

DỰ BÁO XÁC SUẤT NGẮN HẠN CHO CÔNG SUẤT ĐIỆN GIÓ
DỰA TRÊN MÔ HÌNH BIẾN ĐỘNG NGẪU NHIÊN

*Đỗ Phương Nhung**, *Nguyễn Trọng Khiêm†*, *Đặng Minh Quân†*,
Đặng Hoàng Anh†, *Quách Thị Hạnh†*, *Nguyễn Thị Phi Doan†*
Email: *dphung@hou.edu.vn*

Ngày tòa soạn nhận được bài báo: 01/11/2023
Ngày phản biện đánh giá: 17/06/2024
Ngày bài báo được duyệt đăng: 26/06/2024

DOI: 10.59266/houjs.2024.417

Tóm tắt: Năng lượng gió đã trở thành một giải pháp thay thế rất khả thi để cung cấp điện. Do đó, dự báo xác suất điện gió, thứ không những dự báo các giá trị chính xác, mà còn cho thêm nhiều thông tin về tính bất định của các giá trị được dự báo, ngày càng quan trọng. Tuy nhiên, hầu hết các phương pháp dự báo hiện tại chỉ tập trung vào dự báo điểm mà không chú ý đến dự báo mật độ. Do đó, ứng dụng của các kết quả này vào các vấn đề ngẫu nhiên của hệ thống điện còn hạn chế. Bài báo này trình bày một phương pháp dự báo xác suất điện gió mới gọi là mô hình biến động ngẫu nhiên với độ chính xác cao cho dự báo mật độ. Phương pháp phân cấp Bayesian được sử dụng để ước tính mô hình đề xuất. Mô hình biến động ngẫu nhiên được sử dụng trong lập mô hình và dự báo điện gió dựa trên dữ liệu thực tại Nhật Bản vào năm 2005. Độ chính xác của mô hình biến động ngẫu nhiên cho cả dự báo chính xác và dự báo xác suất với các khoảng thời gian dự báo khác nhau sẽ được phân tích và so sánh với mô hình ARMA.

Từ khóa: Mô hình ARMA; Bayesian; Dự báo xác suất; Mô hình biến động ngẫu nhiên; Dự báo công suất điện gió.

Bảng ký hiệu

$\propto$	tỉ lệ thuận với
$\sim$	có phân phối như
A'	Ma trận (vecto) chuyển vị của A

$diag(a)$	Ma trận đường chéo giá trị a
P	Giá trị xác suất
$N(b, c)$	Phân phối chuẩn
$IG(\alpha, \beta)$	Phân phối Gamma nghịch đảo

I. Đặt vấn đề

Với sự cạn kiệt của các nguồn nhiên liệu hóa thạch, năng lượng gió ngày càng được tích hợp nhiều hơn vào các hệ thống

điện. Tuy nhiên, tính bất định của điện gió làm thay đổi các vấn đề truyền thống trong vận hành và điều khiển hệ thống điện, chẳng hạn như vận hành lưới điện

* Trường Đại học Mở Hà Nội
† Trường Điện – Điện tử, Đại học Bách Khoa Hà Nội

thời gian thực, chất lượng điện năng, lựa chọn tổ máy vận hành, công suất hệ thống truyền tải và nâng cấp trong tương lai, chào giá thị trường điện[1-4]. Do đó, việc xây dựng các phương pháp dự báo điện gió là cần thiết, đặc biệt là dự báo xác suất.

Trong những năm gần đây, đã có nhiều nghiên cứu trong lĩnh vực dự báo điện gió với phạm vi thời gian dự báo từ vài giây đến vài tuần, cụ thể là phương pháp vật lý và thống kê. Cách tiếp cận vật lý điển hình là mô hình dự báo số trị yêu cầu nhiều dữ liệu đầu vào như tốc độ và hướng gió, nhiệt độ, áp suất, bề mặt và độ gồ ghề nên mô hình dự báo trở nên phức tạp và cần nhiều thời gian hơn để giải quyết. Do đó, cách tiếp cận vật lý này phù hợp với dự báo dài hạn (hơn một ngày) [5].

Ngược lại, phương pháp thống kê sử dụng dữ liệu lịch sử về điện gió để xây dựng mô hình dự báo và huấn luyện mô hình dựa trên sự chênh lệch giữa giá trị dự báo và giá trị thực tế. Các mô hình dựa trên chuỗi thời gian và các phương pháp dựa trên mạng nơtron nhân tạo (NN) là hai loại phương pháp này. Cả hai đều dễ dàng, không tốn kém và chính xác để dự báo ngắn hạn [6-8]. Theo [5], các mô hình dựa trên chuỗi thời gian chính xác hơn NN.

Mô hình trung bình động hồi quy tự động (ARMA) là loại mô hình dựa trên chuỗi thời gian phổ biến nhất để dự đoán tốc độ hoặc sức mạnh của gió trong tương lai [9-11]. Mặc dù thực hiện dễ dàng trong cả dự báo điểm và xác suất, mô hình ARMA giả định sự biến động có điều kiện liên tục trong khi điều này không đúng trong thế giới thực. Giả định này dẫn đến biến động của dữ liệu đã cho càng cao thì phương sai ước tính của mô hình càng

cao. Kết quả là, điều này làm giảm chất lượng của dự báo xác suất.

Trong bài báo này, mô hình biến động ngẫu nhiên (SV) dựa trên mô hình ARMA được đề xuất. Kế thừa những ưu điểm từ mô hình ARMA; phương pháp đề xuất cũng xem xét sự thay đổi của phương sai có điều kiện theo thời gian. Bên cạnh đó, sự kết hợp giữa phương pháp phân cấp Bayesian và mô phỏng Markov Chain Monte Carlo được áp dụng để không chỉ ước tính các tham số của mô hình mà còn tính toán mật độ được dự đoán. Điều này làm cho việc tính toán linh hoạt hơn. Độ chính xác của phương pháp này sẽ được so sánh với mô hình ARMA cho cả dự báo điểm và dự báo xác suất.

II. Phương pháp nghiên cứu

2.1. Mô hình biến động ngẫu nhiên (SV)

Trong mô hình ARMA(p, q) truyền thống, hành vi của chuỗi thời gian cố định được mô tả như sau:

$$y_t = \varphi_0 + \sum_{i=1}^p \varphi_i y_{t-i} + \sum_{j=1}^q \psi_j u_{t-j} + u_t \quad (1)$$

Trong đó, y_t là giá trị quan sát được tại t ; u_t là phần sai số tại t với giá trị trung bình bằng không và phương sai không đổi; p là bậc của phần tự hồi quy (AR); q là bậc của phần trung bình động (MA); φ_i , ψ_j là hệ số của AR, MA. Trong mô hình này, độ biến động của chuỗi thời gian là không đổi vì phương sai của số hạng sai số được giả định là không đổi.

Trong mô hình này, độ biến động của chuỗi thời gian là không đổi vì phương sai của số hạng sai số được giả định là không đổi:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \psi u_{t-1} + u_t \quad (2)$$

$$u_t = e^{h_t/2} \varepsilon_t \sim N(0, e^{h_t}) \quad (3)$$

Trong đó, h_t được gọi là biến động logarit của chuỗi thời gian và lần lượt phát triển theo mô hình tự hồi quy với độ trễ 1 (AR(1)) như sau:

$$h_t = \mu_h + \varphi_h(h_{t-1} - \mu_h) + \varepsilon_t^h \quad (2)$$

$$\varepsilon_t^h \sim N(0, \sigma_h^2) \quad (3)$$

Trong đó, μ_h là trung bình của chuỗi thời gian h_t ; φ_h là tham số của mô hình AR(1); ε_t^h là phần sai số.

Sai số u_t và ε_t^h được giả định là độc lập với nhau. Giá trị ban đầu u_0 và h_0 được đặt bằng không. Bên cạnh đó, ta cũng giả sử rằng $h_1 \sim N(\mu_h, \sigma_h^2 / (1 - \varphi_h^2))$.

Dễ dàng nhận ra rằng do sự hiện diện của biến động logarit h_t , biến động của y_t thay đổi theo thời gian.

2.2. Phương pháp phân cấp Bayesian để ước tính mô hình biến động ngẫu nhiên

2.2.1. Cơ sở lý thuyết của phương pháp phân cấp Bayesian

Trong phần trước, mô hình SV được giới thiệu là một mô hình dự báo. Trước khi sử dụng mô hình này để dự báo dữ liệu trong tương lai, cần phải khớp nó với dữ liệu lịch sử và ước tính các tham số của nó.

Một cách tiếp cận rất cơ bản để ước tính tham số là phương pháp ước lượng hợp lý cực đại (Maximum Likelihood-ML) tìm các tham số dựa trên việc tối đa hóa hàm khả năng của các tham số và dữ liệu đã cho. Mặc dù phương pháp này rất hữu ích, nhưng nó vẫn có một số nhược điểm. Ví dụ: ước tính ML có thể không xác định được nếu kích thước mẫu nhỏ. Cũng khó tìm cực đại của hàm hợp lý khi có nhiều

tham số trong mô hình [12]. Trong những trường hợp này, việc sử dụng Phương pháp phân cấp Bayesian là cần thiết.

Gọi x và lần lượt là dữ liệu và tham số cần ước lượng. Phương pháp Bayesian được triển khai dựa trên việc coi θ là một biến ngẫu nhiên. Hàm mật độ xác suất $f(\theta)$ trước khi có dữ liệu quan sát đầu tiên được gọi là phân phối trước. Phân phối có điều kiện $f(x|\theta)$ được gọi là hàm khả năng Bayes. Phân phối sau $f(\theta|x)$, mô tả θ có tính đến dữ liệu x đã cho, tỷ lệ thuận với tích của hàm tiên nghiệm và khả năng:

$$f(\theta|x) \propto f(x|\theta)f(\theta)$$

Trong khi phương pháp ML tối đa hóa hàm khả năng của tham số θ và quan sát x , phương pháp Bayes tìm thấy tham số θ tối đa hóa phân phối sau $f(\theta|x)$.

Nếu mô hình được điều chỉnh có nhiều tham số, ví dụ: hai tham số α và β ; một hệ thống phân cấp thường được sử dụng để chỉ định mô hình như sau:

$$\alpha \sim f(\alpha)$$

$$(\beta|\alpha) \sim f(\beta|\alpha)$$

$$(x|\alpha, \beta) \sim f(x|\alpha, \beta)$$

Hàm mật độ xác suất chung được tính đơn giản bằng tích của phân phối có điều kiện $f(x|\alpha, \beta)$, $f(\beta|\alpha)$, và các phân phối trước $f(\alpha)$:

$$f(x, \alpha, \beta) \propto f(x|\alpha, \beta)f(\beta|\alpha)f(\alpha)$$

Để xác định hàm phân phối sau $f(\alpha|x)$ và chúng tôi coi hàm phân phối chung $f(x, \alpha, \beta)$ là hàm của α và β đối với x cố định và các tham số khác. Các kết quả như phân phối sau biên, giá trị trung bình hoặc phương sai của α và β có thể thu được bằng mô phỏng Markov Chain Monte Carlo (MCMC), cụ thể là bộ lấy mẫu Gibbs, theo các bước sau [12] : 1)

Lấy α từ $f(\alpha)$; 2) Lấy β từ $f(\beta | \alpha, x)$; 3) Lấy α từ $f(\alpha | \beta, x)$.

Có thể thấy rằng phân phối sau không chỉ phụ thuộc vào hàm khả năng Bayes mà còn phụ thuộc vào phân phối trước. Cụ thể, việc chọn phân phối trước không cung cấp thông tin và mô phỏng với cỡ mẫu lớn sẽ làm giảm tác động của phân phối trước về 0.

2.2.2. Ước tính mô hình biến động ngẫu nhiên

Trong phần này, phương pháp phân cấp Bayes được áp dụng để ước lượng mô hình biến động ngẫu nhiên được mô tả trong phần II-A với giá trị quan sát $y_0, y_1, y_2, \dots, y_T$.

$$f(h, \varphi, \psi, \mu_h, \varphi_h, \sigma_h^2 | y) \propto f(y | h, \varphi, \psi) f(h | \mu_h, \varphi_h, \sigma_h^2) \times f(\varphi) f(\psi) f(\mu_h) f(\varphi_h) f(\sigma_h^2) \quad (6)$$

Các phân phối sau được lấy theo trình tự từ mẫu MCMC như sau : 1) $f(\varphi | y, h, \psi)$; 2) $f(h | y, \varphi, \psi, \mu_h, \varphi_h, \sigma_h^2)$; 3) $f(\sigma_h^2 | h, \mu_h, \varphi_h)$ và $f(\mu_h | h, \sigma_h^2, \varphi_h)$; 4) $f(\varphi_h | h, \mu_h, \sigma_h^2)$; 5) $f(\psi | y, h, \varphi)$.

• Bước một: Lấy từ

Biểu thức (1) được viết lại dưới dạng ma trận như sau:

$$y = X_\varphi + H_\psi u \quad (7)$$

Trong đó:

$$u = [u_1 \ u_2 \ \dots \ u_T]' \sim N(0, S_y)$$

$$S_y = \text{diag}(e^{h_1}, e^{h_2}, \dots, e^{h_T});$$

$$X = \begin{bmatrix} 1 & y_0 \\ 1 & y_1 \\ \vdots & \vdots \\ 1 & y_{T-1} \end{bmatrix}; H_\psi = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ \psi & 1 & \dots & 0 & 0 \\ 0 & \psi & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \psi & 1 \end{bmatrix}$$

$$\text{Đặt } \tilde{y} = H_\psi^{-1} y \text{ và } \tilde{X} = H_\psi^{-1} X, \quad (7)$$

được viết lại như sau:

$$\tilde{y} = \tilde{X} \varphi + u \quad (8)$$

$$\text{Khi đó, } (\tilde{y} | \varphi, h, \psi) \sim N(\tilde{X} \varphi, S_y)$$

Đặt $y = [y_1 \ y_2 \ \dots \ y_T]'$, $\varphi = [\varphi_0 \ \varphi_1]'$, $h = [h_1 \ h_2 \ \dots \ h_T]'$ chúng tôi giả sử phân phối trước cho các tham số của mô hình như sau:

$$\varphi \sim N(0, \Sigma_\varphi) \text{ với } |\phi_1| < 1$$

$$\psi \sim N(\psi_0, \sigma_\psi^2) \text{ với } |\psi| < 1$$

$$\mu_h \sim N(\mu_{h0}, \sigma_{\mu_h}^2)$$

$$\varphi_h \sim N(\varphi_{h0}, \sigma_{\varphi_h}^2) \text{ với } |\phi_h| < 1$$

$$\sigma_h^2 \sim IG(\alpha_h, \beta_h)$$

Trong đó, $|\phi_1| < 1$ là điều kiện dừng cho (2), $|\psi| < 1$ để khả nghịch. Ngoài ra, $|\phi_h| < 1$ là điều kiện dừng cho (4).

Trong phân phối trước nêu ở trên, hằng số $\Sigma_\varphi, \psi_0, \sigma_\psi^2, \mu_{h0}, \sigma_{\mu_h}^2, \varphi_{h0}, \sigma_{\varphi_h}^2, \alpha_h, \beta_h$, được chỉ định bởi người sử dụng. Phân phối trước chung là:

Theo [12], ta có phân bố sau:

$$(\varphi | y, h, \psi) \sim N(\hat{\varphi}, D_\varphi) \quad (9)$$

Trong đó:

$$D_\varphi = (\tilde{X}' S_y^{-1} \tilde{X} + \Sigma_\varphi^{-1})^{-1}$$

$$\hat{\varphi} = D_\varphi \tilde{X}' S_y^{-1} \tilde{y}$$

Biểu thức (9) có ràng buộc $|\phi_1| < 1$.

• Bước hai:

$$\text{Lấy từ } f(h | y, \varphi, \psi, \mu_h, \varphi_h, \sigma_h^2)$$

Để tính phân bố mật độ $f(h | y, \varphi, \psi, \mu_h, \varphi_h, \sigma_h^2)$, trước tiên cần xác định $\tilde{y} = \tilde{y} - \tilde{X} \varphi$. Từ (8), ta có phương trình sau:

$$\tilde{y} = u \quad (10)$$

Để thấy, $\tilde{y}$ là một vector cột có T phần tử

$$\tilde{y} = [\tilde{y}_1 \ \tilde{y}_2 \ \dots \ \tilde{y}_T]'$$

Do đó, hàng thứ t của (10) được viết lại như sau:

$$\tilde{y}_t = u_t \quad (11)$$

Từ (3) và (10), đặt $y_t^* = \ln(\ddot{y}_t)^2$ và $\varepsilon_t^* = \ln(\varepsilon_t)^2$, ta có:

$$y_t^* = h_t + \varepsilon_t^* \quad (12)$$

Theo [13], hàm mật độ xác suất của có thể được xấp xỉ bằng mô hình hỗn hợp Gaussian bảy thành phần:

$$s_t \in \{1, 2, \dots, 7\} \quad (13)$$

$$\mathbb{P}(s_t = i) = p_i \quad (14)$$

$$(\varepsilon_t^* | s_t = i) \sim N(\mu_i - 1.2704, \sigma_i^2) \quad (15)$$

Trong đó, s_i là chỉ số thành phần; p_i, μ_i, σ_i^2 lần lượt là xác suất hỗn hợp, giá trị trung bình và phương sai của thành phần thứ i được đưa ra Bảng 1.

Bảng 1. Các tham số của mô hình hỗn hợp Gaussian

Thành phần	Xác suất hỗn hợp	Trung bình	Phương sai
1	0.00730	-10.12999	5.79596
2	0.10556	-3.97281	2.61369
3	0.00002	-8.56686	5.17950
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24556	1.79518	0.34023
7	0.25750	-1.08819	1.26261

Đặt

$$s = [s_1 \ s_2 \ \dots \ s_T]',$$

$$y^* = [y_1^* \ y_2^* \ \dots \ y_T^*]',$$

Ta có phân phối sau chung là:

$$f(h, s, \mu_h, \varphi_h, \sigma_h^2 | y^*) \propto f(y^* | s, h) f(h | \mu_h, \varphi_h, \sigma_h^2) \times f(s) f(\mu_h) f(\varphi_h) f(\sigma_h^2) \quad (20)$$

Từ (17), (19) và (20), ta thấy:

$$(h | y^*, \varphi_h, \mu_h, \sigma_h^2) \sim N(\hat{h}, D_h) \quad (17)$$

Trong đó:

$$D_h = (H' \Sigma_h^{-1} H + \Sigma_\varepsilon^{-1})^{-1}$$

$$\hat{h} = D_h (H^{-1} \tilde{\varphi}_h \mu_h + \Sigma_\varepsilon^{-1} (y^* - \mu_\varepsilon))$$

$$(\sigma_h^2 | h, \mu_h, \varphi_h) \sim IG(\alpha_h + \frac{T}{2}, \beta_h + \frac{1}{2} (Hh - \tilde{\varphi}_h \mu_h)' (Hh - \tilde{\varphi}_h \mu_h)) \quad (23)$$

$$\text{và } \varepsilon^* = [\varepsilon_1^* \ \varepsilon_2^* \ \dots \ \varepsilon_T^*]'$$

Từ (12), ta có:

$$y^* = h + \varepsilon^* \quad (16)$$

Trong đó $(\varepsilon^* | s) \sim N(\mu_\varepsilon, \Sigma_\varepsilon)$ với giá trị trung bình và phương sai như sau:

$$\mu_\varepsilon = [\mu_{s_1} - 1.2704 \ \dots \ \mu_{s_T} - 1.2704]'$$

$$\Sigma_\varepsilon = \text{diag}(\sigma_{s_1}^2, \sigma_{s_2}^2, \dots, \sigma_{s_T}^2)$$

Như vậy, ta có:

$$(y^* | s, h) \sim N(h + \mu_\varepsilon, \Sigma_\varepsilon) \quad (17)$$

Mặt khác, (4) được viết lại dưới dạng ma trận như sau:

$$Hh = \tilde{\varphi}_h \mu_h + \varepsilon_h \quad (18)$$

Trong đó:

$$\tilde{\varphi}_h = [1 \ 1 - \varphi_h \ 1 - \varphi_h \ \dots \ 1 - \varphi_h]'$$

$$\varepsilon_h = [\varepsilon_1^h \ \varepsilon_2^h \ \dots \ \varepsilon_T^h] \sim N(0, \Sigma_h)$$

$$\Sigma_h = \text{diag}(\sigma_h^2, \sigma_h^2, \dots, \sigma_h^2)$$

$$H = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ -\varphi_h & 1 & \dots & 0 & 0 \\ 0 & -\varphi_h & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -\varphi_h & 1 \end{bmatrix}$$

Có thể thấy rằng:

$$(h | \varphi_h, \mu_h, \sigma_h^2) \sim N(H^{-1} \tilde{\varphi}_h \mu_h, (H' \Sigma_h^{-1} H)^{-1}) \quad (19)$$

• Bước ba: Lấy từ $f(\sigma_h^2 | h, \mu_h, \varphi_h)$

và $f(\mu_h | h, \sigma_h^2, \varphi_h)$

Từ (18), ta có:

$$(Hh | \varphi_h, \mu_h, \sigma_h^2) \sim N(\tilde{\varphi}_h \mu_h, \Sigma_h) \quad (22)$$

Theo [12], cho h , phân phối gamma nghịch đảo như sau:

Phân phối sau của là:

$$(\mu_h|h, \sigma_h^2, \varphi_h) \sim N(\hat{\mu}_h, D_{\mu_h}) \quad (24)$$

Trong đó:

$$D_{\mu_h} = (\tilde{\varphi}_h' \Sigma_h^{-1} \tilde{\varphi}_h + \sigma_{\mu_h}^{-2})^{-1}$$

$$\hat{\mu}_h = D_{\mu_h} (\tilde{\varphi}_h' \Sigma_h^{-1} H h + \sigma_{\mu_h}^{-2} \mu_{h0})$$

- Bước bốn: Lấy từ $f(\varphi_h|h, \mu_h, \sigma_h^2)$

Đầu tiên, đặt $X_h = [h_0 \ h_1 \ \dots \ h_{T-1}]'$,

(4) được viết lại dưới dạng ma trận như sau:

$$\begin{aligned} \ln f(\psi|y, h, \varphi) &\propto \ln f(\psi) + \ln f(y|\varphi, h, \psi) \\ &\propto \ln f(\psi) - \frac{1}{2} (\tilde{y} - \tilde{X}\varphi)' S_y^{-1} (\tilde{y} - \tilde{X}\varphi) \end{aligned} \quad (27)$$

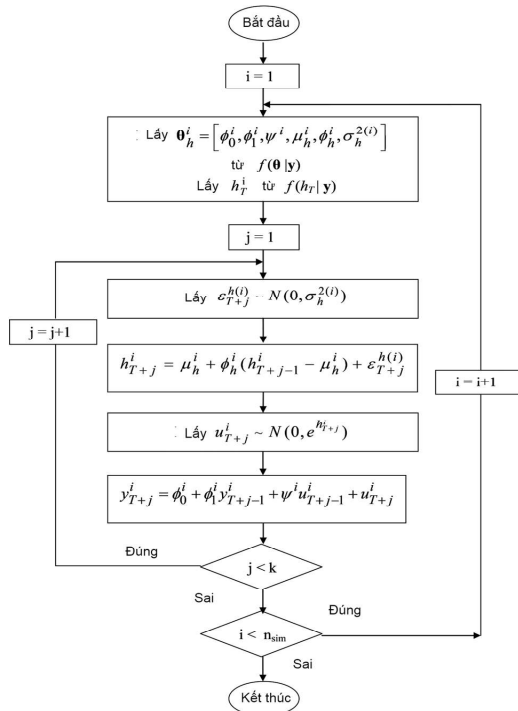
Dễ dàng tính được ψ bằng cách tối đa (27).

2.3. Dự báo

Giả sử $\theta = [\varphi_0, \varphi_1, \psi, \mu_h, \varphi_h, \sigma_h^2]$, phân phối dự đoán sau để dự đoán dữ liệu y trong k bước thời gian từ $T+1$ đến $T+k$ được cho bên dưới:

$$\begin{aligned} f(y_{T+1:T+k}|y) &= \int_{h_T, \theta} \left(\int_{h_{T+1:T+k}} f(y_{T+1:T+k}|h_{T+1:T+k}) \right. \\ &\quad \times f(h_{T+1:T+k}|y, \theta, h_T) d(h_{T+1:T+k})) \times f(\theta, h_T|y) d(\theta, h_T) \end{aligned} \quad (28)$$

Hình 1. Lưu đồ thuật toán dự báo



$$h - \mu_h = (X_h - \mu_h)\varphi_h + \varepsilon_h \quad (25)$$

Với h , phân phối sau của φ_h là:

$$(\varphi_h|h, \mu_h, \sigma_h^2) \sim N(\hat{\varphi}_h, D_{\varphi_h}) \quad (26)$$

Trong đó:

$$D_{\varphi_h} = ((X_h - \mu_h)' \Sigma_h^{-1} (X_h - \mu_h) + \sigma_{\varphi_h}^{-2})^{-1}$$

$$\hat{\varphi}_h = D_{\varphi_h} ((X_h - \mu_h)' \Sigma_h^{-1} (h - \mu_h) + \sigma_{\varphi_h}^{-2} \varphi_{h0})$$

- Bước năm: Lấy từ $f(\psi|y, h, \varphi)$

Từ (6), với y, h, φ , hàm logarit khả năng của ψ là:

III. Kết quả và thảo luận

3.1. Bộ dữ liệu

Phương pháp đề xuất được sử dụng để lập mô hình và dự báo điện gió dựa trên dữ liệu thực được thu thập cứ mỗi 5 phút tại trang trại gió ở Nhật Bản vào năm 2005. Nhà máy điện gió này có công suất 51,75MW.

Dựa trên mô hình ARMA, mô hình SV cũng yêu cầu dữ liệu đầu vào phải cố định. Tuy nhiên, công suất gió không cố

định nên cần phải chuyển đổi dữ liệu ban đầu sang dạng mới có tính cố định. Hai phép biến đổi thường được sử dụng là biến đổi chênh lệch và logarit chênh lệch giữa hai dữ liệu được quan sát liên tiếp [14]. Bây giờ, chúng tôi tính toán chuỗi dữ liệu chênh lệch theo phương trình sau:

$$r_t = P_t - P_{t-1} \quad (29)$$

Trong đó, r_t là chênh lệch tại thời điểm t ; P_t và P_{t-1} lần lượt là công suất gió tại thời điểm t và $t-1$

Đồ thị của chuỗi chênh lệch trong Hình 2 cho thấy chuỗi thời gian này là cố định. Như vậy, mô hình SV sẽ được sử dụng để mô hình hóa chuỗi chênh lệch của điện gió. Dựa trên các chuỗi chênh lệch dự báo và dữ liệu đo được, có thể dễ dàng tính toán năng lượng gió trong tương lai.

3.2. Kết quả ước tính

Đầu tiên, chúng tôi đặt phân phối trước của tham số với $\Sigma_\varphi = \text{diag}(50, 50)$, $\psi_0 = 0$, $\sigma_\psi^2 = 50$, $\mu_{h0} = 0$, $\sigma_{\mu_h}^2 = 5$, $\varphi_{h0} = 0.9$, $\sigma_{\varphi_h}^2 = 1$, $\alpha_h = 10$, $\beta_h = 0.5$. Sau đó lấy 10000 mẫu theo thuật toán được mô tả trong Phần II, phân phối hậu nghiệm thu được được thể hiện trong Hình 3. Ngoài ra, giá trị trung bình hậu nghiệm và độ lệch chuẩn được minh họa trong Bảng 2. Hình 4 trình bày giá trị trung bình sau của h_t cũng như các khoảng tin cậy trong 24 giờ đầu tiên của các mẫu. Mất khoảng 250 giây để ước tính mẫu SV với 4320 mẫu trên Intel® Core™ i5-3320M 2.6GHz (RAM 4GB).

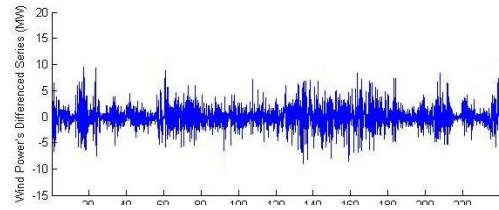
Bảng 2. Giá trị trung bình và độ lệch chuẩn của các tham số

Thông số	β_0	β_1	ψ	μ_h	ϕ_h	σ_h^2
Trung bình	-0.0294	-0.2085	0.1681	1.0854	0.9522	0.1237
Độ lệch chuẩn	0.029	0.2398	0.2402	0.1421	0.0092	0.0215

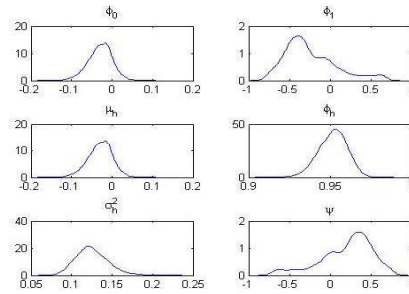
Dự báo trước 15 phút (3 bước) được thực hiện trong hơn 24 giờ được trình bày trong Hình 5. Khoảng dự báo từ 10% đến 90% được minh họa bằng bóng mờ.

Độ chính xác của dự báo điểm sử dụng mô hình SV được đánh giá bằng sai số căn quân phương (RMSE):

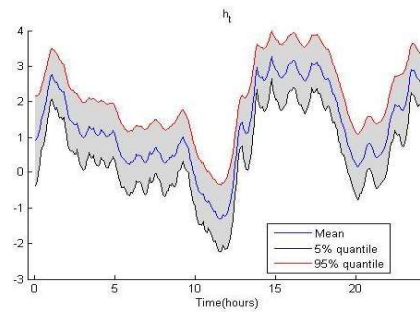
$$\text{RMSE} = \frac{1}{P_{\text{norm}}} \sqrt{\frac{1}{N} \sum_{n=1}^N (P_n - \hat{P}_n)^2} \cdot 100\% \quad (30)$$



Hình 2. Chuỗi dữ liệu chênh lệch của công suất gió.

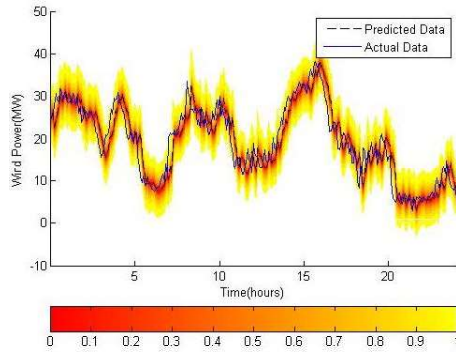


Hình 3. Hàm phân phối của các tham số



Hình 4. Giá trị trung bình và khoảng tin cậy của h_t

Trong đó P_n và $\hat{P}_n$ lần lượt là công suất gió thực tế và dự đoán tại thời điểm n ; P_{norm} là công suất lắp đặt của nhà máy điện gió; N là thời gian dự báo.



Hình 5. Kết quả dự báo 15 phút tới

Dự báo xác suất được đánh giá đơn giản theo tiêu chuẩn Coverage Rates. Theo tiêu chuẩn này, chúng tôi tính toán tần suất mà giá trị thực tế nằm trong khoảng tin cậy phần trăm α của dự báo. Vùng phủ

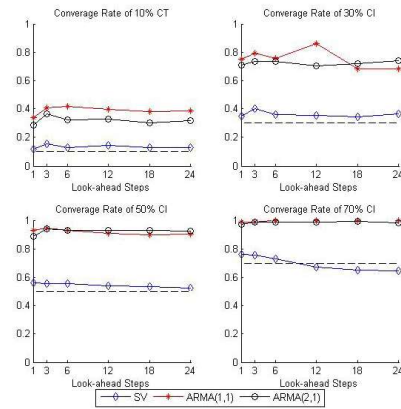
Bảng 3. Sai số căn quân phương (%)

Steps	1	3	6	12	18	24
SV model	3.59	5.69	8.35	12.69	16.17	19.00
ARMA(1,1)	3.75	5.89	8.62	13.12	16.69	20.47
ARMA(2,1)	5.61	12.47	14.70	15.91	20.46	23.54

Dựa trên tiêu chuẩn RMSE và Coverage Rate, mô hình SV được đánh giá và so sánh với mô hình ARMA(1,1) và ARMA(2,1) với các bước nhìn trước là 1, 3, 6, 12, 18, 24 tương ứng với thời gian tới lần lượt là 5 phút, 15 phút, 0,5 giờ, 1 giờ, 1,5 giờ và 2 giờ.

Kết quả về tiêu chuẩn RMSE cho cả hai mô hình được mô tả trong Bảng. III. Rõ ràng từ bảng này là độ chính xác của dự báo điểm sẽ giảm nếu khoảng thời gian dự báo tăng từ trước 1 bước lên trước 24 bước. Tuy nhiên, sai số của mô hình SV nhỏ hơn nhau trong mọi khoảng thời gian. Điều này có nghĩa là mô hình SV chính xác hơn hai mô hình ARMA trong dự báo điểm.

sóng chính xác tương ứng với tần suất khoảng $\alpha\%$. Nếu tần suất lớn hơn (hoặc nhỏ hơn) so với phần trăm α , điều đó có nghĩa là mật độ dự báo sau quá rộng (hoặc hẹp) [15].



Hình 6. Hệ số Coverage Rate của các mô hình SV, ARMA(1,1), ARMA(2,1) tương ứng với khoảng tin cậy là 10%, 30%, 50% and 70%

Hình 6 cho thấy tỷ lệ bao phủ của cả hai mô hình ARMA cao hơn nhiều so với α -phần trăm với α bằng 10%, 30%, 50% và 70% cho tất cả các khoảng thời gian dự báo. Như vậy, dự báo mật độ của các mô hình này là quá rộng. Do đó, việc đánh giá độ bất định của dữ liệu dự đoán sẽ không chính xác. Trong khi đó, tỷ lệ bao phủ của mô hình SV gần bằng α -phần trăm cho tất cả các khoảng thời gian dự báo. Kết quả này cho thấy mô hình SV chính xác hơn các mô hình khác trong dự báo xác suất.

IV. Kết Luận

Một cách tiếp cận dự báo năng lượng gió xác suất ngắn hạn được trình bày trong bài báo này. Sự kết hợp của

Phương pháp phân cấp Bayesian và mô phỏng Markov Chain Monte Carlo được sử dụng để mô hình hóa phương pháp đề xuất cũng được thảo luận. Phương pháp đề xuất là thử nghiệm trên dữ liệu thực tế tại Nhật Bản. Kết quả cho thấy tính hiệu quả và chính xác của phương pháp dự báo này không chỉ đối với dự báo điểm mà cả dự báo xác suất.

Tài liệu tham khảo

- [1]. F. Bouffard and F. D. Galiana, *Stochastic security for operations planning with significant wind power generation*, IEEE Trans. Power Syst., vol. 23, no. 2, pp. 306–316, May 2008.
- [2]. P. Pinson, G. Papaefthymiou, B. Klockl, J. Verboomen, *Dynamic sizing of energy storage for hedging wind power forecast uncertainty*, IEEE Power & Energy Society General Meeting, 26–30 July 2009.
- [3]. N. Qin, C. L. Bak, H. Abildgaard, and Z. Chen, *Multi-Stage Optimization-Based Automatic Voltage Control Systems Considering Wind Power Forecasting Errors*, IEEE Trans on Power Syst, vol. 32, no. 2, pp. 1073 – 1088, March 2017.
- [4]. Nhung N. H, Nakanashi. Y., *Stochastic Dynamic Power Flow Analysis Based on Stochastic Response Surface Method and ARMA-GARCH Model*, IEEE ISGT Conference, April, 2017.
- [5]. S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, *A review of wind power forecasting methods with different time horizons*, in Proc. North Amer. Power Symp., Arlington, TX, USA, Sep. 26–28, 2010, pp. 1–8.
- [6]. J. Yan, X. Zhao, K. Li, *On temporal resolution selection in time series wind power forecasting*, UKACC 11th International Conference on Control (CONTROL), 31 Aug.-2 Sept. 2016.
- [7]. E. Cadenas, and W. Rivera, *Wind speed forecasting in the South Coast of Oaxaca, Mexico*, Renewable Energy, vol. 32, no. 12, pp. 2116–2128, Oct. 2007.
- [8]. E. Cadenas, O.A. Jaramillo, and W. Rivera, *Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method*, Renewable Energy, vol. 35, no. 5, pp. 925–930, May 2010.
- [9]. M. S. Miranda, R. W. Dunn, *One hour ahead wind speed prediction using a Bayesian methodology*, IEEE PES proceedings, IEEE Summer meeting, 2006.
- [10]. S. Gao, Y. He, H. Chen, *Wind Speed Forecast for Wind Farms Based on ARMA-ARCH Model*, International Conference on Sustainable Power Generation and Supply, 6–7 April 2009.
- [11]. J. Jeon and J. W. Taylor, *Short-term density forecasting of wave energy using ARMA-GARCH models and Kernel density estimation*, International Journal of Forecasting.
- [12]. D. P. Kroese, J. C. C. Chan, *Statistical Modeling and Computation*, Springer, 2014, pp. 227–256.
- [13]. Kim, S., N. Shepherd, S. Chib. *Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models*. Review of Economic Studies, 65(3):361–393, 1998.
- [14]. Muhammad Naeem, *A comparison of electricity spot prices simulation using ARMA-GARCH and mean-reverting models*, Master's Thesis, Lappeenranta University of Technology, 2010.
- [15]. Elena-Ivona D., C. Hurlin, J. Madkour, *Testing Interval Forecasts: A GMM-Based Approach*, Journal of Forecasting, Volume 32, Issue 2, pp 97–110, March 2013.

SHORT-TERM PROBABILITY FORECASTING FOR WIND POWER OUTPUT BASED ON STOCHASTIC FLUCTUATION MODEL

*Do Phuong Nhung[‡], Nguyen Trong Khiem[§], Dang Minh Quan[§],
Dang Hoang Anh[§], Quach Thi Hanh[§], Nguyen Thi Phi Doan[§]*

Abstract: *Wind power has become a very viable alternative solution to provide electricity. Consequently, probabilistic wind power forecast, which provides not only predictive deterministic wind power but also information about the uncertainty of the predicted values, becomes more important. However, most existing forecast approaches focus on point forecasts without considering density forecasts. Thus, the application of their results in the power system's stochastic problems is limited. This paper presents a new probabilistic wind power forecast approach called the Stochastic Volatility model, which has high accuracy for density forecasting. The Hierarchical Bayesian Method is used to estimate the proposed model. The stochastic volatility model is applied to forecast wind power based on accurate data from Japan in 2005. The accuracy of the Stochastic Volatility model for both deterministic and probability forecasts with different forecast horizons will be analyzed and compared to the ARMA model.*

Keywords: *ARMA model, bayesian, probabilistic forecast, stochastic volatility, wind power forecast.*

[‡] Hanoi Open University

[§] School of Electrical and Electronic Engineering, University of Science and Technology