

Bài báo khoa học

Tự động dò tìm bộ thông số tối ưu của mô hình thủy văn HEC–HMS bằng thuật toán SCE–UA

Nguyễn Phước Sinh^{1,2*}, Nguyễn Trường Huy^{1,3}, Nguyễn Thế Hùng¹

¹ Khoa Xây dựng Công trình thủy, Đại học Bách Khoa, Đại học Đà Nẵng; phuocsinhbk@gmail.com; nthuy@dut.udn.vn; ngthung@dut.udn.vn.

² Công ty cổ phần Sông Ba, 573 Núi Thành, Hải Châu, Đà Nẵng; sinhnnp@songba.vn.

³ Department of Civil Engineering, McGill University, 817 Sherbrooke Street West, Montreal, Quebec H3A 2K6, Canada; huy.nguyen5@mail.mcgill.ca.

*Tác giả liên hệ: phuocsinhbk@gmail.com; Tel.: +84–905868028

Ban Biên tập nhận bài: 2/8/2022; Ngày phản biện xong: 5/9/2022; Ngày đăng bài: 25/9/2022

Tóm tắt: Việc hiệu chỉnh và kiểm định mô hình thủy văn HEC–HMS theo quy trình thử sai (trial–and–error) thường mất nhiều thời gian và bộ thông số tìm được thường không phải tối ưu. Bài báo này trình bày kết quả phát triển một chương trình cho phép dò tìm tự động bộ thông số tối ưu của mô hình HEC–HMS dựa trên thuật toán SCE–UA. Trước hết phương pháp Latin Hypercube Sampling được sử dụng để lấy giá trị tham biến rộng khắp không gian nghiệm một cách hiệu quả nhất. Sau đó, SCE–UA được sử dụng để dò tìm nghiệm tối ưu thông qua chọn lọc và tiến hóa dựa trên các giá trị mẫu ban đầu. Nghiệm tối ưu được chọn dựa trên mật thoả hiệp Pareto và đánh giá độ tin cậy bằng phương pháp GLUE. Chương trình đã và đang được ứng dụng cho hồ thủy điện Krông H’răng (tỉnh Đắk Lắk). Số liệu thực đo từ 18 trong số 33 trận lũ trong giai đoạn 2016–2021 được sử dụng để hiệu chỉnh và thu hẹp không gian nghiệm ban đầu giúp quá trình dò tìm được nhanh chóng hơn, đồng thời phân tích độ nhạy và xác định ba thông số chủ đạo Tp , CN , Tc nhằm giảm số lượng tham biến (từ 50 xuống còn 18 tham biến). Dựa trên kết quả này, chương trình được kiểm định với 5 trận lũ tiếp theo và cập nhật tự động trong dự báo theo thời gian thực trên 10 trận lũ còn lại. Chương trình đạt hiệu quả dự báo rất tốt đến bước thời gian $t + 4$ giờ, các chỉ số đánh giá đạt mức cao ($KGE > 0,8$; $VE < 10\%$) và kết quả luôn nằm trong vùng tin cậy $Q_{5\%}$ – $Q_{95\%}$.

Từ khóa: HEC–HMS; Dò tìm tự động thông số mô hình; SCE–UA; GLUE; Dự báo lũ thời gian thực; Krông H’răng.

1. Giới thiệu

Các mô hình toán thủy văn như HEC–HMS (Hydrologic Engineering Center–Hydrologic Modeling System) mô phỏng chu trình thủy văn theo cách đơn giản hoá các đặc tính vật lý của hiện tượng tự nhiên thông qua các (hệ) phương trình toán học [1]. Cụ thể, mô hình HEC–HMS mô phỏng mưa–dòng chảy trên lưu vực thông qua 4 thành phần chính: (i) mô hình mưa–dòng chảy (ii) mô hình nước ngầm (iii) mô hình tổn thất và (iv) mô hình truyền lũ trên sông. Bốn thành phần chính này được mô phỏng trên HEC–HMS thông qua các thông số mô hình. Trong khi một số thông số có thể đo đạc được (như diện tích lưu vực, độ dài sông...) thì số khác rất khó hoặc gần như không thể đo lường được (như Muskingum X , Snyder Tc ...). Những thông số này phải được hiệu chỉnh cho mỗi lưu vực khác nhau sao cho kết quả dòng chảy mô phỏng từ mô hình phù hợp với dòng chảy thực đo [2]. Quá trình hiệu chỉnh và kiểm định mô hình thường được thực hiện thủ công bằng phương pháp thử sai (trial–

and-error). Phụ thuộc vào kinh nghiệm của người làm mô hình mà việc thử sai có thể tiêu tốn nhiều hay ít nguồn lực [3]. Ngoài ra, bộ thông số tìm được thường không phải tối ưu và tiêu chí đánh giá sự phù hợp thường rất giới hạn hay chỉ đơn mục tiêu. Việc này thúc đẩy nhu cầu nghiên cứu, phát triển các thuật toán và chương trình máy tính cho phép hiệu chỉnh mô hình một cách tự động và dò tìm bộ thông số tối ưu theo đa mục tiêu nhằm tiết kiệm thời gian và mang lại hiệu quả cao. Điều này hoàn toàn phù hợp và có ý nghĩa lớn đối với nhu cầu dự báo và cảnh báo lũ theo thời gian thực, cũng như việc quản lý các công trình thủy lợi-thủy điện trong điều kiện các hiện tượng mưa lũ cực đoan ngày càng gia tăng [4].

Hiện nay có khá nhiều thuật toán tìm nghiệm tối ưu toàn cục khác nhau. Một số thuật toán thông dụng như thuật toán di truyền (*Genetic Algorithms*) phát triển bởi Michalewicz (1992), giải thuật tối ưu bầy đàn (*Particle Swarm Optimization*) đề xuất bởi Kennedy và Eberhart (1995), thuật toán tiến hóa xáo trộn phức hợp của Trường Đại học Arizona (*Shuffled Complex Evolution-University of Arizona*, thường viết tắt SCE-UA) tạo ra bởi Duan (1992). Đối với các bài toán thủy văn, các hệ phương trình thường ở dạng phi tuyến và bao gồm rất nhiều tham biến thì SCE-UA được xem là một trong những thuật toán mạnh mẽ và hiệu quả nhất cho việc hiệu chỉnh thông số mô hình [5-6]. SCE-UA đã được nhiều nghiên cứu áp dụng thành công với các mô hình thủy văn thông dụng như MIKE-NAM [7], SWAT [8], SWMM [9]. Một số nhà phát triển phần mềm cũng đã tích hợp sẵn SCE-UA vào chức năng dò tìm tự động thông số mô hình như SWAT-CUP hay MIKE-AUTOCAL.

Những phiên bản gần đây của mô hình HEC-HMS như 4.8 (01/2021), 4.9 (01/2022) được Trung tâm kỹ thuật thủy văn thuộc Quân đội Hoa Kỳ (USACE) trang bị một số thuật toán dò tìm như Univariate-Gradient, Nelder and Mead thông qua chức năng Trial Run [10] nhằm giúp người làm mô hình dò tìm bộ thông số mô hình nhanh và hiệu quả hơn. Tuy vậy việc sử dụng phải thông qua thiết lập thủ công, tốn nhiều thời gian. Một dạng sửa đổi của phương pháp Gauss-Marquardt-Levenberg dựa trên SCE-UA hiện ở trạng thái “được đề xuất là tiềm năng” trong một báo cáo nội bộ của USACE [11] và vẫn chưa được tích hợp vào bản chính thức. Số ít nghiên cứu ứng dụng phương pháp dò tìm tối ưu khác như Particle Swarm Optimization để thiết lập thủ tục dò tìm tối ưu bộ thông số mô hình cho HEC-HMS song hiệu quả chưa cao [12].

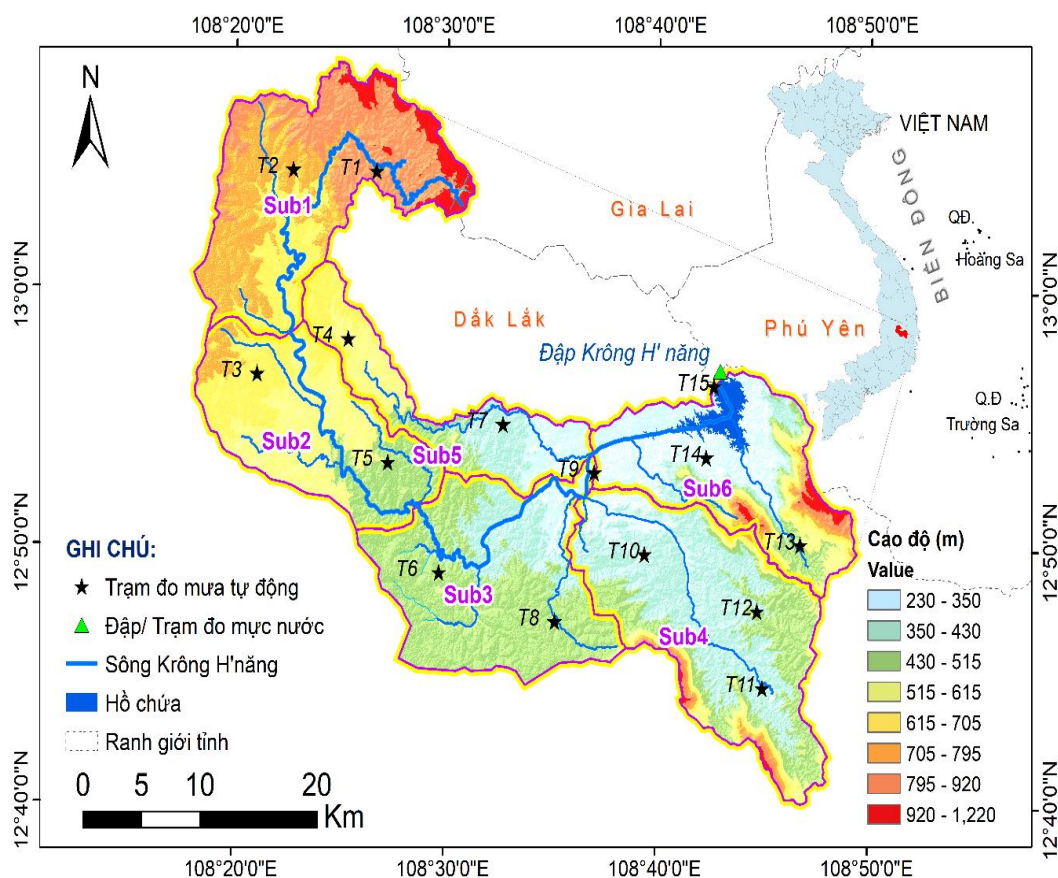
Bài báo này trình bày kết quả phát triển một chương trình cho phép dò tìm tự động bộ thông số tối ưu của mô hình HEC-HMS dựa trên thuật toán SCE-UA, nhằm cập nhật tự động bộ thông số mô hình trong dự báo đường quá trình lũ về hồ chứa theo thời gian thực. Hồ chứa được áp dụng trong nghiên cứu này là hồ Krông H’ năng với lưu vực có diện tích 1.168 km² nằm ở địa phận tỉnh Đắk Lắk (xem mục 2). Để thực hiện dò tìm, trước hết phương pháp lấy mẫu siêu khối Latin (*Latin Hypercube Sampling, LHS*) [13] được sử dụng để lấy mẫu giá trị tham biến rộng khắp không gian nghiệm một cách hiệu quả nhất. Sau đó thuật toán SCE-UA được sử dụng để dò tìm nghiệm tối ưu thông qua việc chọn lọc và tiến hóa dựa trên các giá trị lấy mẫu ban đầu. Thông thường bài toán là đa mục tiêu, nghiệm tối ưu do đó được chọn trên mặt thoả hiệp Pareto và đánh giá độ tin cậy bằng phương pháp ước tính độ tin cậy tổng quát (*Generalized Likelihood Uncertainty Estimation, GLUE*) [14] (xem mục 3). Dữ liệu lưu vực Krông H’ năng với 33 trận lũ quan trắc giai đoạn 2016–2021 được sử dụng để thử nghiệm, trong đó 18 trận lũ dùng để hiệu chỉnh, phân tích độ nhạy, xác định các thông số chủ đạo và không gian nghiệm ban đầu; 5 trận lũ tiếp theo sử dụng để kiểm định chương trình; 10 trận lũ còn lại để thử nghiệm hiệu suất chương trình khi dự báo lũ theo thời gian thực. Kết quả chi tiết và thảo luận trình bày ở mục 4. Các kết quả chính được tóm tắt lại ở phần cuối của bài báo (mục 5).

2. Phạm vi nghiên cứu và dữ liệu

Công trình thủy điện Krông H’ năng nằm trên địa bàn hai tỉnh Đắk Lắk và Phú Yên, trong đó toàn bộ lưu vực thuộc địa phận tỉnh Đắk Lắk (Hình 1). Hồ chứa có dung tích toàn bộ 165,78.10⁶ m³, đập được thiết kế với lũ thiết kế có tần suất $Q_{p0,1\%} = 8.234$ m³/s [15]. Sông

Krông H'ăng (hay Ea Krông H'ăng) là phụ lưu của sông Ba – lưu vực sông lớn nhất ở khu vực Nam Trung Bộ.

Hệ thống quan trắc và tính toán tự động mưa, mực nước, và lưu lượng về hồ thủy điện Krông H'ăng với tần suất 15 phút được thiết lập trên lưu vực từ tháng 9/2016. Tính đến hết tháng 12/2021, hệ thống đã quan trắc được 33 trận lũ với lưu lượng đỉnh từ 180 đến 2.710 m³/s. Lưu ý rằng đối với hồ Krông H'ăng, giá trị ngưỡng lưu lượng được xem là lũ về hồ không được quy định cụ thể trong hồ sơ thiết kế cũng như quy trình vận hành. Trong nghiên cứu này, nhóm tác giả sử dụng tất cả các trận mưa–lũ có thời gian mưa kéo dài hơn 12 giờ và có đồ thị lũ hoàn chỉnh (gồm các yếu tố: dòng chảy cơ bản, nhánh lũ lên, đỉnh lũ, nhánh lũ xuống). Đồ thị các trận lũ thực đo thể hiện trên Hình 2. Các trận lũ thông thường kéo dài trung bình khoảng 42 giờ và lên tới 102 giờ. Lưu lượng đỉnh lũ trung bình vào khoảng 850 m³/s và lớn nhất là 2.710 m³/s. Tổng lượng lũ trung bình khoảng 55 triệu m³ và lớn nhất 192 triệu m³. Một số thống kê khác về cường độ mưa–lũ được thể hiện chi tiết trong Bảng 1.

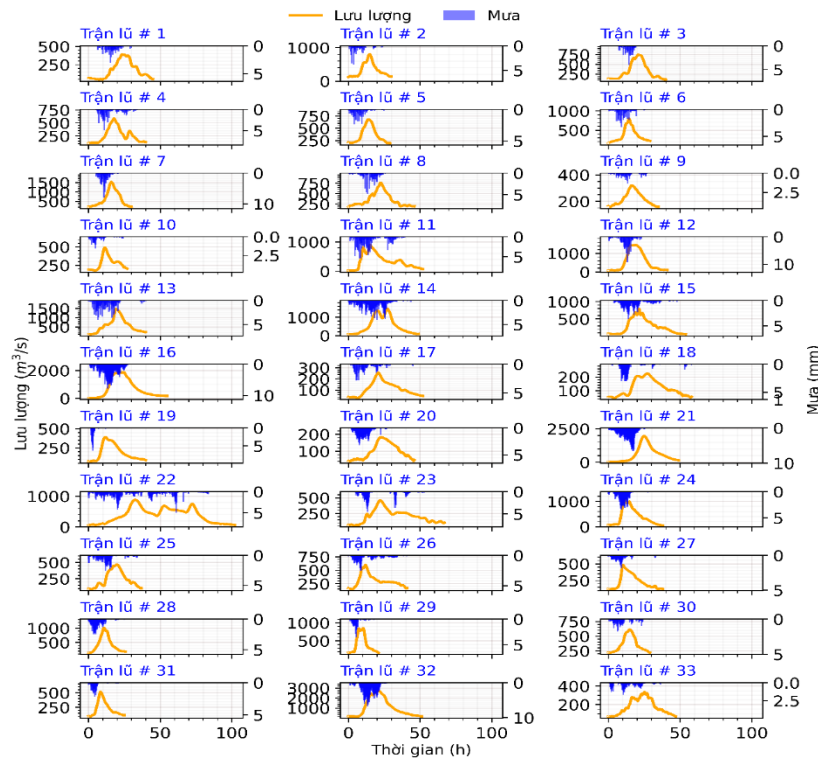


Hình 1. Bản đồ lưu vực hồ thủy điện Krông H'ăng. Các điểm hình sao hiển thị vị trí của 15 trạm đo mưa (T1 đến T15), điểm hình tam giác là vị trí đập. Dải màu (xanh đến đỏ) thể hiện cao độ địa hình lưu vực.

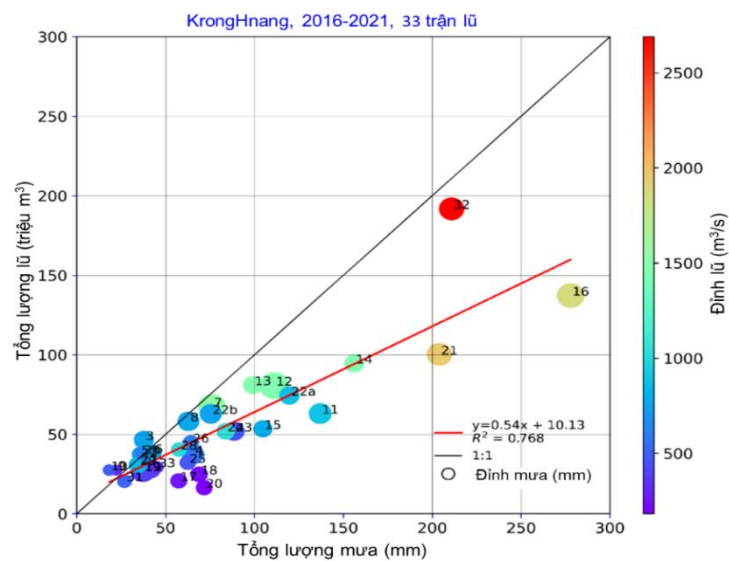
Để có thể hiểu sâu hơn đặc điểm mưa–lũ trên lưu vực Krông H'ăng, tương quan giữa tổng lượng mưa, tổng lượng lũ, lưu lượng đỉnh lũ và độ lớn đỉnh mưa được phân tích chi tiết trên Hình 3. Thông thường quan hệ giữa mưa–dòng chảy là phi tuyến vì phụ thuộc rất nhiều yếu tố liên quan đến đặc điểm lưu vực và khả năng thấm, trữ nước ở các tầng khác nhau của lưu vực. Lưu vực Krông H'ăng cũng không phải ngoại lệ. Tuy vậy, nếu xấp xỉ mối quan hệ tổng lượng mưa và tổng lượng lũ trên lưu vực Krông H'ăng bằng một tương quan tuyến tính, kết quả hệ số tương quan đạt mức tốt $R^2 = 0,768$. Từ đây có thể dùng mối quan hệ này để ước tính hay dự báo tổng lượng lũ khi tính được hay dự báo được tổng lượng mưa trên lưu vực.

Bảng 1. Thống kê cường độ mưa–lũ ở lưu vực Krông H'nh từ số liệu 33 trận mưa–lũ.

Thông số	Nhỏ nhất	Trung bình	Lớn nhất
Thời gian trận lũ (giờ)	21,0	42,0	102,0
Tổng lượng mưa (mm)	18,0	82,7	278,0
Cường độ mưa (mm/ giờ)	0,6	1,9	5,1
Thời gian xuất hiện đỉnh mưa (giờ)	3,0	9,9	20,0
Lưu lượng đỉnh lũ Q^p (m ³ /s)	182,3	849,0	2.710
Tổng lượng lũ (triệu m ³)	16,5	54,6	191,8
Thời gian xuất hiện đỉnh lũ (giờ)	8,5	18,3	32,8
Thời gian trễ giữa đỉnh mưa–đỉnh lũ (giờ)	1,0	8,0	15,5
Lưu lượng ban đầu Q_0 (m ³ /s)	20,0	110,0	221,0



Hình 2. Đồ thị 33 trận lũ thực đo trên lưu vực Krông H'nh trong giai đoạn 2016–2021.



Hình 3. Tương quan giữa tổng lượng mưa (trục x), tổng lượng lũ (trục y), đỉnh lũ (dải màu) và đỉnh mưa (kích thước điểm), đường thẳng màu đỏ thể hiện mối quan hệ tuyến tính giữa tổng lượng mưa và tổng lượng lũ, với hệ số tương quan chặt chẽ $R^2 = 0,768$.

3. Phương pháp nghiên cứu

3.1. Dò tìm bộ thông số tối ưu của mô hình HEC–HMS tự động bằng thuật toán SCE–UA

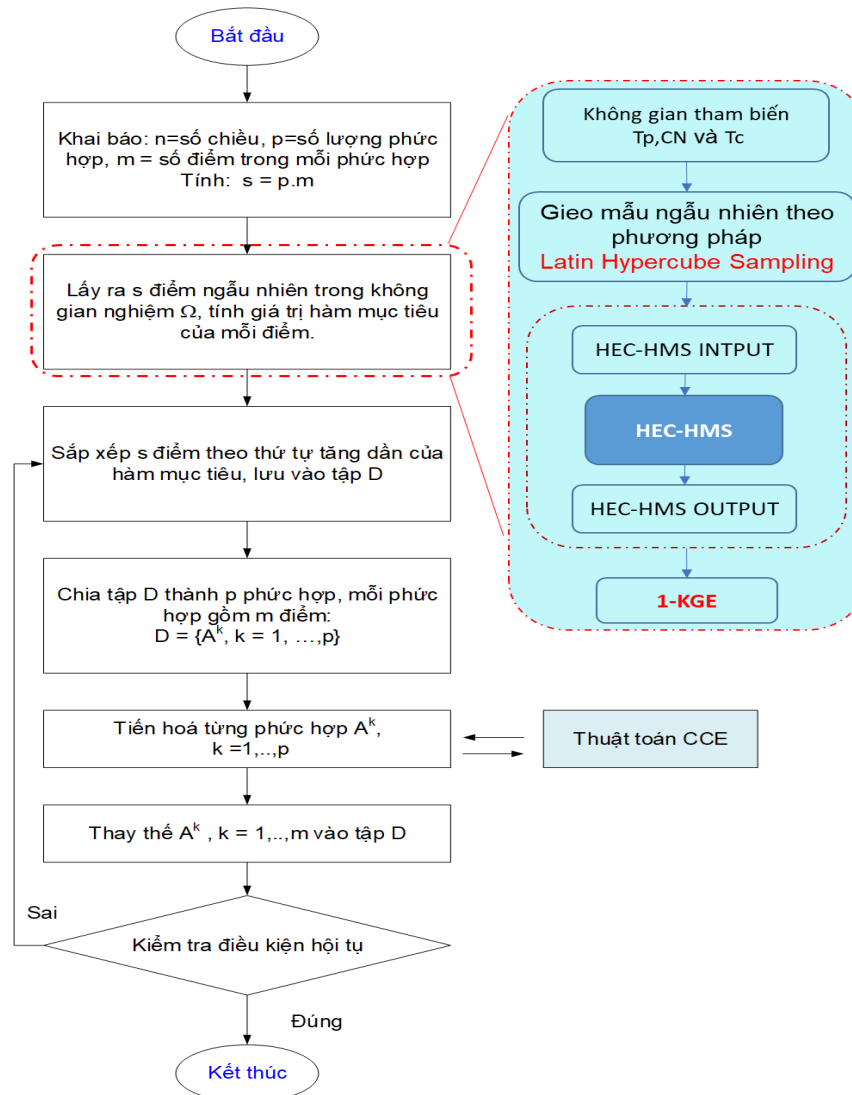
Để có thể dò tìm một cách tự động, trước hết cần chọn không gian khả nghiệm tham biến của mô hình HEC–HMS, nghĩa là xác định số lượng tham biến (số chiều n) và khoảng giá trị của chúng. Đối với mỗi tham biến, khoảng giá trị thích hợp được đưa ra trong phần hướng dẫn sử dụng mô hình HEC–HMS [10]. Khoảng giá trị của mỗi tham biến này khá rộng và cần được thu hẹp cho mỗi lưu vực nghiên cứu cụ thể dựa trên kinh nghiệm sử dụng mô hình và dữ liệu quan trắc trên lưu vực đó (Bảng 4). Việc thu hẹp không gian khả nghiệm giúp tiết kiệm khối lượng tính toán và làm giảm thời gian hội tụ một cách đáng kể vì số lượng tham biến thường rất lớn [9, 16].

Khi đã xác định được không gian khả nghiệm cho lưu vực nghiên cứu, chúng ta cần lấy s tổ hợp tham biến, với $s = p(2n + 1)$ phụ thuộc vào số lượng tham biến n và số lượng phức hợp (complexes) p , để có thể chạy mô hình HEC–HMS và so sánh kết quả mô phỏng dòng chảy từ s tổ hợp với giá trị lưu lượng thực đo để đánh giá chất lượng kết quả mô phỏng. Vì số lượng mẫu thường bị giới hạn, để tiết kiệm thời gian tính toán cần chọn một phương pháp giúp lấy mẫu một cách hiệu quả. Trong nghiên cứu này sử dụng phương pháp lấy mẫu siêu khối Latin (*Latin Hypercube Sampling, LHS*) do phương pháp này vượt trội hơn gieo mẫu ngẫu nhiên vì có xét đến vị trí các mẫu đã gieo [13].

Sau khi đã gieo mẫu (với s tổ hợp tham biến ban đầu), chương trình tự động HEC–HMS mô phỏng mưa–dòng chảy với bộ thông số mô hình theo s tổ hợp tham biến, sau đó tính toán giá trị hàm mục tiêu theo chỉ số đánh giá hiệu quả Kling–Gupta (*KGE*) tương ứng với mỗi tổ hợp. Thuật toán SCE–UA tiến hành sắp xếp s tổ hợp theo thứ tự giá trị *KGE* tăng dần và lưu trữ vào tập D sau đó chia tập D này thành p phức hợp A^k (với $k = 1, \dots, p$). Các phức hợp A^k được tiến hóa một cách độc lập theo phương pháp Downhill Simplex trong khối thuật toán Competitive Complex Evolution (*CCE*). Tiếp theo, SCE–UA loại bỏ các nghiệm xấu (có giá trị *KGE* nhỏ) và xáo trộn điểm giữa các phức hợp A^k đã tiến hoá thành một tập mẫu D mới. Sự tiến hoá và xáo trộn này sẽ được lặp lại đến khi điều kiện hội tụ được thỏa mãn hay đạt số lượng chu trình lặp (*repetitions*) do người dùng khai báo. Nhờ sự xáo trộn điểm giữa các phức hợp, thuật toán SCE–UA rất hiệu quả trong việc giải quyết hiện tượng mắc kẹt nghiệm ở các vùng cực trị địa phương (*local optima*). Nghiệm cuối cùng của thuật toán là đường mô phỏng lưu lượng Q_{opt} (*global optimum*) có chỉ số đánh giá *KGE* tốt nhất và tổ hợp tham biến tương ứng.

Chương trình tự động dò tìm bộ thông số tối ưu của mô hình HEC–HMS dựa trên thuật toán SCE–UA được phát triển theo sơ đồ giải thuật mô tả trên Hình 4. Nghiên cứu này sử dụng các thư viện mã nguồn mở như spotpy, hydrooval, smt. sampling_methods... và lập trình trên ngôn ngữ Python.

Để cải thiện hiệu suất dò tìm nghiệm tối ưu, các tác giả chia quá trình dò tìm tối ưu thành 2 giai đoạn (2 *phase*): giai đoạn 1 tìm kiếm thô với 1.000 chu trình lặp (*repetitions* = 1.000), nghiệm tìm được ở giai đoạn 1 dùng làm nghiệm ban đầu cho quá trình tìm kiếm chi tiết ở giai đoạn 2 với 5.000 chu trình lặp (*repetitions* = 5.000). Để tính đến sự không chắc chắn, một vùng mở rộng 25% giá trị khoảng nghiệm của tham biến về hai phía khi tìm kiếm giai đoạn 2 và giá trị giới hạn vật lý của tham biến trên HEC–HMS (Bảng 4) cũng được thêm vào để đảm bảo ý nghĩa của nghiệm tìm được.



Hình 4. Sơ đồ thuật toán chương trình dò tìm bộ thông số tối ưu HEC–HMS trên nền tảng thuật toán SCE–UA. Khối chương trình bao bởi đường nét đứt màu đỏ bên phải được phát triển mới để liên kết HEC–HMS vào thuật toán gốc của tác giả Duan (1992).

3.2. Phân tích độ tin cậy dự báo

SCE–UA sử dụng chiến lược tối ưu hóa toàn cục bộ thông số dựa trên hàm mục tiêu đơn biến, phương pháp này tìm cách xác định một bộ thông số “tốt nhất” mà bỏ qua những điểm không chắc chắn từ các nguồn liên quan đến dữ liệu, cấu trúc mô hình và sự không chắc chắn của các thông số. Khi không xét đến sai số do quan trắc dữ liệu (mưa, lưu lượng) thì sai số trong ước tính giá trị thông số là nguồn sai số lớn nhất có thể dẫn đến sai số lớn kết quả đầu ra mô hình [17–18]. Nghiên cứu này sử dụng phương pháp ước tính độ tin cậy tổng quát (*Generalized Likelihood Uncertainty Estimation, GLUE*) do Beven và Binley đề xuất năm 1992 [14] để phân tích độ tin cậy dự báo từ kết quả mô hình.

GLUE thực hiện quy trình mô phỏng Monte Carlo, mục tiêu của phương pháp GLUE là tìm ra một tập hợp “có thể chấp nhận được” của các tham số trong không gian nghiệm trên dữ liệu và điều kiện cho trước. Ở đây sử dụng kết quả mô phỏng HEC–HMS trong quá trình tiến hoá bằng SCE–UA (5000 lần ở giai đoạn 2). Dùng ngưỡng $KGE = 0,65$ để loại bỏ phần nghiệm xấu, giữ lại các nghiệm tốt để xác định hàm phân phối tích lũy (*Cumulative Distribution Function, CDF*).

Như vậy, kết quả dự báo của chương trình được đưa ra bởi giá trị nghiệm tối ưu Q_{opt} của SCE–UA và độ không chắc chắn từ CDF, được chọn ở mức độ tin cậy 5% đến 95%,

đồng thời thể hiện được phân phối xác suất (hậu nghiệm) của các tham biến chủ đạo (Tp , CN và Tc).

3.3. Tiêu chí đánh giá chất lượng của mô hình dự báo

Chất lượng mô hình dự báo được đánh giá từ sự sai khác giữa kết quả mô hình và số liệu quan trắc tại điểm nút kiểm soát thông qua một số chỉ tiêu như sau:

Sai lệch về đỉnh lũ (*peak error*, PE %):

$$PE = 100 \cdot \frac{Q_{sim}^p - Q_{obs}^p}{Q_{obs}^p} \quad (1)$$

Sai lệch về tổng lượng lũ (*volume error*, VE %):

$$VE = 100 \cdot \frac{Vol_{sim} - Vol_{obs}}{Vol_{obs}} \quad (2)$$

Chỉ số hiệu quả Kling–Gupta (*Kling–Gupta Efficiency*, KGE):

$$KGE = 1 - \sqrt{(R - 1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\overline{Q_{sim}}}{\overline{Q_{obs}}} - 1\right)^2} \quad (3)$$

với:

$$R = \frac{\sum_{i=1}^n (Q_{obs}^i - \overline{Q_{obs}}) \cdot (Q_{sim}^i - \overline{Q_{sim}})}{\sqrt{\sum_{i=1}^n (Q_{obs}^i - \overline{Q_{obs}})^2} \sqrt{\sum_{i=1}^n (Q_{sim}^i - \overline{Q_{sim}})^2}} \quad (4)$$

Sai số trung phương tương đối (*Relative Root Mean Square Error*, $RRMSE$):

$$RRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (Q_{obs}^i - Q_{sim}^i)^2}}{\overline{Q_{obs}}} \quad (5)$$

Trong đó Q (m^3/s) là lưu lượng lũ; $\overline{Q}$ (m^3/s) là giá trị lưu lượng trung bình; Q^p (m^3/s) là lưu lượng đỉnh lũ; σ (m^3/s) là độ lệch chuẩn của chuỗi lưu lượng; Vol (m^3) là tổng lượng lũ; n là tổng số điểm dữ liệu; các chỉ số dưới *sim* ký hiệu chuỗi mô phỏng và *obs* ký hiệu chuỗi thực đo. Một số nghiên cứu chỉ ra tiêu chí đánh giá chất lượng mô hình thông qua các chỉ số có thể tham khảo ở Bảng 2.

Vì không có chỉ số nào có thể xem xét tất cả đặc điểm của đường quá trình lũ (như thời gian xuất hiện đỉnh lũ, lưu lượng đỉnh lũ, hình dạng lũ và độ lớn trận lũ...), nên cần kết hợp các chỉ tiêu để đánh giá, tức chọn nghiệm theo hàm đa mục tiêu thay vì đơn mục tiêu để có thể thu được kết quả tốt nhất. Trong nghiên cứu này, hai hàm mục tiêu được dùng để chọn nghiệm tốt nhất của mô hình trên mặt thỏa hiệp Pareto là: (i) chỉ số hiệu quả KGE là lớn nhất và (ii) sai lệch tổng lượng lũ VE (%) là nhỏ nhất. Sở dĩ các tác giả lựa chọn chỉ số đánh giá tổng hợp KGE do bản thân chỉ số này bao gồm hệ số tương quan R , độ lệch chuẩn và giá trị trung bình (xem biểu thức (3) và (4)). Đối với sai lệch tổng lượng lũ VE (%), đây là một chỉ số quan trọng khi dự báo lũ cho các hồ chứa bởi nó ảnh hưởng rất lớn đến việc vận hành tích nước, xả lũ của hồ chứa.

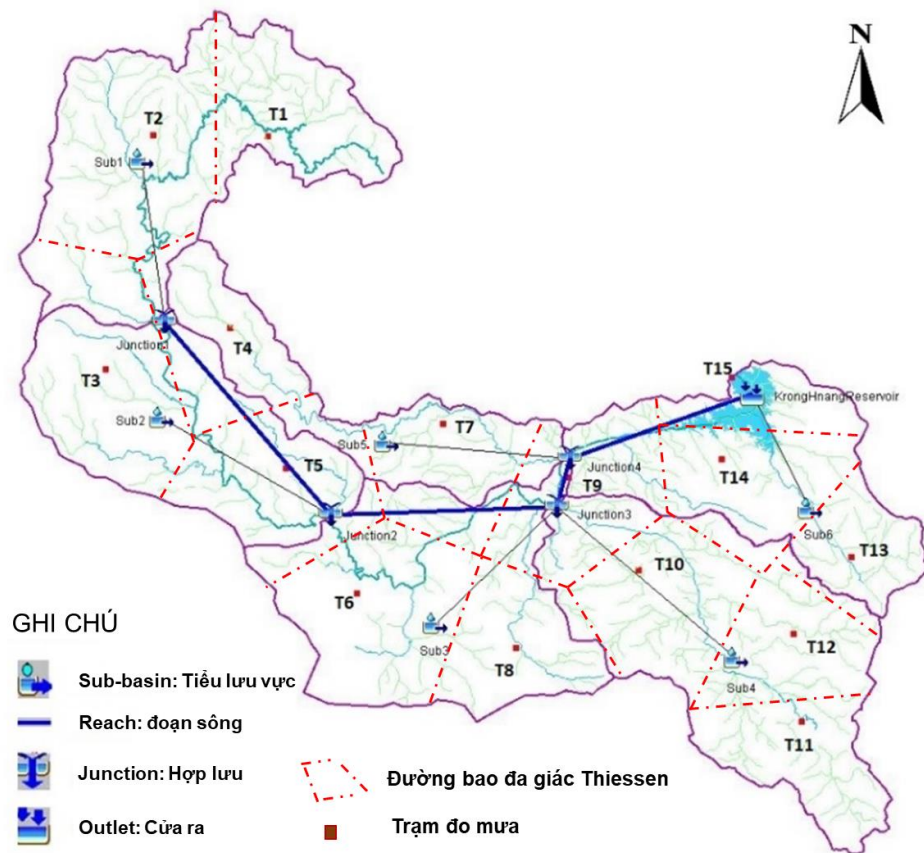
Bảng 2. Tham khảo các tiêu chí đánh giá độ tốt mô hình thông qua các chỉ số.

Chỉ số	Kém	Trung bình	Tốt	Rất tốt	Tham khảo
KGE	< 0,5	0,5 đến 0,7	0,7 đến 0,8	0,8 đến 1,0	[19-20]
VE (%)	> 20	10 đến 15	5 đến 10	0 đến 5	[21]
PE (%)	> 20	10 đến 15	5 đến 10	0 đến 5	[21]
R^2	< 0,4	0,4 đến 0,7	0,7 đến 0,85	0,85 đến 1,0	[22]
$RRMSE$	> 0,7	0,6 đến 0,7	0,5 đến 0,6	0 đến 0,5	[23, 24]

4. Kết quả và thảo luận

4.1. Thiết lập mô hình HEC-HMS cho lưu vực Krông H' năng

Lưu vực Krông H' năng với diện tích 1.168 km² được chia thành 6 tiểu lưu vực (từ Sub1 đến Sub6), sông chính dài 130 km, trong đó đoạn từ cửa ra tiểu lưu vực Sub1 đến hồ chứa được chia thành 4 đoạn sông (từ Reach1 đến Reach4) dựa trên đặc điểm địa hình và thủy hệ của lưu vực. Các thành phần tiểu lưu vực, đoạn sông kết nối với nhau bằng các hợp lưu (*Junction*) tạo thành hệ thống liên kết chặt chẽ từ thượng nguồn đến cửa ra lưu vực (*Outlet*). Mô hình lưu vực Krông H' năng được thiết lập trên HEC-HMS thể hiện ở Hình 5. Diện tích các tiểu lưu vực và chiều dài các đoạn sông là các thông số vật lý có thể đo đạc được, chi tiết ở Bảng 3. Trên lưu vực Krông H' năng có 15 trạm đo mưa tự động, số liệu mưa các trạm được tính chuyển về mưa bình quân các tiểu lưu vực (*sub-basin*) bằng phương pháp đa giác Thiessen.



Hình 5. Mô hình lưu vực Krông H' năng trên HEC-HMS với 6 tiểu lưu vực từ Sub 1 đến Sub 6. Các trạm đo mưa tự động được ký hiệu từ T1 đến T15. Trên hình đập Krông H' năng nằm phía hạ lưu của lưu vực gần trạm đo mưa T15. Nét đứt màu đỏ thể hiện các đường bao của đa giác Thiessen.

Bảng 3. Diện tích các tiểu lưu vực và chiều dài các đoạn sông chính.

Tiểu lưu vực	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6
Diện tích (km ²)	232,00	180,00	224,00	260,00	101,00	171,00
Đoạn sông	Reach1	Reach2	Reach3	Reach4		
Chiều dài (km)	25,97	25,88	2,96	15,05		

4.2. Không gian khả nghiệm

Khoảng giá trị các tham biến của mô hình HEC-HMS rất rộng do được USACE thiết kế để phù hợp với nhiều dạng lưu vực khác nhau (xem Bảng 4). Ngoài ra số lượng tham biến là

khá lớn với tổng cộng $n = 50$ tham biến (6 tiểu lưu vực $\times$ 7 tham biến + 4 đoạn sông $\times$ 2 tham biến). Do đó, việc dò tìm thường mất rất nhiều nguồn lực và thời gian. Để có thể dò tìm tự động một cách hiệu quả thì cần thu hẹp không gian khả nghiệm cho mỗi tham biến và cần giảm bớt số lượng tham biến (hay còn gọi là số chiều không gian). Trong mục này các tác giả trình bày việc giới hạn không gian khả nghiệm thông qua việc hiệu chỉnh thủ công để dò tìm khoảng giá trị phù hợp cho từng lưu vực, từ đó thu hẹp lại khoảng không gian khả nghiệm của mỗi tham biến. Việc giảm số lượng tham biến thông qua phân tích độ nhạy được trình bày ở mục tiếp theo.

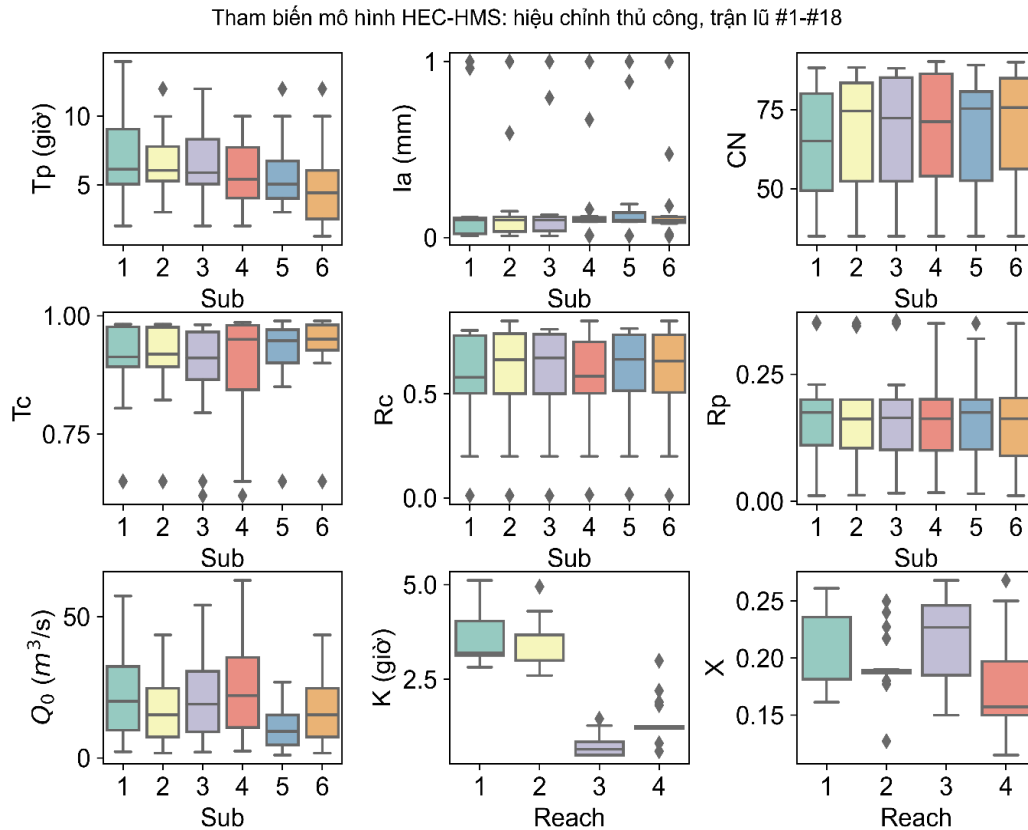
Bảng 4. Các tham biến chính, ý nghĩa và phạm vi giá trị của chúng trên HEC–HMS và cho lưu vực Krông H'nh từ kết quả hiệu chỉnh thủ công 18 trận lũ.

Quá trình	Phương pháp	Tham biến	Ý nghĩa	Giá trị trên HEC–HMS [10]	Giá trị lưu vực Krông H'nh
Mưa – dòng chảy	Snyder UH	T_p (giờ)	Thời gian tập trung nước của lưu vực	0 – 500	0,5 – 16,0
		T_c	Hệ số điều chỉnh độ lớn đỉnh và hình dạng lũ	0 – 1,0	0,3 – 0,95
Nước ngầm	Recession	Q_0 (m ³ /s)	Lưu lượng ban đầu	0 – 100.000	Thực đo
		R_c	Hệ số suy giảm nước ngầm	0 – 1,0	0,2 – 0,85
		R_p	Hệ số xác định độ lớn nước ngầm theo đỉnh lũ	0 – 1,0	0,1 – 0,45
Thấm	SCS Curve Number	I_a (mm)	Tổn thất ban đầu	0 – 500	0 – 20
		CN	Số hiệu đường cong	0 – 100	35 – 88
Truyền lũ	Muskingum	K (giờ)	Thời gian truyền lũ	0 – 150	0,5 – 6,0
		X	Hệ số điều chỉnh	0 – 0,5	0,05 – 0,28

Số liệu 18 trận lũ (từ #1 đến #18) được dùng để tìm khoảng giới hạn cho mỗi tham biến của bộ thông số mô hình HEC–HMS (T_p , T_c , I_a , CN , ...) cho riêng từng trận theo phương pháp thử–sai và lựa chọn nghiệm dựa trên các tiêu chí đánh giá mô hình. Việc hiệu chỉnh thủ công mặt khác giúp người sử dụng mô hình hiểu được độ nhạy của các tham biến khác nhau. Giới hạn không gian tham biến của mô hình HEC–HMS cho lưu vực Krông H'nh dựa trên việc hiệu chỉnh 18 trận lũ được thể hiện trên Hình 6. Tổng hợp các chỉ số đánh giá độ tốt mô hình ở Bảng 5 cho thấy mô hình mô phỏng tốt ở hầu hết các chỉ số. Chỉ số đánh giá hiệu quả Kling–Gupta thấp nhất đạt được $KGE = 0,66$. Cá biệt, các trận lũ số 2, 6, 11, 17 có một chỉ số không đạt mức tốt (được in đậm) trong khi các chỉ số khác đều đạt. Điều này khá là bình thường khi không thể cùng một lúc thỏa mãn hết được tất cả các tiêu chí đặt ra.

Bảng 5. Chỉ số đánh giá độ tốt mô hình HEC–HMS từ hiệu chỉnh thủ công 18 trận lũ (từ #1 đến #18).

TT	KGE	R ²	RRMSE	PE %	VE %	TT	KGE	R ²	RRMSE	PE%	VE%
1	0,868	0,957	0,189	7,48	10,50	10	0,849	0,820	0,158	–2,97	4,78
2	0,815	0,856	0,253	–19,47	8,89	11	0,776	0,889	0,293	16,12	5,56
3	0,934	0,966	0,138	–1,40	–6,03	12	0,936	0,980	0,146	5,93	5,72
4	0,872	0,808	0,257	–0,30	7,52	13	0,929	0,935	0,190	–8,16	4,90
5	0,925	0,953	0,104	–8,16	2,61	14	0,915	0,949	0,213	–0,79	6,53
6	0,660	0,882	0,285	2,66	14,27	15	0,947	0,947	0,183	4,54	–1,72
7	0,905	0,959	0,159	–2,78	–6,04	16	0,977	0,988	0,106	4,20	–0,54
8	0,744	0,920	0,191	–0,67	–0,63	17	0,934	0,945	0,142	–11,19	4,16
9	0,958	0,972	0,054	–0,25	–3,64	18	0,875	0,943	0,139	9,53	1,34



Hình 6. Không gian tham biến mô hình HEC-HMS lưu vực Krông H' năng từ kết quả hiệu chỉnh thủ công 18 trận lũ (từ #1 đến #18).

4.3. Phân tích độ nhạy các tham biến

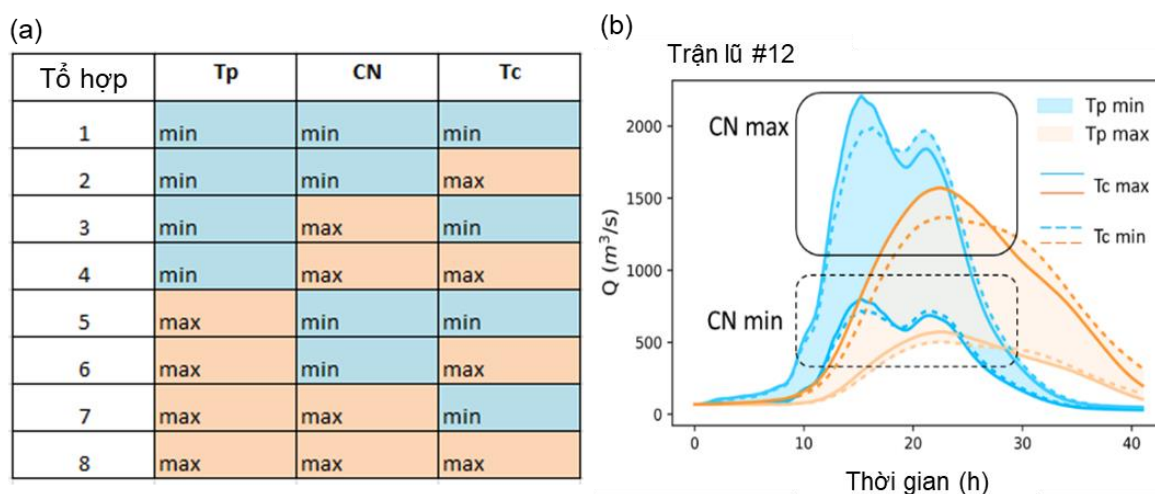
Mô hình HEC-HMS có tất cả 9 loại tham biến khác nhau như trình bày ở mục 4.2. Việc hiệu chỉnh thủ công cho thấy 5 trong số 9 loại tham biến nhạy hơn rất nhiều so với các loại tham biến còn lại. Năm loại tham biến này bao gồm: độ trễ tiêu chuẩn (*standard lag*) T_p và hệ số điều chỉnh (*coefficient*) T_c trong phương pháp đường đơn vị Snyder; số hiệu đường cong (*curver number*) CN và tổn thất ban đầu (*initial loss*) I_a trong phương pháp SCS Curver Number; và thời gian truyền lũ (*travel time*) K trong phương pháp truyền lũ Muskingum. Nhóm tác giả phân tích độ nhạy và đánh giá mức độ ảnh hưởng của mỗi loại tham biến thông qua 2 cực trị *min* và *max* trong khoảng khả nghiệm đến kết quả mô phỏng bằng cách thực hiện hết tất cả mô phỏng từ các tổ hợp có thể. Ví dụ, đối với bộ 3 tham biến nhạy nhất T_p , CN và T_c (dựa trên kinh nghiệm hiệu chỉnh thủ công) sẽ tạo ra tất cả $2^3 = 8$ tổ hợp tham biến (Hình 7a). Kết quả mô phỏng của 8 tổ hợp tham biến được trình bày ở Hình 7b.

Kết quả phân tích cho thấy khi T_p chuyển từ *min* sang *max* làm thay đổi hình dạng lũ một cách rõ rệt. T_p nhỏ lũ lên nhanh và xuống nhanh với đỉnh lũ nhọn trong khi T_p lớn làm đỉnh lũ xuất hiện chậm hơn hẳn, lũ lên chậm, đỉnh lũ bẹt ra và đi xuống chậm hơn. Nếu giữ nguyên T_p và thay đổi CN từ *min* sang max , tham biến này ảnh hưởng đến lưu lượng đỉnh lũ rõ rệt. CN_{max} cho giá trị đỉnh lũ lớn trong khi CN_{min} làm giảm đỉnh lũ do đó quyết định độ lớn tổng lượng lũ. Đối với tham biến T_c , không ảnh hưởng lớn như T_p và CN , nhưng T_c cũng gây ảnh hưởng đến độ lớn và hình dạng đỉnh lũ (tương quan với hình dạng đồ thị mưa).

Tương tự, nhóm tác giả thực hiện mô phỏng với $2^5 = 32$ tổ hợp (khi xét thêm cả tham biến K và I_a). Kết quả cho thấy tham biến Muskingum K của các đoạn sông ảnh hưởng đến thời điểm xuất hiện đỉnh lũ, khi kết hợp với T_p làm thay đổi hình dạng và độ lớn lũ, tuy

vậy phạm vi thay đổi giá trị K của các đoạn sông là khá nhỏ (Hình 6) nên nó ảnh hưởng nhỏ đến kết quả mô phỏng. Đối với tổn thất ban đầu Ia của các tiểu lưu vực chỉ ảnh hưởng đến phạm vi chân lũ, Hình 6 cho thấy 90% khoảng giá trị của Ia khá nhỏ chỉ vào khoảng 0,05 đến 0,28 mm. Lý do là khi vào mùa lũ chính vụ, thông thường đất đã gần đạt trạng thái bão hòa do các trận mưa liên tiếp trước đó nên tổn thất ban đầu Ia thường nhỏ. Do đó, tham số Ia trở nên ít ảnh hưởng. Điều này khác biệt hoàn toàn với bài toán mô phỏng liên tục dòng chảy nhiều năm khi tổn thất ban đầu Ia đóng một vai trò đặc biệt quan trọng.

Từ những kết quả này, các tác giả chọn giữ lại 3 loại tham biến chủ đạo là Tp , CN , và Tc . Việc này giúp giảm đáng kể số chiều của không gian tham biến từ $n = 50$ xuống còn $n = 18$. Đây là kết quả quan trọng giúp giảm đáng kể việc lấy mẫu ban đầu và giúp đẩy nhanh tốc độ dò tìm bộ thông số mô hình tối ưu thông qua SCE-UA. Ngoài 3 tham biến chủ đạo, tham biến lưu lượng ban đầu Q_0 được lấy theo giá trị quan trắc tại thời điểm bắt đầu mô phỏng, 5 tham biến còn lại (Ia , Rp , Rc , K , X) lấy bằng giá trị trung vị (*median*) từ hiệu chỉnh thủ công 18 trận lũ.

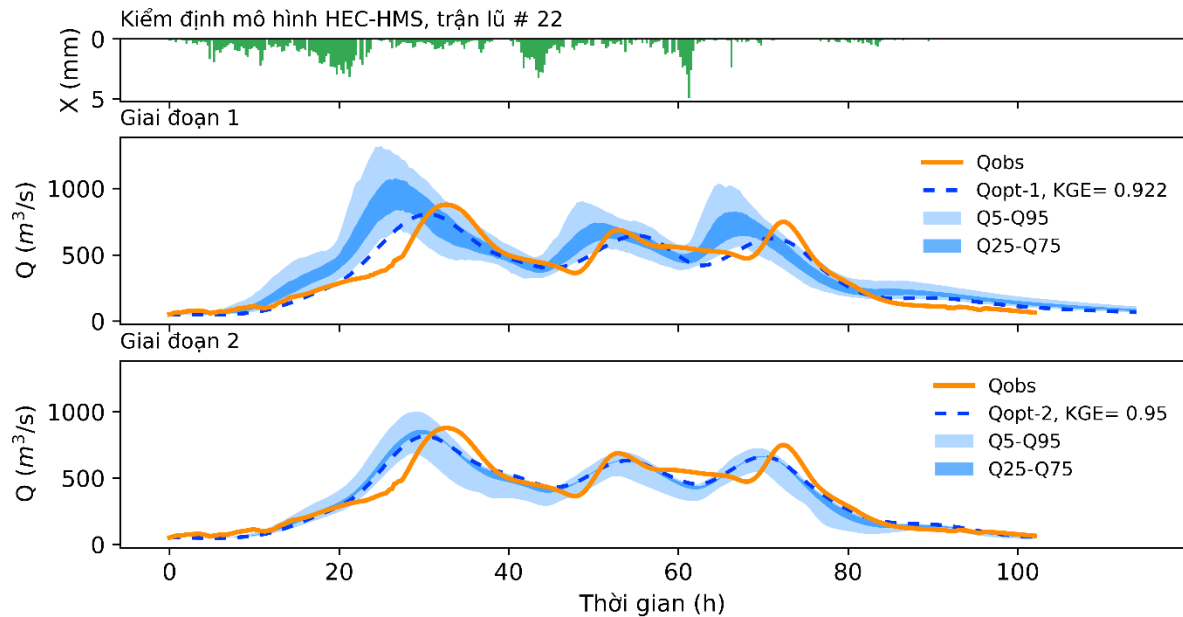


Hình 7. Đánh giá ảnh hưởng và độ nhạy các tham biến chủ đạo đến kết quả mô phỏng, minh họa trên biểu đồ trận lũ số 12.

4.4. Kết quả kiểm định chương trình

Chương trình tự động dò tìm bộ thông số tối ưu của mô hình HEC-HMS được kiểm định trên dữ liệu của 5 trận lũ (từ #19 đến #23) để xem xét hiệu quả mô phỏng toàn trận lũ dựa trên không gian nghiệm ban đầu của bộ 3 tham biến chủ đạo Tp , CN và Tc có được từ hiệu chỉnh thủ công 18 trận lũ. Hình 8 dưới đây thể hiện các nghiệm của quá trình dò tìm tự động theo thuật toán SCE-UA, thực hiện qua hai giai đoạn cho trận lũ #22, đây là trận lũ có hình dạng phức tạp nhất trong số 33 trận lũ (Hình 2).

Đối với trận lũ #22, hiệu quả dò tìm đạt rất tốt với chỉ số KGE tăng từ giai đoạn 1 (Gđ1) sang giai đoạn 2 (Gđ2), độ hội tụ của hàm mục tiêu $1 - KGE$ tốt. Nghiệm của hai giai đoạn được chọn trên mặt thoả hiệp Pareto với hai hàm mục tiêu: (i) chỉ số hiệu quả KGE là lớn nhất (tương đương với $1 - KGE$ nhỏ nhất) và (ii) sai số tổng lượng lũ VE (%) nhỏ nhất. Kết quả mô phỏng tương ứng với bộ thông số tối ưu và vùng tin cậy $Q_{5\%}$ – $Q_{95\%}$ và $Q_{25\%}$ – $Q_{75\%}$ sau khi đã loại nghiệm xấu theo phân tích GLUE cho cả hai giai đoạn. Kết quả trên Hình 8 chỉ ra rằng vùng tin cậy được thu hẹp đáng kể và nghiệm tối ưu cũng được cải thiện từ $KGE = 0,92$ lên $KGE = 0,95$.



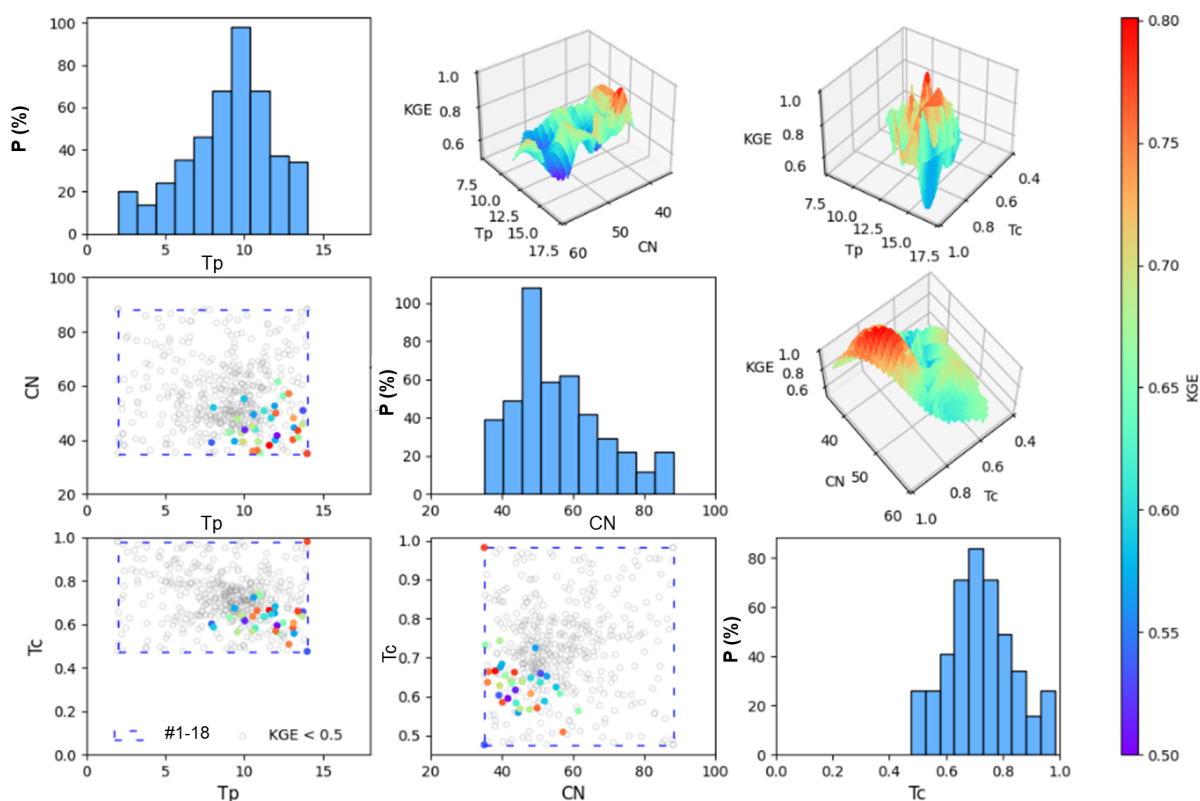
Hình 8. Kết quả dò tìm tự động thông số mô hình toàn trận lũ có hình dạng phức tạp như trận lũ #22 theo phương pháp SCE–UA, thực hiện qua 2 giai đoạn. Kết quả mô hình Q_{opt} nhận được từ SCE–UA và vùng tin cậy theo phương pháp GLUE. Hiệu quả dò tìm rất tốt với chỉ số KGE (Gđ1) = 0,92 tăng lên KGE (Gđ2) = 0,95 và vùng tin cậy cũng được thu hẹp đáng kể.

Ngoài ra, sau khi thực hiện phân tích GLUE, dạng phân bố xác suất (hậu nghiệm) của các tham biến cũng được xác định. Kết quả phân phối xác suất của 3 tham biến chủ đạo Tp , CN và Tc của tiểu lưu vực Sub1, từ phân tích GLUE trận lũ #22 thể hiện ở Hình 9. Trong đó, hai biểu đồ dạng phân bố điểm (*scatter*) thể hiện phân bố nghiệm từng đôi một ($Tp-CN$, $Tp-Tc$, $CN-Tc$) trong không gian hai chiều (2D) và ba biểu đồ dạng không gian ba chiều (3D) ở phía đối diện đường chéo thể hiện thêm chỉ số đánh giá mô hình KGE (có giá trị từ 0,5 đến 0,8 tương ứng dải màu từ tím đến đỏ). Biểu đồ dạng 3D minh họa rõ hơn các cực trị địa phương và cực trị toàn cục trong không gian nghiệm của các tham biến. Ba biểu đồ dạng cột (*histogram*) thể hiện phân bố xác suất của từng tham biến Tp , CN và Tc .

Quá trình này được thực hiện tương tự cho các trận lũ khác. Tổng hợp các chỉ số đánh giá kết quả kiểm định chương trình tự động dò tìm bộ thông số mô hình tối ưu cho 5 trận lũ qua hai giai đoạn thể hiện chi tiết ở Bảng 6.

Bảng 6. Chỉ số đánh giá kết quả kiểm định 5 trận lũ (từ #19 đến #23) qua hai giai đoạn dò tìm tối ưu.

Trận lũ	KGE		R^2		RRMSE		PE %		VE%	
	Gđ1	Gđ2	Gđ1	Gđ2	Gđ1	Gđ2	Gđ1	Gđ2	Gđ1	Gđ2
19	0,918	0,987	0,959	0,974	0,141	0,089	5,48	1,17	7,12	-0,23
20	0,537	0,783	0,858	0,863	0,245	0,210	-19,53	-4,37	-13,36	-6,65
21	0,935	0,985	0,958	0,970	0,200	0,172	17,49	9,29	-3,89	-0,15
22	0,922	0,950	0,900	0,904	0,202	0,199	8,81	7,61	0,05	-0,03
23	0,886	0,882	0,797	0,781	0,246	0,237	8,73	13,46	-0,68	0,67

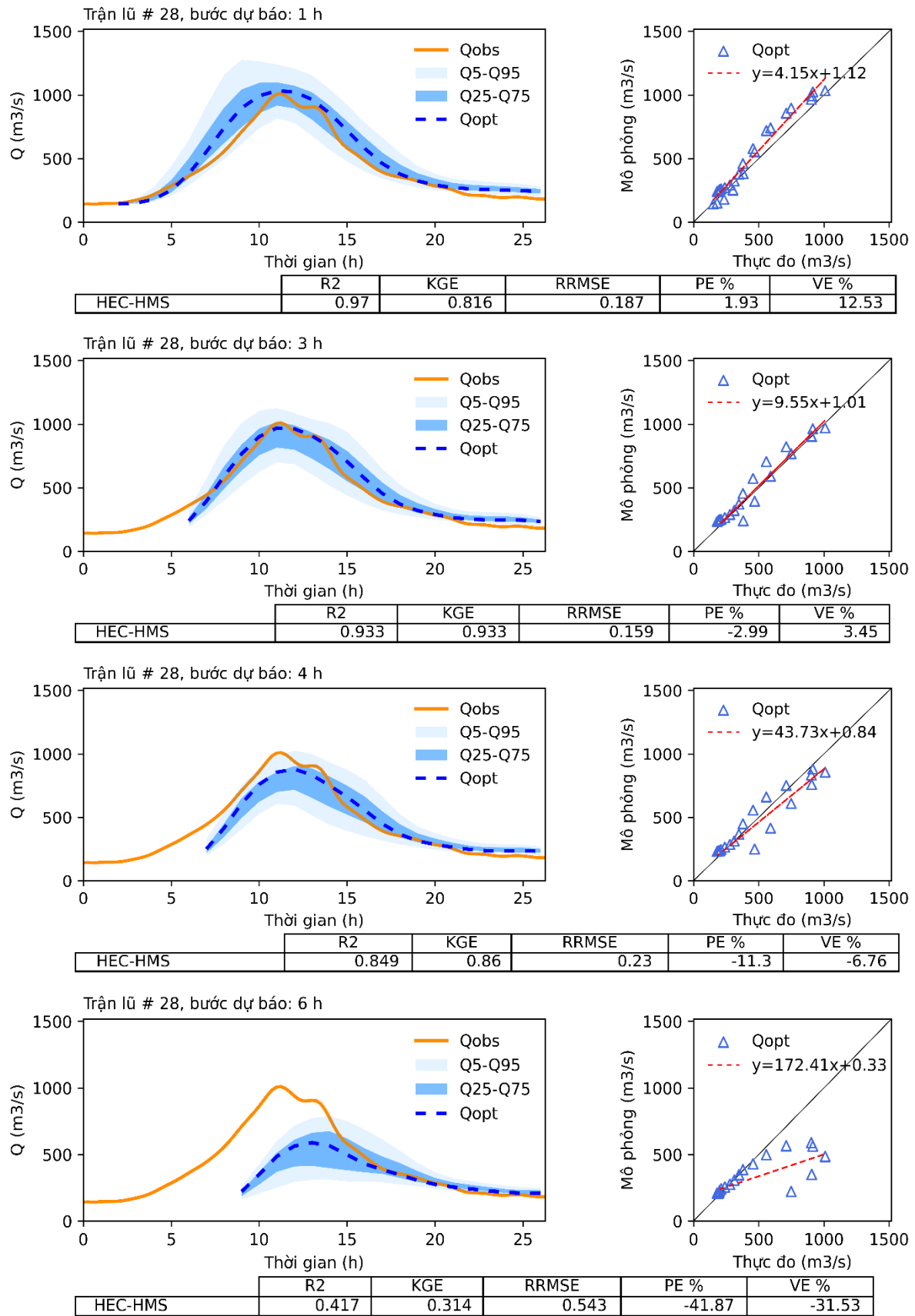


Hình 9. Phân phối xác suất hậu nghiệm của ba tham biến chủ đạo T_p , CN và T_c của tiểu lưu vực Sub1 từ phân tích GLUE trên kết quả giai đoạn 1 của trận lũ #22. Biểu đồ cột thể hiện phân bố xác suất của ba tham biến chủ đạo. Màu sắc thay đổi từ tím đến đỏ thể hiện tăng dần của chỉ số KGE. Biểu đồ dạng 3D thể hiện các cực trị địa phương và cực trị toàn cục trong không gian nghiệm của các tham biến.

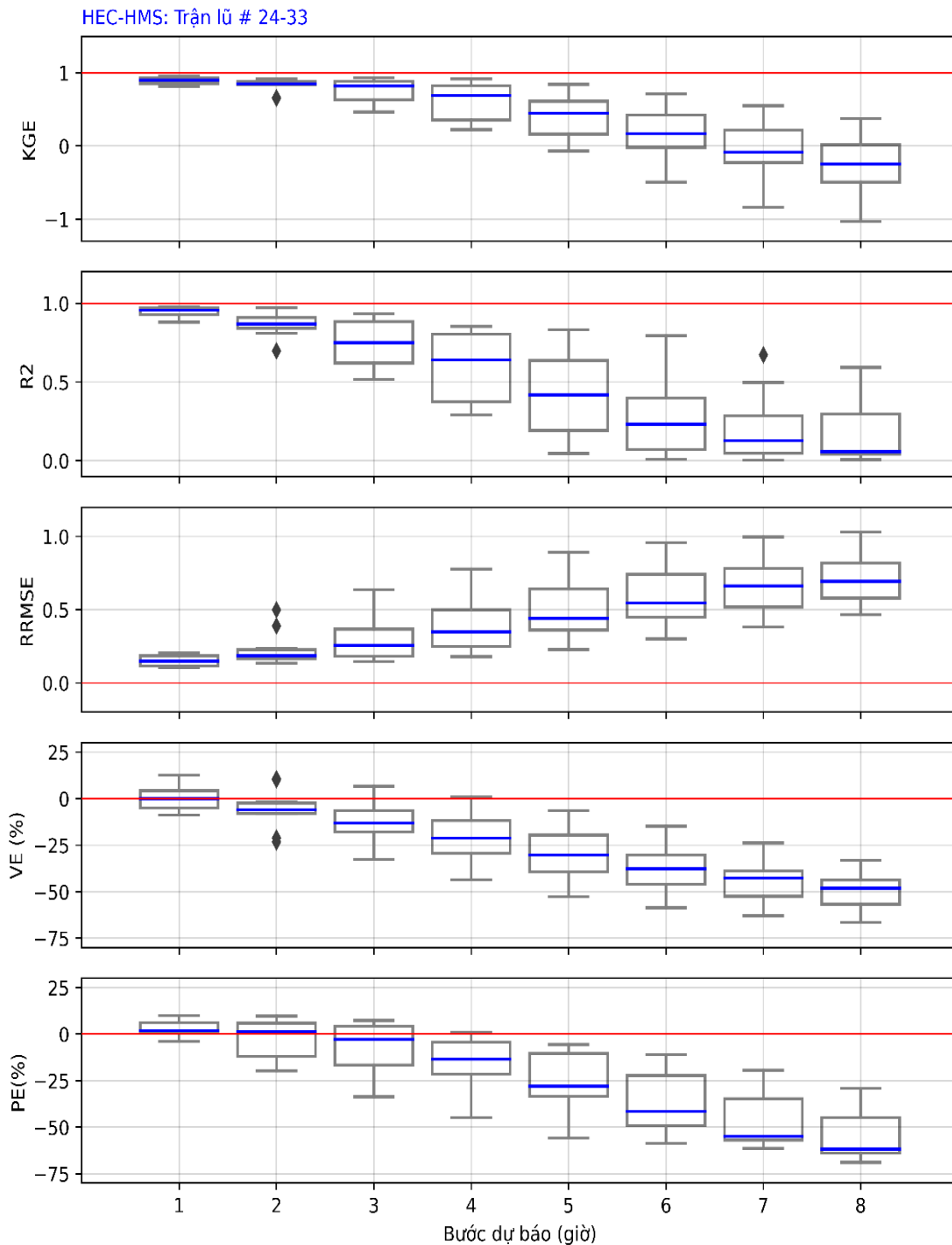
4.5. Kết quả thử nghiệm theo thời gian thực mùa lũ 2021

Dựa trên kết quả hiệu chỉnh và kiểm định, chương trình dò tìm giá trị tối ưu cho các tham biến và cập nhật tự động trong dự báo theo thời gian thực cho 10 trận lũ (từ số #24 đến #33) để đánh giá hiệu quả làm việc của chương trình. Hình 10 thể hiện kết quả dự báo ở các bước thời gian $t + 1$ giờ đến $t + 6$ giờ của trận lũ #28. Tổng hợp kết quả đánh giá chất lượng của mô hình dự báo ở các bước dự báo (*lead-time*) thể hiện chi tiết trên Hình 11, các chỉ số đạt mức rất tốt đến bước $t + 4$ giờ với $KGE > 0,85$; sai số tổng lượng lũ $VE < 10 \%$... và kết quả luôn nằm trong vùng tin cậy $Q_{5\%}$ – $Q_{95\%}$.

Từ thời gian dự báo dài hơn $t + 6$ giờ, mô hình HEC–HMS cho lưu vực Krông H' năng có độ tin cậy kém dần, kết quả này tương đồng với phân tích tương quan mưa–lũ. Lưu vực Krông H' năng với thời gian tập trung nước và truyền về cửa ra lưu vực trong 4 đến 6 giờ, sau khoảng này ($> t + 6$ giờ), do tạm thời mô hình chưa sử dụng thông tin từ mưa dự báo nên độ chính xác và khoảng tin cậy của mô hình không cao. Để tăng thời gian dự báo (tức dự báo sớm hơn), có thể tích hợp thêm thông tin mưa từ các nguồn như mưa radar hay kết quả mưa dự báo của các mô hình số trị toàn cầu, hiệu chỉnh với số liệu đo mưa mặt đất trên lưu vực để đảm bảo độ tin cậy mưa dự báo, từ đó làm đầu vào bổ sung cho mô hình HEC–HMS, đây là hướng giải quyết vấn đề trong nghiên cứu tiếp theo.



Hình 10. Kết quả thử nghiệm dự báo lũ trận lũ #28 ở các bước dự báo (lead-time) từ $t + 1$ đến $t + 6$ giờ. Trên hình bên trái, đường liền màu cam là lưu lượng thực đo Q_{obs} , đường nét đứt màu xanh là nghiệm tối ưu Q_{opt} . Vùng màu thể hiện các khoảng tin cậy $Q_{5\%}$ – $Q_{95\%}$ và $Q_{25\%}$ – $Q_{75\%}$ của kết quả dự báo.



Hình 11. Tổng hợp các chỉ tiêu đánh giá sai số mô hình cho 10 trận lũ (từ #24 đến #33), ở các bước dự báo (*lead-time*) từ $t + 1$ đến $t + 8$ giờ với độ chính xác đạt rất tốt trong khoảng $t + 4$ giờ trở lại ($KGE > 0,8$). Từ $t + 6$ giờ chất lượng dự báo kém dần. Đường màu đỏ thể hiện giá trị tốt nhất của chỉ số đánh giá.

5. Kết luận

Nghiên cứu này đã xây dựng mới chương trình tự động hoá dò tìm bộ thông số tối ưu của mô hình HEC-HMS trên nền tảng thuật toán SCE-UA, một trong những thuật toán dò tìm tối ưu mạnh nhất hiện nay, thay thế cho thủ tục dò tìm thủ công, không phụ thuộc và tránh sai sót chủ quan của người dùng, phù hợp cho việc dự báo lũ theo thời gian thực.

Dữ liệu mưa-lũ trên lưu vực hồ thủy điện Krông H'nh (tỉnh Đắk Lắk) với 33 trận lũ quan sát từ tháng 9/2016 đến tháng 12/2021 đã được sử dụng để hiệu chỉnh, kiểm định và đánh giá độ chính xác của mô hình HEC-HMS. Cụ thể, số liệu thực đo 18 trận lũ được sử dụng nhằm (i) tìm ra không gian khả nghiệm của các tham biến và (ii) giảm bớt số lượng tham biến từ việc phân tích độ nhạy của các loại tham biến khác nhau. Ba loại tham biến chủ đạo Tp , CN và Tc được xác định, việc này giúp giảm đáng kể số chiều không gian tìm kiếm từ 50 xuống còn 18 chiều. Dựa trên kết quả này, chương trình được kiểm định trên dữ liệu của 5 trận lũ (từ #19 đến #23) và cập nhật tự động trong dự báo theo thời gian thực cho 10 trận lũ (từ #24 đến #33) về hồ chứa Krông H'nh. Kết quả cho thấy chương trình có khả năng dự báo rất tốt đến bước thời gian $t + 4$ giờ. Các chỉ số đánh giá đạt mức cao ($KGE > 0,85$; sai số tổng lượng lũ $VE < 10 \%$) và kết quả luôn nằm trong vùng tin cậy $Q_{5\%}$ – $Q_{95\%}$. Bên cạnh đó, nghiên cứu cũng chỉ ra rằng cần bổ sung thêm thông tin mưa dự báo từ các nguồn như radar hay kết quả mưa dự báo của các mô hình số trị toàn cầu đã được hiệu chỉnh với mưa thực đo trên lưu vực để tăng độ tin cậy ở các bước dự báo sớm hơn, mang lại sự chủ động, tin cậy và hiệu quả hơn cho các kế hoạch ứng phó với mưa lũ.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: N.P.S., N.T.H.; Lựa chọn phương pháp nghiên cứu: N.P.S., N.T.H.; Xử lý số liệu: N.P.S., N.T.H.; Viết bản thảo bài báo: N.P.S., N.T.H.; Chỉnh sửa bài báo: N.P.S., N.T.H., N.T.H.

Lời cảm ơn: Nguyễn Phước Sinh được tài trợ bởi Tập đoàn Vingroup – Công ty CP và hỗ trợ bởi chương trình học bổng đào tạo thạc sĩ, tiến sĩ trong nước của Quỹ Đổi mới sáng tạo Vingroup (VINIF), Viện Nghiên cứu Dữ liệu lớn (VinBigdata), mã số VINIF.2021.ThS.97. Các tác giả trân trọng cảm ơn Công ty Cổ phần Sông Ba đã cung cấp số liệu sử dụng trong nghiên cứu này.

Lời cam đoan: Tập thể tác giả cam đoan bài báo này là công trình nghiên cứu của tập thể tác giả, chưa được công bố ở đâu, không được sao chép từ những nghiên cứu trước đây; không có sự tranh chấp lợi ích trong nhóm tác giả.

Tài liệu tham khảo

1. USACE. Hydrologic Modeling System HEC-HMS Technical Reference Manual. US Army Corps of Engineers, 2000, no. Mach. 10/01/2021. Available online: [https://www.hec.usace.army.mil/software/hec-hms/documentation/HEC-HMS_Technical%20Reference%20Manual_\(CPD-74B\).pdf](https://www.hec.usace.army.mil/software/hec-hms/documentation/HEC-HMS_Technical%20Reference%20Manual_(CPD-74B).pdf)
2. Beven, K. Rainfall-runoff modelling: The Primer, 2nd Edition. John Wiley & Sons, Ltd, 2010, pp.18.
3. Che, D.; Mays, L.W. Development of an Optimization/Simulation Model for Real-Time Flood-Control Operation of River-Reservoirs Systems. *Water Resour. Manage.* **2015**, 29, 3987–4005. Doi: 10.1007/s11269-015-1041-8.
4. NOAA. Climate Extremes Index (CEI), Climate Services and Monitoring Division. U.S. 2021. 25/02/2022. Available online: <https://www.ncdc.noaa.gov/extremes/cei/definition>.
5. Naeini, M.R.; Analui, B.; Gupta, H.V.; Duan, Q.; Soroosliian, S. Three decades of the Shuffled Complex Evolution (SCE-UA) optimization algorithm: Review and applications. *Sci. Iran.* **2019**, 26, 2015–2031. Doi: 10.24200/sci.2019.21500.
6. Duan, Q.; Sorooshian, S.; Gupta, H.V. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* **1992**, 428, 1015–1031.
7. Zakermoshfegh, M.; Ghodsian, M.; Neishabouri, S.A.A.S.; Shakiba, M. River flow forecasting using neural networks and auto-calibrated NAM model with shuffled complex evolution. *J. Appl. Sci.* **2008**, 8, 1487–1494.
8. Eckhardt, K.; Arnold, J.G. Automatic calibration of a distributed catchment model. *J. Hydro.* **2001**, 251, 103–109. Doi:10.1016/S0022-1694(01)00429-2.

9. Lee, S.; Kang, T. Analysis of Constrained Optimization Problems by the SCE–UA with an Adaptive Penalty Function. *J. Comput. Civ. Eng.* **2016**, *30*, 04015035(1–11). Doi:10.1061/(ASCE)CP.1943–5487.0000493.
10. USACE. HEC–HMS User’s Manual. US Army Corps of Engineers Hydrologic Engineering Center, 2021, no. December. 15/12/2021. Available online: <https://www.hec.usace.army.mil/confluence/hmsdocs/hmsum/4.7/release-notes/v-4-7-0-release-notes>.
11. Skahill, B.E. Potential Improvements for HEC–HMS Automated Parameter Estimation. U.S. Army Engineer Research and Development Center, 2006, ERDC/CHL TR–06–13. 20/5/2022. Available online: <https://apps.dtic.mil/sti/pdfs/ADA455236.pdf>.
12. Kamali, B.; Mousavi, S.J. Automatic Calibration of HEC–HMS Model Using Multi–Objective Fuzzy Optimal Models. *Civ. Eng. Infrastruct. J.* **2014**, *47*, 1–12. Doi: 10.7508/CEIJ.2014.01.001.
13. McKay, M.D.; Beckman, R.J.; Conover, W.J. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*. **1979**, *21*, 239–245. Doi: 10.2307/1268522.
14. Beven, K.; Binley, A. The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Processes* **1992**, *6*, 279–298.
15. Quy trình vận hành liên hồ chứa sông Ba. Ban hành kèm theo quyết định số 878/QĐ–TTg ngày 17/8/2018 của Thủ tướng Chính phủ.
16. Duan, Q.; Sorooshian, S.; Gupta, V.K. Optimal use of the SCE–UA global optimization method for calibrating watershed models. *J. Hydrol.* **1994**, *158*, 265–284. Doi: 10.1016/0022-1694(94)90057-4.
17. Liu, Y.; Gupta, H.V. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water. Resour. Res.* **2007**, *43*, W07401. Doi: 10.1029/2006WR005756.
18. Madsen, H. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Adv. Water Resour.* **2003**, *26*, 205–216. Doi: 10.1016/S0309–1708(02)00092–1.
19. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. Doi: 10.1016/j.jhydrol.2009.08.003.
20. Andersson, J.C.M.; Arheimer, B.; Traoré, F.; Gustafsson, D.; Ali, A. Process refinements improve a hydrological model concept applied to the Niger River basin. *Hydrol. Processes* **2017**, *31*, 4540–4554. Doi: 10.1002/hyp.11376.
21. Quy định kỹ thuật đánh giá chất lượng dự báo, cảnh báo thủy văn. Ban hành theo Thông tư 42/2017/TT–BTNMT ngày 23/10/2017 của Bộ Tài nguyên và Môi trường.
22. Cameron, C. and Windmeijer, F. A. G. An R–squared measure of goodness of fit for some common nonlinear regression models. *J. Econom.* **1997**, *77*, 329–342. Doi: 10.1016/s0304-4076(96)01818-0.
23. Singh, J.; Knapp, H.V.; Arnold, J.G.; Demissie, M. Hydrological modeling of the Iroquois River watershed using HSPF and SWAT. *J. Am. Water. Resour. Assoc.* **2005**, *41*, 343–360. Doi: 10.1111/j.1752-1688.2005.tb03740.x.
24. Ritter, A.; Muñoz–Carpena, R. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness–of–fit assessments. *J. Hydrol.* **2013**, *480*, 33–45. Doi: 10.1016/j.jhydrol.2012.12.004.

Automatic calibration of HEC–HMS model using Shuffled Complex Evolution (SCE–UA) algorithm

Phuoc Sinh Nguyen^{1,2*}, Truong Huy Nguyen^{1,3}, The Hung Nguyen¹

¹ Faculty of Water Resources Engineering, University of Science and Technology–The University of Da Nang, Vietnam; phuocsinhbk@gmail.com; nthuy@dut.udn.vn; ngthung@dut.udn.vn

² Song Ba JSC, 573 Nui Thanh, Hai Chau, Da Nang; sinhnnp@songba.vn

³ Department of Civil Engineering, McGill University, 817 Sherbrooke Street West, Montreal, Quebec H3A 2K6, Canada; huy.nguyen5@mail.mcgill.ca

Abstract: Verification and calibration of the HEC–HMS hydrological model using a trial–and–error procedure usually costs modelers a lot of time and effort and more importantly, the set of parameters found is often not optimal. This paper presents the results of developing an automatic program that allows the automatic search of the optimal set of parameters of the HEC–HMS model based on the SCE–UA algorithm. First, the Latin Hypercube sampling method is used to efficiently obtain the parameter values widely across the feasible solution space. Then the SCE–UA method is used to search for the optimal solution through complex shuffling and evolution based on the initial sampling values from the Latin Hypercube Sampling method. The problem is usually multi–objective, so the optimal solution is therefore selected based on a Pareto front and evaluated for uncertainty by the GLUE method. The program has been and is being applied to the Krông H'nh hydropower reservoir in Dak Lak province. The authors use real data measured from 18 out of 33 floods observed in the period 2016–2021 to narrow the initial feasible solution space and to reduce the parameter dimensions (from 50 to 18) by identifying the three governing parameters Tp , CN , and Tc through sensitivity analysis. This helps to enhance the search speed and the convergence of the optimal solution. Based on this result, the program searches the optimal value for the parameters and updates them automatically in the real–time forecast. The results of the validation of the next 5 floods and testing of the remaining 10 floods give good results up to the time step $t + 4$ hours. The evaluation indicators are high ($KGE > 0.85$, volume error $VE < 10\%$) and the result is always in the confidence range $Q_{5\%}$ – $Q_{95\%}$.

Keywords: HEC–HMS; Auto calibration; SCE–UA; Real–time flood forecasting; Krông H'nh.