# Performance Comparison of Three Multi-objective Optimization Algorithms on Calibration of Hydrological Model

Yuhui Wang[1], Xiaohui Lei[2], Yunzhong Jiang[1,2], Hao Wang[1,2]
[1]School of Environmental Science and Engineering, Donghua University, Shanghai, P.R. China
[2]China Institute of Water Resources and Hydropower Research, Beijing, P.R. China

***Abstract*—in this essay, the comparison of three multi-objective optimization approaches MOPSO, NSGA-II and MOSCEM-UA has been carried out in the auto calibration of hydrological model- Hymod. By carrying out the calibration on two objectives of high flow and low flow objective functions, the Pareto front can be drawn. The performance of the three optimization algorithms is analyzed depending on three criterions that are the optimization time cost, Pareto front spacing rate and the dominating rate. Through analyzing the comparison results of MOSCEM-UA and NSGA-II with MOPSO method, the performance of convergence rate, Pareto non-dominant spacing rate and iteration speed are ideally expected. The simulation result with MOPSO algorithm is reasonable in the high flow and low flow process. The prediction area drawn from the optimization result indicates the reliability of the model. Meanwhile, the model uncertainty is also discussed to some extent.**

**Keywords-*MOPSO;NSGA II;MOSCEM-UA; Pareto front; Hymod model; spacing rate; dominant rate***

## I. INTRODUCTION

Because of the difficulty to directly acquire the parameters of underlying surface information, the simulation with conceptual hydrological model on a natural basin has made it difficult to obtain ideal parameter set. The optimization of the model structure and parameter value of the conceptual hydrological model constitutes an important factor to determine the effect of the simulation. Generally speaking, hydrological model can be manually adjusted to obtain the ideal parameter values, but the workload of manual calibration is significant. The efficiency of adjustment can be greatly enhanced by selecting rational auto calibration approach and using high-speed iteration of computer. There are several approaches in multi-objective optimization in terms of hydrological parameter optimization: Genetic Algorithm (GA) and its improvement SA-GA and NSGA-II; SCE-UA [1] and its improvement MOSCEM-UA [2] etc.

Particle Swarm Optimization (PSO) is a kind of heuristic algorithm emerging in recent year, whose basic idea stems from the flock flight path when searching for food. PSO algorithm has been greatly developed since it appeared (by Kennedy and Eberhart [3] et al.), and MOPSO algorithm has also been improved to some extent. The development from ordinary PSO to MOPSO has some problems to be solved such as how to keep global optimum location of each swarm generation and the spacing rate of non-dominant solutions etc. In this essay, an improved MOPSO method, that is, based on external archive and crowding distance (CD), has been introduced to the auto calibration of hydrological model-Hymod. The Pareto front approximation ratio, Pareto solution non-dominant spacing rate under the same iteration times are analyzed through the comparison of MOSCEM-UA and NSGA-II with MOPSO method. Meanwhile, the model uncertainty is also discussed to some further extent.

NSGA-II is an improved algorithm by adding hybridization encouraged mechanism and realized in elitist non-dominated sorting genetic algorithm (NSGA-II). This mechanism uses the normalized distance to evaluate the difference among genes in a population. It has a better sorting algorithm, incorporates elitism and no sharing parameter needs to be chosen a priori. Bekele, E. G. and J. W. Nicklow [4] (2007) et al. has introduced NSGA-II in the parameter optimization of SWAT (Soil and Water Assessment Tool). The algorithm has proved its capability of incorporating multiple objectives into the process of calibration and also employing parameterization in order to reduce the number of calibration parameters.

Another widely used algorithm is the MOSCEM-UA (Multi-objective Shuffled Complex Evolution Metropolis). The MOSCEM-UA method is known as a general-purpose global optimization approach designed to infer the traditional "best" parameter sets and the underlying posterior distribution within a single iteration. Luc Feyen, Jasper A. Vrugt et al (2003) [5] has applied SCEM-UA (single objective) in the optimization and uncertainty assessment of hydrologic model parameters for the probability-distribution. The algorithm operates through merging the strengths of the Metropolis algorithm, controlled random search, competitive evolution,

978-1-4244-5961-2/10/$26.00 ©2010 IEEE 

and complex shuffling in order to continuously update the proposal distribution and evolve the sampler to the posterior target distribution. The possibilities and limitations of the two algorithms to evaluate the behavior of model parameters have been compared.

Although the aforementioned optimization methods have performed quite well in the literature, the comparison among MOPSO, NSGA-II and MOSCEM-UA has not been conducted to get a better one in the calibration of a hydrological model Hymod. In this essay, time consumption under same iteration and population size, Pareto solutions dominant rate and spacing rate are taken as three criterions to compare the performance on the calibration of hydrologic model Hymod.

## II. CONCEPTUAL HYDROLOGICAL MODEL-HYMOD

In Hymod model the imaginary catchment (Figure 1) can be considered as an infinite amount of points without interaction between these them. And this theory is fully based on Moore's concept [6]. Take one point for an instance, this point has a certain column of water storage capacity (referred as *Cmax*), of which a portion is filled up as initial water storage. Other variables related to this point are precipitation (*mm)* and evaporation (*mm*) rates in a certain period of time. The excess of water storage are regarded as catchment's runoff, when the water storage *Cmax* is filled up.

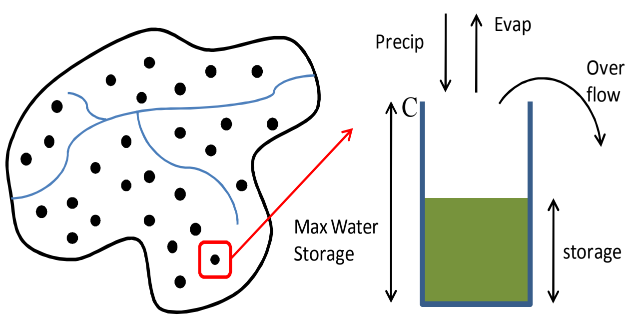



Figure 1. Precipitation-Runoff mechanism of Hymod model

Values of parameters such as soil structure, soil texture and water storage capacities among different points can varies spatially within the range of the catchment. The water storage capacities distribution function can be defined as:

$$F(C)=1-\left(1-\frac{C}{C_{\max}}\right)^{B_{\exp}} \quad 0\le C\le C_{\max} \qquad (1)$$

Where:

$F$ is the cumulative chance of a certain water storage capacity if a point is selected [-];

$C_{\max}$ is the largest water storage capacity of a point among all the points with in the catchment [mm];

$B_{\exp}$ is the degree of spatial variability in the water storage capacities[-].

When a catchment water storage capacity is partially filled up with water, the precipitation that falls and exceeds $C_{\max}$ can be directed through three linear quick flow tanks as the flow can not infiltrate into the soil. The flow rate of between these tanks depends on the constants RQ.

If the remaining precipitation that exceeds the water storage capacity of points with a lower capacity than $C_{\max}$, it can be divided into quick flow tanks and slow flow tank depending on the constant *Alpha*. Some part of water in the catchment evaporates when there is enough water available. The remaining water after outflow and evaporation becomes the water storage for the next time step. The detailed review of the model concept can be found in the literature[6]

Theatrically, the 5 model parameters of Hymod model can be defined with a reasonable range (Table I).

TABLE I. DEFAULT BOUNDARY OF PARAMETERS IN HYMOD MODEL

| *parameter* | *default* | *minimum* | *maximum* |
|---|---|---|---|
| $C_{\max}$ | 250 | 200 | 500 |
| $B_{\exp}$ | 0.3 | 0.01 | 2 |
| *Alpha* | 0.9 | 0.5 | 0.99 |
| *Rs* | 0.02 | 0.01 | 0.2 |
| *Rq* | 0.5 | 0.3 | 0.7 |

## III. MULTIOBJECTIV OPTIMIZATION ALGORITHM

### A. *Multi-Objective Optimization*

The recent studies on natural computation algorithm have shown that the population-based algorithm are potential candidate to solve multi-objective optimization problems and can be efficiently used to eliminate most of the difficulties of classical single object method. A multi-objective problem of a hydrological model can be described as to seek for best parameter set or variables that are able to minimize or maximize the objective vector. The formula of the objective function can be expressed as follows:

$$x=[x1,x2,\cdots,x_n]^T$$
$$\text{minimize: } y=f(x)=\{f_1(x),f_2(x),\cdots,f_m(x),\} \qquad (2)$$
$$\text{s.t.}: x\in S=\{x\mid g_j(x)\le 0,\ j=1,2,\cdots,p\}$$

Where: $x$ is the input decision variable vector; $y$ is the objective vector; $g_j(x)$ is the $j^{th}$ constrains; $S$ is the feasible solution domain.

### B. *Multiobjective particle swarm optimization*

External archive is a method proposed by Raquel et al. [7], which based on the single-objective PSO using an external group (Archive) to save non-dominant solutions of each iteration and preventing their loss. It guarantees that the algorithm has a better convergence, and can timely replace the original value when a dominant found in the next iteration. After Crowding Distance is introduced, non-dominant solution with the best crowding distance can be selected from the archive as the global solution. The steps are listed as follows

***Step1***(initialization): Set the counter *G=0* and initialize the randomly *n* particles with uniform probability selection over the optimized parameter search space. Similarly, initial velocities of all particles are also randomly generated with uniform probability over the dimension. Then the archive volume is set to *Nset* where non-dominants can be stored.

***Step2***(crowding distance):each particle in the initial population is evaluated. And search for the non-dominated solutions and form the non-dominated global set using crowding distance.

***Step3***(Archiving): store the non-dominants in the archive if there is enough position left. If not, check the previous particles in the archive if they are dominated by the current particle. If true, replace them. Conduct crowding distance in the archive and updates the particles in the archive.

***Step4***(velocity updating):particle velocity in the dimension is updated depending on constrains equation. If any particle violates the velocity limits, set its velocity equal to the limit or reset it by multiplying a negative number to the opposite direction.

***Step5***(position updating):each particle changes its position according to the updated velocity.

***Step6***(time updating): update the counter *G* by checking it under the iteration limit *Gen.*

***Step7***(drawing Pareto):collect the non-dominants in the archive and draw the Pareto front with the evaluated objective functions.

***Step8***(stopping criteria):if the number of iterations exceeds its maximum then stop, else go to step 2.

### *C. Non-dominated Sorting Gentic Algorithm II*

NSGA II, compared with NSGA [8] (Srinivas and Deb, 1994), was improved to have a better sorting algorithm, incorporating elitism and having no need to choose a priori. The algorithm takes advantage of a fast non-dominating sorting approach to discriminate solution based on the concept of Pareto dominance and optimality. The population is stored into each front once the population is initialized. Front of the current iteration is non-dominant and would be replaced by individuals of the next iteration, and the front goes on. Each individual in each front are assigned with rank fitness values, and the crowding distance will be calculated for each individual. It measures the close distance between the individual and its neighbors. The crowding distance helps each front to have a good diversity in the population. The parents are selected form the population through binary tournament. Only the individuals with better crowding distance will be chosen to generate off-springs from crossover and mutation. Then the off-springs and the current population will be combined for next selection based on rank and crowding distance. The algorithm can be reviewed in detail in the literature [10].

### *D. Multi-Objective Shuffled Complex Evolution Metropolis*

SCEM-UA (Shuffled Complex Evolution Metropolis) algorithm is an improvement based on SCE-UA [9]. SCEM-UA adopts Metropolis sampling method in the variable space. MOSCEM-UA [6] (Multi-Objective Shuffled Complex Evolution Metropolis), however, uses initialized evolutionary Pareto dominants that get close to the uniform distribution and then solves the multi-objective optimization. The algorithm is able to infer most likely variable sets and their underlying posterior probability distribution within a single iteration.

### *E. Performance Criterion used in optimization*

Three criteria of Pareto optimal are proposed by Zitzler[10] et al. to assess the performance of the algorithms: GD (Generation Distance) between solution set drawn by the algorithm and the ideal Pareto set; non-dominant Pareto front spacing rate SP (Spacing); the ratio of number of solution sets drawn by algorithm that do not belong to Pareto optimized solution by ER (Error). However, due to the complexity of hydrological models, the ideal Pareto solutions of the objectives cannot be drawn, and therefore when comparing the three optimization methods under the same iteration times, Pareto solution dominant rate, Pareto front spacing rate and iteration time can be used as evaluation criteria.

*a) The spacing rate of non-dominant solutions:*

Spacing rate is an important criterion to evaluate the Pareto front. It can be written as:

$$SP = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\bar{d} - d_i)}^{2} \tag{3}$$

Where: $n$ is the number of non-dominant solutions, $d_i$ stands for the distance of the individual between its neighbor solutions. Under ideal circumstance, *SP*=0.

*b) Non-dominant solutions dominant rate:*

Under the same iteration times, the dominant rate refers the ratio of how many individuals in Pareto (drawn from optimization method I) are dominated by the other. The equation can be expressed as:

$$C_{1,2} = \frac{d_{dominant}}{d} \times 100\% \tag{4}$$

Where: $d_{dominatt}$ is the number of dominated individuals by Pareto set II; $d$ is the number of individuals in Pareto I.

## IV. Case Study and Comparison

### *A. The Objective Function of Hymod model*

Different objective functions can be used to describe different characteristics of hydrological process, which is the primary means to assess the efficiency of hydrological model measured values and simulation values. This essay has selected 2 model objective functions and requiring that the objectives are to be reduced to minimum. The objectives are expressed as follows:

*a) Objective function for high flow* $F_H(\theta)$:

$$F_H(\theta) = \frac{\sum_{i=1}^{N} (Q_{o,i} - Q_{s,i})^2}{\sum_{i=1}^{N} (Q_{o,i} - \overline{Q_o})^2} \tag{5}$$

b) *Objective function for low flow* $F_L(\theta)$:

$$F_L(\theta) = \frac{\sum_{i=1}^{N} (\log(Q_{o,i}) - \log(Q_{s,i}))^2}{\sum_{i=1}^{N} (\log(Q_{o,i}) - \log(\overline{Q_o}))^2} \tag{6}$$

Where: $Qs$ is simulation value; $Qo$ is observed data; $\overline{Qo}$ is the mean value of observed data; $\theta$ is model parameter space.

### B. Basin Status

Xiangjiaping sub-basin (Figure 5) in this study is selected from Xun River, a major upper stream tributary of Han River. The control area of Xiangjiaping hydro-station is 6397 km$^2$, and research data includes 3 years' rainfall, measured runoff, and evaporation from year 1982 to 1984. The calculation time step of Hymod model is one day.

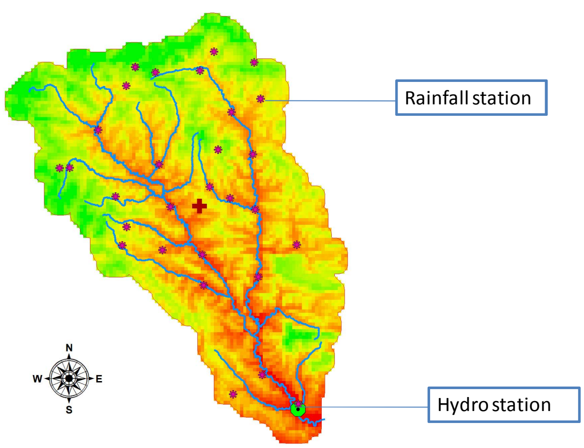



Figure5. The topography and stations of XiangJiapin sub-basin

### C. Comparative Analysis of Optimization Results

In order to discuss the performance of the optimization methods, set the same iteration (*Iter=10000*) and the population size *P=200*. The time consumption during iteration, Pareto dominating rate and spacing rate of obtained Pareto individuals are taken into consideration. The experiment was carried out using a PC with cup 2.40GH quadrupled and memory of 4G. Considering the randomness of optimization methods, the test result was the mean value on a five-run process. The calculated results are showed in the table below (Table 2), where $I$ is the iteration times, $\overline{T}_5\,(s)$ is the average running time of concerned optimization method, $\overline{C}_x\,(\%)$ is the average dominated rate ratio of selected optimization method, $\overline{SP}_5$ is the average Pareto front spacing rate of the selected optimization.

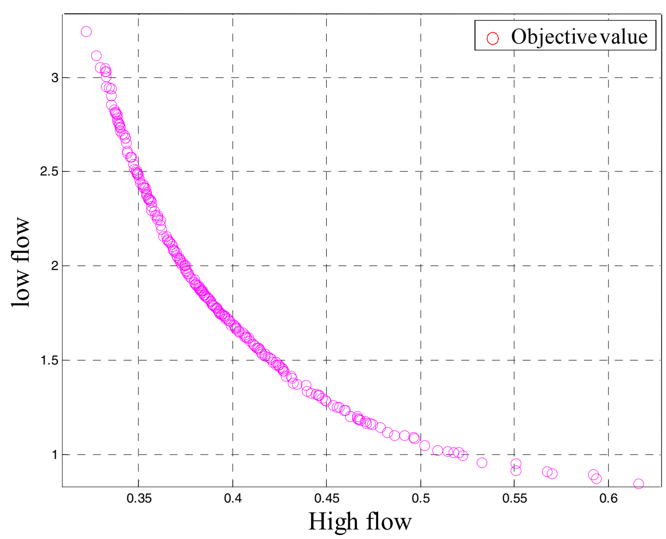
From the table, it can be seen that when iteration times is relatively low, the iteration convergence of MOPSO is faster, and the dominated rate is lower. Because the crowding distance method is applied, the spacing rate of non-dominant solutions is ideal. When iteration times are greatly increased, the archive volume of MOPSO exceeds and the crowding distance of the particle in next iteration will compare with every particle in archive. Thus the iteration calculation time will be increased. The other two methods NSGA-II and MOSCEM-UA do not perform in the spacing rate of Pareto solutions, but when iteration times increases their generation distance are relatively ideal. On the whole, the performance of MOPSO is quite suitable for Hymod model calibration with less iteration.

### D. Analysis hydrologic process with MOPSO

Here we take MOPSO as an example to analyze the result of optimization. The archive capacity of MOPSO method is set to *Nset=300*, number of particles *P=50*, iteration *Gen=10000.* Figure 3(a) indicates the Pareto front of optimization result depending on the high flow objective and the low flow objective. From Pareto front it can be seen that there exists constraint relationship between the two objective functions, so there is no perfect parameter set that can simultaneously fit both the two objectives. Figure 3(b) indicates the runoff forecast range drawn from corresponding parameters of Pareto solutions. It can be seen that the observed flow data is basically included in the prediction range, showing the degree of uncertainty of the parameters. Figure 3(c) and Figure 3(d) indicate that under constraint and non-constraint condition respectively, the normalized values of corresponding parameter value of Pareto solutions, among which the thick solid line indicates shows the parameter when high flow objective is the optimal, the thick dashed line indicates the parameter when low flow objective is the optimal. It can be seen that under non-constraint condition the optimized result of model parameter $C_{\max}$ may be beyond its physical range, which is not conform to the reality. This also tells that the further requirement of improvement on the Hymod model structure should be made.

TABLE II. TIME USED, PARETO DOMINATED RATE AND SPACING RATE UNDER THE SAME ITERATION

| **Optimization Method** | ***Iteration*** | $\overline{T}_5$ **(s)** | $\overline{C}_x$ **(%)** | | $\overline{SP}_5$ |
|---|---|---|---|---|---|
| MOPSO | 100 | 30.78 | $\overline{C}_{SCE} = 10.13$ | $\overline{C}_{NSGA} = 69.80$ | 0.0161 |
| MOSCEM-UA | 100 | 13.27 | $\overline{C}_{PSO} = 66.27$ | $\overline{C}_{NSGA} = 80.27$ | 0.0606 |
| NSGA-II | 100 | 18.63 | $\overline{C}_{PSO} = 11.80$ | $\overline{C}_{SCE} = 13.20$ | 0.0178 |
| MOPSO | 1000 | 213.45 | $\overline{C}_{SCE} = 32.72$ | $\overline{C}_{NSGA} = 62.55$ | 0.0172 |
| MOSCEM-UA | 1000 | 150.34 | $\overline{C}_{PSO} = 41.73$ | $\overline{C}_{NSGA} = 72.27$ | 0.0346 |
| NSGA-II | 1000 | 184.36 | $\overline{C}_{PSO} = 9.00$ | $\overline{C}_{SCE} = 29.40$ | 0.0187 |
| MOPSO | 5000 | 635.55 | $\overline{C}_{SCE} = 31.95$ | $\overline{C}_{NSGA} = 37.62$ | 0.0381 |
| MOSCEM-UA | 5000 | 557.75 | $\overline{C}_{PSO} = 38.67$ | $\overline{C}_{NSGA} = 59.87$ | 0.0235 |
| NSGA-II | 5000 | 889.56 | $\overline{C}_{PSO} = 21.21$ | $\overline{C}_{SCE} = 36.40$ | 0.0184 |
| MOPSO | 10000 | 2434.62 | $\overline{C}_{SCE} = 27.20$ | $\overline{C}_{NSGA} = 31.24$ | 0.0243 |
| MOSCEM-UA | 10000 | 1923.53 | $\overline{C}_{PSO} = 29.73$ | $\overline{C}_{NSGA} = 60.80$ | 0.0304 |
| NSGA-II | 10000 | 1828.35 | $\overline{C}_{PSO} = 15.40$ | $\overline{C}_{SCE} = 33.40$ | 0.0180 |




Figure(a). Pareto front of high flow and low flow objectives

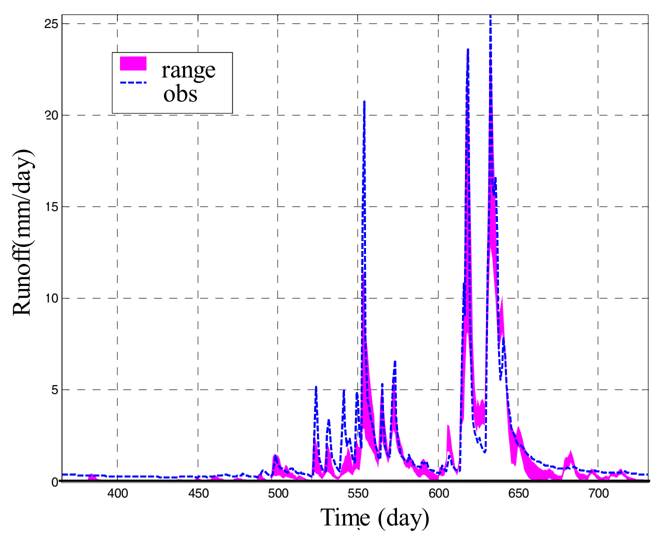



Figure3(b). Runoff forecast range drawn from corresponding parameter of Pareto solutions

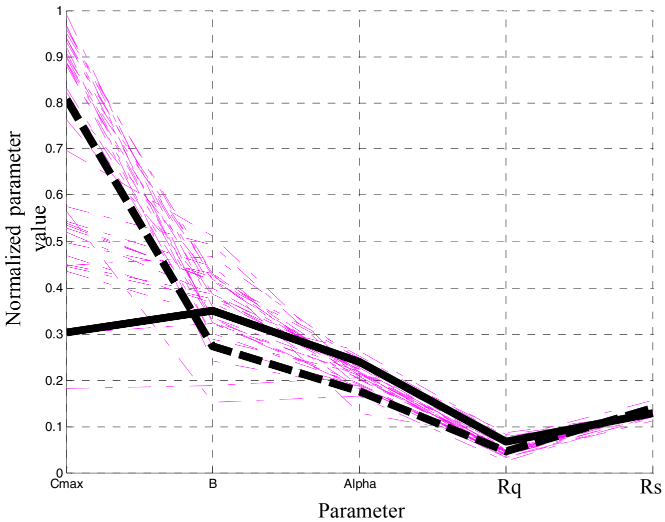



Figure3(c). Normalized value of parameters corresponding Pareto solutions with constrains

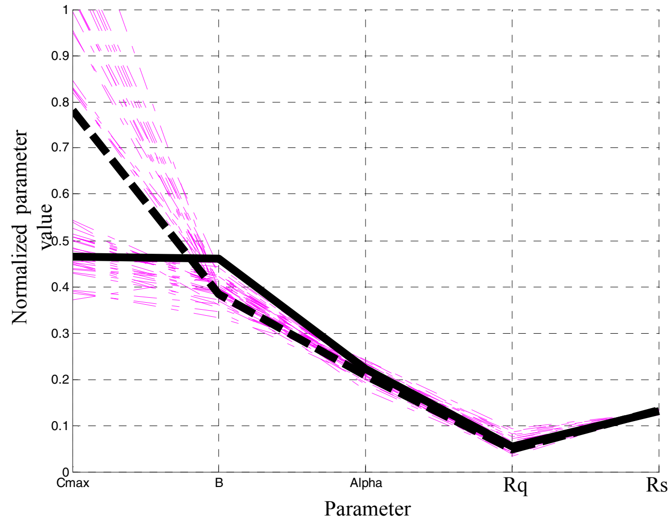



Figure3(d). Normalized value of parameters corresponding Pareto solutions without constrains

## V. CONCLUSIONS

*a)* Through the comparison of the three optimization methods, it can be concluded that MOPSO is characterized with well performance in spacing rate of Pareto front, faster convergence speed under the same iteration times and low dominated rate of Pareto solutions.

*b)* When MOPSO is applied to Hymod, the spacing rate of obtained Pareto front is more ideal, and so is the generation distance. The MOPSO can perform a good distributed Pareto front.

*c)* From the optimized results of model parameters through MOPSO it can be seen that the model parameters have a certain degree of uncertainty. In the meantime through non-constraint scheme it can be seen that the model parameter may be beyond its physical range, which further illustrate that the model structure needs improvement.

In a word, the MOPSO algorithm, compared with the other two algorithms, is the most suitable optimization method on the calibration of Hymod model.

## ACKNOWLEDGMENT

This paper is sponsored by the following projects: 1)by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (Grant No. 50721006); 2) by "The 11th Five-Year Plan" National Major Technology for South-to-North Middle Route Water Diversion and Management Key Technology Research (Grant No. 2006BAB04A07).

## REFERENCE

[1] Duan Q, V K Gupta, S Sorooshian. "Effective and efficient global optimization for conceptual rainfall-runoff models", Water Resource Research, vol.28, pp.1015-1031, 1992.

[2] Bastola S, H Ishidaira. "Regionalization of hydrological model parameters under parameter uncertainty: A case study involving TOPMODEL and basins across the globe", Journal of Hydrology, vol.357 (3-4), pp.188-206, 2008.

[3] J Kennedy, R C Eberhart, "Particle swarm optimization" Proc IEEE International Conference on Neural Networks, pp.1942-1948, 1995.

[4] Bekele, E. G. and J. W. Nicklow, "Multi-objective automatic calibration of SWAT using NSGA-II", Journal of Hydrology, vol.341 (3-4), pp.165-176, 2007.

[5] Feyen, L., J. A. Vrugt, et al., "Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model", Journal of Hydrology, vol. 332(3-4), pp. 276-289, 2007.

[6] Vreng, A. B. A. d. ,"Parameter optimaization of the HYMOD model using SCEM-UA and MOSCEM-UA." University of Amsterdam, pp. 1-39, 2006.

[7] C R Raquel, P C Naval, "An effective use of crowding distance in multi-objective particle swarm optimization", In Proc. of Genetic and Evolutionary Computation Conference, pp.257-264, 2005.

[8] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., "A fast and elitist multi-objective genetic algorithm: NSGA-II", IEEE Transactions on Evolutionary Computation, vol. 6(2), pp. 182–197, 2002.

[9] Duan Q, GUPTA V K, SOROOSHIAN S., "Shuffled complex evolution approach for effective and efficient global minimization", Journal of Optimization Theory and Applications, vol. 76 (3), pp. 501-521, 1993.

[10] Zitzler E, Deb K, Thielel, "Comparison of multi-objective evolutionary algorithms: empirical results", IEEE Transaction on Evolutionary Computation, vol. 18(2), pp. 173-195, 2008.