

MÔ PHỎNG MỰC NƯỚC CỰC ĐOAN TẠI VÙNG TÀU CỎ XÉT ĐẾN SỰ THAY ĐỔI VỀ XU THẾ CỦA CHUỖI SỐ LIỆU

Võ Văn Tiền¹, Nguyễn Đăng Tính², Lê Thị Hòa Bình², Đặng Đồng Nguyên²

Tóm tắt: Trong nghiên cứu này, xu thế biến đổi của chuỗi dữ liệu mực nước cực đoạn tại trạm Vũng Tàu qua giai đoạn 1980-2019 được đánh giá qua kiểm định Mann-Kendall. Bên cạnh đó, tính không dừng trong dữ liệu mực nước cực đoạn cũng được xem xét. Kết quả từ nghiên cứu chỉ ra rằng, mực nước cực đoạn tại trạm Vũng Tàu có xu hướng tăng khá mạnh mẽ trong giai đoạn 1980-2019. Bên cạnh đó, giá trị thiết kế của mực nước cực đoạn dựa trên giả thiết về tính dừng nhỏ hơn so với các giá trị mực nước dựa trên giả thiết về tính không dừng.

Từ khóa: Mực nước cực đoạn, Tính dừng, Mực nước thiết kế, Tính không dừng, Trạm Vũng Tàu

1. ĐẶT VẤN ĐỀ

Những năm gần đây, tình trạng xâm nhập mặn ngày càng lan rộng và trở nên nghiêm trọng, ảnh hưởng trực tiếp đến sản xuất nông nghiệp, tác động lớn đến đời sống nhân dân và nền kinh tế - xã hội. Tại đồng bằng sông Cửu Long (ĐBSCL), xâm nhập mặn trở nên gay gắt hơn bao giờ hết. Cuối năm 2015 và những tháng đầu năm 2016, diễn biến xâm nhập mặn tại ĐBSCL được đánh giá là nặng nề nhất trong 100 năm qua. Trên sông Tiền và sông Hậu, độ mặn trên 45‰, xâm nhập sâu tới 70 km tính từ cửa sông, có nơi thậm chí lên đến 85 km.

Riêng hệ thống sông Vàm Cỏ (Vàm Cỏ Đông, Vàm Cỏ Tây), tình trạng hạn hán và xâm nhập mặn cũng không kém phần nghiêm trọng. Từ giữa tháng 11/2019, xâm nhập mặn bắt đầu xuất hiện và xâm nhập sâu vào hệ thống sông Vàm Cỏ thời gian xuất hiện mặn sớm hơn gần 01 tháng so với cùng kỳ 2018-2019 và sớm hơn nửa tháng so với cùng kỳ 2015-2016. Đến ngày 28/04/2020, do ảnh hưởng kỳ triều cường kết hợp nắng nóng nên độ mặn trên các sông Vàm Cỏ tiếp tục tăng.

Có thể nhận định rằng, triều cường là một trong những nguyên nhân quan trọng ảnh hưởng đến mức độ xâm nhập mặn của hệ thống sông. Do đó, nghiên cứu tần suất về mực nước triều sẽ cung

cấp những luận cứ khoa học để mô tả và phân tích tình trạng hạn hán và xâm nhập mặn trên các lưu vực sông.

Tổng quan các nghiên cứu về mực nước cực đoạn (extreme water level) cho thấy, giá trị mực nước cực đoạn thiết kế có ý nghĩa hết sức quan trọng cho việc lập kế hoạch và thiết kế các công trình phòng chống lũ (Arns & cộng sự, 2013; Katz, 2013). Thông thường, việc đánh giá mực nước cực đoạn được dựa trên những phân tích thống kê theo lý thuyết giá trị cực đoạn (Arns & cộng sự, 2015; Bulteau & cộng sự, 2015; Mudersbach & Jensen, 2010). Trong đó, chuỗi số liệu mực nước được giả định là có tính đồng nhất hoặc tính dừng. Có nghĩa là các tham số thống kê của hàm phân phối được chọn không thay đổi theo thời gian (Katz, 2013). Tuy nhiên, trong bối cảnh biến đổi khí hậu, nước biển dâng, giả định về tính dừng (stationary) trong chuỗi số liệu mực nước có thể không còn phù hợp nữa. Quả thực, tính không dừng (nonstationary) trong chuỗi số liệu mực nước cực đoạn đã được nghiên cứu và chứng minh trong rất nhiều nghiên cứu trên thế giới (Arns & cộng sự, 2015; Méndez & cộng sự, 2007; Menéndez & Woodworth, 2010; Mudersbach & Jensen, 2010; Serafin & Ruggiero, 2014; Skjong & cộng sự, 2013).

Trong nghiên cứu này, các tác giả sẽ mô phỏng mực nước cực đoạn tại trạm Vũng Tàu và dự đoán giá trị mực nước cực đoạn tương ứng với chu kỳ

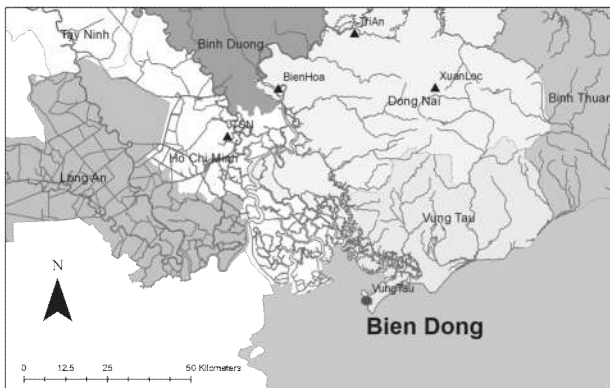
¹ Công ty TNHH Tư vấn Thiết kế XD Vĩnh Hưng, tỉnh Long An

² Phân hiệu trường Đại học Thủy lợi tại tỉnh Bình Dương

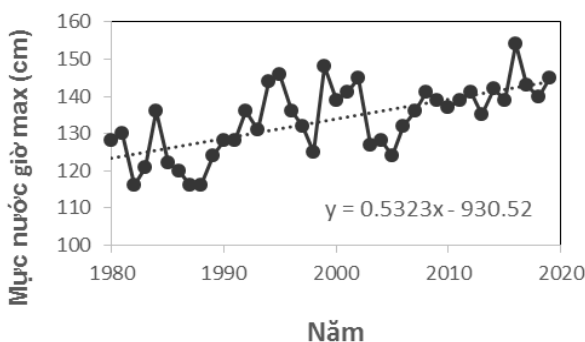
lập lại là 10, 50 và 100 năm. Bên cạnh đó, tính không dừng sẽ được áp dụng trong phân tích tần suất của mực nước cực đoan bằng cách xem xét biến thời gian trong hàm phân phối xác suất Generalized Extreme Value (GEV). Mô hình phù hợp nhất dùng để mô phỏng mực nước cực đoan tại Vũng Tàu sẽ được đưa ra qua các phân tích, so sánh giữa các mô hình. Kết quả từ nghiên cứu này sẽ là tài liệu quan trọng cung cấp cho các bài toán mô phỏng về xâm nhập mặn trên các lưu vực sông Mê Kông và sông Vàm Cỏ.

2. SỐ LIỆU SỬ DỤNG

Trong nghiên cứu này, số liệu mực nước giờ từ năm 1980 đến năm 2019 tại trạm đo Vũng Tàu thu thập từ Trung tâm dữ liệu Khí tượng Thủy văn Quốc gia được sử dụng để phân tích tần suất. Hình 1 thể hiện vị trí của trạm mực nước Vũng Tàu. Hình 2 thể hiện sự biến đổi của mực nước giờ lớn nhất của trạm Vũng Tàu và xu thế tuyến tính.



Hình 1. Vị trí trạm mực nước Vũng Tàu



Hình 2. Mực nước giờ lớn nhất tại trạm Vũng Tàu

3. PHƯƠNG PHÁP NGHIÊN CỨU

3.1 Kiểm định phi tham số Mann-Kendall

Hiện nay, giả định về tính không dừng trong

chuỗi số liệu khí tượng thủy văn đã được nhiều nhà khoa học nghiên cứu và ủng hộ. Do đó, kiểm tra tính không dừng trong chuỗi số liệu là bước quan trọng trong phân tích 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, qua đó phát hiện tính không dừng trong chuỗi số liệu.

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.

3.2 Hàm phân phối xác suất GEV

Các hàm phân phối xác suất như hàm Gumbel, Log-Normal, Pearson, GEV, Pareto, v.v., thường được sử dụng để mô tả dữ liệu khí tượng thủy văn. Trong đó, hàm GEV và Pareto thường được sử dụng nhiều nhất trong phân tích tần suất của các giá trị khí tượng thủy văn cực đoan như mưa, mực nước và hạn hán. Trong nghiên cứu này, chúng tôi sử dụng hàm phân phối xác suất GEV để mô tả dữ liệu mực nước giờ lớn nhất cho trạm Vũng Tàu. Giả sử $x = x_1, x_2, x_3, \dots, x_n$ thể hiện mực nước giờ lớn nhất hàng năm của n biến ngẫu nhiên độc lập và phân phối giống nhau, hàm phân phối tích lũy 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, μ (location), σ (scale) và ξ (shape) thể hiện các tham số thống kê của hàm GEV. Khi

chuỗi số liệu được xem là có tính dừng, giá trị của các tham số là hằng số (Mô hình GEV-S). Trong trường hợp chuỗi số liệu được coi là không dừng, giá trị của tham số sẽ biến đổi theo biến số. Trong nghiên cứu này, tham số μ được biểu diễn như là một biến số theo thời gian (Mô hình GEV-NS):

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

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

Phương pháp ước lượng hợp lý cực đại đã và đang được sử dụng rộng rãi trong việc ước lượng các tham số của hàm GEV. Trong trường hợp chuỗi số liệu được coi là không dừng, hàm likelihood có thể được biểu diễn dưới dạng hàm của các tham số $\mu_0, \mu_1, \sigma, \xi$. Giả sử $x_1, x_2, \dots, x_n$ là chuỗi số liệu mực nước giờ lớn nhất của n năm, hàm log-likelihood có thể được viết như sau:

$$\begin{aligned}\text{Khi } \xi \neq 0, \\ L(\mu, \sigma, \xi | X) = -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log \left[1 + \xi \left(\frac{x_i - (\mu_0 + \mu_1 \times (t))}{\sigma}\right)\right] - \sum_{i=1}^n \left[1 + \xi \left(\frac{x_i - (\mu_0 + \mu_1 \times (t))}{\sigma}\right)\right]^{-\frac{1}{\xi}} \\ 1 + \xi \left(\frac{x_i - (\mu_0 + \mu_1 \times (t))}{\sigma}\right) > 0\end{aligned}\quad (5)$$

$$\begin{aligned}\text{Khi } \xi = 0, \\ L(\mu, \sigma | X) = -n \log \sigma - \sum_{i=1}^n \left(\frac{x_i - (\mu_0 + \mu_1 \times (t))}{\sigma}\right) - \sum_{i=1}^n \exp\left[-\left(\frac{x_i - (\mu_0 + \mu_1 \times (t))}{\sigma}\right)\right]\end{aligned}\quad (6)$$

Như vậy, cực đại hóa hàm hợp lý tương đương với cực đại hóa hàm logarithm của nó. Lúc này để ước lượng các tham số, bài toán trở thành tìm giá trị nhỏ nhất của hàm mất mát (negative log-likelihood) thay vì tìm giá trị lớn nhất (maximum-likelihood).

3.3 Lựa chọn mô hình thích hợp

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 mực nước cực đoạn. Hai chỉ số AIC và BIC được tính theo công thức sau:

$$AIC = -2(\log\text{likelihood}) + 2p \quad (7)$$

$$BIC = -2(\log\text{likelihood}) + p \ln(n) \quad (8)$$

Trong đó p là số lượng tham số trong mô hình, n là độ lớn mẫu.

Kiểm định likelihood ratio cho thấy xu hướng đáng kể trong chuỗi dữ liệu của mô hình mang tính không dừng so với mô hình mang tính dừng bằng cách so sánh negative log-likelihood của 2 mô hình đó. Giả sử mô hình mang tính dừng và không dừng được ký hiệu lần lượt là M_0 và M_1 , thì negative log-likelihood của 2 mô hình M_0 và M_1 được viết như sau:

$$l_0(M_0) = -\log L(x_1, x_2, \dots, x_n; \mu, \sigma, \xi) \quad (9)$$

$$l_1(M_1) = -\log L(x_1, x_2, \dots, x_n; \mu_0, \mu_1, \sigma, \xi) \quad (10)$$

Với giả thiết không có xu hướng trong chuỗi số liệu $\mu_1=0$, kiểm định likelihood ratio dựa trên 2 lần chênh lệch giữa $l_0(M_0)$ và $l_1(M_1)$ và tuân theo phân phối Chi bình phương với 1 bậc tự do, ký hiệu là χ^2 (1). Kiểm định likelihood ratio được viết như sau:

$$2\{l_0(M_0) - l_1(M_1)\} \sim \chi^2(1) \quad (11)$$

3.4 Tính toán giá trị mực nước thiết kế

Khi mô hình phù hợp nhất để mô phỏng tài liệu mực nước cực đoạn được lựa chọn, các giá trị mực nước cực đoạn (Z_T) tương ứng với chu kỳ lặp lại (T -year) 10, 50 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, chúng tôi dựa trên cách tiếp cận rủi ro thấp của Cheng & AghaKouchak (2014), bằng cách lấy 95% (95 percentile) giá trị của tham số μ (phương trình 12) để tính toán giá trị mực nước cực đoạn ứng với các chu kỳ lặp lại trong nghiên cứu này.

$$\hat{\mu}_{95} = Q_{95}(\hat{\mu}_{t1}, \hat{\mu}_{t2}, \dots, \hat{\mu}_{tn}) \quad (12)$$

Giá trị mực nước cực đoạn tương ứng với chu kỳ lặp lại T được đưa ra bởi Coles & cộng sự (2001) như sau:

$$Z_T = \begin{cases} \hat{\mu} - \frac{\sigma}{\xi} \left[1 - \left\{-\log\left(1 - \frac{1}{T}\right)\right\}^{\xi}\right], & \text{for } \xi \neq 0 \\ \hat{\mu} - \sigma \log\left\{-\log\left(1 - \frac{1}{T}\right)\right\}, & \text{for } \xi = 0 \end{cases} \quad (13)$$

Các tính toán trong nghiên cứu này được xử

lý bằng phần mềm R studio với ngôn ngữ lập trình R.

4. KẾT QUẢ

Bảng 1 thể hiện kết quả của kiểm định Mann-Kendall. Qua đó cho thấy chuỗi số liệu mực nước giờ lớn nhất mang xu thế tăng đáng kể qua giai đoạn 1980-2019, 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	369	7330	0.48	1.72×10^{-5}

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 (GEV-S) và không dừng (GEV-NS) qua phương pháp ước lượng hợp lý cực đại.

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

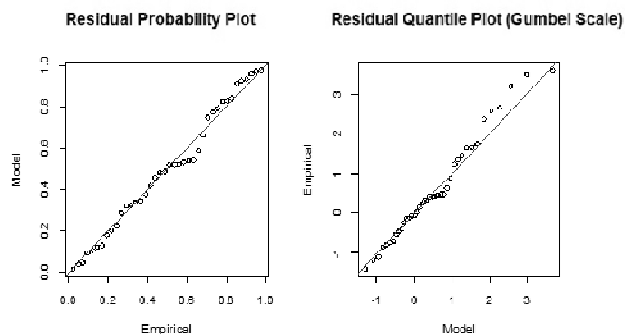
Tham số	Location		Scale	Shape
	μ_0	μ_1	σ	ξ
GEV-S	130.72		9.66	-0.35
GEV-NS	119.66	22.37	6.39	-0.15

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 mực nước cực đoan trong nghiên cứu này (Bảng 3). Kết quả của kiểm định likelihood ratio cũng chỉ ra rằng mô hình GEV-NS phù hợp hơn so với mô hình GEV-S với giá trị p-value là 1.4×10^{-5} . Bên cạnh đó, các biểu đồ xác suất Probability - Probability (PP) và phân vị Quantile - Quantile (QQ), dùng để kiểm tra sự phù hợp của mô hình được chọn, cũng cho thấy rằng mô hình GEV-NS cho kết quả giữa số liệu thực đo và mô hình tương đối phù hợp so với mô hình GEV-S.

Bảng 3. Giá trị của các chỉ số AIC và BIC

Chỉ số	GEV-S	GEV-NS
AIC	375	367
BIC	380	374

Bảng 4 thể hiện các giá trị mực nước cực đoan tương ứng với chu kỳ lặp lại là 10, 50 và 100 năm. Kết quả cho thấy rằng khi xét cùng một chu kỳ lặp lại, giá trị mực nước cực đoan từ mô hình GEV-NS lớn hơn từ mô hình GEV-S. Nói cách khác giả thiết về tính dừng trong chuỗi số liệu mực nước có thể dẫn đến việc đánh giá thấp các sự kiện mực nước cực đoan. Điều này có thể ảnh hưởng đến việc quy hoạch, thiết kế các công trình thủy lợi nói chung và công trình phòng chống ngập lụt nói riêng.



Hình 3. Residual PP và QQ plot cho mô hình GEV-NS



Hình 4. PP và QQ plot cho mô hình GEV-S

Bảng 4. Giá trị mực nước cực đoan (mm)

Chu kỳ lặp lại (năm)	10	50	100
GEV-S	145.79	151.33	152.87
GEV-NS	153.11	159.78	162.15

5. KẾT LUẬN

Trong nghiên cứu này, xu thế biến đổi của mực nước cực đoan tại trạm Vũng Tàu qua giai đoạn 1980-2019 đã được đánh giá qua kiểm định Mann-Kendall. Giả thiết về tính không dừng trong chuỗi số liệu mực nước cực đoan cũng được xem xét. Từ kết quả đạt được có thể đưa ra các kết luận sau: i) Mực nước cực đoan tại trạm Vũng Tàu có xu thế tăng đáng kể qua giai đoạn nghiên cứu; ii) Mô hình GEV-NS dựa trên hàm phân phối xác suất GEV và giả thiết về tính không dừng trong

chuỗi dữ liệu mực nước hoàn toàn phù hợp để mô phỏng mực nước cực đoan cho trạm Vũng Tàu; iii) Các giá trị mực nước thiết kế dựa trên giả thiết về tính không dừng (GEV-NS) lớn hơn so với các giá trị mực nước dựa trên giả thiết về tính dừng; iv). Từ các kết quả trên, mô hình GEV-NS được kiến nghị sử dụng rộng rãi trong khu vực nghiên cứu và vùng lân cận để cung cấp thông tin, dữ liệu cho các bài toán mô phỏng mực nước, mô phỏng mặn trong sông Vàm Cỏ hoặc lưu vực sông Mê Kông.

TÀI LIỆU THAM KHẢO

- Akaike, H. (1974). *A new look at the statistical model identification*. IEEE transactions on automatic control, 19(6), 716-723.
- Arns, A. & cộng sự. (2015). *The impact of sea level rise on storm surge water levels in the northern part of the German Bight*. Coastal Engineering, 96, 118-131.
- Arns, A. & cộng sự. (2013). *Estimating extreme water level probabilities: a comparison of the direct methods and recommendations for best practise*. Coastal Engineering, 81, 51-66.
- Bulteau, T. & cộng sự. (2015). *How historical information can improve estimation and prediction of extreme coastal water levels: application to the Xynthia event at La Rochelle (France)*. Natural Hazards and Earth System Sciences, 15(6), 1135-1147.
- 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. & cộng sự. (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.
- Mann, H. B. (1945). *Nonparametric Tests Against Trend*. Econometrica, 13(3), 245-259. doi:DOI: 10.2307/1907187
- Méndez, F. J. & cộng sự. (2007). *Analyzing monthly extreme sea levels with a time-dependent GEV model*. Journal of Atmospheric and Oceanic Technology, 24(5), 894-911.
- Menéndez, M. & Woodworth, P. L. (2010). *Changes in extreme high water levels based on a quasi - global tide - gauge data set*. Journal of Geophysical Research: Oceans, 115(C10).
- Mudersbach, C. & Jensen, J. (2010). *Nonstationary extreme value analysis of annual maximum water levels for designing coastal structures on the German North Sea coastline*. Journal of Flood risk management, 3(1), 52-62.
- Schwarz, G. (1978). *Estimating the dimension of a model*. The annals of statistics, 6(2), 461-464.

- Serafin, K. A.& Ruggiero, P. (2014). *Simulating extreme total water levels using a time - dependent, extreme value approach*. Journal of Geophysical Research: Oceans, 119(9), 6305-6329.
- Skjong, M. & cộng sự. (2013). *Statistics of Extreme Sea Levels for Locations along the Norwegian Coast*. Journal of Coastal Research, 29(5), 1029-1048.

Abstract:

NONSTATIONARY EXTREME VALUE ANALYSIS FOR ESTIMATION OF DESIGN WATER LEVEL OF VUNG TAU STATION

In this study, the trend in the extreme water level of Vung Tau station is indicated by Mann-Kendall test. Besides, the extreme water level time series is modelled under nonstationary condition. The results show that the increasing trend has been found in extreme water level data of Vung Tau station over the period of 1980-2019. Besides, the design water level estimates under the stationary condition are lower than those under the nonstationary condition in the study area.

Keywords: Extreme water level, Stationary, Design water level, Nonstationary, Vung Tau station

Ngày nhận bài: 21/12/2020

Ngày chấp nhận đăng: 28/12/2020