

## **PHÂN TÍCH TẦN SUẤT MƯA CỰC ĐOAN CHO THÀNH PHỐ HỒ CHÍ MINH CÓ XEM XÉT ĐẾN SỰ BIẾN ĐỘNG CÁC ĐẶC TRƯNG THỐNG KÊ THEO THỜI GIAN**

**Lê Thị Hòa Bình<sup>1</sup>, Đặng Đồng Nguyên<sup>1</sup>**

**Tóm tắt:** *Mưa cực đoạn, một trong những nguyên nhân gây ra ngập lụt tại thành phố Hồ Chí Minh, đã chứng kiến một sự gia tăng về tần suất xuất hiện và cường độ trong vài thập kỷ qua. Mặc dù tính không dừng (nonstationary) trong dữ liệu mưa cực đoạn đã được nghiên cứu và chứng minh trong rất nhiều nghiên cứu trên thế giới, tuy nhiên nghiên cứu về tính không dừng trong dữ liệu mưa cực đoạn tại thành phố Hồ Chí Minh chưa nhận được nhiều sự quan tâm từ các nhà khoa học. Trong nghiên cứu này, chúng tôi sẽ xem xét đến tính không dừng trong dữ liệu mưa 24h tại trạm Tân Sơn Hòa. Kết quả từ nghiên cứu chỉ ra rằng, mưa cực đoạn tại trạm Tân Sơn Hòa có xu hướng tăng khá mạnh mẽ trong giai đoạn 1982-2018. Bên cạnh đó, giá trị thiết kế của mưa cực đoạn dựa trên giả thiết về tính dừng nhỏ hơn đáng kể so với các giá trị mưa dựa trên giả thiết về tính không dừng trong dữ liệu mưa.*

**Từ khóa:** Mưa cực đoạn, Tp.HCM, Mưa thiết kế, Tính không dừng.

### **1. GIỚI THIỆU**

Trong vài thập kỷ trở lại đây, bên cạnh những thách thức lớn liên quan đến gia tăng nhanh về dân số, đô thị hóa và công nghiệp hóa, thành phố Hồ Chí Minh (Tp. HCM) còn phải đối mặt với những vấn đề liên quan đến khí hậu, ví dụ như sự gia tăng về tần suất và cường độ của các sự kiện mưa cực đoạn. Các trận mưa cực đoạn cũng được xem là một trong các nguyên nhân dẫn đến tình trạng ngập úng diễn ra hàng năm ở tại trung tâm kinh tế lớn nhất cả nước này. Nghiên cứu về mưa cực đoạn, cũng như phân tích xu thế, độ lớn của các trận mưa cực đoạn là vấn đề hết sức cần thiết, nhằm cung cấp thông tin, dữ liệu cho việc quy hoạch, xây dựng cơ sở hạ tầng và tính toán thiết kế nhằm giải quyết vấn đề ngập úng lâu nay tại Tp. HCM.

Thông thường, khi xem xét đến chuỗi số liệu khí tượng thủy văn (ví dụ như lượng mưa), cấu trúc của chuỗi số liệu được giả định là có tính dừng. Có nghĩa là các tham số thống kê của các hàm phân phối được chọn không thay đổi theo thời gian (Katz, 2013). Tuy nhiên, trong bối cảnh

biến đổi khí hậu, giả định về tính dừng (stationary) trong chuỗi số liệu khí tượng thủy văn có thể không còn phù hợp nữa (Khaliq, Ouarda, Ondo, Gachon, & Bobée, 2006; Sugahara, Da Rocha, & Silveira, 2009). Thay vào đó, tính không dừng (nonstationary) nên được xem xét đến, nhất là trong phân tích tần suất của các sự kiện cực đoạn. Bên cạnh đó, các giá trị khí tượng thủy văn tính toán (ví dụ như mưa thiết kế) dựa trên giả thiết về tính dừng của chuỗi số liệu thường là nhỏ hơn so với các giá trị tính toán dựa trên giả thiết tính không dừng (Agilan & Umamahesh, 2017; Cheng & AghaKouchak, 2014), do đó sẽ ảnh hưởng đến việc tính toán thiết kế, cũng như hiệu quả làm việc của các công trình.

Trong nghiên cứu này, phân tích tần suất của mưa cực đoạn tại trạm Tân Sơn Hòa (Tp. HCM), cũng như tính toán giá trị mưa thiết kế tương ứng với chu kỳ lặp lại (return period) là 2, 20 và 100 năm được xét đến. Chuỗi số liệu mưa sẽ được kiểm tra xu thế tăng hay giảm dựa trên kiểm định Mann-Kendall. Tính không dừng sẽ được áp dụng trong phân tích tần suất của mưa cực đoạn bằng cách xem xét biến thời gian trong hàm phân phối xác suất. Các phân tích, so sánh giữa các mô hình sẽ được đưa ra để lựa chọn mô hình phù hợp

---

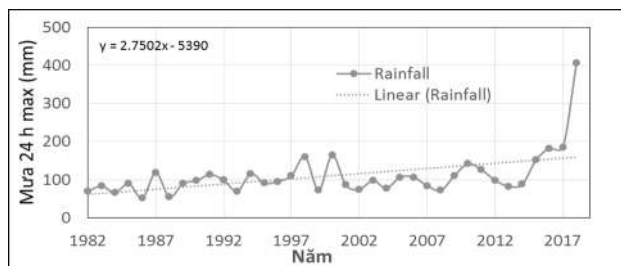
<sup>1</sup> Bộ môn Kỹ thuật Tài nguyên nước và Môi trường, Đại học Thủy lợi phân hiệu Bình Dương

nhất dùng để mô phỏng mưa cực đoan tại vùng nghiên cứu.

## 2. SỐ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

### 2.1. Số liệu thu thập

Trong nghiên cứu này, số liệu mưa 24h lớn nhất được trích xuất từ chuỗi số liệu mưa giờ từ 1982-2018 tại trạm Tân Sơn Hòa thu thập từ Trung tâm tư liệu Khí tượng Thủy văn Quốc gia được sử dụng để phân tích tần suất. Hình 1 thể hiện sự biến đổi của mưa 24h lớn nhất Tân Sơn Hòa và xu thế tuyến tính.



Hình 1. Mưa 24h lớn nhất tại trạm Tân Sơn Hòa và xu thế tuyến tính

### 2.2. Phương pháp nghiên cứu

Phương pháp nghiên cứu trong bài báo này được thực hiện theo trình tự như sau:

- Kiểm định phi tham số Mann-Kendall sẽ được áp dụng để đánh giá xu thế trong chuỗi số liệu mưa 24h lớn nhất
- Tính không dừng sẽ được xem xét trong hàm phân phối xác suất Generalized Extreme Value (GEV)
- Các chỉ số AIC (The Akaike Information Criterion), BIC (The Bayesian Information criterion), và kiểm định likelihood ratio (likelihood ratio test) được dùng để lựa chọn mô hình phù hợp nhất
- Giá trị mưa cực đoan tương ứng với chu kỳ lặp lại là 2, 20 và 100 năm sẽ được dự đoán dựa trên mô hình phù hợp nhất.

#### Kiểm định phi tham số Mann-Kendall

Thông thường, phân tích xu hướng được sử dụng để phát hiện tính không dừng trong chuỗi số liệu khí tượng thủy văn. Kiểm định Mann-Kendall (Kendall, 1962; Mann, 1945), là kiểm định phi tham số, thường được sử dụng rộng rãi để phân tích các xu hướng đơn điệu trong chuỗi dữ liệu. Kết quả từ kiểm định Mann-Kendall cho biết giá

trị của *Tau* (*Tau value*), nó cho biết chuỗi số liệu có xu hướng tăng hay giảm.

Giả thiết một chuỗi dữ liệu thời gian  $(x_1, x_2, x_3, \dots, x_n)$  biểu diễn  $n$  điểm dữ liệu,  $x_i$  biểu diễn số liệu tại thời điểm  $i$ ,  $x_j$  biểu diễn số liệu tại thời điểm  $j$ . Chỉ số thống kê Mann-Kendall  $S$  được tính như sau:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sign}(x_j - x_i) \quad (1)$$

Trong đó,  $\text{sign}(x)$  được xác định như sau:  $\text{sign}(x) = 1$  nếu  $x > 0$ ,  $\text{sign}(x) = 0$  nếu  $x = 0$  và  $\text{sign}(x) = -1$  nếu  $x < 0$ . Giá trị ban đầu của thống kê Mann-Kendall  $S$  là 0 tương ứng với việc không tồn tại xu hướng.

Giá trị của *Tau* được xác định bởi công thức sau (Chandler & Scott, 2011):

$$Tau = \frac{2S}{N(N-1)} \quad (2)$$

Với giá trị  $Tau > 0$ , chuỗi số liệu thể hiện xu thế tăng, ngược lại khi  $Tau < 0$ , chuỗi số liệu thể hiện xu thế giảm.

#### Lựa chọn hàm phân phối xác suất

Hiện nay có rất nhiều hàm phân phối xác suất được sử dụng để mô tả dữ liệu mưa cực đoan, ví dụ như hàm Gumbel, Log-Normal, Pearson, GEV, Pareto, v.v. Trong đó, hàm GEV và Pareto thường được sử dụng nhiều nhất trong phân tích tần suất của các hiện tượng thời tiết cực đoan như mưa, bão và lũ lụt. Trong nghiên cứu này, hàm phân phối xác suất GEV được sử dụng để phân tích dữ liệu mưa 24h lớn nhất cho trạm Tân Sơn Hòa. Giả sử  $x = x_1, x_2, x_3, \dots, x_n$  thể hiện lượng mưa 24h lớn nhất hàng năm của  $n$  biến ngẫu nhiên độc lập và phân phối giống nhau, hàm phân phối lũy tích của GEV được thể hiện ở phương trình sau:

$$G(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\},$$

$$1 + \xi \left(\frac{x-\mu}{\sigma}\right) > 0, \sigma > 0 \quad (3)$$

Trong đó,  $\mu$  (*location*),  $\sigma$  (*scale*) và  $\xi$  (*shape*) thể hiện các tham số thống kê của hàm GEV (thông tin chi tiết về các tham số  $\mu$ ,  $\sigma$  và  $\xi$  vui lòng có thể tham khảo trong Coles, Bawa, Trenner, and Dorazio (2001)). Khi chuỗi số liệu

được xem là có tính dừng, giá trị của các tham số là hằng số. Trong trường hợp chuỗi số liệu được coi là không dừng, giá trị của tham số sẽ biến đổi theo biến số (ví dụ như thời gian, hoặc yếu tố khí hậu). Trong nghiên cứu này, tham số  $\mu$  và  $\sigma$  được biểu diễn như là một biến số theo thời gian:

$$\begin{aligned} \mu(t) &= \mu_0 + \mu_1 \times (t); \\ \sigma(t) &= \sigma_0 + \sigma_1 \times (t); \quad \xi(t) = \xi \end{aligned} \quad (4)$$

Các tham số của hàm GEV sẽ được ước lượng qua phương pháp ước lượng hợp lý cực đại (maximum-likelihood estimation).

#### Lựa chọn mô hình thích hợp nhất

Trong nghiên cứu này, chỉ số AIC (Akaike, 1974), BIC (Schwarz, 1978) và kiểm định likelihood ratio ( $p$ -value) sẽ được dùng để lựa chọn mô hình thích hợp nhất. Mô hình với giá trị của AIC, BIC và  $p$ -value nhỏ hơn thì được xem là mô hình tốt hơn được lựa chọn để mô tả mưa cực đoan. Bên cạnh đó, các biểu đồ xác suất Probability - Probability (PP) và phân vị Quantile - Quantile (QQ) sẽ được sử dụng để kiểm tra sự phù hợp của mô hình được chọn. Hai chỉ số AIC và BIC được tính theo công thức sau:

$$AIC = -2(\log\text{likelihood}) + 2k \quad (5)$$

$$BIC = -2(\log\text{likelihood}) + k \ln(n) \quad (6)$$

Trong đó  $k$  là số lượng tham số trong mô hình,  $n$  là độ lớn mẫu

#### Tính toán giá trị mưa tần suất thiết kế

Khi mô hình phù hợp nhất để mô phỏng tài liệu

mưa cực đoan được lựa chọn, các giá trị mưa cực đoan ( $Z_T$ ) tương ứng với chu kỳ lặp lại ( $T$ -year) 2, 20 và 100 năm sẽ được tính toán. Đối với mô hình dựa trên giả thiết về tính không dừng của chuỗi số liệu, các tham số của hàm phân phối xác suất sẽ biến đổi theo thời gian. Do đó, các tác giả dựa trên cách tiếp cận rủi ro thấp của Cheng et al., 2014, bằng cách lấy 95% (95 percentile) giá trị của tham số  $\mu$  và  $\sigma$  (phương trình 7 và 8) để tính toán giá trị mưa cực đoan ứng với các chu kỳ lặp lại trong nghiên cứu này.

$$\hat{\mu}_{95} = Q_{95}(\hat{\mu}_{t1}, \hat{\mu}_{t2}, \dots, \hat{\mu}_{tm}) \quad (7)$$

$$\hat{\sigma}_{95} = Q_{95}(\hat{\sigma}_{t1}, \hat{\sigma}_{t2}, \dots, \hat{\sigma}_{tm}) \quad (8)$$

Giá trị mưa cực đoan tương ứng với chu kỳ lặp lại  $T$  được đưa ra bởi Coles et al. (2001) như sau:

$$z_{T=} \begin{cases} \mu - \frac{\sigma}{\xi} \left[ 1 - \left\{ -\log \left( 1 - \frac{1}{T} \right) \right\}^{-\xi} \right], & \text{for } \xi \neq 0 \\ \mu - \sigma \log \left\{ -\log \left( 1 - \frac{1}{T} \right) \right\}, & \text{for } \xi = 0 \end{cases} \quad (9)$$

Các tính toán trong nghiên cứu này được xử lý bằng phần mềm R studio với ngôn ngữ lập trình R.

### 3. KẾT QUẢ VÀ THẢO LUẬN

Như đã đề cập ở phần trên, kiểm định Mann-Kendall được dùng để xác định xu thế trong chuỗi số liệu mưa cực đoan. Khi  $Tau$  mang giá trị âm, chuỗi số liệu thể hiện xu thế giảm, ngược lại khi  $Tau$  mang giá trị dương, chuỗi số liệu thể hiện xu thế tăng. Bảng 1 cho thấy chuỗi số liệu mưa 24h lớn nhất mang xu thế tăng đáng kể qua từng năm, và thỏa mãn mức ý nghĩa  $\alpha = 0.05$  (xác suất phạm sai lầm không quá 5%).

**Bảng 1. Kết quả kiểm định Mann-Kendall (p. value- Mức ý nghĩa)**

Mann-Kendall	$S$	$Var(S)$	$Tau$	$p. value$
Giá trị	240	5846	0.36	0.001773

Bảng 2 thể hiện giá trị của các tham số của 2 mô hình phân bố xác suất dựa trên giả thiết về tính

dừng (SGEV) và không dừng (NSGEV) qua phương pháp ước lượng hợp lý cực đại.

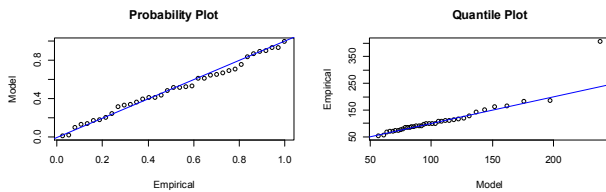
**Bảng 2. Tham số của hàm phân phối xác suất GEV**

Tham số	Location		Scale		Shape
	$\mu_0$	$\mu_1$	$\sigma_0$	$\sigma_1$	$\xi$
SGEV	86.23		26.37		0.24
NSGEV	63.13	1.36	14.01	0.50	0.22

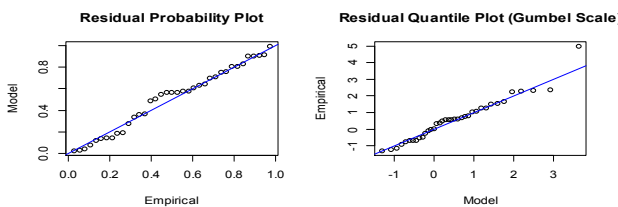
Các chỉ số AIC, BIC đều chỉ ra rằng hàm phân phối xác suất GEV dựa trên giả thiết về tính không dừng trong chuỗi số liệu được xem là phù hợp hơn cho mô phỏng giá trị mưa cực đoạn trong nghiên cứu này (Bảng 3). Thêm vào đó, các biểu đồ PP và QQ cũng xác nhận mô hình NSGEV cho kết quả giữa số liệu thực đo và mô hình tương đối phù hợp so với mô hình SGEV (Hình 2 và 3). Kết quả của kiểm định likelihood ratio cũng chỉ ra rằng mô hình NSGEV phù hợp hơn so với mô hình SGEV với giá trị  $p$ -value là 0.002099.

**Bảng 3. Chỉ số để lựa chọn mô hình tốt nhất**

Chỉ số	SGEV	NSGEV
AIC	375	367
BIC	380	374



Hình 2. PP và QQ plot cho mô hình SGEV



Hình 3. Residual PP và QQ plot cho mô hình NSGEV

Các giá trị mưa cực đoạn tương ứng với chu kỳ lặp lại 2, 20 và 100 năm được thể hiện ở Bảng 4. Kết quả cho thấy rằng giả thiết về tính dừng trong chuỗi số liệu mưa có thể dẫn đến việc đánh giá thấp các sự kiện mưa cực đoạn. Do đó, nếu sử dụng giá trị thiết kế dựa theo tính dừng trong chuỗi số liệu có thể sẽ dẫn đến những thiệt hại nghiêm trọng, nhất là đối với các công trình phòng chống ngập lụt. Điều này cũng có ý nghĩa quan trọng đối với các nhà hoạch định chính sách trong việc quy hoạch thiết kế các công trình phòng

chống thiên tai, cũng như việc dự đoán các trận mưa cực đoạn gây ra tình trạng ngập lụt ở vùng nghiên cứu.

**Bảng 4. Giá trị mưa cực đoạn (mm)**

Chu kỳ lặp lại (năm)	2	20	100
SGEV	96.30	200.14	306.84
NSGEV (95%)	111.50	217.60	355.37

#### 4. KẾT LUẬN

Bài báo đã đánh giá xu thế biến đổi của mưa cực đoạn giai đoạn 1982-2018 cho trạm Tân Sơn Hòa, cũng như dự đoán giá trị mưa cực đoạn tương ứng với các chu kỳ lặp lại khác nhau, dựa theo giả thiết về tính không dừng trong dữ liệu mưa 24h. Từ kết quả đạt được, có thể đưa ra các kết luận sau:

- Mưa cực đoạn có xu hướng tăng mạnh trong giai đoạn 1982-2018
- Mô hình NSGEV dựa trên hàm phân phối xác suất GEV và giả thiết về tính không dừng trong chuỗi dữ liệu mưa hoàn toàn phù hợp để mô phỏng mưa cực đoạn 24h tại khu vực nghiên cứu.
- Các giá trị mưa thiết kế dựa trên giả thiết về tính dừng (SGEV) nhỏ hơn so với các giá trị mưa dựa trên giả thiết về tính không dừng trong dữ liệu mưa
- Từ kết quả nghiên cứu này, mô hình NSGEV được kiến nghị nên sử dụng rộng rãi trong khu vực Tp.HCM nhằm cung cấp thông tin, dữ liệu cho việc tính toán thiết kế, xây dựng các công trình chống ngập tại Tp.HCM.
- Hạn chế trong nghiên cứu này có thể là chỉ xem xét các tham số  $\mu$  và  $\sigma$  của hàm phân phối xác suất GEV như một biến số theo thời gian mà chưa xem xét đến các yếu tố khác. Do đó, tác động của các yếu tố mang tính chất vùng, địa phương lên mưa cực đoạn tại Tp. HCM sẽ được xem xét và đánh giá trong nghiên cứu tiếp theo.

## TÀI LIỆU THAM KHẢO

- Agilan, V., & Umamahesh, N. (2017). *Modelling nonlinear trend for developing non - stationary rainfall intensity–duration–frequency curve*. International Journal of Climatology, 37(3), 1265-1281.
- Akaike, H. (1974). *A new look at the statistical model identification*. IEEE transactions on automatic control, 19(6), 716-723.
- Chandler, R., & Scott, M. (2011). *Statistical methods for trend detection and analysis in the environmental sciences: John Wiley & Sons*.
- Cheng, L., & AghaKouchak, A. (2014). *Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate*. Scientific reports, 4, 7093.
- Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values (Vol. 208): Springer*.
- Katz, R. W. (2013). *Statistical methods for nonstationary extremes Extremes in a Changing Climate (pp. 15-37): Springer*.
- Kendall, M. G. (1962). *Rank correlation methods*. New York: Hafner Publishing Company.
- Khaliq, M., Ouarda, T., Ondo, J.-C., Gachon, P., & Bobée, B. (2006). *Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review*. Journal of Hydrology, 329(3), 534-552.
- Mann, H. B. (1945). *Nonparametric Tests Against Trend*. Econometrica, 13(3), 245-259. doi: DOI: 10.2307/1907187
- Schwarz, G. (1978). *Estimating the dimension of a model*. The annals of statistics, 6(2), 461-464.
- Sugahara, S., Da Rocha, R. P., & Silveira, R. (2009). *Non - stationary frequency analysis of extreme daily rainfall in Sao Paulo, Brazil*. International Journal of Climatology, 29(9), 1339-1349.

### Abstract:

#### NONSTATIONARY EXTREME VALUE ANALYSIS FOR ESTIMATION OF DESIGN RAINFALL IN HO CHI MINH CITY

*In Ho Chi Minh City (HCMC), heavy rainfall, which is considered as a main cause of inundation, witnessed an increase in frequency and magnitude in last few decades. Although nonstationarity in extreme rainfall has been proved in many places of the world, research into nonstationarity feature in extreme rainfall in HCMC has not been paid attention thoroughly. In this study, the extreme rainfall timeseries is modelled under nonstationary condition. The results show that the increasing trend has been found in extreme rainfall data of Tan Son Hoa station over the period of 1982-2018. Besides, the design rainfall estimates under the stationary condition are lower than those under the nonstationary condition in the study area.*

**Keywords:** Extreme rainfall, HCMC, Design rainfall, Nonstationary

---

Ngày nhận bài: 11/4/2020

Ngày chấp nhận đăng: 02/6/2020