

SO SÁNH MƯỜI MỘT PHƯƠNG PHÁP PHÂN BỐ TỐC ĐỘ GIÓ KHÁC NHAU TRÊN CƠ SỞ ĐÁNH GIÁ MỨC ĐỘ PHÙ HỢP CHO NGHIÊN CỨU CHẾ ĐỘ GIÓ VỊNH NHA TRANG

Trần Văn Chung*, Ngô Mạnh Tiến, Nguyễn Hoàng Thái Khang
Viện Hải dương học, Viện Hàn lâm KHCNVN
* tvanchung@gmail.com

Tóm tắt. Trong công trình này, mười một dạng hàm phân bố để mô tả xác suất tốc độ gió được thử nghiệm và đánh giá, tìm ra hàm phân bố tốc độ gió điển hình cho vịnh Nha Trang. Phân bố tốc độ gió 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. Theo kết quả phân tích, phân bố Logistic phù hợp tốt nhất cho chế độ gió tháng 1, 2 và 10; phân bố Gamma phù hợp tốt nhất cho chế độ gió tháng 3, 7; phân bố Log-Logistic phù hợp tốt nhất cho chế độ gió tháng 4, 6, 8, 9 và 12; phân bố InverseGaussian phù hợp tốt nhất tháng 5 và phân bố Nakagami phù hợp tốt nhất cho tháng 11. Hàm phân bố cực trị (Extreme) thường cho kết quả không phù hợp cho chế độ gió vịnh Nha Trang.

Từ khoá: Tốc độ gió, phân bố xác suất, phân bố tích lũy, hoa gió, các tiêu chí ước lượng.

1. Giới thiệu

Để nghiên cứu các quá trình thủy động lực tác động lên vịnh Nha Trang, tác động của chế độ gió là một trong những tham số chính được quan tâm trước tiên. Như đã biết, gió trên bề mặt biển là một tham số điều khiển quan trọng để xác định các thông số của nhiều quá trình tương tác giữa khí-biển như thông lượng ẩn và cảm nhiệt biển-khí và tốc độ chuyển giao biển - khí của carbon dioxide, dòng động lượng và ứng suất gió trên lớp bề mặt biển. Gió gây sóng bề mặt, các dòng chảy biển và các hiện tượng trôi bề mặt (rác trôi, tràn dầu,...). Gió còn là nhân tố quan trọng trong quá trình xói mòn và thành tạo địa hình. Trong các mô hình động lực học biển, tham số ứng suất gió được đưa vào và là nhân tố chính quyết định khả năng trao đổi động lực học trên bề mặt biển. Đặc biệt chế độ gió phản ánh đúng tính chất cục bộ khu vực nghiên cứu quyết định nhiều đến tính khả thi của kết quả mô phỏng.

Chế độ gió tại vịnh Nha Trang khá đặc biệt, chịu ảnh hưởng nhiều của đặc tính địa phương với ảnh hưởng của nhiều đảo che chắn lớn nhỏ. Do vậy, để nắm bắt được quy luật đúng cho phân bố tốc độ gió tác động đến vịnh Nha Trang là công việc với nhiều ý nghĩa. Nhằm tận dụng tốt nguồn dữ liệu quan trắc nhiều năm hiện có của Đài khí tượng thủy văn khu vực Nam Trung Bộ và cập nhật các phương pháp số tối ưu trên thế giới trong việc xác định các hàm phân bố cho tốc độ gió tại vịnh Nha Trang. Đây là một việc làm có ý nghĩa, bước quan trọng cho đánh giá các tác động của gió đến quá trình thủy động lực trong vịnh Nha Trang.

2. Đối tượng và phương pháp nghiên cứu

2.1. Đối tượng nghiên cứu

Để đánh giá tính chất địa phương của các trường gió, chúng tôi đã chọn khu vực có trạm đo gió tại Nha Trang (109°12'E; 12°13'N). Trước đây, chúng tôi đã sử dụng nguồn số liệu

gió tại các trạm đo khí tượng với chuỗi số liệu đo từ 1987 đến 2007, tần suất đo số liệu là 6 tiếng một lần tại các giờ trong ngày 1, 7, 13, 19 giờ (Trần Văn Chung và cs., 2020). Tuy nhiên, có một điểm cần lưu ý, là từ năm 1987- 2000 (14 năm), số liệu thực đo chỉ cung cấp hướng gió theo 8 hướng chính; từ năm 2001 trở đi cùng với sự cải tiến phương pháp và thiết bị đo nên đã cung cấp chính xác hơn hướng gió tác động lên khu vực Nha Trang theo 16 hướng. Để tránh các sai số trong xác định hướng gió đúng cho vịnh Nha Trang, với độ chính xác theo 16 hướng nên chuỗi số liệu phân tích được chúng tôi chọn bắt đầu từ năm 2001 (Hình 1).

2.2. Phương pháp nghiên cứu

Để tìm quy luật phân bố chế độ gió tại vùng biển Nha Trang, chúng tôi đã sử dụng 11 hàm phân bố có 02 tham số điển hình với hình dạng phân bố tốc độ gió 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 gió tại khu vực. Dưới đây là tổng quan ngắn gọn về các loại hàm phân bố và giá trị của 2 tham số thể hiện trên các hàm phân bố được thể hiện trên Bảng 1.

2.2.1. Tổng quan các loại hàm phân bố

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).

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

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

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

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

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

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

- Hàm mật độ phân bố: $y = f(x|a, b) = \frac{1}{b^a\Gamma(a)} x^{a-1} e^{-\frac{x}{b}}$

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

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

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

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

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

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

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:

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

- Hàm phân bố tích lũy:
$$F(x) = \Phi\left(\frac{\sqrt{x} - \sqrt{\frac{1}{\beta}}}{\gamma}\right), x > 0, \gamma > 0$$

Φ là hàm phân bố tích lũy của phân bố chuẩn chuẩn hóa (standard normal distribution) được xác định:
$$\Phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Phân bố Gaussian nghịch đảo (Inverse Gaussian distribution). Phân bố Gaussian nghịch đảo còn được gọi là phân bố Wald. Các phân bố Gaussian nghịch đảo có nhiều điểm giống với các phân bố Gaussian chuẩn (chuẩn tắc), dẫn ra từ các ứng dụng trong suy luận thống kê (Floks & Chhikara, 1978).

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

- Hàm phân bố tích lũy:

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

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

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

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

Phân bố Log-Logistics. Phân bố Log-Logistic 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. 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:

$$z = (\log(x) - \mu)/\sigma$$

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

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 là: $f(x|s, \sigma) = I_0\left(\frac{x s}{\sigma^2}\right) \frac{x}{\sigma^2} e^{-\left(\frac{x^2 + s^2}{2\sigma^2}\right)}$

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: $p = F(x|s, \sigma) = 1 - Q_1\left(\frac{s}{\sigma}, \frac{x}{\sigma}\right)$

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.

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

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

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

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

Với γ là hàm Gamma không hoàn chỉnh dưới:

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

Phân bố giá trị cực trị (Extreme Value Distribution)

- Hàm mật độ xác suất đối với phân bố giá trị cực trị (Coles và cs., 2001) với tham số vị trí μ , tham số tỷ lệ σ là:

$$y = f(x|\mu, \sigma) = \sigma^{-1} \exp\left(\frac{x - \mu}{\sigma}\right) \exp\left(-\exp\left(\frac{x - \mu}{\sigma}\right)\right)$$

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

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 μ	Độ lệch chuẩn σ
2	Lognormal	Giá trị trung bình của logarit μ	Độ lệch chuẩn của các giá trị logarit σ
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	Birnbaum-Saunders	Tham số tỷ lệ β	γ tham số hình dạng
6	Inverse Gaussian	Tham số tỷ lệ μ	λ tham số hình dạng

STT	Loại phân bố	Tham số 1 (A)	Tham số 2 (B)
7	Logistic	Giá trị trung bình μ	σ tham số tỷ lệ
8	Loglogistic	Giá trị trung bình của logarit μ	Tham số tỷ lệ của giá trị logarit σ
9	Rician	Tham số không trung tâm s	Tham số tỷ lệ σ
10	Nakagami	Tham số hình dạng μ	Tham số tỷ lệ ω
11	Extreme	Tham số vị trí μ	Tham số tỷ lệ σ

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.

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 tốc độ gió, một số thông số kiểm định được sử dụng và mô tả ngắn gọn như sau:

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 tốc độ gió được đưa ra bởi: $MABE = \frac{1}{n} \sum_{i=1}^n |E_i - O_i|$

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 tốc độ gió. 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.

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:

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

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 (χ^2). Kiểm định chi-bình phương (χ^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-bình phương 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 và thảo luận

Toàn bộ sản phẩm phân tích về hướng và tốc độ gió được thể hiện khá chi tiết trên Bảng 2. Để có thể nhìn thấy trực quan hơn về kết quả phân tích, chúng tôi đã thể hiện kết quả phân tích này ra dạng hoa gió (Hình 1).

Bảng 2. Bảng phân cấp chế độ gió theo 16 hướng và 11 cấp độ gió (2001 - 2014)

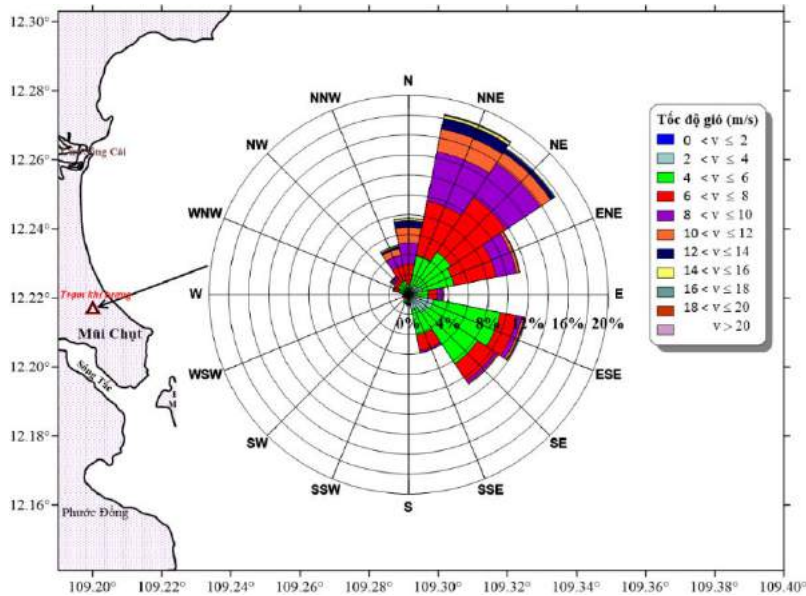
Hướng	0 < v ≤ 2	2 < v ≤ 4	4 < v ≤ 6	6 < v ≤ 8	8 < v ≤ 10	10 < v ≤ 12	12 < v ≤ 14	14 < v ≤ 16	16 < v ≤ 18	18 < v ≤ 20	v > 20	Tốc độ trung bình (m/s)	Tổng số liệu	Tỷ lệ (%)
N	8	32	148	412	436	324	132	40	4	4	0	9,4	1540	7,5
NNE	12	108	684	1148	1032	476	224	56	16	8	4	8,6	3768	18,5
NE	32	192	820	1352	840	260	88	16	4	0	0	7,7	3604	17,7
ENE	16	276	660	880	424	60	12	4	0	0	0	6,9	2332	11,4
E	8	160	236	184	104	24	4	0	0	0	0	6,4	720	3,5
ESE	24	500	1388	356	96	48	16	4	0	0	0	5,7	2432	11,9
SE	12	436	1256	420	64	24	4	0	0	0	0	5,7	2216	10,9
SSE	16	304	532	324	40	4	0	0	0	0	0	5,7	1220	6,0
S	4	100	64	20	20	4	16	0	0	0	0	5,8	228	1,1
SSW	4	60	116	16	4	0	0	0	0	0	0	5	200	1,0
SW	4	16	28	24	12	4	0	4	0	0	0	6,7	92	0,5
WSW	0	20	60	32	8	12	0	0	0	0	0	6,4	132	0,6
W	0	12	40	16	16	8	4	4	4	0	0	7,8	104	0,5
WNW	12	68	124	88	24	16	0	0	0	0	0	6	332	1,6
NW	12	64	184	100	68	16	8	0	0	4	0	6,6	456	2,2
NNW	12	108	168	308	224	140	48	24	8	0	0	8,2	1040	5,1
													20416	100,0

Tổng số liệu	176	2456	6508	5680	3412	1420	556	152	36	16	4			
Tỷ lệ (%)	0,9	12,0	31,9	27,8	16,7	7,0	2,7	0,7	0,2	0,1	0,0			100,0

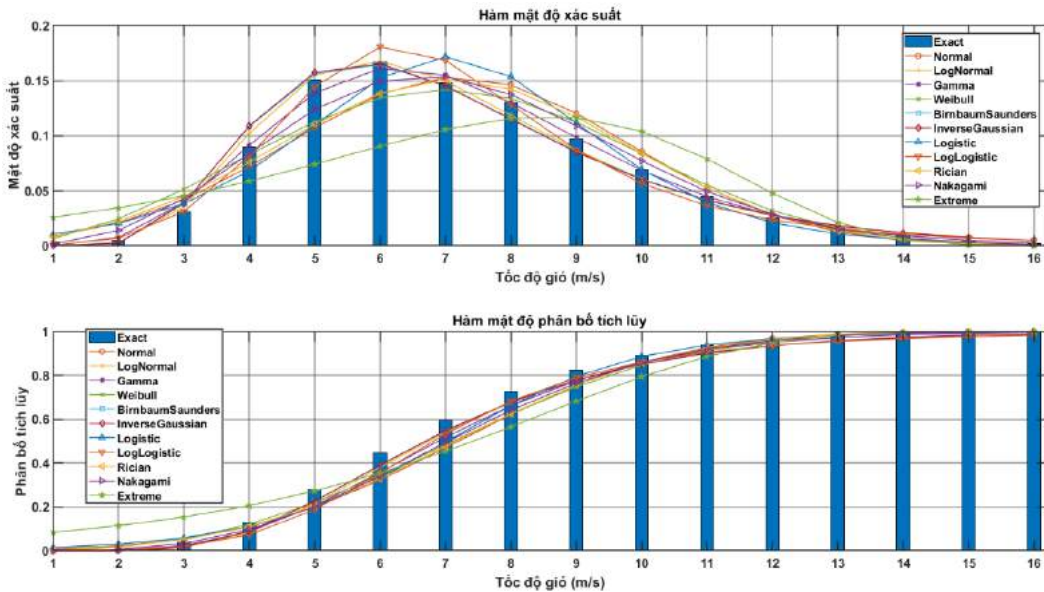
*) Ghi chú: v là tốc độ gió (m/s) tại độ cao 10 m so với mực nước biển.

Phân tích trên Bảng 2 và Hình 1, đã thể hiện đầy đủ số lượng số liệu cho các tần suất xuất hiện của hướng và tốc độ gió. Có 4 phân loại tốc độ gió có tần suất xuất hiện nhiều nhất (trên 12 %), đó là trong khoảng (4 < v ≤ 6) (m/s) với 6.508 số liệu được tìm thấy chiếm 31,9 %; tiếp theo là tốc độ gió trong khoảng (6 < v ≤ 8) (m/s) với 5.680 số liệu, chiếm 27,8 %; tốc độ gió (8 < v ≤ 10) (m/s) với 3.412 số liệu được tìm thấy, chiếm 16,7 % và tốc độ gió (2 < v ≤ 4) (m/s) với 2.456 số liệu được phân tích chiếm 12,0 %. Các loại gió xuất hiện thường xuyên này từ 10 m/s trở xuống, chiếm tới 89,3 % số liệu được phân tích.

Về đặc trưng hướng gió phổ biến, chế độ gió vịnh Nha Trang chịu ảnh hưởng tính khu vực thể hiện khá rõ qua các hướng gió thường xuyên xuất hiện (Hình 1). Cụ thể, khi khu vực chịu tác động của mùa gió Đông Bắc, khu vực vịnh Nha Trang chiếm ưu thế với 03 hướng gió là gió Bắc Đông Bắc (18,5 %); Đông Bắc (17,7 % và Đông Đông Bắc (11,4 %), trong khi tác động của mùa gió Tây Nam thể hiện qua 03 hướng gió, đó là Đông Đông Nam (11,9 %), Đông Nam (10,9 %) và Nam Đông Nam (6,0 %) (Hình 1, Bảng 2).



Hình 1. Hoa gió tại vịnh Nha Trang (2001 - 2014)



Hình 2. Các dạng hàm phân bố của tốc độ gió điển hình theo nhiều năm

Tuy nhiên, để có thể phản ánh đúng quy luật gió, các phân tích cho từng phương pháp trích lọc tham số theo các hàm phân bố đã được thực hiện. Kết quả của các loại phân bố thể hiện trên mật độ xác suất và phân bố tích lũy so với số liệu quan sát được trình bày trên Hình 2. Để đánh giá cho tính phù hợp của các phương pháp sử dụng, các tiêu chí kiểm định đã được thực hiện trên Bảng 3.

Căn cứ vào kết quả đánh giá sai số trên Bảng 3 qua các tính toán được phân tích theo tháng và năm (Hình 2) cho 11 phương pháp số. Chúng ta dễ dàng có thể chọn lựa chọn được phương pháp thích hợp cho nghiên cứu phân bố gió cho vịnh Nha Trang, 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 10 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 (χ^2) (chọn giá trị nhỏ nhất).

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

Tháng	Tham số	Normal	Log-Normal	Gamma	Weibull	Birnbaum Saunders	InverseGaussian	Logistic	Log-Logistic	Rician	Nakagami	Extreme
1	MABE	0,012	0,014	0,011	0,015	0,015	0,015	<u>0,008</u>	0,009	0,012	0,010	<u>0,026</u>
	RMSE	0,061	0,080	0,064	0,079	0,085	0,086	<u>0,041</u>	0,044	0,060	0,059	<u>0,131</u>
	R ²	0,938	0,893	0,931	0,896	0,877	0,876	<u>0,972</u>	0,968	0,939	0,942	<u>0,711</u>
2	χ^2	0,004	0,0069	0,004	0,007	0,008	0,008	<u>0,002</u>	0,002	0,004	0,004	<u>0,019</u>
	MABE	0,007	0,014	0,010	0,009	0,014	0,015	<u>0,007</u>	0,011	0,007	0,007	<u>0,022</u>
	RMSE	0,039	0,073	0,052	0,052	0,076	0,077	<u>0,036</u>	0,057	0,038	0,040	<u>0,114</u>
3	R ²	0,976	0,912	0,955	0,955	0,904	0,901	<u>0,979</u>	0,946	0,976	0,974	<u>0,785</u>
	χ^2	0,001	0,0052	0,002	0,002	0,006	0,006	<u>0,0005</u>	0,003	0,001	0,001	<u>0,013</u>
	MABE	0,012	0,011	<u>0,0102</u>	0,012	0,011	0,011	0,013	0,013	0,011	0,010	<u>0,021</u>
4	RMSE	0,072	0,059	<u>0,054</u>	0,068	0,060	0,061	0,081	0,070	0,069	0,057	<u>0,117</u>
	R ²	0,900	0,932	<u>0,944</u>	0,910	0,929	0,927	0,873	0,905	0,908	0,936	<u>0,732</u>
	χ^2	0,004	0,0031	<u>0,003</u>	0,004	0,003	0,003	0,006	0,004	0,004	0,003	<u>0,012</u>
5	MABE	0,022	0,012	0,014	0,028	0,012	0,012	0,015	<u>0,008</u>	0,022	0,018	<u>0,040</u>
	RMSE	0,110	0,059	0,069	0,132	0,062	0,062	0,069	<u>0,036</u>	0,108	0,089	<u>0,203</u>
	R ²	0,865	0,960	0,947	0,803	0,957	0,957	0,946	<u>0,986</u>	0,869	0,911	<u>0,538</u>
6	χ^2	0,012	0,0033	0,004	0,019	0,004	0,004	0,004	<u>0,001</u>	0,012	0,008	<u>0,048</u>
	MABE	0,015	0,004	0,006	0,016	0,00337	<u>0,00335</u>	0,013	0,008	0,015	0,011	<u>0,030</u>
	RMSE	0,089	0,020	0,037	0,094	0,0193	<u>0,0192</u>	0,081	0,043	0,086	0,062	<u>0,177</u>
7	R ²	0,905	0,995	0,984	0,895	<u>0,9955</u>	0,9956	0,922	0,978	0,913	0,954	<u>0,625</u>
	χ^2	0,008	0,0003	0,001	0,009	0,0003	<u>0,0002</u>	0,006	0,002	0,007	0,004	<u>0,035</u>
	MABE	0,016	0,009	0,009	0,022	0,009	0,009	0,012	<u>0,006</u>	0,016	0,012	<u>0,039</u>
8	RMSE	0,072	0,049	0,040	0,093	0,054	0,054	0,053	<u>0,026</u>	0,070	0,051	<u>0,168</u>
	R ²	0,942	0,973	0,982	0,903	0,967	0,967	0,968	<u>0,993</u>	0,945	0,971	<u>0,684</u>
	χ^2	0,006	0,0027	0,002	0,009	0,003	0,003	0,002	<u>0,001</u>	0,005	0,003	<u>0,031</u>
9	MABE	0,013	0,009	<u>0,005</u>	0,020	0,009	0,009	0,011	0,008	0,013	0,008	<u>0,038</u>
	RMSE	0,059	0,042	<u>0,025</u>	0,082	0,044	0,045	0,045	0,035	0,057	0,035	<u>0,163</u>
	R ²	0,962	0,981	<u>0,993</u>	0,927	0,979	0,978	0,978	0,986	0,965	0,986	<u>0,710</u>
10	χ^2	0,003	0,0018	<u>0,001</u>	0,007	0,002	0,002	0,001	0,001	0,003	0,001	<u>0,029</u>
	MABE	0,018	0,012	0,010	0,021	0,013	0,013	0,010	<u>0,007</u>	0,017	0,013	<u>0,036</u>

Tháng	Tham số	Normal	Log-Normal	Gamma	Weibull	Birnbaum Saunders	InverseGaussian	Logistic	Log-Logistic	Rician	Nakagami	Extreme
	RMSE	0,088	0,070	0,058	0,103	0,079	0,080	0,055	0,037	0,086	0,069	0,189
	R ²	0,918	0,948	0,964	0,887	0,934	0,932	0,968	0,985	0,921	0,950	0,619
	χ^2	0,008	0,0053	0,004	0,012	0,007	0,007	0,003	0,002	0,008	0,005	0,041
9	MABE	0,026	0,012	0,015	0,030	0,013	0,012	0,018	0,007	0,026	0,021	0,044
	RMSE	0,147	0,069	0,089	0,155	0,072	0,072	0,104	0,047	0,145	0,120	0,245
	R ²	0,787	0,954	0,922	0,764	0,949	0,950	0,894	0,978	0,795	0,859	0,412
	χ^2	0,025	0,0053	0,009	0,028	0,006	0,006	0,011	0,002	0,024	0,016	0,076
	MABE	0,008	0,027	0,020	0,012	0,031	0,032	0,007	0,017	0,010	0,015	0,017
	RMSE	0,043	0,139	0,103	0,062	0,160	0,163	0,036	0,092	0,050	0,078	0,094
10	R ²	0,962	0,607	0,783	0,921	0,478	0,463	0,973	0,830	0,949	0,877	0,822
	χ^2	0,001	0,0195	0,010	0,003	0,026	0,027	0,001	0,008	0,002	0,006	0,010
	MABE	0,010	0,010	0,006	0,010	0,010	0,010	0,010	0,009	0,009	0,007	0,022
11	RMSE	0,049	0,056	0,039	0,050	0,059	0,060	0,050	0,047	0,046	0,036	0,107
	R ²	0,943	0,925	0,963	0,939	0,917	0,914	0,940	0,948	0,949	0,969	0,724
	χ^2	0,002	0,0035	0,002	0,003	0,004	0,004	0,002	0,002	0,002	0,001	0,012
12	MABE	0,012	0,007	0,008	0,016	0,007	0,007	0,012	0,006	0,012	0,010	0,027
	RMSE	0,082	0,043	0,050	0,099	0,045	0,045	0,068	0,037	0,081	0,065	0,158
	R ²	0,894	0,971	0,962	0,849	0,968	0,968	0,929	0,979	0,898	0,935	0,612
Cả năm	χ^2	0,007	0,002	0,003	0,010	0,002	0,002	0,005	0,001	0,007	0,005	0,028
	MABE	0,013	0,005	0,003	0,012	0,006	0,006	0,012	0,006	0,012	0,008	0,027
	RMSE	0,068	0,024	0,019	0,064	0,031	0,032	0,065	0,033	0,064	0,041	0,140
	R ²	0,913	0,989	0,994	0,925	0,982	0,981	0,923	0,980	0,925	0,970	0,636
	χ^2	0,005	0,0005	0,0003	0,004	0,001	0,001	0,004	0,001	0,004	0,002	0,020

*) Ghi chú: Chữ **in đậm gạch chân** là thỏa mãn tốt nhất và **chữ in nghiêng** là ít thỏa mãn nhất.

Kết luận

Phân tích chế độ gió thực đo nhiều năm tại khu vực vịnh Nha Trang, cho thấy khi chịu tác động của mùa gió Đông Bắc, khu vực vịnh Nha Trang chiếm ưu thế với 03 hướng gió là gió Bắc Đông Bắc (18,5 %); Đông Bắc (17,7 % và Đông Đông Bắc (11,4 %), trong khi tác động của mùa gió Tây Nam thể hiện qua 03 hướng gió, đó là Đông Đông Nam (11,9 %), Đông Nam (10,9) và Nam Đông Nam (6,0 %).

Trong 11 phương pháp số về dạng hàm phân bố được sử dụng cho phân tích chế độ gió vịnh Nha Trang cho thấy khả năng thích ứng cụ thể trong đánh giá mức độ sai số cho phân tích theo từng tháng và cho cả năm. Tùy vào mức độ ứng dụng các tương tác gió mà chọn lựa các hàm phù hợp cho dự báo. Tuy nhiên, qua các phân tích cho thấy, các dự báo nên sử dụng các hàm phân bố theo tháng sẽ đạt kết quả chính xác cao. Cụ thể, trên cơ sở phân tích và so sánh hiệu quả của 11 phương pháp số, có một số kết quả như sau: sử dụng phân bố Gamma cho đánh giá chế độ gió chung cho cả năm; trong tháng 1,2 nên sử dụng hàm Logistic; tháng 3, 7 nên dùng hàm Gamma; tháng 4, 6, 8, 9 và 12 nên sử dụng hàm Log-Logistic; tháng 5 nên sử dụng hàm InverseGaussian; trong tháng 10 nên sử dụng hàm Logistic và tháng 11 nên sử dụng hàm Nakagami. Phương pháp phân bố cực trị (Extreme) thường cho kết quả không tốt cho dự báo chế độ gió ở vịnh Nha Trang và riêng tháng 10 và 11 không nên sử dụng phương pháp InverseGaussian.

Lời cảm ơn: Nhóm tác giả xin cảm ơn đề tài mã số KHCBBI.01/20-22: “*Tác động của các quá trình hải dương đến quần xã sinh vật phù du vùng thềm lục địa Nam Trung Bộ Việt Nam*”, đã cung cấp một phần kinh phí và các đồng nghiệp đã đóng góp ý kiến hoàn thiện bài báo. Công trình này chào mừng Kỷ niệm 100 năm thành lập Viện Hải dương học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

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, 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.
- Coles, S., Bawa, J., Trenner, L., Dorazio, P., 2001. An introduction to statistical modeling of extreme values, Vol. 208. Springer, New York.
- 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 Chhikara, R.S., 1978. The inverse gaussian distribution and its statistical application - A Review. *Journal of the Royal Statistical Society. Series B*, Vol. 40, No. 3, 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.
- 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*, 3rd edition. ISBN-10: 0131471228; ISBN-13: 9780131471221, Pearson, 832 pp.
- 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, 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.

COMPARISON ELEVEN DIFFERENT METHODS OF WIND SPEED
DISTRIBUTION WITH A NOVEL GOODNESS OF FIT METRIC FOR STUDYING
WIND REGIME FOR NHA TRANG BAY

Tran Van Chung*, Ngo Manh Tien and Nguyen Hoang Thai Khang
Institute of Oceanography, VAST
* tvanchung@gmail.com

Abstract. In this work, eleven distribution functions to describe the wind speed probability are tested and evaluated, finding a typical wind speed distribution function in Nha Trang bay. The final wind speed distribution was selected based on multiple criteria and the proposed distribution outperformed the other parametric models tested and optimally described. According to the analysis results, the Log-Logistic distribution is best suited for the wind regimes of January, February, and October; the Gamma distribution is best suited for the wind regime in March and July; the Log-Logistic distribution is best suited for wind regimes in April, June, August, September, and December; the InverseGaussian distribution is best suited for May; and the Nakagami distribution is best suited for November. The extreme distribution function often gives inaccurate results forecasting the wind regime for Nha Trang bay.

Keywords: Wind speed, probability distribution, cumulative distribution function, wind rose, evaluation criterias.