

## ƯỚC LƯỢNG CÁC PHÂN BỐ ĐỘ CAO SÓNG CÓ NGHĨA CHO NGHIÊN CỨU CHẾ ĐỘ SÓNG VÙNG BIỂN PHÚ YÊN

Trần Văn Chung\*, Cao Văn Nguyễn, Ngô Mạnh Tiến

Viện Hải Dương Học (IO), Viện Hàn Lâm Khoa học và Công nghệ Việt Nam (VAST)

\*Email: [tvanchung@gmail.com](mailto:tvanchung@gmail.com)

Ngày nhận bài: 14/03/2022; Ngày nhận đăng: 10/11/2022

### Tóm tắt

Trong công trình này, một số mô hình mô tả xác suất độ cao sóng được thử nghiệm và đánh giá, nhằm tìm ra hàm phân bố độ cao sóng điển hình cho vùng biển Phú Yên. Phân bố độ cao sóng được lựa chọn cuối cùng dựa trên nhiều tiêu chí và phân bố được đề xuất vượt trội hơn so với các mô hình tham số khác được thử nghiệm và mô tả một cách tối ưu.

**Từ khóa:** Độ cao sóng có nghĩa, phân bố xác suất, phân bố tích lũy, hoa sóng, các tiêu chí ước lượng.

### Probabilistic models of significant wave height deployed to study the coastal zone wave regimes in Phu Yen

Tran Van Chung, Cao Van Nguyen, Ngo Manh Tien

Institute of Oceanography (IO), Vietnam Academy of Science and Technology (VAST)

Received: March 14, 2022; Accepted: November 10, 2022

### Abstract

In this work, several models describing the wave height probability are tested and evaluated, aiming at finding the typical wave height distribution function in Phu Yen coastal zone. The final wave height distribution was selected based on multiple criteria and the proposed distribution outperformed the other parametric models tested and optimally described.

**Keywords:** Significant wave height, probability distribution, cumulative distribution function, wave rose, evaluation criteria.

### 1. Mở đầu

Thông tin về độ cao sóng trong thời gian dài khá quan trọng và được sử dụng nhiều trong thiết kế xác suất và đánh giá khả năng chịu tải của công trình biển, các tương tác ven biển với sóng biển, ước tính khả năng vận chuyển trầm tích theo thời gian dài và sự thay đổi hình thái của khu vực ven bờ, đánh giá năng lượng sóng, .... Đối với vùng ven bờ, sóng là yếu tố cơ bản quyết định đến dạng đường bờ, địa hình đáy, đến việc thiết kế các công trình cảng, luồng rạch và các công trình bảo vệ bờ biển. Có thể nói, độ cao sóng có nghĩa là một biến ngẫu nhiên thể hiện các thay đổi trong miền không gian và thời gian. Ngoài việc ước tính các tham số thống kê tiêu chuẩn của độ cao sóng, hàm mật độ xác suất tương ứng sẽ cung cấp cái nhìn sâu sắc hơn về các đặc tính chế độ sóng, do việc lựa chọn hàm mật độ xác

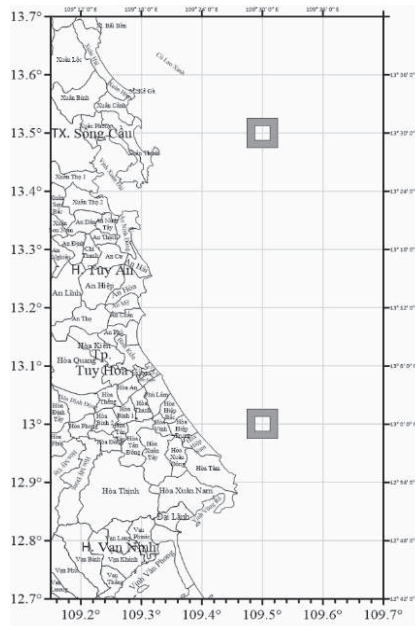
suất thích hợp đã mô tả phân tích tần số liên quan đến tác động trạng thái biển (Soukissian, 2021). Hơn nữa, như cũng được đề cập bởi Lucas và Soares (2015), có thể thuận tiện hơn nếu có các mô hình lý thuyết để kết hợp với các mô hình khác trong bất kỳ quy trình tính toán nào. Việc sử dụng mô hình xác suất tham số thay vì sử dụng trực tiếp dữ liệu rời rạc cũng được chứng minh bởi thực tế là mô hình lý thuyết có thể đại diện tốt hơn cho một số mức xác suất không có trong mẫu, do đó cho phép ngoại suy và lấp đầy dữ liệu. Theo giới thiệu của Soukissian (2021) thì thảo luận về độ không đảm bảo thống kê liên quan đến các thủ tục ước lượng của phân bố xác suất tham số cho độ cao sóng có nghĩa có thể được tìm thấy trong (Guedes Soares và Henriques, 1996).

Phú Yên với bờ biển dài 189 km, có nhiều vịnh, bãi, vũng, đầm phá, gành, ... Ngoài ra, Phú Yên có nhiều vùng bãi triều nước lợ, cửa sông, đầm phá, vịnh rất thuận lợi cho phát triển nuôi trồng thủy sản xuất khẩu. Do đó hiểu biết được quy luật sóng của vùng là việc làm thiết thực, trong nghiên cứu quá trình tương tác sóng biển trong vùng biển Phú Yên, phục vụ cho định hướng phát triển du lịch, cảnh báo các tác động do sóng đến vùng.

## **2. Phương pháp tính và thông tin nguồn số liệu sử dụng**

### **2.1. Đối tượng nghiên cứu cho vùng biển Phú Yên**

Thông tin về đặc trưng sóng (độ cao, chu kỳ và hướng) được cập nhật từ: [http://oos.soest.hawaii.edu/erddap/grid-dap/NWW3\\_Global\\_Best.html](http://oos.soest.hawaii.edu/erddap/grid-dap/NWW3_Global_Best.html) với độ phân giải 0,5 độ theo từng giờ (số liệu tính thống kê được chọn phân tích từ ngày 01/01/ 2011 đến tháng 31/12/2021 theo tần suất 01 giờ/số liệu (tương ứng với 11 năm). Đây là bộ số liệu trong chương trình hợp tác với NOAA/NCEP và NWS Honolulu, Đại học Hawaii đã triển khai mô hình 7 ngày WaveWatch III (WW3) quy mô toàn cầu với dự báo 5 ngày theo giờ ở độ phân giải khoảng 50 km hoặc 0,5 độ. Mô hình toàn cầu chịu sự tác động của gió của Hệ thống Dự báo Toàn cầu (GFS) của NOAA/NCEP. Mô hình này được thiết kế để thu nhận sóng biển quy mô lớn và cung cấp các điều kiện biên quang phổ cho Hawaii và các mô hình WW3 khu vực Thái Bình Dương khác. Bộ số liệu này đã được chúng tôi sử dụng cho đầu vào trong mô phỏng tương tác sóng-dòng vùng biển Nha Trang (Trần Văn Chung và cs., 2020).



**Hình 1.** Vị trí trích xuất dữ liệu sóng cho vùng biển Phú Yên

Trong vùng biển Phú Yên, chúng tôi phân tích hàm phân bố sóng từ 2 vị trí được lấy trung bình từ nguồn số liệu tính sóng liên tục trong 11 năm theo tần suất 01 giờ/số liệu. Vị trí nghiên cứu chế độ tại vùng biển Phú Yên được trích xuất từ chuỗi số liệu sóng trong 11 năm tại khu vực thể hiện trên Hình 1.

## 2.2. Phương pháp nghiên cứu

Để tìm quy luật phân bố sóng tại vùng biển Phú Yên, chúng tôi đã sử dụng 11 hàm phân bố có 02 tham số điển hình với hình dạng phân bố độ cao sóng vùng nghiên cứu và dựa vào các tiêu chí thống kê để đánh giá độ chính xác nhằm tìm 01 hàm phân bố tối ưu có thể thỏa mãn quy luật phân bố sóng tại khu vực.

### 2.2.1. Tổng quan các loại hàm phân bố

#### 2.2.1.1. Phân bố chuẩn (normal distribution)

Phân bố chuẩn, còn gọi là phân bố Gauss, là một phân bố xác suất cực kì quan trọng trong nhiều lĩnh vực (Johnson và cs., 2002).

$$\text{- Hàm mật độ xác suất chuẩn: } f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

trong đó  $\mu$  là giá trị trung bình,  $\sigma$  là độ lệch chuẩn, xác định cụ thể như sau:  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ;

$$\sigma = \left[ \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2}$$

$$\text{- Hàm phân bố tích lũy chuẩn: } F(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

#### 2.2.1.2. Phân bố loga chuẩn (LogNormal distribution)

Phân bố loga chuẩn, đôi khi được gọi là phân bố Galton (Johnson và cs., 1994), là một phân bố xác suất mà logarit của nó có phân bố chuẩn. Các tham số phân bố chuẩn được xác định:  $\mu = \log(m^2/\sqrt{v} + m^2)$ ;  $\sigma = \sqrt{\log(v/m^2 + 1)}$

Trong đó  $\mu$ : giá trị trung bình của lô-ga-rit;  $\sigma$ : độ lệch chuẩn của giá trị logarit;  $m$ : giá trị trung bình và  $v$ : phương sai .

Theo Johnson và cs. (1994) và Weisstein (2020) các hàm của phân bố lognormal được xác định như sau:

$$\text{Hàm mật độ xác suất: } f(x|\mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}$$

$$\text{Hàm phân bố tích lũy: } F(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^x \frac{1}{t} \exp\left\{-\frac{(\log t - \mu)^2}{2\sigma^2}\right\} dt, \text{ cho } x > 0$$

### 2.2.1.3. Phân bố Gamma

Theo Hogg và cs. (2005) và Olver (2010), hàm phân bố Gama được xác định như sau:

$$\text{- Hàm mật độ xác suất: } f(x|a, b) = \frac{1}{b^a\Gamma(a)} x^{a-1} e^{-\frac{x}{b}}$$

trong đó: a là tham số tỷ lệ; b là tham số hình dạng.

$$\text{Hàm gamma được xác định bởi tích phân: } \Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$$

$$\text{- Hàm phân bố tích lũy gamma: } F(x|a, b) = \frac{1}{b^a\Gamma(a)} \int_0^x t^{a-1} e^{-\frac{t}{b}} dt$$

### 2.2.1.4. Phân bố Weibull

Hàm phân bố Weibull được viết (Johnson và cs., 1994):

$$\text{- Hàm mật độ xác suất: } f(x|a, b) = \frac{b}{a} \left(\frac{x}{a}\right)^{b-1} e^{-\left(\frac{x}{a}\right)^b}$$

$$\text{- Hàm phân bố tích lũy: } F(x|a, b) = 1 - e^{-\left(\frac{x}{a}\right)^b}$$

Trong đó a là tham số tỉ lệ và b là tham số hình dạng.

### 2.2.1.5. Phân bố Rayleigh

Phân bố Rayleigh là một trường hợp đặc biệt của phân bố Weibull (Papoulis và Pillai, 2001). Nếu A và B là các tham số của phân bố Weibull, thì phân bố Rayleigh với tham số b tương đương với phân bố Weibull với tham số  $A = \sqrt{2}b$  và  $B=2$

$$\text{- Hàm mật độ xác suất: } f(x|b) = \frac{x}{b^2} e^{-\frac{x^2}{2b^2}}$$

$$\text{- Hàm phân bố tích lũy: } F(x|b) = 1 - e^{-x^2/(2b^2)}$$

### 2.2.1.6. Hàm phân bố Birnbaum-Saunders:

Birnbaum và Saunders (1969) đã giới thiệu hàm phân bố hai tham số mô tả tuổi thọ của kim loại, chịu tác động theo chu kỳ. Phân bố Birnbaum-Saunders (Leiva, 2016) được hiển thị bên dưới:

$$\text{- Hàm mật độ xác suất: } f(x) = \frac{\left(\frac{\sqrt{x} + \sqrt{\beta}}{\sqrt{\beta} + \sqrt{x}}\right)}{2\gamma x} \phi\left(\frac{\left(\frac{\sqrt{x} - \sqrt{\beta}}{\sqrt{\beta} + \sqrt{x}}\right)}{\gamma}\right)$$

Với tham số tỷ lệ  $\beta > 0$  và tham số hình dạng  $\gamma > 0$ , đối với  $x > 0$

$$\text{- Hàm phân bố tích lũy: } F(x) = \Phi\left(\frac{\sqrt{x} - \sqrt{\beta}}{\gamma}\right), x > 0, \gamma > 0$$

$\Phi$  là hàm phân bố tích lũy của phân bố chuẩn chuẩn hóa (standard normal distribution)

được xác định:  $\Phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

### 2.2.1.7. Phân bố Gaussian nghịch đảo (Inverse Gaussian Distribution)

Phân bố Gaussian nghịch đảo còn được gọi là phân bố Wald. Các phân bố Gaussian

ngược đảo có nhiều điểm giống với các phân bố Gaussian chuẩn (chuẩn tắc), dẫn ra từ các ứng dụng trong suy luận thống kê (Flocks and Chhikara, 1978).

- Hàm mật độ xác suất:  $f(x|\mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{-\frac{\lambda}{2\mu^2 x}(x - \mu)^2\right\}$

Trong đó  $\mu > 0$  là tham số trung bình và  $\lambda > 0$  là tham số tỷ lệ

- Hàm phân bố tích lũy:

$$F(x|\mu, \lambda) = \frac{1}{2} \left\{ 1 + \operatorname{erf} \left[ \sqrt{\frac{\lambda}{2x}} \left( \frac{x}{\mu} - 1 \right) \right] \right\} + \frac{1}{2} e^{2\lambda/\mu} \left\{ 1 - \operatorname{erf} \left[ \sqrt{\frac{\lambda}{2x}} \left( \frac{x}{\mu} + 1 \right) \right] \right\}$$

Trong đó: erf là hàm sai số. Hàm erf của x sẽ được viết:  $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$

#### 2.2.1.8. Phân bố Logistic

Phân bố logistic được sử dụng cho các mô hình tăng trưởng và trong hồi quy logistic. Nó có phần đuôi dài hơn và kurtosis cao hơn so với phân bố chuẩn (von Seggern, 1993).

- Hàm mật độ xác suất:  $f(x|\mu, \sigma) = \frac{\exp\left\{\frac{x-\mu}{\sigma}\right\}}{\sigma(1+\exp\left\{\frac{x-\mu}{\sigma}\right\})^2}$

- Hàm phân bố tích lũy:  $F(x|\mu, \sigma) = \frac{1}{1+\exp\left\{\frac{x-\mu}{\sigma}\right\}}$

$\mu$  là giá trị trung bình;  $\sigma$  là tham số tỷ lệ

#### 2.2.1.9. Phân bố Log-Logistics

Phân bố *loglogistic* là một phân bố xác suất mà logarit của nó có phân bố logistic, còn được gọi là phân bố Fisk trong các ứng dụng kinh tế học. Phân bố này được sử dụng trong thủy văn để mô hình hóa dòng chảy và lượng mưa. Hàm mật độ và hàm tích lũy của phân bố Log-Logistics được thể hiện như dưới đây (Shoukri và cs., 1988; Ashkar và Mahdi (2006))

- Hàm mật độ xác suất:  $f(x|\mu, \sigma) = \frac{1}{\sigma} \frac{1}{x} \frac{e^z}{(1+e^z)^2}; x \geq 0,$

Trong đó:  $z = \frac{\log(x)-\mu}{\sigma}$ ;  $\mu$ : giá trị trung bình của lô-ga-rit;  $\sigma$ : tham số tỷ lệ của các giá trị logarit

- Hàm phân bố tích lũy:  $F(x|\mu, \sigma) = \frac{1}{1+\left(\frac{x}{\sigma}\right)^{-\mu}}$

#### 2.2.1.10. Phân bố "Rician":

Trong lý thuyết xác suất, phân bố Rice hoặc phân bố Rician (hoặc, ít phổ biến hơn, phân bố Ricean) là phân bố xác suất về độ lớn của một biến ngẫu nhiên chuẩn hai biến đối xứng tròn, có thể có giá trị trung bình khác 0 (không trung tâm).

- Hàm mật độ xác suất:  $f(x|s, \sigma) = I_0\left(\frac{xs}{\sigma^2}\right) \frac{x}{\sigma^2} e^{-\left(\frac{x^2+s^2}{2\sigma^2}\right)}$

với tham số không trung tâm  $s \geq 0$  và tham số tỷ lệ  $\sigma > 0$ , với  $x > 0$ .  $I_0$  là hàm Bessel được sửa đổi bậc 0 của loại đầu tiên.

- Hàm phân bố tích lũy:  $F(x|s, \sigma) = 1 - Q_1\left(\frac{s}{\sigma}, \frac{x}{\sigma}\right)$

Trong đó  $Q_1$  là hàm Marcum Q [Cantrel và Ojha (1987); Shnidman (1989)]

$$Q_s(a, b) = \frac{1}{a^{s-1}} \int_b^\infty x^s \exp\left(-\frac{x^2 + a^2}{2}\right) I_{s-1}(ax) dx$$

Trong đó  $b \geq 0$  và  $a, s > 0$  và  $I_{s-1}$  là hàm Bessel điều chỉnh của bậc đầu tiên  $s-1$ .

$$I_{s-1}(ax) \text{ được xác định bởi: } I_{s-1}(ax) = \left(\frac{ax}{2}\right)^{s-1} \sum_{k=0}^{\infty} \frac{\left(\frac{-(ax)^2}{4}\right)^k}{k! \Gamma(s+k)}$$

### 2.2.1.11. Hàm phân bố Nakagami

Hàm phân bố Nakagami được thể hiện theo Laurenson (1994) như sau:

- Hàm mật độ xác suất:  $f(x|\mu, \omega) = 2 \left(\frac{\mu}{\omega}\right)^\mu \frac{1}{\Gamma(\mu)} x^{(2\mu-1)} e^{-\frac{\mu}{\omega}x^2}$

Với tham số hình dạng  $\mu$  và tham số tỷ lệ  $\omega > 0$ , cho  $x > 0$ .

- Hàm phân bố tích lũy:  $F(x|\mu, \omega) = \frac{\gamma(\mu, \frac{\mu}{\omega}x^2)}{\Gamma(\mu)}$

Với  $\gamma$  là hàm gamma không hoàn chỉnh dưới:

$$\gamma(\mu, \frac{\mu}{\omega}x^2) = \left(\frac{\mu}{\omega}x^2\right)^\mu \Gamma(\mu) e^{-\left(\frac{\mu}{\omega}x^2\right)} \sum_{k=0}^{\infty} \frac{\left(\frac{\mu}{\omega}x^2\right)^k}{\Gamma(\mu+k+1)}; \Gamma(\cdot) \text{ là hàm Gamma}$$

### 2.2.2. Ước lượng các tham số

Chúng tôi đã sử dụng cách thức thực hiện đã được giới thiệu bởi Dookie và cs. (2018) với một số kỹ thuật được đưa ra để ước lượng tham số (điền hình Carta và cs., 2009). Trong công trình này, các tham số cho các phân bố khác nhau này được ước tính bằng cách sử dụng phương pháp khả năng xảy ra tối đa, phương pháp này chọn làm giá trị ước tính của nó là giá trị tham số tối đa hóa xác suất của dữ liệu quan sát (Leon-Garcia, 2008). Phương pháp này được sử dụng phổ biến vì các công cụ ước lượng kết quả thường không chệch và phù hợp về mặt tiệm cận. Phương pháp này cũng cung cấp lợi thế của sự đơn giản trong thực hiện. Trong bảng 1 trình bày các tham số của các loại phân bố.

**Bảng 1:** Các hàm phân bố có 02 tham số được mô tả

| STT | Loại phân bố      | Tham số 1 (A)                        | Tham số 2 (B)                                  |
|-----|-------------------|--------------------------------------|--|
| 1   | Normal            | Giá trị trung bình $\mu$             | Độ lệch chuẩn $\sigma$                         |
| 2   | Lognormal         | Giá trị trung bình của lôgarit $\mu$ | Độ lệch chuẩn của các giá trị lôgarit $\sigma$ |
| 3   | Gamma             | Tham số hình dạng $a$                | Tham số tỷ lệ $b$                              |
| 4   | Weibull           | Tham số tỷ lệ $a$                    | Tham số hình dạng $b$                          |
| 5   | Rayleigh          | Tham số tỷ lệ $b$                    |  |
| 6   | Birnbaum-Saunders | Tham số tỷ lệ $\beta$                | Tham số hình dạng $\gamma$                     |
| 7   | Inverse Gaussian  | Tham số tỷ lệ $\mu$                  | Tham số hình dạng $\lambda$                    |
| 8   | Logistic          | Giá trị trung bình $\mu$             | Tham số tỷ lệ $\sigma$                         |
| 9   | Loglogistic       | Giá trị trung bình của lôgarit $\mu$ | Tham số tỷ lệ của giá trị lôgarit $\sigma$     |
| 10  | Rician            | Tham số không trung tâm $s$          | Tham số tỷ lệ $\sigma$                         |
| 11  | Nakagami          | tham số hình dạng $\mu$              | tham số tỷ lệ $\omega$                         |

### 2.2.3. Tiêu chí thống kê được sử dụng để đánh giá tính phù hợp

Để đánh giá mức độ phù hợp của các hàm phân bố nói trên trong việc mô tả xác suất độ cao sóng, một số thông số kiểm định được sử dụng và mô tả ngắn gọn như sau:



### 2.2.3.1. Sai số độ lệch tuyệt đối trung bình (MABE)

MABE cung cấp đại lượng trung bình của tổng sai số độ lệch tuyệt đối giữa tần số ước tính và quan sát của độ cao sóng có nghĩa được đưa ra bởi:  $MABE = \frac{1}{n} \sum_{i=1}^n |E_i - O_i|$

### 2.2.3.2. Căn bậc hai của sai số bình phương trung bình (RMSE)

RMSE đại diện cho độ chính xác của phân bố bằng cách đo sự không phù hợp trung bình giữa các giá trị của tần suất ước tính và quan sát của độ cao sóng có nghĩa. Nó được đưa ra bởi:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (E_i - O_i)^2}$$

Giá trị cao của RMSE cho biết phân bố ước tính có vấn đề và giá trị nhỏ cho thấy phân bố ước tính phù hợp với dữ liệu.

### 2.2.3.3. Hệ số tương quan ( $R^2$ )

Hệ số xác định  $R^2$  xác định mối quan hệ tuyến tính giữa các giá trị tính toán từ phân bố Weibull và các giá trị tính toán từ dữ liệu đo được. Giá trị hệ số tương quan ( $R^2$ ) cho thấy dữ liệu đủ gần để khớp với đường hồi quy. Mối tương quan có thể được viết là:

$$R^2 = \frac{\sum_{i=1}^n (E_i - O_i)^2 - \sum_{i=1}^n (E_i - \bar{E})(O_i - \bar{O})}{\sum_{i=1}^n (O_i - \bar{O})^2}$$

Giá trị của  $R^2$  tiến tới giá trị 1 (đạt đến tương quan hoàn hảo), trong khi giá trị 0 cho thấy hai tập dữ liệu hoàn toàn khác nhau.

### 2.2.3.4. Kiểm định Chi-bình phương ( $\chi^2$ )

Kiểm định Chi-bình phương ( $\chi^2$ ) được sử dụng để xác định xem có tồn tại bất kỳ sự khác biệt nào giữa tần số mong đợi và tần số quan sát được hay không. Trong kiểm định giả thuyết thống kê này, phân bố là phân bố chi-bình phương khi giả thuyết rỗng là đúng. Biểu thức chi-square như được hiển thị bên dưới trong phương trình:  $\chi^2 = \sum_{i=1}^n \left[ \frac{(E_i - O_i)^2}{O_i} \right]$

## 3. Kết quả nghiên cứu

Để phân tích bộ số liệu sóng, bài báo đã phân tích bộ dữ liệu sóng trung bình tại 2 vị trí thuộc trong phạm vi vùng biển Phú Yên với tần suất 01 giờ/số liệu bắt đầu từ ngày 01/01/2011 đến 31/12/2021. Toàn bộ sản phẩm phân tích về hướng và độ cao sóng được thể hiện khá chi tiết trên bảng 2 và dạng hoa sóng (hình 2).

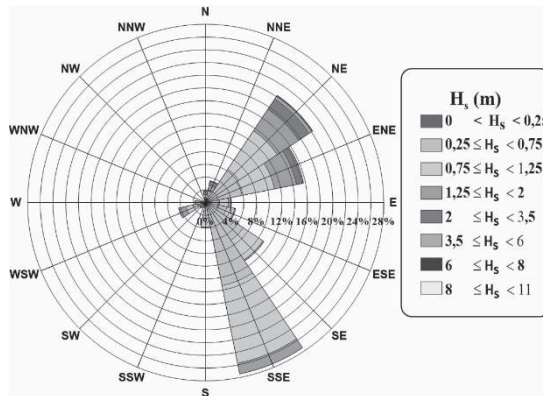
**Bảng 2.** Bảng phân cấp chi tiết độ cao sóng có nghĩa tại vùng biển Phú Yên

| Hướng sóng | $0 < H_s < 0,25$ | $0,25 \leq H_s < 0,75$ | $0,75 \leq H_s < 1,25$ | $1,25 \leq H_s < 2$ | $2 \leq H_s < 3,5$ | $3,5 \leq H_s < 6$ | $6 \leq H_s < 8$ | $8 \leq H_s < 11$ | $H_{th}$ (m) | Tổng số liệu | Chiếm tỷ lệ (%) |
|------------|------------------|------------------------|------------------------|---------------------|--------------------|--------------------|------------------|-------------------|--------------|--------------|-----------------|
| N          | 1                | 39                     | 116                    | 126                 | 55                 | 3                  | 0                | 0                 | 1.4          | 340          | 1.4             |
| NNE        | 4                | 175                    | 278                    | 212                 | 165                | 17                 | 1                | 0                 | 1.4          | 852          | 3.6             |
| NE         | 32               | 1411                   | 1847                   | 1028                | 489                | 78                 | 15               | 0                 | 1.2          | 4900         | 20.4            |
| ENE        | 118              | 1396                   | 1219                   | 840                 | 185                | 94                 | 9                | 0                 | 1.1          | 3861         | 16.1            |
| E          | 17               | 624                    | 235                    | 66                  | 21                 | 9                  | 6                | 1                 | 0.8          | 979          | 4.1             |
| ESE        | 11               | 902                    | 218                    | 18                  | 1                  | 0                  | 0                | 0                 | 0.6          | 1150         | 4.8             |
| SE         | 6                | 1654                   | 939                    | 36                  | 0                  | 0                  | 0                | 0                 | 0.7          | 2635         | 11.0            |
| SSE        | 7                | 2761                   | 3343                   | 431                 | 6                  | 0                  | 0                | 0                 | 0.8          | 6548         | 27.3            |
| S          | 8                | 800                    | 133                    | 9                   | 0                  | 0                  | 0                | 0                 | 0.6          | 950          | 4.0             |
| SSW        | 0                | 31                     | 18                     | 1                   | 0                  | 0                  | 0                | 0                 | 0.7          | 50           | 0.2             |

| Hướng sóng      | $0 < H_s < 0,25$ | $0,25 \leq H_s < 0,75$ | $0,75 \leq H_s < 1,25$ | $1,25 \leq H_s < 2$ | $2 \leq H_s < 3,5$ | $3,5 \leq H_s < 6$ | $6 \leq H_s < 8$ | $8 \leq H_s < 11$ | $H_{tb}$ (m) | Tổng số liệu | Chiếm tỷ lệ (%) |
|-----------------|------------------|------------------------|------------------------|---------------------|--------------------|--------------------|------------------|-------------------|--------------|--------------|-----------------|
| SW              | 0                | 79                     | 126                    | 36                  | 8                  | 4                  | 1                | 0                 | 1.1          | 254          | 1.1             |
| WSW             | 0                | 143                    | 558                    | 317                 | 16                 | 0                  | 0                | 0                 | 1.1          | 1034         | 4.3             |
| W               | 0                | 52                     | 208                    | 145                 | 2                  | 0                  | 0                | 0                 | 1.1          | 407          | 1.7             |
| WNW             | 0                | 0                      | 0                      | 0                   | 0                  | 0                  | 0                | 0                 | 0            | 0            | 0.0             |
| NW              | 0                | 1                      | 4                      | 0                   | 0                  | 0                  | 0                | 0                 | 0.9          | 5            | 0.0             |
| NNW             | 0                | 5                      | 6                      | 2                   | 0                  | 0                  | 0                | 0                 | 0.8          | 13           | 0.1             |
| Cấp sóng        | I                | II                     | III                    | IV                  | V                  | VI                 | VII              | VIII              |              | 23978        | 100.0           |
| Tổng số liệu    | 204              | 10073                  | 9248                   | 3267                | 948                | 205                | 32               | 1                 |              |              |                 |
| Chiếm tỷ lệ (%) | 0.9              | 42.0                   | 38.6                   | 13.6                | 4.0                | 0.9                | 0.1              | 0.0               | 100.0        |              |                 |

\*) Ghi chú:  $H_s$ : độ cao sóng có nghĩa (m);  $H_{tb}$ : độ cao sóng có nghĩa được tính trung bình

Kết quả phân tích trên Bảng 2 và Hình 2, cho thấy, độ cao sóng theo cấp sóng có 3 cấp có tần xuất xuất hiện nhiều nhất (trên 13%), đó là cấp sóng cấp II ( $0,25 \leq H_s < 0,75$ )(m) với 10.073 số liệu được tìm thấy chiếm tới 42,0%; tiếp theo là độ cao sóng với cấp độ sóng III ( $0,75 \leq H_s < 1,25$ )(m) với 9.248 số liệu, chiếm 38,6% và độ cao sóng cấp độ IV ( $1,25 \leq H_s < 2$ )(m) với 3.267 số liệu, chiếm 13,6%. Về đặc trưng hướng sóng phổ biến, chế độ sóng vùng biển Phú Yên chịu ảnh hưởng tính khu vực thể hiện khá rõ qua các hướng sóng thường xuyên xuất hiện (Hình 2). Cụ thể, qua các hướng sóng điển hình cho mùa gió Đông Bắc với 02 hướng chính là Bắc Đông Bắc (27,3%) và Đông Bắc (20,4%), trong khi tác động của mùa gió Tây Nam với 2 hướng chính thể hiện qua 02 hướng sóng: Nam Đông Nam (27,3%) và Đông Nam (11,0%) (Hình 2, Bảng 2).



Hình 2. Hoa sóng được phân tích tại vùng biển Phú Yên

Trong 11 năm gần đây (2011 – 2021), theo kết quả phân tích chỉ có 01 lần độ cao sóng đạt cấp XIII (sóng rất mạnh), sự xuất hiện độ cao sóng lịch sử là do ảnh hưởng của cơn bão số 12 (cơn bão Damrey) xảy ra vào lúc 07 giờ 04/11/2017 với độ cao sóng có nghĩa đạt 8,4m, hướng sóng  $91.5^0$  và chu kỳ sóng 13s.

Bảng 3. Đánh giá tính phù hợp qua các hàm phân bố theo các tiêu chí thống kê

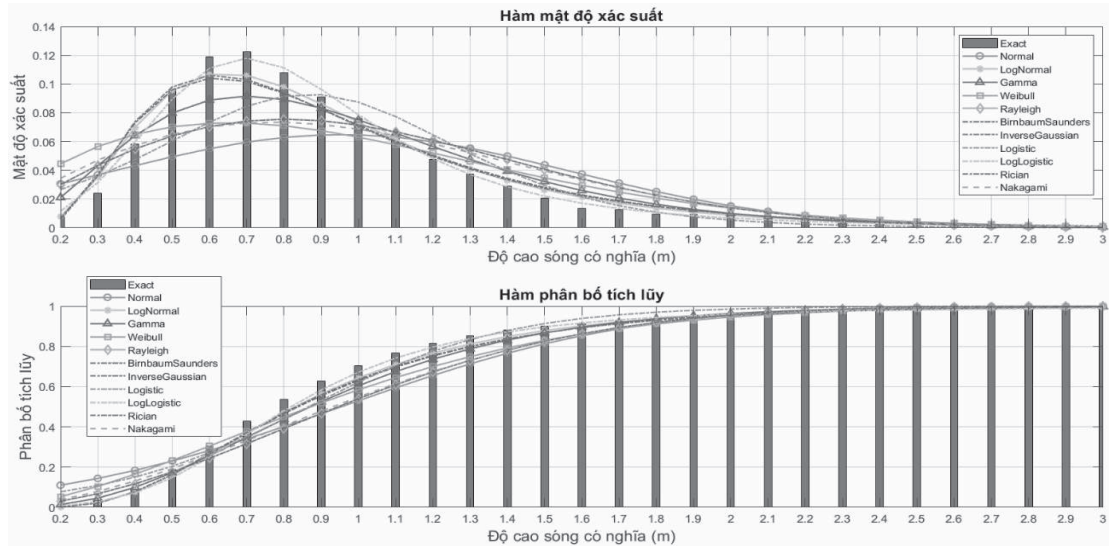
|      |        |           |       |         |          |                  |                 |          |             |        |          |
|------|--------|-----------|-------|---------|----------|------------------|-----------------|----------|-------------|--------|----------|
| Type | Normal | LogNormal | Gamma | Weibull | Rayleigh | BirnbaumSaunders | InverseGaussian | Logistic | LogLogistic | Rician | Nakagami |
|------|--------|-----------|-------|---------|----------|------------------|-----------------|----------|-------------|--------|----------|



|              |       |       |       |        |       |       |        |       |        |       |         |
|--------------|-------|-------|-------|--------|-------|-------|--------|-------|--------|-------|---------|
| (1)          |       |       |       |        |       |       |        |       |        |       |         |
| A            | 6,148 | 0,508 | 2,554 | 1,729  | 8,027 | 0,531 | 31,624 | 2,689 | 0,280  | 8,026 | 128,875 |
| B            | 9,543 | 2,116 | 3,736 | 10,785 | 0,000 | 8,370 | 9,543  | 8,663 | 2,102  | 0,266 | 0,929   |
| MABE         | 0,015 | 0,004 | 0,007 | 0,012  | 0,012 | 0,005 | 0,004  | 0,009 | 0,002  | 0,012 | 0,012   |
| RMSE         | 0,130 | 0,031 | 0,062 | 0,104  | 0,101 | 0,040 | 0,038  | 0,083 | 0,016  | 0,101 | 0,103   |
| R2           | 0,644 | 0,979 | 0,919 | 0,769  | 0,785 | 0,966 | 0,969  | 0,856 | 0,994  | 0,785 | 0,776   |
| ( $\chi^2$ ) | 0,020 | 0,001 | 0,004 | 0,013  | 0,011 | 0,002 | 0,002  | 0,008 | 0,0003 | 0,011 | 0,012   |

\*) **Ghi chú:** Chữ **in đậm** là thỏa mãn tốt nhất và **chữ in nghiêng** là ít thỏa mãn nhất. <sup>(1)</sup>: giá trị A và B được ghi chú chi tiết trên Bảng 1

Căn cứ vào kết quả đánh giá sai số trên Bảng 3 và thể hiện trên dạng biểu đồ cho 11 phương pháp số được phân tích. Trên cơ sở 4 tiêu chí đánh giá gồm: Sai số độ lệch tuyệt đối trung bình (MABE) (lựa chọn phù hợp với giá trị nhỏ nhất trong 11 phương pháp số); căn bậc hai của sai số bình phương trung bình (RMSE) (chọn giá trị nhỏ nhất); hệ số tương quan ( $R^2$ ) (lựa chọn giá trị gần với “1” nhất) và kiểm định Chi-bình phương ( $\chi^2$ ) (chọn giá trị nhỏ nhất). Kết quả cho thấy: Hàm phân bố theo **LogLogistic** phù hợp nhất cho trong khi đó hàm phân bố chuẩn (**Normal** distribution) cho kết quả không tốt.



**Hình 3.** Các dạng hàm phân bố của độ cao sóng cho vùng biển Phú Yên (2011 – 2021)

**4. Kết luận**

Kết quả phân tích chế độ sóng cho vùng biển Phú Yên trong 11 năm gần nhất (2011 – 2021), thể hiện qua độ cao sóng điển hình trong vùng biển Phú Yên dưới 2 m chiếm tới 95,1% số liệu được phân tích. Độ cao sóng thể hiện sự ảnh hưởng khá rõ của chế độ gió mùa, với gió trường gió mùa Đông Bắc, khu vực chịu ảnh hưởng của 02 hướng sóng chính là Bắc Đông Bắc (27,3%) và Đông Bắc (20,4%). Trong khi đó, ảnh hưởng của mùa gió Tây Nam được thể hiện qua 2 hướng sóng chủ đạo là Nam Đông Nam (27,3%) và Đông Nam (11,0%).

Bài báo đã cung cấp được bộ dữ liệu các tham số phục vụ cho dự báo chế độ sóng cho vùng biển Phú Yên theo 11 dạng phân bố hai tham số. Tùy vào mức độ ứng dụng có thể chọn lựa các hàm phân bố phù hợp cho dự báo. Tuy nhiên, ở góc độ chuyên môn và từ những phân tích ở trên, đối với phân bố sóng cho vùng biển Phú Yên nên chọn hàm phân bố LogLogistic để làm điều kiện ban đầu ngoài khơi cho mô phỏng dự báo chế độ sóng sóng đổ ven bờ, không nên dùng hàm phân bố chuẩn để dự báo độ cao sóng cho vùng Phú Yên.

**Lời cảm ơn:** Nhóm tác giả xin cảm ơn: Đề tài nhiệm vụ KH & CN tỉnh Phú Yên “Xây dựng bản đồ nhạy cảm môi trường vùng ven bờ phục vụ các sự cố tràn dầu tỉnh Phú Yên”, đã cung

cấp một phần kinh phí cho nghiên cứu; các đồng nghiệp đã góp ý và hỗ trợ giúp chúng tôi hoàn thành bài báo này □

#### TÀI LIỆU THAM KHẢO

- Ashkar, F. and Mahdi, S. (2006). *Fitting the log-logistic distribution by generalized moments*. Journal of Hydrology, 328 (3–4): 694–703.
- Birnbaum, Z. W. and Saunders, S. C. (1969). *A new family of life distributions*. Journal of Applied Probability, 6 (2): 319–327, doi:10.2307/3212003, JSTOR 3212003.
- Cantrell, P. E., and A. K. Ojha (1987). *Comparison of Generalized Q-Function Algorithms*. IEEE Transactions on Information Theory, Vol. IT-33, July, 1987, pp. 591–596.
- Carta, J.A., Ramirez, P., Velazquez, S. (2009). *A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands*. Sustain. Energy Rev. 13(5), 933-966.
- Dookie, I., Rocke, S., Singh, A., Ramlal, C. J. (2018). *Evaluating wind speed probability distribution models with a novel goodness of fit metric: a Trinidad and Tobago case study*. International Journal of Energy and Environmental Engineering 9(3): 323-339.
- Floks, J.L. and R.S. Chhikara (1978). *The Inverse Gaussian Distribution and its Statistical Application – A Review*. Journal of the Royal Statistical Society. Series B, Vol. 40, No. 3, pp. 263-289.
- Guedes-Soares, C. and Henriques, A.C. (1996). *Statistical uncertainty in long-term distributions of significant wave height*. J. Offshore Mech. Arct. Eng., 11, 284–291.
- Hogg, R.V., McKean, J.W., Craig, A.T. (2005). *Introduction to mathematical statistics*. Pearson Education, Prentice Hall.
- Johnson, N. L.; Kotz, S.; Balakrishnan, N. (1994). *14: Lognormal Distributions, Continuous univariate distributions. Vol. 1*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics (2nd ed.), New York: John Wiley & Sons, ISBN 978-0-471-58495-7, MR 1299979
- Johnson, N.L., Kotz, S., Balakrishnan, N. (1994). *Continuous univariate distributions, vol. 1*. Wiley, New York, p. 163 (1994)
- Johnson, N.L., Kotz, S., Balakrishnan, N. (2002). *Continuous multivariate distributions, vol. 1, models and applications, vol. 59*. Wiley, New York (2002).
- Laurenson, D. (1994). *Nakagami Distribution*. Indoor Radio Channel Propagation Modelling by Ray Tracing Techniques. Retrieved 2007-08-04.
- Leiva, V (2016). *The Birnbaum-Saunders Distribution*. Academic Press, Cambridge (book section Genesis of the Birnbaum-Saunders Distribution, pp. 1-15)
- Leon-Garcia, A. (2008). *Probability, statistics and random processes for electrical engineering*. Pearson.
- Lucas, C., Guedes Soares, C. (2015). *Bivariate distributions of significant wave height and mean wave period of combined sea states*. Ocean. Eng. 106, 341-353.
- Olver, F.W. (2010). *NIST Handbook of Mathematical Functions Hardback and CD-ROM*. Cambridge University Press, Cambridge.
- Papoulis, A. and Pillai, S. (2001). *Probability, Random Variables and Stochastic Processes*. ISBN 0073660116, ISBN 9780073660110.
- Shnidman, D. A. (1989). *The Calculation of the Probability of Detection and the Generalized Marcum Q-Function*. IEEE Transactions on Information Theory, Vol. IT-35, March, 1989, pp. 389–400.
- Shoukri, M.M.; Mian, I.U.M.; Tracy, D.S. (1988). *Sampling Properties of Estimators of the Log-Logistic Distribution with Application to Canadian Precipitation Data*. The Canadian Journal of Statistics, 16 (3): 223–236.
- Soukissian, T. H. (2021). *Probabilistic modelling of significant wave height using the extended generalized inverse Gaussian distribution*. Ocean Engineering 230: 109061.
- Trần Văn Chung, Ngô Mạnh Tiến và Võ Văn Quang (2020). *Phát triển mô hình sóng-dòng qua cặp mô hình FEM và SWAN tại vịnh Nha Trang*. Tạp chí Khoa học & Công nghệ Đại học Duy Tân, ISSN 1859-4905, 06(43), 52-59.
- Von Seggern, D. (1993). *CRC Standard Curves and Surfaces*. Boca Raton, FL: CRC Press, p. 250, 1993.
- Weisstein, E.W. (2020). *Log Normal Distribution*. Mathworld.wolfram.com. Retrieved 2020-09-13.