

CHỌN HÀM PHÂN PHỐI XÁC SUẤT ĐẠI DIỆN CHO PHÂN PHỐI MƯA 1 NGÀY MAX Ở VIỆT NAM

**Nguyễn Trường Huy¹, Nguyễn Hoàng Lâm¹, Võ Ngọc Dương¹,
Phạm Thành Hưng¹, Nguyễn Chí Công¹**

Tóm tắt: Thông tin về tần suất và cường độ mưa cực hạn là vô cùng quan trọng trong việc thiết kế và quản lý các công trình xây dựng. Thông tin này có được thông qua việc phân tích tần suất mưa (PTTS). Thách thức đặt ra là hiện nay có rất nhiều hàm phân phối xác suất (HPPXS) khác nhau được sử dụng rộng rãi trong việc PTTS nhưng vẫn chưa có sự nhất trí chung về việc nên sử dụng HPPXS nào. Do đó, trong thực tế tính toán, một số HPPXS thông dụng sẽ được lựa chọn và so sánh mức độ khớp để chọn ra hàm thích hợp nhất. Bài báo này so sánh mức độ khớp của 07 HPPXS phổ biến sử dụng số liệu mưa ngày cực hạn từ 155 trạm mưa toàn Việt Nam. Kết quả phân tích dựa trên các dạng đồ thị và các chỉ tiêu thống kê cho thấy hàm phân phối chuẩn tổng quát là HPPXS thích hợp nhất cho việc mô tả mưa ngày cực hạn ở Việt Nam. Hai hàm Pearson loại III và giá trị cực hạn tổng quát cũng cho giá trị gần như tương đồng.

Từ khóa: hàm phân phối xác suất, đường tần suất, mưa lớn nhất, phân tích tần suất.

1. GIỚI THIỆU

Việc thiết kế, quản lý và vận hành hệ thống các công trình xây dựng khác nhau như hệ thống thoát nước đô thị, hệ thống hồ chứa và đập dâng, hệ thống tiêu thoát nước cho cây trồng, v.v. đòi hỏi các thông tin xác suất về tần suất, thời lượng và cường độ của mưa cực hạn. Những thông tin này thường có được thông qua việc phân tích tần suất mưa (PTTS) (Chow, 1964). Để tiến hành PTTS, trước hết cần trích xuất dữ liệu mưa cực hạn (MCH) từ chuỗi dữ liệu đo mưa đầy đủ (WMO, 2009). Thông thường chuỗi giá trị lớn nhất năm, có được bằng cách trích xuất các giá trị mưa lớn nhất hàng năm, được sử dụng rộng rãi trong thực tế. Một phương pháp khác sử dụng chuỗi giá trị vượt ngưỡng, có được bằng cách trích xuất tất cả các giá trị vượt trên một ngưỡng nhất định. Phương pháp thứ hai ít được ưa chuộng hơn do những bất cập trong việc lựa chọn giá trị ngưỡng trích xuất (WMO, 2009). Sau khi đã trích xuất chuỗi MCH, bước tiếp theo là lựa chọn một hàm phân phối xác suất (HPPXS) thích hợp có khả năng mô tả tốt chuỗi MCH thực đo. Đây là một bước

quan trọng và cũng là một trong những thách thức lớn nhất. Việc lựa chọn HPPXS không phù hợp có thể dẫn đến cường độ mưa thiết kế thiên lớn hoặc thiên bé so với thực đo. Thực tế có rất nhiều các HPPXS khác nhau được đề xuất cho việc PTTS các biến cực trị thủy văn (Chow, 1964; Stedinger et al., 1993; WMO, 2009). Tuy nhiên, cho đến nay vẫn chưa có một sự nhất trí chung về việc nên sử dụng HPPXS nào. Việc lựa chọn một HPPXS thích hợp, do đó, thường phụ thuộc vào các đặc trưng của chuỗi dữ liệu thực đo tại các trạm. Trong thực tế, một số HPPXS thông dụng sẽ được lựa chọn và so sánh mức độ khớp (MDK) chuỗi dữ liệu thực đo để chọn ra HPPXS thích hợp nhất (ARR, 2015; Nguyen et al., 2002; Wilks, 1993).

Trong bài báo này, 07 HPPXS hiện đang được sử dụng rộng rãi ở rất nhiều quốc gia khác nhau trên thế giới (WMO, 2009) sẽ được khảo sát, phân tích và so sánh để chọn ra HPPXS tốt nhất sử dụng để miêu tả và PTTS của chuỗi mưa 1 ngày max ở Việt Nam. Các HPPXS này bao gồm: hàm giá trị cực hạn tổng quát (GEV), hàm logistic tổng quát (GLO), hàm phân phối chuẩn tổng quát (GNO), hàm pareto tổng quát (GPA), hàm giá trị cực hạn loại I Gumbel (GUM), hàm Log-Pearson (LP3) và hàm Pearson loại III (PE3)

¹ Khoa Xây dựng Thủy lợi - Thủy điện, Trường Đại học Bách Khoa - Đại học Đà Nẵng.

được trình bày trong phần 2. Cơ sở dữ liệu là 155 trạm quan trắc mưa lớn trải khắp toàn quốc. Kết quả khảo sát, phân tích và so sánh các HPPXS dựa trên việc sử dụng các dạng đồ thị và các tiêu chí thống kê khác nhau được trình bày trong phần 3. Phần 4 trình bày tóm lược lại các kết quả đạt được và đưa ra kết luận.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Các HPPXS thường dùng trong phân tích tần suất thủy văn

Rất nhiều HPPXS khác nhau từ hai đến năm tham số đã được đề xuất cho việc PTTS các biến thủy văn cực trị như hàm GEV, GUM, GLO, GNO, GPA, LP3 và PE3, hàm gamma tổng quát, hàm Beta-Kappa và Beta-Pareto, hàm Kappa, hàm Hyphen, hàm Wakeby (Chow, 1964; Hosking và Wallis, 1997; Stedinger et al., 1993; Wilks, 1993; WMO, 2009). Một vài trường hợp đặc biệt của các phân bố này cũng được dùng rộng rãi như phân phối chuẩn hay chuẩn-log hai tham số, hàm mũ, hàm logistics. Các HPPXS này thường được quy vào họ các hàm phân phối như họ phân phối chuẩn, họ phân phối cực trị, họ Gamma, họ Beta, họ Pareto, họ Hyphen, và nhiều họ khác (Bobée và Ashkar, 1991; WMO, 2009). Thông thường, các HPPXS với nhiều tham số (bốn hay năm tham số) thường có MĐK tốt hơn so với các phân phối ít tham số (hai hay ba tham số). Tuy nhiên các phân phối nhiều tham số thường ít được sử dụng hơn do có nhiều bất cập. Trước hết, việc ước tính tham số trở nên khó khăn hơn và khối lượng tính toán cũng nhiều hơn. Ngoài ra, việc có quá nhiều tham số mặc dù giúp mô tả dữ liệu tốt hơn nhưng mặt trái của nó là làm cho các hàm phân phối trở nên quá cứng nhắc và có thể dẫn đến sự kém chính xác khi sử dụng để ngoại suy các cực trị có tần suất vượt thấp so với các phân phối ít tham số. Việc ngoại suy các cực trị tương ứng với các tần suất thiết kế là một trong những yêu cầu quan trọng trong thực hành do chiều dài các mẫu dữ liệu đo thường ngắn hơn nhiều so với chiều dài tính toán cần thiết. Do đó, trong thực tế sử dụng, các HPPXS với hai hoặc ba tham số thường được ưa chuộng hơn (ARR, 2015; Nguyen, et al 2002; Wilks, 1993).

Nghiên cứu này do đó chỉ tập trung phân tích

và so sánh MĐK của các HPPXS hai và ba tham số được sử dụng phổ biến trong PTTS (ARR, 2015; Nguyen et al., 2002; Wilks, 1993) và có khả năng áp dụng vào PTTS mưa lớn nhất ở Việt Nam, bao gồm hàm GNO, PE3 và LP3, GEV, GUM, GLO và GPA. Trong số các HPPXS vừa đề cập, chỉ có hàm phân phối GUM chứa 02 tham số và đây cũng là một trường hợp đặc biệt của hàm GEV khi tham số hình dạng $\kappa = 0$. Các hàm còn lại đều là hàm 03 tham số.

2.2. Phương pháp ước tính tham số của các HPPXS

Có nhiều cách thức ước tính tham số khác nhau gồm phương pháp moment, phương pháp khả năng lớn nhất, phương pháp moment trọng số xác suất và phương pháp L-moment (Chow, 1964; Hosking và Wallis, 1997; Stedinger, et al., 1993). Các phương pháp này khác nhau ở trọng số mà mỗi phương pháp gán cho các phần tử trong toàn chuỗi dữ liệu, trọng số lớn hơn có thể gán cho các cực trị ở gần phần đuôi hay phần giữa của hàm mật độ xác suất. Phương pháp khả năng lớn nhất cho phép ước tính tham số gần như tối ưu cho một vài HPPXS. Tuy nhiên phương pháp này đòi hỏi khối lượng tính toán lớn do phải dùng phương pháp giải lặp để tìm nghiệm, đồng thời nó cũng rất nhạy khi sử dụng các phương pháp số để tìm nghiệm. Phương pháp L-moment là sự kết hợp tuyến tính các trọng số khác nhau của phương pháp moment trọng số xác suất và cho kết quả gần như không sai lệch. Phương pháp L-moment cho kết quả ổn định hơn nhiều so với phương pháp moment khi có sự tồn tại của các giá trị ngoại lai – là các giá trị cực lớn hay cực nhỏ, và lớn hơn hay nhỏ hơn nhiều lần so với các giá trị còn lại trong chuỗi dữ liệu. Trong nhiều trường hợp, phương pháp L-moment cho kết quả ước tính tham số hữu hiệu hơn nhiều so với phương pháp khả năng lớn nhất (Hosking và Wallis, 1997; Stedinger et al., 1993). Do đó bài báo này sử dụng phương pháp L-moment để ước tính tham số cho tất cả các HPPXS được chọn để khảo sát và so sánh (Hosking và Wallis, 1997).

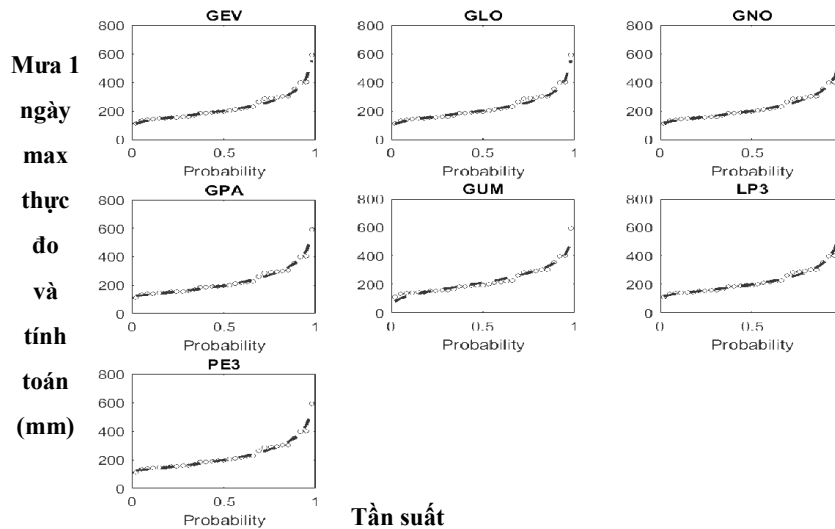
2.3. Các tiêu chí đánh giá mức độ khớp của một HPPXS

Để đánh giá MĐK của các HPPXS một cách tốt nhất, cả phương pháp so sánh trực quan đồ thị và phương pháp so sánh dùng các tiêu chí thống kê được áp dụng. Về phần đồ thị: (i) đồ thị hàm phân bố lũy tích (đồ thị CDF, xem hình 1) và (ii) đồ thị điểm vi phân (đồ thị Q-Q, xem hình 2) được sử dụng. Để xây dựng hai đồ thị này, trước hết cần chọn công thức tính tần suất kinh nghiệm lũy tích. Có rất nhiều công thức khác nhau, tuy nhiên, công thức của Cunnane (1978) cho kết quả điểm phân vị gần như không sai lệch cho rất nhiều các HPPXS khác nhau (Helsel và Hirsch, 2002). Do đó bài báo này sử dụng công thức Cunnane – xem công thức [1].

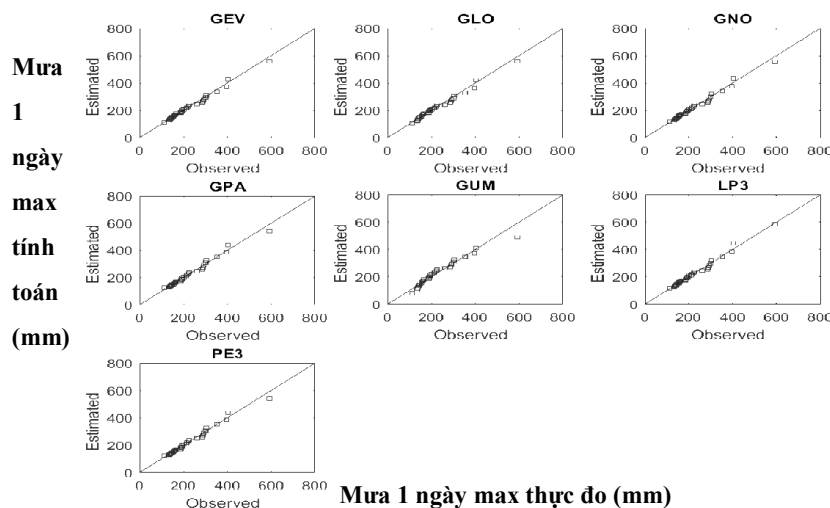
Cả hai đồ thị CDF và Q-Q đều rất hữu dụng trong việc quan sát và so sánh liệu số liệu thực đo có thuộc HPPXS được giả thiết hay không. Để quan sát MĐK của một HPPXS thì có thể sử dụng đồ thị CDF hay đồ thị Q-Q. Tuy nhiên, để so sánh MĐK giữa các HPPXS khác nhau thì đồ thị Q-Q cho cái nhìn rõ ràng hơn (so sánh hình 1 và hình 2).

$$p_{i:n} = \frac{i - 0.4}{n + 0.2} \quad (1)$$

Trong đó: p_i là tần suất lũy tích kinh nghiệm của phần tử thứ i trong mẫu số liệu thực đo có chiều dài n được sắp xếp theo thứ tự từ giá trị nhỏ nhất đến lớn nhất.



Hình 1. Đồ thị CDF của số liệu thực đo (các điểm vòng tròn) và tính toán (đường nét đứt) của 07 HPPXS tính cho trạm mưa Đà Nẵng



Hình 2. Đồ thị Q-Q của số liệu thực đo và tính toán (các điểm ô vuông) của trạm mưa Đà Nẵng. Một cách lý tưởng (tính toán trùng với thực đo), các điểm ô vuông sẽ nằm trên đường thẳng có độ dốc 1:1

Mặc dù cả hai đồ thị CDF và Q-Q đều là công cụ đặc trưng, trong nhiều trường hợp, sự khác nhau trong kết quả tính toán giữa các HPPXS là rất nhỏ và việc so sánh trực quan trở nên khó khăn. Ví dụ hình 2 thể hiện đồ thị Q-Q cho trạm mưa Đà Nẵng. Có thể dễ dàng nhận thấy phân phối GUM có MĐK kém hơn nhiều so với các hàm còn lại, tuy nhiên rất khó để nhận xét và so sánh các hàm còn lại. Do đó cần thiết sử dụng thêm các chỉ tiêu thống kê khác nhau để so sánh sự khác biệt. Có rất nhiều các chỉ tiêu thống kê để đánh giá xem liệu rằng một chuỗi số liệu thực đo có thuộc về một HPPXS được giả thiết hay không (WMO, 2009), như chỉ tiêu Chi-bình phương (χ^2), tỉ số hợp lệ, phân tích Kolmogorov - Smirnov, phân tích Anderson-Darling, và rất nhiều các chỉ tiêu khác như sai số trung bình bình phương, sai số tuyệt đối, v.v... Để thuận tiện cho việc tính toán, bài báo này sử dụng bốn chỉ tiêu thống kê thường hay được sử dụng rộng rãi để so sánh MĐK (Nguyen et al., 2002), bao gồm (1) căn bậc hai của sai số tuyệt đối bình phương trung bình (root mean square error, RMSE), (2) căn bậc hai của sai số tương đối bình phương trung bình (root mean square relative error, RRMSE), (3) sai số tuyệt đối lớn nhất (maximum absolute error, MAE), và (4) hệ số tương quan tính toán-thực đo (correlation coefficient, CC).

$$RMSE = \left\{ \frac{\sum (x_i - y_i)^2}{(n - m)} \right\}^{\frac{1}{2}} \quad (2)$$

$$RRMSE = \left[\frac{1}{(n - m) \sum \left\{ \frac{(x_i - y_i)}{x_i} \right\}^2} \right]^{\frac{1}{2}} \quad (3)$$

$$MAE = \max(|x_i - y_i|) \quad (4)$$

$$CC = \frac{\sum \{(x_i - \bar{x})(y_i - \bar{y})\}}{\left\{ \sum [(x_i - \bar{x})^2] \sum [(y_i - \bar{y})^2] \right\}^{\frac{1}{2}}} \quad (5)$$

Trong đó x_i và y_i là các giá trị thực đo; là các giá trị tính toán từ HPPXS giả thiết; $i = 1, 2, \dots, n$ với n là chiều dài mẫu; m là số tham số của mỗi HPPXS; \bar{x} và \bar{y} là giá trị trung bình của chuỗi số liệu thực đo và số liệu tính toán.

Để so sánh MĐK của các HPPXS khác nhau

sau khi đã tính toán xong các chỉ tiêu thống kê, một sơ đồ xếp hạng được sử dụng để xếp hạng MĐK của các HPPXS. Thứ hạng được gán cho mỗi HPPXS tương ứng với mỗi chỉ tiêu thống kê. Một HPPXS bất kì cho kết quả RMSE, RRMSE và MAE thấp nhất hay CC cao nhất sẽ có thứ hạng 1. Trong trường hợp hai HPPXS bất kì cho kết quả giống nhau, thứ hạng trung bình được sử dụng. Ví dụ hạng 1.5 được sử dụng cho hai HPPXS bất kì cùng xếp thứ 1. Sau khi đã xếp hạng các HPPXS theo các chỉ tiêu khác nhau, thứ hạng tổng cộng tương ứng với mỗi HPPXS và mỗi chỉ tiêu thống kê sẽ được tính toán và so sánh để chọn ra HPPXS có MĐK tốt nhất.

3. ÁP DỤNG CHO DỮ LIỆU MƯA CỰC HẠN Ở VIỆT NAM

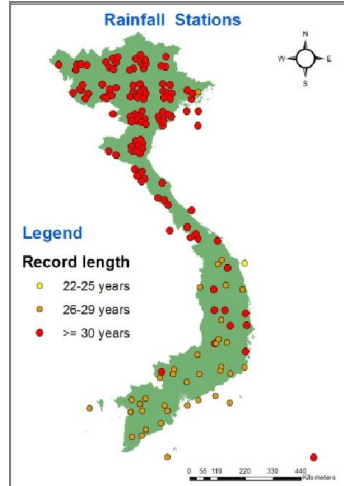
3.1. Cơ sở dữ liệu

Tổng cộng 155 trạm mưa được khảo sát nhằm cung cấp cái nhìn tổng quát về việc chọn HPPXS thích hợp nhất cho việc PTTS mưa 1 ngày max ở Việt Nam. Các trạm quan trắc mưa này được chọn dựa trên chất lượng của trạm đo, chiều dài quan trắc, và khả năng đại diện cho sự phân bố mưa theo không gian tại các vùng khác nhau. Cụ thể, hơn 3/4 số trạm nghiên cứu có thời gian quan trắc trên 30 năm, và 1/4 còn lại có thời gian quan trắc tối thiểu 26 năm, duy nhất 01 trạm Lý Sơn có thời gian quan trắc 22 năm. Số liệu của tất cả các trạm được cập nhật đến năm 2006 và thời gian đo đạc nằm trong khoảng từ năm 1975-2006. Các trạm đo mưa này nằm trải rộng trên toàn quốc, từ Bắc vào Nam và từ Tây sang Đông. Vị trí và sự phân bố của các trạm được thể hiện ở hình 3A.

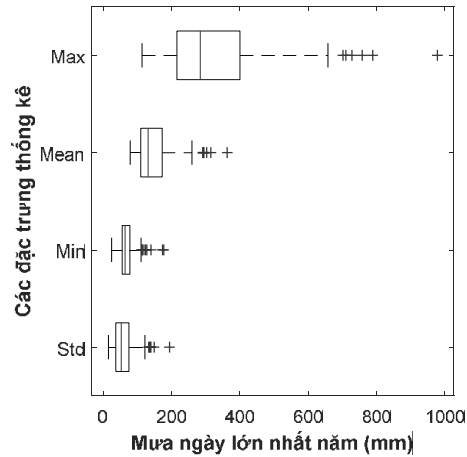
Các đặc trưng thống kê của 155 mẫu dữ liệu mưa 1 ngày max được trình bày trong hình 3B. Đối với mỗi mẫu, các đặc trưng giá trị lớn nhất (max), giá trị trung bình (mean), giá trị nhỏ nhất (min) và độ lệch chuẩn (std) được tính toán. Giá trị thống kê của tất cả 155 mẫu với 4 đặc trưng max, mean, min và std được tổng hợp lại dưới dạng 04 biểu đồ hộp chuẩn (Helsel và Hirsch, 2002) trong hình 3B. Đối với mỗi biểu đồ hộp, chiều rộng của hộp là khoảng cách giữa các tứ vị phân vị (interquartile range, IQR) –

chính là sự khác biệt giữa điểm phân vị 25% (Q1) và 75% (Q3). Đường gạch bên trong thân mỗi hộp thể hiện giá trị trung vị (median value), hay điểm phân vị 50% (Q2). Phần râu (whiskers) phía bên phải và phía bên trái của mỗi hộp kéo dài đến điểm dữ liệu lớn nhất và

nhỏ nhất trong phạm vi $1.5 \cdot IQR$ tính từ cạnh phải và cạnh trái tương ứng của mỗi hộp. Các giá trị nằm ngoài phạm vi hay chiều dài râu của mỗi hộp chính là các giá trị ngoại lai và được thể hiện bằng dấu “+”.



(A)



(B)

Hình 3. (A) Vị trí và phân bố của 155 trạm mưa nghiên cứu và (B) Biểu đồ hộp chuẩn các đặc trưng thống kê của 155 mẫu mưa 1 ngày max nghiên cứu

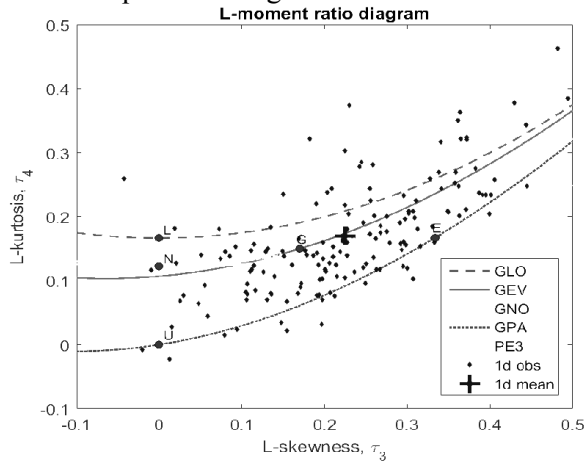
3.2. Kết quả

Một trong những cách thức để chọn sơ bộ HPPXS đại diện cho tất cả các tài liệu mưa 1 ngày max là sử dụng biểu đồ tỉ số L-momen. Biểu đồ này thể hiện mối quan hệ giữa hệ số thiên lệch τ_3 (L-skewness) và hệ số độ nhọn chuẩn τ_4 (L-kurtosis) tính theo phương pháp L-moment. Trên biểu đồ này mỗi phân phối với ba tham số như GEV, GLO, GNO (hay LN3), GPA được thể hiện bằng một đường cong duy nhất. Đường cong này thể hiện sự thay đổi của tham số hình dạng (shape parameter). Các phân bố hai tham số được biểu diễn bằng điểm. Ví dụ: GUM được thể hiện bằng điểm màu đỏ (G) trên đường cong GEV. Biểu đồ tỉ số L-moment của 155 trạm mưa nghiên cứu được thể hiện trên hình 4. Sự phân tán của tất cả các điểm trên biểu đồ xung quanh các HPPXS khác nhau chỉ ra rằng không có một HPPXS nào có thể thỏa mãn tốt tất cả các chuỗi mưa 1 ngày max từ tất cả các trạm. Trong trường hợp như thế này, nếu giá trị trung bình của hệ số thiên lệch τ_3 và hệ số độ nhọn chuẩn τ_4 tính theo phương pháp L-moment

của tất cả các trạm nằm gần với phân phối nào nhất thì phân phối đó sẽ được chọn là HPPXS đại diện cho tất cả các chuỗi dữ liệu. Mặc dù phương pháp này khá chủ quan nhưng nó cũng cho một đánh giá sơ bộ về HPPXS đại diện. Biểu đồ tỉ số L-moment của 155 trạm mưa khảo sát cho thấy giá trị trung bình τ_3 và τ_4 của tất cả các trạm nằm gần với 2 phân bố GNO và GEV.

Để chọn ra HPPXS đại diện từ 07 HPPXS được khảo sát thông qua việc so sánh MDK, như đã trình bày ở mục 2, trước hết đồ thị Q-Q được sử dụng để đánh giá bằng trực quan (xem hình 2). Kết quả quan sát cho thấy phần phía trái và phần giữa của chuỗi dữ liệu đều được mô phỏng khá chính xác bởi tất cả các HPPXS nói chung, riêng phần đuôi bên phải đều có thể bị ước lượng cao, ước lượng thấp, hay tương đối gần với giá trị thực đo bởi tất cả các HPPXS. Đồng thời dựa trên việc phân tích đồ thị Q-Q giữa các HPPXS, có thể dễ dàng nhận thấy khả năng mô phỏng của hàm GUM cho kết quả kém hơn các hàm khác. Điều này hoàn toàn có thể lý giải do hàm GUM chỉ có 02 tham số. Việc thiếu

tham số hình dạng làm cho khả năng miêu tả chuỗi dữ liệu của GUM kém hơn so với các hàm 03 tham số. Tuy nhiên rất khó để kết luận sự khác biệt giữa các HPPXS còn lại, cũng như mức độ ưu việt của mỗi hàm. Do đó, các chỉ tiêu đánh giá được trình bày trong mục 2.3 được sử dụng cùng với sơ đồ xếp hạng để có thể so sánh kết quả MĐMT giữa các HPPXS tốt hơn.



Hình 4. Biểu đồ tỉ số L-moment $\tau_3 \sim \tau_4$ của các trạm mưa (các điểm màu xanh). Giá trị trung bình τ_3 và τ_4 kí hiệu bằng “+”. Các điểm đặc biệt trên các đường cong phân phối thể hiện các phân phối 02 tham số $L =$ phân phối Logistics, $G =$ phân phối Gumbel (cực trị loại I), $N =$ phân phối chuẩn, $E =$ phân phối hàm mũ, $U =$ phân phối đều

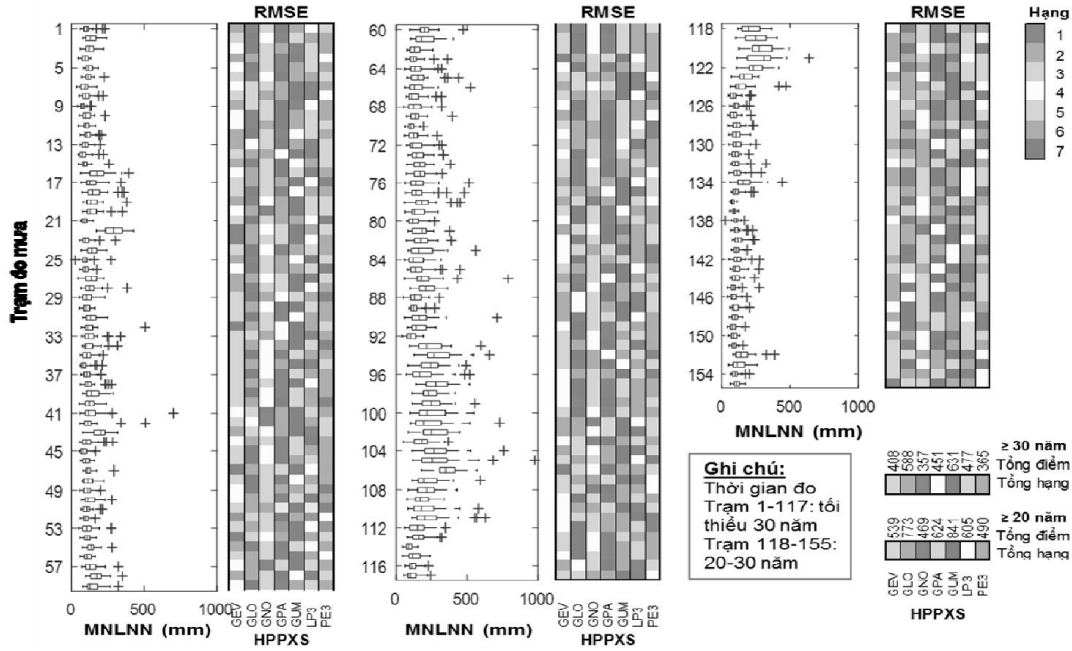
Kết quả so sánh và xếp hạng MĐK dựa trên tiêu chí RMSE của tất cả 07 HPPXS của tất cả các trạm được trình bày trong hình 5 như một minh chứng. Có thể dễ dàng nhận thấy từ kết quả rằng không có một HPPXS nào có thể cho kết quả MĐK dựa trên tiêu chí RMSE luôn tốt nhất hay luôn vượt trội hơn các HPPXS còn lại cho tất cả các mẫu nghiên cứu. Bất cứ hàm nào cũng có thể cho kết quả tốt nhất, trung bình hay kém nhất tùy thuộc vào đặc trưng của các mẫu dữ liệu mưa 1 ngày max được thể hiện thông qua biểu đồ hộp chuẩn đặt theo phương ngang nằm phía bên trái của tiêu chí RMSE trên hình 5. Kết quả tương tự cũng được quan sát thấy khi sử dụng các tiêu chí đánh giá khác như RRMSE, MAE và CC: không có một HPPXS nào là hoàn hảo nhất cho tất cả các mẫu dữ liệu. Tuy nhiên dựa trên kết quả quan sát từ phần đông các trạm, hay dựa trên kết quả tổng xếp hạng từ tổng tỉ số

của tất cả các trạm, và qua các tiêu chí đánh giá khác nhau cũng như mức độ ổn định qua các chiều dài mẫu khác nhau, có thể dễ dàng nhận thấy rằng ba phân phối GUM, GLO và GPA cho kết quả kém nhất (xem hình 6). LP3 nếu chỉ xét đơn thuần chỉ tiêu RRMSE thì nó cho kết quả tốt nhất nhưng nếu dựa trên đánh giá toàn diện từ cả 04 tiêu chí thì LP3 chỉ đứng ở mức độ trung bình. Đối với ba phân bố còn lại, có thể thấy rằng GNO luôn cho kết quả tốt hơn GEV và PE3 ở tất cả các chỉ tiêu, ngoại trừ chỉ tiêu MAE, ở tiêu chí này phân bố PE3 cho kết quả tốt hơn. GEV và PE3 gần như tương đồng, GEV có thể tốt hơn PE3 ở tiêu chí này nhưng lại thua PE3 ở tiêu chí khác (xem hình 6). Tuy nhiên, một điều cần lưu ý là sự khác biệt tuyệt đối giữa các giá trị thực (không phải điểm xếp hạng) tính theo các phân phối GNO, PE3 và GEV cho mỗi trạm cho cả 04 tiêu chí là rất nhỏ. Do đó, về cơ bản cả ba phân bố GNO, PE3 và GEV đều có thể được sử dụng, nhưng GNO có thể được xem như là HPPXS thích hợp nhất cho việc PTTS mưa 1 ngày max ở Việt Nam.

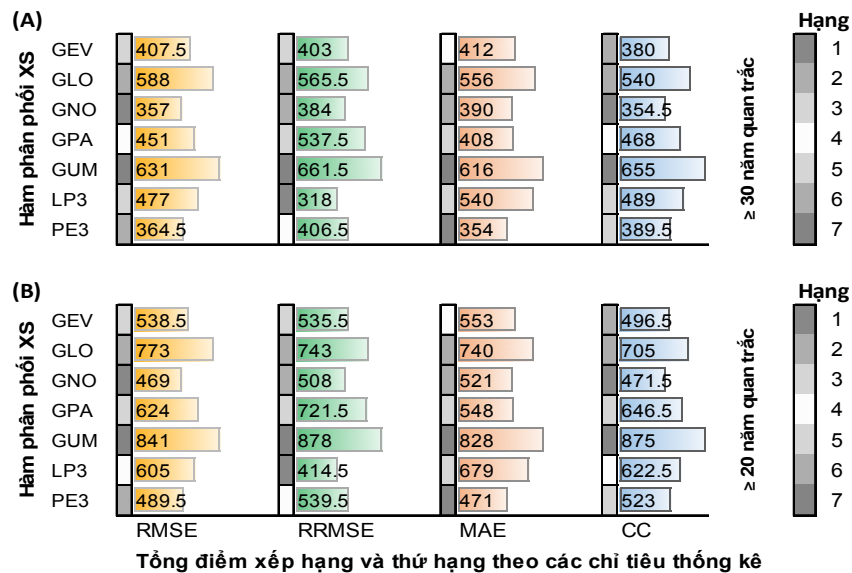
4. KẾT LUẬN

Bài báo này trình bày một cách có hệ thống cách thức để chọn ra HPPXS thích hợp nhất nhằm sử dụng cho việc PTTS mưa 1 ngày max. Trong đó, 07 HPPXS được sử dụng phổ biến ở các nước trên thế giới khi PTTS các biến thủy văn cực trị được khảo sát, bao gồm: hàm giá trị cực hạn tổng quát (GEV), lôgistic tổng quát (GLO), hàm phân phối chuẩn tổng quát (GNO), hàm pareto tổng quát (GPA), hàm giá trị cực hạn loại I Gumbel (GUM), Log-Pearson (LP3) và Pearson loại III (PE3). Kết quả tính toán và phân tích dựa trên các tiêu chí đánh giá khác nhau bao gồm sử dụng đồ thị Q-Q và các chỉ tiêu thống kê RMSE, RRMSE, MAE và CC cho thấy rằng không có một HPPXS nào luôn có thể thoả mãn tốt tất cả các mẫu dữ liệu từ 155 trạm đo mưa ở khắp Việt Nam. Tuy nhiên, xét một cách tổng thể, hàm GNO cho kết quả tốt nhất và ổn định nhất. Do đó, GNO có thể xem là HPPXS thích hợp nhất cho việc mô tả mưa 1 ngày max ở Việt Nam. Hai hàm PE3 và GEV cũng cho giá trị gần như tương đồng với GNO

và cũng có thể được sử dụng cho việc PTTT mưa 1 ngày max ở Việt Nam.



Hình 5. Kết quả xếp hạng các HPPXS tại mỗi trạm và tổng cộng dựa trên tiêu chí RMSE



Hình 6. Kết quả tổng điểm và tổng hạng của các HPPXS dựa trên các tiêu chí RMSE, RRMSE, MAE và CC cho (A) trạm quan trắc tối thiểu 30 năm và (B) tất cả trạm quan trắc

TÀI LIỆU THAM KHẢO

- ARR, 2015. Australian Rainfall & Runoff – A Guide to Flood Estimation. Book 2: Rainfall Estimation. Engineers Australia. Accessed Jan 2016, link: <http://book.arr.org.au/>
- Bobée, B., Ashkar, F., 1991. *The gamma family and derived distributions applied in hydrology*. Water resources publications, Colorado, USA, 203 pp.
- Chow, V.T., 1964. *Handbook of Applied Hydrology*. McGraw-Hill, New York, USA.

- Cunnane, C., 1978. "Unbiased plotting positions" - A review". J. Hydrol., 37(3): 205-222.
- Helsel, D.R., Hirsch, R.M., 2002. *Statistical Methods in Water Resources Techniques of Water Resources Investigations, Book 4, chapter A3*. U.S. Geological Survey, US, 522 pp.
- Hosking, J.R.M., Wallis, J.R., 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge, UK, 224 pp.
- Nguyen, V.-T.-V., Tao, D., Bourque, A., 2002. "On selection of probability distributions for representing annual extreme rainfall series". 9ICUD, ASCE Library, USA
- Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., 1993. *Frequency Analysis of Extreme Events, chapter 18 in Handbook of Hydrology*. McGraw-Hill, New York, USA.
- Wilks, D.S., 1993. "Comparison of three-parameter probability distributions for representing annual extreme and partial duration precipitation series". Water Resour. Res., 29(10):3543-3549.
- WMO, 2009. *Guide to hydrological practices, vol.II: Management of water resources and application of hydrological practices, 6th edition*, Geneva, Switzerland, 302 pp.

Abstract:

**ON SELECTION OF A PARENT PROBABILITY DISTRIBUTION FOR DESCRIBING
THE DISTRIBUTION OF DAILY ANNUAL MAXIMUM RAINFALL SERIES
IN VIETNAM**

Information on the probability of extreme rainfall occurrence and amount is of critical importance for the design and management of various hydraulic structures. This information is commonly obtained using the rainfall frequency analysis (RFA) method – a statistical approach where an appropriate statistical model is selected based on the best fit to the observed data. The common issue is that many statistical models are available, however, there is no general agreement as to which model should be used. Therefore, in practice, a number of popular models are selected and compared for their goodness-of-fit degrees. This paper assesses the performance of 07 statistical models in fitting the daily annual maximum rainfall series (AMS) from a network of 155 rain-gauges in Vietnam. Results based on various numerical and graphical goodness-of-fit criteria have indicated that no distribution can serve as the best distribution for all stations. However, overall, the Generalized Normal (GNO) were the best models and could be recommended as the most suitable model for describing the distribution of daily AMS in Vietnam. The Pearson Type III (PE3) and the Generalized Extreme Values (GEV) distributions provided almost identical results to the GNO distribution and can also be used for RFA.

Keywords: probability distribution, rainfall, annual maximum, frequency analysis.

BBT nhận bài: 16/11/2016

Phản biện xong: 06/3/2017