

ẢNH HƯỞNG CỦA ĐỘ DÀI CHUỖI SỐ LIỆU ĐẦU VÀO ĐẾN KẾT QUẢ MÔ PHỎNG NHIỆT ĐỘ KHÔNG KHÍ BẰNG MẠNG NƠN NHÂN TẠO (ANN) TẠI ĐỒNG BẰNG NAM BỘ

TRẦN TRÍ DŨNG

Trường Đại học Công nghiệp Thành phố Hồ Chí Minh
trantridung@iuh.edu.vn

Tóm tắt. Mạng nơon nhân tạo (ANN) được sử dụng để mô phỏng nhiệt độ không khí cho trạm khí tượng Nhà Bè. Bộ số liệu sử dụng gồm 5 yếu tố đầu vào với 1 yếu tố đầu ra là nhiệt độ không khí. Cấu trúc ANN trong phần mềm Matlab được thiết kế gồm 2 lớp ẩn với 3 cấp số lượng nơon (2, 5 và 8) trong mỗi lớp ẩn và hàm chuyển tansig. Những độ dài chuỗi số liệu khác nhau từ 1 tháng đến 48 tháng khi khảo sát đã cho kết quả R từ 0.8318 đến 0.9673. Giá trị R thay đổi không theo quy luật khi độ dài của chuỗi số liệu từ 4 tháng trở xuống nhưng lại mang xu hướng giống nhau cho cả 3 cấu trúc ANN khi số liệu dài bằng 6 tháng hoặc hơn. Với cùng cấu trúc ANN, độ dài chuỗi số liệu tăng không đảm bảo cho giá trị R tăng. Sự sai lệch trong kết quả mô phỏng xảy ra mạnh hơn ở những đoạn đỉnh thấp hay cao của chuỗi số liệu, nhất là khi cấu trúc ANN có số lượng nơon ít trong lớp ẩn.

Từ khóa. Nhiệt độ không khí, yếu tố khí tượng, độ dài chuỗi số liệu, huấn luyện, mạng nơon nhân tạo (ANN), mô phỏng.

EFFECT OF THE LENGTH OF THE INPUT DATA SERIES ON SIMULATION RESULTS OF AIR TEMPERATURE BY ARTIFICIAL NEURAL NETWORK (ANN) IN NAM BO PLAIN

Abstract. Artificial neural network (ANN) was used to simulate air temperature for Nha Be meteorological station. The data set used includes 5 inputs with an output element of air temperature. ANN structures in Matlab software were designed of 2 hidden layers with 3 levels of neuron number (2, 5 and 8) in each hidden layer and tansig transfer function. Data series lengths varied from 1 month to 48 months giving R from 0.8318 to 0.9673. R values did not follow any certain rule when the length of the data series was no longer than 4 months but bear the same tendency for all three ANN structures when the data was equal or longer than 6 months. With the same ANN structure, the increase in the data series length did not guarantee an increase in R value. The deviation in simulation results from measured data occurred more strongly in the sections of the low or high peaks of the data series, especially when the ANN structure has a small number of neurons in the hidden layer.

Keywords. air temperature, meteorological factors, length of data series, training, artificial neural network (ANN), simulation.

1 MỞ ĐẦU

Nhiệt độ không khí có ảnh hưởng hết sức quan trọng trong rất nhiều lĩnh vực kinh tế - xã hội như sản xuất nông nghiệp, chăn nuôi, giao thông vận tải, y tế. Là một nhân tố khí tượng chủ yếu, nhiệt độ không khí tác động mạnh mẽ tới cuộc sống và hoạt động của con người. Cho đến nay, trên thế giới đã có nhiều công trình được công bố về đặc điểm hình thành cũng như thay đổi của nhiệt độ không khí ở nhiều quốc gia. Trong số đó, một số nghiên cứu đã nêu rõ vai trò của các yếu tố khí tượng liên quan cần được đưa vào phép tính toán hay kỹ thuật dự báo nhiệt độ không khí như độ ẩm tương đối không khí, tổng bức xạ mặt trời, lượng mưa, lượng bốc hơi và tốc độ gió [2, 5, 6, 17].

30 ẢNH HƯỞNG CỦA ĐỘ DÀI CHUỖI SỐ LIỆU ĐẦU VÀO ĐẾN KẾT QUẢ MÔ PHỎNG NHIỆT ĐỘ KHÔNG KHÍ BẰNG MẠNG NƠN NHÂN TẠO (ANN) TẠI ĐỒNG BẰNG NAM BỘ

Để mô phỏng những hiện tượng trong tự nhiên, các nhà khoa học đã sử dụng nhiều cách tiếp cận khác nhau như ARIMA hay hồi quy đa biến. Đặc biệt, mạng nơron nhân tạo (ANN) là một trong những phương pháp tiên tiến có khả năng mô phỏng hữu hiệu các quá trình diễn biến phức tạp và đã thể hiện hết sức hiệu quả trong thực tế. ANN là một hệ thống huấn luyện dựa trên tổ hợp nơron thần kinh với các kết nối nhằm thu nhận và xử lý nguồn dữ liệu thông tin đưa vào. Sức mạnh của kỹ thuật ANN một phần nằm ở khả năng tìm ra được mối quan hệ phi tuyến phức tạp và ẩn giữa nhiều yếu tố mà đôi khi rất khó có thể đạt được bằng những phương pháp khác. Rất nhiều tài liệu trong nước và nước ngoài ở các lĩnh vực như khí tượng, tài nguyên nước hay môi trường đã đề cập khá chi tiết đến những kết quả tốt thu được từ phương pháp này [1, 3, 4, 7, 8, 9, 10, 13, 14, 15, 18]. Ở nước ta, kỹ thuật ANN cũng đã được áp dụng hiệu quả trong một số lĩnh vực khoa học và đời sống như công tác điều tra nghiên cứu những đặc trưng khí hậu [11, 12, 16].

Với đà phát triển của cuộc cách mạng công nghiệp 4.0, kỹ thuật ANN ngày càng nhận được sự quan tâm lớn hơn. Bởi đây là phương pháp có ứng dụng rộng rãi và đã chứng minh được hiệu quả trên thực tế, việc đánh giá ảnh hưởng các đặc tính số liệu sử dụng đến kết quả ước lượng và mô phỏng bằng ANN là một vấn đề hữu ích. Do đặc điểm hoạt động, kỹ thuật ANN thường khai thác số liệu để từ đó đưa ra kết luận và dẫn đến kết quả sẽ phụ thuộc vào đặc tính số liệu được cung cấp. Chính vì vậy, độ chính xác của kết quả thu được từ việc sử dụng ANN sẽ có thể được nâng cao khi áp dụng chuỗi số liệu hợp lý.

Hiện nay, số lượng các công trình đã được công bố về công tác mô phỏng những thông số khí tượng ở miền Đông Nam Bộ bằng công cụ ANN tạm thời còn chưa nhiều. Đặc biệt, những tác động của đặc điểm số liệu đầu vào đến hiệu quả mô phỏng theo kỹ thuật nêu trên cũng đang còn ít được đề cập tới. Nghiên cứu này đánh giá tác động của độ dài chuỗi số liệu khí tượng khi sử dụng các cấu trúc ANN khác nhau tới kết quả mô phỏng nhiệt độ không khí tại trạm khí tượng Nhà Bè thuộc khu vực Thành phố Hồ Chí Minh. Những kết quả thu được sẽ góp phần làm rõ thêm hoạt động của ANN trong nghiên cứu khí tượng ở nước ta, đồng thời có thể dùng làm tài liệu tham khảo cho các nhà quản lý hay nhà khoa học khi cần mô phỏng số liệu khí tượng cho các mục đích khác nhau.

2 SỐ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Số liệu

Trạm khí tượng Nhà Bè thuộc lưới trạm điều tra cơ bản của Trung tâm Khí tượng Thủy văn Quốc gia. Trạm được đưa vào hoạt động từ ngày 15/11/2012 và bắt đầu cung cấp số liệu quan trắc từ ngày 1/12/2012. Các yếu tố quan trắc bao gồm những yếu tố khí tượng cơ bản và các hiện tượng thời tiết.

Tọa độ địa lý của trạm: $106^{\circ}43'41''$ Kinh độ Đông, $10^{\circ}39'36''$ Vĩ độ Bắc.



Hình 1. Vị trí Trạm khí tượng Nhà Bè (Nguồn: Here.DigitalGlobe)

Số liệu được sử dụng trong bài báo này là kết quả quan trắc cho các yếu tố khí tượng của Trạm khí tượng Nhà Bè thuộc giai đoạn 1/1/2013 đến 31/12/2016. 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 nhằm tăng cường hiệu quả xử lý của các cấu trúc ANN trong phần mềm Matlab. Theo [9], công thức chuyển đổi số liệu trước khi đưa vào mô phỏng 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.

Phương pháp nghiên cứu

Mạng Nơron nhân tạo (ANN) trong phần mềm Matlab được sử dụng để mô phỏng nhiệt độ không khí trung bình ngày ($^{\circ}\text{C}$) bằng cách sử dụng 5 yếu tố đầu vào bao gồm độ ẩm tương đối không khí trung bình ngày (%), tổng bức xạ mặt trời trong ngày (W/m^2), tốc độ gió trung bình trong ngày (m/s), lượng bốc hơi ngày (mm), ngày trong năm. Các thông số đầu vào trên được chọn bởi chúng có mối liên quan vật lý mật thiết đến nhiệt độ không khí cũng như sự phổ biến của chúng trong các công trình nghiên cứu khác trên thế giới về chủ đề.

Công tác mô phỏng được bắt đầu bằng cách cho ANN học trên bộ số liệu đầu vào (gồm 5 thông số đầu vào và 1 thông số đầu ra là nhiệt độ không khí). Sau đó, mối quan hệ vừa tìm được sẽ được áp dụng với 5 thông số đầu vào để mô phỏng lại thông số đầu ra (nhiệt độ không khí). Sự trùng khớp giữa chuỗi giá trị nhiệt độ không khí mô phỏng (tính toán) và thực đo thể hiện cho hiệu quả của công tác mô phỏng.

Cấu trúc ANN trong phần mềm Matlab sử dụng để mô phỏng nhiệt độ không khí trung bình ngày được thiết kế gồm 2 lớp ẩn với số lượng nơron trong mỗi lớp ẩn là như nhau và 3 cấp số lượng nơron (2, 5 và 8). Lý do số lượng nơron trong lớp ẩn tối đa là 8 được lựa chọn là vì theo một kết quả nghiên cứu trước của tác giả, khi tăng số nơron trong lớp ẩn lên hơn nữa thì hiệu suất mô phỏng đối với nhiệt độ không khí trung bình ngày tại Trạm Nhà Bè sẽ tăng lên không đáng kể. Trong nghiên cứu này, các cấu trúc ANN đều sử dụng thuật toán huấn luyện Levenberg-Marquardt backpropagation và số giai đoạn huấn luyện tối đa là 1000. Hàm “tansig” được sử dụng trong các lớp trung gian và hàm đầu ra là “purelin”.

Công tác mô phỏng được tiến hành với các độ dài chuỗi số liệu khác nhau: 1 tháng, 2 tháng, 4 tháng, 6 tháng, 12 tháng, 18 tháng, 24 tháng, 36 tháng và 48 tháng (chung mốc thời điểm đầu là 1/1/2013). Giới hạn thời gian 48 tháng trong nghiên cứu này bị hạn chế bởi độ dài tối đa của bộ số liệu thu thập được. Việc chia ra các độ dài đoạn số liệu khác nhau cho phép đánh giá ảnh hưởng của độ dài chuỗi số liệu đầu vào đến độ chính xác trong mô phỏng nhiệt độ không khí. 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 công thức 1 ở trên) và sau khi hoàn thành mô phỏng sẽ được chuyển đổi ngược lại.

Độ chính xác trong mô phỏng hiện tượng của các cấu trúc ANN khác nhau được đánh giá chủ yếu bằng hệ số tương quan (R) có tham khảo sai số tuyệt đối trung bình (MAE) và căn bậc hai của trung bình bình phương sai số (RMSE) giữa kết quả mô phỏng và thực đo.

$$R = \frac{\sum_{i=1}^k (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\left[\sum_{i=1}^k (y_i - \bar{y})^2 \right] \left[\sum_{i=1}^k (\hat{y}_i - \bar{\hat{y}})^2 \right]} \quad (2)$$

$$\text{MAE} = \frac{1}{k} \sum_{i=1}^k |y_i - \hat{y}_i| \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^k (y_i - \hat{y}_i)^2}{k}} \quad (4)$$

trong đó:

y_i - giá trị thực đo;

\hat{y}_i - giá trị mô phỏng;

\bar{y} - trung bình các giá trị thực đo;

$\bar{\hat{y}}$ - trung bình các giá trị mô phỏng;

k - tổng số lượng các giá trị;

i - số thứ tự giá trị.

3 KẾT QUẢ VÀ THẢO LUẬN

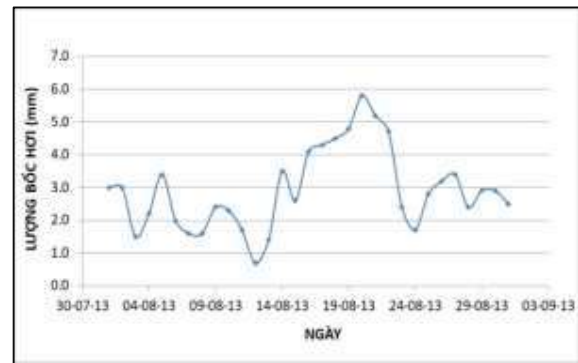
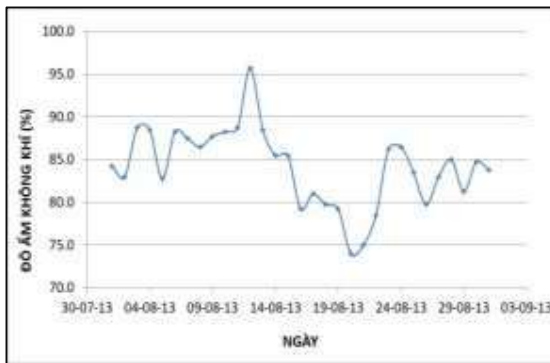
Kiểm tra chất lượng số liệu đưa vào mô phỏng

Bởi độ tin cậy khi sử dụng công cụ ANN trong mô phỏng nhiệt độ không khí tại trạm khí tượng Nhà Bè phụ thuộc nhiều vào chất lượng của số liệu sử dụng nên ta cần tiến hành công tác kiểm tra số liệu thô trước khi bắt đầu mô phỏng. Mục tiêu nhằm loại bỏ những số liệu không hợp lý có thể do nhiều nguyên nhân khách quan và chủ quan khác nhau gây ra.

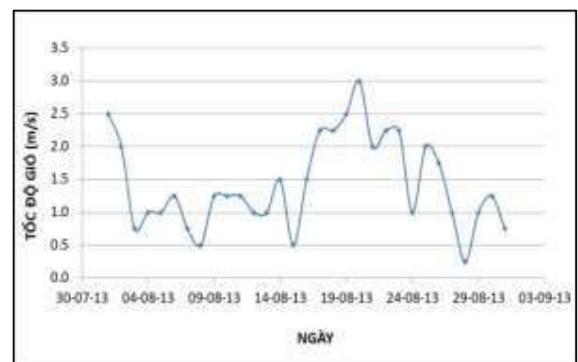
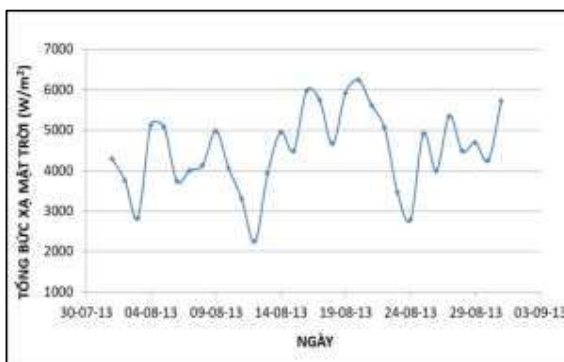
Số liệu đầu vào cho công tác mô phỏng được cung cấp bởi Đài khí tượng thủy văn khu vực Nam Bộ. Chuỗi số liệu khí tượng 4 năm (2013-2016) đã được kiểm tra để loại bỏ những số liệu kém chất lượng như giá trị quá cao hoặc quá thấp không hợp lý. Kết quả kiểm tra cho thấy số liệu đều có chất lượng tốt, ngoại trừ ngày 22/9/2013 có giá trị tổng lượng bức xạ ngày bị khuyết (bởi nguyên nhân không rõ). Số liệu ngày nêu trên đã bị loại khỏi tính toán ở các bước tiếp theo. Những thống kê về số liệu đầu vào được thể hiện trong bảng 1; các đồ thị số liệu 1 tháng đại diện (tháng 8 năm 2013) được thể hiện trong các hình 2 ÷ 6.

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

Số thứ tự	Tên yếu tố	Đơn vị đo	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	%	96.50	58.75	79.99	6.30
2	Lượng bốc hơi ngày	mm	7.60	0.40	3.04	1.15
3	Tổng lượng bức xạ mặt trời ngày	w/m ²	7212.00	543.00	4545.96	1196.51
4	Tốc độ gió trung bình ngày	m/s	4.25	0.00	1.22	0.64
5	Nhiệt độ không khí trung bình ngày	°C	31.78	22.53	27.79	1.50

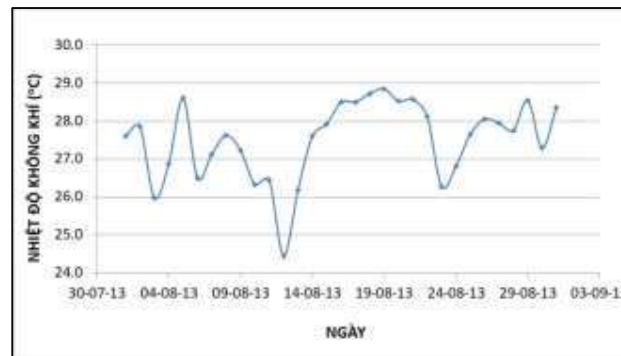


Hình 2. Độ ẩm không khí đo tại Trạm Nhà Bè T.8/2013 Hình 3. Lượng bốc hơi ngày đo tại Trạm Nhà Bè T.8/2013



Hình 4. Bức xạ mặt trời đo tại Trạm Nhà Bè T.8/2013

Hình 5. Tốc độ gió ngày đo tại Trạm Nhà Bè T.8/2013



Hình 6. Nhiệt độ không khí đo tại Trạm Nhà Bè tháng 8/2013

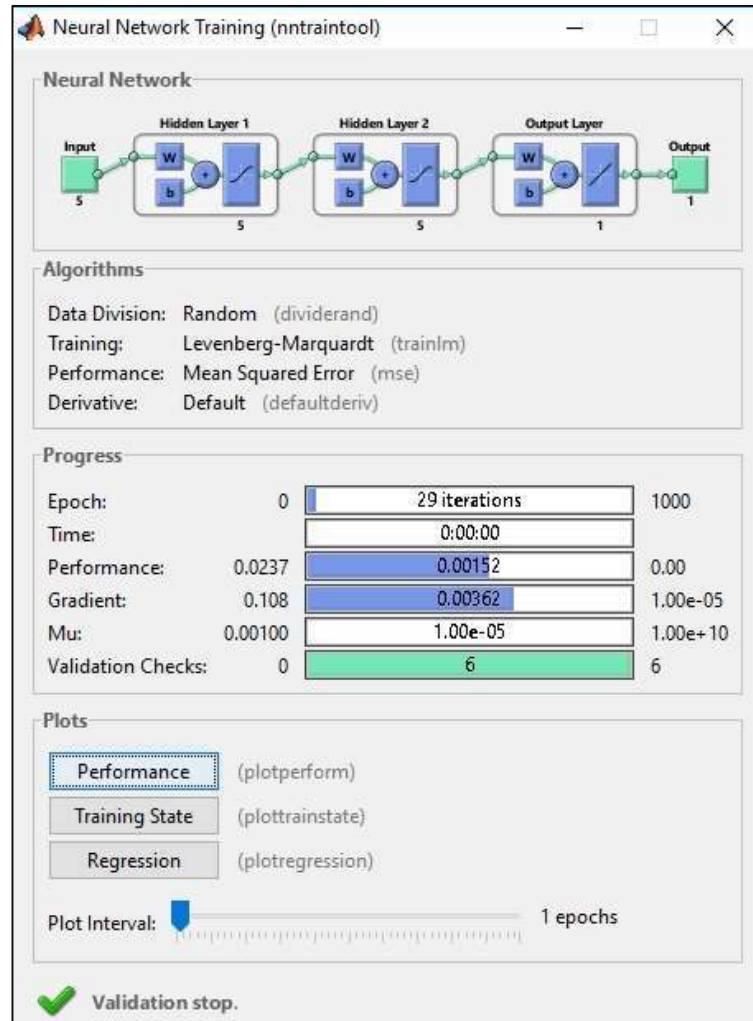
Mô phỏng nhiệt độ không khí bằng công cụ ANN

Sau khi qua kiểm tra chất lượng ở bước trên, tổng số mẫu số liệu khí tượng thực đo ngày đưa vào mô phỏng là 1459. Số lượng mẫu cụ thể sử dụng cho mỗi kịch bản chuỗi số liệu phụ thuộc vào số ngày trong (các) tháng thành phần nằm trong chuỗi, như tháng 1 có 31 ngày... Mỗi mẫu số liệu gồm thông tin của 5 yếu tố đầu vào (độ ẩm tương đối không khí trung bình ngày, lượng bốc hơi ngày, tổng lượng bức xạ mặt trời ngày, tốc độ gió trung bình ngày, ngày trong năm) với 1 yếu tố đầu ra là nhiệt độ không khí trung bình ngày. Các số liệu đã được phần mềm Matlab lựa chọn ngẫu nhiên theo tỷ lệ 70% số lượng mẫu để phục vụ cho các mục đích huấn luyện mạng, 15% cho xác nhận tổng hợp mạng và tránh hiện tượng quá ngưỡng, 15% cho việc kiểm tra độc lập.

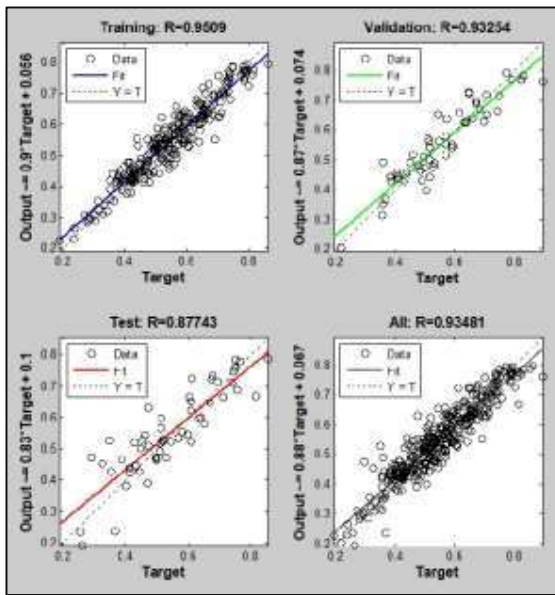
Việc mô phỏng được tiến hành 1000 lần cho mỗi tổ hợp cấu trúc ANN và độ dài chuỗi số liệu mô tả ở trên, trong đó số liệu đầu vào được phần mềm lựa chọn ngẫu nhiên theo tỷ lệ 70/15/15 để mô phỏng. Cho từng cấu trúc ANN, lần chạy cho giá trị hệ số tương quan R tổng (giữa giá trị nhiệt độ không khí mô phỏng và thực đo) lớn nhất được ghi nhận. Tổng hợp kết quả mô phỏng dựa trên những giá trị hệ số tương

34 ẢNH HƯỞNG CỦA ĐỘ DÀI CHUỖI SỐ LIỆU ĐẦU VÀO ĐẾN KẾT QUẢ MÔ PHỎNG NHIỆT ĐỘ KHÔNG KHÍ BẰNG MẠNG NƠN NHÂN TẠO (ANN) TẠI ĐỒNG BẰNG NAM BỘ

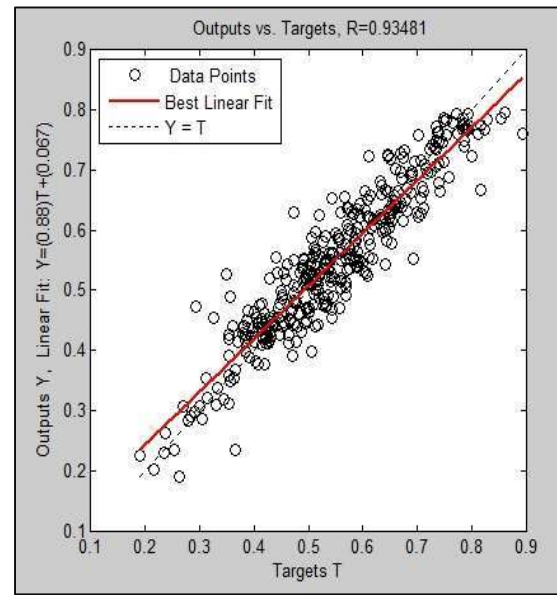
quan (R) và sai số tuyệt đối trung bình (MAE) tương ứng cho những tổ hợp của cấu trúc ANN và độ dài chuỗi số liệu khác nhau như liệt kê trong bảng 2 và các hình 7 ÷ 11 dưới đây.



Hình 7. Mô phỏng ANN cho cấu trúc 5-5-5-1 và độ dài chuỗi số liệu 12 tháng



Hình 8. Kết quả R với cấu trúc 5-5-5-1 và 12 tháng số liệu



Hình 9. Kết quả R tổng với cấu trúc 5-5-5-1 và 12 tháng số liệu



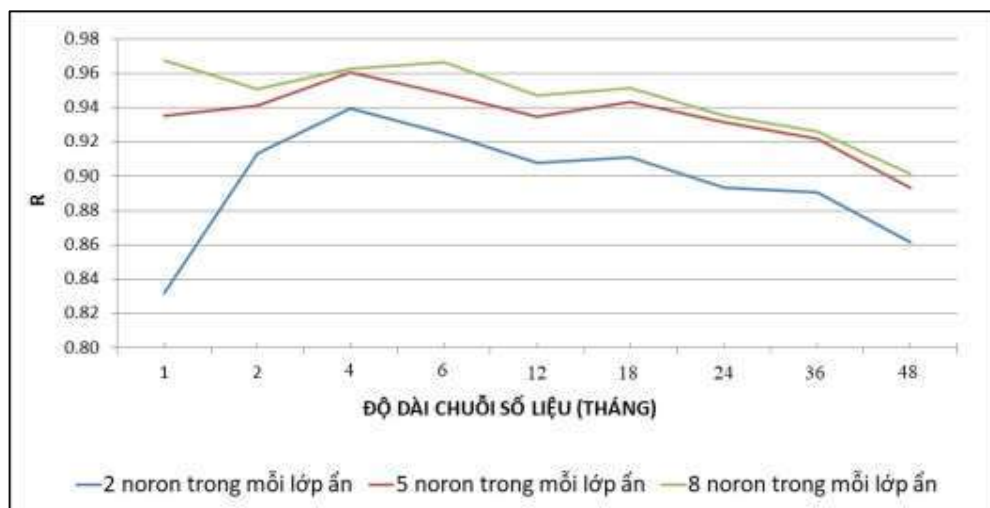
Hình 10. So sánh kết quả mô phỏng nhiệt độ không khí sử dụng cấu trúc ANN 5-5-5-1 và 12 tháng số liệu đầu vào với dữ liệu đo hiện trường cho giai đoạn 1/5/2013 - 30/6/2013

Bảng 2. Kết quả mô phỏng nhiệt độ không khí tại trạm khí tượng Nhà Bè với các độ dài chuỗi số liệu khác nhau và mạng ANN 2 lớp ẩn

Tên thông số	Độ dài chuỗi số liệu (tháng)	Số nơon trong mỗi lớp ẩn		
		2	5	8
R	1	0.8318	0.9352	0.9673
	2	0.9132	0.9408	0.9508
	4	0.9397	0.9605	0.9627
	6	0.9251	0.9483	0.9665
	12	0.9077	0.9348	0.9468
	18	0.9110	0.9431	0.9512

36 ẢNH HƯỞNG CỦA ĐỘ DÀI CHUỖI SỐ LIỆU ĐẦU VÀO ĐẾN KẾT QUẢ MÔ PHỎNG NHIỆT ĐỘ KHÔNG KHÍ BẰNG MẠNG NƠON NHÂN TẠO (ANN) TẠI ĐỒNG BẰNG NAM BỘ

	24	0.8930	0.9311	0.9350
	36	0.8903	0.9218	0.9260
	48	0.8617	0.8930	0.9013
RMSE	1	0.56791	0.54208	0.54568
	2	0.69337	0.69714	0.65917
	4	0.98359	1.00957	0.98752
	6	0.94714	0.85939	0.85878
	12	0.71227	0.65399	0.61600
	18	0.75924	0.64285	0.61211
	24	0.74799	0.63000	0.61776
	36	0.72661	0.62536	0.60924
	48	0.76076	0.67481	0.65107
MAE	1	0.4692	0.4348	0.4380
	2	0.5879	0.5889	0.5551
	4	0.7727	0.8359	0.8285
	6	0.7512	0.6845	0.7081
	12	0.5783	0.5209	0.4995
	18	0.5990	0.5018	0.4864
	24	0.6001	0.5084	0.4954
	36	0.5741	0.4943	0.4846
	48	0.5709	0.5145	0.4985



Hình 11. Kết quả giá trị R thu được từ công tác mô phỏng nhiệt độ không khí với độ dài chuỗi số liệu khác nhau (trục hoành không theo tỷ lệ)

Kết quả mô phỏng đã cho thấy một số điểm quan trọng sau:

- Giá trị R thu được qua mô phỏng thay đổi trong khoảng từ 0.8318 đến 0.9673. Hệ số tương quan R nhỏ nhất khi số liệu đầu vào ngắn nhất (1 tháng) và cấu trúc ANN có số nơon ít nhất (2) trong lớp ẩn, chỉ đạt 0.8318 (hay $R^2 = 0.6918$). Dù rằng giá trị R nhỏ nhất vẫn đủ lớn để chấp nhận độ chính xác của kết quả mô phỏng nhiệt độ không khí (khi R^2 đạt mức từ $0.6 \div 0.7$) nhưng kết quả cũng cho thấy không nên sử dụng cấu trúc ANN có số nơon trong lớp ẩn bằng 2 (tức 5-2-2-1) khi độ dài chuỗi số liệu biến đổi mà nên sử dụng số nơon trong lớp ẩn ít nhất bằng 5. Bên cạnh đó, trong đồ thị, những điểm mô phỏng bị sai lệch nhiều so với số liệu thực đo thường nằm ở các đoạn đỉnh thấp hay cao của chuỗi số liệu, điều này đặc biệt rõ hơn khi số lượng nơon trong lớp ẩn của cấu trúc ANN nhỏ.

- Với cùng một độ dài chuỗi dữ liệu, cấu trúc ANN (2 lớp) nào có số lượng nơon trong lớp ẩn lớn hơn sẽ cho giá trị R cao hơn. Điều này đã thể hiện rõ sức ảnh hưởng của số lượng nơon trong mô phỏng. Tuy nhiên, giá trị R tăng chậm dần (thể hiện qua các đường kết quả R gần nhau hơn trên đồ thị hình 11) khi số lượng nơon trong lớp ẩn tăng lên.

- Với cùng một cấu trúc ANN, độ dài chuỗi số liệu đưa vào mô phỏng tăng không chắc sẽ dẫn đến sự tăng lên của giá trị R hay độ chính xác của mô phỏng (ví dụ: 48 tháng số liệu cho R thấp hơn 6 tháng số liệu với mọi cấu trúc ANN được khảo sát). Điều này xảy ra có thể do những số liệu khác biệt mới đưa vào chuỗi dài hơn lại mang tính biến động khác biệt nhiều so với quy luật vốn có của các chuỗi dữ liệu ngắn, từ đó gây khó khăn cho quá trình mô phỏng và ảnh hưởng đến độ chính xác của kết quả.

- Khi độ dài của chuỗi số liệu đầu vào ngắn từ 4 tháng trở xuống, sự biến đổi R trong mô phỏng không tuân theo một quy luật nhất định. Nhưng với các chuỗi số liệu có độ dài từ 6 tháng trở lên, sự biến đổi trong giá trị R thu được từ cả 3 cấu trúc ANN rất đồng nhất về xu hướng (cùng tăng hay cùng giảm) dù biên độ thay đổi của R từ các cấu trúc ANN nêu trên có khác nhau.

4 KẾT LUẬN

Công tác mô phỏng nhiệt độ không khí tại trạm khí tượng Nhà Bè đã cho thấy tầm quan trọng của độ dài chuỗi số liệu đầu vào đến độ chính xác của kết quả mô phỏng khi áp dụng những cấu trúc ANN khác nhau. Sự biến đổi của giá trị R trong mô phỏng từ chỗ không tuân theo một quy luật nhất định khi độ dài của chuỗi số liệu đầu vào ngắn dưới 4 tháng đến chỗ có xu hướng giống nhau khi sử dụng các chuỗi có độ dài từ 6 tháng trở lên được thể hiện rõ cho 3 cấu trúc ANN đã lựa chọn. Tuy nhiên, độ dài chuỗi số liệu đưa vào mô phỏng tăng lên chưa chắc sẽ dẫn đến sự tăng lên của giá trị R tức chất lượng mô phỏng (trong nghiên cứu này thậm chí còn có xu hướng giảm dần) khi sử dụng cùng một cấu trúc ANN bởi kết quả rất có thể còn phụ thuộc vào độ phức tạp của dữ liệu đầu vào.

Một kết quả khác từ công tác mô phỏng nhiệt độ không khí là sự sai lệch giữa kết quả mô phỏng với kết quả đo thường xảy ra mạnh hơn ở đoạn thuộc các đỉnh thấp hay cao trong đồ thị của chuỗi số liệu. Đặc biệt, điều này càng thể hiện rõ rệt hơn trong trường hợp cấu trúc ANN với số lượng nơon ít trong lớp ẩn được sử dụng để mô phỏng.

Nghiên cứu này đã sử dụng 4 năm số liệu thực đo tại một trạm khí tượng miền Đông Nam Bộ trong quá trình mô phỏng. Dù đánh giá đã được thực hiện với những độ dài khác nhau của chuỗi số liệu nhưng vị trí khảo sát mới chỉ giới hạn ở một trạm khí tượng nên cần được thử nghiệm thêm bằng số liệu thu thập từ những trạm khí tượng ở các địa phương có đặc điểm tự nhiên khác nhau để có thể tổng quát hóa. Mặt khác, các tổ hợp thông số khí tượng cũng như cấu trúc ANN thay đổi cũng có thể sẽ có tác động tới mức độ ảnh hưởng của độ dài chuỗi số liệu đến kết quả mô phỏng nhiệt độ không khí và cũng cần được điều tra thêm trong tương lai.

LỜI CẢM ƠN

Tác giả xin cảm ơn Trường Đại học Công nghiệp thành phố Hồ Chí Minh đã tạo mọi điều kiện thuận lợi để cho nghiên cứu này có thể hoàn thành. Rất cảm ơn các đồng nghiệp ở Viện Khoa học Công nghệ và Quản lý Môi trường đã hỗ trợ tác giả trong quá trình thực hiện công việc. Xin cảm ơn Đài khí tượng thủy văn khu vực Nam Bộ đã cung cấp số liệu cho nghiên cứu.

TÀI LIỆU THAM KHẢO

- [1] Aqil M., Kita I., Yano A., Nishiyama S., Neural networks for real time catchment flow modeling and prediction, *Water Resources Management*, 21, pp. 1781-1796, 2007.
- [2] Baboo S.S. and Shereef I. K., An Efficient Weather Forecasting System using Artificial Neural Network, *International Journal of Environmental Science and Development*, 1(4), pp. 321-325, 2010.
- [3] Đào Nguyên Khôi và Huỳnh Ái Phương, Mô phỏng dòng chảy lưu vực sông Sêrêpôk với mạng nơ-ron nhân tạo, *Science & Technology Development*, 19: T3-2016, pp. 114-124, 2016.
- [4] Granger R. J. and Hedstrom N., Modeling hourly rates of evaporation from small lakes, *Hydrology and Earth System Science*, 15, pp. 267-277, 2011.
- [5] Hayati M, and Mohebi Z., 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.
- [6] Kaur A., Singh H., Artificial Neural Networks in Forecasting Minimum Temperature, *International Journal of Electronics & Communication Technology*, 2(3), pp. 101-105, 2011.
- [7] Kleiber W., Katz R. W. and Rajagopalan B., Daily minimum and maximum temperature simulation over complex terrain. *The Annals of Applied Statistics*, Vol. 7, No. 1, pp. 588 - 612, 2013.
- [8] Kumar M., Bandyopadhyay A., Raghuvanshi N.S., and Singh R., Comparative study of conventional and artificial neural network-based ETo estimation models, *Irrigation Science*, 26, pp. 531-545, 2008.
- [9] Lê Văn Nghinh, Hoàng Thanh Tùng, Nguyễn Ngọc Hải, Nghiên cứu ứng dụng mạng nơron thần kinh vào dự báo lũ các sông ở tỉnh Bình Định và Quảng Trị, *Tạp chí Khoa học kỹ thuật Thủy lợi và Môi trường*, Số 14 (8/2006), trang 65-70, 2006.
- [10] Narvekar M., Fargose P., Daily weather forecasting using Artificial Neural Network, *International Journal of Computer Applications (0975-8887)*, Volume 121-No.22, pp. 9-13, 2015.
- [11] Nguyễn Quang Hoan, Phạm Thị Trang, Hoàng Hồng Công, Nguyễn Thị Huyền, Dự báo thời tiết ứng dụng mạng nơron nhân tạo và thuật toán Bayes, *Tạp chí Khoa học & Công nghệ*, Số 13, trang 39-43, 2017.
- [12] Nguyễn Tân Ân, Nguyễn Quang Hoan, Hệ dự báo thời tiết với ứng dụng của mạng nơron nhân tạo, *Tạp chí Khoa học & Công nghệ*, 90(02), trang 65-70, 2012.
- [13] Raza K., Jothiprakash V., Multi-output ANN model for prediction of seven meteorological parameters in a weather station, *Journal of The Institution of Engineers (India): Series A*, 95(4), pp. 221-229, 2014.
- [14] Seyam M., Mogheir Y., Application of Artificial Neural Networks model as analytical tool for groundwater salinity, *Journal of Environmental Protection*, 2, pp. 56-71, 2011.
- [15] Tan S.B.K., Shuy E.B., and Chua L.H.C., Modelling hourly and daily open-water evaporation rates in areas with an equatorial climate, *Hydrological Processes*, 21, pp. 486-499, 2007.
- [16] Trần Thị Vân, Hà Dương Xuân Bảo, Đinh Thị Kim Phương, Nguyễn Thị Tuyết Mai và Đặng Thị Mai Nhung, Đặc điểm môi trường nhiệt và diễn biến đảo nhiệt đô thị bề mặt khu vực bắc thành phố Hồ Chí Minh, *Tạp chí Khoa học Trường Đại học Cần Thơ*, 49(A), trang 11-20, 2017.
- [17] Ustaoglu B., Cigizoglu H. K. and Karacaa M., Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods, *Meteorological Applications*, 15, pp. 431- 445, 2008.
- [18] Zhu S., Nyarko E. K. and Hadzima-Nyarko M., Modelling daily water temperature from air temperature for the Missouri River, *PeerJ*, 6:e4894, 2018.

Ngày nhận bài: 17/04/2019

Ngày chấp nhận đăng: 08/06/2019