_1,g) = 1-\left(\dfrac{x_1}{g}\right)^2$ $g(x_{m+1},\ldots,x_N) = g_{\min} + (g_{\max}-g_{\min})\left(\dfrac{\sum_{i=m+1}^{N} x_i - \sum_{i=m+1}^{N} x_i^{\min}}{\sum_{i=m+1}^{N} x_i^{\max} - \sum_{i=m+1}^{N} x_i^{\min}}\right)^{\gamma}$	$0 \le x_1, x_2 \le 1$ m=1; N=2; $g_{min}=1$; $g_{max}=2$; γ=0.25	x_i^{min} and x_i^{max} are, respectively, the minimum and maximum values of the variable x_i, g_{min} and g_{max} are, respectively, the minimum and maximum values of the function g(.). Difficulty: γ is the parameter responsible for the introduction of bias in the function values.
$f_5(x_1,x_2) = (1+10x_2)\times\left[1-\left(\dfrac{x_1}{1+10x_2}\right)^{\alpha} - \dfrac{x_1}{1+10x_2}\sin(2\pi q x_1)\right]$	$0 \le x_1, x_2 \le 1$ q=4; α=2	

A sensitivity analysis of the parameters of the aforementioned algorithms was performed in order to determine their values to be later used in the calibration of the hydrologic model. This sensitivity analysis is not presented here, but can be found in Barros (2007).

5.1.1 Uni-objective Minimization

In order to compare the performance between HBMO and PSO for minimization, both algorithms were used to minimize several test functions. The goals were to access both the ability of each algorithm to identify the optimum and their speed of convergence in terms of number of evaluation needed to reach the optimum. The initial condition was kept the same for both algorithms to guarantee a fair comparison.

For the uni-objective minimization of test functions, the following parameters of the HBMO algorithm were used: a. Initial population size = 100; b. Number of mating flights = 50; c. Number of queens = 20; d. Number of drones = 1; e. Minimum randomness factor of the drone = 10%; and, f. Number of offspring per queen = 4. For the PSO algorithm, the following parameters were used: a. Size of population = 100; b. Number of iterations = 50; c. Maximum speed of the particle = 1 (function 4) and 0.1 (functions 2, 3 and 5); and, d. c1 = c2 = 1 and w varying from 0.95 to 0.4 until the algorithm reaches 70% of the maximum number of iterations, keeping the smallest value in the following iterations.

Once defined the algorithms' parameters, five applications were then made starting from the same initial conditions and parameter values for both algorithms as mentioned before. For each function, convergence was illustrated using graphics that show the value of the objective function at each iteration:

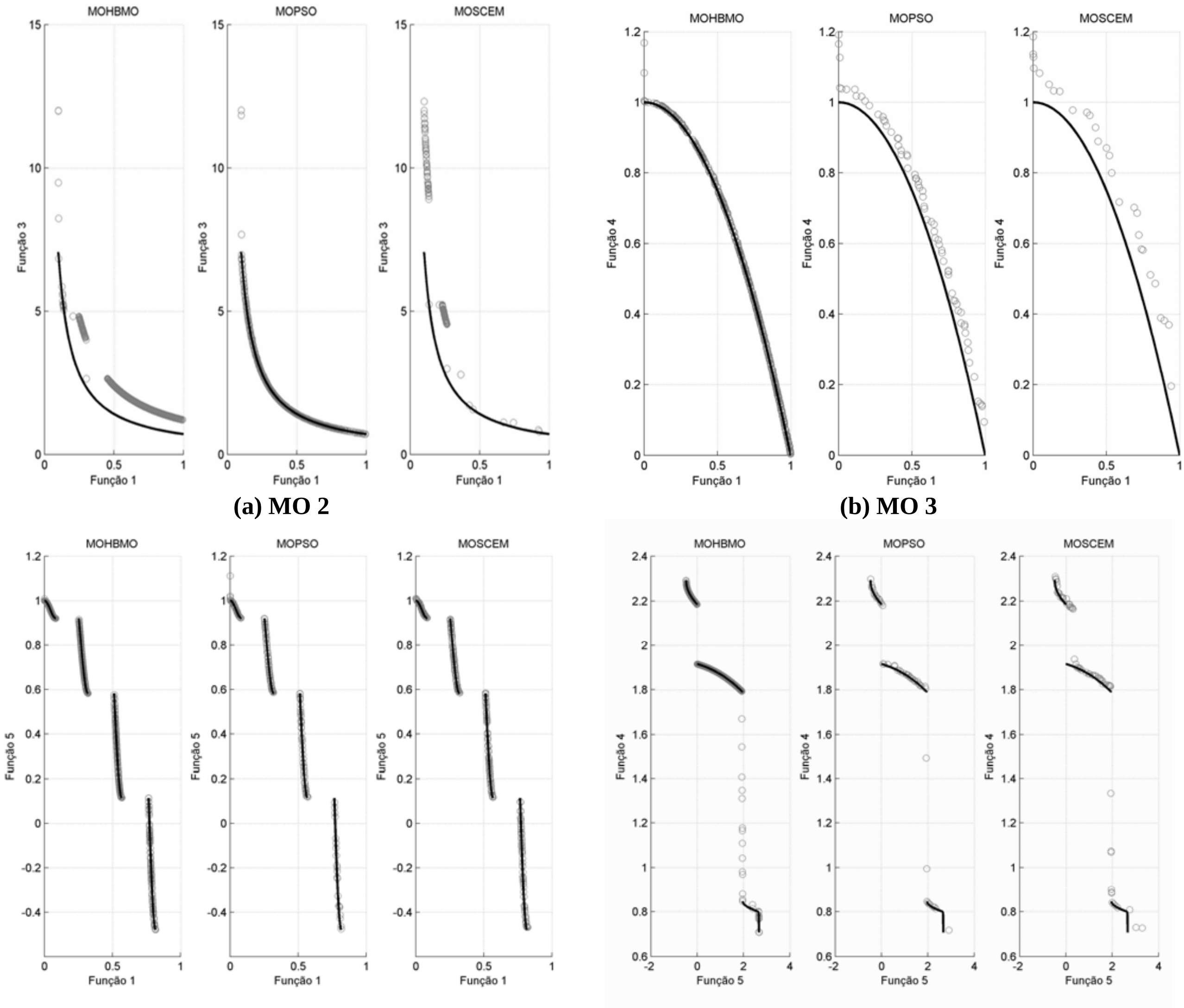
- Function 2: Both algorithms have similar convergence with function values almost equal at the end of 700 iterations;
- Function 3: PSO converged slower than HBMO. While HBMO rapidly reached the global optimum by 1500th evaluation, PSO did not identify the global minimum;
- Function 4: For this problem, it was noted that PSO was more robust than HBMO, since PSO identified the global optimum for all cases but one, while HBMO identified the global optimum for two cases only;
- Function 5: Although HBMO converged faster than PSO, both algorithms identified easily the global optimum.

5.1.2 *Multiobjective Optimization*

The comparative evaluation of the algorithms in multiobjective problems employs multiobjective versions of the algorithms tested in the last section: the algorithms MOHBMO, MOPSO and MOSCEM, the multiobjective versions of HBMO, PSO and SCEM, respectively. The non utilization of the algorithm SCEM is justified by the fact this algorithm does not work with the concept of objective function.

For the minimization of the multiobjective problems, the following parameters for MOHBMO algorithm were used: a. Initial population size = 100; b. Number of mating flights = 100; c. Number of queens = 20; d. Number of drones = 1; e. Minimum randomness factor of the drone = 1%; and, f. Number of offspring per queen = 4. The other parameters were used as described previously in section 5.1.1. For MOPSO algorithm, it was employed the following parameters a. Size of population = 100; b. Number of iterations = 100; c. Maximum speed of the particle = 0.5. The other MOPSO parameters were those recommended by Nascimento *et al.* (2007). For MOSCEM, it was used the following parameters: a. Size of population = 100; b. Number of complexes = 2; c.

Number of Objective Function evaluations = 10,000. Figure 4 presents the true Pareto Fronts and those identified by the three algorithms MOHBMO, MOPSO and MOSCEM for the multiobjective problems 2–5.



(a) MO 2 (b) MO 3

(c) MO 4 (d) MO 5

Figure 4 – True Pareto Fronts and those identified by the algorithms MOHBMO, MOPSO and MOSCEM for the multiobjective problems 2–5.

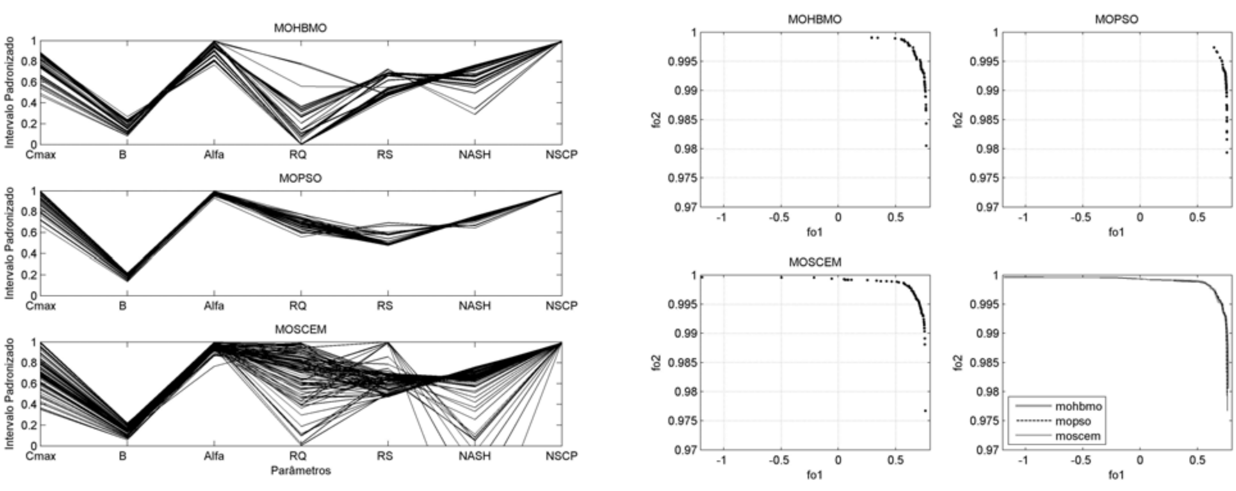
MOHBMO and MOSCEM algorithms experienced problems in identifying the Pareto Front in the presence of singularities (MO_2), while MOPSO was able to fill in the front adequately. With respect to the presence of bias in the objective function (MO_3), MOHBMO algorithm had the best performance and was the only algorithm to properly fill the Pareto Front. For the Multiobjective problem 4 (MO_4), all three algorithms had similar performance, while for the Multiobjective problem 5 (MO_5) MOHBMO was able to better identify and adequately fill the Pareto Front relative to the other two algorithms.

5.2 – Hydrologic Models Calibration

This section evaluates the performance of the HBMO and MOHBMO algorithms relative to other relatively well known evolutionary algorithms (PSO, MOPSO, MOSCEM). The MOHBMO, a multiobjective version of the HBMO algorithm, was proposed in this paper for the calibration the hydrologic model HYMOD. The difficulty in calibrating daily hydrologic models for semi-arid regions is due mainly to the poor quality of the available hydrologic data as well as the inexistence of information regarding small reservoirs, which can affect tremendously in the streamflow generation process, in particular for low-flow periods. Another difficulty is the predominant convective nature of the precipitation regime over the rainy season, which makes even more necessary the existence of a dense monitoring network along with high quality information. This is very difficult to guarantee on a daily basis. For the calibration study, the longest record was used, leaving out part of the record for verification. The use of a long record for calibration makes difficult the identification of reliable parameters (Yapo *et al.*, 1996), but the goal of the study was the evaluation of the performance of the optimization algorithms. The Nash-Suttcliffe coefficient was employed as fitting criteria applied to daily streamflow series (of_1), characteristic points of the flow-duration curve (of_2), peak flows (of_3) and monthly volume series (of_4):

$$of = \max_{\theta}\left\{1-\left(\sum_{i=1}^{N}\left(Q_i-\hat{Q}_i(\theta)\right)^2\right)\Big/\left(\sum_{i=1}^{N}\left(Q_i-\overline{Q}\right)^2\right)\right\} \tag{14}$$

Figure 5 presents the result for the calibration of HYMOD applied to the stream gage station 34750000, selected from those stations that achieved better results among the 21 stations employed. The calibration was carried out for the 32 years of the station data, but only a period is shown here. This Figure shows the sets of optimal parameters (Figure 5a) and the Pareto Fronts (Figure 5b) identified by the multiobjective algorithms using objective functions 1 and 2. Generally, MOHBMO and MOSCEM were able to identify the Pareto Front with a good density of points and coverage of its limits. The MOPSO experienced difficulty, as already noticed with the test functions, to identify points close to Pareto Front limits with adequate density. For some stations, the Pareto Front identified by the MOPSO was completely or partially dominated by the other Pareto Fronts identified by MOHMO and MOSCEM.



(a) (b)

Figure 5 – Optimum solutions identified by the algorithms MOHBMO, MOSCEM and MOPSO using OFs 1 and 2 in the calibration of the model HYMOD for the stream gage station 34750000: (a) Set of optimal parameters; (b) Identified Pareto Fronts.

Figure 6 presents the observed hydrograph and those associated with the optimal solution set of the Pareto Front identified by the MOHBMO algorithm, but in this case, differently from Figure 5, the objectives are of_1 and of_3. The bold black line represents a "trade-off" solution between the two objectives. Should be noticed that the algorithms MOHBMO and MOSCEM did not experienced any difficulty in determining the Pareto Front for the case of high correlated objectives and fractioned fronts. For these two algorithms the Pareto Front filling was appropriate and extended to its range limits.

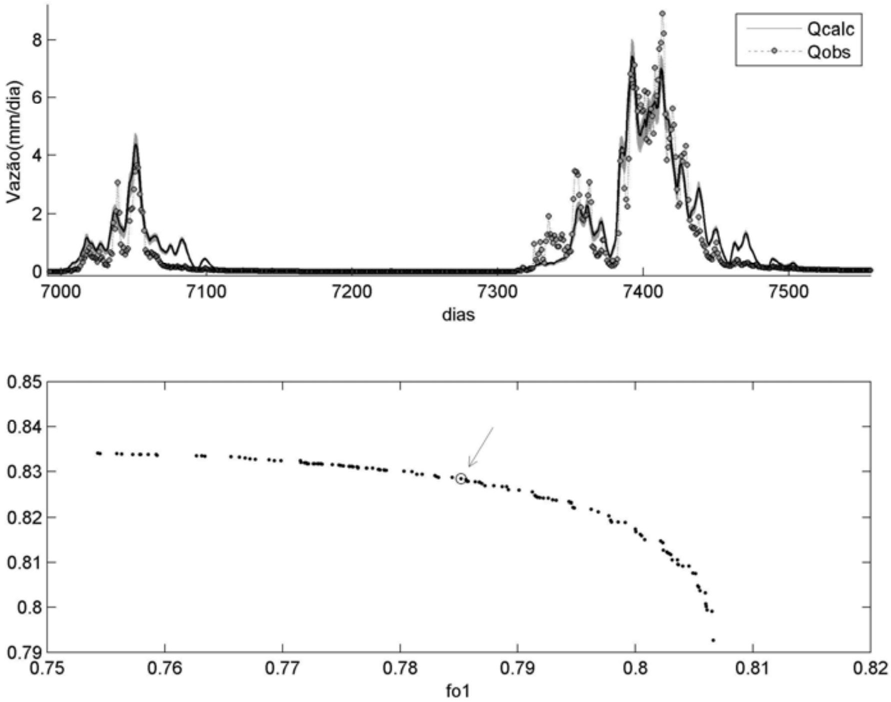


Figure 6 – Observed hydrograph and optimal hydrographs associated to Pareto Front points for the objective functions OF_1 and OF_3. Dotted line represents the observed hydrograph while the continuous line represents the solution correspondent to the point of the Pareto Front indicated by the arrow (Hydrographs associated to other points in the Pareto Front are represented by grey lines). The identified Pareto Front is presented below the hydrographs.

5 – CONCLUSION

The optimization algorithms used in this paper were tested with mathematical functions for which the optima are known (optimum solution for the single-objective case and the Pareto Fronts for the multiobjective case), and afterwards used in the calibration of the hydrologic model HYMOD. For the test functions, the algorithms PSO and MOPSO had better performance when compared to the other algorithms in the presence of singularities of the objective function. In general, HBMO and MOHBMO were superior to PSO and MOPSO/MOSCEM respectively, in terms of appropriate filling of the Pareto Front and identification of the elements of the Pareto Front in their limits. The PSO and MOPSO performed worse when bias was introduced in the objective function.

Several problems identified in section 5.2 difficult the calibration process. For posterior analysis, special attention should be given to the definition of the calibration period. The calibration period is not necessarily the longest, but that with better quality information regarding the hydrological processes captured by the HYMOD model.

The efficiency of the proposed algorithm MOHBMO in identifying the Pareto Front during the calibration of the HYMOD model was compared to those fronts identified by MOPSO and MOSCEM. In general, MOHBMO and MOSCEM presented a better performance relative to MOPSO, but MOSCEM, besides the higher computing time when compared to the other two algorithms, offers additional information regarding the parametric uncertainty of the model. The choice of the objective function within a multiobjective optimization framework should be carefully made, since this choice can dramatically affect the dynamics of the hydrologic simulation output. It is necessary take more advantage from the multiobjective optimization by exploring different characteristics of the observed hydrographs to be preserved.

REFERENCES

Abbass, H.A. (2001). "*A Pleometrosis MBO Approach to Satisfiability*". Proceeding of International Conference on Computational Intelligence for Modeling, Control and Automation, CIMCA2001, Las Vegas, USA.

BARROS, F.V.F.. (2007). "*Uso de Algoritmos Evolucionários na Calibração de Modelos Hidrológicos e na Operação de Sistemas de Reservatórios*". Dissertação de Mestrado apresentada à Universidade Federal do Ceará, como parte dos requisitos necessários para obtenção do título de Mestre em Engenharia Civil na área de concentração em Recursos Hídricos e Saneamento. Fortaleza, Ceará.

BASTIDAS, L.A.; GUPTA, V.K.; S. SOROOSHIAN; W.J. SHUTTLEWORTH; Z.L. YANG. (1999). *"Sensitvity Analysis of a Land Surface Scheme using Multi-Criteria Methods"*. Journal of Geophysical Research, 104(D16), 19, pp. 481-490.

BAUMGARTNER, U.; C. MAGELE; W. RENHART. (2004). *"Pareto Optimality and Particle Swarm Optimization"*. IEEE Transactions on Magnetics, 40(2), pp.1172–1175.

BOYLE, D.P.; H.V. GUPTA; S. SOROOSHIAN. (2001). *"Toward Improved Calibration Models: Combining the Strengths of Manual and Automatic Methods"*. Water Resources Research, 36(12), pp.3663-3674.

BOYLE, D.P.; H.V. GUPTA; S. SOROOSHIAN. (2001). *"Toward Improved Streamflow Forecsts: Value of Semi-distributed Modeling"*. Water Resources Research, 37(11), pp. 2749-2759.

COELLO, A.C.C; G.T. PULIDO; M.S. LECHUGA. (2004). *"Handling Multiple Objectives With Particle Swarm Optimization"*. IEEE Transactions on Evolutionary Computation, 8(3), pp. 256–279.

DAS, I.; J. DENNIS. (1997). "*A Closer Look at Drawbacks of Minimizing Weighted Sums of Objectives for Pareto Set Generation in Multicriteria Optimization Problems*". Structural Optimization, 14(1), pp. 63-69.

DUAN, Q. (2002). *"Global Optimization for Watershed Model Calibration". in Calibration of Watershed Models*. AGU, Washington DC, pp. 89-104.

DUAN, Q; S. SOROOSHIAN; V.K. GUPTA. (1994). *"Optimal use of the SCE-UA Global Optimization Method for Calibrating Watershed Models"*. Journal of Hydrology, 158, pp. 265-284.

DUAN, Q; V.K. GUPTA; S. SOROOSHIAN. (1993). *"A shuffled complex evolution approach for effective and efficient global minimization"*. J. Optim. Theory Appl., 76(3), pp. 501-521.

DUAN, Q; V.K. GUPTA; S. SOROOSHIAN. (1992). *"Effective and efficient global optimization for conceptual rainfall-runoff models"*. Water Resources Research, 28(4), pp. 1015-1031.

GUPTA, V.K.; S. SOROOSHIAN; P.O. YAPO. (1998). *"Toward Improved Calibration of Hydrologic Models: Multiple and Noncommensurable Measures of Information"*. Water Resources Research, 34(4), pp. 751-763.

HADDAD, O.B.; A. AFSHAR; M.A. MARINO. (2006). "*Honey Bees Mating Optimization Algorithm (HBMO); Honey-Bees Mating Optimization (HBMO) Algorithm: A new heuristic approach for water resources optimization*." Water Resources Management, Vol. 20, Nro. 5 (22 October 2006), pp. 661-680.

HU, X.; R.C. EBERHART; Y. SHI. (2003). "*Particle swarm with extended memory for multiobjective optimization*". In Proceedings of the 2003 IEEE Swarm Intelligence Symposium, Indianapolis, Indiana, USA, IEEE Service Center, pp. 193–197.

JIN, Y.; T. OKABE; B. SENDHOFF. (2001). *"Dynamic weighted aggregation for evolutionary multiobjective optimization: Why does it work and how?*". In Lee Spector, Erik D. Goodman, Annie Wu, W.B. Langdon, Hans-Michael Voigt, Mitsuo Gen, Sandip Sen, Marco Dorigo, Shahram Pezeshk, Max H. Garzon, and Edmund Burke, editors, Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'2001), San Francisco, California, Morgan Kaufmann Publishers, pp. 1042–1049.

KNOWLES, J.D; D. W. CORNE. (1999). "*The Pareto Archived Evolution Strategy: A New Baseline Algorithm for Multiobjective Optimization*". In 1999 Congress on Evolutionary Computation, Washington D.C., IEEE Service Center, pp. 98-105.

KNOWLES, J.D; D. W. CORNE. (2000). *"Approximating the Nondominated Front Using the Pareto Archived Evolution Strategy"*. Evolutionary Computation, 8(2), pp. 149-172.

LACERDA, E. G. M.; A. C. P. L. F. CARVALHO. (1999). *"Introdução aos Algoritmos Genéticos"*. In C. O. Galvão e M. J. S. Valença, Orgs., Sistemas Inteligentes: Aplicações a Recursos Hídricos e Ciências Ambientais. Porto Alegre, Ed. Universidade/UFRGS/ABRH, 246 p.

MOORE, R.J. (1985). "*The probability-distributed principle and runoff production at point and basin scale*". Hydrological Sciences Journal 30(2), pp. 273-297.

NASCIMENTO, L.S.V.; D.S. REIS JR; E.S. MARTINS. (2007), "*Avaliação do Algoritmo Evoluvionário MOPSO na Calibração Multiobjetivo do Modelo SMAP no Estado do Ceará*", In: Anais do XVII Simpósio Brasileiro de Recursos Hídricos, 25 – 29 de novembro, São Paulo, SP.

SCHAFFER, J.D. (1985). *"Multiple Objective Optimization with Vector Evaluated Genetic Algorithms"*. In Genetic Algorithms and their Applications: Proceedings of the First International Conference on Genetic Algorithms, Lawrence Erlbaum, pp. 93-100.

SEBER, G.A.F., Multivariate Observations, Wiley, New York, 1984.

SPATH, H. (1985) Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples, translated by J. Goldschmidt, Halsted Press, New York, 226 pp.

VRUGT, J.A.; H.V. GUPTA; L.A. BASTIDAS; W. BOUTEN; S. SOROOSHIAN. (2003). *"Effective and Efficient Algorithm for Multiple Objective Optimization of Hydrological Model"*. Water Resources Research, 39(8), doi:10.1029/2002WR001746.

XIAO-HUA, Z.; M. HONG-YUN; J. LI-CHENG. (2005). *"Intelligent particle swarm optimization in multiobjective optimization*". In Congress on Evolutionary Computation (CEC'2005), Edinburgh, Scotland, UK, IEEE Press, pp. 714–719.

WAGENER, T.; D.P. BOYLE; M.J. LEES; H.S. WHEATER; H.V. GUPTA; S. SOROOSHIAN. (2001). *"A Framework for Development and Application of Hydrological Models"*. Hydrol. Earth Sys. Sci., 5(1), pp. 13-26.

YAPO, P.O; H.V. GUPTA; S. SOROOSHIAN. (1998). *"Multiobjective Global Optimization for Hydrologic Models"*. Journal of Hydrology, 204, pp. 83-97.

YAPO, P.O; H.V. GUPTA; S. SOROOSHIAN. (1996). "Automatic Calibration of Conceptual Rainfall-Runoff Models: Sensitivity to Calibration Data". Journal of Hydrology, 181, pp. 23-48.