

ỨNG DỤNG KỸ THUẬT PHÂN RÃ KHÔNG GIAN MỤC TIÊU GIẢI BÀI TOÁN ƯỚC TÍNH THÔNG SỐ TRONG MÔ HÌNH THỦY VĂN THÔNG SỐ PHÂN BỐ

Bùi Đình Lập¹, Trần Hồng Thái², Phạm Thị Hương Lan³

Tóm tắt: Việc xác định được tập thông số tối ưu trong hệ thống mô hình thủy văn thông số phân bố là rất cần thiết nhằm giải quyết các bài toán thực tiễn trong lĩnh vực thủy văn và tài nguyên nước. Bài báo này trình bày kết quả nghiên cứu, ứng dụng kỹ thuật phân rã không gian mục tiêu nhằm tìm lời giải tối ưu toàn cục cho các thông số trong mô hình thủy văn phân bố. Các kết quả đạt được trên bài toán thử nghiệm đã cho thấy, kỹ thuật phân rã không gian mục tiêu dựa trên tập hướng tham chiếu đã giúp bài toán có khả năng tìm kiếm trên toàn bộ POF (front tối ưu Pareto), đảm bảo sự đa dạng về mẫu trong không gian mục tiêu.

Từ khóa: Tối ưu đa mục tiêu, mô hình thủy văn phân bố, ước tính thông số.

1. MỞ ĐẦU

Mô hình thủy văn thông số phân bố là một hệ thống có cấu trúc rất phức tạp, hầu hết các thành phần tham gia vào hệ thống như mưa, thấm, bốc hơi..., đều là các thành phần phi tuyến được biểu diễn bởi các phương trình vi phân chứa cả biến không gian và thời gian. Số lượng thông số cần xác định trong hệ thống là rất lớn và rất khó xác định. Đặc biệt bài toán ước tính thông số cho hệ thống sẽ trở nên phức tạp hơn và đòi hỏi khối lượng tính toán gia tăng theo cấp số nhân khi số chiều trong không gian thông số cần tìm và không gian mục tiêu gia tăng. Mặc dù vậy đối với bài toán ước tính thông số đơn mục tiêu (hàm mục tiêu vô hướng), người hiệu chỉnh hệ thống vẫn có thể tìm được tập thông số chấp nhận được dựa trên phương pháp thử sai (phương pháp đang sử dụng phổ biến ở nước ta hiện nay). Tuy nhiên chất lượng bộ thông số tìm được theo phương pháp thử sai lại phụ thuộc rất lớn vào mức độ hiểu biết về hệ thống thực và mô hình của người hiệu chỉnh, thường mang nặng tính chủ quan và chất lượng

thấp, đặc biệt là phải mất rất nhiều thời gian trong quá trình hiệu chỉnh.

Để giải bài toán ước tính tối ưu thông số trong mô hình thủy văn đã có rất nhiều công trình nghiên cứu ngoài nước được thực hiện trong hơn 2 thập kỷ vừa qua nhằm xây dựng được giải thuật tối ưu hoặc chỉ ra các phương pháp hiệu chỉnh mô hình tự động nhanh hơn và hiệu quả hơn, các giải thuật đạt được từ công trình điển hình có thể kể đến như: Giải thuật tiến hóa sáo trộn phức hợp của trường đại học Arizona (shuffled complex evolution-SCE-UA) (Duan, et al 1994), thuật toán MOCOM-UA (Yapo, et al 1998); thuật toán MOSCEM-UA (Vrugt, et al 2003). Các thành tựu đạt được trong lĩnh vực nghiên cứu, ứng dụng lý thuyết toán tối ưu vào bài toán thủy văn trong hơn 2 thập kỷ qua là rất lớn, tuy nhiên các báo cáo về kết quả nghiên cứu và ứng dụng từ các công trình nghiên cứu kể trên cho thấy vẫn còn tồn tại nhiều vấn đề cần phải tiếp tục nghiên cứu hoàn thiện để nâng cao hơn nữa hiệu quả của của thuật toán như: thời gian hiệu chỉnh còn rất lớn (đặc biệt là khi áp dụng cho mô hình thủy văn thông số phân bố); bài toán có xu hướng hội tụ nhanh khi số lượng thông số cần tìm lớn hoặc các hàm mục

¹ Trung tâm Dự báo khí tượng thủy văn quốc gia

² Tổng cục Khí tượng Thủy văn

³ Trường Đại học Thủy lợi

tiêu có tương quan cao; các thuật toán thường có hiệu quả không cao khi các mục tiêu hiệu chỉnh lớn;... Bài báo này trình bày kết quả nghiên cứu, ứng dụng kỹ thuật phân rã không gian mục tiêu nhằm tìm lời giải tối ưu toàn cục cho các thông số trong mô hình thủy văn thông số phân bố.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1 Nội dung phương pháp

a) Bài toán tối ưu thông số

Theo (Trần Hồng Thái, 2005) hệ thống cần được hiệu chỉnh phù hợp với thực tế thông qua số liệu quan trắc trước khi ứng dụng vào thực tế. Nếu chúng ta ký hiệu hệ thống là $F^{(k)}(t_j, w, p)$ và số liệu quan trắc là z_{kj} với $k \in M^w$ (M^w là không gian biến trạng thái, các biến mà ở đó số liệu quan trắc được sử dụng để ước lượng thông số), biến quan trắc z_{kj} có thể là mực nước hoặc lưu lượng có được tại các trạm đo theo thời gian t_j với $j=1, \dots, n_m$. Khi đó biến quan trắc có thể được mô tả như sau:

$$z_{kj} = F^{(k)}(t_j, w, X) + \epsilon_{kj}$$

Trong đó $w \in \mathbb{R}^{2 \times N}$ là không gian biến trạng thái của hệ thống; X là không gian véc tơ thông số của hệ thống; ϵ_{kj} là lỗi quan trắc với giả thiết là tuân theo luật phân phối chuẩn $\epsilon_{kj} \sim N(0, \sigma_{kj}^2)$, ở đây σ_{kj} là độ lệch chuẩn của biến quan trắc z_{kj} .

Bài toán ước tính thông số trong hệ thống mô phỏng quá trình mưa, dòng chảy được xác định như sau: xác định tập giá trị thông số $X(X_1, \dots, X_N)$ với điều kiện ràng buộc (2) để hàm tối ưu (1) đạt giá trị cực tiểu

$$\begin{cases} \min_{X, w} \|E(X)\|_2^2 = \min_{X, w} \sum_{k \in M^w} \sum_{j=1}^{n_m} \frac{(z_{kj} - F^{(k)}(t_j, w, X))^2}{\sigma_{kj}^2} & (1) \\ X_i^{\min} \leq X \leq X_i^{\max}, i = 1 \dots N & (2) \end{cases}$$

Mục tiêu của bài toán hiệu chỉnh thông số mô hình là tìm tập véc tơ thông số X sao cho chuỗi thời gian $E(X)$ đạt cực tiểu, trong thực tế chuỗi $E(X)$ thường được thay thế bằng một hoặc nhiều các chỉ tiêu thống kê (như chỉ số SLS, Nash,

PDIF, DRMS,...) được biết đến như là các hàm mục tiêu hiệu chỉnh ký hiệu $F(X)$. Đối với bài toán hiệu chỉnh thông số tối ưu đa mục tiêu (MOP), bài toán được mô tả như sau:

$$\min F(x) = (f_1(x), \dots, f_m(x))^T \quad (3)$$

$$\text{Sao cho: } \begin{cases} g_j(x) \geq 0, j = 1, \dots, J \\ h_k = 0, k = 1, \dots, K \\ x \in \Omega \end{cases}$$

Trong đó:

- J và K là số của bất phương trình và phương trình ràng buộc tương ứng
- $\Omega \subseteq \mathbb{R}^n$ là không gian thông số (không gian biến quyết định),
- $x = (x_1, \dots, x_n)^T \in \Omega$ là véc tơ thông số
- $F: \Omega \rightarrow \mathbb{R}^m$ là hàm ánh xạ từ không gian thông số Ω sang không gian Euclid m chiều.
- $\mathbb{R}^m$ gọi là không gian mục tiêu.

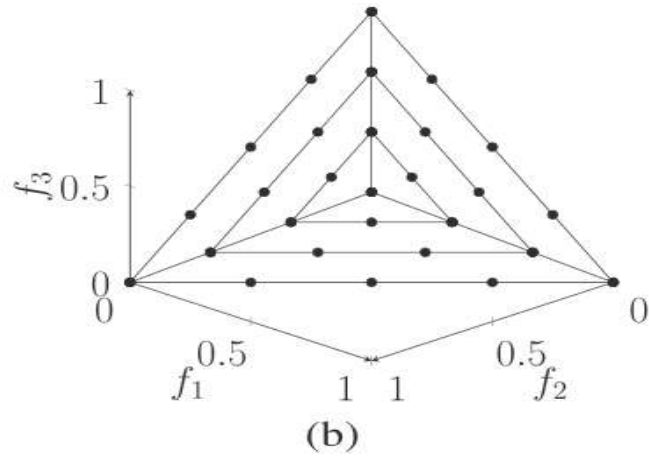
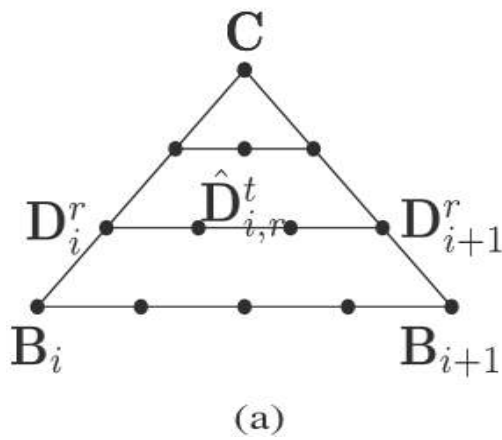
Lựa chọn tốt nhất hiện nay để giải bài toán trên là phát triển các thuật toán theo hướng tiến hóa đa mục tiêu (MOEAs). Khi các hàm mục tiêu mâu thuẫn nhau, sẽ không tồn tại bất kỳ một điểm nào trong không gian Ω để các hàm mục tiêu đồng thời đạt cực trị, cách duy nhất là phải cân bằng giữa các hàm mục tiêu, vấn đề này có thể được giải quyết bằng điều kiện tối ưu Pareto. Khi ấy tập lời giải của bài toán tối ưu đa mục tiêu sẽ chứa đựng tất cả các véc tơ quyết định mà ở đó các véc tơ mục tiêu tương ứng của chúng không thể cải thiện tốt hơn trong bất kỳ chiều nào mà không làm suy giảm một chiều khác.

b) Giải bài toán theo phương pháp phân rã không gian mục tiêu

Các tiếp cận để giải bài toán tối ưu đa mục tiêu trong bài báo này được thực hiện dựa trên nền tảng kết hợp các thành tựu đã đạt được của thuật toán tối ưu đơn mục tiêu trong lĩnh vực thủy văn và tài nguyên nước SCE-UA (Duan, et al 1994) và thuật toán tối ưu đa mục tiêu SPEA/R (Jiang, et al 2017). Không gian mục tiêu trong bài toán sẽ được chia thành các vùng nhỏ độc lập dựa trên tập hướng tham chiếu (W), cách làm này sẽ giúp bài

toán có khả năng tìm kiếm trên toàn bộ POF (front tối ưu Pareto), đảm bảo sự đa dạng về mẫu trong không gian mục tiêu. Việc xây dựng tập tham chiếu W được thực hiện như sau: Đầu tiên tâm hướng tham chiếu C được xác định qua số hàm mục tiêu M $C(1/M, \dots, 1/M)$, sau đó tọa độ của các điểm B_i (điểm cắt trục thứ i trong không gian mục tiêu) được xác định là $B_i = (b_{i1}, \dots, b_{iM})$ sao cho $b_{i1} = 1$ và $b_{ij} = 0$ cho tất cả $\{j \neq i \mid 1 \leq j \leq M, 1 \leq i \leq M\}$. Như vậy đơn vị đơn hình có thể được chia tới M đơn hình nhỏ hơn ký hiệu là $Simp(i)$, ở

đó biên của các $Simp(i)$ được xác định bởi các điểm C, B_i và B_{i+1} (xem Hình 1). Trên mỗi $Simp(i)$, các điểm trên cạnh CB_i và CB_{i+1} sẽ được sinh trước dựa theo công thức $D_i^r = C + \frac{r}{k}(B_i - C)$ trong đó k là số lớp từ đỉnh C tới cạnh B_iB_{i+1} , $r \in \{1, \dots, k\}$, sau đó các điểm trên các lớp được tạo ra theo công thức $\hat{D}_{i,r}^t = D_i^r + \frac{t}{r+1}(D_{i+1}^r - D_i^r)$ với $t \in \{1, \dots, r\}$, r là số tham chiếu cho lớp rth của $Simp(i)$.



Hình 1. Các giao điểm của hướng tham chiếu và một đơn vị đơn hình.

(a) các hướng tham chiếu trên một $Simp(i)$ và (b) là các hướng trong không gian 3 mục tiêu (28 hướng được sinh ra bởi 3 lớp) (Jiang, et al 2017)

Như vậy có thể nhận thấy với k-lớp tổng số hướng tham chiếu cho bài toán tối ưu M mục tiêu có thể được xác định theo công thức như sau:

$$H_M^k = \sum_{i=1}^M \left\{ \sum_{r=1}^k r + k \right\} + 1 = \frac{Mk(k+3)}{2} + 1 \quad (4)$$

Tại mỗi hướng tham chiếu $w^i \in W$, $i \in \{1, \dots, H_M^k\}$, thuật toán xác định không gian mục tiêu tương ứng Ψ^i thông qua định nghĩa sau:

$$\Psi^i = \{\tilde{F}(x) \in \Omega_f \mid \langle \tilde{F}(x), w^i \rangle \leq \langle \tilde{F}(x), w^j \rangle\} \quad (5)$$

Trong đó $i \in \{1, \dots, H_M^k\}$, $x \in \Omega_f$, $\tilde{F}(x)$ là vectơ mục tiêu đã được chuẩn hóa của x, và $\langle \tilde{F}(x), w^i \rangle$ là góc nhọn tạo bởi $\tilde{F}(x)$ và w^i . Sử

dụng định nghĩa này có thể dễ dàng xác định được số các cá thể đang có trong không gian Ψ^i .

Bài toán đã được mã hóa thành chương trình tính toán tối ưu đa mục tiêu với tên gọi MSCE-UA có khả năng hoạt động song song trên nhiều Server hoạt động trong môi trường hệ điều hành Linux và hệ điều hành máy tính hiệu năng cao Cray XC 40 Series (Bùi Đình Lập, 2021).

2.2. Thử nghiệm giải thuật

Để đảm bảo giải thuật hoạt động hiệu quả, chính xác trước khi mã hóa và triển khai ứng dụng trong thực tế. Trong nội dung này chúng tôi tiến hành thử nghiệm giải thuật thông qua các bài toán test mẫu được đề xuất bởi (Zitzler, et al 2000) và (Deb, et al 2002). Bốn bài toán test với độ khó

khác nhau đã được lựa chọn để tham gia đánh giá mức độ hiệu quả và tính ổn định của giải thuật, trong đó: bài toán ZDT1, ZDT6 thực hiện trong điều kiện tối ưu đồng thời 2 hàm mục tiêu với tập tối ưu Pareto front là hàm lồi và hàm phân bố

không đều trên Pareto front; bài toán DTLZ1, DTLZ5 được thực hiện trong điều kiện tối ưu đồng thời 3 hàm mục tiêu.

Mô tả toán học của 4 bài toán test được trình bày như trong Bảng 1.

Bảng 1. Các bài toán tham gia thử nghiệm

Bài toán thử nghiệm và các hàm mục tiêu	Giới hạn biến	Lời giải tối ưu
Bài toán ZDT1: $f_1(x) = x_1; \quad f_2(x) = g * h$ $g(x_2, \dots, x_m) = 1 + 9 * \sum_{i=2}^m x_i / (m - 1)$ $h(f_1, g) = 1 - \sqrt{f_1 / g}$	[0,1]	$x_1 \in [0,1]$ $x_i = 0$ $i = 2, \dots, m$
Bài toán ZDT6: $f_1(x_1) = 1 - \exp(-4x_1) \sin^6(6\pi x_1); \quad f_2(x) = g * h$ $g(x_2, \dots, x_m) = 1 + 9 * ((\sum_{i=2}^m x_i) / (m - 1))^{0.25}$ $h(f_1, g) = 1 - (f_1 / g)^2$	[0,1]	$x_1 \in [0,1]$ $x_i = 0$ $i = 2, \dots, m$
Bài toán DTLZ1: $\begin{cases} \text{Minimize } f_1(x) = \frac{1}{2} x_1 x_2 \dots x_{M-1} (1 + g(x_M)), \\ \text{Minimize } f_2(x) = \frac{1}{2} x_1 x_2 \dots (1 - x_{M-1}) (1 + g(x_M)), \\ \vdots \\ \text{Minimize } f_{M-1}(x) = \frac{1}{2} x_1 (1 - x_2) (1 + g(x_M)), \\ \vdots \\ \text{Minimize } f_M(x) = \frac{1}{2} (1 - x_1) (1 + g(x_M)), \\ \text{subject to } 0 \leq x_i \leq 1, \quad \text{for } i = 1, 2, \dots, n. \end{cases}$	[0,1]	$x_i = 0.5$ $\forall x_i \in x_M$ $\sum_{m=1}^M f_m^* = 0.5$
Bài toán DTLZ5: $\left. \begin{aligned} \text{Min. } f_1(x) &= (1 + g(x_M)) \cos(\theta_1 \pi / 2) \dots \cos(\theta_{M-2} \pi / 2) \cos(\theta_{M-1} \pi / 2), \\ \text{Min. } f_2(x) &= (1 + g(x_M)) \cos(\theta_1 \pi / 2) \dots \cos(\theta_{M-2} \pi / 2) \sin(\theta_{M-1} \pi / 2), \\ \text{Min. } f_3(x) &= (1 + g(x_M)) \cos(\theta_1 \pi / 2) \dots \sin(\theta_{M-2} \pi / 2), \\ &\vdots \\ \text{Min. } f_M(x) &= (1 + g(x_M)) \sin(\theta_1 \pi / 2), \end{aligned} \right\}$ $\text{with } \theta_i = \frac{\pi}{4(1 + g(x_M))} (1 + 2g(x_M)x_i), \quad \text{for } i = 2, 3, \dots, (M - 1)$ $g(x_M) = \sum_{x_i \in x_M} (x_i - 0.5)^2,$ $0 \leq x_i \leq 1, \quad \text{for } i = 1, 2, \dots, n.$	[0,1]	$x_i = 0.5$ $\forall x_i \in x_M$ $\sum_{m=1}^M f_m^2 = 0.5$

Công cụ MatLab đã được sử dụng để mô phỏng và đánh giá mức độ hiệu quả của giải thuật

để xuất thông qua các bài toán test mẫu được đề xuất trong bảng 1. Trong nội dung thử nghiệm

này, chúng tôi lựa chọn số biến quyết định bằng 4, số hàm mục tiêu bằng 2 cho bài toán ZDT1, ZDT6 và bằng 3 cho bài toán DTLZ1, DTLZ5. Các bài toán kiểm thử được thực hiện với kích cỡ mẫu bằng 250 cho bài toán ZDT1, ZDT6 và bằng 1000 cho bài toán DTLZ1, DTLZ5, biến dừng của thuật toán là 10000 lần.

▪ Các chỉ số sử dụng để đánh giá hiệu quả của giải thuật đề xuất:

1-Chỉ số IGD (Inverted Generational Distance): Sử dụng để đánh giá chất lượng của tập lời giải P trong bài toán kiểm thử. Giả thiết P^* là tập hợp các điểm phân phối đều trên Pareto front trong không gian mục tiêu, và P là một xấp xỉ Pareto front trong không gian mục tiêu tìm được. Khoảng cách giữa P^* và P có thể được xác định như sau:

$$IGD(P^*, P) = \frac{\sum_{v \in P^*} d(v, P)}{|P^*|}$$

Trong đó $d(v, P)$ là khoảng cách Euclidean nhỏ nhất từ điểm v tới P. Như vậy có thể thấy giá trị IGD càng nhỏ thì hiệu quả của thuật toán càng tốt.

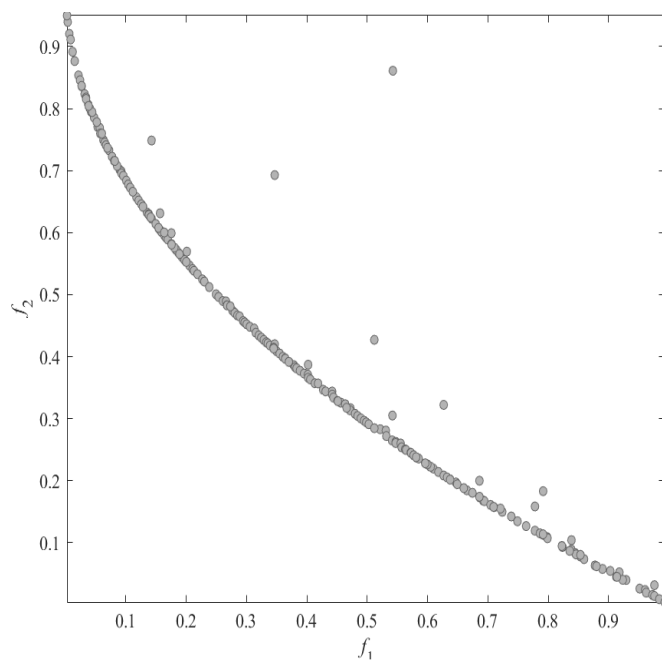
2- Chỉ số HV (Hypervolume): sử dụng để quan sát độ lớn của các lời giải xấp xỉ trội S trong không gian mục tiêu và biên tạo bởi điểm tham chiếu $R = (R_1, \dots, R_M)^T$ trên tất cả các điểm trên mặt tối ưu Pareto.

$$HV(S) = \text{Leb} \left(\bigcup_{x \in S} [f_1(x), R_1] \times \dots \times [f_M(x), R_M] \right)$$

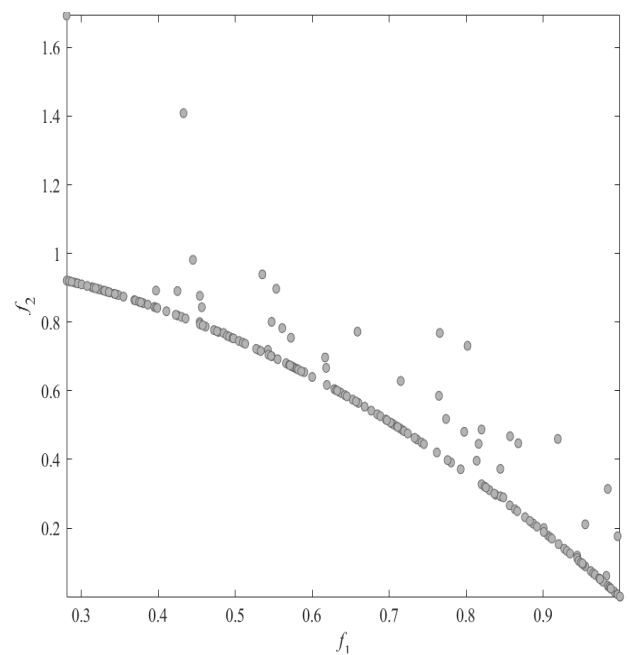
3. KẾT QUẢ VÀ THẢO LUẬN

Kết quả mô phỏng tập Pareto front trên bài toán kiểm tra ZDT1 và ZDT6 xem Hình 2, trên bài toán kiểm tra DTLZ1 và ZDT5 xem Hình 3. Kết quả thử nghiệm giải thuật trên 4 bài toán kiểm tra Hình 2 và Hình 3 cho thấy rõ ràng, thuật toán đề xuất có thể đạt được một sự phân bố khá tốt của các điểm (lời giải) trên tập Pareto front. Không gian lời giải cũng đã chỉ ra được tính đa dạng của các lời giải trên các hàm mục tiêu cạnh tranh của tập Pareto front.

Bảng 2 cũng cho thấy các chỉ số đánh giá mức độ hiệu quả của giải thuật IGD, HV là khá tốt, toàn bộ các chỉ số IGD đều có giá trị nhỏ hơn 0.1, chỉ số HV nhỏ hơn 0.8.

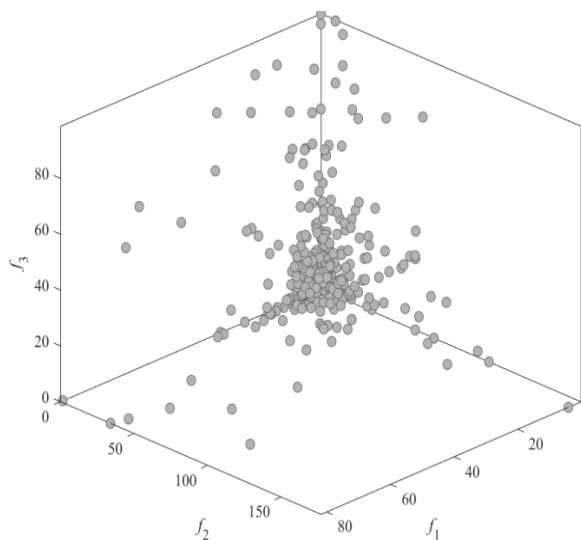


a) Tập Pareto front trên bài toán ZDT1

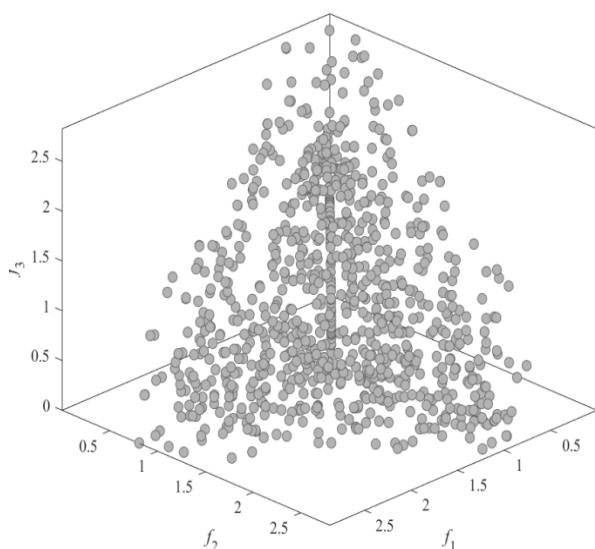


b) Tập Pareto front trên bài toán ZDT6

Hình 2. Kết quả mô phỏng tập Pareto front qua bài toán kiểm tra a) ZDT1 và b) ZDT6



a) Tập Pareto front trên bài toán DTLZ1



b) Tập Pareto front trên bài toán ZDT5

Hình 3. Kết quả mô phỏng tập Pareto front qua bài toán kiểm tra a) DTLZ1 và b) DTLZ5

Bảng 2. Kết quả chỉ số đánh giá mức độ hiệu quả của giải thuật IGD, HV (trung bình và độ lệch chuẩn) đạt được qua các bài toán kiểm tra

ZDT1		ZDT6		DTLZ1		DTLZ5	
IGD	HV	IGD	HV	IGD	HV	IGD	HV
4.9621e-3	7.1880e-1	6.9891e-2	3.1083e-1	4.6169e-2	7.6382e-1	2.9604e-2	7.6382e-1
(4.33e-4)	(7.32e-4)	(5.30e-2)	(6.19e-2)	(2.31e-2)	(8.01e-2)	(4.50e-3)	(8.01e-2)

4. KẾT LUẬN

Các kết quả thử nghiệm đạt được trên các bài toán thử nghiệm đã cho thấy, kỹ thuật phân rã không gian mục tiêu dựa trên tập hướng tham chiếu đã giúp bài toán có khả năng tìm kiếm trên toàn bộ POF (front tối ưu Pareto), đảm bảo sự đa dạng về mẫu trong không gian mục tiêu. Giải

thuật đề xuất mới MSCE-UA có thể ứng dụng để tối ưu thông số cho các mô hình thủy văn thông số phân bố. Sự thành công của công trình nghiên cứu này đã làm giảm được một lượng lớn thời gian, công sức và loại bỏ được tính chủ quan của người hiệu chỉnh mô hình so với phương pháp hiệu chỉnh thử sai.

TÀI LIỆU THAM KHẢO

- Đình Lập, B. (2021). *Nghiên cứu phát triển mô hình thủy văn thông số phân bố trong dự báo lũ cho các lưu vực sông ở Việt Nam*.
- A.Vrugt, J., V.Gupta, H., A.Bastidas, L., Bouten, W., & Sorooshian, S. (2003). *Effective and efficient algorithm for multiobjective optimization of hydrologic models*. Water Resources research, 39.
- Deb, K., Thiele, L., Laumanns, M., & Zitzler, E. (2002). *Scalable multi-objective optimization test problems*. Evolutionary Computation, 1.
- Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). *Optimal use of the SCE-UA global optimization method for calibrating watershed models*. Journal of Hydrology.

- Hồng Thái, T. (2005). *Numerical Methods for Parameter Estimation and Optimal Control of the Red River Network*. Heidelberg university.
- Jiang, S., & Yang, S. (2017). *A Strength Pareto Evolutionary Algorithm Based on Reference Direction for Multiobjective and Many-Objective Optimization*. IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, 21.
- Yapo, P. O., Gupta, H. V., & Sorooshian, S. (1998). *Multi-objective global optimization for hydrologic models*. Journal of Hydrology, 204, 83-97.
- Zitzler, E., Deb, K., & Thiele, L. (2000). *Comparison of Multiobjective Evolutionary Algorithms: Empirical Results*. Evolutionary Computation, 8, 173-195.

Abstract:

**APPLICATION OF OBJECTIVES SPACE DECOMPOSITION TECHNIQUE SOLVE
THE PARAMETER ESTIMATION PROBLEM IN THE DISTRIBUTED
HYDROLOGICAL MODEL**

Determining an optimal parameter set in distributed-hydrological model is very essential to solve practical issues in the hydrology and water resources. This article presents the results of application of spatial decomposition technique to find the globally optimal model parameters in the distributed hydrological model. The results achieved on the basin and the experimental problem have shown that the spatial decomposition technique based on the reference direction has helped the problem to be searchable on the entire POF (Pareto-Optimal Front), ensuring diversity of sample in objective space.

Keywords: Multiobjective optimization, distributed hydrological model, parameter estimation

Ngày nhận bài: 22/12/2020

Ngày chấp nhận đăng: 23/2/2021