

ỨNG DỤNG THUẬT TOÁN SCE TỐI ƯU HÓA TỰ ĐỘNG CÁC THÔNG SỐ CỦA MÔ HÌNH MƯA – DÒNG CHẢY

ThS. Nguyễn Đức Hạnh, CN. Hoàng Thị Mỹ Linh

Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

ThS. Phùng Đức Chính - Viện Khoa học Khí tượng Thủy văn Và Môi trường

Bài báo trình bày kết quả nghiên cứu, xây dựng chương trình tự động xác định các thông số của mô hình NAM (MDC-TUH) trên cơ sở sử dụng thuật toán tối ưu toàn cầu SCE (Shuffled Complex Evolution); đồng thời tiến hành đánh giá ảnh hưởng của việc tính mưa bình quân lưu vực bằng các phương pháp tính trọng số trạm đo mưa khác nhau đến kết quả mô phỏng dòng chảy. Kết quả áp dụng thử nghiệm cho lưu vực sông Ba (tính đến trạm An Khê) cho thấy khả năng tự động dò tìm các thông số là tối ưu và mô phỏng dòng chảy cho kết quả tốt.

Từ khóa: Thuật toán SCE, tối ưu hóa tự động, mô hình mưa – dòng chảy.

1. Đặt vấn đề

Dựa vào sự hiểu biết về các hiện tượng thủy văn, cùng với sự hỗ trợ đắc lực của công nghệ thông tin, các mô hình mưa – dòng chảy xuất hiện ngày càng nhiều và mô phỏng ngày càng chính xác. Các mô hình này đều được đặc trưng bởi các thông số. Các thông số này có thể được xác định theo phương pháp thử sai bằng cách lần lượt gán các giá trị cho các thông số đến khi tìm được bộ thông số sao cho có sự phù hợp nhất giữa đường quá trình lưu lượng quan trắc và tính toán. Nhưng mỗi thông số lại có mức độ ảnh hưởng khác nhau đến kết quả mô phỏng; hơn nữa các thông số cũng ảnh hưởng lẫn nhau, do đó để tìm được bộ thông số tối ưu đòi hỏi nhiều thời gian, công sức, kinh nghiệm của người hiệu chỉnh về độ nhạy của các thông số cũng như tác động qua lại giữa chúng. Vì vậy, cần xây dựng một thủ tục tự động hiệu chỉnh các thông số của mô hình một cách hiệu quả.

Chính nhờ sự khách quan, nhanh chóng và dễ thực hiện của việc tự động tối ưu các thông số của mô hình mà phương pháp này đã được áp dụng rộng rãi và thêm vào trong phần lớn các mô hình mưa – dòng chảy. Kết quả tối ưu tự động không phụ thuộc vào người hiệu chỉnh mà chỉ phụ thuộc vào: thuật toán tối ưu; hàm mục tiêu; dữ liệu hiệu chỉnh; cấu trúc và tính phức tạp của mô hình; khả năng xác định thông số và sự tương tác giữa các thông số [1].

Trong những năm gần đây, nhiều nghiên cứu đã được thực hiện để tự động tối ưu các thông số của

các mô hình quan niệm cũng như mô hình thủy động lực. Trong lĩnh vực này, các thuật toán dựa vào tập hợp mẫu như thuật toán di truyền (GA), chiến lược phát triển (ES), tiến triển phức hợp được xáo trộn (SCE)[2],... đã cho thấy những hiệu quả trong việc tối ưu. Trong một số nghiên cứu, khi so sánh thuật toán SCE với các phương pháp khác như: phương pháp tìm kiếm ngẫu nhiên, phương pháp kết hợp, phương pháp Multi Start – Simplex [2]; phương pháp SGA, GA [3], phương pháp GA, PSO [4],... cho thấy rằng thuật toán SCE là hiệu quả nhất. Ngoài ra, thuật toán này cũng đã được áp dụng thành công cho nhiều mô hình khác nhau như: mô hình Pitman [5]; mô hình Tank [4]; mô hình Xinanjiang [6], MIKE/NAM [7], [8]... Do đó nghiên cứu này sử dụng thuật toán SCE để tự động tối ưu các thông số của mô hình.

2. Thuật toán SCE (Shuffled Complex Evolution)

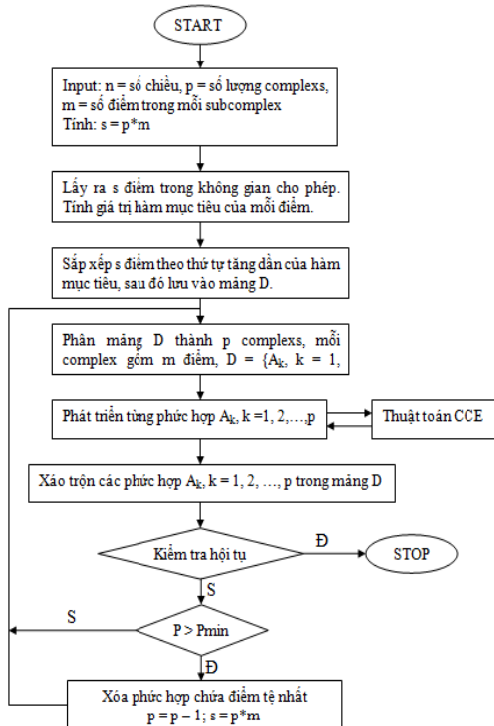
a. Mô tả thuật toán

Thuật toán tiến triển phức hợp được xáo trộn của trường Đại học Arizona (SCE-UA) được đề xuất để giải quyết vấn đề ước lượng tham số trong mô hình mưa dòng chảy quan niệm [2]. Hiện nay, phương pháp này được sử dụng rộng rãi trong nhiều lĩnh vực kĩ thuật khác nhau, đặc biệt trong các mô hình thủy văn.

Phương pháp này được xây dựng dựa trên 3 khái niệm: sự kết hợp giữa phương pháp Downhill Simplex của Nealder Mead [9] với các khái niệm của phương pháp tìm kiếm ngẫu nhiên có kiểm soát

[10]; sự phát triển cạnh tranh [11]; xáo trộn các phức hợp. Sự kết hợp 3 khái niệm này cho phép có thể giải nhanh và tránh được các cực trị địa phương, phù hợp để giải các bài toán phi tuyến phức tạp. Các nghiên cứu đã khẳng định phương pháp này có tính ổn định và hội tụ tốt ([2], [12], [13],...)

Phương pháp SCE được khởi tạo bằng cách chọn các tham số p và m (p là số phức hợp, m là số điểm trong mỗi phức hợp). Không gian mẫu s là mẫu lấy ngẫu nhiên trong không gian khả thi của các thông số sử dụng một phân phối xác suất đồng nhất và tính toán giá trị hàm mục tiêu tại mỗi điểm đó. Các điểm trong s đều được sắp xếp theo thứ tự tốt dần giá trị của hàm mục tiêu. Các điểm này sẽ được chia thành p phức hợp, mỗi phức hợp gồm m điểm. Mỗi phức hợp sẽ tiến hóa một cách độc lập theo phương pháp Downhill Simplex [9]. Tiếp theo, tiến hành xáo trộn và kết hợp các điểm trong các phức hợp đã phát triển thành một tập mẫu mới dựa trên thông tin của tập mẫu ban đầu. Sự phát triển và xáo trộn này sẽ được lặp đi lặp lại cho tới khi nào các tiêu chí hội tụ được thỏa mãn.



Hình 1. Sơ đồ khối thuật toán SCE

b. Các bước của thuật toán

1) Phương pháp SCE

Bước 1. Khởi tạo tham số: p , m với $p \geq 1, m \geq n +$

1. Trong đó: p là số phức hợp, m là số điểm trong mỗi phức hợp, n là số thông số hiệu chỉnh thì $s = p*m$ là không gian mẫu.

Bước 2. Tạo ra nhóm giá trị: Lấy ra ngẫu nhiên s điểm $x_i [x_1, x_2, \dots, x_s]$ trong không gian khả thi của các thông số. Sau đó tính giá trị f_i của hàm mục tiêu.

Bước 3. Sắp xếp: Sắp xếp s điểm x theo thứ tự tăng dần ý nghĩa giá trị của hàm mục tiêu và lưu trong mảng $D, D = [(x_i, f_i), i=1, 2, \dots, s]$, ứng với $i=1$ giá trị của hàm mục tiêu là kém nhất.

$$A_k = \left[x_j^k, f_j^k \mid x_j^k = x_{k-p(j-1)}, f_j^k = f_{k+p(j-1)} \right]$$

Bước 4. Phân nhóm: Chia mảng D thành p phức hợp A_1, A_2, \dots, A_p với mỗi phức hợp gồm m điểm, sao cho:

Bước 5. Phát triển từng phức hợp một: Mỗi phức hợp $A_k, k = 1, 2, \dots, p$ được phát triển theo thuật toán CCE (Competitive Complex Evolution) [14].

Bước 6. Xáo trộn các phức hợp: Kết hợp các điểm trong các phức hợp đã được phát triển trong bước 5 thành một tập mẫu duy nhất và sắp xếp lại tập mẫu này theo thứ tự tăng dần ý nghĩa hàm mục tiêu. Phân nhóm lại tập mẫu vào p phức hợp theo cách thức ở bước 4.

Bước 7. Kiểm tra điều kiện hội tụ. Nếu thỏa mãn điều kiện thì dừng lại, nếu không thì tiếp tục.

Bước 8. Kiểm tra số lượng phức hợp. Nếu số lượng nhỏ nhất của các phức hợp được đề nghị p_{min} nhỏ hơn p , xóa phức hợp chứa điểm kém nhất, đặt $p = p - 1, s = p*m$. Quay lại bước 4. Nếu $p_{min} = p$ thì quay lại bước 4.

2) Cách lựa chọn thông số của thuật toán

Phương pháp SCE chứa nhiều thành phần ngẫu nhiên và tất định. Các thông số của thuật toán cần được lựa chọn cẩn thận vì nó ảnh hưởng đến hiệu quả và tính tối ưu của thuật toán. Các thông số đó là số phức hợp p , số điểm trong mỗi phức hợp m , số điểm mới mà mỗi phức hợp con tạo ra là α và số bước phát triển β .

Về mặt lý thuyết, m có thể lấy bất kì sao cho $m \geq n$. Nhưng nếu số lượng điểm trong mỗi phức hợp quá nhỏ thì cách thức tìm kiếm này sẽ tương tự như thủ tục tìm kiếm thông thường và khả năng tìm kiếm toàn cầu sẽ bị suy yếu. Còn nếu m quá lớn thì sẽ mất nhiều thời gian để tính toán mà hiệu quả đạt

NGHIÊN CỨU & TRAO ĐỔI

được cũng không được tốt hơn nhiều.

Số điểm trong mỗi phức hợp q , có thể thay đổi trong khoảng từ 2 đến m .

Số lượng điểm mới mà mỗi phức hợp con được tạo ra trước khi nó được đưa lại vào phức hợp cũ, α 1. Nếu bằng 1 thì chỉ có một trong số những điểm ban đầu được thay thế, còn nếu lớn hơn 1 thì việc tìm kiếm sẽ thiên mạnh mẽ hơn về tìm kiếm địa phương của không gian tham số.

Số bước phát triển của thuật toán β 0. Nếu β là nhỏ, khu phức hợp sẽ được xáo trộn thường xuyên, còn nếu β lớn thì mỗi phức hợp sẽ bị thu vào một nhóm nhỏ, làm mất đi hiệu quả của tìm kiếm toàn cầu.

Q. Duan và các cộng sự [4] đã kiến nghị giá trị của các thông số trong phương pháp SCE như sau:

Số trạm đo mưa				
STT	Tên trạm	File số liệu	Trọng số	Tối ưu
1	An Khê	Nhập	0.145	<input type="checkbox"/>
2	Kbang	Nhập	0.855	<input type="checkbox"/>
3		Nhập		<input type="checkbox"/>
4		Nhập		<input type="checkbox"/>
5		Nhập		<input type="checkbox"/>
6		Nhập		<input type="checkbox"/>
7		Nhập		<input type="checkbox"/>
8		Nhập		<input type="checkbox"/>
9		Nhập		<input type="checkbox"/>
10		Nhập		<input type="checkbox"/>

TIẾP TỤC

Hình 2. Bảng nhập số liệu mưa

số điểm của mỗi phức hợp $m = 2n + 1$, số điểm trong mỗi phức hợp con $q = n + 1$, $\alpha = 1$, $\beta = 2n + 1$. Như vậy, tham số duy nhất mà người sử dụng cần đưa vào là số phức hợp p .

3. Xây dựng chương trình tính toán

Chương trình MDC-TUH là chương trình tự động tối ưu các thông số của mô hình NAM được xây dựng dựa trên cơ sở lý thuyết của mô hình NAM và thuật toán SCE.

Chương trình này được xây dựng bằng ngôn ngữ lập trình Visual Basic - Microsoft Visual Studio Express 2012, với giao diện trực quan và thuận lợi cho người sử dụng. Từ dữ liệu đầu vào là số liệu mưa, bốc hơi và lưu lượng quan trắc chương trình sẽ tự động xử lý số liệu, tính toán và cho ra dữ liệu đầu ra là lưu lượng tính toán và đồ thị biểu diễn mối quan hệ giữa các giá trị thực đo và tính toán.

Input		Mô phỏng		Kết quả		Thoát	
Lưu vực							
Tên lưu vực		An Khê					
Diện tích		1394					
Nhập file dữ liệu							
STT	Lưu lượng mưa (mm)	Bốc hơi (mm)	Lưu lượng (m ³ /s)	Q mô phỏng (m ³ /s)			
1	3.1764	4.5	113	25.192...			
2	4.8483	4.4	86.2	25.049...			
3	3.7159	5	79.6	24.907...			
4	5.0281	4.8	72.3	24.766...			
5	8.6274	3.3	68.3	24.626...			
6	29.2834	5.3	98	24.486...			
7	5.3072	5.1	119	24.384...			
8	7.2363	4.8	82.9	24.269...			
9	5.5838	3.9	75.4	24.151...			
10	30.7011	4.2	201	24.149...			
11	26.7806	3.6	518	24.407...			
12	12.6541	4.2	317	77.713...			
13	6.776	4.6	208	60.789...			

Nhập điều kiện ban đầu					
U/Umáx	0.2				
L/Lmáx	0.5				
QOF	0				
QIF	0				
BF	1				
Các thông số					
Thông số	Min	Max	Fit	Giá trị	
Umáx	5	35	<input type="checkbox"/>	11.532860	
Lmáx	50	350	<input type="checkbox"/>	86.726702	
CQOF	0.01	0.99	<input type="checkbox"/>	0.8215269	
CKIF	50	1000	<input type="checkbox"/>	518.46870	
TOF	0	0.8	<input type="checkbox"/>	0.6624705	
TIF	0	0.8	<input type="checkbox"/>	0.1147979	
CK12	3	72	<input type="checkbox"/>	25.563442	
TG	0	0.7	<input type="checkbox"/>	0.5796620	
CKBF	500	5000	<input type="checkbox"/>	4226.3990	
Hệ số Nash - Shuffle 86.3407110119545					

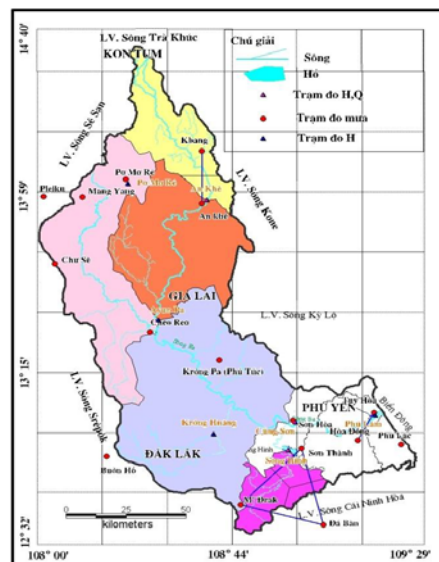
Hình 3. Nhập số liệu đầu vào cho chương trình

4. Ứng dụng chương trình tính toán cho lưu vực sông Ba tính đến trạm An Khê

a. Giới thiệu lưu vực nghiên cứu

Lưu vực sông Ba là một trong 9 lưu vực sông lớn của Việt Nam, là con sông lớn nhất vùng ven biển miền Trung. Lưu vực sông có dạng dài và hẹp, nằm trong phạm vi ranh giới của 3 tỉnh Tây Nguyên là Gia Lai, Kon Tum, Đăk Lăk và tỉnh duyên hải miền Trung là Phú Yên (Hình 4).

Lưu vực nghiên cứu nằm ở thượng lưu lưu vực sông Ba (tính đến trạm An Khê) với diện tích khổng lồ là 1394 km².



Hình 4. Bản đồ lưu vực sông Ba

b. Số liệu đầu vào

Trên khu vực nghiên cứu có trạm thủy văn An Khê là trạm thủy văn cấp I và 02 trạm đo mưa là An Khê và KBang.

Các số liệu đầu vào sử dụng trong nghiên cứu này gồm:

- Số liệu mưa ngày giai đoạn 2002-2005 tại 2 trạm An Khê và KBang.
- Số liệu bốc hơi ngày giai đoạn 2002-2005 tại trạm An Khê.
- Số liệu lưu lượng ngày giai đoạn 2002-2005 tại trạm An Khê.

c. Hiệu chỉnh kiểm định chương trình

Chương trình MDC-TUH sẽ tự động dò tìm bộ thông số tối ưu ứng với 3 trường hợp tính trọng số trạm đo mưa theo các phương pháp khác nhau:

- Phương pháp đa giác Thiesson: trọng số trạm đo mưa được xác định dựa vào bản đồ lưu vực.
- Phương pháp tương quan: trọng số trạm đo mưa được xác định dựa vào việc phân tích tương quan giữa lưu lượng quan trắc tại vị trí cửa ra của lưu vực (Q) và lượng mưa của các trạm đo mưa trên hoặc xung quanh lưu vực đó (Xi).
- Phương pháp tối ưu tự động: Mỗi trọng số được coi như một thông số của mô hình và được xác định theo phương pháp tối ưu tự động. Khi đó số thông số của mô hình sẽ cộng thêm n – 1 thông số, với n là số trạm đo mưa.

Từ đó tiến hành so sánh sự ảnh hưởng của các phương pháp tính trọng số trạm đo mưa đến kết quả mô phỏng.

Chỉ tiêu được sử dụng để đánh giá chương trình là hệ số Nash – Sutcliffe:

Trong đó:

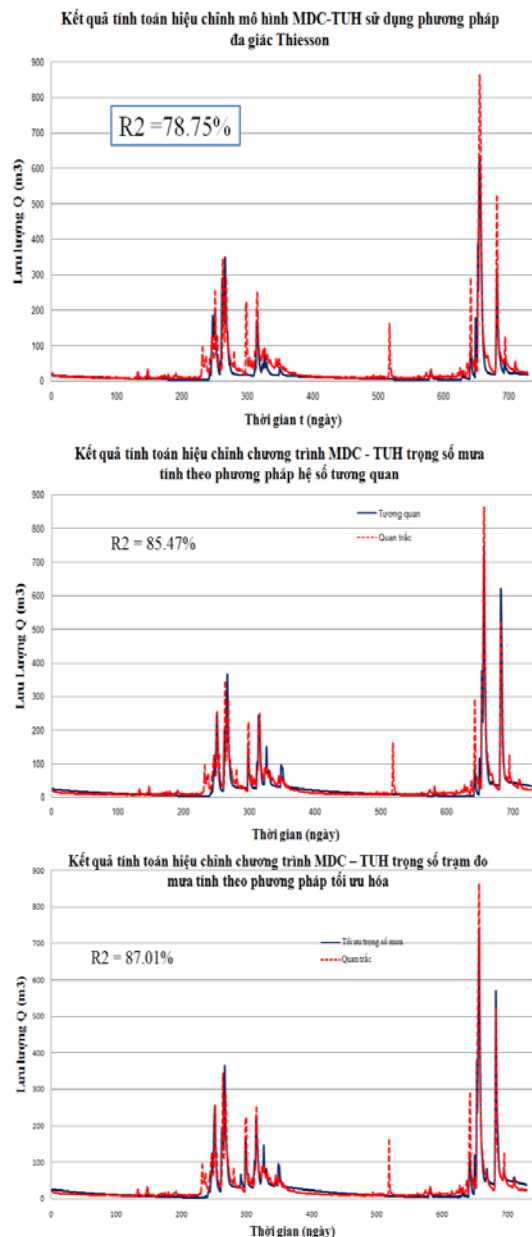
$$R2 = 1 - \frac{\sum_{i=1}^N (Q_{obsi} - Q_{simi})^2}{\sum_{i=1}^N (Q_{obsi} - \overline{Q_{obs}})^2}$$

Qobsi là lưu lượng quan trắc được ở thời điểm i.
 Qsimi là lưu lượng tính toán được ở thời điểm i.
 $\overline{Q_{obs}}$ là lưu lượng quan trắc trung bình.

Hiệu chỉnh chương trình

Sử dụng số liệu ngày giai đoạn 2002 - 2003 để hiệu chỉnh chương trình.

Dựa vào kết quả tính toán có thể thấy rằng, tính mưa bình quân theo phương pháp đa giác Thiesson cho kết quả thấp nhất, đạt loại khá với R2 = 78m75%, phương pháp tối ưu tự động trọng số mưa cho kết quả mô phỏng tốt nhất với R2 = 87m01% (Bảng 1, Hình 5).



Hình 5. Kết quả hiệu chỉnh chương trình MDC-TUH giai đoạn 2002 - 2003

Bảng 1. Trong số trạm đo mưa tính theo các phương pháp khác nhau (tính toán hiệu chỉnh)

Trạm	Phương pháp		
	Đa giác Thiesson	Hệ số tương quan	Tối ưu hóa trọng số mưa
An Khê	0.145	0.449	0.434
Kbang	0.855	0.551	0.566
Hệ số Nash (%)	78.75	85.47	87.01

Kiểm định chương trình

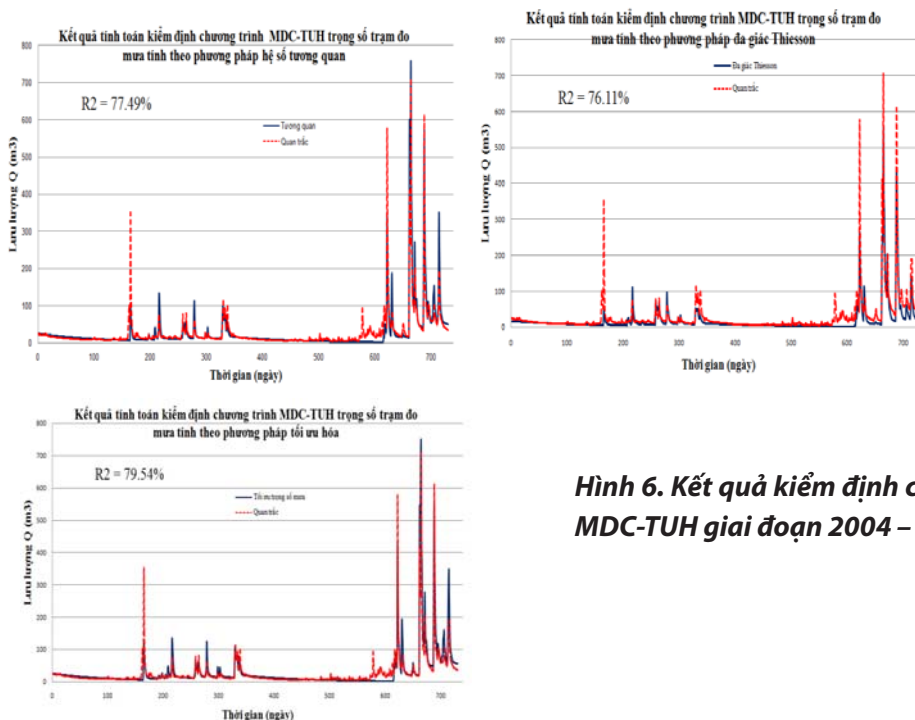
Sử dụng số liệu ngày giai đoạn 2004 - 2005 để kiểm định chương trình.

Kết quả kiểm định cho thấy kết quả mô phỏng đều đạt loại khá, trong đó phương pháp tối ưu tự động trọng số mưa cho kết quả tốt nhất, có hệ số

Nash là 79,54%, cho kết quả thấp nhất là phương pháp đa giác Thiesson với hệ số Nash là 76,11% (Bảng 2, Hình 6).

Bảng 2. Trong số trạm đo mưa tính theo các phương pháp khác nhau (tính toán kiểm định)

Trạm	Phương pháp		
	Đa giác Thiesson	Hệ số tương quan	Tối ưu hóa trọng số mưa
An Khê	0.145	0.449	0.434
Kbang	0.855	0.551	0.566
Hệ số Nash (%)	76.11	77.49	79.54



Hình 6. Kết quả kiểm định chương trình MDC-TUH giai đoạn 2004 –

So sánh chương trình MDC-TUH với các mô hình mưa dòng chảy khác:

Nghiên cứu này đã tiến hành so sánh khả năng

mô phỏng dòng chảy của các mô hình mưa – dòng chảy khác nhau đối với lưu vực An Khê trên sông Ba sử dụng số liệu năm 2002 – 2003. Kết quả được thể hiện ở bảng 3.

Bảng 3. Bảng giá trị hàm mục tiêu của các mô hình mưa dòng chảy áp dụng tính toán hiệu chỉnh cho lưu vực An Khê năm 2002-2003

Mô hình	MDC -TUH	Saccamento	Tank	AWBM	SMAR	Sym Hyd	MIKE /NAM
Hệ số Nash (%)	87.01	68.78	73.04	78.69	60	76.12	82.8

Từ bảng trên có thể thấy, chương trình MDC-TUH áp dụng lý thuyết mô hình NAM và thuật toán SCE có khả năng mô phỏng dòng chảy và dò tìm bộ thông số tối ưu tốt nhất trong số các mô hình mưa dòng chảy nêu trên đối với lưu vực An Khê với hệ số Nash là 87,01%.

5. Kết luận

Nghiên cứu này đã xây dựng được chương trình

MDC-TUH trên ngôn ngữ lập trình Visual Basic 2012 dựa trên cơ sở lý thuyết của mô hình NAM và sử dụng thuật toán SCE để tối ưu các thông số của chương trình. Chương trình được áp dụng để tính toán cho lưu vực sông Ba tính đến trạm An Khê đều cho kết quả mô phỏng đạt hệ số Nash trên 87%. Tuy nhiên, chương trình cũng cần phải kiểm chứng với nhiều lưu vực khác nhau để kiểm tra tính đúng đắn, ổn định và hiệu quả của chương trình.

Tài liệu tham khảo

1. Thian Yew Gan và Getu Fana Biftu. *Effects of Model Complexity and Structure, Parameter Interactions and Data on Watershed Modeling* (2003).
2. Duan, Q., Sorooshian, S., Gupta, *Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour and Research*, 28 (1992) 1015.
3. Wang, Y. C., Yu, P. S. & Yang, T. C. *Comparison of genetic algorithms and shuffled complex evolution approach for calibrating distributed rainfall-runoff model. Hydrological Processes* 24, 1015–1026.
4. Kouk King Kouk, *Parameter Optimization methods for calibrating TANK model and Neural Network Model for Rainfall – Runoff Modeling*, (2010).
5. J.G.Ndiritu, *Automatic Calibration of Pitman model using the Shuffled Complex Evolution Method*, 2009.
6. Dong-mei Xu, Wen-chuan Wang, Kwok-wing Chau, Chun-tiang Cheng, *Comparison of three global optimization Algorithms for calibration of the Xinanjiang method parameters. Journal of Hydroinformatics*, (2013) 174.
7. Soon Thiam KHU. *Automatic Calibration of NAM model with Multi-Objectives Consideration* (1998).
8. H. Madsen. *Automatic calibration of a conceptual rainfall – runoff model using multiple objectives. Journal of hydrology*, 235 (2000) 276.
9. J. A. Nelder, R. Mead. *A simplex method for function minimization, Computer Journal*, 7 (1965) 308.
10. Price, W. L., *Global optimizationa lgorithmsf or a CAD workstation, J. Optira. Theory Appl.*, 55(1), 133-146, (1987).
11. Holland, J. H., *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975.
12. Sorooshian S, Duan Q Y, Gupta V K. *Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of Hydrology*, 158 (1994) 265.
13. Giha Lee, Yasuto Tachikawa, and Kaoru Takara, *Analysis Hydrogic Model Parameter Characteristics using Automatic Global Optimization Method*, (2006).