

ỨNG DỤNG TRÍ TUỆ NHÂN TẠO TRONG ĐÁNH GIÁ MỨC ĐỘ LIÊN QUAN THEO MÙA GIỮA ĐỘ ẨM TƯƠNG ĐỐI KHÔNG KHÍ VÀ CÁC YẾU TỐ KHÍ HẬU KHÁC TẠI KHU VỰC NAM BỘ

TRẦN TRÍ DŨNG

Viện Khoa học công nghệ và Quản lý môi trường - Trường Đại học Công nghiệp Thành phố Hồ Chí Minh

Tác giả liên hệ: trantridung@iuh.edu.vn

DOIs: <https://www.doi.org/10.46242/jstih.v77i5.5235>

TÓM TẮT. Nghiên cứu này đánh giá mức độ quan hệ của độ ẩm tương đối không khí và những đại lượng khí hậu khác tại các trạm khí tượng Nhà Bè, Cần Thơ và Rạch Giá thuộc khu vực Nam Bộ theo hai mùa khô - mưa. Việc dùng Mạng nơron nhân tạo (ANN) để mô phỏng độ ẩm cho bộ số liệu khí hậu ngày trong giai đoạn 2013 - 2020 đã cho kết quả khá tốt với $0.84 < R < 0.94$ (hay $0.71 < R^2 < 0.88$). Đặc biệt, kết quả chứng minh rằng mọi chuỗi R_{max} thu được đều không tuân theo phân bố chuẩn đối với 1000 lần mô phỏng dù rằng một số chuỗi đã thể hiện xu hướng tuân theo các phân bố khác như Johnson Transformation (phổ biến hơn cả), 3 - Parameter Weibull và Smallest Extreme Value. Kết quả đánh giá đồ thị theo phương pháp khuỷu tay đã cho thấy dạng ANN với 1 lớp ẩn có 8 nơron và hàm chuyển Tanh là cấu trúc phù hợp cho mô phỏng độ ẩm không khí ở cả 3 trạm. Kết quả tính toán thể hiện rằng giá trị độ ẩm trong mùa khô liên hệ mạnh nhất với lượng bốc hơi ngày, rồi đến độ ẩm trung bình ngày hôm trước. Tuy nhiên, đối với mùa mưa thì cường độ quan hệ theo thứ tự giảm dần lại là nhiệt độ trung bình và kế đó là lượng bốc hơi ngày. Như vậy, lượng bốc hơi ngày chính là yếu tố quan trọng bậc nhất có tương quan đến độ ẩm không khí tại khu vực khảo sát trong suốt cả năm.

Từ khóa: mô phỏng, liên quan, độ ẩm không khí, mạng nơron nhân tạo (ANN)

1. MỞ ĐẦU

Độ ẩm không khí có vai trò hết sức quan trọng trong thiên nhiên. Nó tham gia tích cực vào vòng tuần hoàn nước tự nhiên, từ đó tác động đến điều kiện khí hậu, nên luôn có ảnh hưởng đến hoạt động sản xuất cũng như đời sống của con người và các loài sinh vật. Thông kê đã chứng minh rằng độ ẩm thường xuyên tác động trực tiếp hay gián tiếp đến năng suất, sản lượng cây trồng - vật nuôi cũng như chất lượng môi trường nói chung [1, 2, 3, 4]. Nhất là giai đoạn hiện nay, cường độ biến đổi khí hậu với xu hướng nhiệt độ tăng và hiện tượng hạn hán xảy ra mạnh mẽ thì tác động từ độ ẩm không khí đến sản xuất nông nghiệp ở nhiều nơi trên thế giới lại càng không thể xem nhẹ. Đó là lý do việc đánh giá quy luật thay đổi của độ ẩm không khí đã và đang được nghiên cứu trong suốt nhiều năm, qua, nhất là đối với những quốc gia có nền sản xuất nông nghiệp phát triển lâu đời như Việt Nam.

Trên thế giới đã có nhiều công trình nghiên cứu đặc điểm biến thiên, cách xác định và mối quan hệ giữa độ ẩm không khí với những yếu tố môi trường khác. Romps (2014) đã đề xuất một mô hình giải tích cho độ ẩm tương đối nhiệt đới sử dụng mối quan hệ Clausius - Clapeyron, cân bằng thủy tĩnh và cân bằng nước khối lượng lớn. Đặc biệt tác giả này cũng đã chứng minh được hiệu quả tốt của công thức trong bối cảnh toàn bộ vùng nhiệt đới chỉ với một số diễn giải lại các biến [5]. Köhli và cộng sự cũng đánh giá sự phụ thuộc của độ ẩm không khí vào những yếu tố khác như tia vũ trụ neutron trên mặt đất [6], trong khi Eccel lại đi sâu nghiên cứu mối quan hệ giữa độ ẩm với nhiệt độ và lượng mưa [7].

Ở Việt Nam, một số nghiên cứu có liên quan đến độ ẩm không khí đã lần lượt được công bố. Nhận xét về ảnh hưởng của hiện tượng biến đổi khí hậu đến điều kiện tự nhiên trong bối cảnh chung toàn quốc, Nguyễn Văn Thắng và cộng sự đã kết luận rằng nền nhiệt độ của nước ta sẽ tiếp tục tăng lên trong thế kỷ 21, đi kèm chỉ số ẩm ướt hàng năm giảm dần [8]. Trong một nghiên cứu khác, Hoàng Phương Hồng sử dụng kết quả đánh giá thống kê các dãy số liệu quan trắc trong 10 năm tại 3 trạm khí tượng thủy văn nằm cách xa nhau để khảo sát sự biến thiên độ ẩm trên cao của Việt Nam [9].

Bên cạnh những phương pháp tính toán thông dụng như hồi quy tuyến tính, hiện giờ con người đã có cơ hội áp dụng các kỹ thuật mới và tiên tiến trí tuệ nhân tạo (AI) trong lĩnh vực khí tượng - môi trường. Trong đó, mạng nơron nhân tạo (ANN) thuộc mảng học sâu thường được coi là rất có triển vọng đối với công tác mô phỏng. Quej và cộng sự (2021) đã thu được kết quả tốt khi tiến hành đánh giá các mô hình trí tuệ nhân tạo sử dụng các yếu tố đầu vào là nhiệt độ, lượng mưa và độ ẩm tương đối trong môi trường ẩm

và khá ẩm tại 5 trạm khí tượng ở Mexico cho việc dự đoán lượng bốc hơi nước tham chiếu ngày [10]. Nyakuri và cộng sự (2023) đã phát triển một hệ thống Trí tuệ nhân tạo (AI) và Internet vạn vật (IoT) dựa trên dữ liệu đo tại đồng ruộng và thời tiết theo thời gian thực, đồng thời cho phép tương tác với trang trại thông qua điện thoại thông minh. Ưu điểm của hệ thống này nằm trong khả năng phân tích để lên lịch tưới tiêu cũng như bón phân giúp tối ưu hóa năng lượng và lượng nước sử dụng [11]. Ở nước ta, Nguyễn Phúc Hiếu và cộng sự (2024) cũng đã phát huy sức mạnh của AI trong công tác xây dựng mô hình thể hiện chỉ số chất lượng không khí tại khu vực thành phố Hồ Chí Minh. Trong đó, độ ẩm không khí là một trong số 5 biến khí tượng dùng cho dự báo khi áp dụng các mô hình học máy như SVR (Support Vector Regression), MLP (Multi-layer Perceptron) và RF (Random Forest) [12].

Theo như tác giả được biết, tuy đã có nhiều đề tài liên quan đến chủ đề mô phỏng số liệu khí tượng bằng ANN được công bố, nhưng việc đánh giá mức độ liên quan theo mùa giữa độ ẩm không khí và các yếu tố khí hậu khác là chủ đề hiện còn rất ít được nghiên cứu. Nghiên cứu này có mục đích xác định hiệu quả của việc dùng mạng nơ-ron nhân tạo (ANN) để mô phỏng độ ẩm tương đối không khí với số liệu khí hậu tại 3 trạm khí tượng Nhà Bè, Cần Thơ và Rạch Giá đặc trưng cho các vùng Đông và Tây Nam Bộ nước ta. Cho mỗi mùa trong năm, sự tác động của từng yếu tố khí hậu (thể hiện mức liên quan) đến sai số của công tác mô phỏng độ ẩm không khí cũng được xem xét.

Những kết quả thu được có thể làm sáng tỏ thêm tiềm năng áp dụng ANN để mô phỏng điều kiện khí hậu tại khu vực Nam Bộ nước ta, đồng thời xác định các yếu tố khí hậu quan trọng nhất có liên quan đến độ ẩm không khí theo từng mùa mưa - khô trong năm. Đối với công tác quan trắc, những kết luận trên rất có ích trong việc phục hồi số liệu khí hậu khi số liệu bị mất hoặc không thu thập liên tục được bởi những nguyên nhân khác nhau. Bên cạnh đó, việc xác định yếu tố chủ đạo chi phối hay liên quan đến độ ẩm không khí cũng có thể giúp cho công tác tăng cường hiệu quả những hoạt động sản xuất nông nghiệp trong thực tế ở nước ta.

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

2.1. Khu vực và số liệu nghiên cứu

a. Khu vực nghiên cứu

Nằm ở phía Nam nước ta, vùng Nam Bộ mang những nét điển hình cho dạng khí hậu nhiệt đới gió mùa với thời gian bức xạ và tổng tích ôn khá lớn. Nhìn chung, khí hậu năm có thể phân thành hai mùa chính là mùa khô kéo dài từ tháng 12 tới tháng 4 năm sau, mùa mưa gồm những tháng còn lại.

Về giáng thủy, đặc điểm nổi bật là lượng mưa phân bố không đồng đều. Trong đó, mưa mang xu hướng giảm dần từ vùng giáp Thành phố Hồ Chí Minh xuống khu vực phía Tây - Tây Nam.

b. Số liệu

- Các trạm khí tượng:

Cả 3 trạm khí tượng nghiên cứu đều duy trì việc quan trắc những yếu tố khí tượng cơ bản và các hiện tượng thời tiết. Khoảng cách theo đường thẳng giữa hai trạm Nhà Bè và Cần Thơ gần bằng 125 km, trong khi khoảng cách giữa hai trạm Rạch Giá và Cần Thơ gần bằng 75 km.

Bảng 1. Thông tin cơ bản về các trạm khí tượng nghiên cứu

Tên trạm khí tượng	Tọa độ	
	Kinh độ Đông	Vĩ độ Bắc
Nhà Bè	106° 43' 41"	10° 39' 36"
Cần Thơ	105° 46' 06.48"	10° 01' 36.85"
Rạch Giá	105° 04' 56.79"	10° 00' 31.08"

- Số liệu:

Số liệu được sử dụng trong bài báo này là kết quả quan trắc ở 3 trạm khí tượng nêu trên, được cung cấp bởi Đài khí tượng Thủy văn khu vực Nam Bộ. Bộ số liệu có độ dài 8 năm từ 1/1/2013 đến 31/12/2020, tổng cộng là 2922 ngày đo. Những thông số khí hậu đưa vào nghiên cứu gồm nhiệt độ không khí, lượng bốc hơi, tốc độ gió, áp suất không khí, lượng mưa, ngày trong năm (thể hiện tính chất tuần hoàn theo chu kỳ của khí hậu).

2.2 Phương pháp nghiên cứu

Nhằm đạt mục đích đề ra, các bước nghiên cứu được tiến hành theo trình tự như sau:

Bước 1: Kiểm tra chất lượng số liệu

Có rất nhiều nguyên nhân gây ra những giá trị không hợp lý ở bộ số liệu đo hiện trường. Các nguyên nhân này có thể là khách quan hay do bởi con người.

Vì chất lượng của bộ số liệu đầu vào sẽ quyết định rất nhiều đến tính đúng đắn và mức tin cậy của kết quả mô phỏng nên chúng ta cần kiểm tra kỹ số liệu thô ban đầu. Bước này có tác dụng loại bỏ những số liệu không hợp lý (quá cao hay thấp, mang giá trị âm), giúp đảm bảo thêm tính chính xác của tính toán ở các bước sau. Tác giả sử dụng các hàm số kết hợp với đồ thị trong phần mềm bảng tính Excel thuộc bộ Microsoft Office để thực hiện kiểm tra chất lượng số liệu thô.

Bước 2: Xác định cấu trúc ANN phù hợp cho công tác mô phỏng độ ẩm không khí

Kế tiếp, tác giả sử dụng công cụ Neural Network Toolbox trong phần mềm Matlab (MathWorks, Inc.) để chạy mô phỏng ANN. Những cấu trúc mạng nơron được khảo sát sẽ có từ 1 đến 2 lớp ẩn với số lượng nơron thay đổi từ 2 đến 16 trong mỗi lớp ẩn.

Bộ số liệu đo theo ngày tại 3 trạm khí tượng ở giai đoạn 2013 - 2020 được dùng cho huấn luyện mạng nơron. Số liệu đã được Matlab chia ngẫu nhiên theo tỷ lệ 70 - 15 - 15 (mặc định) tương ứng cho ba mục đích là huấn luyện mạng, xác nhận và tránh quá ngưỡng, kiểm tra độc lập. Trong nghiên cứu này, những cấu trúc ANN sử dụng hàm “Tansig” trong các lớp trung gian và hàm “Purelin” ở đầu ra, với số giai đoạn huấn luyện tối đa là 1000. Mô phỏng áp dụng thuật toán huấn luyện Levenberg - Marquardt backpropagation (lan truyền ngược). Giá trị R tổng quan của mỗi lần huấn luyện chỉ được chấp nhận nếu sai lệch ít hơn 5% so với R của kiểm tra độc lập. Các biến đầu vào gồm nhiệt độ không khí, áp suất không khí, tốc độ gió, lượng bốc hơi, lượng mưa, ngày trong năm và biến đầu ra là độ ẩm không khí.

Sau khi huấn luyện, mạng sẽ được áp dụng để ước tính độ ẩm không khí với chính dữ liệu thực đo đã dùng ở bước học. Những mức biến thiên trong độ chính xác mô phỏng độ ẩm không khí phụ thuộc vào cấu trúc ANN sử dụng được thể hiện qua giá trị hệ số tương quan R.

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

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.

Đồng thời, chuỗi giá trị kết quả $R_{\max}$ thu được từ 1000 lần mô phỏng cũng được đánh giá theo dạng phân bố chi phối chủ đạo. Phần mềm Minitab (Minitab, LLC) được sử dụng nhằm xác định dạng phân bố tương thích nhất cho các chuỗi $R_{\max}$. Những dạng phân bố được nghiên cứu là: Chuẩn (Normal), Mũ (Exponential), Mũ 2 tham số (2 - Parameter Exponential), Chuyển đổi Box - Cox (Box - Cox Transformation), Gamma, Gamma 3 tham số (3-Parameter Gamma), Weibull, Weibull 3 tham số (3 - Parameter Weibull), Log chuẩn (Lognormal), Log chuẩn 3 tham số (3 - Parameter Lognormal), Giá trị cực tiểu (Smallest Extreme Value), Giá trị cực đại (Largest Extreme Value), Logistic, Loglogistic, Loglogistic 3 tham số (3 - Parameter Loglogistic), Chuyển đổi Johnson (Johnson Transformation). Trong đó, chuyển đổi Johnson sẽ đưa một chuỗi số liệu cho trước về dạng phân bố chuẩn bằng cách chọn một trong số các họ phân bố ở bảng 2 sau [13].

Bảng 2. Chuyển đổi Johnson

Họ Johnson	Hàm chuyển đổi	Khoảng
S_B	$\gamma + \eta \ln [(x - \epsilon) / (\lambda + \epsilon - x)]$	$\eta, \lambda > 0, -\infty < \gamma < \infty, -\infty < \epsilon < \infty, \epsilon < x < \epsilon + \lambda$
S_L	$\gamma + \eta \ln (x - \epsilon)$	$\eta > 0, -\infty < \gamma < \infty, -\infty < \epsilon < \infty, \epsilon < x$
S_U	$\gamma + \eta \sinh^{-1} [(x - \epsilon) / \lambda]$ $\sinh^{-1}(x) = \ln [x + \sqrt{1 + x^2}]$	$\eta, \lambda > 0, -\infty < \gamma < \infty, -\infty < \epsilon < \infty, -\infty < x < \infty$

trong đó:

S_B - Phân bố họ Johnson với biến bị giới hạn (B)

S_L - Phân bố họ Johnson với biến Lognormal (L)

S_U - Phân bố họ Johnson với biến không bị giới hạn (U)

Phân bố Weibull 3 tham số (3 - parameter Weibull):

$$f(x) = \frac{\beta}{\alpha^\beta} (x - \lambda)^{\beta-1} \exp \left[- \left(\frac{x - \lambda}{\alpha} \right)^\beta \right] \quad (2)$$

với: $x > \lambda$, $\alpha > 0$, $\beta > 0$

Phân bố Giá trị cực tiểu (Smallest Extreme Value):

$$f(x) = \frac{1}{\theta} \exp \left(\frac{x - \zeta}{\theta} \right) \exp \left[- \left(\frac{x - \zeta}{\theta} \right) \right] \quad (3)$$

với: $\theta > 0$

Sự phù hợp của mỗi cấu trúc ANN sẽ phụ thuộc vào 2 tiêu chí là nhỏ gọn, song vẫn đảm bảo độ chính xác cho mô phỏng. Đó là một cấu trúc với số lớp ẩn và lượng nơron trong lớp ẩn là nhỏ nhất mà nếu tăng thêm lên thì khả năng mô phỏng chính xác cũng sẽ tăng không đáng kể. Cấu trúc ANN phù hợp nhất được xác định bằng phương pháp đồ thị khuỷu tay (elbow-plot) [14] đối với đồ thị $R_{\max}$. Theo đó, nguyên lý đồ thị khuỷu tay chọn số nơron phù hợp nhất (giá trị theo trục hoành) của giao điểm giữa 2 đường thẳng kẻ chong khít cho các điểm trên 2 nhánh cong cơ bản chi phối đồ thị $R_{\max}$.

Bước 3: Đánh giá mức ảnh hưởng của mỗi yếu tố đầu vào tới độ ẩm không khí

Mạng nơron đã qua huấn luyện vừa tìm ra sẽ được sử dụng để đánh giá mức ảnh hưởng của các biến đầu vào tới độ ẩm không khí. Theo đó, sự thay đổi giá trị $R_{\max}$ trong kết quả mô phỏng độ ẩm được đánh giá cho các kịch bản rút dần từng biến khí hậu khỏi bộ số liệu đầu vào. Kết quả tính toán được tiếp tục xử lý bằng công thức dưới đây

$$A_j = \frac{R_{bd} - R_j}{R_{bd}} \times 100\% \quad (4)$$

trong đó:

A_j - hệ số thể hiện độ ảnh hưởng của biến thứ j (%);

R_j - giá trị $R_{\max}$ của kết quả mô phỏng sau khi rút một biến thứ j nào đó ra khỏi bộ số liệu đầu vào;

R_{bd} - giá trị $R_{\max}$ của kết quả mô phỏng khi còn đầy đủ các biến đầu vào.

Công thức trên cho thấy giá trị của hệ số A_j tỷ lệ thuận với mức ảnh hưởng của yếu tố khí tượng đầu vào thứ j đến độ chính xác trong kết quả mô phỏng độ ẩm không khí.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Kiểm tra chất lượng và thống kê cơ bản các bộ số liệu khí hậu ngày

Chuỗi số liệu khí hậu đo hàng ngày trong 8 năm (2013 - 2020) tại 3 trạm đã được rà soát để xác định sự tồn tại của số liệu có vấn đề như giá trị quá cao, quá thấp hoặc không hợp lý (mang dấu âm do máy đo bị trục trặc kỹ thuật...). Kết quả cho thấy các bộ số liệu đo tại cả 3 trạm đều có chất lượng tốt với những thống kê chính thể hiện trong bảng 3 dưới đây.

Bảng 3. Thống kê cho những yếu tố khí hậu đo được (theo ngày) tại 3 trạm khí tượng nghiên cứu

Tên yếu tố khí hậu	Trạm	Đơn vị đo	Thống kê			
			Giá trị lớn nhất	Giá trị nhỏ nhất	Giá trị trung bình	Độ lệch chuẩn
Nhiệt độ không khí trung bình	1	°C	31.20/31.77/31.77	22.52/24.45/22.52	27.53/28.14/27.84	1.46/1.23/1.38
	2		30.83/31.15/31.15	21.93/24.15/21.93	27.42/27.86/27.64	1.47/1.22/1.37
	3		31.67/32.18/32.18	21.27/25.00/21.27	27.49/28.57/28.03	1.58/1.26/1.53
	1	hPa	1016.42/1012.57/1016.42	1004.90/1003.27/1003.27	1010.84/1008.26/1009.54	1.82/1.64/2.16

Áp suất không khí trung bình	2		1016.20/1012.37/1016.20	1004.73/1003.80/1003.80	1010.73/1008.23/1009.47	1.89/1.56/2.13
	3		1016.55/1012.45/1016.55	1004.87/1003.85/1003.85	1010.84/1008.37/1009.59	1.86/1.52/2.10
Lượng bốc hơi	1	mm	7.60/5.80/7.60	0.40/0.30/0.30	3.32/2.41/2.86	1.09/0.86/1.08
	2		5.60/5.80/5.80	0.70/0.40/0.40	3.08/2.47/2.77	0.81/0.84/0.88
	3		6.70/6.80/6.80	0.80/0.50/0.50	3.34/2.91/3.12	0.90/1.02/0.98
Tốc độ gió trung bình	1	m/s	5.50/3.75/5.50	0/0/0	1.37/1.18/1.27	0.76/0.64/0.71
	2		3.00/4.00/4.00	0/0/0	1.19/1.46/1.33	0.45/0.65/0.57
	3		5.50/9.25/9.25	0/0.50/0	1.88/3.58/2.74	0.66/1.67/1.53
Tổng lượng mưa	1	mm	227.60/140.80/227.60	0/0/0	1.32/7.32/4.34	8.61/14.29/12.20
	2		97.50/111.40/111.40	0/0/0	1.49/7.45/4.49	6.35/13.01/10.68
	3		78.70/145.80/145.80	0/0/0	2.45/9.77/6.13	8.47/16.71/13.77
Độ ẩm tương đối trung bình	1	%	95.00/96.50/96.50	58.75/63.25/58.75	76.21/82.69/79.47	5.94/4.93/6.35
	2		94.25/96.50/96.50	61.25/66.75/61.25	77.12/83.73/80.45	5.00/4.93/5.96
	3		94.75/94.75/94.75	61.00/65.50/61.00	77.39/82.25/79.84	4.62/4.67/5.24

**Chú thích:* trạm khí tượng Nhà Bè được ký hiệu là trạm 1, trạm khí tượng Cần Thơ được ký hiệu là trạm 2 và trạm khí tượng Rạch Giá được ký hiệu là trạm 3; trong mỗi ô các giá trị A/B/C thể hiện tương ứng cho kết quả thống kê số liệu cho mùa khô/mùa mưa/chung cả năm được làm tròn đến 2 chữ số lẻ sau dấu thập phân.

3.2. Xác định cấu trúc ANN phù hợp để mô phỏng độ ẩm không khí

Tác giả đã dùng phần mềm Matlab để chạy huấn luyện ANN với 1 đến 2 lớp ẩn mang hàm chuyển Tansig. Trong đó, từng cấu trúc 2, 4, 6, 8, 10, 12, 14 và 16 nơron ở mỗi lớp ẩn được khảo sát. Cho mỗi cấu trúc ANN nêu trên, mô phỏng được lặp lại 1000 lần với việc ghi nhận lần chạy cho giá trị lớn nhất của hệ số tương quan R giữa giá trị thực đo và mô phỏng ($R_{\max}$).

Kết quả cho thấy rằng việc mô phỏng độ ẩm không khí sử dụng ANN đối với số liệu cả 3 trạm đều có độ chính xác tương đối cao với giá trị $R_{\max}$ nằm trong khoảng từ 0.84 đến 0.94.

a) Kiểm tra dạng phân bố của các chuỗi $R_{\max}$ thu được khi mô phỏng

Kết quả kiểm tra dạng phân bố của các chuỗi $R_{\max}$ thu được khi mô phỏng ANN sử dụng phần mềm Minitab được trình bày trong bảng 4. Trong khi kiểm định, giá trị $p > 0.05$ và giá trị thống kê Anderson - Darling (AD) thấp thể hiện rằng phân bố là phù hợp.

Bảng 4. Kiểm tra dạng phân bố sử dụng phần mềm Minitab cho các chuỗi $R_{\max}$ thu được khi mô phỏng ANN ở 3 trạm khí tượng nghiên cứu

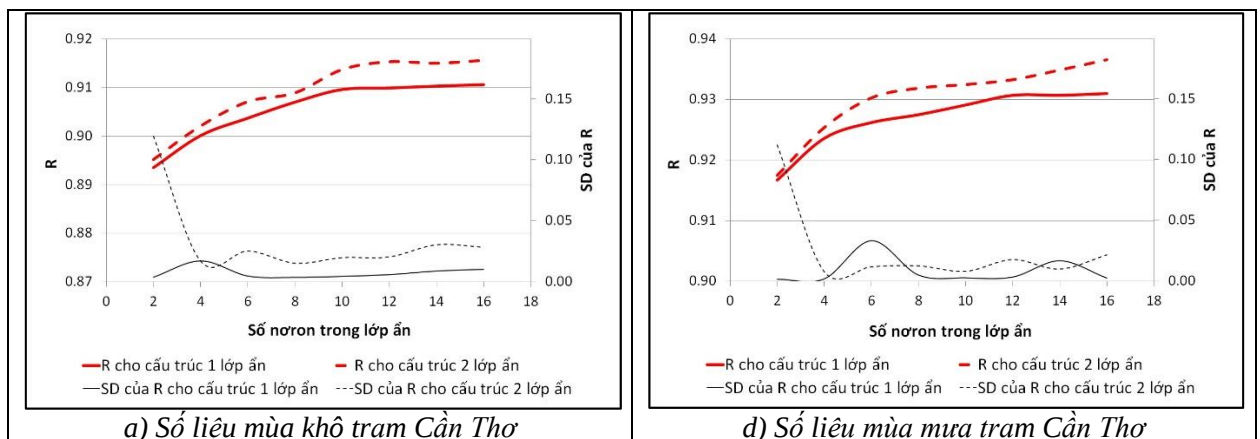
Mùa	Cấu trúc ANN sử dụng cho mô phỏng		Dạng phân phối phù hợp cho chuỗi $R_{\max}$		
	Số lớp ẩn	Số nơron trong mỗi lớp ẩn	Trạm Nhà Bè	Trạm Cần Thơ	Trạm Rạch Giá
Khô	1	2	-	-	-
		4	-	Johnson Transformation	-
		6	Johnson Transformation	Johnson Transformation, Smallest Extreme Value	Johnson Transformation
		8	Johnson Transformation	Johnson Transformation	Johnson Transformation
		10	-	Johnson Transformation	Johnson Transformation
		12	-	-	Johnson Transformation
		14	-	-	Johnson Transformation
		16	-	Johnson Transformation	-

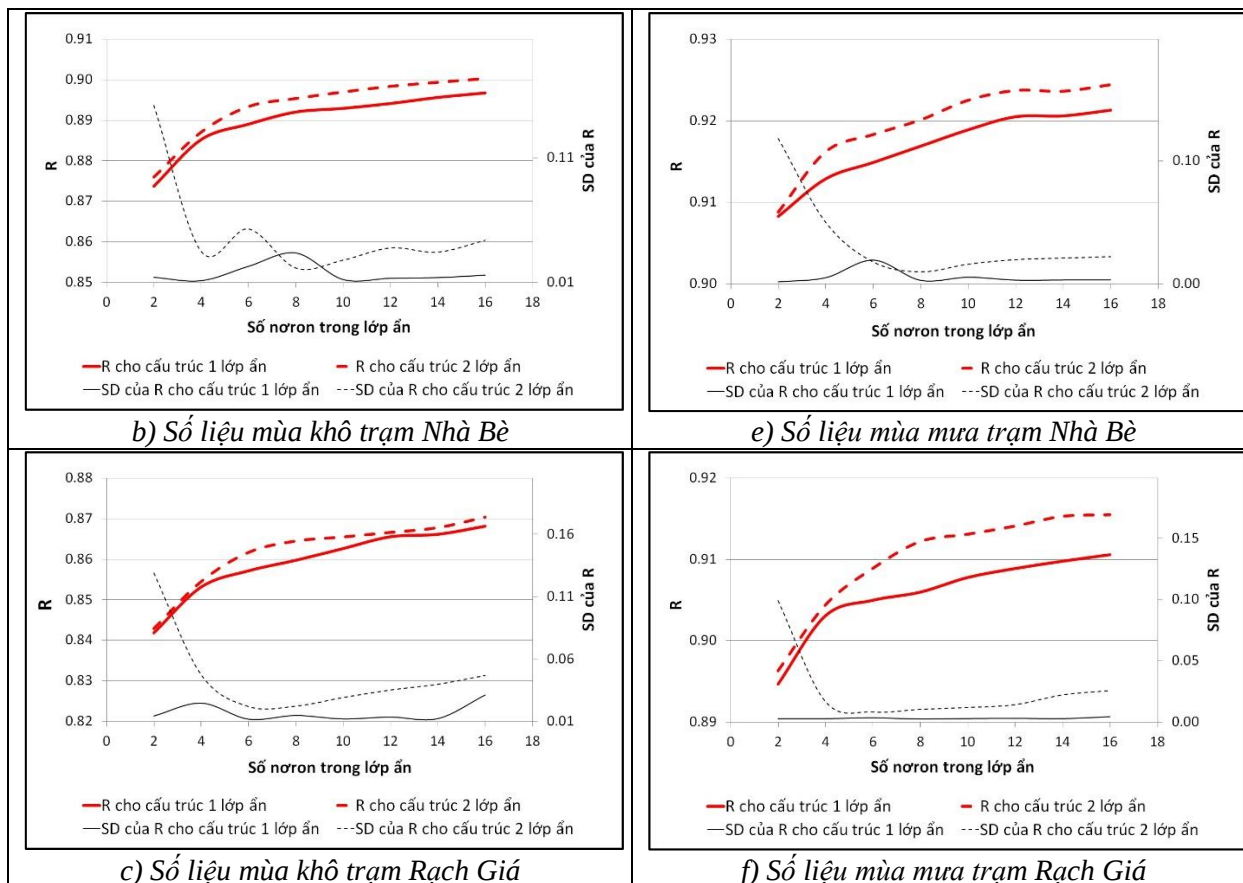
ỨNG DỤNG TRÍ TUỆ NHÂN TẠO TRONG ĐÁNH GIÁ MỨC ĐỘ LIÊN QUAN THEO MÙA...

Mưa	2	2	-	-	-
		4	-	-	-
		6	-	-	Johnson Transformation
		8	-	-	-
		10	-	-	-
		12	-	-	-
		14	-	-	-
		16	-	-	-
	1	2	-	-	-
		4	-	-	-
		6	-	-	Johnson Transformation
		8	Johnson Transformation	-	Johnson Transformation
		10	Johnson Transformation	-	Johnson Transformation
		12	Johnson Transformation, 3 - Parameter Weibull	-	Johnson Transformation
		14	Johnson Transformation	-	Johnson Transformation
		16	-	Johnson Transformation	-
	2	2	-	-	-
		4	-	-	-
		6	-	-	-
		8	-	-	-
		10	-	-	-
		12	-	-	-
		14	-	-	-
		16	-	-	-

Chú thích: Nếu một chuỗi $R_{\max}$ có nhiều hơn 1 dạng phân bố phù hợp thì các phân bố này được sắp xếp theo thứ tự phù hợp từ cao đến thấp hơn.

b) Xác định cấu trúc ANN phù hợp cho mô phỏng độ ẩm không khí

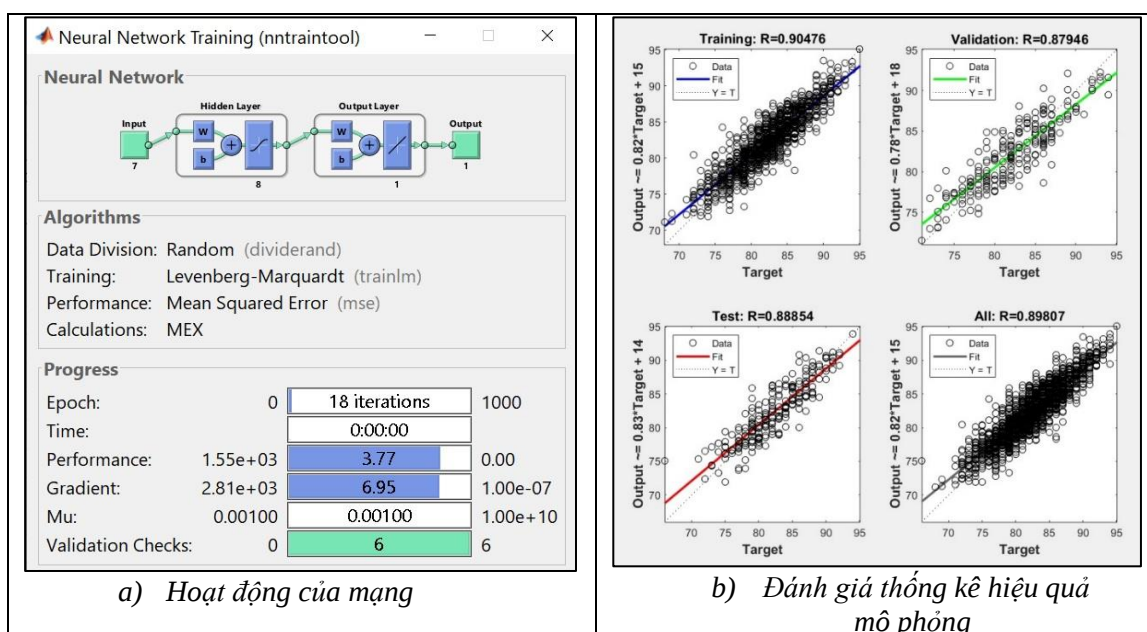




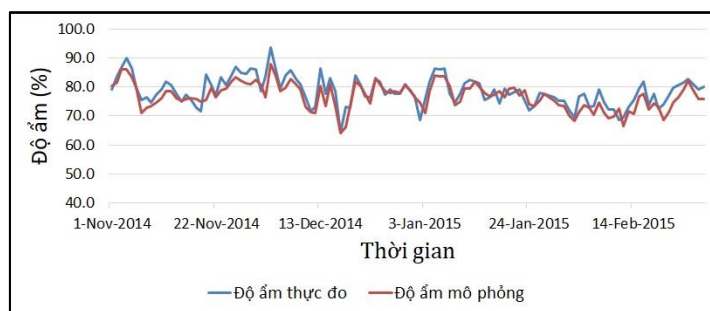
Hình 1. Biểu đồ kết quả mô phỏng ANN

*Chú thích: SD là độ lệch chuẩn.

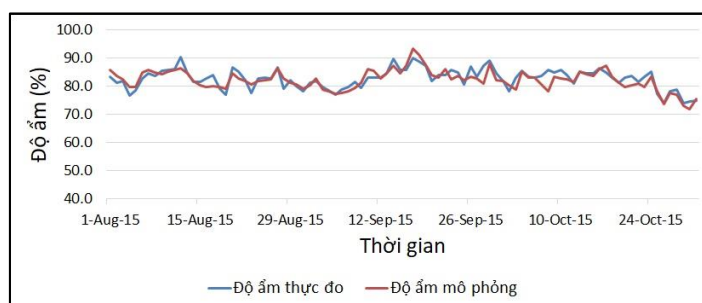
Cho mục tiêu mô phỏng độ ẩm không khí ở 3 trạm đang xem xét, kết quả phân tích bằng phương pháp đồ thị khuỷu tay đối với đồ thị $R_{\max}$ cho thấy cấu trúc ANN truyền thẳng có 1 lớp ẩn với 8 nơron dùng hàm truyền Tansig là đáp ứng tốt nhất cả 2 tiêu chí nhỏ gọn cũng như đạt mức R cao. Một số kết quả cụ thể khi sử dụng cấu trúc ANN với 1 lớp ẩn và 8 nơron để mô phỏng số liệu tại các trạm khí tượng được thể hiện trong các hình 2, 3 và 4 dưới đây.



Hình 2. Mô phỏng bằng mạng ANN cấu trúc 7 - 8 - 1



Hình 3. So sánh kết quả mô phỏng độ ẩm bằng mạng ANN cấu trúc 7 - 8 - 1 với số liệu thực đo trong một số tháng mùa khô giai đoạn 2014 - 2015 cho trạm Nhà Bè



Hình 4 So sánh kết quả mô phỏng độ ẩm bằng mạng ANN cấu trúc 7 - 8 - 1 với số liệu thực đo trong một số tháng mùa mưa giai đoạn 2015 cho trạm Rạch Giá

c) Đánh giá mức ảnh hưởng của mỗi yếu tố đầu vào tới độ ẩm không khí

Cấu trúc ANN 1 lớp ẩn với 8 neuron tìm ra ở bước trên đã được sử dụng để đánh giá mức ảnh hưởng của mỗi yếu tố đầu vào tới độ ẩm không khí. Mô phỏng được thực hiện 1000 lần độc lập cho từng trường hợp nghiên cứu. Kết quả kiểm tra hệ số A_j biểu hiện cho mức giảm trong hệ số tương quan ($R_{\max}$) mô phỏng độ ẩm thu được khi tuần tự rút từng biến khỏi số liệu đầu vào cho 3 trạm khí tượng được trình bày trong bảng 5 sau.

Bảng 5. Hệ số A_j thu được khi rút riêng mỗi yếu tố khí hậu khỏi bộ số liệu đầu vào so với khi còn nguyên cả 7 yếu tố

Mùa	Yếu tố được rút ra	A_j (%)			
		Trạm Cần Thơ	Trạm Nhà Bè	Trạm Rạch Giá	Trung bình cho cả 3 trạm
Khô	Ngày trong năm	0.573	1.939	0.244	0.92
	Nhiệt độ trung bình	0.937	0.975	1.512	1.14
	Áp suất trung bình	0.805	0.661	0.675	0.71
	Lượng bốc hơi ngày	7.078	7.769	9.108	7.99
	Tốc độ gió trung bình	0.077	0.426	0.174	0.23
	Lượng mưa ngày	0.761	0.583	1.687	1.01
	Độ ẩm trung bình ngày trước đó	1.312	2.175	1.535	1.67
Mưa	Ngày trong năm	0.485	0.611	0.530	0.54
	Nhiệt độ trung bình	3.677	4.657	5.574	4.64
	Áp suất trung bình	0.593	0.196	0.662	0.48
	Lượng bốc hơi ngày	2.469	1.614	1.424	1.84
	Tốc độ gió trung bình	0.065	0.164	0.519	0.25
	Lượng mưa ngày	0.323	0.185	0.331	0.28
	Độ ẩm trung bình ngày trước đó	0.000	0.338	0.232	0.19

3.4 Thảo luận

So với những phương pháp mô phỏng cổ điển, ANN dù là phương pháp mới nhưng đã được chứng minh trong thực tiễn là có khả năng mô tả tốt các dạng quan hệ phức tạp. Kết quả mô phỏng độ ẩm không khí sử dụng ANN cho 3 trạm khí tượng đều đạt mức chính xác tương đối cao với giá trị $R_{\max}$ nằm trong khoảng từ 0.84 đến 0.94. Đây là kết quả khá hợp lý nếu xét đến mối quan hệ hỗ tương chặt chẽ trong cùng hệ thống giữa các yếu tố khí hậu với nhau và khả năng mô phỏng tốt hiện tượng với hàm phi tuyến của ANN.

Kết quả khảo sát sử dụng phần mềm Minitab đối với dạng phân bố của các chuỗi $R_{\max}$ thu được khi mô phỏng ANN cho thấy rằng tất cả các chuỗi $R_{\max}$ đều không thuộc về dạng phân bố chuẩn. Một số chuỗi $R_{\max}$ thể hiện xu hướng tuân theo phân bố Johnson Transformation (phổ biến hơn cả), 3 - Parameter Weibull và Smallest Extreme Value. Dù kết quả không thể hiện rõ ràng tính quy luật nào trong mối liên hệ giữa số lớp ẩn / số nơron trong mỗi lớp ẩn và quy luật phân bố của những chuỗi $R_{\max}$, nhưng đối với $R_{\max}$ của 2 lớp ẩn thì gần như hoàn toàn không có bất kỳ dạng phân bố nào (trừ 1 trường hợp cho Trạm Rạch Giá) trong số 16 dạng phân bố đưa vào khảo sát thỏa mãn. Theo như cố gắng tìm hiểu của tác giả, kết quả này hiện chưa thấy được công bố trong các nghiên cứu về mô phỏng ANN và rất nên được điều tra sâu hơn nữa để tìm hiểu chi tiết quy luật biến thiên của R khi sử dụng ANN nhằm khai thác tối đa sức mạnh của phương pháp.

Như dự đoán, đồ thị của $R_{\max}$ đã cho thấy các cấu trúc ANN với 2 lớp ẩn cho giá trị $R_{\max}$ thường cao hơn so với cấu trúc 1 lớp ẩn của cùng số nơron trong lớp ẩn. Song dù độ lệch chuẩn của chuỗi $R_{\max}$ có xu hướng giảm dần khi cấu trúc ANN phức tạp hơn nhưng điều này vẫn đôi khi bị vi phạm kể cả đối với các cấu trúc dùng đến 16 nơron. Đặc biệt, hiện tượng độ lệch chuẩn của chuỗi $R_{\max}$ có giá trị rất cao đã quan sát thấy trong mọi trường hợp sử dụng 2 nơron trong lớp ẩn. Điều này cho thấy các cấu trúc 2 nơron thường là không đủ mạnh để duy trì tính ổn định trong mô phỏng độ ẩm không khí và nên tránh sử dụng.

Cấu trúc ANN phù hợp luôn phải đảm bảo duy trì chất lượng mô phỏng mối quan hệ giữa các biến đầu vào và ra mà đôi khi xảy ra rất phức tạp. Theo tác giả tìm hiểu, việc thiết kế cấu trúc ANN để đạt hiệu suất mô phỏng cao nhưng đồng thời tránh khỏi hiện tượng quá ngưỡng vẫn là mối quan tâm hàng đầu của các nhà khoa học. Qua điều tra này, cấu trúc ANN có 1 lớp ẩn với 8 nơron dùng hàm chuyển Tanh đã được xác định là phù hợp để mô phỏng độ ẩm không khí cho 3 trạm ở Đồng bằng Nam Bộ đưa vào khảo sát. Số lượng 8 nơron trong lớp ẩn rút ra từ nghiên cứu này xấp xỉ bằng số nơron sử dụng bởi Abudu và cộng sự [15], song thấp hơn số nơron mà Abhishek và cộng sự [16], cũng như Nezhad và cộng sự [17] đã dùng để mô phỏng. Dù vậy, so sánh trên cũng chỉ là tương đối bởi sự khác biệt ngay trong số lượng và tên biến trong bộ số liệu của từng nghiên cứu, cộng thêm khả năng tồn tại những dị biệt khác như việc lựa chọn hàm truyền khi mô phỏng.

Cấu trúc ANN có 1 lớp ẩn với 8 neuron đã được sử dụng để đánh giá mức ảnh hưởng của từng yếu tố đầu vào tới độ ẩm không khí với 1000 lần chạy độc lập cho mỗi trường hợp nghiên cứu. Căn cứ vào mức thay đổi trung bình trong $R_{\max}$ (%) cho cả 3 trạm khi rút mỗi yếu tố so với khi còn nguyên mọi yếu tố đầu vào, chúng ta có thể thấy rằng mức ảnh hưởng tới độ ẩm trong mùa khô thể hiện mạnh nhất (nếu chỉ xét 2 yếu tố quan trọng nhất đứng hàng đầu) ở lượng bốc hơi ngày và kế đó là độ ẩm trung bình ngày trước đó. Trong mùa mưa, mức ảnh hưởng theo thứ tự mạnh yếu là nhiệt độ trung bình và kế đó là lượng bốc hơi ngày. Như vậy, lượng bốc hơi ngày là yếu tố vô cùng quan trọng liên hệ chặt chẽ với giá trị độ ẩm không khí ở cả 2 mùa tại Đồng bằng Nam Bộ.

Về mặt cơ chế vật lý, đối với mùa khô, hiện tượng độ ẩm trung bình ngày trước đó có tầm quyết định chỉ đứng sau lượng bốc hơi ngày có thể là biểu hiện rõ ràng cho tính kế thừa của chuỗi số liệu độ ẩm không khí từ những thời điểm nằm kế trước đó. Tuy nhiên, đối với mùa mưa thì vai trò hàng đầu lại thuộc về nhiệt độ trung bình và điều này có thể được giải thích rằng trong mùa mưa thì lượng nước từ mặt đất hay thảm thực vật cung cấp cho quá trình bốc hơi là không bị thiếu (so với mùa khô khi có rất ít trận mưa xảy ra) và do đó nhiệt lượng gây bốc hơi mới là yếu tố quyết định đến lượng hơi nước trong không khí. Các quan sát trên thể hiện rất giống nhau ở 3 trạm khí tượng và nhấn mạnh rằng sự khác biệt trong đặc điểm biến thiên giữa 2 mùa mưa - khô là vô cùng quan trọng đối với quy luật thay đổi của độ ẩm không khí tại nhiều vị trí quan trắc nằm xa nhau của Đồng bằng Nam Bộ.

4. KẾT LUẬN

Việc sử dụng phương pháp ANN để mô phỏng độ ẩm tương đối không khí sử dụng bộ số liệu khí hậu thực đo hàng ngày cho giai đoạn 2013 - 2020 ở các trạm khí tượng Nhà Bè, Cần Thơ và Rạch Giá đã

cho kết quả khả quan. Độ chính xác mô phỏng cho 1000 lần lặp ngẫu nhiên đạt mức khá cao với giá trị $0.84 < R < 0.94$ (hay $0.71 < R^2 < 0.88$). Phân tích cho thấy cấu trúc ANN với 1 lớp ẩn có 8 nơron sử dụng hàm chuyển Tansig là phù hợp để mô phỏng độ ẩm không khí cho cả 3 trạm bởi khi cấu trúc mạng phức tạp hơn nữa thì giá trị $R_{\max}$ tăng không đáng kể. Kết quả phân tích thể hiện rằng tất cả các chuỗi $R_{\max}$ thu được đều không tuân theo phân bố chuẩn, trong đó một số chuỗi đã thể hiện xu hướng tuân theo các phân bố khác như Johnson Transformation (phổ biến hơn cả), 3 - Parameter Weibull và Smallest Extreme Value.

Kết quả tính toán bằng cách rút dần từng yếu tố đầu vào cho thấy giá trị độ ẩm trong mùa mưa có quan hệ mạnh nhất với nhiệt độ trung bình và kế đó là lượng bốc hơi ngày. Song đối với mùa khô, mối liên hệ mạnh nhất của độ ẩm lại là với lượng bốc hơi ngày, rồi đến độ ẩm trung bình ngày trước đó. Như vậy trong suốt cả năm, lượng bốc hơi ngày chính là yếu tố quan trọng bậc nhất có tương quan đến độ ẩm không khí ở khu vực đồng bằng Nam Bộ nước ta.

Nghiên cứu này còn có điểm hạn chế là mới chỉ giới hạn cho 3 Trạm khí tượng cho nên khả năng tổng quát hóa cần được tiếp tục điều tra thêm bằng các đánh giá ở những khu vực khác. Hơn nữa, nếu điều kiện cho phép có số liệu đo thì những yếu tố khí hậu khác như tổng bức xạ mặt trời cũng có thể được bổ sung thêm vào quá trình mô phỏng. Đồng thời, việc nghiên cứu ảnh hưởng hoàn lưu đến độ ẩm tương đối sẽ hết sức hữu ích nếu có đủ số liệu đo đạc theo không gian.

LỜI CẢM ƠN

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 bài báo. Xin cảm ơn Đài khí tượng Thủy văn khu vực Nam Bộ đã cung cấp số liệu dùng cho nghiên cứu.

TÀI LIỆU THAM KHẢO

- [1] P. V. Trường, N. Đ. Văn, N. Đ. Núi và T. A. Tuấn, “Ảnh hưởng của các yếu tố khí hậu đối với quá trình bốc thoát hơi trong đới thông khí vùng đồng bằng ven biển Ninh Thuận”, *Tạp chí Tài nguyên nước*, 1, trang 51 - 59, 2021.
- [2] T. V. Tuấn, B. Đ. Hoàn và N. X. Sang, “Nghiên cứu ảnh hưởng của điều kiện khí tượng đến hàm lượng PM_{10} và $PM_{2.5}$ trong không khí xung quanh cao tốc Pháp Vân - Cầu Giẽ”, *Tạp chí khoa học công nghệ hàng hải*, 66 (4), trang 99 - 102, 2021.
- [3] E. Arulmozhi, J. K. Basak, T. Sihalath, J. Park, H. T. Kim, and B. E. Moon, Machine Learning-Based Microclimate Model for Indoor Air Temperature and Relative Humidity Prediction in a Swine Building, *Animals*, 11 (1), 222, 2021.
- [4] M. Benzaama, K. Touati, Y. E. Mendili, M. L. Guern, F. Streiff, and S. Goodhew, Machine Learning-Based Indoor Relative Humidity and CO_2 Identification Using a Piecewise Autoregressive Exogenous Model: A Cob Prototype Study, *Energies*, 17(1), 243, 2024.
- [5] D. M. Romps, “An Analytical Model for Tropical Relative Humidity”, *Journal of Climate*, 27(19, pp. 7432 - 7449, 2014.
- [6] M. Köhli, J. Weimar, M. Schrön et al., “Soil Moisture and Air Humidity Dependence of the Above-Ground Cosmic-Ray Neutron Intensity”, *Frontiers in Water*, 2, pp. 1 - 15, 2021.
- [7] E. Eccel, “Estimating air humidity from temperature and precipitation measures for modelling applications”, *Meteorological Applications*, 19, pp. 118 - 128, 2012.
- [8] N. V. Thắng, N. T. Hiệu, P. T. Hương và N. S. Giai, “Đánh giá tác động của biến đổi khí hậu đến điều kiện tự nhiên và tài nguyên khí hậu”, *Tạp chí Khí tượng Thủy văn*, 595, trang 7 - 13, 2010.
- [9] H. P. Hồng, “Chế độ độ ẩm trên cao của Việt Nam”, *Tạp chí Khí tượng Thủy văn*, 426 (6), trang 30 - 35, 1996.
- [10] V. H. Quej, C. D. L. C. Castillo, J. Almorox et al., “Evaluation of artificial intelligence models for daily prediction of reference evapotranspiration using temperature, rainfall and relative humidity in a warm sub-humid environment”, *Italian Journal of Agrometeorology*, 1, pp. 49 - 63, 2022.
- [11] J. P. Nyakuri, V. Ishimwe, J. L. Uwimana et al., “AI Based Real-Time Weather Condition Prediction with Optimized Agricultural Resources”, *European Journal of Technology*, 7 (2), pp. 36 - 49, 2023.
- [12] N. P. Hiếu, Đ. D. H. Vân và Đ. N. Khôi, “Xây dựng mô hình mô phỏng chỉ số chất lượng không khí cho thành phố Hồ Chí Minh bằng phương pháp học máy”, *Tạp chí Khí tượng Thủy văn*, 761, trang 13-24, 2024.
- [13] Minitab, LLC website: <https://www.minitab.com/en-us/> (truy cập ngày 10/8/2024)
- [14] M. Lavielle, Using penalized contrasts for the change-point problem, *Signal Processing*, 85(8), pp. 1501 - 1510, 2005.
- [15] 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.
- [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] E. F. Nezhad, G. F. Ghalhari, F. Bayatani, Forecasting Maximum Seasonal Temperature Using Artificial Neural Networks “Tehran Case Study”, *Asia-Pacific Journal of*

ARTIFICIAL INTELLIGENCE APPLICATION FOR ASSESSMENT OF SEASONAL RELATIONSHIP BETWEEN AIR RELATIVE HUMIDITY AND OTHER CLIMATE FACTORS IN NAMBO PLAIN

TRAN TRI DUNG

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

ABSTRACT

This study evaluated the relationship between air relative humidity and other climatic variables at Nha Be, Can Tho and Rach Gia meteorological stations in the Southern region during both dry - rainy seasons. The use of Artificial Neural Networks (ANN) to simulate humidity for daily climate data sets in the period 2013 - 2020 provided quite good results with $0.84 < R < 0.94$ (or $0.71 < R^2 < 0.88$). In particular, the results demonstrated that all obtained $R_{\max}$ series for 1000 simulations did not follow normal distribution, although some series have shown a tendency to follow other distributions such as Johnson Transformation (the most common), 3-Parameter Weibull and Smallest Extreme Value. Elbow method of graphical evaluation demonstrated that the ANN with 1 hidden layer containing 8 neurons and Tansig transfer function was the most suitable structure for simulating air humidity at all 3 stations. The calculation results proved that humidity value in dry season is most strongly related to daily evaporation, followed by the previous day's average humidity. However, for rainy season, the relationship strength in decreasing order was average temperature, followed by daily evaporation. Thus, daily evaporation is the most important factor with a stable relationship to air humidity in survey area throughout the year.

Keywords: *simulation, relation, air humidity, Artificial Neural Network (ANN) Atmospheric Sciences, 55, pp. 145-153, 2019.*

Ngày gửi bài: 21/10/2024

Ngày chấp nhận đăng: 14/03/2025