

ẢNH HƯỞNG CỦA VIỆC LỰA CHỌN THÔNG SỐ ĐẦU VÀO ĐẾN KẾT QUẢ MÔ PHỎNG NHIỆT ĐỘ KHÔNG KHÍ DÙNG MẠNG NƠON NHÂN TẠO TẠI CÁC TRẠM KHÍ TƯỢNG CẦN THƠ VÀ NHÀ BÈ

TRẦN TRÍ DŨNG

*Viện Khoa học công nghệ và Quản lý môi trường, Trường Đại học Công nghiệp thành phố Hồ Chí Minh;
trantridung@iuh.edu.vn*

Tóm tắt. Nghiên cứu này đánh giá ảnh hưởng của các bộ số liệu đầu vào khác nhau gồm một số yếu tố khí tượng cơ bản đến kết quả mô phỏng nhiệt độ không khí tại 2 trạm khí tượng Cần Thơ và Nhà Bè bằng mạng nơon nhân tạo (ANN). Những cấu trúc ANN với số lượng nơon khác nhau cho giá trị hệ số tương quan (R) tổng quát trong khoảng khá rộng ($0.5156 \div 0.9658$). Kết quả cho thấy mức 6 nơon trong lớp ẩn của cấu trúc ANN (1 lớp ẩn) với hàm chuyển tansig là phù hợp để mô phỏng nhiệt độ không khí tại các trạm trên. Độ chính xác trong mô phỏng đạt mức cao hơn với bộ số liệu đầu vào chứa nhiệt độ không khí đặc trưng ngày đo tại trạm khác và ngày trong năm, qua đó thể hiện sự chi phối khí hậu tâm vĩ mô khá mạnh đến nhiệt độ không khí ở cả 2 trạm. Sự quan trọng của việc lựa chọn các thông số đầu vào được chứng minh ở mức dao động nhỏ hơn trong giá trị R tổng quát từ việc lựa chọn các cấu trúc ANN so với từ việc thay đổi thông số trong bộ số liệu đầu vào.

Từ khóa: nhiệt độ không khí, lựa chọn thông số đầu vào, mạng nơon nhân tạo (ANN), mô phỏng.

EFFECTS OF INPUT PARAMETER SELECTION ON ARTIFICIAL NEURAL NETWORK SIMULATION RESULTS FOR AIR TEMPERATURE AT CAN THO AND NHA BE METEOROLOGICAL STATIONS

Abstract. This study assessed the effects of different input data sets including several basic meteorological factors on air temperature simulation results using Artificial Neural Network (ANN) at Can Tho and Nha Be meteorological stations. ANN structures with different number of neurons provided overall correlation coefficient value (R) in a quite wide range ($0.5156 \div 0.9658$). The results also showed that ANN structure with a single hidden layer of 6 neurons using tansig transfer function is suitable to simulate air temperature at the above-mentioned stations. Simulation accuracy was higher with the input data containing the typical air temperature measured at the same day at another station and the Day of year, thereby revealing strong macroscopic climate influence to air temperature at both stations. The importance of input parameter selection was demonstrated via a smaller variation in overall R value caused by switching ANN structures compared to that from changing parameters of the input data set.

Keywords: air temperature, input parameter selection, Artificial Neural Network (ANN), simulation.

1. MỞ ĐẦU

Nhiệt độ không khí là một yếu tố khí tượng điển hình đặc trưng cho các vùng địa lý, có vai trò đặc biệt quan trọng trong rất nhiều lĩnh vực kinh tế - xã hội như nông - lâm nghiệp hay y tế cộng đồng. Trên thế giới, các nhà khoa học đã nỗ lực đo đạc và điều tra sự biến đổi trong nhiều năm của nhiệt độ không khí để tìm ra bản chất của quá trình hình thành cũng như xác định các phương pháp hiệu quả sao cho có thể ước tính giá trị của nó. Đặc biệt, một số nghiên cứu tập trung vào việc xác định các yếu tố khí tượng quan trọng nhất cần được đưa vào phép tính toán nhiệt độ không khí (như tốc độ gió, độ ẩm tương đối không khí, bức xạ mặt trời) và ảnh hưởng của chúng tới sự hình thành một nhiệt độ nhất định từ việc phân tích số liệu đo thực địa [1, 2].

Mạng nơon nhân tạo (ANN) là một trong những công cụ khoa học tiên tiến thuộc mảng học sâu có ứng dụng rộng rãi trong nhiều ngành nghề để mô phỏng và dự báo. Các công trình về ứng dụng ANN đã được công bố có đối tượng nghiên cứu hết sức đa dạng, trong đó bao gồm nhiều hiện tượng tự nhiên có liên quan đến nhiệt độ không khí như bốc thoát hơi [3, 4, 5, 6, 7] hay trực tiếp mô phỏng - dự báo các yếu tố thời tiết

khác nhau [8, 9, 10]. Trong công trình của mình, Baboo và Shereef đã xác nhận mạng nơron lan truyền ngược với sức mạnh phát hiện các mối quan hệ phi tuyến có tiềm năng cao để dự đoán nhiệt độ không khí dựa trên bộ số liệu khí tượng tổng hợp đầu vào nhờ ưu điểm cải thiện độ hội tụ và giảm dao động [2]. Kaur và Singh đã sử dụng mạng nơron nhiều lớp (MLP) với thuật toán lan truyền ngược để huấn luyện mạng trên số liệu thực đo phục vụ công tác mô phỏng nhiệt độ tối thiểu cho thành phố Chandigarh, Ấn Độ với độ chính xác khá cao [11]. Ustaoglu cùng cộng sự khi tiến hành mô phỏng nhiệt độ không khí cho khu vực Marmara ở Thổ Nhĩ Kỳ đã kết luận rằng một số phương pháp mạng nơron nhân tạo phổ biến (FFBP, RBF, GRNN) đều có tiềm năng mô phỏng - dự báo với giá trị hệ số tương quan giữa kết quả mô phỏng và thực đo tương đối cao ($R > 0.9$) [12].

Kết quả của những công trình được công bố giúp khẳng định các mô hình sử dụng ANN có ưu điểm nổi bật là phát hiện và nắm bắt được mối quan hệ phức tạp giữa nhiều yếu tố tác động lẫn nhau mà đôi khi rất khó có thể đạt được bằng những phương pháp khác. Bên cạnh đó, một số kết quả nghiên cứu cũng đã chứng minh rằng kỹ thuật sử dụng trong mô phỏng dù là một [9, 10] hay kết hợp nhiều [13] phương pháp để hạn chế yếu điểm mỗi cách tiếp cận, cách thức tiền xử lý số liệu đầu vào [14, 15], các thuật toán hay cấu trúc ANN khác nhau [7, 16, 17] đều có thể có ảnh hưởng đáng kể đến độ chính xác của kết quả mô phỏng.

Ở Việt Nam, một số nghiên cứu cũng đã mô tả hiệu quả của kỹ thuật ANN trong công tác điều tra khí tượng - thủy văn phục vụ đời sống và sản xuất. Nguyễn Quang Hoan và cộng sự khi phát triển hệ thống dự báo nhiệt độ tỉnh Hải Dương sử dụng thuật toán lan truyền ngược và thuật toán Bayes [18] hay dự báo thời tiết tỉnh Vĩnh Phúc [19] đều đã thu được kết quả khả quan với tốc độ dự báo nhanh. Trong lĩnh vực thủy văn, Đào Nguyên Khôi và Huỳnh Ái Phương kết luận rằng không những mô hình ANN có khả năng mô phỏng khá chính xác lưu lượng dòng chảy lưu vực sông Sêrêpôk mà việc lựa chọn các thông số đầu vào cho mô hình ANN cũng có thể gây ảnh hưởng đáng kể đến kết quả mô phỏng [20].

Dù đã thu được nhiều thành quả khả quan, việc gia tăng độ chính xác khi ước lượng và mô phỏng thông số bằng ANN cũng vẫn tiếp tục là mối quan tâm đối với các nhà khoa học. Bên cạnh đó, tuy có nhiều đề tài liên quan đến chủ đề mô phỏng số liệu khí tượng bằng ANN, việc đánh giá ảnh hưởng của bộ số liệu đầu vào đến mô phỏng nhiệt độ không khí - có xét đến độ ổn định của kết quả mô phỏng khi số liệu đầu vào được chia ngẫu nhiên - là chủ đề còn ít được đề cập tới. Mục tiêu của nghiên cứu này là đánh giá độ ổn định khi sử dụng mạng nơron nhân tạo (ANN) trong công tác mô phỏng nhiệt độ không khí có sử dụng bộ số liệu thực đo với các yếu tố về mặt bản chất vật lý có khả năng tác động tới nhiệt độ tại các trạm khí tượng Cần Thơ và Nhà Bè đại diện cho 2 khu vực Tây và Đông Nam Bộ nước ta. Ảnh hưởng của các nhóm thông số khí tượng đầu vào (được thể hiện qua sự tuyển chọn có mục đích các bộ dữ liệu đầu vào) cũng như mức độ ảnh hưởng của từng thông số đến độ chính xác của công tác mô phỏng đã được khảo sát. Kết quả của nghiên cứu góp phần làm rõ thêm khả năng ứng dụng ANN trong mô phỏng số liệu khí tượng tại địa bàn Nam Bộ Việt Nam nói riêng và tiềm năng sử dụng các thông số đầu vào khác nhau (kể cả ở trạm đo cách xa) để mô phỏng nhiệt độ không khí ở một trạm khí tượng.

2. KHU VỰC, SỐ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Khu vực nghiên cứu



Hình 1. Vị trí các điểm nghiên cứu: a) Hai trạm khí tượng ở khu vực Nam Bộ; b) Trạm khí tượng Cần Thơ (Nguồn: Google map)

Nằm ở phía Nam nước ta, nhìn chung địa hình Nam Bộ khá bằng phẳng với phía đông và đông nam giáp biển Đông, phía tây giáp Vịnh Thái Lan, phía bắc và tây bắc giáp Campuchia, phía đông bắc giáp với Duyên Hải Nam Trung Bộ và Tây Nguyên. Về mặt khí hậu, khu vực Nam Bộ có khí hậu đặc trưng của miền nhiệt đới gió mùa và cận xích đạo, thời gian bức xạ dài. Nhiệt độ và tổng tích ôn tương đối cao với biên độ nhiệt ngày đêm giữa các tháng thấp. Về cơ bản, khí hậu trong năm có thể được chia thành hai mùa chủ yếu là mùa mưa (từ tháng 5 đến tháng 11) và mùa khô (từ tháng 12 tới tháng 4 năm sau) [21].

Đối tượng nghiên cứu của đề tài này là nhiệt độ không khí tại Trạm khí tượng Cần Thơ và Trạm khí tượng Nhà Bè. Cả 2 trạm đều thuộc lưới trạm điều tra cơ bản của Trung tâm Khí tượng Thủy văn Quốc gia, với mục tiêu quan trắc những yếu tố khí tượng cơ bản và các hiện tượng thời tiết.

Trạm khí tượng Cần Thơ ở vùng Tây Nam Bộ có tọa độ địa lý: 105° 46' 06.48" Kinh độ Đông; 10° 01' 36.85" Vĩ độ Bắc. Trạm khí tượng Nhà Bè ở vùng Đông Nam Bộ có tọa độ địa lý: 106°43'41" Kinh độ Đông; 10°39'36" Vĩ độ Bắc. Khoảng cách theo đường thẳng giữa 2 trạm xấp xỉ 130 km (hình 1).

2.2. Số liệu

Số liệu sử dụng trong bài báo này là kết quả quan trắc cho các yếu tố khí tượng của các trạm khí tượng nêu trên được cung cấp bởi Đài khí tượng Thủy văn khu vực Nam Bộ. Chuỗi số liệu có thời gian 5 năm thuộc giai đoạn 15/1/2013 đến 31/12/2017 với tổng số 1812 ngày đo. Các thiết bị sử dụng để đo đạc thông số như sau:

- Gió: đo bằng cột gió cao 12 m, máy gió Vild.
- Bốc hơi: đo bằng ống đo bốc hơi Piche.
- Nhiệt độ: đo bằng nhiệt kế khô 126906.

Nhằm đảm bảo độ chính xác khi sử dụng công cụ ANN để mô phỏng nhiệt độ không khí đo được tại trạm khí tượng Cần Thơ (hay Nhà Bè), công tác kiểm tra số liệu trước khi bắt đầu mô phỏng đã được thực hiện nhằm loại bỏ những số liệu không hợp lý có thể do nhiều nguyên nhân khách quan và chủ quan khác nhau gây ra (như vấn đề trục trặc về thiết bị đo ngoài thực địa). Thông số được kiểm tra cho các trạm bao gồm: nhiệt độ không khí trung bình ngày và các yếu tố khí tượng cơ bản cùng ngày (độ ẩm tương đối không khí trung bình ngày, tổng lượng bốc hơi ngày, tốc độ gió trung bình ngày).

Các số liệu ban đầu sau khi kiểm tra đã được chuyển đổi về dạng số liệu trong khoảng giá trị từ 0.05 đến 0.95 cho phù hợp với cấu trúc ANN trong Matlab nhằm tăng cường hiệu quả xử lý. Theo [12], công thức chuyển đổi số liệu trước khi đưa vào mô phỏng (tiền xử lý) có dạng như sau:

$$X'_t = \frac{0.9(X_t - X_{\min})}{(X_{\max} - X_{\min})} + 0.05 \quad (1)$$

trong đó: X_t - giá trị thực; X_{\max} - giá trị cực đại của X_t ; X_{\min} - giá trị cực tiểu của X_t ; X'_t - giá trị sau khi chuyển đổi.

Khi đã hoàn thành mô phỏng, các giá trị đã chuyển đổi được tính toán để trả lại những giá trị ban đầu của chúng như sau:

$$X_t = \frac{(X_{\max} - X_{\min})(X'_t - 0.05)}{0.9} + X_{\min} \quad (2)$$

2.3. Phương pháp

Mạng nơron nhân tạo (ANN) nhiều lớp dạng truyền thẳng - lan truyền ngược (feedforward - backpropagation) - một kiểu mạng hoạt động hiệu quả trong Neural Network Toolbox của phần mềm Matlab - theo những cấu trúc điển hình được sử dụng để mô phỏng nhiệt độ không khí trạm Cần Thơ (hay Nhà Bè). Số liệu đầu vào là các bộ thông số khí tượng cơ bản và Ngày trong năm (đặc trưng cho chu trình mùa của biến đổi thời tiết trong năm). Trước khi huấn luyện mạng nơron nhân tạo, số liệu đầu vào cho mô phỏng đã được chuyển đổi về dạng số liệu trong khoảng giá trị từ 0.05 đến 0.95 theo như công thức (1) đã mô tả ở trên.

Sau khi được huấn luyện, ANN nhiều lớp được dùng để chạy mô phỏng với nhiều yếu tố đầu vào và một yếu tố đầu ra. Mỗi cấu trúc ANN được đặc trưng bởi số lượng lớp ẩn, số lượng nơron trong mỗi lớp ẩn và đặc điểm hàm chuyển. Trong nghiên cứu này, tất cả các lớp ẩn đều được bố trí số lượng nơron như nhau (nếu cấu trúc ANN có hơn 1 lớp ẩn) và sử dụng hàm chuyển "tansig"; riêng lớp ra sử dụng hàm chuyển mặc định của Matlab là "purelin". Độ chính xác trong mô phỏng hiện tượng của các cấu trúc ANN khác

nhau được đánh giá bằng hệ số tương quan (R), sai số tuyệt đối trung bình (MAE) có tham khảo căn bậc hai của trung bình bình phương sai số (RMSE) giữa kết quả mô phỏng và giá trị thực đo.

$$R = \frac{\sum_{i=1}^k (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^k (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^k (\hat{y}_i - \bar{\hat{y}})^2}} \quad (3)$$

$$MAE = \frac{1}{k} \sum_{i=1}^k |y_i - \hat{y}_i| \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^k (y_i - \hat{y}_i)^2}{k}} \quad (5)$$

trong đó: y_i - giá trị thực đo (°C); \hat{y}_i - giá trị mô phỏng (°C); \bar{y} - trung bình các giá trị thực đo (°C); $\bar{\hat{y}}$ - trung bình các giá trị mô phỏng (°C); k - tổng số lượng các giá trị; i - số thứ tự giá trị.

Cấu trúc mạng nơon nhân tạo (ANN) phù hợp sẽ phải đồng thời thỏa mãn được 2 yêu cầu cơ bản: nhỏ gọn (số lượng lớp ẩn, số nơon trong lớp ẩn) nhưng vẫn có hiệu quả mô phỏng đủ tốt tức là cho giá trị R cao và MAE thấp.

Bên cạnh việc nghiên cứu giá trị lớn nhất của R cho mỗi cấu trúc ANN với kích bản bộ số liệu cho trước, việc đánh giá tính ổn định (xét mức dao động của giá trị R) trong kết quả mô phỏng khi chạy mô phỏng nhiều lần với sự phân chia số liệu ngẫu nhiên cũng được khảo sát thông qua kiểm định sự khác biệt của giá trị trung bình trong chuỗi R thu được trên phần mềm SPSS. Đánh giá giả thiết phương sai mẫu bằng nhau được tiến hành bằng các kiểm định Levene, Welch và Brown-Forsythe; phân tích Post Hoc sử dụng kiểm định Games-Howell trong trường hợp phương sai mẫu không bằng nhau [22]. Để hỗ trợ nhận định kết quả phân tích thống kê trong trường hợp số lượng mẫu khá lớn (n = 1000), mức độ ảnh hưởng (Effect size) cũng được đánh giá và xếp hạng theo thang giá trị như Cohen (1988) và Parllant (2016) đề xuất.

$$\eta^2 = \frac{SSB}{SST} \quad (6)$$

trong đó: η^2 - hệ số đánh giá mức độ ảnh hưởng (không đơn vị); SSB - tổng bình phương phản ánh độ khác nhau giữa các nhóm; SST - tổng bình phương độ khác nhau cho toàn bộ mẫu.

Mức độ ảnh hưởng của các thông số khí tượng đầu vào đến kết quả mô phỏng nhiệt độ không khí được ước lượng bằng phương pháp rút mỗi thông số ra khỏi mô hình và chạy lại quá trình huấn luyện - mô phỏng sử dụng chính cấu trúc ANN đã chọn đó. Kết quả tính toán được tiếp tục xử lý theo công thức sau:

$$A_j = \frac{MAE_j - MAE_{bd}}{MAE_{bd}} \times 100\% \quad (7)$$

trong đó: A_j - hệ số đánh giá mức độ ảnh hưởng của yếu tố thứ j (%); MAE_j - sai số tuyệt đối trung bình của kết quả mô phỏng bằng cấu trúc ANN đang xét sau khi rút một yếu tố thứ j ra khỏi dữ liệu đầu vào (°C); MAE_{bd} - sai số tuyệt đối trung bình của kết quả mô phỏng bằng cấu trúc ANN đang xét khi còn đầy đủ các yếu tố đầu vào (°C).

Theo công thức trên, giá trị của hệ số đánh giá độ ảnh hưởng A_j càng cao thì yếu tố khí tượng đầu vào thứ j càng có ảnh hưởng mạnh đến độ chính xác (làm tăng sai số tuyệt đối trung bình) trong kết quả mô phỏng nhiệt độ không khí bằng công cụ ANN.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Đặc điểm số liệu đưa vào mô phỏng

Sau khi kiểm tra chất lượng số liệu, tổng số các số liệu (ngày) đưa vào tính toán cho giai đoạn 15/1/2013 đến 31/12/2017 là 1812 và được thống kê trong bảng 1.

Bảng 1. Thống kê các yếu tố khí tượng đo được tại trạm khí tượng Cần Thơ và Nhà Bè

Số thứ tự	Tên yếu tố	Đơn vị đo	Trạm khí tượng	Giá trị lớn nhất	Giá trị nhỏ nhất	Giá trị trung bình	Độ lệch chuẩn
1	Độ ẩm tương đối không khí trung bình ngày	%	Cần Thơ	96.25	61.25	80.76	5.91
			Nhà Bè	96.50	58.75	80.16	6.27
2	Tổng lượng bốc hơi ngày	mm	Cần Thơ	5.80	0.60	2.73	0.87
			Nhà Bè	7.60	0.30	2.97	1.15
3	Tốc độ gió trung bình ngày	m/s	Cần Thơ	4.00	0.00	1.42	0.54
			Nhà Bè	4.50	0.00	1.22	0.65
4	Nhiệt độ không khí trung bình ngày	°C	Cần Thơ	31.15	21.90	27.62	1.41
			Nhà Bè	31.78	22.53	27.78	1.45

3.2. Khảo sát hoạt động của các cấu trúc ANN dưới tác động của các bộ số liệu đầu vào khác nhau

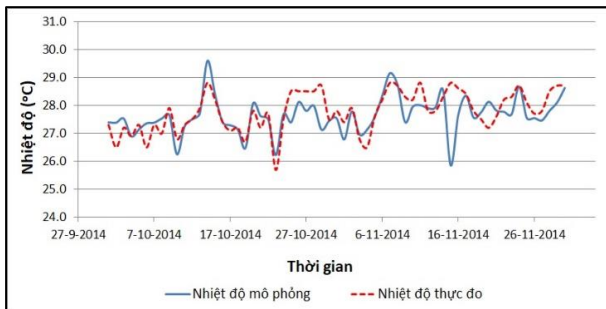
Trong nghiên cứu này, loại bộ số liệu dùng cho các tính toán mô phỏng nhiệt độ không khí tại một trạm khí tượng (Cần Thơ hoặc Nhà Bè) sẽ được lặp lại tương tự (kiểu đối xứng) cho trạm còn lại. Mục đích để khảo sát kết quả thu được có ổn định trong không gian hay không khi các điểm khảo sát nằm xa nhau. Hoạt động của công cụ ANN trong Matlab đã được khảo sát để đánh giá kết quả mô phỏng cho nhiệt độ không khí trung bình ngày tại mỗi trạm khí tượng. Tính toán được tiến hành trên tổng số 1812 mẫu số liệu khí tượng thực đo ngày, các bộ số liệu đầu vào cho mạng ANN chia thành 4 loại bộ khác nhau như mô tả trong bảng 2.

Bảng 2. Các loại bộ số liệu đưa vào mô phỏng

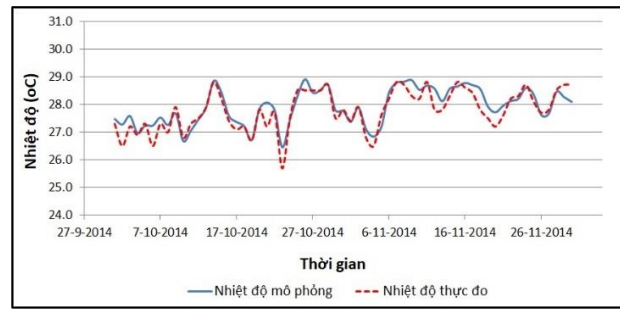
Loại	Yếu tố khí tượng	Ý nghĩa thể hiện
I	Các yếu tố khí tượng đo cùng ngày tại cùng trạm khí tượng đang khảo sát gồm độ ẩm tương đối không khí trung bình ngày, tổng lượng bốc hơi ngày, tốc độ gió trung bình ngày	Ảnh hưởng song song từ các yếu tố khí tượng khác đến nhiệt độ không khí cùng ngày ở cùng địa điểm
II	Nhiệt độ không khí trung bình ngày trong 1, 2 và 3 ngày trước tại cùng trạm khí tượng đang khảo sát	Ảnh hưởng kế thừa từ nhiệt độ không khí trung bình ngày từ vài ngày trước đến nhiệt độ trung bình ngày ở cùng địa điểm
III	Nhiệt độ không khí trung bình ngày tại trạm khí tượng khác (tức trạm còn lại) nằm cách xa với trạm đang khảo sát và Ngày trong năm	- Mỗi quan hệ mang tính vĩ mô của khu vực chi phối nhiệt độ không khí trung bình ngày ở nhiều địa điểm khác biệt về mặt không gian - Thông số Ngày trong năm thể hiện tính chu kỳ của biến đổi thời tiết trong năm chi phối đến nhiệt độ
IV	Dạng tổng hợp gồm cả 3 loại bộ số liệu (I, II và III) trên	Ảnh hưởng của cả 3 loại trên lên nhiệt độ không khí ngày tại trạm đang khảo sát

Trong nghiên cứu này, mỗi bộ số liệu đầu vào đã được phần mềm Matlab chia ngẫu nhiên theo tỷ lệ 70% số lượng mẫu cho việc huấn luyện mạng, 15% cho việc xác nhận tổng hợp mạng và tránh hiện tượng quá ngưỡng, 15% cho việc kiểm tra độc lập. Cấu trúc ANN nhiều lớp sử dụng thuật toán huấn luyện lan truyền ngược Levenberg-Marquardt (trainlm), hàm huấn luyện weight/bias là “learnsgdm” và được đánh giá hiệu quả bởi hàm MSE (mặc định của Matlab). Đối với mỗi kiểu cấu trúc ANN, tất cả các lớp ẩn đều được bố trí số lượng nơon như nhau (nếu có hơn 1 lớp ẩn) và sử dụng hàm chuyển “tansig”; riêng lớp ra luôn sử dụng hàm chuyển mặc định của Matlab là “purelin”.

Các cấu trúc ANN một hoặc hai lớp ẩn với số lượng nơon khác nhau (2, 4, 6, 8, 10 và 12) trong mỗi lớp ẩn đã được đánh giá để xác định hiệu suất mô phỏng. Mỗi cấu trúc ANN được sử dụng để chạy mô phỏng 1000 lần và giá trị cao nhất của R thu được từ nhiều lần chạy một cấu trúc sẽ là tiêu chuẩn đánh giá cho hiệu suất mô phỏng của cấu trúc đó đối với bộ số liệu đầu vào đang khảo sát. Kết quả mô phỏng được trình bày trong các hình 2, 3, 4 và 5.

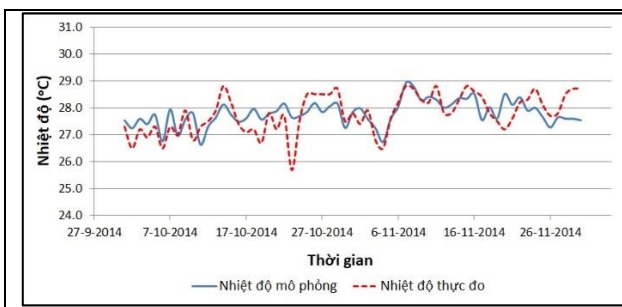


a) Chỉ dùng số liệu khí tượng cùng ngày trạm Cần Thơ

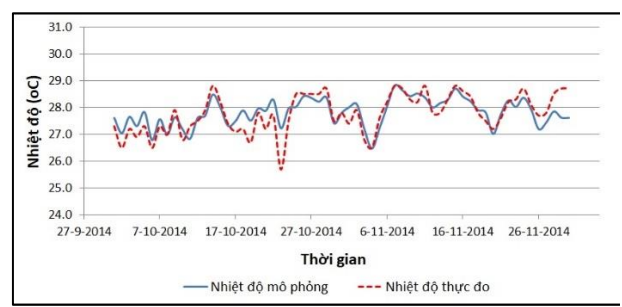


b) Dùng bộ số liệu hỗn hợp 8 yếu tố

Hình 2. So sánh kết quả mô phỏng nhiệt độ không khí ở trạm Cần Thơ với cấu trúc ANN có 1 lớp ẩn gồm 6 nơron cho giai đoạn 1/10/2014 - 30/11/2014 khi dùng các bộ số liệu đầu vào khác nhau.



a) Chỉ dùng số liệu khí tượng cùng ngày trạm Nhà Bè



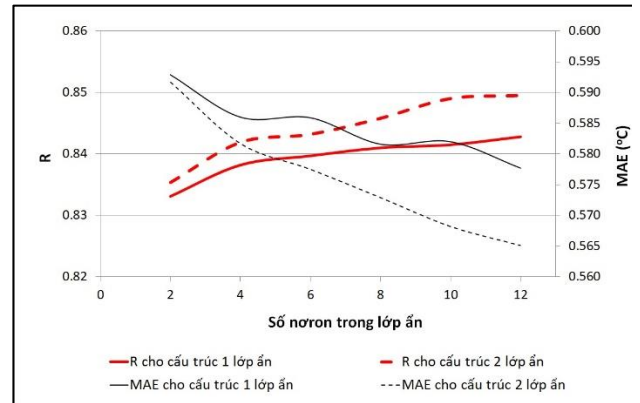
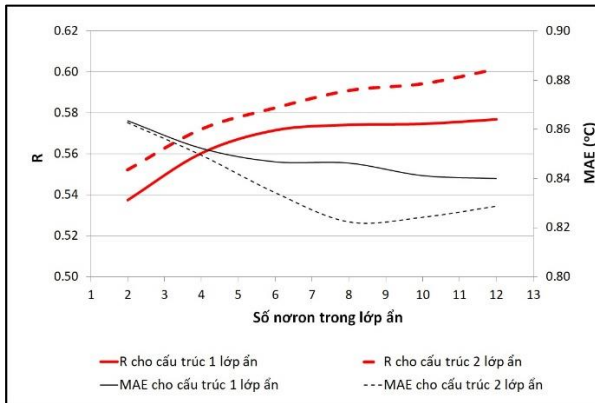
b) Dùng bộ số liệu hỗn hợp 8 yếu tố

Hình 3. So sánh kết quả mô phỏng nhiệt độ không khí ở trạm Nhà Bè với cấu trúc ANN có 1 lớp ẩn gồm 6 nơron cho giai đoạn 1/10/2014 - 30/11/2014 khi dùng các bộ số liệu đầu vào khác nhau.

Dưới ảnh hưởng của các loại bộ số liệu đầu vào, nhìn chung bộ số liệu sử dụng nhiệt độ không khí trung bình ngày tại trạm khí tượng khác và ngày trong năm (loại III) cho kết quả mô phỏng tốt hơn (R cao, MAE thấp) đối với cùng số lớp ẩn và số nơron trong mỗi lớp ẩn so với các bộ số liệu loại I và II đối với cả 2 trạm khảo sát. Khoảng giá trị R thu được cũng không có sự khác biệt nhiều ở 2 trạm cho mỗi loại bộ số liệu đầu vào. Không ngoài dự kiến, bộ số liệu tổng hợp (loại IV) với đầy đủ 8 thông số đầu vào đã cho kết quả mô phỏng nhiệt độ không khí tại mỗi trạm tốt nhất tức kết quả mô phỏng gần với số liệu thực đo nhất.

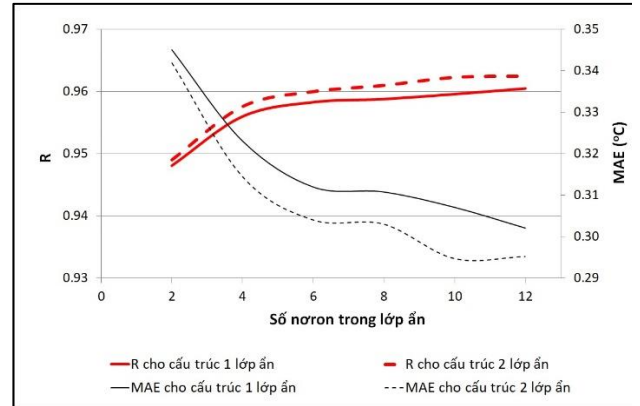
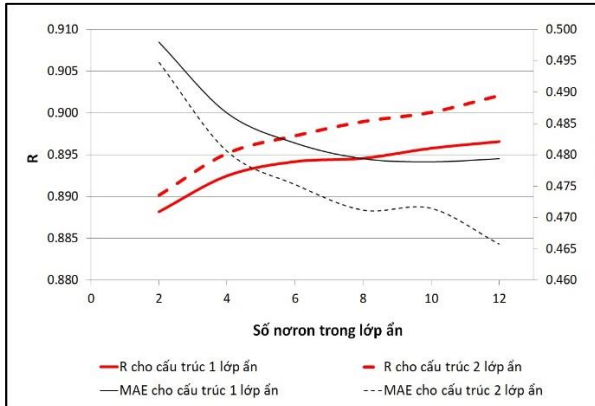
Kết quả mô phỏng thu được cho thấy giá trị hệ số tương quan (R) tổng quát lớn nhất cho quá trình mô phỏng với hàm tansig dao động trong khoảng $0.5375 \div 0.9625$ cho trạm Cần Thơ và $0.5156 \div 0.9658$ cho trạm Nhà Bè; trong khi giá trị sai số tuyệt đối trung bình (MAE) dao động trong khoảng $0.2948 \div 0.8635$ cho trạm Cần Thơ và $0.2860 \div 0.9109$ cho trạm Nhà Bè. Đối với mỗi bộ số liệu đầu vào, kết quả mô phỏng cũng đã thể hiện khá rõ xu hướng tăng của giá trị hệ số tương quan và xu hướng giảm của giá trị sai số tuyệt đối trung bình khi số lượng lớp ẩn và/hoặc số lượng nơron trong mỗi lớp ẩn tăng lên tức là chất lượng mô phỏng được cải thiện tốt hơn (đạt độ chính xác cao hơn) khi cấu trúc ANN phức tạp hơn. Tuy vậy, độ dốc các đường cong trên đồ thị cho thấy mức độ biến đổi của các giá trị thống kê trên lại không đều và cường độ thay đổi của chúng thường cũng giảm dần theo chiều tăng của số lượng lớp ẩn và/hoặc số lượng nơron trong mỗi lớp ẩn cho cả 2 trạm khảo sát.

176 ẢNH HƯỞNG CỦA VIỆC LỰA CHỌN THÔNG SỐ ĐẦU VÀO ĐẾN KẾT QUẢ MÔ PHỎNG NHIỆT ĐỘ KHÔNG KHÍ DÙNG MẠNG NƠON NHÂN TẠO TẠI CÁC TRẠM KHÍ TƯỢNG CẦN THƠ VÀ NHÀ BÈ



a) Số liệu khí tượng đầu vào là thông số khí tượng cùng ngày ở trạm Cần Thơ

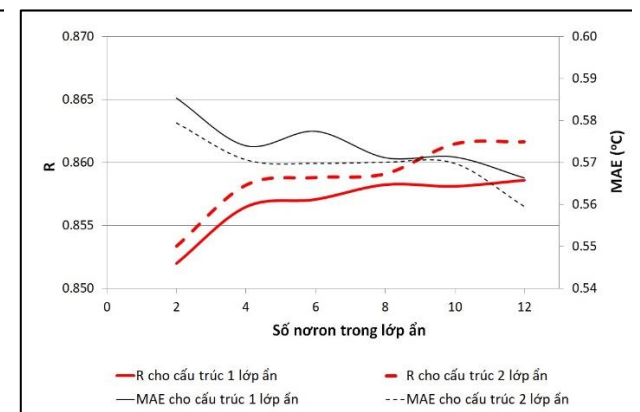
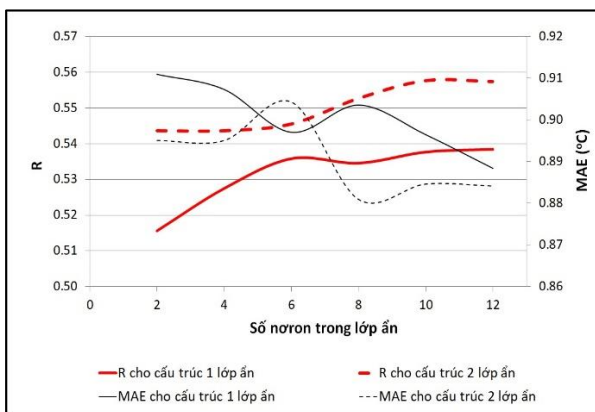
b) Số liệu khí tượng đầu vào là nhiệt độ các ngày trước ở trạm Cần Thơ



c) Số liệu khí tượng đầu vào là nhiệt độ cùng ngày ở trạm Nhà Bè

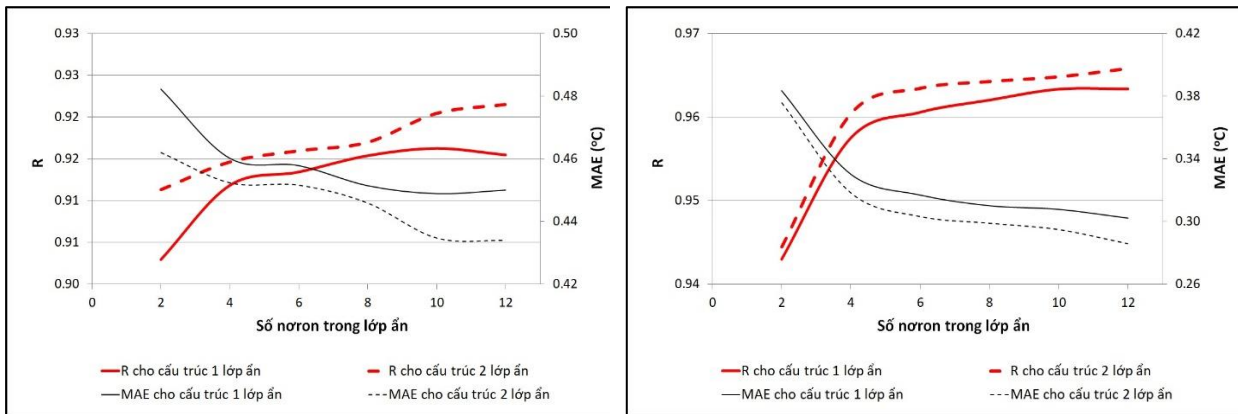
d) Số liệu khí tượng đầu vào gồm cả 3 loại trên

Hình 4. Giá trị hệ số tương quan (R) lớn nhất và sai số tuyệt đối trung bình (MAE) khi sử dụng mạng ANN để mô phỏng nhiệt độ không khí trạm Cần Thơ với các loại bộ số liệu khác nhau.



a) Số liệu khí tượng đầu vào là thông số khí tượng cùng ngày ở trạm Nhà Bè

b) Số liệu khí tượng đầu vào là nhiệt độ các ngày trước ở trạm Nhà Bè



c) Số liệu khí tượng đầu vào là nhiệt độ cùng ngày ở trạm Cần Thơ d) Số liệu khí tượng đầu vào gồm cả 3 loại trên

Hình 5. Giá trị hệ số tương quan (R) lớn nhất và sai số tuyệt đối trung bình (MAE) khi sử dụng mạng ANN để mô phỏng nhiệt độ không khí trạm Nhà Bè với các loại bộ số liệu khác nhau.

Bên cạnh công tác khảo sát giá trị lớn nhất của hệ số R, việc đánh giá độ ổn định trong kết quả mô phỏng khi bộ số liệu được chia ngẫu nhiên cho 3 mục đích (huấn luyện, tổng hợp, kiểm tra) đối với cấu trúc ANN cũng hết sức quan trọng. Từ 1000 lần chạy mô phỏng với bộ số liệu được chia ngẫu nhiên, các kịch bản tổ hợp của cấu trúc ANN (số lớp ẩn, số nơon trong lớp ẩn) và bộ số liệu đầu vào thay đổi đã cho ra những chuỗi giá trị R trong mô phỏng khác nhau. Sự khác biệt của giá trị trung bình trong chuỗi R đã được đánh giá trên phần mềm SPSS để ước tính mức ổn định và hiệu quả mô phỏng của cấu trúc ANN khi bộ số liệu đầu vào thay đổi. Do phương sai mẫu không bằng nhau theo kết quả từ các kiểm định Levene, Welch và Brown-Forsythe ($p < 0.05$), phân tích Post Hoc với độ tin cậy 95% đã được tiến hành với kiểm định Games-Howell. Ngoài ra, mức độ ảnh hưởng được phân loại theo Cohen cũng được trình bày kèm theo trong bảng 3.

Bảng 3. Kết quả đánh giá ảnh hưởng bởi số nơon trong mỗi lớp ẩn đến hiệu quả mô phỏng (R) của các cấu trúc ANN

Trạm khí tượng	Bộ số liệu đầu vào	Số lớp ẩn	Kiểm định phương sai bằng nhau (p)			Kết quả Post Hoc (Games-Howell test) với độ tin cậy 95% cho R trung bình của các cặp số lượng nơon	Mức độ ảnh hưởng (Effect size)	
			Levene	Welch	Brown-Forsythe		Cohen η^2	Đánh giá
Cần Thơ	Số liệu khí tượng cùng ngày ở trạm	1	<0.0005	<0.0005	<0.0005	$p_{10-12} = 0.096$	0.21	Lớn
		2	<0.0005	<0.0005	<0.0005	$p_{8-10} = 0.872$	0.22	Lớn
	Số liệu nhiệt độ các ngày trước ở trạm	1	0.002	<0.0005	<0.0005	$p_{8-10} = 0.999$; $p_{8-12} = 0.059$; $p_{10-12} = 0.111$	0.26	Lớn
		2	<0.0005	<0.0005	<0.0005	$p_{6-8} = 0.149$; $p_{8-10} = 0.382$	0.05	Trung bình - nhỏ
	Số liệu nhiệt độ cùng ngày ở trạm Nhà Bè	1	<0.0005	<0.0005	<0.0005		0.57	Lớn
		2	<0.0005	<0.0005	<0.0005		0.06	Trung bình

	Cả 3 loại số liệu đầu vào	1	<0.000	<0.000	<0.0005		0.31	Lớn
		2	<0.000	<0.000	<0.0005	$p_{6-8} = 0.511$	0.06	Trung
Nhà Bè	Số liệu khí tượng cùng ngày ở trạm	1	<0.000 5	<0.000 5	<0.0005	$P_{4-6} = 0.330; P_{6-8} = 0.556; p_{10-12} =$	0.11	Trung bình - lớn
		2	<0.000 5	<0.000 5	<0.0005	$p_{6-8} = 0.996; P_{8-10} = 0.128; p_{10-12} =$	0.17	Lớn
	Số liệu nhiệt độ các ngày trước ở trạm	1	0.006	<0.000	<0.0005	$p_{8-10} = 0.413$	0.31	Lớn
		2	<0.000 5	<0.000 5	<0.0005	$P_{4-6} = 0.352; P_{6-8} = 0.450; p_{8-10} = 0.278; p_{10-12} =$	0.04	Trung bình - nhỏ
	Số liệu nhiệt độ cùng ngày ở trạm Cần Thơ	1	<0.000 5	<0.000 5	<0.0005		0.10	Trung bình - lớn
		2	<0.000	<0.000	<0.0005		0.06	Trung
	Cả 3 loại số liệu đầu vào trên	1	<0.000	<0.000	<0.0005		0.77	Lớn
		2	<0.000 5	<0.000 5	<0.0005	$P_{6-8} = 0.823; p_{8-10} = 0.888; p_{10-12} =$	0.08	Trung bình - lớn

*Chú thích: p_{a-b} thể hiện thống kê khi so sánh R trung bình giữa hai cấu trúc ANN với số lượng a và b nơron trong lớp ẩn; cột kết quả post hoc chỉ thể hiện các mức $p > 0.05$ (sự khác biệt không có ý nghĩa thống kê), các kết quả so sánh khác biệt trung bình R còn lại đều có ý nghĩa thống kê.

Kết quả thu được cho thấy trong gần như tất cả các trường hợp đang khảo sát, sự khác biệt về giá trị trung bình của chuỗi R khi số nơron trong lớp ẩn của ANN tăng có ý nghĩa thống kê ở các cấp nhỏ hơn hoặc bằng 4 nơron (trừ 2 trường hợp). Khi số nơron trong lớp ẩn từ 6 trở lên (đặc biệt khi lớn hơn 8) thì trong nhiều trường hợp mức tăng giá trị trung bình R trong mô phỏng không còn có ý nghĩa thống kê. Kết hợp với các hình 4 và 5, điều này cho thấy khi tăng số nơron từ mức 8 trở lên thì nhiều khả năng cấu trúc ANN không thể chiết xuất thêm hay xử lý hữu hiệu hơn những thông tin có trong bộ số liệu đầu vào trong quá trình mô phỏng nhiệt độ không khí. Nói cách khác, mức 6 nơron trong lớp ẩn ANN (1 lớp ẩn) thường là đủ trong mô phỏng nhiệt độ không khí tại các trạm Cần Thơ và Nhà Bè.

Một điểm quan trọng rút ra từ kết quả phân tích này là dấu có sự khác biệt có ý nghĩa thống kê về giá trị R khi số nơron tăng cho cùng bộ số liệu đầu vào, sự khác biệt này lại nhỏ hơn nhiều so với sự thay đổi của R khi thay đổi bộ số liệu đầu vào (giữa những bộ số liệu loại I ÷ IV trên). Điều này cho thấy tầm quan trọng đặc biệt của việc chọn thông số đầu vào khi mô phỏng nhiệt độ không khí.

3.3. Khảo sát mức độ ảnh hưởng của các yếu tố đầu vào đến độ chính xác trong mô phỏng

Nói tiếp bước trên, cấu trúc ANN với 1 lớp ẩn và 6 nơron trong lớp ẩn (8-6-1) được sử dụng để mô phỏng nhiệt độ không khí từ bộ số liệu đầu vào tổng hợp (loại IV) bởi cấu trúc này cho kết quả mô phỏng đủ tốt (trạm Cần Thơ: $R = 0.9583$ hay $R^2 = 0.9183$, MAE = 0.3120 với giá trị hệ số tương quan cho giai đoạn kiểm tra $R_{test} = 0.9566$; trạm Nhà Bè: $R = 0.9606$ hay $R^2 = 0.9228$, MAE = 0.3170 với giá trị hệ số tương quan cho giai đoạn kiểm tra $R_{test} = 0.9573$). Mặt khác, khi tăng số nơron trong lớp ẩn lên hơn nữa thì mức tăng trong hiệu quả mô phỏng (tức R tăng, MAE giảm) không đáng kể về mặt ý nghĩa thực tế (dù rằng có thể vẫn có ý nghĩa thống kê cho R tăng ở cấp sau đó). Vì vậy, cấu trúc này đã được đánh giá là phù hợp để sử dụng cho đánh giá mức độ ảnh hưởng của các yếu tố khí tượng đầu vào đến kết quả mô phỏng nhiệt độ không khí ở mỗi trạm (phân tích độ nhạy) bằng phương pháp lần lượt rút từng thông số đầu vào khỏi bộ số liệu.

Kết quả đánh giá mức độ ảnh hưởng của mỗi thông số khí tượng đầu vào đến độ chính xác của kết quả mô phỏng bằng phương pháp loại yếu tố đó ra khỏi bộ số liệu đầu vào được trình bày trong bảng 4. Căn cứ vào độ lớn của hệ số đánh giá độ ảnh hưởng A_j , các thông số khí tượng trong vai trò số liệu đầu vào đã được sắp xếp theo mức độ ảnh hưởng của chúng đến độ chính xác của kết quả mô phỏng nhiệt độ không khí. Ở bảng này, giá trị trong cột “Mức ảnh hưởng” cho một yếu tố đầu vào càng nhỏ thì việc loại bỏ yếu tố đó ra khỏi quá trình mô phỏng sẽ càng làm giảm (tức giảm R) mạnh hơn đến độ chính xác của kết quả mô phỏng so với giá trị thực đo.

Bảng 4. Mức độ ảnh hưởng của các yếu tố khí tượng đến độ chính xác của kết quả mô phỏng nhiệt độ không khí tại trạm Cần Thơ theo MAE

<i>Trạm khí tượng</i>	<i>Tên yếu tố</i>	$A_j(\%)$	<i>Mức ảnh hưởng</i>
Cần Thơ	Nhiệt độ không khí đặc trưng ngày trạm Nhà Bè	24.74	1
	Độ ẩm không khí đặc trưng ngày	14.78	2
	Nhiệt độ không khí đặc trưng 1 ngày trước	6.94	3
	Ngày trong năm	4.24	4
	Lượng bốc hơi ngày, tốc độ gió trung bình ngày, nhiệt độ không khí đặc trưng 2 và 3 ngày trước	< 0.5	Rất nhỏ
Nhà Bè	Nhiệt độ không khí đặc trưng ngày trạm Cần Thơ	24.18	1
	Độ ẩm không khí đặc trưng ngày	18.93	2
	Nhiệt độ không khí đặc trưng 1 ngày trước	4.56	4
	Ngày trong năm	11.49	3
	Lượng bốc hơi ngày, tốc độ gió trung bình ngày, nhiệt độ không khí đặc trưng 2 và 3 ngày trước	< 2.0	Rất nhỏ

Kết quả đánh giá mức độ ảnh hưởng của từng yếu tố khí tượng đầu vào đã cho thấy nhiệt độ không khí đặc trưng ngày trạm khác và độ ẩm không khí đặc trưng ngày tại trạm nghiên cứu là các yếu tố có mức độ ảnh hưởng lớn nhất đến sai số MAE trong kết quả mô phỏng nhiệt độ không khí tại mỗi trạm khí tượng (khi loại bỏ làm tăng giá trị MAE ở mức khoảng 24 % và 14 ÷ 18 %, tương ứng). Các yếu tố nhiệt độ không khí đặc trưng 1 ngày trước và ngày trong năm đều có mức ảnh hưởng nhất định nhưng lại không ổn định giữa 2 trạm nghiên cứu. Trong khi đó, những yếu tố như lượng bốc hơi ngày, tốc độ gió trung bình ngày, nhiệt độ không khí đặc trưng 2 và 3 ngày trước tại cùng vị trí có mức độ ảnh hưởng nhỏ nhất (khi loại bỏ chỉ làm tăng giá trị MAE ở mức dưới 2 %).

3.4 Thảo luận

Kết quả mô phỏng nhiệt độ không khí trung bình ngày tại trạm khí tượng Cần Thơ và Nhà Bè sử dụng nhiệt độ không khí trung bình ngày tại trạm khí tượng khác cách đó khoảng 130 km (và ngày trong năm) cho hệ số tương quan ở mức cao (R xấp xỉ 0.91 hay R^2 xấp xỉ 0.82). Điều này cho thấy có tiềm năng sử dụng thông tin nhiệt độ không khí trung bình ngày tại một địa điểm khác khá xa trong không gian đạt hiệu quả tốt trong mô phỏng; ở đây, độ chính xác thậm chí có thể còn cao hơn so với khi sử dụng số liệu các yếu tố khí tượng cùng ngày hoặc một vài ngày trước ở ngay tại địa điểm nghiên cứu. Nguyên nhân của hiện tượng trên rất có thể là do sự chi phối của các quy luật biến đổi nhiệt độ trên diện tích rộng dưới ảnh hưởng của bức xạ mặt trời trong cùng thời điểm. Nhận định này có tiềm năng áp dụng thực tế quan trọng nếu được kiểm chứng thêm, bởi việc sử dụng nhiệt độ không khí ở địa điểm khác (như dùng số liệu của trạm đo Nhà Bè để mô phỏng số liệu của các trạm đo ở Cà Mau, Bình Thuận) hỗ trợ cho các loại thông tin khí tượng ở cùng trạm có thể sẽ giúp tiết kiệm kinh phí quan trắc trong khi vẫn duy trì hiệu quả công tác thu thập và xử lý số liệu quan trắc nhiệt độ không khí, đặc biệt trong trường hợp bị mất số liệu (do thiết bị đo đặc bị hỏng...).

Kết quả nghiên cứu thu được cũng cho thấy rằng việc sử dụng số lượng lớn hơn các yếu tố đầu vào không phải lúc nào cũng tăng (ở mức đáng kể) độ chính xác trong mô phỏng của mô hình mạng nơ-ron. Điều này phù hợp với nhận xét của Đào Nguyên Khôi và Huỳnh Ái Phương [20], Nezhad và cộng sự [23].

Khoảng dao động của giá trị R thu được từ việc lựa chọn các cấu trúc ANN (nhỏ hơn 11%) đa số nhỏ hơn so với khoảng dao động R thu được từ việc thay đổi bộ số liệu đầu vào (lớn hơn 42%). Bên cạnh đó, đối với bộ số liệu đầu vào đầy đủ thông tin (bộ số liệu loại IV), ngay cả những cấu trúc ANN đơn giản với số lớp ẩn và số nơron ít vẫn cho giá trị R cao hơn so với cấu trúc ANN phức tạp nhưng lại được cung cấp bộ dữ liệu đầu vào ít thông tin hơn (các bộ số liệu loại I ÷ III). Hiện tượng trên khi được thể hiện ổn định cho cả 2 trạm khí tượng đã chứng minh khá rõ tầm quan trọng của bộ số liệu đầu vào trong mô phỏng. Theo như tác giả được biết, điều này hiện còn ít được đề cập và phân tích trong các nghiên cứu khác.

Kết quả mô phỏng đã cho thấy hiệu quả khai thác thông tin của mạng ANN có xu hướng bắt đầu tăng chậm dần khi cấu trúc ANN tăng dần độ phức tạp đến một mức nào đó (với những bộ số liệu trong nghiên cứu này là khoảng 6 nơron trong lớp ẩn cho mạng 1 lớp ẩn). Tuy vậy, các mạng với số nơron bằng 2 ÷ 4 trong lớp ẩn thường khá yếu để xử lý thông tin nên không thật phù hợp để mô phỏng và vẫn cần số nơron nhiều hơn để mô phỏng hữu hiệu cho các bộ số liệu khác nhau. Lý do có thể là khả năng khai thác thông tin từ dữ

liệu của mạng ANN không những chỉ là hàm số của chính cấu trúc ANN đó mà còn của cả độ phức tạp của các mối tương quan tồn tại trong những thông số đầu vào cụ thể. Mặc dù rất khó để so sánh cấu trúc ANN đơn giản nhất đủ để mô phỏng tốt trong nghiên cứu này với các nghiên cứu khí tượng của các tác giả khác do sự khác biệt trong bộ dữ liệu đầu vào sử dụng, kết quả số lượng 6 nơon trong lớp ẩn thu được ít hơn kết quả tương tự của Abudu và cộng sự [7], Abhishek và cộng sự [16], nhưng vẫn cao hơn của Nezhad và cộng sự [23].

Kết quả đánh giá ảnh hưởng bởi số lượng nơon trong mỗi lớp ẩn đến hiệu quả mô phỏng (R tổng quát trung bình) của các cấu trúc ANN bằng công cụ thống kê cho kết quả sự khác biệt có ý nghĩa đối với nhiều cấp số lượng nơon, đặc biệt là khi số nơon ít. Tuy nhiên, giá trị thực tế của sự khác biệt này cũng nên được cân nhắc kỹ bằng việc xét đến những nhược điểm gây ra bởi sự gia tăng số lượng nơon dẫn đến tăng độ phức tạp của mạng ANN so với lợi ích thu được của việc tăng giá trị R (bằng chứng cho thấy sự thay đổi không quá lớn của giá trị R đối với các bộ số liệu đầu vào cho trước cố định trong nghiên cứu này khi số nơon đã đạt đến mức nào đó).

4. KẾT LUẬN VÀ KIẾN NGHỊ

Nghiên cứu này đã góp phần chứng minh rằng việc sử dụng mạng nơon nhân tạo có triển vọng đem lại những kết quả khả quan trong công tác mô phỏng số liệu nhiệt độ không khí ở khu vực Nam Bộ và việc cân nhắc lựa chọn thông số đầu vào là vô cùng quan trọng. Kết quả mô phỏng với những bộ số liệu đo ở các trạm khí tượng Cần Thơ và Nhà Bè cho thấy giá trị hệ R tổng quát với hàm chuyển tansig dao động trong khoảng khá rộng ($0.5156 \div 0.9658$). Kết quả cũng chỉ ra mức 6 nơon trong lớp ẩn của cấu trúc ANN (1 lớp ẩn) với hàm chuyển tansig thường là đủ trong mô phỏng nhiệt độ không khí tại các trạm khí tượng Nam Bộ trên.

Độ chính xác trong mô phỏng nhiệt độ không khí cao hơn với bộ số liệu đầu vào chứa nhiệt độ không khí đặc trưng ngày của trạm khác và ngày trong năm, qua đó gián tiếp thể hiện sự chi phối vĩ mô trong khu vực khá mạnh đến nhiệt độ không khí ở cả 2 trạm đặc trưng cho các miền Đông và Tây Nam Bộ này. Khi sử dụng phương pháp rút dần từng yếu tố, kết quả đã chỉ rõ nhiệt độ không khí đặc trưng ngày tại trạm khác và độ ẩm không khí đặc trưng ngày tại trạm đang xem xét là các yếu tố có mức độ ảnh hưởng lớn nhất đến sai số MAE (làm tăng giá trị MAE ở mức khoảng 24 % và trên 14 % tương ứng nếu loại bỏ mỗi yếu tố này). Đặc biệt, tầm quan trọng của việc lựa chọn kỹ càng bộ số liệu cho mô phỏng còn thể hiện rõ khi mức dao động của giá trị R nhỏ hơn từ việc lựa chọn các cấu trúc ANN so với từ việc thay đổi bộ số liệu đầu vào.

Dù đạt được mục tiêu đề ra nhưng nghiên cứu này còn một số hạn chế có thể được khắc phục trong tương lai. Bởi số liệu thu thập được để sử dụng mới chỉ giới hạn cho 2 trạm khí tượng Nhà Bè và Cần Thơ tức là chịu ảnh hưởng đặc thù của khí hậu Nam Bộ, để kết quả mô phỏng khí tượng mang tính tổng quát cao hơn, các nghiên cứu sau có thể mở rộng thêm đến những vị trí khác nằm cách xa nhau với chuỗi số liệu dài hơn. Bên cạnh đó, các bộ thông số khí tượng khác nhau cũng có thể được đưa vào đánh giá trong vai trò số liệu đầu vào.

LỜI CẢM ƠN

Tác giả xin cảm ơn Viện Khoa học Công nghệ và Quản lý Môi trường, các đồng nghiệp đã hỗ trợ trong quá trình thực hiện nghiên cứu này.

TÀI LIỆU THAM KHẢO

- [1] M. Hayati, and Z. Mohebi, Application of Artificial Neural Networks for Temperature Forecasting, *World Academy of Science, Engineering and Technology International Journal of Electrical and Computer Engineering*, 1(4), pp. 662-666, 2007.
- [2] S. S. Baboo and I. K. Shereef, An Efficient Weather Forecasting System using Artificial Neural Network, *International Journal of Environmental Science and Development*, 1(4), pp. 321-325, 2010.

- [3] G. Landeras, A. Ortiz-Barredo, and J. J. López, Forecasting weekly evapotranspiration with ARIMA and artificial neural network models, *Journal of Irrigation and Drainage Engineering*, 135(3), pp. 323-334, 2009.
- [4] K. P. Sudheer, A. K. Gosain, and K. S. Ramasastri, Estimating actual evapotranspiration from limited climatic data using neural computing technique, *Journal of Irrigation and Drainage Engineering*, 129(3), pp. 214-218, 2003.
- [5] M. Kumar, A. Bandyopadhyay, N. S. Raghuvanshi, and R. Singh, Comparative study of conventional and artificial neural network-based ET_o estimation models, *Irrigation Science*, 26, pp. 531-545, 2008.
- [6] O. Kissi, Generalized regression neural networks for evapotranspiration modeling, *Hydrological Sciences*, 51(6), pp. 1092-1105, 2006.
- [7] S. Abudu, A. S. Bawazir, and J. P. King, Infilling missing daily evapotranspiration data using neural networks, *Journal of Irrigation and Drainage Engineering*, 136 (5), pp. 317-325, 2010.
- [8] A. R. Gurung, Forecasting Weather System Using Artificial Neural Network (ANN): A Survey Paper, *International Journal of Latest Engineering Research and Applications (IJLERA)*, 2(10), pp. 42-50, 2017.
- [9] I. Odeh, Temperature Prediction in Jordan using ANN, *International Journal of Computer Science and Information Technologies*, 7 (1), pp. 378-383, 2016.
- [10] M. Narvekar, P. Fargose, Daily weather forecasting using Artificial Neural Network, *International Journal of Computer Applications (0975-8887)*, Volume 121-No.22, pp. 9-13, 2015.
- [11] A. Kaur, H. Singh, Artificial Neural Networks in Forecasting Minimum Temperature, *International Journal of Electronics & Communication Technology*, 2(3), pp. 101-105, 2011.
- [12] B. Ustaoglu, H. K. Cigizoglu and M. Karaca, Forecast of daily mean, maximum and minimum temperature time series by three Artificial neural network methods, *Meteorological Application*, 15, pp. 431-445, 2008.
- [13] G. P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing*, 50, pp. 159-175, 2003.
- [14] H. Anysz, A. Zbiciak, N. Ibadov, The influence of input data standardization method on prediction accuracy of artificial neural networks, *Procedia Engineering*, 153, pp. 66-70, 2016.
- [15] X. Zhang, Q. Zhang, G. Zhang, Z. Nie, Z. Gui and H. Que, A Novel Hybrid Data-Driven Model for Daily Land Surface Temperature Forecasting Using Long Short-Term Memory Neural Network Based on Ensemble Empirical Mode Decomposition, *International Journal of Environmental Research and Public Health*, 15, p. 1032, 2018.
- [16] K. Abhishek, M. P. Singh, S. Ghosh, A. Anand, Weather forecasting model using Artificial Neural Network, *Procedia Technology*, 4, pp. 311-318, 2012.
- [17] S. A. Kakar, N. Sheikh, A. Naseem, S. Iqbal, A. Rehman, A. Kakar, B. A. Kakar, H. A. Kakar, B. Khan, Artificial Neural Network based Weather Prediction using Back Propagation Technique, *International Journal of Advanced Computer Science and Applications*, 9 (8), pp. 462-470, 2018.
- [18] Nguyễn Quang Hoan, Phạm Thị Trang, Hoàng Hồng Công, Nguyễn Thị Huyền, Dự báo thời tiết ứng dụng mạng nơon nhân tạo và thuật toán Bayes, *Tạp chí Khoa học & Công nghệ*, Số 13, trang 39 - 43, 2017.
- [19] Nguyễn Tân Ân và Nguyễn Quang Hoan, Hệ dự báo thời tiết với ứng dụng của mạng nơon nhân tạo, *Tạp chí Khoa học & Công nghệ*, 90(02), trang 65-70, 2012.
- [20] Đào Nguyên Khôi và Huỳnh Ái Phương, Mô phỏng dòng chảy lưu vực sông Sêrêpôk với mạng nơ-ron nhân tạo, *Phát triển Khoa học & Công nghệ*, 19(3), trang 114-124, 2016.
- [21] Mô tả về Nam Bộ: https://vi.wikipedia.org/wiki/Nam_B%E1%BB%99 (truy cập tháng 4/2020)
- [22] IBM SPSS Statistics Base 26, 2019.
- [23] E. F. Nezhad, G. F. Ghalhari, F. Bayatani, Forecasting Maximum Seasonal Temperature Using Artificial Neural Networks “Tehran Case Study”. *Asia-Pacific Journal of Atmospheric Sciences*, 55, pp. 145-153, 2019.

Ngày nhận bài: 24/03/2020

Ngày chấp nhận đăng: 09/07/2021