

# ĐÁNH GIÁ MỨC TỰ TƯƠNG QUAN TRONG SỐ LIỆU KHÍ HẬU TẠI HAI TRẠM KHÍ TƯỢNG Ở ĐỒNG BẰNG NAM BỘ SỬ DỤNG HÀM TỰ TƯƠNG QUAN RIÊNG PHẦN (PACF) VÀ MẠNG NƠON NHÂN TẠO (ANN)

TRẦN TRÍ DŨNG

Viện Khoa học công nghệ và Quản lý môi trường - Đại học Công nghiệp Thành phố Hồ Chí Minh  
trantridung@iuh.edu.vn

DOIs: <https://doi.org/10.46242/jstiuh.v68i02.5088>

**Tóm tắt.** Nghiên cứu này sử dụng Mạng nơon nhân tạo (ANN) và biểu đồ hàm tự tương quan riêng phần (PACF) để đánh giá mức tự tương quan trong chuỗi thời gian của những yếu tố khí hậu giai đoạn 2014 - 2017 đo tại các trạm khí tượng Cần Thơ và Nhà Bè. Cấu trúc ANN được sử dụng có 2 lớp ẩn dạng n - 25 - 25 - 1, thuật toán huấn luyện lan truyền ngược Levenberg-Marquardt với hàm chuyển “tansig” cho các lớp ẩn. Kết quả cho thấy vấn đề tự tương quan tồn tại ở mức có ý nghĩa thống kê đối với tất cả các yếu tố khí hậu đưa vào xem xét, nhưng thường tắt dần (với khoảng cách tối đa 5 - 7 ngày, riêng tổng lượng mưa lên đến 30 ngày). Trừ yếu tố tổng lượng mưa, giá trị hệ số tương quan lớn nhất ( $R_{max}$ ) của 1000 lần mô phỏng ANN độc lập là khá hiệu quả để đánh giá mức độ tương quan cho số liệu ngày trễ (hay năm trước) đầu tiên, dù còn chưa xác định rõ được mức ảnh hưởng đến các ngày năm xa hơn. Để ANN phát huy tác dụng, giá trị PACF của một ngày trễ (hay năm trước) so với ngày nghiên cứu có thể sẽ phải lớn hơn 0.4 và cao hơn ít nhất 2 lần so với giá trị PACF lớn kế nó.

**Từ khóa.** tự tương quan, yếu tố khí hậu, mạng nơon nhân tạo (ANN), hàm tự tương quan riêng phần (PACF), mô phỏng.

## I. MỞ ĐẦU

Nhiều lĩnh vực hoạt động của con người chịu ảnh hưởng của diễn biến khí hậu xảy ra hàng năm. Bên cạnh đó, một số nghiên cứu đã cho thấy rằng hiện tượng biến đổi khí hậu toàn cầu hiện đang diễn ra ngày một mạnh mẽ hơn ở khắp mọi quốc gia trên thế giới. Đối với nước ta, ảnh hưởng của các kịch bản biến đổi khí hậu như đã công bố năm 2016 cũng thể hiện rõ xu thế biến đổi cực đoan của những yếu tố khí hậu. Chúng đã và đang gây ra tác động tiêu cực đến sản xuất nông nghiệp cũng như các mặt khác trong đời sống của con người [1]. Do đó, công tác theo dõi và đánh giá xu thế biến động của những yếu tố khí hậu cơ bản như nhiệt độ không khí, gió, mưa là vô cùng cấp thiết hiện nay.

Trên thế giới đã có nhiều nghiên cứu tập trung vào việc xác định quy luật biến động và phương pháp tính toán các yếu tố khí hậu quan trọng. Việc thu thập và phân tích số liệu đo đạc được thực hiện cho các địa điểm thiên nhiên như hồ nước hoặc theo diện rộng như các vùng địa lý [2, 3, 4], ngoài ra còn được hỗ trợ bằng việc dùng các bộ phương tiện quan trắc hiện đại khác nhau như trạm khí tượng [5, 6]. Những nghiên cứu này đã mô tả những khía cạnh trong biến động của những yếu tố khí hậu điển hình cũng như cho thấy rằng diễn biến của chúng đôi khi xảy ra rất phức tạp.

Công tác đánh giá định lượng và mô phỏng chính xác các yếu tố khí hậu luôn là một thách thức không nhỏ đối với các nhà khoa học. Ngoài các cách tiếp cận tính toán phổ biến hay được sử dụng như phương pháp hồi quy đa biến, mạng nơon nhân tạo (ANN) hiện là một trong những công cụ khoa học tiên tiến thuộc mảng học sâu có tiềm năng trong công tác mô phỏng - dự báo. Kết quả của những công trình được công bố đã cho thấy các mô hình sử dụng ANN có thể phát hiện và nắm bắt được mối quan hệ phức tạp giữa nhiều yếu tố mà đôi khi rất khó đạt được bằng những phương pháp khác. Các công trình nghiên cứu về ứng dụng ANN có đối tượng hết sức đa dạng, như hiện tượng bốc thoát hơi [7, 8, 9] hay nhiệt độ không khí [10, 11]. Ví dụ như ở Ấn Độ, Kaur và Singh đã sử dụng mạng nơon nhiều lớp (MLP) với thuật toán lan truyền ngược để huấn luyện mạng trên số liệu thực đo phục vụ công tác nghiên cứu mô phỏng nhiệt độ tối thiểu cho thành phố Chandigarh. Kết quả thu được chỉ ra rằng nhiệt độ tối thiểu có khả năng được dự đoán với độ chính xác khá cao nhờ công cụ mạng nơon nhân tạo [12]. Để ước tính sự phân bố không gian

và nguy cơ ô nhiễm kim loại nặng trong đất nông nghiệp ở Thổ Nhĩ Kỳ, Günal và cộng sự đã nghiên cứu khả năng kết hợp tự tương quan không gian với các mô hình trí tuệ nhân tạo và thu được kết quả khả quan [13]. Trong một nghiên cứu khác, Kim và cộng sự cũng đã phát triển mô hình học máy kết hợp tự tương quan không gian và đi đến kết luận rằng việc lựa chọn mô hình hoạt động tốt nhất chỉ dựa trên lỗi dự đoán thấp nhất là không phù hợp cho mô hình dự đoán không gian, do đó việc phân tích kỹ lưỡng dữ liệu dự đoán bằng trực quan là rất cần thiết đối với mô hình địa kỹ thuật dưới bề mặt đất [14].

Ở Việt Nam, một số tác giả cũng đã nghiên cứu việc sử dụng công nghệ thông tin và kỹ thuật ANN phục vụ cho công tác điều tra những đặc trưng khí hậu đối với các khu vực khác nhau. Nguyễn Quang Hoan và cộng sự khi nghiên cứu hệ thống dự báo nhiệt độ tỉnh Hải Dương sử dụng thuật toán lan truyền ngược cùng thuật toán Bayes [15] và dự báo thời tiết tỉnh Vĩnh Phúc [16] đã thu được kết quả rất khả quan với tốc độ dự báo nhanh. Trong nghiên cứu sử dụng mô hình mạng nơron nhân tạo (ANN) để mô phỏng lưu lượng dòng chảy cho lưu vực sông Sêrêpôk, Đào Nguyên Khôi và Huỳnh Ái Phương đã thu được kết quả mô phỏng khá chính xác, đồng thời kết luận rằng việc lựa chọn các thông số đầu vào cho mô hình ANN cũng có ảnh hưởng lớn đến kết quả [17].

Khả năng đánh giá và mô phỏng chuỗi số liệu thời gian dựa trên đặc tính kế thừa (từ số liệu xảy ra trước đó) cho các yếu tố khí hậu là một vấn đề rất hữu ích. Hiểu biết về mối liên quan này có thể giúp chúng ta tiến hành diễn (khôi phục) số liệu bị mất hay đưa ra dự báo dựa trên số liệu thu thập được trong quá khứ. Cho đến nay, trên thế giới, hệ phương pháp phổ biến hay được sử dụng để đánh giá/mô phỏng phục vụ khảo sát tính chất tự tương quan trong chuỗi thời gian là các dạng biểu đồ điển hình như hàm tự tương quan (ACF) và hàm tự tương quan riêng phần (PACF), các mô hình như ARIMA hay GARCH (phi tuyến). Trong nghiên cứu của mình, Jafarian-Namin và cộng sự đã mô tả kết quả phân tích và dự báo khá chính xác nhiệt độ hàng tháng của Tehran (Iran) bằng việc kết hợp sử dụng các biểu đồ ACF, PACF để xác định tính dừng của chuỗi số liệu trước khi áp dụng mô hình ARIMA cho mô phỏng [18]. Ở một nghiên cứu khác, Murat và cộng sự đã phân tích những biểu đồ ACF và PACF chứa độ trễ bao trùm phạm vi thời gian thử nghiệm là 25 năm cho 4 khu vực ở châu Âu. Phương pháp mô phỏng bằng mô hình ARIMAF đã cho phép chứng minh những hành vi và động lực tương tự trong giá trị trung bình hàng ngày của dữ liệu nhiệt độ từ bốn vùng khí hậu được xem xét [19].

Với đà phát triển của cuộc cách mạng công nghiệp 4.0 trên thế giới, những ứng dụng ANN ngày càng nhận được sự quan tâm lớn hơn. Song hiện nay, số lượng các công trình đã được công bố về áp dụng trí tuệ nhân tạo để xử lý số liệu khí hậu ở khu vực Nam Bộ nước ta tạm thời còn chưa nhiều. Đặc biệt, theo như tác giả tìm hiểu, khả năng ước lượng cường độ của mức tự tương quan trong chuỗi thời gian của các yếu tố khí hậu bằng cách áp dụng kỹ thuật ANN hiện đang là vấn đề còn chưa được nghiên cứu kỹ trong các điều tra đặc điểm khí hậu ở Nam Bộ nước ta. Theo hướng nêu trên, nghiên cứu này có mục tiêu sử dụng ANN để đánh giá mức tự tương quan trong chuỗi thời gian của những yếu tố khí hậu tại các trạm khí tượng Cần Thơ và Nhà Bè trong vai trò đại diện cho 2 khu vực Tây và Đông Nam Bộ. Các yếu tố khí hậu được lựa chọn đưa vào khảo sát - về mặt lý thuyết - đều có khả năng chịu ảnh hưởng của hiện tượng tự tương quan. Kết quả khảo sát mức tự tương quan bằng ANN cũng đã được so sánh với kết quả phân tích PACF để đánh giá độ chính xác.

Những kết quả thu được đã góp phần bước đầu làm rõ triển vọng sử dụng ANN trong việc ước lượng quy luật biến thiên cũng như ảnh hưởng trễ theo thời gian của số liệu khí hậu vùng Nam Bộ. Mở rộng ra, kết quả nghiên cứu cũng có thể được dùng để tham khảo thêm trong công tác thiết kế quan trắc, mô phỏng - dự báo khí hậu phục vụ sản xuất nông nghiệp và các lĩnh vực liên quan khác ở nước ta.

## II. 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, khu vực Nam Bộ mang những nét đặc trưng của dạng khí hậu nhiệt đới gió mùa và cận xích đạo. Hàng năm, thời gian bức xạ dài với nhiệt độ và tổng tích ôn khá cao. Tuy nhiên, biên độ nhiệt ngày đêm giữa các tháng lại tương đối thấp.

Về cơ bản, khí hậu trong năm có thể được chia thành hai mùa chủ yếu là mùa khô (từ tháng 12 tới tháng 4 năm sau) và mùa mưa (từ tháng 5 đến tháng 11). Đặc điểm nổi bật là lượng mưa phân bố không

## ĐÁNH GIÁ MỨC TỰ TƯƠNG QUAN TRONG SỐ LIỆU KHÍ HẬU...

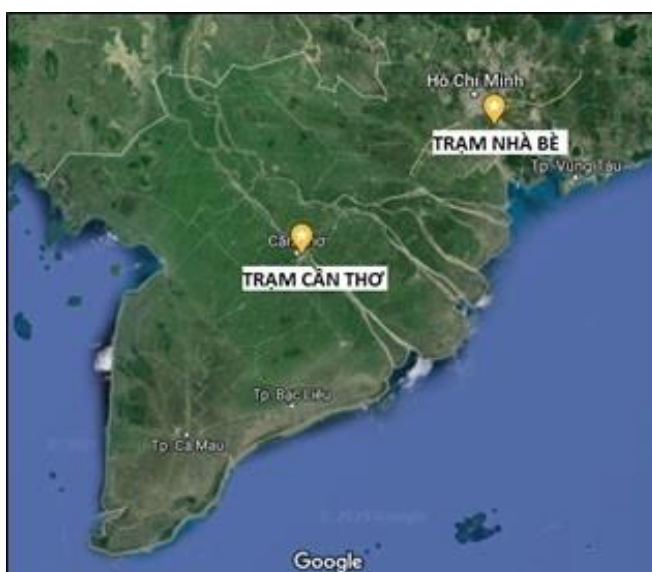
đồng đều. Trong đó, mưa mang xu hướng giảm dần từ vùng giáp Thành phố Hồ Chí Minh xuống khu vực phía Tây - Tây Nam.

### **b. Số liệu**

- Các trạm khí tượng:

Trạm khí tượng Nhà Bè: ở vùng Đông Nam Bộ, thuộc lưới trạm điều tra cơ bản của Trung tâm Khí tượng thủy văn Quốc gia, với tọa độ địa lý: 106°43'41" Kinh độ Đông; 10°39'36" Vĩ độ Bắc. 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.

Trạm khí tượng Cần Thơ: nằm ở vùng Tây Nam Bộ, thuộc lưới trạm điều tra cơ bản của Trung tâm Khí tượng thủy văn Quốc gia, với tọa độ địa lý: 105° 46' 06.48" Kinh độ Đông; 10° 01' 36.85" Vĩ độ Bắc. Ban đầu, trạm do Đài Khí tượng thủy văn Hậu Giang (cũ) quản lý. Từ tháng 6/1993, Đài Khí tượng thủy văn khu vực Nam Bộ được thành lập và quản lý trạm từ đó cho đến nay. Yếu tố quan trắc bao gồm những yếu tố khí tượng cơ bản và các hiện tượng thời tiết. Khoảng cách theo đường thẳng giữa trạm Nhà Bè và Cần Thơ gần bằng 130 km.



Hình 1. Vị trí các trạm khí tượng nghiên cứu ở khu vực Nam Bộ (Nguồn: Google map)

- Số liệu:

Số liệu được sử dụng trong bài báo này là kết quả quan trắc cho các yếu tố khí hậu của 2 trạm khí tượng nêu trên, được cung cấp bởi Đài khí tượng thủy văn khu vực Nam Bộ. Các yếu tố khí hậu đưa vào xem xét gồm: độ ẩm không khí (tương đối) trung bình, tổng lượng bốc hơi, tốc độ gió trung bình, nhiệt độ không khí trung bình, tổng lượng mưa. Bộ số liệu có thời gian 4 năm thuộc giai đoạn 1/1/2014 đến 31/12/2017 với tổng số 1461 ngày đo.

Các thiết bị sử dụng để đo đạc như sau:

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

### **2.2. Phương pháp**

Để thực hiện được các mục tiêu đề ra, tác giả đã thực hiện những bước xử lý và tính toán sau:

- Bước 1: Rà soát chất lượng các bộ số liệu đo ngày để loại bỏ các lỗi có thể xảy ra như số liệu rỗng hay số liệu bị lỗi do thiết bị đo gặp hỏng hóc gây ra, đồng thời thực hiện những phép thống kê cơ bản cho các yếu

tổ khí hậu đầu vào. Những thông số thống kê cơ bản cho mỗi yếu tố khí hậu là giá trị trung bình, các giá trị lớn nhất và nhỏ nhất, phương sai theo ngày để giúp đánh giá tổng quan số liệu.

- Bước 2: Kiểm tra vấn đề tự tương quan trong bộ số liệu cho mỗi yếu tố khí hậu, đồng thời đánh giá cụ thể mức tự tương quan. Phần mềm mã nguồn mở R (do R Development Core Team phát triển) đã được sử dụng trong bước tính toán này.

+ Sự tồn tại của vấn đề tự tương quan: kiểm định Box - Pierce cho phép đánh giá xem số liệu cho mỗi yếu tố khí hậu có xảy ra tự tương quan hay không. Trong đó, tiêu chí giá trị  $p$  nhỏ hơn 0.05 được sử dụng để chấp nhận rằng vấn đề tự tương quan là có ý nghĩa thống kê.

$$BP(k) = n \sum_{k=1}^K \rho_{a,k}^2$$

[1]

$\rho_{a,k}^2$  - hệ số tự tương quan ở độ trễ  $k$  của phần dư  $a_t$  ;

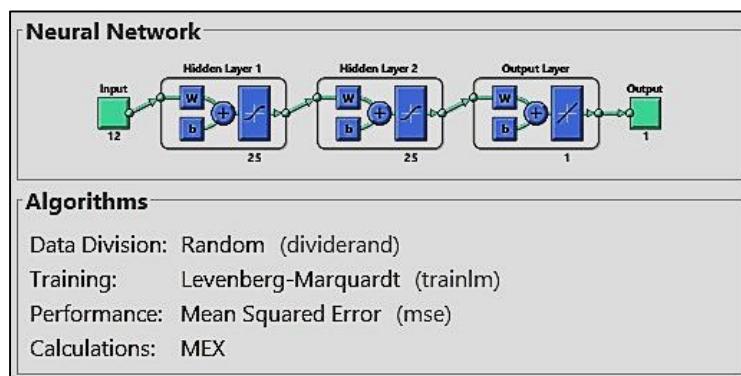
$n$  - số số hạng trong chuỗi sai phân;

$K$  - độ trễ tối đa đang được xem xét.

+ Nếu vấn đề tự tương quan được xác nhận rằng có ý nghĩa thống kê, mức tự tương quan cụ thể sẽ được khảo sát bằng việc phân tích biểu đồ hàm tự tương quan riêng phần (PACF) để xác định mức tự tương quan xảy ra giữa hai mốc thời gian, tức hai mốc ngày trong nghiên cứu này. Sự tự tương quan (cặp) nếu xảy ra giữa một ngày mốc với một ngày nào đó nằm trước nó  $m$  ngày thì cũng sẽ đúng cho ngày sau nó  $m$  ngày nếu lấy ngày nằm sau làm mốc chuẩn để tính giật lùi lại theo cách tương tự.

- Bước 3: Khảo sát mức tự tương quan trong các chuỗi số liệu khí hậu bằng ANN.

Dùng cùng bộ số liệu, công cụ ANN thuộc bộ công cụ Neural Network Toolbox của phần mềm Matlab (MathWorks, Inc.) được sử dụng để mô phỏng giá trị từng yếu tố khí hậu trong mỗi ngày dựa trên giá trị của chính yếu tố đó trong những ngày năm trước. Trong nghiên cứu này, tác giả dùng mạng ANN với cấu trúc 2 lớp ẩn có dạng  $n - 25 - 25 - 1$  (trong đó  $n$  là số biến đầu vào, tức số liệu cho những ngày năm trước ngày mốc). Thuật toán huấn luyện sử dụng là Levenberg - Marquardt back - propagation (lan truyền ngược) và số giai đoạn huấn luyện tối đa là 1000. Hàm chuyển "tansig" được sử dụng trong các lớp trung gian và hàm đầu ra là "purelin". Cấu trúc mạng này đã được tác giả lựa chọn sau một quá trình thử - sai với số liệu nghiên cứu khí hậu thực tế, mục đích nhằm đảm bảo rằng ANN dù tinh gọn nhưng vẫn đủ mạnh mẽ trong việc khai thác hết những thông tin tiềm ẩn giữa bộ biến đầu vào và ra.



Hình 2. Cấu trúc ANN dạng  $n - 25 - 25 - 1$  (cho trường hợp  $n = 12$ )

Các phương án mô phỏng bao gồm phương án đầu vào đủ các ngày trước ngày xem xét và những phương án trong đó thiếu một ngày trước đó nào đó (mỗi lần rút một biến đầu vào khác nhau ra). Mạng ANN được chạy 1000 lần cho mỗi phương án mô phỏng, ví dụ như phương án khuyết số liệu 1 ngày trước cho mô phỏng tốc độ gió, với việc chia ngẫu nhiên số liệu thành 3 phần phục vụ cho 3 tác vụ (học - xác nhận - kiểm tra) theo tỷ lệ 70 - 15 - 15. Các thông số đánh giá kết quả mô phỏng bao gồm giá trị hệ số tương quan  $R$  trung bình,  $R$  lớn nhất ( $R_{max}$ ),  $R$  nhỏ nhất ( $R_{min}$ ), độ chênh lệch giữa  $R_{max}$  và  $R_{min}$  (tính cho  $R_{all}$  của 1000 lần mô phỏng độc lập). Mức giá trị mà thông số đạt được cho mỗi phương án mô phỏng yếu tố khí hậu nào đó sử dụng ANN được tính theo công thức:

$$A = 100 \times \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

[2]

trong đó:

$A$  - mức phần trăm giá trị đạt được của thông số cho phương án mô phỏng nào đó so với giá trị lớn nhất thông số có thể đạt tới (%);

$X$  - giá trị thông số đạt được cho phương án mô phỏng đó;

$X_{\max}$  - giá trị thông số đạt được lớn nhất cho mọi phương án mô phỏng yếu tố khí hậu xem xét;

$X_{\min}$  - giá trị thông số đạt được nhỏ nhất cho mọi phương án mô phỏng yếu tố khí hậu xem xét.

Đối với thông số  $R$ , công thức [2] sẽ cho kết quả  $A$  càng nhỏ khi ngày rút ra có mức tương quan càng cao với ngày mốc xem xét (nghĩa là khi rút ngày đó ra thì lượng thông tin liên quan bị mất mát sẽ càng lớn hơn, dẫn đến mô phỏng kém chính xác và  $R$  giảm nhiều hơn) và ngược lại.

### III. KẾT QUẢ VÀ THẢO LUẬN

#### 3.1. Kết quả

##### a. Kiểm tra chất lượng các bộ số liệu khí hậu ngày

Bởi độ tin cậy khi tính toán phụ thuộc nhiều vào chất lượng của số liệu sử dụng nên 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.

Chuỗi số liệu khí hậu đo hàng ngày trong 4 năm (2014 - 2017) 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. Những thống kê về số liệu đầu vào được thể hiện trong bảng 1.

Bảng 1. Thống kê cho những yếu tố khí hậu đo được (theo ngày) tại trạm khí tượng Nhà Bè (ký hiệu là trạm 1) và trạm khí tượng Cần Thơ (ký hiệu là trạm 2)

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
Độ ẩm không khí trung bình	1	%	96.25	58.75	80.08	6.23
	2		96.00	61.25	80.62	5.91
Lượng bốc hơi	1	mm	6.60	0.30	2.90	1.11
	2		5.40	0.60	2.73	0.86
Tốc độ gió trung bình	1	m/s	4.50	0.00	1.18	0.63
	2		3.50	0.00	1.41	0.54
Nhiệt độ không khí trung bình	1	°C	31.78	22.53	27.79	1.46
	2		31.15	21.90	27.64	1.42
Tổng lượng mưa	1	mm	126.50	0.00	4.37	12.22
	2		111.40	0.00	4.77	11.12

##### b. Kiểm tra vấn đề tự tương quan trong bộ số liệu

Kết quả đánh giá vấn đề tự tương quan cho các chuỗi thời gian ngày đối với những thông số khí hậu xem xét sử dụng phần mềm  $R$  và kiểm định Box-Pierce được tóm tắt trong bảng 2. Kết quả đánh giá cho thấy biểu hiện của vấn đề tự tương quan luôn ở mức có ý nghĩa thống kê ( $p < 0.05$ ) trong chuỗi số liệu đối với tất cả các yếu tố khí hậu đang xem xét ở cả 2 trạm khí tượng. Như vậy, hiện tượng tự tương quan đã được xác nhận.

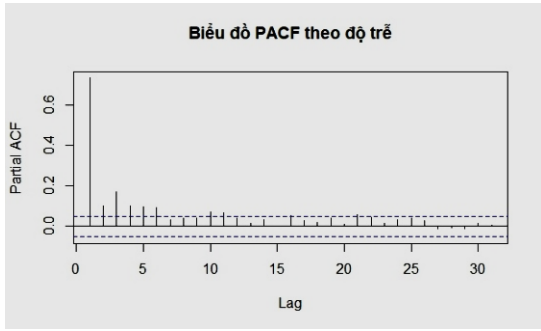
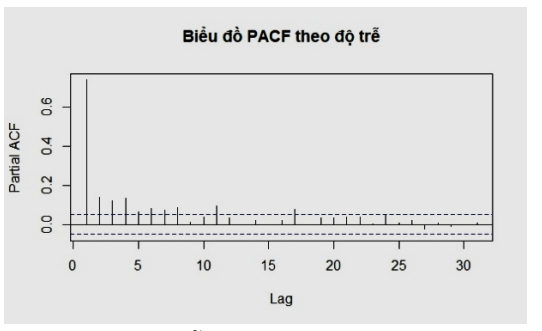
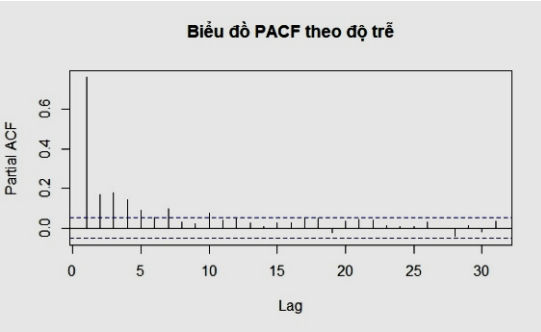
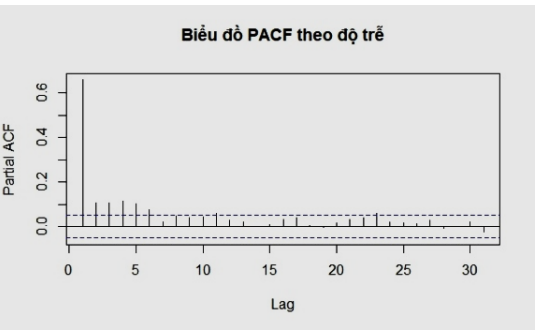
Bảng 2. Kết quả đánh giá vấn đề tự tương quan sử dụng phần mềm R và kiểm định Box-Pierce

Yếu tố khí hậu	Trạm	$\chi$ - bình phương	p	
Độ ẩm trung bình	1	781.27	< 2.2e-16	
	2	797.24		
Lượng bốc hơi	1	841.75		
	2	633.26		
Tốc độ gió trung bình	1	202.08		
	2	311.73		
Nhiệt độ trung bình	1	1058.8		
	2	999.84		
Tổng lượng mưa	1	63.06		1.998e-15
	2	22.528		2.071e-06

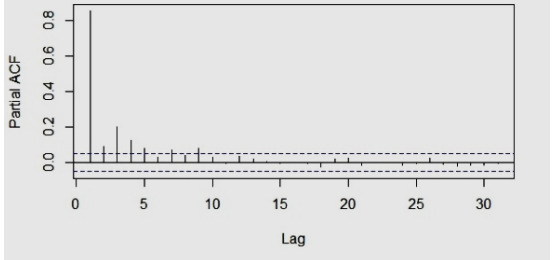
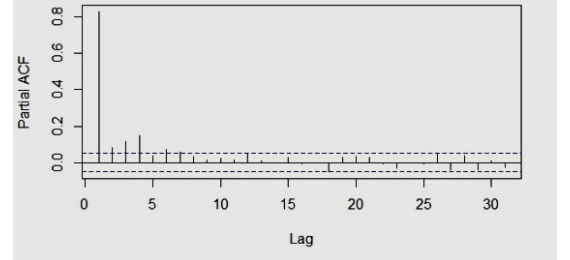
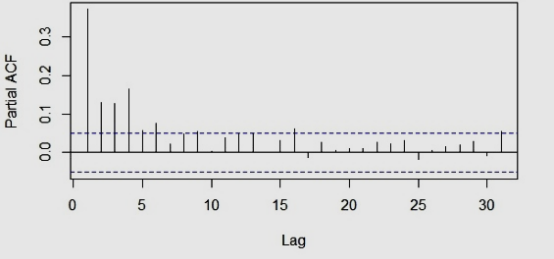
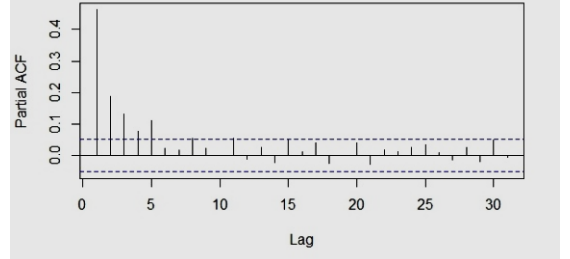
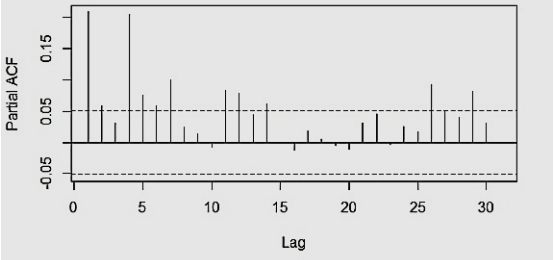
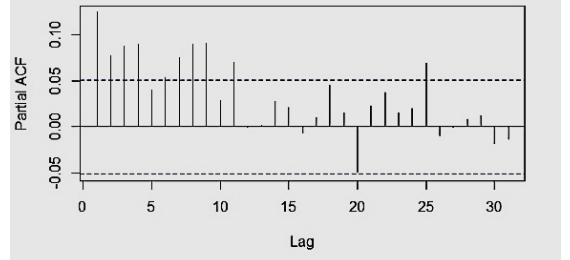
**c. Khảo sát mức tự tương quan bằng biểu đồ PACF**

Kết quả đánh giá cụ thể mức độ tự tương quan trong các chuỗi thời gian (ngày) đối với những thông số khí hậu đưa vào nghiên cứu trên phần mềm R bằng biểu đồ PACF được tóm tắt trong các bảng 3 dưới đây. Trong nghiên cứu này, từ “trễ” (lag) để chỉ thứ tự ngày nằm sau ngày đang xem xét, ví dụ như ngày trễ thứ 5 là nếu đếm xuôi thời gian sẽ là ngày thứ 5 xảy ra sau đó.

Bảng 3. Mức độ tự tương quan trong các chuỗi thời gian ngày đối với những thông số khí hậu thông qua biểu đồ PACF

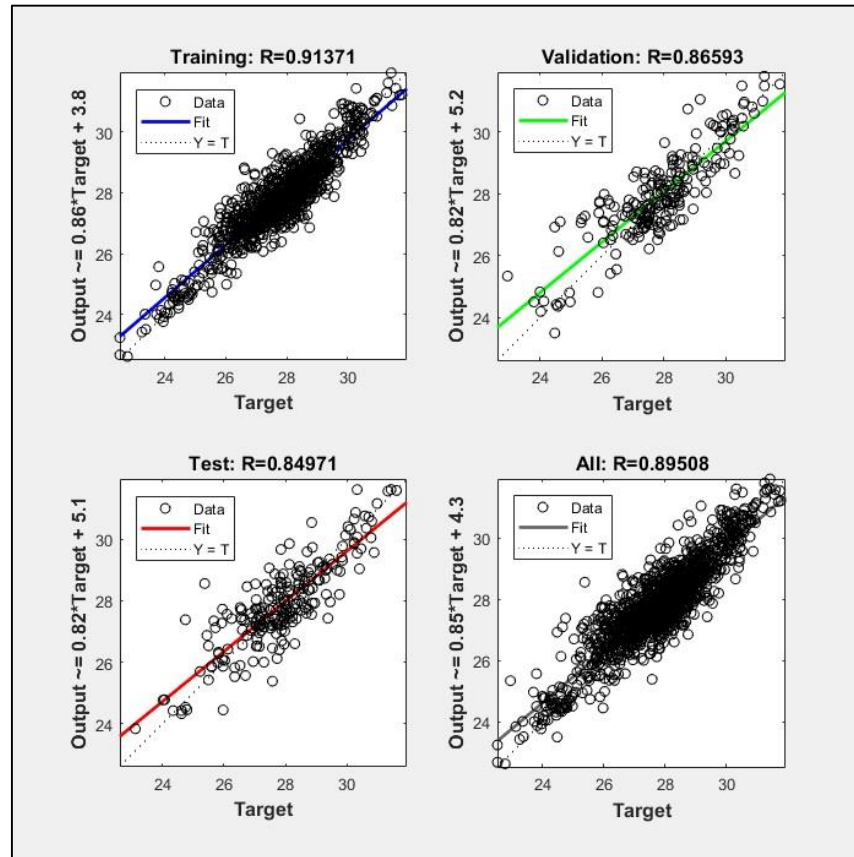
Yếu tố khí hậu	Biểu đồ PACF và nhận xét về số ngày trễ ở mức tương quan có ý nghĩa	
	Trạm 1	Trạm 2
Độ ẩm trung bình	 <p>Ngày trễ có ý nghĩa: 1 - 6</p>	 <p>Ngày trễ có ý nghĩa: 1 - 8, 11</p>
Lượng bốc hơi	 <p>Ngày trễ có ý nghĩa: 1 - 7, 10</p>	 <p>Ngày trễ có ý nghĩa: 1 - 6</p>

ĐÁNH GIÁ MỨC TỰ TƯƠNG QUAN TRONG SỐ LIỆU KHÍ HẬU...

<p>Nhiệt độ trung bình</p>	<p style="text-align: center;"><b>Biểu đồ PACF theo độ trễ</b></p>  <p style="text-align: center;">Ngày trễ có ý nghĩa: 1 - 5, 7, 9</p>	<p style="text-align: center;"><b>Biểu đồ PACF theo độ trễ</b></p>  <p style="text-align: center;">Ngày trễ có ý nghĩa: 1 - 4, 6</p>
<p>Tốc độ gió trung bình</p>	<p style="text-align: center;"><b>Biểu đồ PACF theo độ trễ</b></p>  <p style="text-align: center;">Ngày trễ có ý nghĩa: 1 - 6</p>	<p style="text-align: center;"><b>Biểu đồ PACF theo độ trễ</b></p>  <p style="text-align: center;">Ngày trễ có ý nghĩa: 1 - 5</p>
<p>Tổng lượng mưa 1</p>	<p style="text-align: center;"><b>Biểu đồ PACF theo độ trễ</b></p>  <p style="text-align: center;">Ngày trễ có ý nghĩa: 1 - 7, 11, 12, 14, 26, 29</p>	<p style="text-align: center;"><b>Biểu đồ PACF theo độ trễ</b></p>  <p style="text-align: center;">Ngày trễ có ý nghĩa: 1 - 4, 6 - 9, 11, 25</p>

**d. Kiểm tra mức tự tương quan bằng phương pháp mô phỏng sử dụng ANN**

Mô phỏng ANN được tiến hành cho tất cả các phương án dự kiến và kết quả về giá trị thông số được ghi nhận. Hình 3 dưới đây trình bày một ví dụ cho kết quả R nhận được cho một phương án khi mô phỏng nhiệt độ không khí trung bình tại trạm Nhà Bè.



Hình 3. Kết quả R khi mô phỏng ANN cho nhiệt độ không khí tại trạm Nhà Bè (phương án rút biến ngày thứ 9 ra khỏi nhóm biến đầu vào)

Kết quả thu được khi mô phỏng bằng ANN cho chuỗi thời gian ngày đối với từng yếu tố khí hậu đưa vào nghiên cứu dùng phương pháp tuần tự rút ra mỗi ngày trước đó trong số các biến đầu vào được tóm tắt trong bảng 4 dưới đây.

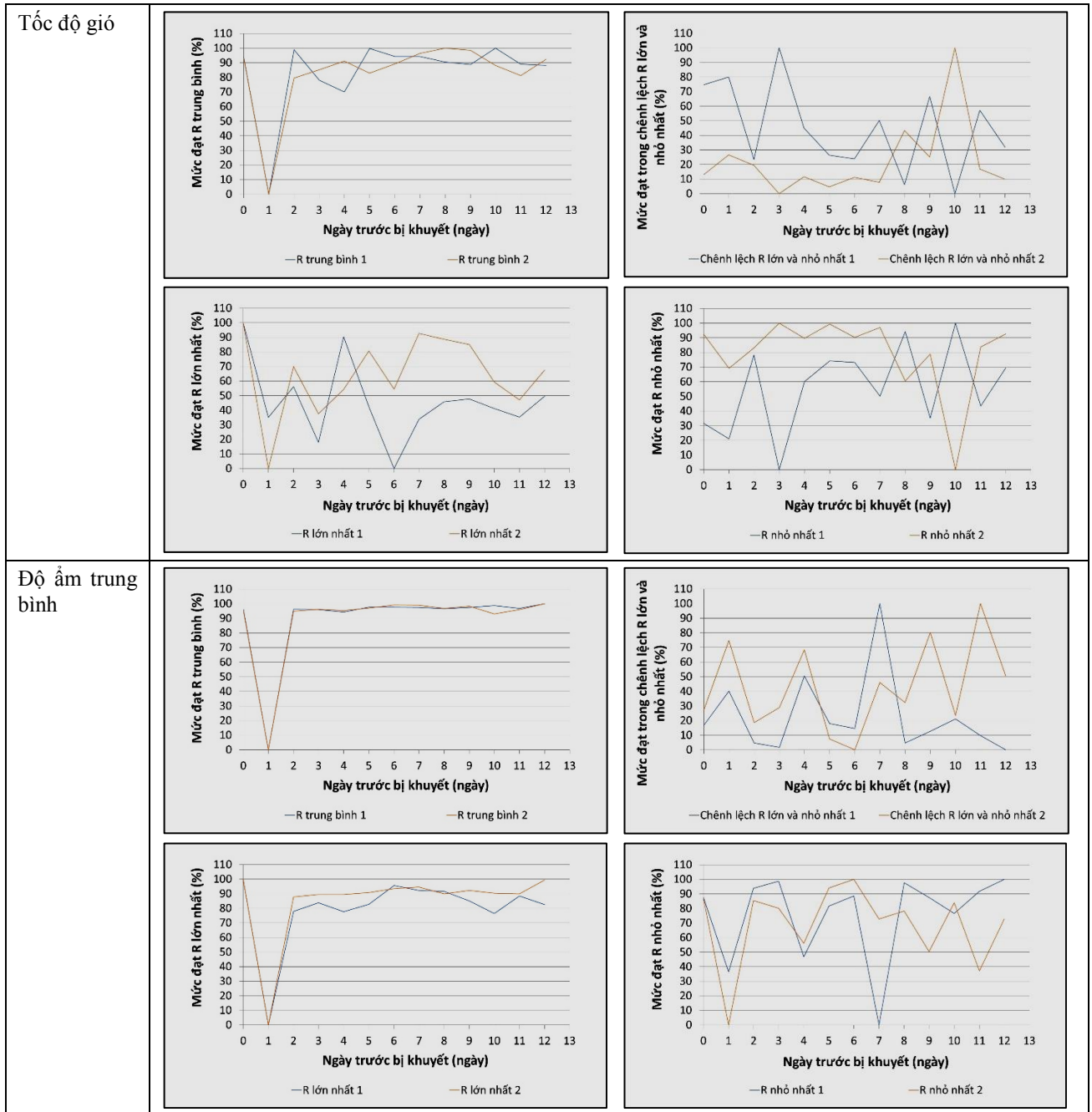
Bảng 4. Kết quả thu được khi mô phỏng bằng ANN chuỗi thời gian ngày cho các yếu tố khí hậu

Yếu tố khí hậu	Các thông số đánh giá cho kết quả chạy mô phỏng bằng ANN	
Nhiệt độ trung bình		



ĐÁNH GIÁ MỨC TỰ TƯƠNG QUAN TRONG SỐ LIỆU KHÍ HẬU...

<p>Lượng bốc hơi</p>		
<p>Tổng lượng mưa</p>		



### 3.2. Thảo luận

Kết quả nghiên cứu thu được cho phép rút ra một số nhận xét quan trọng sau:

- Kết quả phân tích chỉ ra rằng vấn đề tự tương quan tồn tại cho tất cả các yếu tố khí hậu đưa vào nghiên cứu. Điều này cho thấy tính phụ thuộc đặc điểm số liệu của một hay nhiều ngày hôm sau từ số liệu của mỗi ngày bất kỳ tại cả 2 trạm khí tượng khu vực đồng bằng Nam Bộ đưa vào nghiên cứu. Đây có thể là cơ sở cho các lập luận mô phỏng (hay điền số liệu mất) số liệu hiện tại dựa trên số liệu khí hậu quá khứ trong khu vực.

Nhận xét trên có thể hỗ trợ khẳng định thêm kết luận đã được đưa ra bởi Di Cecco và Gouhier (2018) khi các tác giả trên kiểm tra dữ liệu nhiệt độ không khí hàng ngày từ 21 mô hình hoàn lưu chung theo kịch bản phát thải carbon thông thường. Họ đã định lượng các mô hình tự tương quan về không gian và thời gian trong khoảng thời gian giai đoạn 1871 - 2099. Kết quả của nghiên cứu do hai nhà khoa học trên thực hiện cho thấy sự tự tương quan theo thời gian của nhiệt độ không khí thể hiện mức gia tăng ở đa số vùng trên trái đất dưới tác động của biến đổi khí hậu [20]. Tương tự, kết luận của nghiên cứu này cũng

thống nhất với nhận định đưa ra bởi Barnston và cộng sự khi các tác giả này mô tả kết quả phân tích số liệu mùa hè ở Hoa Kỳ kết hợp với mô hình mô phỏng lịch sử giai đoạn 1979 - 2005. Họ đã nhấn mạnh rằng có rất ít hoặc không có thay đổi nào trong sự tự tương quan của nhiệt độ cực đoan giữa quá khứ gần đây và tương lai ở khu vực xem xét [21].

- Đối với tất cả các yếu tố khí hậu trừ tổng lượng mưa, giá trị của yếu tố khí hậu tại ngày nằm ngay trước hay sau đó (ngày trước 1 như tên gọi trong nghiên cứu này) luôn có mối tương quan mạnh mẽ nhất đối với giá trị của yếu tố khí hậu tại ngày đang xem xét. Hiện tượng trên được thể hiện rất rõ trên biểu đồ PACF. Ngoài ra, kết quả cũng cho thấy rằng giới hạn 5 - 7 ngày kể từ ngày đang xét là khoảng trễ mà mức tự tương quan trong chuỗi số liệu của yếu tố khí hậu nào đó thường xảy ra đáng kể, tức giá trị cột PACF vượt ra ngoài hai đường giới hạn đứt nét có giá trị tuyệt đối bằng 0.05 trong các biểu đồ PACF. Điều này đúng với độ ẩm trung bình, lượng bốc hơi, nhiệt độ trung bình, tốc độ gió trung bình. Nguyên nhân gây ra hiện tượng trên rất có thể do xu thế biến thiên đồng điệu theo mùa của đa số các yếu tố khí hậu nên ngày sau sẽ có xu hướng thay đổi rất ít với ngày sát trước đó và mức tương quan thường tắt dần đến mức không đáng kể sau 1 tuần.

Tuy nhiên, riêng đối với tổng lượng mưa, dù số liệu mưa ngày ngay sau đó vẫn có mối tương quan mạnh nhất, nhưng với khoảng cách xa thậm chí đến khoảng 2 đến 4 tuần (15 - 30 ngày) sau thì mối tương quan vẫn còn khá mạnh mẽ với ngày xem xét, chứ không yếu hẳn đi như đã xảy ra với những yếu tố khí hậu khác. Nguyên nhân rất có thể do việc xảy ra mưa cũng như cường độ mưa đã tuân thủ chặt chẽ hơn theo quy luật biến đổi từ từ - trong cùng thời điểm mùa mưa hay nắng của năm - so với các yếu tố khí hậu còn lại.

Việc nắm bắt được cường độ cũng như đặc điểm tắt dần của sự ảnh hưởng tự tương quan theo thời gian đối với một yếu tố khí hậu nào đó có ý nghĩa thực tế khá quan trọng. Những nhận xét trên nếu được kiểm chứng bằng số liệu nghiên cứu ở nhiều trạm đo khác nữa thì có thể giúp ích cho việc thiết kế tần suất quan trắc các yếu tố khí hậu khác nhau sao cho phù hợp với đặc tính từng khu vực địa lý trong khi vẫn đảm bảo tuân thủ những quy định kỹ thuật của công tác quan trắc. Việc nắm vững quy luật nói trên cũng có ích với việc phục hồi số liệu đo thực địa bị mất tại các trạm khí tượng do những nguyên nhân khác nhau gây ra như hư hỏng thiết bị đo do tác động vật lý hay điện tử, máy đo bị hết năng lượng...

- Kỹ thuật ANN có ưu điểm rằng không đòi hỏi chuỗi số liệu theo thời gian phải dừng (Stationary) như yêu cầu của các mô hình thông thường khác mà điển hình là hệ phương pháp Box-Jenkins. Bằng cách sử dụng mạng ANN với kỹ thuật loại bỏ từng ngày số liệu nằm trước so với ngày mô phỏng, tác giả đã khảo sát sự biến thiên trong kết quả mô phỏng giá trị từng yếu tố khí hậu khi rút theo thứ tự số liệu mỗi ngày trước đó (khảo sát đến 12 ngày sau theo gọi ý tầm ảnh hưởng từ kết quả PACF). Trừ yếu tố tổng lượng mưa, kết quả cho thấy giá trị hệ số tương quan lớn nhất ( $R_{max}$ ) khi chạy 1000 lần mô phỏng độc lập là thông số khá ổn định trong việc chỉ ra mức độ tương quan của số liệu ngày nằm trước đầu tiên, nhưng lại không nêu chính xác được mức độ ảnh hưởng (tương quan) của các ngày xảy ra trước lâu hơn đối với số liệu mỗi yếu tố khí hậu trong ngày đang xem xét. Các thông số đánh giá mô phỏng ANN khác như R trung bình của nhiều lần mô phỏng,  $R_{min}$ , chênh lệch giữa  $R_{max}$  và  $R_{min}$  cũng đều không chỉ ra được mức tương quan thống nhất với kết quả PACF (dùng phần mềm R nói ở phần trên). Lưu ý rằng giá trị PACF ngày nằm trước (hay sau) đầu tiên đối với 4 yếu tố khí hậu gồm độ ẩm trung bình, lượng bốc hơi, nhiệt độ trung bình, tốc độ gió trung bình đều ở mức giá trị 0.4 - 0.8 và thường lớn gấp 2 - 5 lần so với giá trị PACF lớn kế nó.

Việc không một thông số mô phỏng ANN nào, kể cả  $R_{max}$ , có hiệu quả rõ rệt khi đánh giá mức độ tự tương quan cho tổng lượng mưa tại 2 trạm khí tượng Nhà Bè và Cần Thơ có thể liên quan đến độ nhạy trong việc phát hiện tương quan của phương pháp còn chưa cao hoặc chưa nắm bắt được đặc tính ngẫu nhiên phức tạp của các hiện tượng thời tiết. Điều này có thể thấy rõ khi kết hợp việc quan sát biểu đồ PACF: giá trị PACF trong ngày trễ đầu tiên của tổng lượng mưa chỉ đạt khoảng 0.15 và đặc biệt là chênh lệch không nhiều so với giá trị PACF lớn kế nó.

Tóm lại, hướng tiếp cận chạy mô phỏng ANN kết hợp đánh giá sự biến thiên trong thông số  $R_{max}$  cho phép phát hiện ở dạng đề xuất cho mức tự tương quan trong chuỗi số liệu thời gian của một số yếu tố khí hậu, tuy nhiên độ nhạy hiện còn chưa cao. Phân tích cũng cho thấy mức chênh lệch PACF của các ngày trễ (hay nằm trước) so với ngày nghiên cứu để phương pháp này phát huy tác dụng có thể phải nằm trong ngưỡng tối thiểu là trên 2 lần so với giá trị PACF lớn kế nó, đồng thời giá trị PACF muốn được phát hiện cũng có khả năng cần phải lớn hơn 0.4. Ở mức giá trị và sự chênh lệch PACF thấp (như yếu tố tổng lượng mưa) thì hiệu quả phát hiện còn yếu. Với tiềm năng hỗ trợ và phối hợp của ANN với những công cụ tính

toán truyền thống như chuỗi Box-Jenkins, đây là vấn đề cần được nghiên cứu kỹ hơn và cải tiến thêm để có thể thu được kết quả tin cậy ổn định trong thực tế. Về những nhận xét này, tác giả không tìm được thông tin tương tự về cùng chủ đề ở các nghiên cứu đã công bố khác nên không thể so sánh đánh giá sự khác biệt hay tương đồng.

- Giá trị R trung bình tính cho 1000 lần mô phỏng độc lập sử dụng mỗi cấu trúc ANN với cùng bộ số liệu cho thấy ảnh hưởng của việc chia số liệu ngẫu nhiên theo tỷ lệ được định trước (ở đây là 70 - 15 - 15) và cường độ tự tương quan của số liệu đưa vào tính toán. Theo lý thuyết, khi rút số liệu một ngày nào đó ra khỏi bộ số liệu biến đầu vào sẽ làm giảm một lượng thông tin đóng góp nhất định đến việc mô phỏng biến đầu ra và dẫn đến giá trị R sẽ phải giảm đi. Nhưng điều này lại không thể hiện rõ ràng trong kết quả mô phỏng cho một số trường hợp đối với các yếu tố tốc độ gió và tổng lượng mưa.

Trong mô phỏng ANN, nếu tính đến khoảng giá trị R thu được thấp khi mô phỏng 2 yếu tố khí hậu nêu trên (vào khoảng 0.5 - 0.7) so với mức R của các yếu tố khí hậu còn lại (thường là 0.75 - 0.85) thì rất có thể mức R thể hiện cho cường độ tương quan thấp có liên hệ chặt chẽ đến tính ổn định trong kết quả khảo sát khi chạy mô phỏng nhiều lần. Nguyên nhân có khả năng do mối liên hệ vốn yếu hơn trong bộ số liệu ban đầu đã dẫn đến không những mức giá trị R thấp mà tính ổn định của mô phỏng cũng kém hơn khi bộ số liệu bị chia ngẫu nhiên nhiều lần. Đánh giá này có thể hữu ích khi xem xét việc chấp nhận độ lớn tối thiểu của giá trị R nhận được lúc mô phỏng, sao cho phải đảm bảo được độ ổn định trong kết quả mô phỏng nhiều lần tức gián tiếp nâng cao độ tin cậy với bộ số liệu đầu vào được phân chia ngẫu nhiên.

#### IV. KẾT LUẬN

Công tác đánh giá hiện tượng tự tương quan cho số liệu đo tại 2 trạm khí tượng Nhà Bè và Cần Thơ đã chứng minh rằng vấn đề tự tương quan tồn tại ở mức có ý nghĩa thống kê đối với tất cả các yếu tố khí hậu đưa vào xem xét. Tuy nhiên, mức tự tương quan lại không duy trì được theo thời gian và thường tắt dần khi cặp ngày xem xét càng xa nhau. Ngoại lệ là trường hợp tổng lượng mưa có mức tự tương quan xa đến 30 ngày (khoảng 4 tuần), tất cả các yếu tố khí hậu còn lại đều chỉ giới hạn phạm vi thời gian xảy ra tự tương quan tối đa là 5 - 7 ngày kể từ ngày đang xét.

Với kỹ thuật loại bỏ từng ngày số liệu đầu vào (xảy ra sớm so với ngày mô phỏng) khi đưa vào mô phỏng ANN, tác giả đã khảo sát sự biến thiên trong kết quả mô phỏng giá trị từng yếu tố khí hậu. Trừ yếu tố tổng lượng mưa, kết quả cho thấy giá trị hệ số tương quan lớn nhất ( $R_{max}$ ) của 1000 lần mô phỏng độc lập là thông số khá ổn định khi đánh giá mức độ tương quan cho số liệu ngày năm kế đó, dù vẫn chưa xác định chính xác mức ảnh hưởng (tự tương quan) của các ngày năm xa hơn. Kết hợp với những biểu đồ PACF, điều này còn cho phép đi đến lập luận rằng mức chênh lệch PACF của các ngày trước (hay trễ) so với ngày nghiên cứu để phương pháp dùng ANN phát huy tác dụng có thể phải nằm trong ngưỡng tối thiểu là trên 2 lần so với giá trị PACF lớn kế nó, đồng thời giá trị PACF muốn được phát hiện có khả năng cũng cần phải lớn hơn 0.4. Bên cạnh đó, kết quả nghiên cứu cũng cho thấy mối liên hệ tự tương quan nếu ở mức yếu trong bộ số liệu khí hậu ban đầu sẽ dẫn đến không chỉ mức giá trị R thấp mà tính ổn định trong kết quả mô phỏng cũng kém hơn khi bộ số liệu được phân chia ngẫu nhiên nhiều lần để mô phỏng.

Nghiên cứu này hiện mới chỉ dùng 4 năm số liệu thực đo tại 2 trạm khí tượng Nam Bộ. Do đó, để có thể tổng quát hóa vấn đề, kết quả rất cần được thử nghiệm và kiểm tra thêm bằng chuỗi số liệu có thời gian dài hơn thu thập từ những trạm khí tượng ở các địa phương có đặc điểm tự nhiên khác nhau. Bên cạnh đó, những cấu trúc và kỹ thuật ANN khác cũng rất cần được bổ sung đánh giá thêm, nhất là cụ thể hóa các tiêu chuẩn để làm rõ khoảng thời gian trễ nào còn có ý nghĩa trong mô phỏng, để khai thác triệt hơn tiềm năng của bộ công cụ này trong khảo sát hiện tượng tự tương quan.

#### 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.

#### TÀI LIỆU THAM KHẢO

## ĐÁNH GIÁ MỨC TỰ TƯƠNG QUAN TRONG SỐ LIỆU KHÍ HẬU...

- [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] 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.
- [5] O. Kissi, “Generalized regression neural networks for evapotranspiration modeling”, *Hydrological Sciences*, 51(6), pp. 1092-1105, 2006.
- [6] 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.
- [7] G. Landaras, 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] M. Hayati, and Z. Mohebi, Application of Artificial Neural Networks for Temperature Forecasting, *World Academy of Science, Engineering and Technology International Journal of Electrical and Computer Engineering*, 1(4), pp. 662-666, 2007.
- [11] S. S. Baboo and I. K. Shereef, An Efficient Weather Forecasting System using Artificial Neural Network, *International Journal of Environmental Science and Development*, 1(4), pp. 321-325, 2010.
- [12] A. Kaur, H. Singh, Artificial Neural Networks in Forecasting Minimum Temperature, *International Journal of Electronics & Communication Technology*, 2(3), pp. 101-105, 2011.
- [13] E. Günal, M. Budak, M. Kılıç, B. Cemek, and M. Sirri, Combining spatial autocorrelation with artificial intelligence models to estimate spatial distribution and risks of heavy metal pollution in agricultural soils, *Environmental Monitoring and Assessment*, 195(2), 2023.
- [14] H. Kim, K. B. A. Mawuntu, T. Park, H. Kim, J. Park, and Y. Jeong, Spatial Autocorrelation Incorporated Machine Learning Model for Geotechnical Subsurface Modeling, *Applied Science*, 13, p. 4497, 2023.
- [15] 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.
- [16] 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.
- [17] N. K. Dao and A. P. Huynh, Flow simulation for Serepok river basin with Artificial neural network, *Science & Technology Development*, 19(3), pp. 114-124, 2016.
- [18] S. Jafarian-Namin, D. Shishebori, and A. Goli, Analyzing and Predicting the Monthly Temperature of Tehran using ARIMA Model, Artificial Neural Network, and Its Improved Variant, *Journal of Applied Research on Industrial Engineering*, 2023 (in press; doi: 10.22105/jarie.2023.356297.1502).
- [19] M. Murat, I. Malinowska, M. Gos and J. Krzyszczyk, Forecasting daily meteorological time series using ARIMA and regression models, *International Agrophysics*, 32, pp. 253-264, 2018.
- [20] G. J. Di Cecco, T. C. Gouhier, Increased spatial and temporal autocorrelation of temperature under climate change, *Scientific Reports*, 8, 14850, 2018 (<https://doi.org/10.1038/s41598-018-33217-0>).
- [21] A. G. Barnston, B. Lyon, E. D. Coffel, and R. M. Horton, Daily Autocorrelation and Mean Temperature/Moisture Rise as Determining Factors for Future Heat-Wave Patterns in the United States, *American Meteorological Society*, 59, pp. 1735-1754, 2020.

**ASSESSMENT OF CLIMATE DATA AUTO - CORRELATION AT TWO  
METEOROLOGICAL STATIONS IN NAM BO PLAIN USING PARTIAL  
AUTOCORRELATION FUNCTION (PACF) AND ARTIFICIAL NEURAL NETWORK  
(ANN)**

TRAN TRI DUNG

*Institute of Environmental Science, Engineering and Management - Industrial University of Ho Chi Minh  
City*

*trantridung@iuh.edu.vn*

**Abstract.** This study used Artificial Neural Networks (ANN) and Partial autocorrelation function (PACF) graphs to evaluate auto - correlation level in time series of climate factors measured during 2014 - 2017 period at Can Tho and Nha Be meteorological stations. The ANN structure has 2 hidden layers of the form  $n - 25 - 25 - 1$ , Levenberg-Marquardt backpropagation training algorithm with "tansig" transfer function for hidden layers. The results had demonstrated that auto - correlation problem exists at a statistically significant level for all climate factors considered, but is often damped (with a maximum interval of 5 - 7 days, especially up to 30 days for total rain). Except for total rainfall factor, the maximum correlation coefficient value ( $R_{max}$ ) of 1000 independent ANN simulations was relatively effective to evaluate the degree of correlation for the first lagged (or earlier) day data, although it was still not able to clearly determine the influence of further days. For ANN detection to be effective, the PACF value of a day lagged (or earlier) to the date considered will probably have to be higher than 0.4 and at least 2 times higher than the next largest PACF value.

**Key-words.** auto - correlation, climate factors, Artificial Neural Network (ANN), Partial autocorrelation function (PACF), simulation.

*Ngày gửi bài: 05/12/2023*

*Ngày chấp nhận đăng: 02/04/2024*