

Nghiên cứu dự báo tốc độ gió khu vực Quảng Trị sử dụng mạng nơ-ron nhân tạo NARX

Study on wind energy forecasting of Quang Tri region using artificial neural networks NARX

Nguyễn Đức Hạnh¹, Nguyễn Văn Hùng¹, Phùng Văn Thái¹, Trương Văn Quảng², Tống Thị Lý²,
Trịnh Trọng Chương^{2*}

¹Viện năng lượng, ²Trường Đại học Công nghiệp Hà nội

*Corresponding author E-mail: chuongtt@hau.edu.vn

Abstract

Quang Tri Province has a rich source of wind energy, but effectively harnessing this valuable resource requires consideration of several factors. One of these is wind forecast databases, which are crucial to optimizing wind energy utilization. The purpose of this study is to develop a wind speed forecasting model based on artificial neural networks using the NARX structure. Our model was constructed using historical data on wind speed and meteorological factors in Quang Tri. A local monitoring station's wind speed is used as an input variable for the NARX model. The MAPE and MAE metrics indicate that the forecasting model is highly effective, even under changing weather conditions. The proposed model has also been compared with models that use additional correlated input data, including wind direction, temperature, pressure and humidity, at various altitudes of 40m, 60m, 80m. The study also shows that choosing appropriate NARX model parameters can potentially expand the early forecast range up to 7 days. As a result of this validation, not only is the accuracy of the model determined, but also how the model can be used in order to harness wind energy efficiently and sustainably. Local wind energy management systems can benefit from this forecasting model by improving resource utilization and minimizing risks.

Keywords: Mạng thần kinh nhân tạo (ANN), NARX (Nonlinear Autoregressive with Exogenous Inputs), MAE (Mean Absolute Error), Mean Absolute Percentage Error (MAPE)

Tóm tắt

Tỉnh Quảng Trị có nguồn năng lượng gió phong phú, tuy nhiên, để khai thác hiệu quả nguồn tài nguyên quý giá này, nhiều yếu tố cần được xem xét. Trong đó, cơ sở dữ liệu dự báo về gió giữ vị trí then chốt trong việc khai thác hiệu quả nguồn năng lượng. Nghiên cứu này phát triển một mô hình dự báo tốc độ gió dựa trên mạng nơ-ron nhân tạo với cấu trúc NARX. Để xây dựng mô hình, nghiên cứu đã thu thập và phân tích dữ liệu quá khứ về tốc độ gió cùng các yếu tố khí tượng tại trạm quan trắc địa phương. Mô hình NARX được huấn luyện đầu tiên với các biến đầu vào là tốc độ gió quá khứ, đầu ra là tốc độ gió dự báo. Kết quả đánh giá bằng chỉ số MAPE và MAE cho thấy mô hình dự báo rất hiệu quả, ngay cả trong điều kiện thời tiết thay đổi. Mô hình đề xuất cũng đã được so sánh với các mô hình có sử dụng thêm cả các dữ liệu tương quan đầu vào, bao gồm hướng gió, nhiệt độ, áp suất và độ ẩm, ở nhiều độ cao khác nhau 40m, 60m, 80m. Nghiên cứu cũng chỉ ra việc lựa chọn các thông số của mô hình NARX phù hợp có khả năng cho phép mở rộng khoảng dự báo sớm lên tới 7 ngày. Việc thực hiện kiểm tra này không chỉ giúp xác định độ chính xác của mô hình mà còn góp phần tối ưu hóa việc khai thác nguồn năng lượng gió một cách hiệu quả và bền vững. Mô hình dự báo này có thể áp dụng vào các hệ thống quản lý năng lượng gió tại địa phương, nhằm cải thiện hiệu quả sử dụng tài nguyên và giảm thiểu rủi ro trong sản xuất năng lượng tái tạo.

1. Giới Thiệu

Năng lượng tái tạo đang trở thành lĩnh vực ngày càng được khai thác phổ biến trong hệ thống năng lượng toàn cầu. Dự báo, năng lượng tái tạo sẽ tăng trưởng 7,1% mỗi năm trong hai thập kỷ tới [1]. Tại Việt Nam, năng lượng tái tạo đã và đang đóng góp một phần đáng kể vào hệ thống năng lượng quốc gia. Tính đến hết năm 2021, tổng công suất lắp đặt các nguồn năng lượng tái tạo đạt 20.670 MW, chiếm 27% tổng công suất lắp đặt toàn hệ thống (76.620 MW). Sản lượng điện từ nguồn năng lượng tái tạo đã đạt 31,508 tỷ kWh, tương đương 12,27% tổng sản lượng điện sản xuất toàn hệ thống [2]. Trong vài thập kỷ qua, nước ta ghi nhận tốc độ tăng trưởng kinh tế nhanh chóng với mức tăng trưởng trung bình 6,2% mỗi năm từ 2002 đến 2022 và GDP bình quân đầu người tăng 3,6 lần, đạt gần 3.700 USD, nhưng đi kèm là những thách thức về an ninh năng lượng và ô nhiễm môi trường. Thách thức đặt ra là tiến tới nguồn năng lượng sạch và bền vững, đa dạng hóa nguồn cung cấp. Tốc độ tăng trưởng nhu cầu tiêu thụ năng lượng hiện khoảng 9,5% mỗi năm. Giai đoạn 2016-2017, mức tiêu thụ năng lượng của Việt Nam ghi nhận hơn 184 tỷ kWh/năm, trong khi hệ thống điện quốc gia chỉ có khả năng sản xuất khoảng 170 tỷ kWh mỗi năm theo báo cáo của Tập đoàn Điện lực Việt Nam [3]. Chính phủ và Tập đoàn Điện lực Việt Nam đã hoạch định kế hoạch và định hướng phát triển cho ngành năng lượng trong vài thập kỷ tới, với mục tiêu tăng gấp đôi sản lượng điện lên

129.500 MW vào năm 2030 [4]. Theo quy hoạch điện VII và VIII, tỷ trọng nguồn năng lượng tái tạo nói chung, và năng lượng gió nói riêng, sẽ đứng thứ hai, như trong Bảng 1:

Bảng 1: Tỷ trọng nguồn năng lượng của nước ta đến năm 2030

Nguồn năng lượng	Tỷ trọng
Hydropower	19,5%
Nhiệt điện (từ than)	20%
Điện gió trên bờ	14,5%
Điện gió ngoài khơi	4%
Điện mặt trời	5%
Pin lưu trữ	0,2%
Điện nhập khẩu	3,3%
Nhiệt điện LNG	14,9%
Điện sinh khối, điện rác	1,5%

Việt Nam có tiềm năng năng lượng gió cao, với hơn 39% tổng diện tích được ước tính có tốc độ gió trung bình hàng năm vượt quá 6 m/s ở độ cao 65 m, tương đương với công suất 512 GW. Đặc biệt, gần 8% diện tích Việt Nam được xếp hạng có tiềm năng gió rất tốt, với tốc độ gió ở độ cao 65 m dao động từ 7 đến 8 m/s [5]. Đây là một nguồn năng lượng khổng lồ mà thiên nhiên đã ban tặng, và chúng ta cần tận dụng để phát triển đất nước một cách bền vững.

Tuy nhiên, bên cạnh tiềm năng to lớn, các nhà khai thác điện gió trên đất liền cũng phải đối mặt với nhiều thách thức. Mặc dù có một số lợi thế trong việc phát triển cơ sở hạ tầng, nhưng lượng gió lại phân bố không đều cả về không gian và thời gian. Theo khảo sát như tại *Bảng 2*, tại các tỉnh Nam

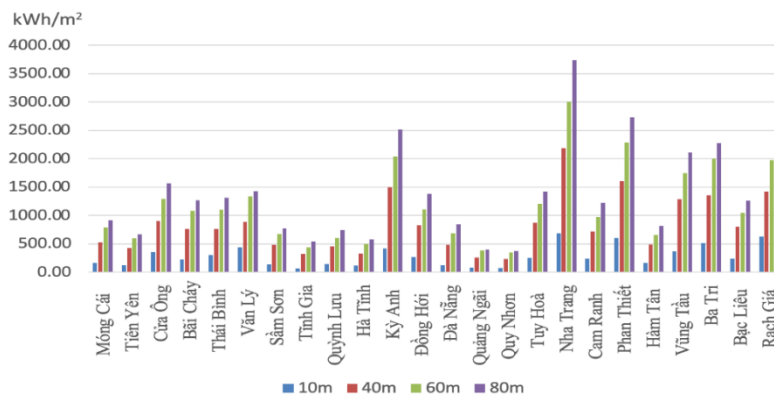
Trung Bộ và Tây Nguyên cho thấy có tiềm năng gió khá lớn. Hiện tại, hầu hết các nhà máy điện gió ở Việt Nam có quy mô vừa và nhỏ như *Bảng 3*, do chúng ta vẫn đang trong giai đoạn đầu của ngành công nghiệp này và vẫn còn nhiều cơ hội để phát triển nguồn năng lượng gió tại khu vực [6].

Bảng 2: Tiềm năng của năng lượng gió GW

Khu vực	Trên đất liền
Đông Bắc	4.6
Tây Bắc	2.8
Đồng bằng sông Hồng	1.5
Bắc Trung bộ	0.3
Nam Trung bộ	16.8
Tây Nguyên	12.5
Đông Nam bộ	3.3
Đồng bằng sông Cửu Long	0.2
<i>Tổng cộng</i>	<i>42.0</i>

Bảng 3: Một số nhà máy điện gió của Việt Nam 2024

Dự Án	Công Suất (MW)	Sản lượng điện hàng năm (triệu KWh)
Điện gió V1-2 Trà Vinh	48	163
Hướng Linh 1,2	60	244,7
Hướng Phùng 1	30	81,1
Mũi Dinh	37,6	105
Bình Thạnh	30	85
Hướng Hiệp	30	126,3
Đàm Nai	40	110



Hình 1: Phân bố gió ven biển theo độ cao (VIASEE)

Các nhà đầu tư và các chuyên gia phải đối mặt với nhiều thách thức trong đó chúng ta có thể kể đến như công nghệ kỹ thuật hay nguồn nhân lực. Trước tiên chúng ta sẽ phải đối mặt với các thách thức chung là vấn đề kỹ thuật và sau đó là những vấn đề từ nội tại chúng ta [7], [8]. Vì Việt Nam là đất nước thuộc vùng nhiệt đới gió mùa nên có khí hậu biến đổi theo từng khoảng thời gian trong năm [9]. Việc khảo sát bước đầu để mở rộng hay đầu tư dự án mới rất quan trọng nó ảnh hưởng trực tiếp đến sản lượng hay độ tin cậy của hệ thống [10], [11]. Để đánh giá đầu tư phát triển các trang trại gió, các phương pháp dựa vào kinh nghiệm và các phương pháp hồi quy dữ liệu thông thường được sử dụng [12], [13].

Một số phương pháp hiện đại hơn cũng được triển khai như sử dụng các hình ảnh vệ tinh trực tiếp để dự đoán. Một số mô hình phổ biến bao gồm mạng nơ-ron hồi tiếp (RNN) và

mạng nơ-ron tích chập (CNN), giúp dự đoán tốc độ và hướng gió với độ chính xác cao dựa trên dữ liệu thời tiết lịch sử. Ngoài ra, các phương pháp như học tăng cường (Reinforcement Learning) cũng được nghiên cứu để tối ưu hóa các hệ thống năng lượng gió. Tuy nhiên, các biện pháp trên chưa thể hiện được sự hiệu quả của mình do các hạn chế về độ tin cậy hay chi phí quá cao so với doanh nghiệp hoặc một số các quốc gia. Trong đó một vấn đề rất lớn đó là xây dựng cơ sở dữ liệu bản đồ gió cho các khu vực có nguồn gió lớn đang đặt ra những thách thức. Vì bản đồ gió không những thể hiện phân bố gió về cả thời gian và không gian. *Hình 1* dưới đây là biểu đồ phân bố gió thể hiện tổng năng lượng gió khác nhau tại một số vùng ven biển của Việt Nam ở độ cao 10 m, 40 m, 60 m và 80 m trên cả nước. Tuy nhiên, việc đánh giá dữ liệu phân bố gió này mang tính tổng quan và chưa chi tiết cho các vùng miền cụ thể.

Để đạt được mục tiêu này, việc xây dựng cơ sở dữ liệu gió chi tiết tại từng khu vực là cần thiết, nhằm cung cấp thông tin chính xác cho công tác dự báo. Cơ sở dữ liệu này không chỉ giúp phân tích các xu hướng biến thiên của năng lượng gió mà còn hỗ trợ trong việc ra quyết định cho các dự án đầu tư và phát triển hệ thống năng lượng gió. Một số phương pháp tiên tiến hiện đang được áp dụng trong dự báo gió đã được trình bày trong *Bảng 4* dưới đây, bao gồm các phương pháp mô phỏng, mạng nơ-ron nhân tạo (ANN), và tối ưu hóa bầy đàn (PSO). Những phương pháp này không chỉ giúp cải thiện độ chính xác của dự báo mà còn giảm thiểu rủi ro trong quá trình vận hành các tuabin gió.

Bảng 4: Một số phương pháp dự báo gió

Phương pháp	Tổ Chức	Bài báo
Mô phỏng	Usability and Challenges of Offshore Wind Energy in Vietnam Revealed by the Regional Climate Model Simulation	[11]
CFS/RSM	Tạp chí khoa học công nghệ Việt nam ISSN 1859-4794	[13]
ANN	Green Technology and Sustainable Development; Renewable Energy	[14] [3] [15]
LSTM	Elsevier	[15] [16]

Mô phỏng có ưu điểm là tính chính xác cao và khả năng đánh giá nhiều yếu tố khí tượng, nhưng cần lượng lớn dữ liệu và thời gian tính toán dài. CFS/RSM cung cấp dự báo dài hạn và thích ứng với điều kiện địa phương, tuy nhiên độ chính xác có thể thay đổi và yêu cầu nguồn lực lớn. ANN có khả năng xử lý các mối quan hệ phi tuyến tính và cải thiện qua thời gian, nhưng thông thường chỉ yêu cầu dữ liệu đầu vào hiện tại để đưa ra dự đoán. Các mối quan hệ giữa các đầu vào và đầu ra có thể không hoàn toàn rõ ràng. Cuối cùng, PSO tối ưu hóa nhanh chóng và dễ áp dụng, nhưng có thể mắc kẹt trong cực tiểu địa phương và cần điều chỉnh tham số.

Mô hình NARX là một phương pháp mạnh mẽ trong dự báo năng lượng gió, kết hợp giữa khả năng hồi quy phi tuyến và các yếu tố ngoại sinh [17]. Ưu điểm của NARX là khả năng nắm bắt các mối quan hệ phức tạp giữa các biến động trong dữ liệu lịch sử và các yếu tố bên ngoài, giúp cải thiện độ chính xác của dự báo. Được tối ưu hóa để xử lý dữ liệu chuỗi thời gian, giúp nó có khả năng tốt hơn trong việc dự đoán các sự kiện tương lai dựa trên các xu hướng đã diễn ra. Tuy nhiên, NARX cũng có một số nhược điểm, bao gồm việc yêu cầu một lượng lớn dữ liệu để huấn luyện và khó khăn trong việc xác định cấu trúc mô hình phù hợp. Điều này có thể dẫn đến việc mô hình không tối ưu hóa hiệu suất nếu dữ liệu đầu vào không đầy đủ hoặc không chính xác. Do đó, việc lựa chọn và điều chỉnh mô hình NARX cần được thực hiện cẩn thận để đạt được kết quả tốt nhất trong dự báo năng lượng gió.

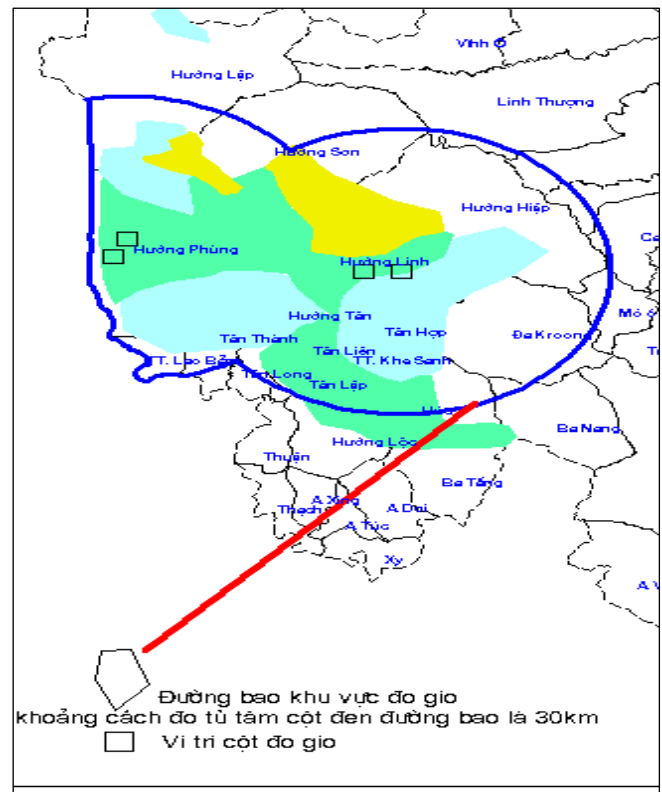
Trong bài báo này, mạng nơ-ron nhân tạo NARX được đề xuất để dự báo tốc độ gió tại tỉnh Quảng Trị, Việt Nam. Trước tiên, dữ liệu thu thập từ một trạm đo sẽ được phân tích, bao gồm các thông tin khí tượng như nhiệt độ, độ ẩm, hướng gió và áp suất tại các độ cao 40 m, 60 m và 80 m. Tiếp theo, mô hình mạng NARX với 10 lớp ẩn sẽ được xây dựng, sử dụng dữ liệu tốc độ gió của tháng 3 năm 2018. Tốc

độ gió sẽ được dự báo cho các khoảng thời gian 10 phút, 30 phút, 60 phút, 180 phút và 360 phút tại từng độ cao 40 m, 60 m và 80 m để đánh giá hiệu suất của mô hình. Để đào tạo mạng nơ-ron, dữ liệu khí tượng trong tháng 4 cùng năm sẽ được sử dụng. Sau khi được đào tạo, mạng nơ-ron sẽ được kiểm chứng về độ chính xác và độ tin cậy với bộ dữ liệu của 11 tháng liên tiếp. Kết quả được đánh giá bằng chỉ số MAPE, MAE và MSE cho thấy mô hình dự báo rất hiệu quả, ngay cả trong điều kiện thời tiết thay đổi. Điều này đặc biệt quan trọng đối với các khu vực có tiềm năng phát triển năng lượng tái tạo như tỉnh Quảng Trị.

2. Xây dựng mô hình mạng NARX dự báo năng lượng gió tại Quảng trị

2.1. Phân tích dữ liệu gió thu thập tại Quảng Trị

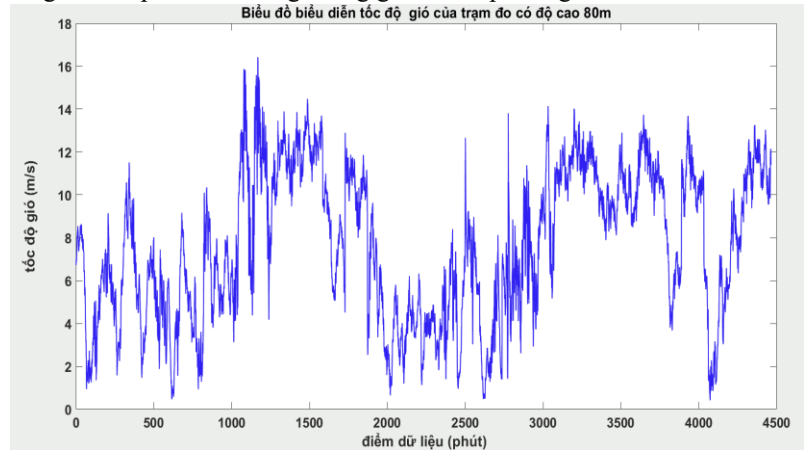
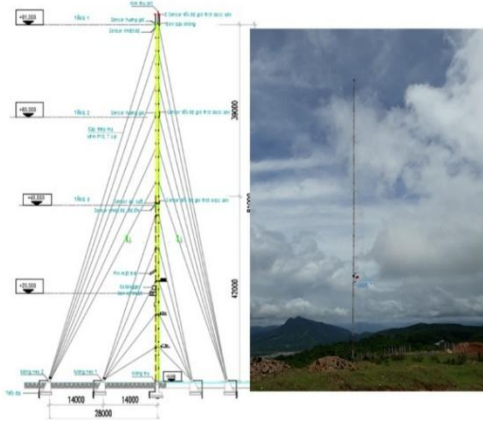
Dữ liệu gió cho nghiên cứu này đã được thu thập từ các trạm đo trên địa bàn tỉnh Quảng Trị trong khoảng thời gian từ tháng 8 năm 2017 đến tháng 7 năm 2018. Các tham số được ghi nhận bao gồm tốc độ gió cực đại, tốc độ gió cực tiểu, tốc độ gió trung bình, hướng gió, nhiệt độ, độ ẩm, và áp suất. Cụ thể, dữ liệu gió sử dụng cho nghiên cứu này được thu thập trạm quan sát địa phương (Trụ đo gió số 1, Dự án điện gió Hướng Linh - Hướng Phùng-Hải Anh), với các cảm biến được lắp đặt ở các độ cao khác nhau: 10 m, 40 m, 60 m và 80 m. Việc ghi nhận dữ liệu chi tiết này giúp đảm bảo tính chính xác và độ tin cậy của mô hình dự báo, đặc biệt trong việc đánh giá tiềm năng năng lượng gió tại khu vực này.



Hình 2: Vị trí đo gió ở độ cao theo số liệu quan trắc tại Quảng Trị

Ngoài ra, sự phân bố dữ liệu từ các trạm đo gió khác tại tỉnh Quảng Trị được thiết kế để đảm bảo tính khách quan và đa dạng. Tốc độ gió được chọn làm biến đầu vào cho mô hình

dự báo dựa trên mối liên hệ chặt chẽ giữa tốc độ gió và công suất của tuabin gió, yếu tố quyết định chủ yếu đến lượng năng lượng thu được từ hệ thống [18]. Tốc độ gió thường được coi là một yếu tố khó xác định bằng các phương pháp đo đạc thông thường. Do đó, nghiên cứu này nhằm cung cấp cái nhìn sâu sắc hơn về các phương pháp dự báo được áp dụng, với mục tiêu nâng cao độ chính xác trong việc dự đoán tiềm năng năng lượng gió. Kết quả của nghiên cứu được xác định và phân tích dựa trên dữ liệu đo gió từ bốn vị trí khác nhau, bao gồm khu vực Hướng Linh và Hướng Phùng. Sự



Hình 3: Đồ thị điểm dữ liệu và tốc độ gió tại độ cao 80m tháng 3 năm 2017

Để đánh giá tiềm năng năng lượng gió phục vụ cho các tuabin gió, trong nghiên cứu này sẽ dự báo tốc độ gió tại các độ cao 40 m, 60 m và 80 m sẽ được dự báo dựa trên các dữ liệu thu thập được từ các độ cao này. Bởi vì tốc độ gió là một yếu tố chính quyết định chính để tính toán năng lượng gió, khi gió mạnh hơn sẽ giúp sản xuất điện nhiều hơn. Hình 3 minh họa dữ liệu đã được ghi lại với chu kỳ lấy mẫu là 10 phút trong tháng 3 năm 2018, tổng cộng có 4.464 điểm dữ liệu được thu thập tại độ cao 80 m. Phân tích biểu đồ cho thấy sự biến thiên đáng kể của tốc độ gió, với biên độ thay đổi lên tới 15 m/s và tốc độ biến thiên cao trong khoảng thời gian ngắn. Điều này cho thấy rằng việc áp dụng các phương pháp toán học thông thường để tính toán quỹ đạo gió có thể không khả thi. Khảo sát dữ liệu thu thập tại các độ cao 40 m và 60 m cũng cho thấy kết quả biến thiên tương tự như ở độ cao 80 m. Để nâng cao độ chính xác trong dự báo, nghiên cứu đã áp dụng mạng nơ-ron nhân tạo, với hy vọng rằng phương pháp này sẽ khắc phục được những hạn chế của các mô hình toán học truyền thống. Với sự phát triển nhanh chóng của năng lượng gió, nhu cầu về các phương pháp và kỹ thuật dự báo tốc độ gió ngắn hạn chính xác và tin cậy đang trở nên cấp thiết.

2.2. Xây dựng cấu trúc mô hình mạng NARX

Mô hình NARX là một loại mạng nơ-ron được thiết kế đặc biệt để xử lý các chuỗi thời gian phi tuyến và dự đoán các thông số đầu ra dựa trên các yếu tố đầu vào bên ngoài. Khác với các mô hình tự hồi quy thông thường, NARX có khả năng kết hợp cả các giá trị trước đó của chính đầu ra và các đầu vào exogenous, cho phép nó nắm bắt được các mối quan hệ phức tạp giữa các biến trong hệ thống. Mô hình NARX có nhiều ưu điểm nổi bật, đặc biệt trong các bài toán dự báo

kết hợp với Bản đồ nguồn gió của Việt Nam, được công bố bởi Ngân hàng Thế giới vào năm 2010, đã chỉ ra tiềm năng gió đáng kể tại các khu vực này. Cụ thể, các vùng có tốc độ gió từ 5,5 m/s trở lên chủ yếu tập trung ở các huyện miền núi như Hướng Hóa, cùng với một số khu vực nhỏ ở huyện Đrông và huyện Cam Lộ. Thêm vào đó, các huyện ven biển của tỉnh Quảng Trị cũng thể hiện tiềm năng gió tương đối cao. Hình 3 minh họa vị trí các cột đo gió và những khu vực có tốc độ gió đạt từ 5,5 m/s trở lên, từ đó cho thấy khả năng phát triển năng lượng gió tại địa phương.

chuỗi thời gian. Đầu tiên, nó có khả năng mô hình hóa mối quan hệ phi tuyến giữa các biến, làm cho nó phù hợp với các hệ thống phức tạp. Thứ hai, NARX sử dụng dữ liệu quá khứ để dự đoán giá trị tương lai, rất hữu ích cho các bài toán tự hồi quy như dự báo thời tiết hoặc tài chính. Thứ ba, mô hình này có thể tận dụng dữ liệu từ các yếu tố bên ngoài để cải thiện độ chính xác dự báo. Hơn nữa, NARX có khả năng tích hợp dễ dàng với các mạng nơ-ron hồi quy, nâng cao khả năng xử lý dữ liệu lớn và phức tạp. Cuối cùng, nhờ vào khả năng học từ dữ liệu quá khứ và yếu tố ngoại sinh, NARX có thể giảm thiểu lỗi dự báo, mang lại kết quả chính xác hơn so với các mô hình tuyến tính thông thường. Những ưu điểm này đã khiến NARX trở thành một công cụ hữu ích trong việc dự báo và phân tích dữ liệu trong nhiều lĩnh vực khác nhau.

Hình 4 trình bày mô hình mạng nơ-ron NARX, tức là mạng nơ-ron phi tuyến tự hồi quy với đầu vào ngoại sinh. Trong đó mô hình minh họa cấu trúc cơ bản của mạng nơ-ron nhân tạo, được cấu thành từ các nút mạng, trong đó các thông số đầu vào được máy tính thu nhận và xử lý thông qua các lớp ẩn để đưa ra các trọng số. Mô hình mạng NARX được phát triển dựa trên cấu trúc đơn giản này.

Mạng NARX có khả năng học để dự đoán các giá trị trong một khoảng thời gian nhất định dựa trên các giá trị trong quá khứ và các giá trị ngoại lai từ các thời điểm khác. Mạng NARX có hai dạng: dạng vòng hở và dạng vòng kín, và trong nghiên cứu này, dạng vòng hở được sử dụng. Ở dạng vòng hở, mạng NARX có hai loại đầu vào và một loại mục tiêu [19].

- Loại đầu vào thứ nhất, $x(t)$ bao gồm các giá trị ngoại lai theo thời gian của các thông số khác ngoài thông số cần dự đoán.

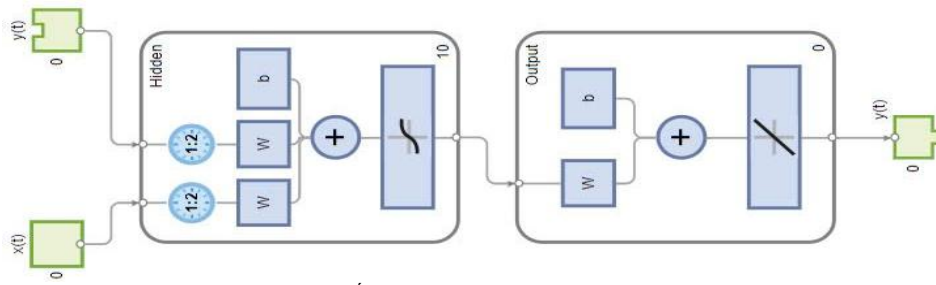
- Loại đầu vào thứ hai, $y(t)$ bao gồm các giá trị đã được đo trong quá khứ của thông số cần dự đoán, đồng thời cũng đóng vai trò là mục tiêu của mạng.

Trong mô hình mạng NARX, multi-layer perceptron được sử dụng để xấp xỉ công thức sau, với f là một hàm phi tuyến:

$$y(t) = f(x(t - T), x(t - 2T), \dots, x(t - nT), y(t - T), y(t - 2T), \dots, y(t - nT)) \tag{1}$$

với $n \in N^*$ là số điểm dữ liệu quá khứ được sử dụng để dự báo, và T là bước thời gian. Hàm f làm hàm thể hiện mối quan hệ của dữ liệu thời điểm hiện tại vào các dữ liệu quá khứ.

Trong nghiên cứu này, mô hình mạng NARX được sử dụng với một lớp ẩn, bao gồm 10 nơ-ron, và số lượng trễ của các đầu vào $x(t)$ và $y(t)$ được thiết lập là 2.



Hình 4: Cấu trúc mạng NARX trong MATLAB

Như vậy, mô hình NARX đề xuất được huấn luyện để dự báo tốc độ gió tại tỉnh Quảng Trị dựa vào nguồn dữ liệu đã được thu thập, một tập dữ liệu đã được tạo ra và biểu diễn dưới dạng ma trận có kích thước 5x1464 để tiến hành huấn luyện cho mô hình. Cấu hình máy tính được sử dụng để đào tạo mô hình là ASUSTeK COMPUTER INC (Intel(R) Core(TM) i7-1355U thế hệ thứ 13, 1700 MHz, 10 lõi). Tập dữ liệu huấn luyện bao gồm 31 ngày, tương ứng với 4164 điểm dữ liệu, với độ trễ (delay) được thiết lập là 2, số lớp ẩn là 1, và số nơ-ron trong mỗi lớp ẩn là 10.

3. Huấn luyện mô hình NARX sử dụng tốc độ gió làm thông số đầu

3.1. Phân chia dữ liệu và huấn luyện mô hình

Trong nghiên cứu thực hiện huấn luyện cho mô hình NARX trong mục này, dữ liệu sử dụng được cung cấp ở mục 2- ứng với bộ dữ liệu đo tại độ cao 80m. Để đánh giá chính xác tính hiệu quả của mô hình nghiên cứu đã thử nghiệm với các kịch bản khác nhau, sử dụng dữ liệu đầu vào là tốc độ gió và dự đoán đầu ra là tốc độ gió.

Cụ thể, trong kịch bản huấn luyện mục này, các khoảng dữ liệu đầu vào được sử dụng là $a = 12, 18, 36, 72, 144, 216$ điểm dữ liệu, tương ứng với khoảng thời gian trước đó là 2 giờ, 3 giờ, 6 giờ, 12 giờ, 1 ngày và 1,5 ngày. Các khoảng dữ liệu mục tiêu để dự đoán được xác định là $b = 1, 3, 6, 18, 36$ điểm dữ liệu, tương ứng với việc dự báo trong các khoảng thời gian 10 phút, 30 phút, 1 giờ, 3 giờ và 6 giờ.

Thời gian dự báo cho mô hình trong Bảng 5 dựa trên lượng đầu vào và đầu ra, trong đó a được sử dụng để chỉ lượng dữ liệu đầu vào cho quá trình huấn luyện, và b là khoảng thời

Nhận xét 1: Sau khi thay đổi số nơ-ron từ 5 đến 15 thì việc chọn số nơ-ron là 10 đã đáp ứng được yêu cầu về độ chính xác từ 80 - 90% đồng thời đơn giản hóa mô hình và đảm bảo khả năng xử lý phù hợp với cấu hình máy tính hiện tại. Nếu số nơ-ron được tăng lên, mô hình sẽ cần nhiều tài nguyên tính toán hơn, có thể dẫn đến tình trạng chậm trễ hoặc treo máy trên các máy tính thông thường.

Nhận xét 2: Số trễ được chọn là 2 vì nó đủ để mô hình học từ dữ liệu có đặc tính động lực học phức tạp một cách hiệu quả và đáp ứng được yêu cầu về độ chính xác. Nếu số trễ quá thấp, mô hình có thể không nắm bắt được mối quan hệ dài hạn giữa các biến. Ngược lại, nếu số trễ quá cao, mô hình có thể trở nên quá phức tạp, dễ dẫn đến hiện tượng “overfitting”. Sự lựa chọn số trễ này được dựa trên tập dữ liệu hiện có sau khi thay đổi độ trễ (delay) từ 2 đến 10, nhằm đảm bảo tính phù hợp và hiệu quả cho mô hình NARX.

gian dự báo tương ứng. Thời gian đào tạo của mô hình được xác định bởi cả a và b , cũng như cấu hình máy tính đã được đề cập trước đó.

Bảng 5 Thời gian đào tạo mạng NARX

$b \backslash a$	1	2	18	36	72	144	216
1	0s	3s	2s	18s	1m 12s	2m 21s	
3	1s	15s	8s	1m 21s	2m 54s	3m 1s	
6	3s	1m	16s	1m 54s	7m 09s	8m 18s	
18		10m 09s	2m 28s	19m 01s	41m 12s	1h 8m 51s	
36			22m 14s	1h 33m 38s	1h 14m 34s		

Nhận xét 3: Một yếu tố quan trọng trong việc đánh giá thời gian đào tạo của mô hình là sự đánh đổi giữa thời gian và độ chính xác. Trong quá trình huấn luyện mạng NARX, nếu mục tiêu là đạt được tốc độ hội tụ nhanh, mô hình có thể đưa ra những kết quả không chính xác. Điều này có thể xảy ra khi mô hình không được huấn luyện đủ lâu để học từ dữ liệu một cách hiệu quả. Do đó, việc cân nhắc giữa thời gian huấn luyện và chất lượng dự đoán là rất cần thiết.

3.2. Kết quả đào tạo mạng NARX

Việc đánh giá mô hình sau huấn luyện dựa vào sai số tuyệt đối trung bình MAE và sai số tương đối trung bình MAPE:

- MAPE đo lường sai số trung bình dưới dạng phần trăm, có thể hiểu rằng đó là mức độ sai lệch giữa giá trị dự đoán và giá trị thực. Tuy nhiên, MAPE có thể không phù hợp nếu giá trị thực tiến tới 0, vì nó có thể làm sai lệch kết quả.

- Trong khi đó MAE lại không phụ thuộc vào thang đo giúp đánh giá trực quan về hiệu suất của mô hình.

Việc kết hợp hai thông số này sẽ đánh giá được một cách tổng quát mô hình và đưa ra nhận xét chính xác rằng mô hình đó có thực sự hiệu quả hay không. Công thức tính sai số tuyệt đối trung bình (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2)$$

Công thức tính sai số tương đối trung bình (MAPE):

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (3)$$

Trong đó N là số điểm dữ liệu khảo sát, y_i là giá trị dữ liệu thực tế, \hat{y}_i là giá trị đầu ra dự báo của mô hình. Dựa trên kết quả đánh giá thông số MAE và MAPE để đánh giá hiệu quả của mô hình đề xuất cho nghiên cứu này. Qua *Bảng 6* và *Bảng 7* ta thấy rằng sai số MAPE và MAE ở ô tương ứng với việc sử dụng 144 điểm dữ liệu (m ngày) làm đầu vào (input) và 6 điểm dữ liệu (một giờ) làm mục tiêu (target) nhỏ nhất.

Bảng 6: Thông số MAPE

b \ a	12	18	36	72	144	216
1	7,883 7	8,600 9	7,743	7,7073	8,2277	8,2615
3	7,823 0	8,370 6	7,6824	7,5747	7,7931	8,0705
6	7,898 0	8,373 2	7,8740	7,7144	7,4511	8,1217
18		8,485 1	8,0592	7,8490	7,8814	8,0185
36			8,5015	8,2704	8,4703	

Bảng 7: Thông số MAE

b \ a	12	18	36	72	144	216
1	0,436 0	0,468 3	0,4349	0,4387	0,4674	0,4568
3	0,439 7	0,460 3	0,4299	0,4219	0,4245	0,4463
6	0,440 7	0,464 1	0,4430	0,4345	0,4172	0,4510
18		0,470 0	0,4514	0,4487	0,4481	0,4580
36			0,4722	0,4659	0,4752	

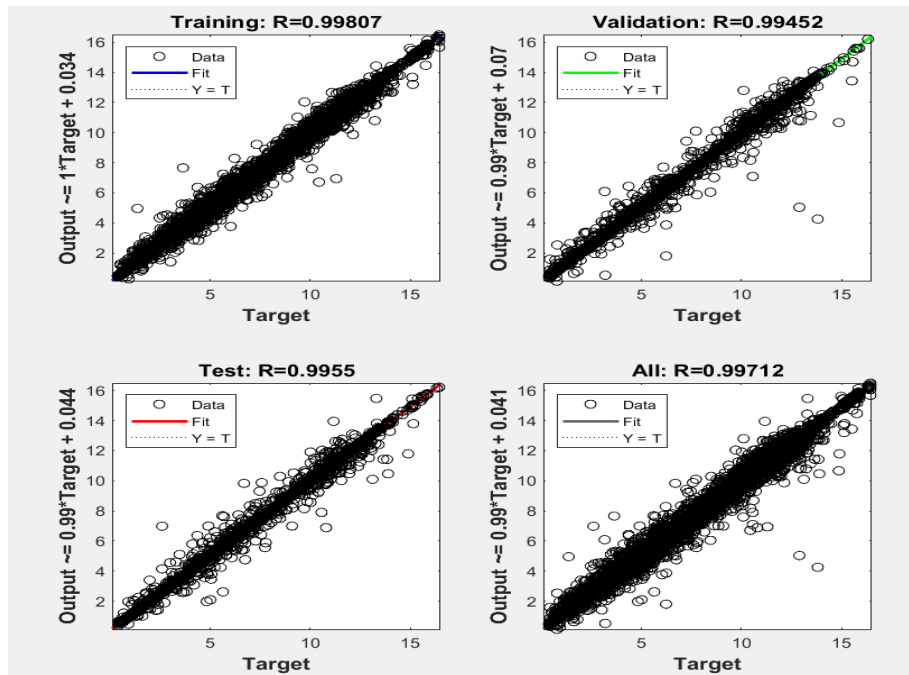
+Kết quả huấn luyện (Training): Sử dụng 70% tập dữ liệu ban đầu cho tập huấn luyện. Trong giai đoạn huấn luyện, mô hình đã được tối ưu hóa để tìm ra các tham số tốt nhất nhằm dự báo chính xác. Số nơ-ron và độ trễ được thay đổi theo kịch bản, bộ trọng số của mạng được cập nhật liên tục bằng thuật toán lan truyền ngược để đạt được giá trị đầu ra mong muốn (sai số tối thiểu nhất có thể). Những tham số tốt nhất chính là bộ trọng số của mạng, số nơ-ron, độ trễ. Kết quả trong *Bảng 7* cho thấy MAE đạt giá trị thấp, chứng tỏ mô hình đã học tốt từ dữ liệu quá khứ và có khả năng dự báo chính xác. Tuy nhiên, cần lưu ý rằng độ chính xác cao trong

tập huấn luyện không đảm bảo hiệu quả khi mô hình dự báo trên dữ liệu mới.

+ Kết quả kiểm tra (Testing): Sử dụng 15% tập dữ liệu ban đầu cho tập kiểm tra. Trên tập kiểm tra, mô hình cho thấy khả năng tổng quát hóa dữ liệu với sai số nhỏ. MAPE trong *Bảng 6* cho thấy mức độ sai lệch giữa dự báo và thực tế trong khoảng chấp nhận, thể hiện mô hình có khả năng dự báo tốt với các dữ liệu không nằm trong tập huấn luyện. Việc duy trì sai số thấp trên cả hai tập huấn luyện và kiểm tra là tín hiệu tích cực cho thấy mô hình không bị quá khớp (overfitting).

+ Kết quả xác thực (Validating): Sử dụng 15% tập dữ liệu ban đầu cho tập xác thực. Trong quá trình xác thực với một tập dữ liệu hoàn toàn mới, mô hình đạt kết quả như *Hình 5* và sai số MAPE như trên *Bảng 6* trong giới hạn mong đợi, tuy nhiên sai số có xu hướng cao hơn so với tập huấn luyện và kiểm tra. Điều này cho thấy mô hình có thể bị ảnh hưởng bởi những biến động không nằm trong phạm vi dự đoán ban đầu. Tuy vậy, kết quả xác thực vẫn nằm trong mức độ chấp nhận, chứng tỏ mô hình có khả năng ứng dụng thực tế. *Hình 5* thể hiện hệ số hồi quy của mạng sử dụng một ngày để dự đoán cho một giờ. Tất cả các giai đoạn đều có hệ số hồi quy R gần như bằng một (tương ứng với độ khớp dữ liệu gần 100%) cho ta thấy qua huấn luyện với dữ liệu của tháng 3, đầu ra của mô hình gần như chính xác với mục tiêu, điều đó chứng minh rằng mô hình đã được huấn luyện tốt và có khả năng dự đoán chính xác hơn rất nhiều. Mỗi biểu đồ đều cho thấy đường hồi quy tuyến tính (Fit), so với đường lý tưởng $Y = T$ (giá trị dự đoán Y bằng giá trị thực T). Mô hình thể hiện độ khớp rất cao, với các điểm dữ liệu nằm sát đường $Y = T$. Điều này cho thấy mô hình đã được huấn luyện tốt và có khả năng dự đoán chính xác giá trị trong các tình huống thực tế. Các kết quả này cho thấy mô hình có tiềm năng áp dụng trong các hệ thống dự báo, đặc biệt trong các bài toán phức tạp như dự báo tốc độ gió hoặc năng lượng gió.

Nhận xét 4: Dựa trên kết quả đánh giá mô hình NARX đã được đề xuất, có thể khẳng định rằng việc sử dụng toàn bộ dữ liệu trong một tháng, với mỗi điểm dữ liệu tương ứng với khoảng thời gian 10 phút đo, yêu cầu một cách tiếp cận tỉ mỉ trong việc lựa chọn cấu trúc đầu vào và mục tiêu đầu ra. Cụ thể, để tối ưu hóa quá trình huấn luyện, nên sử dụng đầu vào có độ dài là một ngày và mục tiêu dự đoán có độ dài là một giờ. Lựa chọn này không chỉ cho phép mô hình nắm bắt các xu hướng và biến động trong dữ liệu theo thời gian ngắn mà còn cung cấp một cái nhìn tổng quát hơn về các yếu tố ảnh hưởng đến tốc độ gió trong một chu kỳ 24 giờ. Sự kết hợp giữa đầu vào một ngày và mục tiêu một giờ được cho là phù hợp để cải thiện độ chính xác và độ tin cậy trong dự báo tốc độ gió, đặc biệt trong bối cảnh khí hậu hiện nay đang diễn biến phức tạp. Hơn nữa, lựa chọn này cũng nhấn mạnh tầm quan trọng của việc cân nhắc giữa độ dài dữ liệu và khả năng xử lý của mô hình, nhằm đảm bảo rằng mô hình có đủ thông tin để đưa ra các dự đoán chính xác mà không bị ảnh hưởng bởi tiếng ồn trong dữ liệu. Từ đó, nhận xét này không chỉ đóng góp vào việc cải thiện mô hình NARX mà còn cung cấp những chỉ dẫn quan trọng cho các nghiên cứu và ứng dụng tương lai trong lĩnh vực dự đoán chuỗi thời gian.



Hình 5: Đồ thị sai số của mô hình sau huấn luyện với dữ liệu tại độ cao 80m

4. Đánh giá hiệu quả mô hình NARX

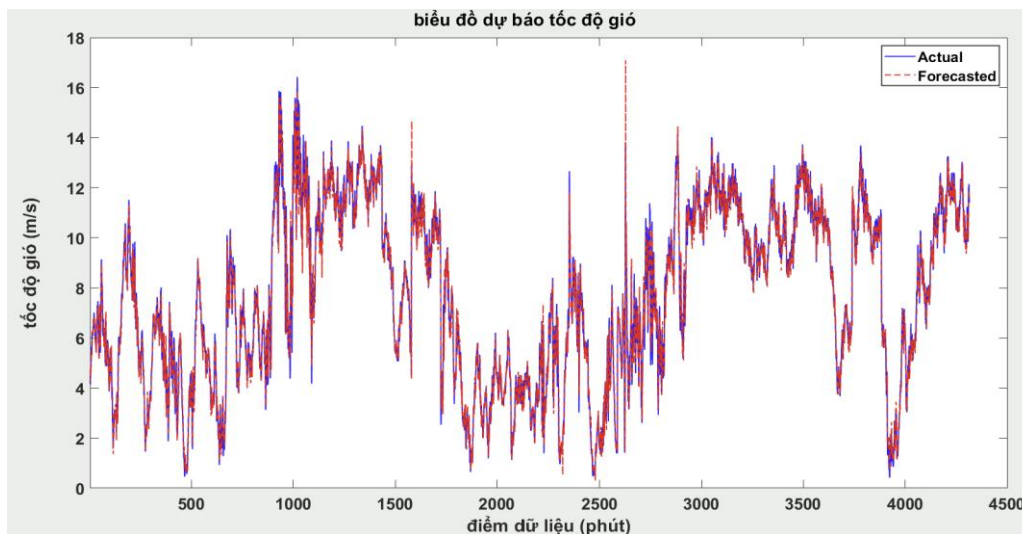
4.1. Đánh giá kết quả dự báo với dữ liệu đầu vào là tốc độ gió tại độ cao 80m

Kết quả 1: Sử dụng một ngày dữ liệu để dự báo sớm một giờ

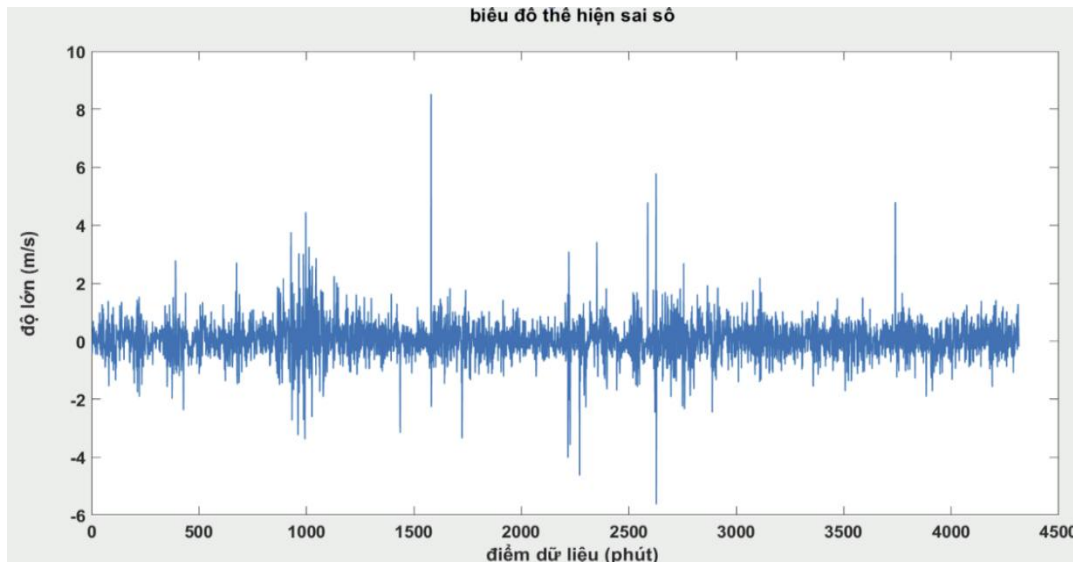
Trong phần này ta sử dụng dữ liệu đo trong tháng 3/2018 để huấn luyện và kiểm tra mô hình dự báo. Đồ thị sai số Hình 6 kết quả tốc độ gió được dự đoán khi sử dụng một ngày dữ liệu để dự đoán cho một giờ. Nhìn vào đồ thị, ta thấy các giá trị sai lệch chủ yếu dao động xung quanh 0, tuy nhiên, có một số điểm xuất hiện sai lệch lớn, đặc biệt ở khoảng 1000 và 2500. Điều này có thể chỉ ra rằng mô hình gặp phải những lỗi đột ngột tại những thời điểm này, hoặc

có thể mô hình không hoạt động hiệu quả tại các điểm dữ liệu cụ thể. Mặc dù vậy, phần lớn các sai số vẫn dao động trong khoảng $(-2; 2)$, cho thấy rằng mô hình duy trì tính ổn định và hoạt động tương đối đều trong hầu hết thời gian. Các kết quả thu nhận được từ MAE bé hơn 0,5 và MAPE bé hơn 8,5, kết quả biểu đồ hồi quy cho thấy mô hình đã được đào tạo tốt, có khả năng dự đoán một cách chính xác và tổng quát Hình 6 thể hiện đồ thị sai số của tốc độ gió dự đoán khi sử dụng một ngày để dự đoán một giờ. Qua đồ thị dự đoán cho ta thấy:

- Đường thực tế nét liền (Actual) biểu diễn tốc độ gió thực tế theo thời gian
- Đường dự đoán nét đứt (Forecasted) biểu diễn giá trị tốc độ gió mà mô hình dự đoán được



Hình 6 Kết quả sử dụng một ngày dữ liệu để dự báo tốc độ gió sớm một giờ



Hình 7 Đồ thị sai số sử dụng một ngày dữ liệu để dự báo trước một giờ

Nhận xét 5: Để đánh giá kết quả mô hình xây dựng có hiệu quả hay không ta dựa trên sự khớp nhau của hai đường thực tế và dự đoán. Ví dụ như ở Hình 7 cho ta thấy mô hình đã dự đoán khá chính xác xu hướng và dao động của tốc độ gió, có một số vị trí không khớp hoàn toàn nhưng nhìn chung mức độ tương đồng là khá tốt. Sai số cục bộ xảy ra khi có sự chênh lệch đáng kể giữa kết quả dự báo và thực tế, nhất là khi dữ liệu có biến động lớn mà mô hình không thể phản ứng kịp. Nguyên nhân có thể do một số nơ-ron trong mạng gặp khó khăn trong cập nhật trọng số, ảnh hưởng đến việc truyền thông tin và giảm hiệu suất dự báo. Hiện tượng quá khớp cục bộ cũng có thể xuất hiện khi mô hình hoạt động tốt trên tập huấn luyện nhưng gặp khó khăn khi dự đoán dữ liệu mới. Ngoài ra, gradient cục bộ có thể biến mất hoặc phát nổ, gây cản trở quá trình cập nhật trọng số. Các biện pháp khắc phục bao gồm: điều chỉnh siêu tham số (hyperparameters), sử dụng kỹ thuật khởi tạo trọng số tốt hơn, tăng cường dữ liệu huấn luyện (Data Augmentation) và chuẩn hóa (Normalization).

Kết quả 2: Sử dụng mô hình được huấn luyện bởi dữ liệu tháng 3 năm 2018 để kiểm tra cho 11 tháng liền kề

Kết quả từ việc sử dụng dữ liệu 11 tháng liền kề để mô hình được huấn luyện bởi dữ liệu tháng 3 năm 2018 cho thấy hiệu quả dự đoán khi kiểm tra với 11 tháng lân cận, được đánh giá dựa trên các chỉ số MAPE và MAE như trình bày trong Bảng 8. Kết quả cho thấy MAPE có sự dao động mạnh giữa các tháng, với giá trị cao nhất là vô cực (inf) vào tháng 5/2018. Điều này cho thấy mô hình có thể đã gặp vấn đề nghiêm trọng trong việc dự đoán trong tháng này và do dữ liệu gió của tháng 5 có những khoảng thời gian lặng gió (tốc độ gió là 0 m/s). Trong khi đó, giá trị MAE vẫn duy trì ở mức thấp và ổn định, với giá trị cao nhất là 0,9407 vào tháng 11/2017.

Nhận xét 6: Mặc dù mô hình có những thời điểm dự đoán không chính xác về phần trăm (MAPE), nhưng sai số tuyệt đối vẫn nằm trong giới hạn chấp nhận được (MAE). Đặc

biệt, các tháng như 8/2017 và 9/2017 ghi nhận giá trị MAPE khá cao, lần lượt là 16,3717 và 12,8647, cho thấy mô hình gặp khó khăn trong việc dự đoán trong thời gian này. Ngược lại, tháng 1/2018 có giá trị MAPE và MAE thấp nhất, lần lượt là 5,9374 và 0,4754, phản ánh độ chính xác cao của mô hình trong tháng đó. Qua đó, có thể khẳng định rằng mặc dù mô hình có những thiếu sót trong một số khoảng thời gian, nhưng nhìn chung, nó vẫn có khả năng hoạt động hiệu quả trong việc dự đoán tốc độ gió, đặc biệt là trong các điều kiện ổn định. Những kết quả này mở ra hướng nghiên cứu tiếp theo để cải thiện độ chính xác của mô hình trong các tình huống khó khăn.

Bảng 8 So sánh MAPE và MAE với 11 tháng liền kề

Thời gian	MAPE	MAE
8/2017	16,3717	0,6451
9/2017	12,8647	0,5912
10/2017	10,8792	0,6783
11/2017	8,4914	0,8407
12/2017	8,0320	0,8165
1/2018	5,9374	0,4754
2/2018	7,7336	0,4962
4/2018	10,1377	0,5378
5/2018	inf	0,4282
6/2018	7,7364	0,6460
7/2018	8,0599	0,7661
8/2018	7,9965	0,7293

4.2. Sử dụng dữ liệu đầu vào kết hợp với các thông số khác tại độ cao 80m

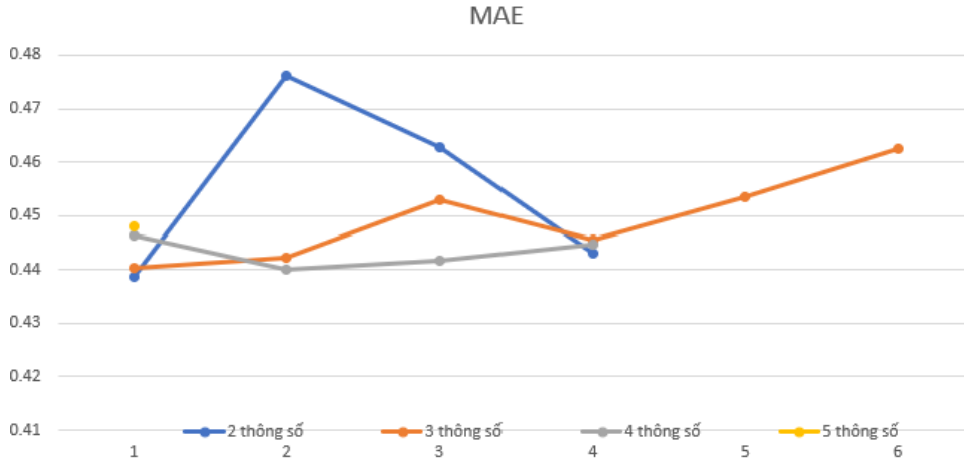
Để cải thiện độ chính xác của mô hình như đã đề cập ở phần trước, chúng ta có thể kết hợp thêm các thông số khác có ảnh hưởng đến tốc độ gió. Trong nghiên cứu này, nghiên cứu đã lựa chọn bổ sung 4 thông số mới vào tập dữ liệu đầu vào, nhằm nâng cao khả năng dự đoán của mô hình. Các thông số được lựa chọn bao gồm hướng gió, độ ẩm, nhiệt độ và áp suất và được đánh số ký hiệu từ 1 đến 5 như trong Bảng 9.

Bảng 9 Các thông số sử dụng dự báo tốc độ gió

1	2	3	4	5
Tốc độ gió	Hướng gió	Độ ẩm	Nhiệt độ	Áp suất

Để đánh giá tác động tổng quát của mô hình và phân tích ảnh hưởng của từng yếu tố đầu vào đối với kết quả đầu ra, một loạt các thử nghiệm đã được thực hiện với tất cả các trường hợp khả thi. Các trường hợp này được trình bày chi tiết trong *Bảng 10*, cho phép nghiên cứu xác định mức độ ảnh hưởng của từng tổ hợp thông số đến dự đoán tốc độ gió.

Mục tiêu của các thử nghiệm này không chỉ là để tối ưu hóa mô hình mà còn nhằm nâng cao độ chính xác của các kết quả dự đoán. Kết quả từ các thử nghiệm này đóng góp vào việc phát triển một mô hình dự báo toàn diện và hiệu quả hơn, giúp cải thiện khả năng dự đoán trong bối cảnh khí hậu biến đổi hiện nay.



Hình 8: Biểu đồ MAE theo các thông số

Việc áp dụng mô hình NARX với khả năng kết hợp các yếu tố hồi quy phi tuyến tính và các biến đầu vào bên ngoài đã chứng minh hiệu quả rõ rệt trong việc xử lý dữ liệu gió phức tạp. *Bảng 10* trình bày kết quả các thông số đánh giá, bao gồm các chỉ số sai số như MAE và MAPE và MSE (Mean Squared Error). Các thông số được thử nghiệm từ 2 đến 5 yếu tố đầu vào, cho thấy sự thay đổi trong hiệu suất của mô hình dự báo.

Dựa trên kết quả và các bước thực hiện được trình bày trong Mục 3, nghiên cứu đã sử dụng 24 giờ dữ liệu để dự đoán cho một giờ tiếp theo. *Bảng 11* đã được tạo ra để tổng hợp các giá trị MAE của các mô hình đã được xây dựng, cung cấp cái nhìn tổng quan về hiệu quả của từng cấu hình mô hình. Những kết quả này không chỉ giúp xác định được các yếu tố đầu vào quan trọng mà còn hỗ trợ trong việc phát triển các chiến lược tối ưu hóa mô hình trong tương lai, từ đó nâng cao khả năng dự đoán tốc độ gió một cách chính xác và đáng tin cậy hơn.

Kết quả được trình bày trong *Bảng 11* cho thấy rằng mô hình thể hiện hiệu suất ấn tượng khi chỉ sử dụng một thông số đầu vào duy nhất, cụ thể là tốc độ gió. Điều này chứng tỏ rằng mô hình có khả năng khai thác triệt để thông tin từ biến tốc độ gió để đưa ra những dự đoán chính xác. Hơn nữa, dữ liệu từ *Bảng 10* cũng củng cố nhận định rằng mô hình này có thể đạt được độ chính xác cao chỉ dựa vào thông số đầu vào tương tự. Sự nhất quán giữa các kết quả trong hai bảng này chỉ ra rằng tốc độ gió là một yếu tố then chốt, có khả năng cung cấp thông tin đầy đủ cần thiết cho mô hình, mặc dù việc tích hợp thêm các yếu tố khác có thể góp phần nâng cao hơn nữa độ chính xác và độ tin cậy của các dự đoán. Từ đó, có thể khẳng định rằng mô hình có tiềm năng tối ưu hóa hiệu suất dự đoán chỉ thông qua việc sử dụng một biến đầu vào, mở ra những hướng nghiên cứu đầy hứa hẹn cho các công trình tiếp theo trong lĩnh vực này.

Bảng 10 Kết quả các thông số đánh giá

Số thông số	Các thông số được kết hợp	MAE	MAPE	MSE
2 thông số	(1) (2)	0.4386	7.9015	0.07
	(1) (3)	0.4761	9.1248	0.0804
	(1) (4)	0.4629	8.7739	0.0788
	(1) (5)	0.443	8.1629	0.0756
	(1) (2) (3)	0.4403	8.1294	0.0706
3 thông số	(1) (2) (4)	0.4424	8.1384	0.0707
	(1) (2) (5)	0.4529	8.3747	0.0744
	(1) (3) (4)	0.4456	8.2001	0.0701
	(1) (3) (5)	0.4535	8.3249	0.0756
4 thông số	(1) (2) (3) (4)	0.4463	8.0012	0.0769
	(1) (2) (3) (5)	0.44	7.9548	0.0643
	(1) (2) (4) (5)	0.4418	7.8944	0.0658
5 thông số	(1) (3) (4) (5)	0.4447	7.9266	0.0664
	(1) (2) (3) (4) (5)	0.4481	7.9092	0.0703

Bảng 11 Tổng hợp MAE theo các thông số

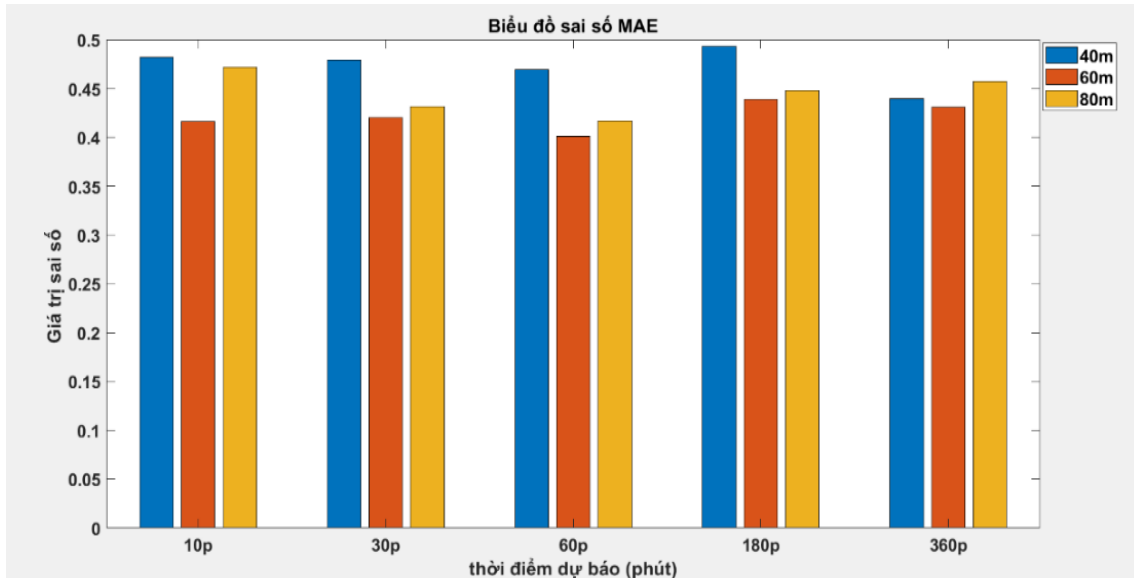
1 thông số	2 thông số	3 thông số	4 thông số	5 thông số
0,4172	0.4386	0.4403	0.4463	0.4481
	0.4761	0.4424	0.44	
	0.4629	0.4529	0.4418	
	0.443	0.4456	0.4447	
		0.4535		
		0.4624		

Nhận xét 7: Nghiên cứu đã chứng minh rằng mô hình NARX không chỉ nâng cao độ chính xác của các dự báo mà còn giảm thiểu sai số so với các mô hình truyền thống. Đáng chú ý, mô hình này hoạt động hiệu quả với chỉ một thông số đầu vào, mà không cần sử dụng thông tin từ quá khứ hay các yếu tố ngoại sinh như áp suất không khí và nhiệt độ. Điều này cho phép mô hình nắm bắt các xu hướng và biến đổi trong dữ liệu gió một cách hiệu quả hơn. Hơn nữa, NARX thể hiện khả năng thích ứng tốt với các điều kiện khí hậu khác nhau, một yếu tố đặc biệt quan trọng trong quy hoạch

năng lượng tái tạo. Sự linh hoạt và độ chính xác của mô hình không chỉ giúp cải thiện chất lượng dự báo mà còn làm cho nó trở thành một công cụ quý giá cho các nhà nghiên cứu và kỹ sư trong lĩnh vực năng lượng gió.

Để đánh giá tính khả thi của một dự án năng lượng gió, việc tối ưu hóa chi phí đầu tư là yếu tố quan trọng, đặc biệt trong bối cảnh xây dựng các tuabin gió ở độ cao lớn, nơi mà chi phí có thể gia tăng đáng kể.

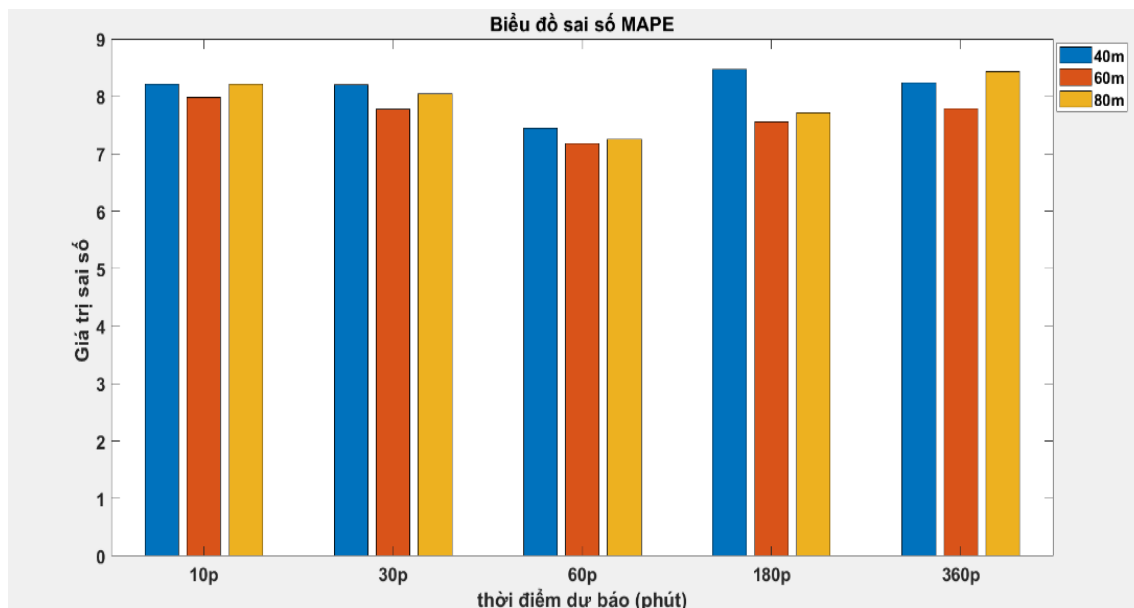
4.3. Đánh giá mô hình NARX cho dữ liệu thu thập ở cả ba độ cao khác nhau 40m, 60m, 80m



Hình 9: Biểu đồ biểu diễn MAE sử dụng dữ liệu 3 độ cao 40m, 60m, 80m

Nghiên cứu này đã áp dụng mô hình dự báo NARX để đánh giá hiệu suất tại ba độ cao khác nhau: 40m, 60m và 80m. Việc lựa chọn mô hình NARX là hợp lý, bởi vì nó có khả năng xử lý các mối quan hệ phi tuyến tính phức tạp giữa các yếu tố đầu vào và đầu ra, điều này rất cần thiết trong bối cảnh khí hậu biến đổi hiện nay. Mô hình này cho phép tích hợp nhiều biến đầu vào, từ đó phản ánh chính xác hơn sự biến đổi của tốc độ gió ở các độ cao khác nhau.

Kết quả thu được từ các thử nghiệm này được trình bày trong Hình 9 và Hình 10, cho thấy mức độ ảnh hưởng của độ cao đến khả năng dự đoán tốc độ gió. Qua đó, nghiên cứu này có thể đưa ra kết luận về hiệu quả của mô hình NARX trong việc tối ưu hóa dự báo năng lượng gió, từ đó hỗ trợ cho các quyết định đầu tư trong lĩnh vực năng lượng tái tạo



Hình 10: Biểu đồ biểu diễn MAPE sử dụng dữ liệu 3 độ cao 40m, 60m, 80m

Nhận xét 8: Các thử nghiệm cũng đã được tiến hành ở độ cao 40m và 60m, trong đó kết quả ở độ cao 60m đã cho thấy độ chính xác cao. Dựa vào dữ liệu quá khứ và kết quả dự báo, việc xây dựng các dự án ở độ cao này có thể được cân nhắc bởi các nhà đầu tư. Qua đó, tầm quan trọng của mô hình đã được chứng minh trong khảo sát đầu tư cho dự án.

5. Mở rộng khoảng thời gian dự báo

Từ kết quả phân tích, đánh giá và thử nghiệm dự báo ở các khoảng thời gian khác nhau, chúng tôi nhận thấy rằng sai số dự báo có xu hướng gia tăng khi khoảng thời gian dự báo kéo dài. Cụ thể, khi dự báo tại cùng một mốc thời gian, việc tăng số điểm dữ liệu đầu vào (inp) có thể giúp giảm sai số. Tuy nhiên, nếu inp tăng quá cao, điều này sẽ dẫn đến việc gia tăng khối lượng tính toán, làm cho mô hình trở nên phức tạp và kéo dài thời gian huấn luyện. Trong nghiên cứu này, chúng tôi đã xác định rằng tỷ lệ giữa inp và out_p nên nằm trong khoảng từ 12 đến 18. Dựa trên kết quả huấn luyện, để có thể dự báo cho một tuần tiếp theo, cần sử dụng dữ liệu từ khoảng 12 đến 18 tuần trước đó. Đồng thời, lượng dữ liệu còn lại phải đủ để mô hình có thể thực hiện quá trình huấn luyện hiệu quả. Nếu giữ nguyên số điểm dữ liệu, với out_p là 1008 và inp khoảng 12096, sẽ tạo ra khối lượng tính toán lớn và không khả thi. Do đó, kịch bản huấn luyện được đề xuất là chia nhỏ 6, 18, và 36 điểm dữ liệu liên tiếp thành một điểm dữ liệu duy nhất, với tỷ lệ inp so với out_p là 12, 15, và 18.

Qua nhiều lần thử nghiệm, chúng tôi đã rút ra được kinh nghiệm quan trọng trong việc tối ưu hóa độ chính xác của mô hình NARX. Việc điều chỉnh thông số độ trễ thời gian (time delay) trong khoảng từ 12 đến 18 bước thời gian đã mang lại những cải thiện đáng kể về hiệu quả và sự ổn định của mô hình. Sự lựa chọn thông số này không chỉ giúp mô hình nắm bắt tốt hơn các mối quan hệ động của dữ liệu mà còn nâng cao khả năng dự đoán chính xác trong các kịch bản khác nhau. Điều này khẳng định rằng việc tối ưu hóa thông số time delay là một yếu tố then chốt trong việc cải thiện hiệu suất của mô hình NARX.

Bảng 12 Tổng hợp MAE theo các thông số NARX dự báo một tuần

Thời gian (phút)	inp	out_p	Delay	MSE	MAE
360	336	28	12	2.188838	0.78203
	402	28	15	2.094092	0.703583
	504	28	18	1.969589	0.597383
180	672	56	12	1.778597	0.750929
	840	56	15	1.355111	0.628446
	1008	56	18	1.168556	0.496876
60	2016	168	12	0.925476	0.544702
	2520	168	15	0.720961	0.491057
	3024	168	18	0.705611	0.492314

Qua quá trình thử nghiệm kéo dài một năm nhằm mục tiêu dự báo cho khoảng thời gian 7 ngày, chúng tôi đã thu thập và phân tích một khối lượng lớn dữ liệu để tối ưu hóa mô hình dự báo. Kết quả nghiên cứu cho thấy rằng việc

lựa chọn thông số kỹ thuật phù hợp là yếu tố then chốt trong việc nâng cao độ chính xác của dự báo.

Trong một số trường hợp trạm đo gặp một số sự cố khiến dữ liệu gửi về bị lỗi, để đảm bảo chuỗi thời gian không bị đứt đoạn theo đúng nguyên tắc của mô hình NARX chúng tôi đã thay thế cách đánh giá mô hình từ MAE và MAPE thành MAE và MSE (Mean Squared Error). Bản chất của sự thay đổi này sẽ không làm thay đổi tính khách quan hay tổng quát của mô hình.

MSE (Mean Squared Error) trong mạng nơ-ron là một chỉ số đo lường độ sai lệch giữa giá trị dự đoán của mô hình và giá trị thực tế. Đây là một trong những hàm mất mát phổ biến, đặc biệt trong các bài toán hồi quy.

Công thức tính MSE:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y_t)^2 \quad (4)$$

Trong đó:

- n là số lượng mẫu.
- y_i là giá trị thực tế của mẫu thứ i.
- y_t là giá trị dự đoán của mẫu thứ i.

MSE có ưu điểm là khuếch đại sai số lớn (vì có bình phương), giúp mô hình quan tâm nhiều hơn đến các dự đoán sai lệch lớn. Tuy nhiên, vì thế nó cũng có nhược điểm là nhạy cảm với các giá trị cực đoan. MSE thường được sử dụng trong quá trình huấn luyện mạng nơ-ron để tối ưu hóa các trọng số của mô hình. Cụ thể, nghiên cứu đã quyết định sử dụng dữ liệu từ 5 năm qua, với tần suất 1 giờ cho mỗi điểm dữ liệu. Điều này không chỉ giúp mô hình có cái nhìn sâu sắc hơn về các xu hướng lịch sử mà còn cung cấp một cơ sở dữ liệu phong phú để cải thiện khả năng dự đoán. Thông số độ trễ (delay) được điều chỉnh thành 1:18, cho phép mô hình nắm bắt và phản ánh tốt hơn những biến động trong dữ liệu theo thời gian. Việc thiết lập các thông số này đã chứng minh hiệu quả trong việc giảm thiểu sai số dự báo, đồng thời tăng cường tính ổn định của mô hình trong các tình huống khác nhau. Nhờ vào những điều chỉnh tối ưu này, mô hình đã thể hiện khả năng dự đoán chính xác hơn đối với các biến đổi ngắn hạn, từ đó mở rộng khả năng ứng dụng của nó trong thực tiễn.

Một nhận xét quan trọng từ nghiên cứu này là sự cần thiết trong việc phải liên tục cập nhật và điều chỉnh mô hình theo các thay đổi trong dữ liệu và điều kiện môi trường. Sự linh hoạt trong thiết kế mô hình sẽ cho phép nó thích ứng tốt hơn với những biến động không lường trước, từ đó nâng cao độ tin cậy của dự đoán. Kết quả nghiên cứu không chỉ khẳng định tính khả thi của phương pháp mà còn cung cấp một khung tham khảo cho các nghiên cứu và ứng dụng tương lai trong lĩnh vực dự báo, đặc biệt là trong các lĩnh vực có tính nhạy cảm cao như dự báo thời tiết và quản lý tài nguyên. Các kết quả của nghiên cứu này sẽ đóng góp tích cực vào việc phát triển các mô hình dự báo hiệu quả hơn, phục vụ cho nhu cầu ngày càng cao trong việc ra quyết định dựa trên dữ liệu.

6. Kết luận

Kết quả nghiên cứu đã chỉ ra rằng việc áp dụng mô hình NARX trong dự báo năng lượng gió mang lại kết quả đáng tin cậy với độ chính xác cao. Đồng thời kết quả cũng chỉ ra nếu dữ liệu dự báo là tốc độ gió thì mô hình đề xuất

chỉ cần sử dụng tốc độ gió quá khứ làm thông số đầu vào để đào tạo mạng và thực hiện các dự đoán, mà không yêu cầu xem xét thêm các yếu tố tương quan tác động khác đến năng lượng của tốc độ gió như nhiệt độ, độ ẩm, hướng gió,... Mô hình nghiên cứu cũng đã thử nghiệm với cho kết quả tốt với các bộ dữ liệu ở các độ cao khác nhau 40m, 60m, 80m. Các kết quả dự báo hoàn toàn có cơ sở góp phần xây dựng cơ sở dữ liệu bản đồ gió tại địa phương, tạo những bước chuyển biến mạnh mẽ trong việc lập kế hoạch và vận hành các trang trại điện gió.

Việc lựa chọn các thông số của mô hình NARX một cách thích hợp không chỉ được ghi nhận đã nâng cao độ chính xác mà còn được đánh giá là đã mở rộng khoảng thời gian dự báo, cho phép mô hình thực hiện các dự đoán hiệu quả hơn trong các khoảng thời gian dài hơn. Kết quả nghiên cứu không chỉ khẳng định tính khả thi của phương pháp mà còn cung cấp một khung tham khảo cho các nghiên cứu và ứng dụng trong tương lai trong lĩnh vực dự báo, đặc biệt là trong các lĩnh vực nhạy cảm như dự báo thời tiết và quản lý tài nguyên. Những phát hiện này được kỳ vọng sẽ đóng góp tích cực vào việc phát triển các mô hình dự báo hiệu quả hơn, đáp ứng nhu cầu ngày càng cao trong việc ra quyết định dựa trên dữ liệu.

Mặc dù đánh giá mô hình cho thấy hiệu suất cao, tuy nhiên vẫn cần cải thiện và tiếp tục nghiên cứu để hoàn thiện như mô hình mới xét đầu ra tốc độ gió, chưa xem xét đến hướng gió đầu ra hay các yếu tố tương quan khác. Các công việc cụ thể dự kiến sẽ nghiên cứu tiếp theo gồm:

- Nội suy dữ liệu cho các vùng lân cận, điều này sẽ giúp tăng cường độ tin cậy và chính xác của dự báo bằng cách sử dụng thông tin từ các trạm khí tượng gần nhất.
- Cải tiến mô hình cũng sẽ được thực hiện nhằm tối ưu hóa khả năng dự báo. Tiếp tục nghiên cứu áp dụng các phương pháp học sâu và tối ưu hóa tham số nhằm cải thiện khả năng dự báo sớm hơn, giúp cho việc lập kế hoạch và vận hành các trang trại điện gió trở nên hiệu quả hơn.

Lời cảm ơn

Bài báo này được tài trợ bởi Bộ Khoa học và Công nghệ, theo nhiệm vụ KH&CN độc lập cấp Quốc gia, mã số: ĐTĐLCN.56/23.

Danh mục tài liệu tham khảo

- [1] P. T. S. Thương, N. T. Trường (2024), *Một số vấn đề về phát triển năng lượng tái tạo ở Việt Nam hiện nay: Thực trạng, tiềm năng và hàm ý giải pháp*, Tạp chí cộng sản, số 3/2024.
- [2] N. C. Dũng (2024), *Việt Nam đẩy mạnh chuyển đổi cơ cấu ngành năng lượng*, Ban Tuyên giáo trung ương, 2024.
- [3] Tập đoàn Điện lực Việt Nam (2019), *Báo cáo kết quả thực hiện kế hoạch sản xuất kinh doanh năm 2018 và 3 năm 2016-2018*, 02/2019. <https://www.evn.com.vn/userfile/User/minhhanh/files/2019/7/Bao%20cao%20SXKD%20nam%20va%203%20nam%20gan%20nha%20t.pdf> [Truy cập 26/02/2025].
- [4] Department of Commerce (2024), *Vietnam Country Commercial Guide: Power Generation, Transmission, and Distribution, 2024*. <https://www.trade.gov/country-commercial-guides/vietnam-power-generation-transmission-and-distribution>. [Truy cập 26/02/2025].
- [5] H. T. Xuân (2021), *Điện gió tại Việt Nam: Nhận diện thách thức và đề xuất giải pháp phát triển*, Tạp chí Công thương Điện tử, 2021.
- [6] H. X. Cơ (2021), *Phát triển điện gió ở Việt Nam: Góc nhìn người trong cuộc*, VIASEE, 2021.
- [7] M. Aitken (2010), *Wind power and community benefits: Challenges and opportunities*, Energy Policy, vol. 38, no. 10, pp. 6066–6075, 2010. <https://doi.org/10.1016/j.enpol.2010.05.062>.
- [8] N. S. Zamir (2023), *Five Critical Challenges of Wind Energy and Potential Solutions*, <https://www.vhive.ai/5-critical-challenges-of-wind-energy-and-potential-solutions/> [Truy cập 26/02/2025].
- [9] Bộ tài nguyên và môi trường (2022), *Báo cáo tiềm năng gió, sóng ngoài khơi tại các vùng biển Việt Nam, 3/2022*. [http://vnmha.gov.vn/upload/files/2022/2-bao-cao-tiem-nang-nang-luong-gio-song-ngoai-khroi-viet-nam-copy-\(2\).pdf](http://vnmha.gov.vn/upload/files/2022/2-bao-cao-tiem-nang-nang-luong-gio-song-ngoai-khroi-viet-nam-copy-(2).pdf) [Truy cập 26/02/2025].
- [10] Vu Dinh Quang, Van Quang Doan, Van Quang Doan, Van Nguyen Dinh, & Nguyen Dinh Duc. (2020). *Evaluation of resource spatial-temporal variation, dataset validity, infrastructures and zones for Vietnam offshore wind energy*. Vietnam Journal of Science, Technology and Engineering, 62(1), 3-16. [https://doi.org/10.31276/VJSTE.62\(1\).03-16](https://doi.org/10.31276/VJSTE.62(1).03-16)
- [11] N. D. Da, M. Herrmann, R. Morrow, F. Niño, N. M. Huan, and N. Q. Trinh (2019), *Contributions of Wind, Ocean Intrinsic Variability, and ENSO to the Interannual Variability of the South Vietnam Upwelling: A Modeling Study*, J. Geophys. Res. Ocean., vol. 124, no. 9, pp. 6545–6574, 2019. <https://doi.org/10.1029/2018JC014647>.
- [12] S. Hanifi, X. Liu, Z. Lin, and S. Lotfian (2020), *A Critical Review of Wind Power Forecasting*, Energies, vol. 13, no. 15, pp. 1–24, 2020.
- [13] C. Skitides and W. G. Früh (2014), *Wind forecasting using Principal Component Analysis*, Renew. Energy, vol. 69, pp. 365–374, 2014. <https://doi.org/10.1016/j.renene.2014.03.068>.
- [14] RasitAta (2015), *Artificial neural networks applications in wind energy systems: a review* Renewable and Sustainable Energy Reviews 49 534–562. <http://dx.doi.org/10.1016/j.rser.2015.04.166>
- [15] Kun Zhou, Sung-Kwun Oh , Witold Pedrycz , Jianlong Qiu, Kisung Seo (2024), *A self-organizing deep network architecture designed based on LSTM network via elitism-driven roulette-wheel selection for time-series forecasting*, Knowledge-Based Systems <https://doi.org/10.1016/j.knosys.2024.111481>
- [16] Muhammad Waqas , Usa Wannasingha Humphries , Bunthid Chueasa , Angkool Wangwongchai, (2024), *Artificial intelligence and numerical weather prediction models: A technical survey*, Natural Hazards Research. <https://doi.org/10.1016/j.nhres.2024.11.004>
- [17] Germánico López, Pablo Arboleya, (2024), *Short-term wind speed forecasting over complex terrain using linear regression models and multivariable LSTM and NARX networks in the Andes*, Renewable Energy <https://doi.org/10.1016/j.renene.2021.10.070>
- [18] Yuzhe Yang , Weiye Song , Shuang Han , Jie Yan, Han Wang, Qiangsheng Dai , Xuesong Huo, Yongqian Liu (2024), *Power forecasting method of ultra-short-term wind power cluster based on the convergence cross mapping algorithm*, Global Energy Interconnection <https://doi.org/10.1016/j.gloi.2024.11.014>
- [19] A. Khosravi, L. Machado, R.O. Nunes, (2018), *Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil*, Applied Energy <https://doi.org/10.1016/j.apenergy.2018.05.043>