

MÔ PHỎNG LƯỢNG BỐC HƠI Ở ĐỒNG BẰNG NAM BỘ BẰNG MẠNG NƠRON NHÂN TẠO, HỒI QUY TUYẾN TÍNH ĐA BIẾN KẾT HỢP KỸ THUẬT KHỞI TẠO SỐ LIỆU NGẪU NHIÊN

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

DOIs: <https://doi.org/10.46242/jstiuh.v62i02.4790>

Tóm Tắt. Nghiên cứu này đánh giá tác động của chuỗi số liệu khí tượng khởi tạo lên kết quả thu được khi sử dụng ANN và hồi quy tuyến tính đa biến để mô phỏng lượng bốc hơi tuần tại các trạm khí tượng Cần Thơ và Nhà Bè thuộc đồng bằng Nam Bộ. Bởi chuỗi số liệu khí tượng thực đo cho các yếu tố khí tượng ở cả 2 trạm đa số đều không tuân theo phân bố chuẩn nên chuỗi số liệu thể hiện các kịch bản khác nhau đã được khởi tạo bằng các kỹ thuật Monte Carlo, Latin Hypercube với mức độ chi tiết 5%, 10% và tứ phân vị dựa trên thống kê cụ thể của số liệu thực đo. Kết quả phân tích cho thấy 2 phương pháp hồi quy tuyến tính đa biến và ANN đều cho độ chính xác mô phỏng lượng bốc hơi ở mức cao ($R > 0.93$ hay $R^2 > 0.87$), trong đó cấu trúc ANN với 1 lớp ẩn có 6 nơron sử dụng hàm chuyển tansig là phù hợp để mô phỏng lượng bốc hơi cho cả 2 trạm. Sự biến động về giá trị trung bình và độ lệch chuẩn trong kết quả mô phỏng lượng bốc hơi thay đổi phụ thuộc nhiều hơn vào việc chọn mức chi tiết trong nấc xác suất phân bố của công tác khởi tạo so với việc chọn kỹ thuật khởi tạo. Trong một số trường hợp, kết quả mô phỏng lượng bốc hơi bằng ANN có giá trị nhỏ nhất mang dấu âm trong khi phương pháp hồi quy tuyến tính đa biến lại không có hiện tượng này.

Từ khóa. mô phỏng, bốc hơi, mạng nơron nhân tạo (ANN), hồi quy tuyến tính đa biến, Monte Carlo, Latin Hypercube

1. MỞ ĐẦU

Bốc hơi là một quá trình đóng vai trò to lớn trong vòng tuần hoàn nước tự nhiên, góp phần quan trọng chi phối lượng nước có thể khai thác, sử dụng cho các hoạt động sản xuất và đời sống của con người. Đặc biệt trong bối cảnh biến đổi khí hậu toàn cầu diễn ra ngày càng mạnh mẽ, lượng thất thoát hơi nước từ mặt đất là yếu tố vô cùng quan trọng tác động đến hiện tượng khô hạn và công tác sản xuất nông nghiệp nói riêng. Ở Việt Nam, ảnh hưởng của các kịch bản biến đổi khí hậu công bố năm 2016 đã thể hiện rõ xu thế biến đổi cực đoan của một số yếu tố khí hậu với tác động tiêu cực tiềm năng đến sản xuất và đời sống [1].

Trên thế giới đã có khá nhiều nghiên cứu tập trung vào việc xác định các yếu tố khí tượng quan trọng nhất cần được đưa vào tính toán bốc hơi (hay bốc thoát hơi) và ảnh hưởng của chúng từ việc phân tích số liệu đo đạc tại các hồ nước [2, 3] hay trạm khí tượng [4, 5]. Những nghiên cứu này đề xuất được một số yếu tố quan trọng như tốc độ gió, độ ẩm tương đối, nhiệt độ không khí, bức xạ mặt trời. Một số điều tra còn chỉ ra rằng mức độ quan trọng của mỗi yếu tố khí hậu ảnh hưởng đến tốc độ bốc hơi có thể thay đổi phụ thuộc vào những điều kiện địa lý đặc thù như giữa vùng khí hậu ôn đới và vùng xích đạo [6].

Bên cạnh việc nắm bắt được tầm quan trọng của các yếu tố khí hậu khác nhau ảnh hưởng tới cường độ bốc hơi trong thực tế, việc ước lượng và mô phỏng chính xác lượng bốc hơi luôn là một thách thức không nhỏ. Ngoài các cách tiếp cận tính toán phổ biến hiện nay như phương pháp hồi quy đa biến, mạng nơron nhân tạo (ANN) hiện là một trong những công cụ khoa học tiên tiến thuộc mảng học sâu có tiềm năng để mô phỏng - 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 bốc thoát hơi [4, 5, 7, 8, 9]. Ở Việt Nam, một số tác giả cũng đã nghiên cứu việc sử dụng công nghệ thông tin và kỹ thuật ANN để phục vụ điều tra những đặc trưng khí hậu cho các khu vực khác nhau [10, 11, 12]. Bên cạnh đó, nhiều kết quả nghiên cứu cũng đã chứng minh rằng kỹ thuật sử dụng trong mô phỏng có ảnh hưởng đáng kể đến độ chính xác của kết quả [13, 14].

Các kỹ thuật mô phỏng như Monte Carlo là bộ công cụ phổ biến để giải quyết nhiều bài toán đa chiều mà nhiều khi rất khó có thể được giải quyết thỏa đáng bằng những cách tiếp cận thông thường. Kottegod và cộng sự sử dụng lượng mưa từ một số trạm đo ở nước Ý để phát triển một quy trình cơ bản cho mô phỏng

Monte Carlo và mô phỏng một loạt các biểu đồ mưa thực tế phục vụ việc mở rộng đến một số khu vực không có trạm đo [15]. Ở Việt Nam, mô phỏng Monte Carlo cũng đã được ứng dụng trong một số lĩnh vực như lập kịch bản đề án xây dựng [16] hay quản lý đất nông nghiệp [17]. Trong một nghiên cứu khí tượng thủy văn, Tô Việt Thắng và cộng sự đã sử dụng thành công phương pháp mô phỏng Monte-Carlo để ước lượng dạng phân phối xác suất phổ biến của số liệu dòng chảy đo đạc đến một số hồ chứa, từ đó cho phép tính toán tối ưu việc phân bổ nguồn nước cho hệ thống hồ thuộc lưu vực sông Vu Gia - Thu Bồn [18]. Khi mô phỏng chất lượng không khí tại thành phố Hồ Chí Minh, tác giả Hồ Quốc Bằng đã sử dụng kỹ thuật Monte-Carlo và xác định mức sai số trong kết quả mô phỏng chiếm khoảng 15% giá trị trung bình [19].

Với đà phát triển của cuộc cách mạng công nghiệp 4.0 trên thế giới, các ứng dụng ANN ngày càng nhận được sự quan tâm lớn hơn. Song cũng giống như bất kỳ dạng trí thông minh nhân tạo nào, độ chính xác của kết quả mô phỏng thu được từ loại kỹ thuật này sẽ phụ thuộc đáng kể vào đặc tính số liệu được cung cấp. Để nắm vững mức tin cậy của quá trình mô phỏng hiện tượng, việc đánh giá ảnh hưởng từ đặc tính số liệu sử dụng đến kết quả mô phỏng bốc hơi và mức độ biến thiên trong kết quả mô phỏng bằng ANN là một vấn đề hết sức hữu ích. Tuy nhiên, theo như tác giả tìm hiểu, đây hiện là vấn đề còn chưa được đề cập và nghiên cứu kỹ trong các điều tra về bốc hơi ở nước ta.

Nghiên cứu này có mục tiêu đánh giá tác động của chuỗi số liệu khí tượng đến kết quả thu được từ việc sử dụng ANN và hồi quy tuyến tính đa biến để mô phỏng lượng bốc hơi tại các trạm khí tượng Cần Thơ và Nhà Bè thuộc đồng bằng Nam Bộ. Các yếu tố khí tượng được lựa chọn trong mô phỏng - về mặt lý thuyết - đều có mối quan hệ vật lý với quá trình bốc hơi. Để khảo sát các các kịch bản số liệu khác nhau, quá trình khởi tạo số liệu đã được hoàn thành với kỹ thuật Monte Carlo và Latin Hypercube. Những kết quả thu được đã góp phần làm rõ thêm hiệu quả hoạt động của một số phương pháp mô phỏng trong nghiên cứu lượng bốc hơi, đặc biệt là ước lượng sự biến thiên trong kết quả mô phỏng số liệu khí tượng cho các mục đích khác nhau. Mở rộng ra, kết quả nghiên cứu cũng có thể được bổ sung vào kho tài liệu thực tế để các nhà khoa học tham khảo, cân nhắc thêm trong công tác mô phỏng - dự báo khí hậu phục vụ sản xuất nông nghiệp và các lĩnh vực liên quan.

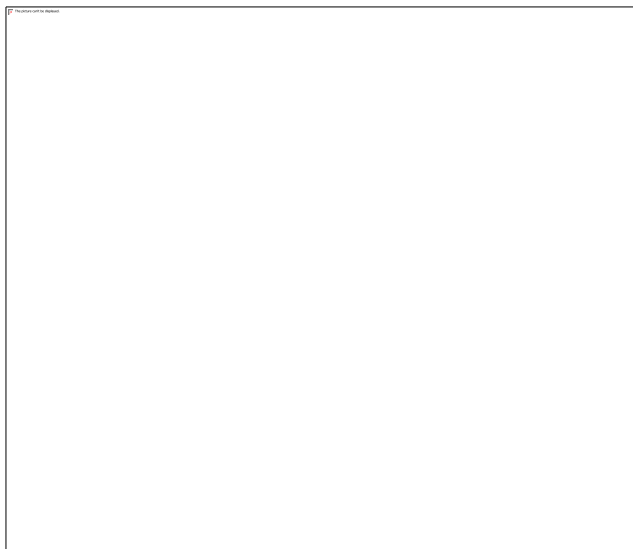
2. SỐ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1 Số liệu

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. 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. Khoảng cách theo đường thẳng giữa 2 trạm xấp xỉ 130 km. Số lượng trạm sử dụng ở nghiên cứu này tuy còn ít nhưng cũng phần nào đại diện cho 2 khu vực địa lý Đông và Tây Nam Bộ nước ta.

Cả 2 trạm đều có nhiệm vụ quan trắc những yếu tố khí hậu cơ bản và các hiện tượng thời tiết. Số liệu được sử dụng trong bài báo này là kết quả quan trắc cho các yếu tố khí hậu của các trạm khí tượng Cần Thơ và Nhà Bè được cung cấp bởi Đài khí tượng Thủy văn khu vực Nam Bộ. Bộ số liệu thực đo có độ dài 4 năm (1/1/2014 đến 31/12/2017) với tổng số 207 tuần đo. Theo mô tả của cơ quan quản lý các trạm, 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.



Hình 1. Vị trí hai trạm khí tượng Nhà Bè và Cần Thơ ở khu vực Nam Bộ (Nguồn: Google map)

Nhằm đảm bảo độ chính xác trong kết quả khi sử dụng công cụ ANN để mô phỏng, việc kiểm tra chất lượng của số liệu đầu vào là rất quan trọng. Công tác kiểm tra số liệu trước khi bắt đầu tính toán đã được thực hiện nhằm loại bỏ những số liệu không hợp lý (như mang giá trị âm) có thể do nhiều nguyên nhân khách quan và chủ quan khác nhau gây ra (như khả năng có vấn đề trực trặc về thiết bị đo thực địa). Những số liệu không hợp lý này nếu không được loại bỏ sẽ có ảnh hưởng làm giảm độ tin cậy trong kết quả vì chúng không thể hiện đúng bản chất mối liên hệ giữa các đại lượng vật lý đang xem xét.

Số liệu thu thập cho các trạm đã được kiểm tra và không thấy xuất hiện số liệu không hợp lý, chỉ có một vài số liệu mất được thay thế bằng trung bình toàn chuỗi số liệu đo trong suốt thời gian còn lại cho cùng thông số. Cho mỗi yếu tố khí hậu, giá trị số liệu tuần là trung bình cộng giá trị đo của các ngày trong tuần cho yếu tố đó. Tổng số tuần số liệu sử dụng đã qua kiểm tra là 207.

2.2 Phương pháp nghiên cứu

Để thực hiện mục tiêu, nghiên cứu đã được tiến hành theo các bước sau:

Bước 1: Đánh giá hiệu quả mô phỏng dùng các phương pháp ANN và hồi quy tuyến tính đa biến cho bộ dữ liệu thực đo.

Tác giả đã so sánh hiệu quả mô phỏng của phương pháp cổ điển (hồi quy tuyến tính đa biến) và kỹ thuật mới (ANN) thông qua mức chính xác (độ lệch) giữa kết quả mô phỏng bốc hơi thu được và số liệu bốc hơi thực đo. Các biến đầu vào phục vụ cho mô phỏng là độ ẩm tương đối không khí (%), nhiệt độ không khí ($^{\circ}\text{C}$), tốc độ gió (m/s), lượng mưa (mm) và biến đầu ra là lượng bốc hơi (mm).

Đầu tiên, phương pháp hồi quy tuyến tính đa biến được áp dụng nhằm tìm ra mối tương quan tuyến tính giữa bộ các biến đầu vào và biến đầu ra cho số liệu đo giai đoạn 2014 - 2017 đã mô tả trên. Trong đó, hệ số VIF (Variable Inflation Factors) được đánh giá để ước lượng hiện tượng đa cộng tuyến nhằm loại bỏ hiện tượng hai hoặc nhiều biến độc lập có tương quan cao với nhau trong mô hình hồi quy.

Kế tiếp, tác giả chạy mô phỏng ANN trên công cụ Neural Network Toolbox của phần mềm Matlab (MathWorks, Inc.) sử dụng mạng nơron 1 đến 2 lớp ẩn với từng cấu trúc có số lượng nơron trong mỗi lớp ẩn của hàm chuyển tansig khác nhau. Mạng nơron được huấn luyện sử dụng số liệu đầu vào thực đo (tuần) cho 2 trạm khí tượng giai đoạn 2014 - 2017. Các số liệu đã được phần mềm Matlab lựa chọn ngẫu nhiên (mặc định) 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. Trong nghiên cứu này, các cấu trúc ANN đều sử dụng thuật toán huấn luyện Levenberg-Marquardt backpropagation (lan truyền ngược) 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”. Mạng đã qua huấn luyện sẽ được dùng để ước tính lượng bốc hơi sử dụng chính dữ liệu thực đo trong bước học. Giá trị hệ số tương quan R (mức liên hệ giữa lượng bốc hơi thực đo và mô phỏng) thu được từ các cấu trúc ANN khác nhau được so sánh để xác định các mốc thay đổi trong độ chính xác phụ thuộc vào độ phức tạp của ANN sử dụng.

Độ chính xác trong mô phỏng hiện tượng của các cấu trúc ANN khác nhau và hồi quy tuyến tính đa biến được đánh giá chủ yếu bằng hệ số tương quan (R) giữa kết quả mô phỏng và thực đo.

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

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.

Bước 2: Đánh giá dạng phân bố chi phối các yếu tố khí hậu và khởi tạo bộ số liệu mang đặc tính phân bố của chuỗi số liệu ban đầu.

Trong khá nhiều trường hợp, về mặt lý thuyết các yếu tố khí hậu như nhiệt độ thường được giả thiết tuân theo quy luật phân bố chuẩn hay một vài dạng phân bố phổ biến khác. Ở nghiên cứu này, để ước lượng dạng phân bố phù hợp nhất cho mỗi yếu tố khí hậu tại 2 trạm Nhà Bè và Cần Thơ trên thực tế, tác giả đã sử dụng phần mềm Minitab (Minitab, LLC) để khảo sát chuỗi số liệu thực đo tuần trong 4 năm. Phân bố được khảo sát bao gồm 14 dạng phổ biến sau: Chuẩn (Normal), Chuyển đổi Box-Cox (Box-Cox Transformation), Log chuẩn (Lognormal), Log chuẩn 3 tham số (3-Parameter Lognormal), Mũ (Exponential), Mũ 2 tham số (2-Parameter Exponential), Weibull, Weibull 3 tham số (3-Parameter Weibull), Giá trị cực tiểu (Smallest Extreme Value), Giá trị cực đại (Largest Extreme Value), Gamma, Gamma 3 tham số (3-Parameter Gamma), Logistic, Loglogistic, Loglogistic 3 tham số (3-Parameter Loglogistic), Chuyển đổi Johnson (Johnson Transformation).

Bên cạnh số liệu khí tượng đo đạc, để đánh giá những kịch bản khí hậu có thể xảy ra, số liệu còn được khởi tạo bằng kỹ thuật khởi tạo ngẫu nhiên là Monte Carlo và Latin Hypercube trên phần mềm Crystal Ball (Oracle, Inc.) căn cứ vào đặc điểm phân bố của mỗi yếu tố khí tượng sử dụng các xác suất phân bố với độ chi tiết khác nhau dựa trên phân bố thu được từ áp dụng phần mềm SPSS (IBM, Inc.). Điểm khác biệt của phương pháp Latin Hypercube là phương pháp lấy mẫu này chia phân phối xác suất giả định thành các khoảng xác suất bằng nhau dẫn đến toàn bộ phạm vi phân phối được lấy mẫu đồng đều hơn so với phương pháp lấy mẫu Monte Carlo thông thường [20].

Đặc điểm phân bố (%) của chuỗi 207 số liệu tuần thực đo gốc theo khoảng giá trị của chuỗi số liệu thực đo được khảo sát trên phần mềm SPSS với các xác suất phân bố 5%, 10%, tứ phân vị (quartile). Kế đó, số liệu được khởi tạo bằng 2 kỹ thuật khởi tạo ngẫu nhiên là Monte Carlo và Latin Hypercube trên phần mềm Crystal Ball dựa trên đặc điểm phân bố vừa tìm được của mỗi yếu tố khí tượng. Số lượng giá trị khởi tạo là 50,000 điểm (tuần) cho mỗi trạm khí tượng với mỗi yếu tố. Với số lượng điểm khởi tạo khá lớn này, các phương án biến đổi trong tổ hợp giá trị các biến đầu vào được tính đến ở mức chi tiết và đa dạng hơn.

Bước 3: Đánh giá sự biến động trong chuỗi kết quả mô phỏng khi sử dụng các loại bộ số liệu đầu vào thực đo và khởi tạo.

Mạng neuron đã qua huấn luyện và công thức hồi quy tuyến tính tìm ra ở bước trên được dùng để tính giá trị biến phụ thuộc với dữ liệu các biến độc lập đã khởi tạo bằng những thuật toán Monte Carlo và Latin Hypercube. Mức biến động về giá trị trong kết quả tính từ các phương pháp ANN và hồi quy tuyến tính đa biến được đánh giá cho các kịch bản sử dụng các loại bộ số liệu đầu vào thực đo và khởi tạo khác nhau.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1 Hiệu quả mô phỏng dùng hồi quy tuyến tính đa biến và ANN cho bộ dữ liệu thực đo

3.1.1 Hồi quy tuyến tính đa biến

Kết quả phân tích hồi quy tuyến tính đa biến sử dụng phần mềm SPSS cho mô phỏng lượng bốc hơi (tuần) tại hai trạm khí tượng được thể hiện trong bảng 1. Các yếu tố khí tượng đầu vào qua kiểm tra đều cho thấy có mức tương quan lẫn nhau ở mức cho phép ($VIF < 10$) nên không yếu tố nào bị loại ra khỏi mô hình. Độ chính xác của mô hình hồi quy tuyến tính đa biến thu được cho cả hai trạm đều ở mức cao (trạm Nhà Bè có $R = 0.952$ hay $R^2 = 0.906$, trạm Cần Thơ có $R = 0.935$ hay $R^2 = 0.875$).

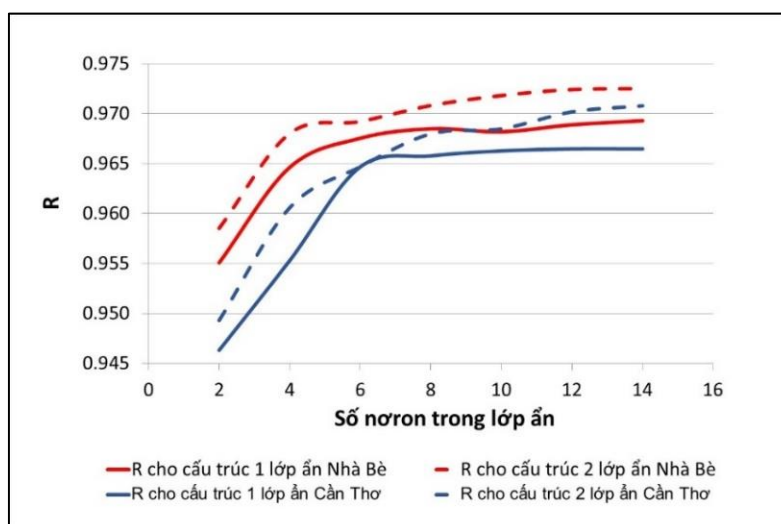
Bảng 1. Kết quả phân tích hồi quy tuyến tính đa biến cho mô phỏng lượng bốc hơi

MÔ PHỎNG LƯỢNG BỐC HƠI...

Trạm	Yếu tố	Hệ số hồi quy không chuẩn hóa	Mức ý nghĩa (Sig.) của hệ số	R/R ² của mô hình	Mức ý nghĩa (Sig.) của mô hình	VIF
Nhà Bè	Độ ẩm lớn nhất	-0.425	0	0.952/0.906	< 0.001	3.307
	Độ ẩm nhỏ nhất	-0.335	0			6.576
	Tốc độ gió trung bình	4.817	0			1.483
	Nhiệt độ lớn nhất	1.279	0			6.84
	Nhiệt độ nhỏ nhất	-0.608	0.019			6.911
	Tổng lượng mưa	-0.015	0			1.549
	Hằng số	47.247	0.000			-
Cần Thơ	Độ ẩm lớn nhất	-0.525	0	0.935/0.875	< 0.001	2.273
	Độ ẩm nhỏ nhất	-0.3	0			6.833
	Tốc độ gió trung bình	3.966	0			1.373
	Nhiệt độ lớn nhất	0.939	0			7.234
	Nhiệt độ nhỏ nhất	-0.541	0.012			5.509
	Tổng lượng mưa	-0.009	0.055			2.248
	Hằng số	63.926	0.000			-

3.1.2 ANN

Tác giả đã chạy huấn luyện ANN trên phần mềm Matlab sử dụng mạng nơron 1 đến 2 lớp ẩn với từng cấu trúc 2, 4, 6, 8, 10, 12 và 14 nơron trong mỗi lớp ẩn của hàm chuyển tansig. Đầu tiên, mạng nơron được huấn luyện sử dụng số liệu đầu vào thực đo (tuần) cho 2 trạm khí tượng giai đoạn 2014 - 2017. Mỗi mẫu số liệu gồm thông tin của 6 yếu tố đầu vào giống như ở phương pháp hồi quy tuyến tính đa biến (xem bảng 2) với 1 yếu tố đầu ra là lượng bốc hơi ngày.



Hình 2. Kết quả mô phỏng lượng bốc hơi (tuần) sử dụng các cấu trúc ANN khác nhau

Việc mô phỏng được tiến hành 1000 lần cho mỗi tổ hợp cấu trúc ANN trong đó số liệu đầu vào được phần mềm lựa chọn ngẫu nhiên theo tỷ lệ 70/15/15 tương ứng cho 3 bước huấn luyện (training), kiểm định (validation) và kiểm tra với số liệu độc lập (test). Cho từng cấu trúc ANN, lần chạy cho giá trị lớn nhất của hệ số tương quan R giữa giá trị mô phỏng và thực đo (ở đây R là R_{all} gộp chung cho cả 3 bước nói trên) được ghi nhận. Mạng qua huấn luyện được dùng để ước tính lượng bốc hơi cho những loại dữ liệu khởi tạo ở bước kế tiếp.

Kết quả mô phỏng lượng bốc hơi bằng ANN cho cả hai trạm đều đạt mức chính xác cao (R_{all} > 0.94). Nhìn chung, cấu trúc ANN với 1 lớp ẩn có 6 nơron sử dụng hàm chuyển tansig là phù hợp (R_{all} ≥ 0.965) để mô phỏng lượng bốc hơi (tuần) tại cả 2 trạm Nhà Bè và Cần Thơ vì khi tăng thêm số lớp ẩn và/hoặc số nơron trong mỗi lớp ẩn lên cao hơn nữa thì hiệu suất mô phỏng (R_{all}) được cải thiện không đáng kể. Điều này có thể quan sát được từ độ dốc các đường đồ thị giá trị R trong hình 2. Nhìn chung, độ chính xác trong mô phỏng thể hiện qua giá trị R của ANN đều cao hơn so với giá trị R của hồi quy tuyến tính đa biến ở cả hai trạm khí tượng đưa vào nghiên cứu.

3.2 Khởi tạo số liệu bằng kỹ thuật khởi tạo ngẫu nhiên

Tổng số ngày số liệu thực đo hiện trường chỉ giới hạn trong 4 năm (1462 ngày hay 207 tuần). Do đó, để đánh giá mức biến động trong kết quả mô phỏng dưới nhiều kịch bản khác nhau trong thực tế có thể xảy ra, việc khởi tạo số liệu đầu vào để tạo chuỗi dài hơn nhiều lần là rất hữu ích và hiệu quả. Điều này còn trở nên đặc biệt quan trọng hơn trong bối cảnh biến đổi khí hậu hiện nay khi ranh giới chuyển mùa thường xuyên biến đổi với nhiều yếu tố khí hậu.

Sự hiểu biết về dạng phân bố số liệu của các yếu tố khí hậu là hết sức quan trọng khi khởi tạo bộ số liệu ngẫu nhiên (kịch bản có thể xảy ra) mới mang những đặc trưng biến đổi cơ bản của bộ số liệu ban đầu. Để phục vụ công tác này, việc đánh giá dạng phân bố chi phối các yếu tố khí hậu sử dụng số liệu đo tuần được thực hiện với phần mềm Minitab cho những dạng phân bố phổ biến nhất đã mô tả trên. Các yếu tố đưa vào khảo sát bao gồm: Độ ẩm tương đối không khí (lớn nhất, nhỏ nhất), Tổng lượng bốc hơi, Tốc độ gió trung bình, Nhiệt độ không khí (lớn nhất, nhỏ nhất), Tổng lượng mưa. Kết quả thu được thể hiện trong bảng 2.

Bảng 2. Kết quả đánh giá dạng phân bố cho các yếu tố khí hậu (tuần) tại hai trạm khí tượng Cần Thơ và Nhà Bè

Tên yếu tố	Trạm khí tượng	Phân bố phù hợp nhất	AD	p	Thông số của phân bố
Độ ẩm lớn nhất	Nhà Bè	Johnson Transformation	0.208	0.865	Location = -0.03595; Scale = 0.96221
	Cần Thơ	Johnson Transformation Weibull Smallest Extreme Value	0.228 0.412 0.408	>0.250 >0.250	Location = 0.00382; Scale = 1.01134 Shape = 39.30899; Scale = 93.24503 Location = 93.27730; Scale = 2.35266
Độ ẩm nhỏ nhất	Nhà Bè	Johnson Transformation	0.313	0.546	Location = 0.05633; Scale = 1.10319
Tổng lượng bốc hơi	Nhà Bè	Johnson Transformation 3-Parameter Weibull	0.344 2.103	0.484 0.102	Location = -0.01289; Scale = 0.98340 Shape = 1.98703; Scale = 13.97951; Threshold = 7.94566
	Cần Thơ	3-Parameter Weibull Box-Cox Transformation Normal	0.262 0.406 0.446	>0.500 0.348 0.280	Shape = 2.88330; Scale = 13.48495; Threshold = 7.10455 Location = 4.34082; Scale = 0.52724 Location = 19.11932; Scale = 4.55940
Tốc độ gió trung bình	Nhà Bè	Lognormal Johnson Transformation Box-Cox Transformation	0.372 0.372 0.372	0.417 0.417 0.417	Location = 0.11395; Scale = 0.33403 Location = 0.00195; Scale = 0.94643 Location = 0.11395; Scale = 0.33403
	Cần Thơ	Johnson Transformation Box-Cox Transformation Lognormal	0.221 0.319 0.319	0.832 0.533 0.533	Location = 0.01846; Scale = 1.00405 Location = 0.31483; Scale = 0.25003 Location = 0.31483; Scale = 0.25003
Nhiệt độ lớn nhất	Nhà Bè	Lognormal	0.421	0.321	Location = 3.49576; Scale = 0.04172
		Box-Cox Transformation	0.421	0.321	Location = 3.49576; Scale = 0.04172
		Gamma	0.451	>0.250	Shape = 577.41404; Scale = 0.05716
		Loglogistic	0.355	>0.250	Location = 3.49538; Scale = 0.02366
Cần Thơ	Box-Cox Transformation	0.282	0.634	Location = 34412.47141; Scale = 4560.07916	
	Normal	0.397	0.366	Location = 32.46225; Scale = 1.45309	
	Logistic	0.190	>0.250	Location = 32.49154; Scale = 0.81241	
	Loglogistic	0.277	>0.250	Location = 3.48057; Scale = 0.02511	
Nhiệt độ nhỏ nhất	Nhà Bè	Johnson Transformation	0.269	0.678	Location = 0.02962; Scale = 1.06938
	Cần Thơ	Johnson Transformation	0.277	0.650	Location = -0.04024; Scale = 0.97687

*Chú giải: bảng chỉ thể hiện các yếu tố khí hậu tuân theo một hay nhiều dạng phân bố (trong bảng nêu thông tin về 3 phân bố có p cao nhất tương ứng với yếu tố đó) còn các yếu tố khí hậu còn lại không tuân theo dạng phân bố nào đạt mức có ý nghĩa về mặt thống kê.

Nhìn chung, căn cứ vào thông kê AD (Anderson-Darling) và giá trị p cho kiểm định mức độ phù hợp (goodness-of-fit test), những dạng phân bố xuất hiện nhiều nhất trong việc chi phối các yếu tố khí hậu là Johnson Transformation, Box-Cox Transformation, Weibull và 3-Parameter Weibull. Phân bố chuẩn có xuất hiện nhưng chỉ ở cấp độ phù hợp thấp hơn (p thấp hơn) so với các phân bố khác trong tất các trường hợp. Riêng tổng lượng mưa tại cả 2 trạm lại không tuân theo bất kỳ phân bố nào trong số 14 dạng đã đưa vào xem xét. Bởi kết quả cho thấy số liệu khí tượng của rất nhiều yếu tố khí hậu tại 2 trạm thường không

có chung một dạng phân bố đặc thù nhất nên tác giả đã không sử dụng một (hay vài) hình thức phân bố chung nào cho khởi tạo bộ số liệu kịch bản ở bước sau. Thay vào đó, mỗi yếu tố khí hậu đã được khảo sát cụ thể để tìm ra phân bố đặc thù cho riêng cho yếu tố đó nhằm phục vụ cho công tác khởi tạo số liệu.

3.3 Đánh giá sự biến động trong chuỗi kết quả mô phỏng khi sử dụng các loại bộ số liệu đầu vào thực đo và khởi tạo

Kết quả mô phỏng dùng mạng neuron nhân tạo sau khi xét đến những giá trị đặc trưng cơ bản của mỗi tập kết quả mô phỏng như giá trị trung bình, độ lệch chuẩn, giá trị lớn nhất (max), giá trị nhỏ nhất (min) cho phép ta đi đến một số nhận định sau:

- Giá trị trung bình trong kết quả mô phỏng sử dụng các bộ số liệu đầu vào khác nhau không thể hiện rõ quy luật biến đổi khi xét chung cho cả hai trạm (đầu độ chênh lệch lớn nhất thường xuất hiện khi sử dụng bộ số liệu khởi tạo Tứ phân vị), song độ lệch chuẩn trong kết quả mô phỏng bốc hơi sử dụng các bộ số liệu khởi tạo và số liệu gốc làm đầu vào thường thấp hơn so với độ lệch chuẩn thực trong số liệu bốc hơi đo đạc hiện trường (gốc). Khi các xác suất phân bố với độ chi tiết tăng dần (5%, 10%, Tứ phân vị tức 25%) - ở cả 2 phương pháp Monte Carlo và Latin Hypercube - kết quả mô phỏng lượng bốc hơi đều có xu hướng có giá trị trung bình cũng như độ lệch chuẩn tăng dần. Những khác biệt trên thể hiện sự ảnh hưởng đến kết quả mô phỏng từ những bộ số liệu đầu vào khác nhau (tức những tình huống có thể xảy ra thu được nhờ khởi tạo nhằm tăng số lượng số liệu khảo sát lên).

- Giá trị lớn nhất trong kết quả mô phỏng sử dụng các bộ số liệu đầu vào khác nhau không thể hiện rõ quy luật biến đổi khi xét chung cho cả hai trạm nhưng cả 2 phương pháp Monte Carlo và Latin Hypercube đều có xu hướng cho kết quả mô phỏng lượng bốc hơi với giá trị lớn nhất tăng dần khi sử dụng dữ liệu khởi tạo có các xác suất phân bố với độ chi tiết tăng dần (5%, 10%, Tứ phân vị) theo nấc xác suất chia nhỏ cho phép trong phần mềm Crystal Ball.

- Giá trị nhỏ nhất trong kết quả mô phỏng lượng bốc hơi có xu hướng thấp hơn khi sử dụng các bộ số liệu khởi tạo so với khi sử dụng số liệu thực đo gốc. Ngoài ra, kết quả mô phỏng lượng bốc hơi với giá trị nhỏ nhất thường giảm dần khi sử dụng dữ liệu khởi tạo có các xác suất phân bố với độ chi tiết tăng dần.

- Kết quả mô phỏng lượng bốc hơi theo tuần có giá trị nhỏ nhất (min) mang giá trị âm (là giá trị lượng bốc hơi không xảy ra trên thực tế) - dù giá trị tất cả số liệu khởi tạo đều không âm - đã xuất hiện ngay cả khi giá trị R_{all} của quá trình học khá cao nhờ sử dụng các cấu trúc ANN phức tạp hơn (nhiều neuron trong lớp ẩn hơn). Trong khi đó, một số cấu trúc ANN đơn giản hơn lại cho kết quả không có giá trị âm của lượng bốc hơi nhỏ nhất với cùng bộ số liệu đầu vào.

Công tác kiểm định sự khác biệt cho các giá trị trung bình, độ lệch chuẩn, giá trị lớn nhất (max), giá trị nhỏ nhất (min) trong kết quả mô phỏng bốc hơi cho các loại bộ số liệu đầu vào gốc và khởi tạo khác nhau được tiến hành sử dụng kiểm định Independent-Samples Kruskal-Wallis trên phần mềm SPSS. Mỗi mẫu đưa vào so sánh là các giá trị mô phỏng cho chỉ tiêu quan tâm sử dụng tất cả các cấu trúc ANN với 1÷2 lớp ẩn và 2÷14 neuron trong mỗi lớp ẩn. Kết quả tính toán được tổng kết lại như mô tả dưới đây.

- Giá trị trung bình và độ lệch chuẩn của những giá trị bốc hơi mô phỏng:

- + Trạm Nhà Bè: giá trị trung bình của những chuỗi kết quả mô phỏng bốc hơi thu được không khác nhau đáng kể về mặt thống kê. Tuy nhiên độ lệch chuẩn của chuỗi giá trị mô phỏng với các nấc xác suất phân bố 5% và 10% bằng cả 2 kỹ thuật khởi tạo ngẫu nhiên (Monte Carlo và Latin Hypercube) đều khác biệt đáng kể so với độ lệch chuẩn của chuỗi số liệu thực đo ban đầu. Đầu độ lệch chuẩn của chuỗi kết quả mô phỏng sử dụng số liệu khởi tạo với nấc xác suất phân bố tứ phân vị bằng cả 2 kỹ thuật khởi tạo không khác so với độ lệch chuẩn của chuỗi mô phỏng sử dụng nấc xác suất phân bố 10% nhưng lại khác đáng kể với nấc xác suất phân bố 5%. Điều này cho thấy nấc chi tiết 5% có dấu hiệu khác cơ bản với tứ phân vị về tác động lên chuỗi số liệu khởi tạo đối với cả 2 kỹ thuật khởi tạo đang nghiên cứu ở trạm này. Sự khác biệt trong trung bình và độ lệch chuẩn thể hiện rõ hơn khi thay đổi độ lớn của cấp khởi tạo số liệu (như 5% qua 10% hay tứ phân vị) song lại không đáng kể khi thay đổi phương pháp khởi tạo (chuyển từ sử dụng phương pháp Monte Carlo qua Latin Hypercube).

- + Trạm Cần Thơ: giá trị trung bình của những chuỗi kết quả mô phỏng bốc hơi từ các chuỗi khởi tạo đều khác ở mức có ý nghĩa về mặt thống kê với trung bình số liệu bốc hơi thực đo gốc ban đầu và kết quả mô phỏng ANN chạy trực tiếp trên chuỗi các yếu tố khí tượng gốc. Tuy nhiên, kết quả mô phỏng bốc hơi bằng ANN sử dụng chuỗi số liệu khí tượng thực đo lại không khác với chuỗi bốc hơi gốc thực đo. Đa số giá trị trung bình của những giá trị mô phỏng từ các chuỗi khởi tạo với các nấc xác suất phân bố 5% và 10% đều giống nhau nhưng lại khác về mặt thống kê với trung bình của những giá trị mô phỏng từ các chuỗi khởi

tạo với nấc xác suất phân bố tứ phân vị. Độ lệch chuẩn của chuỗi giá trị mô phỏng với nấc xác suất phân bố 5% bằng cả 2 kỹ thuật khởi tạo ngẫu nhiên (Monte Carlo và Latin Hypercube) đều khác biệt đáng kể so với độ lệch chuẩn của kết quả mô phỏng sử dụng các số liệu thực đo ban đầu và số liệu tứ phân vị. Dẫu các độ lệch chuẩn của chuỗi mô phỏng sử dụng số liệu khởi tạo với nấc xác suất phân bố tứ phân vị bằng cả 2 kỹ thuật không khác so với độ lệch chuẩn của chuỗi mô phỏng sử dụng nấc xác suất phân bố 10% nhưng lại khác đáng kể với nấc xác suất phân bố 5%. Điều này cho thấy việc chọn mức chi tiết 5% có dấu hiệu khác cơ bản với tứ phân vị về mức độ tác động lên chuỗi số liệu khởi tạo đối với cả 2 kỹ thuật khởi tạo đang nghiên cứu cho bộ số liệu ở trạm Cần Thơ. Sự khác biệt trong trung bình và độ lệch chuẩn thể hiện rõ hơn khi chuyển giữa các mức chính xác của khởi tạo số liệu (như 5% qua 10% hay tứ phân vị) so với khi chuyển phương pháp khởi tạo (chuyển từ sử dụng phương pháp Monte Carlo qua Latin Hypercube).

- Giá trị lớn nhất và nhỏ nhất của những giá trị bốc hơi mô phỏng:

+ Trạm Nhà Bè: đa số giá trị lớn nhất và nhỏ nhất của những giá trị bốc hơi mô phỏng từ các chuỗi khởi tạo đều khác ở mức có ý nghĩa về mặt thống kê với giá trị lớn nhất và nhỏ nhất tương ứng của số liệu thực đo gốc ban đầu và kết quả mô phỏng chạy trực tiếp trên chuỗi các yếu tố khí tượng gốc. Tuy nhiên, giá trị lớn nhất và nhỏ nhất của kết quả mô phỏng ANN sử dụng chuỗi số liệu các yếu tố khí tượng gốc lại không khác với chuỗi số liệu bốc hơi thực đo gốc. Do đặc điểm khởi tạo nên tất cả các giá trị lớn nhất và nhỏ nhất của những giá trị mô phỏng từ các chuỗi khởi tạo với các nấc xác suất phân bố 5%, 10% và tứ phân vị đương nhiên sẽ đều không có sự khác biệt về mặt thống kê.

+ Trạm Cần Thơ: đa số giá trị lớn nhất và nhỏ nhất của những giá trị mô phỏng từ các chuỗi khởi tạo đều khác ở mức có ý nghĩa về mặt thống kê với giá trị lớn nhất và nhỏ nhất tương ứng của số liệu bốc hơi đo gốc ban đầu và kết quả mô phỏng chạy trực tiếp chuỗi số liệu các yếu tố khí tượng gốc. Tuy vậy, giá trị lớn nhất và nhỏ nhất của kết quả mô phỏng ANN sử dụng chuỗi số liệu các yếu tố khí tượng đo gốc lại không khác với chuỗi bốc hơi đo gốc. Do đặc điểm khởi tạo nên tất cả các giá trị lớn nhất và nhỏ nhất trung bình của những giá trị mô phỏng từ các chuỗi khởi tạo với các nấc xác suất phân bố 5%, 10% và tứ phân vị đương nhiên sẽ đều không có sự khác biệt ở mức có ý nghĩa.

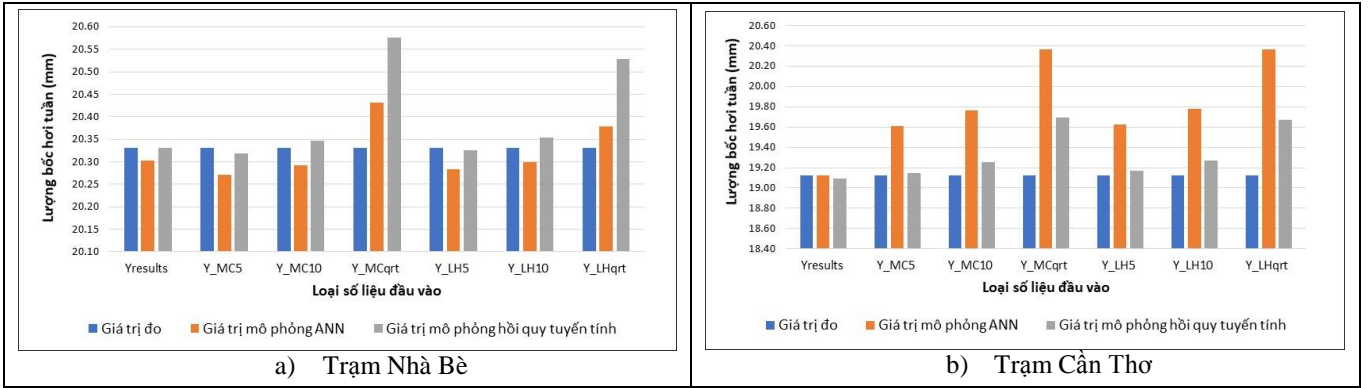
Kết quả mô phỏng lượng bốc hơi dùng phương pháp hồi quy tuyến tính đa biến với các loại chuỗi số liệu khác nhau (bảng 3) cho thấy xu thế hội tụ quanh giá trị trung bình trong kết quả mô phỏng với đa số bộ số liệu khởi tạo cao hơn so với khi sử dụng số liệu các yếu tố khí tượng gốc (độ lệch chuẩn nhỏ hơn trong 10/12 trường hợp). Dù giá trị trung bình trong kết quả mô phỏng bốc hơi không khác nhau nhiều nhưng kết quả thu được khi sử dụng bộ số liệu khởi tạo vẫn thường cao hơn so với kết quả mô phỏng sử dụng số liệu các yếu tố khí tượng gốc. Một điểm đặc biệt khác với kết quả bốc hơi mô phỏng sử dụng ANN là giá trị kết quả nhỏ nhất từ phương pháp hồi quy tuyến tính đa biến không thấy xuất hiện giá trị lượng bốc hơi âm.

Bảng 3. Kết quả mô phỏng lượng bốc hơi dùng hồi quy tuyến tính đa biến với các loại chuỗi số liệu khởi tạo

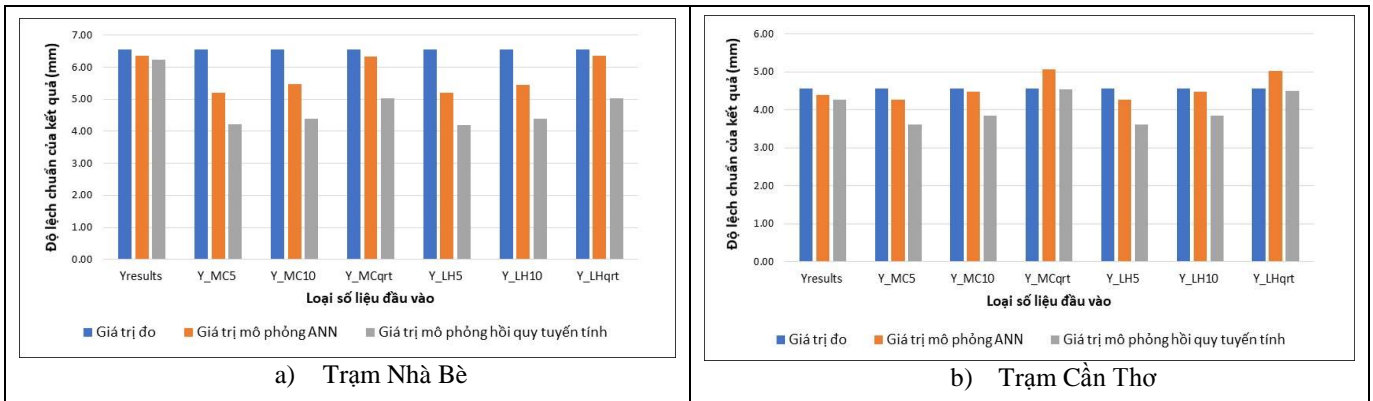
Trạm khí tượng	Biến động trong kết quả mô phỏng lượng bốc hơi (giá trị trung bình / độ lệch chuẩn / max / min) theo loại số liệu đầu vào						
	Số liệu thực đo	Số liệu khởi tạo Monte Carlo			Số liệu khởi tạo Latin Hypercube		
		5%	10%	Tứ phân vị	5%	10%	Tứ phân vị
Nhà Bè	20.33 /	20.32 /	20.35 /	20.58 /	20.33 /	20.35 /	20.53 /
	6.23 /	4.21 /	4.39 /	5.02 /	4.20 /	4.39 /	5.03 /
	36.05 /	37.89 /	38.04 /	39.09 /	38.32 /	39.02 /	40.95 /
	8.82	5.36	3.91	2.82	5.08	3.46	2.80
Cần Thơ	19.10 /	19.15 /	19.26 /	19.70 /	19.17 /	19.27 /	19.67 /
	4.27 /	3.60 /	3.84 /	4.54 /	3.61 /	3.84 /	4.51 /
	28.30 /	35.79 /	34.48 /	38.77 /	38.36 /	37.02 /	38.13 /
	6.59	5.55	5.13	4.54	6.15	5.82	4.33

*Chú giải: “Thực đo” là kết quả cho 207 tuần số liệu thực đo; “5%”, “10%”, “Tứ phân vị” là các kết quả cho 50.000 tuần số liệu khởi tạo.

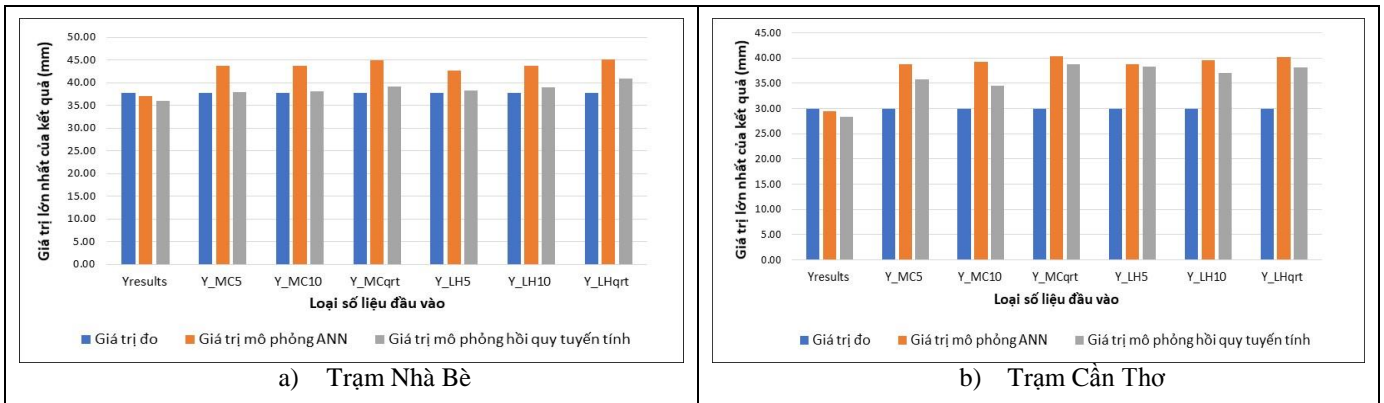
MÔ PHỎNG LƯỢNG BỐC HƠI...



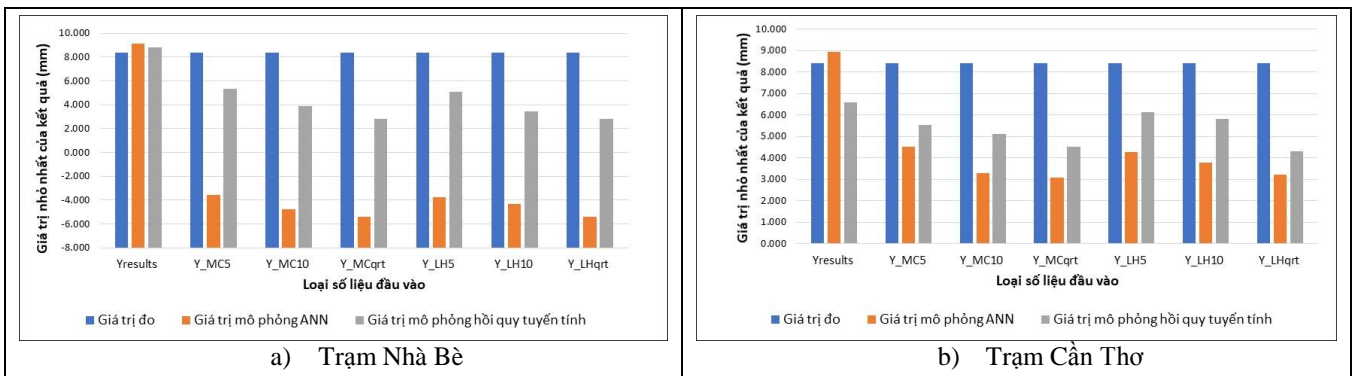
Hình 3. Sự biến đổi trong giá trị trung bình của kết quả mô phỏng bốc hơi sử dụng các bộ số liệu đầu vào khác nhau



Hình 4. Sự biến đổi trong độ lệch chuẩn của kết quả mô phỏng bốc hơi sử dụng các bộ số liệu đầu vào khác nhau



Hình 5. Sự biến đổi trong giá trị lớn nhất của kết quả mô phỏng bốc hơi sử dụng các bộ số liệu đầu vào khác nhau



Hình 6. Sự biến đổi trong giá trị nhỏ nhất của kết quả mô phỏng bốc hơi sử dụng các bộ số liệu đầu vào khác nhau

3.4 Thảo luận

ANN là phương pháp có rất nhiều ưu điểm vượt trội đã được chứng minh trong thực tiễn. Trong một số nghiên cứu đã được công bố [21, 22, 23, 24], kết quả mô phỏng hồi quy tuyến tính đa biến thường có độ chính xác thể hiện qua giá trị hệ số tương quan R thấp hơn so với khi áp dụng ANN. Tuy nhiên ở nghiên cứu này, sự khác biệt về độ chính xác trong mô phỏng giữa 2 phương pháp lại khá nhỏ ($R = 0.952$ hay $R^2 = 0.906$ ở ANN so với $R = 0.935$ hay $R^2 = 0.875$ ở hồi quy đa biến). Điều này xảy ra có khả năng do lượng thông tin giữa các biến đầu vào và biến đầu ra thuận tiện để thể hiện bằng hàm tuyến tính đơn giản, hay nói khác đi khả năng mô phỏng tốt các hàm phi tuyến của ANN có thể đã không có nhiều tác dụng để tạo sự khác biệt với bộ thông số mang đặc trưng vừa nêu. Giá trị hệ số tương quan khá giống nhau không bộc lộ đầy đủ ưu điểm về độ chính xác của phương pháp sử dụng trong mô phỏng vì kết quả đã bị chi phối khá nhiều bởi độ phức tạp của mối quan hệ giữa các biến. Đây có lẽ là điểm nên được cân nhắc điều tra kỹ thêm trong các nghiên cứu tương lai khi đánh giá hiệu quả của việc sử dụng ANN ở bối cảnh cụ thể, bên cạnh một số hạn chế hiện còn tồn tại của phương pháp này như không (hay bị giới hạn) tính năng cung cấp mối liên hệ (hàm số) dưới dạng công thức liên hệ tường minh hay giải thích quan hệ giữa các thông số, vấn đề quá ngưỡng [23] so với các phương pháp phổ biến hiện nay.

Qua điều tra, cấu trúc ANN với 1 lớp ẩn có 6 nơron sử dụng hàm chuyển tansig là phù hợp để mô phỏng lượng bốc hơi cho 2 trạm đưa vào khảo sát. Theo tác giả được biết, số lớp ẩn cùng số lượng nơron trong lớp ẩn cũng như hàm chuyển tối ưu để đạt hiệu suất mô phỏng cao, tránh hiện tượng quá ngưỡng vẫn luôn là mối quan tâm hàng đầu của các nhà khoa học khi sử dụng ANN để mô phỏng các hiện tượng tự nhiên. Kết quả 6 nơron trong lớp ẩn thu được là số lượng nơron ít hơn kết quả tương tự của Abudu và cộng sự [5], Abhishek và cộng sự [14], nhưng vẫn cao hơn của Nezhad và cộng sự [25]. Tuy nhiên, so sánh trên chỉ là tương đối bởi tính không đồng nhất ngay trong số liệu đầu vào và khả năng các điều kiện biên tác động tới quá trình học và mô phỏng.

Một kết quả quan trọng khác thu được từ nghiên cứu này là dẫn rằng có hệ số tương quan giữa kết quả thực đo và mô phỏng cao nhưng ANN lại cho kết quả mô phỏng lượng bốc hơi thấp nhất mang giá trị âm. Đó là một điều không thể xảy ra trong thực tế. Hơn nữa, hai trạm khí tượng Cần Thơ và Nhà Bè cách xa nhau hơn 130 km (theo đường thẳng) có thể đại diện cho hai vùng khác nhau về mặt không gian song kết quả mô phỏng dùng ANN của chúng cũng đều thấy xuất hiện kết quả mô phỏng âm. Tuy nhiên, điều này đã không xảy ra với kết quả mô phỏng dùng hồi quy tuyến tính đa biến với cùng các bộ số liệu khởi tạo. Trên phương diện ứng dụng, sự hợp lý của kết quả về mặt ý nghĩa vật lý là rất quan trọng và đây là điểm cần được quan tâm nghiên cứu thêm khi sử dụng ANN để mô phỏng hay dự báo số liệu khí hậu dùng số liệu khởi tạo.

Giả thiết số liệu tuân theo phân bố chuẩn hay được áp dụng trong thực tế phân tích số liệu không phù hợp với chuỗi số liệu thực đo tuần ở 2 trạm khí tượng Nhà Bè và Cần Thơ. Kết quả phân tích cho thấy phần lớn chuỗi số liệu của các yếu tố khí hậu (tuần) đưa vào xem xét ở 2 trạm gần như không tuân theo phân bố chuẩn và các dạng phân bố xuất hiện nhiều nhất là Johnson Transformation, Box-Cox Transformation, Weibull và 3-Parameter Weibull. Các kết luận trên phù hợp về mặt ý tưởng với nhận định của Lloyd khi tác giả này tìm ra dạng phân bố là log-normal cho tổng lượng mưa trong nghiên cứu của mình [026] hay Alcantara và Ahn khi nhận xét về các dạng phân bố của lượng mưa ngày ở Hàn Quốc là Pearson loại III, Weibull, Gamma [27].

Ưu điểm của việc khởi tạo số liệu để đánh giá các kịch bản khí hậu cũng như mức biến thiên trong kết quả mô phỏng phụ thuộc số liệu đầu vào là rất lớn. Điều đó còn đặc biệt đúng hơn với nhu cầu đánh giá đa kịch bản khi bài toán mô phỏng phức tạp sử dụng số lượng nhiều các biến số thay đổi liên tục theo không gian và thời gian. Ở nghiên cứu này, giá trị trung bình của những giá trị mô phỏng bốc hơi sử dụng các bộ số liệu khởi tạo không khác nhau ở mức có ý nghĩa về mặt thống kê nhưng lại khác với trung bình số liệu thực đo gốc ban đầu và kết quả mô phỏng chạy trực tiếp trên chuỗi số liệu các yếu tố khí tượng gốc. Tuy nhiên, độ lệch chuẩn của chuỗi mô phỏng sử dụng số liệu khởi tạo với xác suất phân bố tứ phân vị bằng cả 2 kỹ thuật không khác so với độ lệch chuẩn của chuỗi mô phỏng sử dụng xác suất phân bố 10% nhưng lại khác đáng với xác suất phân bố 5%. Điều này cho thấy nấc khởi tạo 5% và tứ phân vị có ảnh hưởng khác nhau đến kết quả mô phỏng bốc hơi về độ biến động trong trung bình kết quả. Sự khác biệt trong trung bình và độ lệch chuẩn không thể hiện rõ về mặt thống kê khi chuyển đổi phương pháp khởi tạo (Monte Carlo hay Latin Hypercube). Các nhận xét trên đã chỉ rõ sự biến động cao hơn trong kết quả mô phỏng gây ra bởi

MÔ PHÒNG LƯỢNG BỐC HƠI...

việc lựa chọn mức chi tiết trong nấc xác suất khởi tạo so với từ việc lựa chọn cách phương pháp khởi tạo sẽ sử dụng. Theo như tác giả được biết, vấn đề này hiện còn ít được đề cập trong các nghiên cứu khác.

4. KẾT LUẬN

Kết quả phân tích số liệu khí hậu thực đo tuần cho giai đoạn 2014 - 2017 ở các trạm khí tượng Nhà Bè và Cần Thơ đã cho thấy hai phương pháp hồi quy tuyến tính đa biến và ANN đều cho độ chính xác mô phỏng lượng bốc hơi ở mức cao (giá trị $R > 0.93$ hay $R^2 > 0.87$). Cấu trúc ANN với 1 lớp ẩn có 6 nơron sử dụng hàm chuyển tansig là phù hợp để mô phỏng lượng bốc hơi cho cả 2 trạm bởi khi cấu trúc mạng phức tạp hơn nữa thì giá trị R tăng không đáng kể. Dẫu rằng ANN mô phỏng tốt hơn về tổng thể cho các bộ số liệu thực đo cũng như khởi tạo nhưng trong một số trường hợp kết quả mô phỏng lại xuất hiện các giá trị nhỏ nhất mang dấu âm trong khi phương pháp hồi quy tuyến tính đa biến không có hiện tượng này.

Kết quả thống kê chi rõ ràng số liệu khí tượng thực đo cho các yếu tố khí tượng khác nhau ở cả 2 trạm đều phần lớn đều không tuân theo phân bố chuẩn hoặc nếu có tuân theo thì mức độ phù hợp vẫn kém hơn so với một/vài phân bố khác. Khi thay đổi phương pháp khởi tạo chuỗi số liệu (Monte Carlo hay Latin Hypercube), kết quả mô phỏng lượng bốc hơi thay đổi ít hơn về giá trị trung bình và độ lệch chuẩn so với khi thay đổi độ chi tiết trong nấc xác suất phân bố của công tác khởi tạo.

Kết quả của nghiên cứu có hạn chế là số liệu khí hậu sử dụng mới chỉ giới hạn cho 2 Trạm khí tượng thuộc khu vực Nam Bộ nên khả năng tổng quát hóa cho nhiều trường hợp khác rất cần được tiếp tục điều tra thêm. Mặt khác, các yếu tố khí hậu khác như tổng bức xạ mặt trời cũng có thể được đưa vào bộ số liệu như đối tượng đánh giá.

LỜI CẢM ƠN

Tác giả xin cảm ơn các đồng nghiệp ở Viện Khoa học Công nghệ và Quản lý Môi trường đã luôn hỗ trợ để cho nghiên cứu này có thể hoàn thành. Xin cảm ơn Đài khí tượng Thủy văn khu vực Nam Bộ đã cung cấp số liệu hiện trường sử dụng cho nghiên cứu.

TÀI LIỆU THAM KHẢO

- [1] Bộ Tài nguyên và Môi trường, *Kịch bản biến đổi khí hậu và nước biển dâng cho Việt Nam*, 2016.
- [2] M. E. Keskin and O. Terzi, "Evaporation estimation models for lake Egirdir, Turkey", *Hydrological Processes*, 20, pp. 2381-2391, 2006.
- [3] R. J. Granger and N. Hedstrom, "Modeling hourly rates of evaporation from small lakes", *Hydrology and Earth System Science*, 15, pp. 267-277, 2011.
- [4] O. Kissi, "Generalized regression neural networks for evapotranspiration modeling", *Hydrological Sciences*, 51(6), pp. 1092-1105, 2006.
- [5] 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.
- [6] S. B. K. Tan, E. B. Shuy, and L. H. C. Chua, "Modelling hourly and daily open-water evaporation rates in areas with an equatorial climate", *Hydrological Processes*, 21, pp. 486-499, 2007.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] N. Q. Hoan, P. T. Trang, H. H. Công và N. T. 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.
- [11] N. T. Ân và N. Q. 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.
- [12] T. T. Vân, H. D. X. Bảo, Đ. T. K. Phụng, N. T. T. Mai và Đ. T. M. 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.
- [13] M. Narvekar and P. Fargose, "Daily weather forecasting using Artificial Neural Network", *International Journal of Computer Applications*, 121 (22), pp. 9-13, 2015.
- [14] K. Abhishek, M. P. Singh, S. Ghosh, and A. Anand, "Weather forecasting model using Artificial Neural Network", *Procedia Technology*, 4, pp. 311-318, 2012.

- [15] N. T. Kottegod, L. Natale, and E. Raiteri, "Monte Carlo Simulation of rainfall hyetographs for analysis and design", *Journal of Hydrology*, 519, pp. 1-11, 2014.
- [16] L. Đ. Linh, N. Q. Toán, N. H. Hải và V. K. Cường, "Sử dụng phương pháp mô phỏng Monte Carlo để ước lượng thời gian dự phòng khi lập tiến độ thi công xây dựng", *Tạp chí Khoa học Công nghệ Xây dựng (KHCN XD) – ĐHXD*, 29, trang 63-67, 2016.
- [17] N. H. Thảo và N. H. Trung, "Ứng dụng mô hình Monte Carlo dự báo diện tích đất sản xuất nông nghiệp trong tối ưu hóa đất nông nghiệp", *Tạp chí Khoa học Trường Đại học Cần Thơ*, 55, trang 164-174, 2019.
- [18] T. V. Thắng, N. T. Phong, L. Long và R. Lars, "Nghiên cứu tạo chuỗi số liệu dòng chảy với mô phỏng Monte Carlo phục vụ bài toán phân bổ hợp lý nguồn nước lưu vực sông Vu Gia - Thu Bồn", *Tạp chí khoa học và công nghệ thủy lợi*, 37, trang 1-8, 2017.
- [19] H. Q. Bằng, "Ứng dụng phương pháp Monte - Carlo để ước lượng sai số trong kết quả mô phỏng chất lượng không khí: áp dụng cho trường hợp thành phố Hồ Chí Minh", *Tạp chí phát triển KH&CN*, 14(M1): , trang 78-91, 2011.
- [20] Oracle Corporation, *Crystal Ball User's Guide v.11.1.2*, 2010.
- [21] D. M. L. Mas and D. P. Ahlfeld, "Comparing artificial neural networks and regression models for predicting faecal coliform concentrations", *Hydrological Sciences Journal*, 52(4), pp. 713-731, 2007.
- [22] G. Bandyopadhyay and S. Chattopadhyay, "Single hidden layer artificial neural network models versus multiple linear regression model in forecasting the time series of total ozone", *Int. J. Environ. Sci. Tech.*, 4 (1), pp. 141-149, 2007.
- [23] K. Lee, K. Kim, J. Kang, ..., Y. Lim, "Comparison and Analysis of Linear Regression & Artificial Neural Network", *International Journal of Applied Engineering Research*, 12(20), pp. 9820-9825, 2017.
- [24] B. A. Ganesh, R. J. Ramireddy, N. K. Santosh, and S. R. P. Reddy, "Performance Evaluation Of Linear Regression And Neural Networks On Forecasting Numerical Data Sets", *International journal of Engineering research & Technology*, 03 (1), pp. 1571-1578, 2014.
- [25] E. F. Nezhad, G. F. Ghalhari, and F. Bayatani, "Forecasting Maximum Seasonal Temperature Using Artificial Neural Networks "Tehran Case Study"". *Asia-Pacific Journal of Atmospheric Sciences*, 55, pp. 145-153, 2019.
- [26] P. Lloyd, "On the determination of trends in rainfall", *Water SA*, 35, pp. 237-243, 2009.
- [27] A. L. Alcantara and K. Ahn, "Probability Distribution and Characterization of Daily Precipitation Related to Tropical Cyclones over the Korean Peninsula", *Water*, 12, pp. 1204-1214, 2020.

SIMULATION OF EVAPORATION IN NAM BO PLAIN BY ARTIFICIAL NEURAL NETWORK, MULTIVARIABLE LINEAR REGRESSION IN COMBINATION WITH RANDOM DATA GENERATION TECHNIQUES

TRAN TRI DUNG

Institute of Environmental Science, Engineering and Management, Industrial University of Ho Chi Minh City
trantridung@iuh.edu.vn

Abstract. This study evaluated the impact of generated meteorological data series on the weekly evaporation simulation results using Artificial Neural Network (ANN) and multivariate linear regression at Can Tho and Nha Be meteorological stations in Nam Bo plain, Vietnam. Because most of data series measured for meteorological factors at both stations do not follow the normal distribution, the data series representing different scenarios had been generated using Monte Carlo, Latin Hypercube techniques with 5%, 10% and quartiles steps based on specific statistics of measured data. The analysis results demonstrated that both multivariate linear regression and ANN methods provided high evaporation simulation accuracy ($R > 0.93$ or $R^2 > 0.87$), while the ANN structure with 1 hidden layer having 6 neurons and tansig transfer function is suitable to simulate evaporation for both stations. The variation in mean and standard deviation of evapotranspiration simulation results is more dependent on choosing steps in the probability distribution initiation processes than on initiation technique selection. In several cases, the results of evaporation simulation by ANN have a negative sign while this phenomenon did not occur with the multivariate linear regression method.

Keywords: simulation, evaporation, Artificial Neural Network (ANN), multivariable linear regression, Monte Carlo, Latin Hypercube

Ngày nhận bài: 04/10/2022

Ngày chấp nhận đăng: 10/01/2023