

# ẢNH HƯỞNG CỦA VIỆC SỬ DỤNG HÀM PHẠT ĐẾN KẾT QUẢ KHẢO SÁT ĐIỂM THAY ĐỔI CHO SỐ LIỆU NHIỆT ĐỘ KHÔNG KHÍ TẠI CÁC TRẠM KHÍ TƯỢNG NHÀ BÈ VÀ CẦN THƠ

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*

**Tóm tắt.** Nghiên cứu này có mục tiêu đánh giá sự khác biệt trong kết quả khảo sát điểm thay đổi cho số liệu nhiệt độ không khí thuộc loại trung bình ngày và trung bình tuần khi sử dụng các hàm phạt khác nhau. Gói phần mềm "changepoint" trong bộ phần mềm R được sử dụng để điều tra ảnh hưởng của hàm phạt tới kết quả đánh giá các điểm thay đổi của nhiệt độ không khí giai đoạn 2013 - 2017 tại các trạm khí tượng Nhà Bè và Cần Thơ. Các kết quả thu được bằng phương pháp PELT (Pruned Exact Linear Time) và SegNeigh (Segment Neighbourhood, để kiểm tra) là như nhau. Trong khi đó, số lượng điểm thay đổi được tìm thấy tăng dần theo thứ tự đối với các hàm phạt SIC, HQIC và AIC. Việc áp dụng nguyên tắc đồ thị khuỷu tay và so sánh (trực quan) tính logic của kết quả phân chia điểm thay đổi của từng hàm với đường cong thay đổi nhiệt độ theo thời gian trên đồ thị đã chứng minh rằng hàm SIC cho số lượng điểm thay đổi hợp lý nhất. Kết quả cũng cho thấy việc lựa chọn mức thời gian (ngày hoặc tuần) cho dữ liệu đầu vào sẽ ảnh hưởng đến kết quả điều tra điểm thay đổi. Đối với dữ liệu dạng tuần, các điểm thay đổi trong năm được phát hiện có sự trùng hợp rất cao (83,3 %, với độ lệch cho phép là 1 tuần) giữa các trạm Nhà Bè và Cần Thơ, trong khi tỷ lệ này lại thấp hơn nhiều nếu dữ liệu đầu vào thuộc về dạng ngày.

**Từ khóa.** Nhiệt độ không khí, trạm khí tượng, điểm thay đổi, phạt, SIC, HQIC, AIC.

## EFFECT OF PENALTY FUNCTION USE ON CHANGE - POINT INVESTIGATION RESULTS FOR AIR TEMPERATURE DATA AT NHA BE AND CAN THO METEOROLOGICAL STATIONS

**Abstract.** This study aimed at the difference in change - point detection results for daily and weekly air temperature data using various penalty functions. Package "changepoint" in R was used to investigate the effect of penalty on the results of air temperature change - point assesment by mean for the period of 2013 ÷ 2017 at Nha Be and Can Tho meteorological stations. The results obtained by using PELT (Pruned Exact Linear Time) and SegNeigh (Segment Neighbourhood, for checking purposes) methods were identical. Meanwhile, the number of change - points found was increased gradually for SIC, HQIC and AIC penalty functions, respectively. The use of elbow - plot principle in combination with curve visual evaluation on graph demonstrated that the SIC function gives most reasonable number of change - points. It was also shown that the selection of time level (day or week) for input data will affect change - point investigation results. For weekly data, the change - points during the year were found to have very high coincidence (83.3 %, with the allowed deviation of 1 week) between Nha Be and Can Tho stations while that ratio was much lower if inputs belongs to daily type.

**Keywords.** Air temperature, meteorological stations, change - point, penalty, SIC, HQIC, AIC.

### 1. MỞ ĐẦU

Nhiệt độ không khí - một nhân tố khí tượng tự nhiên - đóng vai trò hết sức quan trọng chi phối thời tiết và có ảnh hưởng đến nhiều lĩnh vực kinh tế - xã hội như sản xuất nông nghiệp, chăn nuôi, giao thông vận tải, y tế. Bởi vậy, việc đánh giá sự thay đổi của yếu tố này là rất hữu ích và luôn nhận được sự quan tâm của các nhà khoa học trong nhiều thập kỷ qua. Nhiều công trình nghiên cứu về sự thay đổi của nhiệt độ không khí cùng các nguyên nhân gây ra đã được công bố trên thế giới. Ở nước ta, các kỹ thuật tiên tiến cũng đã được áp dụng hiệu quả trong hoạt động điều tra nghiên cứu những đặc trưng khí hậu trong đó có nhiệt độ [1, 2].

Bên cạnh những thông tin phổ biến như giá trị trung bình hay các giá trị cực đại, cực tiểu, một đặc trưng quan trọng của chuỗi số liệu chính là thời điểm thay đổi. Do đó, việc xác định điểm thay đổi trong các chuỗi số liệu, trong đó có số liệu thuộc loại khí tượng - viễn thám, cũng ngày càng nhận được sự chú ý lớn hơn trên lý thuyết [3, 4], cũng như trong công tác điều tra thực tiễn [5, 6, 7, 8, 9]. Kiến thức về kỹ thuật xác định điểm thay đổi có ý nghĩa khoa học quan trọng bởi nó cho phép khảo sát chính xác hơn quy luật biến động nhiệt độ hàng năm, từ đó giúp các nhà khoa học và quản lý nắm vững đặc điểm biến đổi đặc trưng của khí hậu ở từng vùng miền. Bên cạnh đó, nhiệt độ còn là yếu tố khí hậu chi phối nhiều hiện tượng khác như bốc hơi hay hạn hán nên những kiến thức này đồng thời cũng là cơ sở thông tin hết sức hữu ích phục vụ cho sản xuất nông nghiệp (như chọn thời điểm phù hợp để cung cấp nước cho đồng ruộng) và những ngành nghề liên quan.

Để xác định điểm thay đổi trong chuỗi số liệu, nhiều tác giả đã mô tả các kỹ thuật khác nhau như kỹ thuật kiểm định Pettitt [10], thay đổi cấu trúc [11], phương pháp dựa trên kiểm định Cramér von Mises [12], kiểm định tỷ số von Neumann và kiểm định phạm vi Buishand [13]. Trong nghiên cứu khảo sát hiệu quả hoạt động của các phương pháp trên số liệu viễn thám, Militino và cộng sự đã đi đến kết luận về nhiều ưu điểm của kiểm định Mann - Kendall đầu rằng kỹ thuật này không thể xác định giai đoạn thay đổi đột ngột như các kỹ thuật khác như kiểm định Pettitt, thay đổi cấu trúc hay phân chia thứ bậc (E.divisive) [11]. Bên cạnh đó, để đối phó với vấn đề số lượng điểm thay đổi tăng đáng kể khi chiều dài các bộ số liệu tăng, một số thuật toán hữu hiệu từng thông qua thực tế kiểm định được khuyến nghị là SegNeigh (Segment Neighbourhood) [14] và PELT (Pruned Exact Linear Time) [15]. Trong nghiên cứu độ cao sóng đo theo giờ ở biển Bắc Atlantic, Canada giai đoạn 2012 ÷ 2015, Killyck và cộng sự [15] đã kết luận phương pháp PELT xác định điểm thay đổi hợp lý (tức là phân đoạn tốt cho những khu vực biến động khác nhau) hơn so với các phương pháp được sử dụng phổ biến khác trong thực tế như BS (Binary Segmentation). Một nghiên cứu khác do Wambui và cộng sự tiến hành cũng đã mô tả phương pháp PELT có độ chính xác cao hơn phương pháp BS và nhanh hơn so với các phương pháp tìm kiếm chính xác khác, cũng như lực của kiểm định duy trì gần như giống nhau ở tất cả các vị trí thay đổi với một kích thước thay đổi nhất định và tăng lên cùng với sự gia tăng kích thước thay đổi trong số liệu [8].

Trong những phương pháp phổ biến để ước lượng điểm thay đổi, hàm phạt thường có vai trò khống chế số lượng các điểm thay đổi trong bộ số liệu. Một số tác giả đã phát triển các hàm phạt như BIC - Bayesian information criterion), AIC (Akaike information criterion), HQIC (Hannan - Quinn information criterion) và áp dụng chúng một cách hữu hiệu trong nhiều nghiên cứu [16, 17, 18, 19, 20, 21]. Các tác giả Gayawan và Ipinoyomi sau khi nghiên cứu trên bộ dữ liệu sinh sản cụ thể theo độ tuổi thu thập tại một số nước châu Phi - châu Âu đã kết luận những mô hình mô phỏng khớp tốt nhất với số liệu thực tế thu được từ các hàm AIC và SIC khác nhau chỉ khoảng 5% [22]. Beaulieu và cộng sự đã sử dụng hàm SIC rất hiệu quả để hỗ trợ công tác phát hiện các biến đổi đột ngột của số liệu khí hậu trong cân nhắc đánh đổi giữa việc tăng cơ hội phát hiện sự thay đổi và giảm rủi ro phát hiện sự thay đổi trong khi thực tế là không có. Các tác giả trên cũng đã mô tả rằng chúng ta có xác suất cao để tìm ra những thay đổi đột ngột với cường độ lớn hơn độ lệch chuẩn trong chuỗi quan sát, nhất là khi thay đổi có cường độ rất lớn (gấp ba lần độ lệch chuẩn) thì chỉ sau 4 năm xác suất của phạm vi phát hiện trong nhiều trường hợp thậm chí có thể đạt 100% [23].

Bởi đây là phương pháp tiếp cận có ứng dụng thiết thực và đã chứng minh được hiệu quả trên các hoạt động thực tiễn, việc đánh giá ảnh hưởng của giá trị phạt đến kết quả ước lượng điểm thay đổi là một vấn đề hữu ích. Bằng việc kiểm nghiệm hàm phạt phù hợp cho các loại số liệu, độ chính xác và tính hợp lý của kết quả thu được từ việc sử dụng kỹ thuật đánh giá điểm thay đổi sẽ có thể được nâng cao hơn.

Hiện nay, số lượng các công trình đã được công bố về xác định điểm thay đổi trong số liệu nhiệt độ ở khu vực Nam Bộ nước ta tạm thời còn chưa nhiều. Đặc biệt, những tác động của hàm phạt trong các công trình loại trên cũng đang còn ít được đề cập tới. Nghiên cứu này có mục tiêu đánh giá tác động của các hàm và khoảng giá trị phạt khác nhau đến kết quả xác định điểm thay đổi trong chuỗi số liệu nhiệt độ quan trắc tại 2 trạm khí tượng Nhà Bè và Cần Thơ trực thuộc Đài Khí tượng Thủy văn Khu Vực Nam Bộ. Những kết quả thu được sẽ góp phần làm rõ thêm triển vọng áp dụng kỹ thuật xác định điểm thay đổi trong nghiên cứu khí tượng ở nước ta, đồng thời có thể sử dụng như tư liệu tham khảo trong công tác đánh giá số liệu khí tượng cho các mục đích khác nhau.

## 2. SỐ LIỆU VÀ PHƯƠNG PHÁP

### Số liệu

Trạm khí tượng Cần Thơ ở vùng Tây Nam Bộ, thuộc lưới trạm điều tra cơ bản của Trung tâm Khí tượng Thủy văn Quốc gia, nằm tại đường 3/2, phường Xuân Khánh, Cần Thơ, với tọa độ địa lý như sau:  $105^{\circ}46'06.48''$  Kinh độ Đông;  $10^{\circ}01'36.85''$  Vĩ độ Bắc. Trạm bắt đầu quan trắc lấy số liệu từ 01/01/1978 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 cho đến nay.

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, nằm tại ấp 3, xã Long Thới, huyện Nhà Bè, với tọa độ địa lý như sau:  $106^{\circ}43'41''$  Kinh độ Đông;  $10^{\circ}39'36''$  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.

Khoảng cách theo đường thẳng giữa 2 trạm xấp xỉ 130 km. Yếu tố quan trắc tại 2 trạm trên bao gồm những yếu tố khí tượng cơ bản và các hiện tượng thời tiết.



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

Số liệu được sử dụng trong bài báo này là kết quả quan trắc cho nhiệt độ không khí ngày của các trạm khí tượng Cần Thơ và Nhà Bè được cung cấp bởi Đài khí tượng Thủy văn khu vực Nam Bộ. Số liệu mỗi tuần được tính toán bằng trung bình cộng số liệu các ngày trong tuần đó. Bộ số liệu có thời gian 5 năm thuộc giai đoạn 1/1/2013 đến 31/12/2017. Thiết bị sử dụng để đo đạc nhiệt độ không khí là nhiệt kế khô 126906.

### Phương pháp

Số liệu thô của nhiệt độ không khí đo được tại 2 trạm khí tượng Nhà Bè và Cần Thơ đã được đưa vào xử lý để khảo sát các điểm thay đổi trong chuỗi số liệu. Điểm thay đổi được đánh giá cho từng năm trong giai đoạn khảo sát (2013 ÷ 2017) với mỗi trạm khí tượng, đồng thời cũng được đánh giá riêng cho 2 loại số liệu trung bình ngày và trung bình tuần.

Số liệu thu thập được kiểm tra chất lượng để loại bỏ các lỗi có thể xảy ra do những nguyên nhân khác nhau như máy đo bị trục trặc trong quá trình hoạt động. Trong đó, một số thống kê mô tả cơ bản được tính toán để khảo sát những đặc điểm chung của sự biến đổi số liệu nhiệt độ theo thời gian trong giai đoạn đang xem xét.

Gói phần mềm “changepoint” của tập thể các tác giả Killick và cộng sự (2014, cập nhật 2016) trong bộ phần mềm R (phiên bản 3.6.2) được áp dụng cho tính toán điểm thay đổi. Trong đó, tiêu chuẩn xác định điểm thay đổi là giá trị trung bình (mean) của nhiệt độ không khí. Để tìm ra các điểm thay đổi, tác giả đã dùng hàm `cpt.mean` trong gói phần mềm nói trên. Những thông số sử dụng cho hàm `cpt.mean` trong nghiên cứu này lần lượt được khảo sát, bao gồm:

- Phương pháp (thuộc loại chính xác): PELT hoặc SegNeigh. Các phương pháp này được chọn bởi khả năng thường xảy ra nhiều điểm thay đổi trong số liệu nhiệt độ của mỗi năm. Đặc biệt, phương pháp PELT có khả năng áp dụng tương thích với nhiều loại hàm phạt hết sức đa dạng. Độ chính xác bằng cách sử dụng PELT cũng cao hơn hẳn so với khi sử dụng các phương pháp khác như Binary Segmentation, dấu rằng tốc độ tính toán thực tế không nhanh hơn chúng;
- Giá trị phạt: sử dụng các hàm SIC (Schwarz information criterion; hoặc BIC - Bayesian information criterion), AIC (Akaike information criterion) và HQIC (Hannan - Quinn information criterion) là những hàm phạt đã từng chứng minh được ưu điểm trong công tác đánh giá điểm thay đổi. Tùy chọn số lượng điểm thay đổi tối đa tiềm năng luôn được đặt ở mức cao nhất mà phần mềm cho phép. Ngoài ra, tùy chọn "CROPS" cũng được khảo sát với giá trị phạt thay đổi trong khoảng  $1 \div 200$  nhằm mục đích liên tục và trực tiếp đánh giá sự biến đổi của số lượng điểm thay đổi theo giá trị phạt.

Hàm SIC:  $SIC = \ln(n)k - 2\ln(L_{max})$  (1)

Hàm AIC:  $AIC = (\frac{2n}{n-k-1})k - 2\ln(L_{max})$  (2)

Hàm HQIC:  $HQIC = 2\ln[\ln(n)]k - 2\ln(L_{max})$  (3)

trong đó:

$n$ : số lượng điểm quan sát hay kích thước mẫu;

$k$ : số thông số sẽ ước lượng;

$L_{max}$ : giá trị tối đa của log-likelihood cho mô hình ước lượng.

Cụ thể, tiêu chí để xác định giá trị phạt phù hợp sẽ là giá trị phạt nhỏ nhất mà khi tăng lên nữa thì số lượng điểm thay đổi sẽ giảm không đáng kể. Phương pháp xác định giá trị phạt phù hợp dùng trong nghiên cứu này là nguyên lý đồ thị khuỷu tay (elbow-plot) [19] áp dụng cho đồ thị thể hiện số lượng điểm thay đổi theo hàm penalty (kết quả từ tùy chọn "CROPS"). Trong đó, giá trị này chính là giá trị theo trục tung của giao điểm giữa 2 đường thẳng được kẻ chồng khít cho 2 nhánh cong trên đồ thị. Bên cạnh đó, việc so sánh (trực quan) tính logic của kết quả phân chia điểm thay đổi của từng hàm với đường cong thay đổi nhiệt độ theo thời gian trên đồ thị để số liệu không bị chia quá chi tiết (đến độ không cần thiết) hay ngược lại quá sơ sài mà bỏ sót điểm thay đổi cũng giúp kiểm tra tính phù hợp của số lượng điểm thay đổi tìm được đối với bộ số liệu đang xem xét.

### 3. KẾT QUẢ VÀ THẢO LUẬN

Kết quả

Công tác kiểm tra số liệu đầu vào đã được thực hiện nhằm loại bỏ những số liệu không hợp lý có thể gây ra bởi những nguyên nhân khác nhau (như vấn đề trục trặc thiết bị đo thực địa). Kết quả kiểm tra đã cho thấy bộ số liệu thô sử dụng không có những giá trị bất thường. Thống kê số liệu nhiệt độ cho từng năm trong thời đoạn nghiên cứu ở 2 trạm khí tượng được thể hiện ở bảng 1.

Bảng 1. Thống kê số liệu nhiệt độ không khí (°C) ngày và tuần tại các trạm khí tượng Nhà Bè và Cần Thơ trong giai đoạn 2013 ÷ 2017

Loại số liệu	Trạm khí tượng	Năm	Giá trị lớn nhất	Giá trị nhỏ nhất	Giá trị trung bình	Độ lệch chuẩn
Ngày	Nhà Bè	2013	31.20	23.98	27.65	1.38
		2014	30.88	22.53	27.56	1.62
		2015	31.45	23.70	27.92	1.59
		2016	31.78	23.80	28.01	1.35
		2017	30.63	22.53	27.69	1.21
	Cần Thơ	2013	30.80	22.80	27.48	1.39
		2014	31.00	21.90	27.47	1.52
		2015	30.60	23.60	27.83	1.56
		2016	31.15	23.48	27.75	1.30
		2017	30.25	22.23	27.52	1.24
Tuần	Nhà Bè	2013	30.26	24.68	27.62	1.21
		2014	30.34	23.51	27.57	1.50
		2015	30.53	24.46	27.94	1.46

		2016	31.39	26.15	28.01	1.19
		2017	29.70	24.89	27.70	0.97
	Cần Thơ	2013	29.79	24.16	27.45	1.19
		2014	30.13	23.37	27.47	1.37
		2015	30.06	24.63	27.86	1.41
		2016	30.53	25.69	27.75	1.07
		2017	29.54	24.03	27.54	0.97

Số liệu thô sau khi qua kiểm tra đã được đưa vào xử lý với gói phần mềm “changepoint”. Kết quả tính toán cho các điểm thay đổi của nhiệt độ không khí hàng năm giai đoạn 2013 ÷ 2017 cho 2 trạm Nhà Bè và Cần Thơ khi sử dụng phương pháp PELT với các hàm phạt khác nhau (SIC, AIC, Hannan-Quinn) được mô tả trong các bảng 2 và 3. Do trong nghiên cứu này các kết quả thu được từ phương pháp PELT và SegNeigh là giống nhau (khi các thông số khác như nhau) nên dưới đây chỉ trình bày kết quả sử dụng phương pháp PELT.

Bảng 2. Kết quả tính toán điểm thay đổi cho các năm sử dụng số liệu ngày

Năm	Trạm Nhà Bè			Trạm Cần Thơ		
	SIC	AIC	Hannan - Quinn	SIC	AIC	Hannan - Quinn
2013	33, 83, 135, 195, 351, 365	32, 54, 63, 72, 83, 112, 113, 131, 158, 162, 168, 195, 205, 221, 225, 247, 282, 320, 351, 365	33, 83, 131, 162, 195, 225, 247, 282, 320, 351, 365	32, 73, 162, 168, 183, 351, 365	4, 30, 52, 54, 80, 90, 96, 112, 113, 136, 158, 162, 168, 176, 183, 191, 196, 206, 212, 222, 225, 250, 263, 266, 334, 351, 365	30, 54, 80, 136, 162, 168, 183, 222, 225, 250, 263, 266, 351, 365
2014	13, 40, 68, 106, 161, 350, 365	13, 25, 40, 65, 72, 109, 118, 123, 151, 155, 161, 198, 212, 236, 260, 335, 363, 365	13, 25, 40, 68, 108, 151, 155, 161, 286, 335, 363, 365	14, 26, 39, 69, 107, 152, 237, 260, 341, 365	14, 26, 39, 68, 73, 107, 119, 123, 127, 152, 162, 172, 176, 202, 206, 237, 260, 270, 297, 341, 351, 357, 362, 365	14, 26, 39, 69, 107, 152, 162, 218, 237, 260, 270, 297, 341, 365
2015	14, 47, 81, 109, 163, 349, 365	14, 47, 57, 81, 109, 164, 177, 199, 200, 209, 213, 245, 250, 260, 275, 297, 349, 365	14, 47, 81, 109, 164, 177, 250, 260, 297, 349, 365	14, 48, 83, 164, 176, 250, 260, 296, 349, 365	3, 14, 48, 57, 83, 110, 139, 164, 177, 199, 202, 209, 223, 250, 257, 260, 297, 307, 312, 349, 356, 365	3, 14, 48, 82, 108, 139, 164, 176, 250, 260, 296, 349, 356, 365
2016	37, 40, 81, 103, 138, 166, 253, 312, 334, 366	37, 40, 81, 101, 123, 136, 144, 170, 181, 207, 219, 253, 291, 311, 334, 345, 349, 366	37, 40, 81, 103, 138, 170, 181, 207, 219, 253, 312, 334, 366	37, 41, 98, 136, 170, 181, 276, 312, 335, 366	24, 26, 37, 41, 79, 97, 104, 136, 142, 160, 166, 170, 182, 187, 228, 254, 257, 276, 312, 335, 345, 354, 362, 366	37, 41, 79, 101, 136, 170, 182, 187, 228, 254, 257, 276, 312, 335, 366
2017	74, 97, 125, 172, 244, 256, 351, 360, 365	14, 51, 61, 72, 97, 125, 172, 196, 217, 245, 256, 274, 276, 286, 314, 322, 323, 342, 351, 355, 358, 360, 365	14, 72, 97, 125, 196, 217, 245, 256, 351, 360, 365	72, 97, 124, 196, 199, 240, 261, 351, 355	25, 51, 72, 91, 96, 124, 147, 162, 171, 196, 199, 209, 227, 240, 263, 268, 342, 351, 356, 358, 360, 365	72, 97, 124, 161, 171, 196, 199, 240, 261, 342, 351, 356, 358, 360, 365

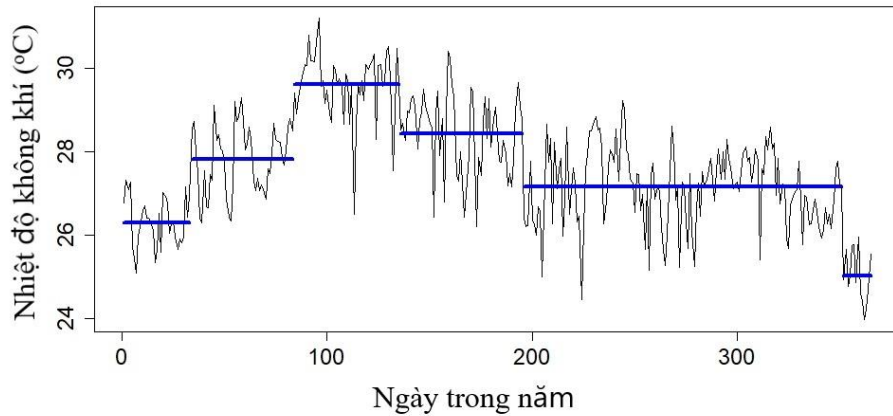
Bảng 3. Kết quả tính toán điểm thay đổi cho các năm sử dụng số liệu tuần

Năm	Trạm Nhà Bè			Trạm Cần Thơ		
	SIC	AIC	Hannan - Quinn	SIC	AIC	Hannan - Quinn
2013	11, 23, 50, 53	6, 12, 23, 50, 53	11, 23, 50, 53	8, 23, 50, 53	8, 23, 50, 53	8, 23, 50, 53
2014	9, 22, 52	5, 9, 22, 49, 52	5, 9, 22, 52	9, 22, 52	4, 9, 21, 49, 52	9, 22, 52
2015	6, 11, 23, 52	6, 11, 23, 52	6, 11, 23, 52	6, 11, 23, 52	6, 11, 23, 52	6, 11, 23, 52
2016	14, 20, 52	14, 20, 35, 52	14, 20, 35, 52	13, 19, 52	13, 19, 48, 52	13, 19, 52
2017	10, 24, 51	10, 24, 50, 51	10, 24, 50, 51	10, 49, 51	10, 18, 50, 51	10, 18, 50, 51

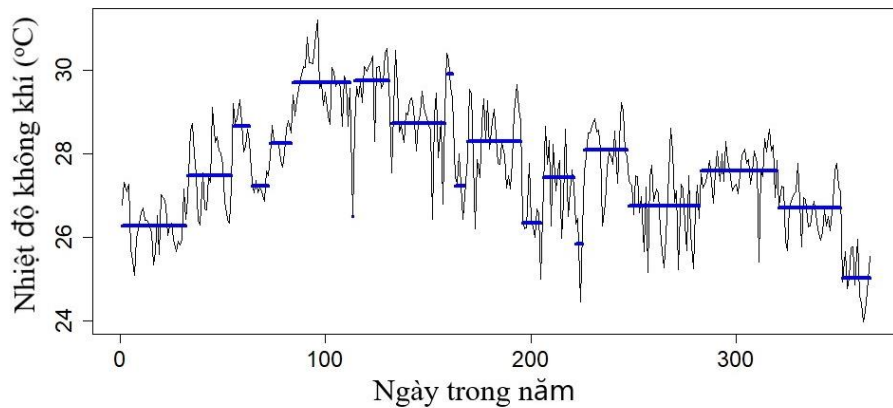
Nhìn chung, số lượng điểm thay đổi thu được gần như luôn khác nhau cho cùng bộ số liệu nhiệt độ (cùng năm, ở cùng trạm khí tượng, với cùng phương pháp tính PELT) khi sử dụng các hàm phạt khác nhau. Đối với bộ số liệu ngày, phụ thuộc vào hàm phạt sử dụng, số lượng điểm thay đổi tìm được dao động mỗi năm

trong khoảng từ 5 đến 26; với bộ số liệu tuần, số lượng điểm thay đổi mỗi năm tìm được dao động trong khoảng từ 2 đến 4.

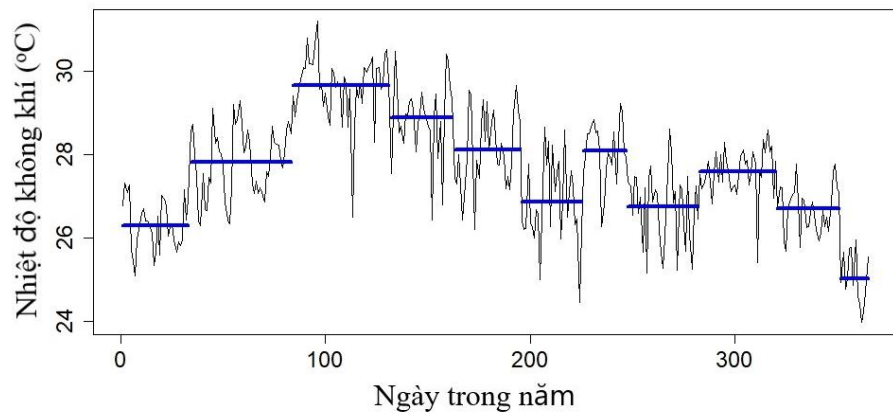
Dẫu sử dụng hàm phạt khác nhau nhưng vì khai thác chung bộ số liệu, số lượng điểm thay đổi hợp lý sẽ phải là một giá trị xác định cho mỗi năm ở mỗi trạm. Kết quả (nêu làm đại diện) cho sự phân bố điểm thay đổi của nhiệt độ không khí năm 2013 ở trạm Nhà Bè được thể hiện trong hình 2 (số liệu ngày) và hình 3 (số liệu tuần). Nhận xét qua quan sát sơ bộ dựa trên hình dạng đồ thị biến đổi nhiệt độ cho ta thấy hàm SIC là phù hợp hơn cả cho công tác xác định điểm thay đổi ở trạm khí tượng xem xét, tức là hàm này cho số lượng điểm thay đổi hợp lý nhất.



2.a)

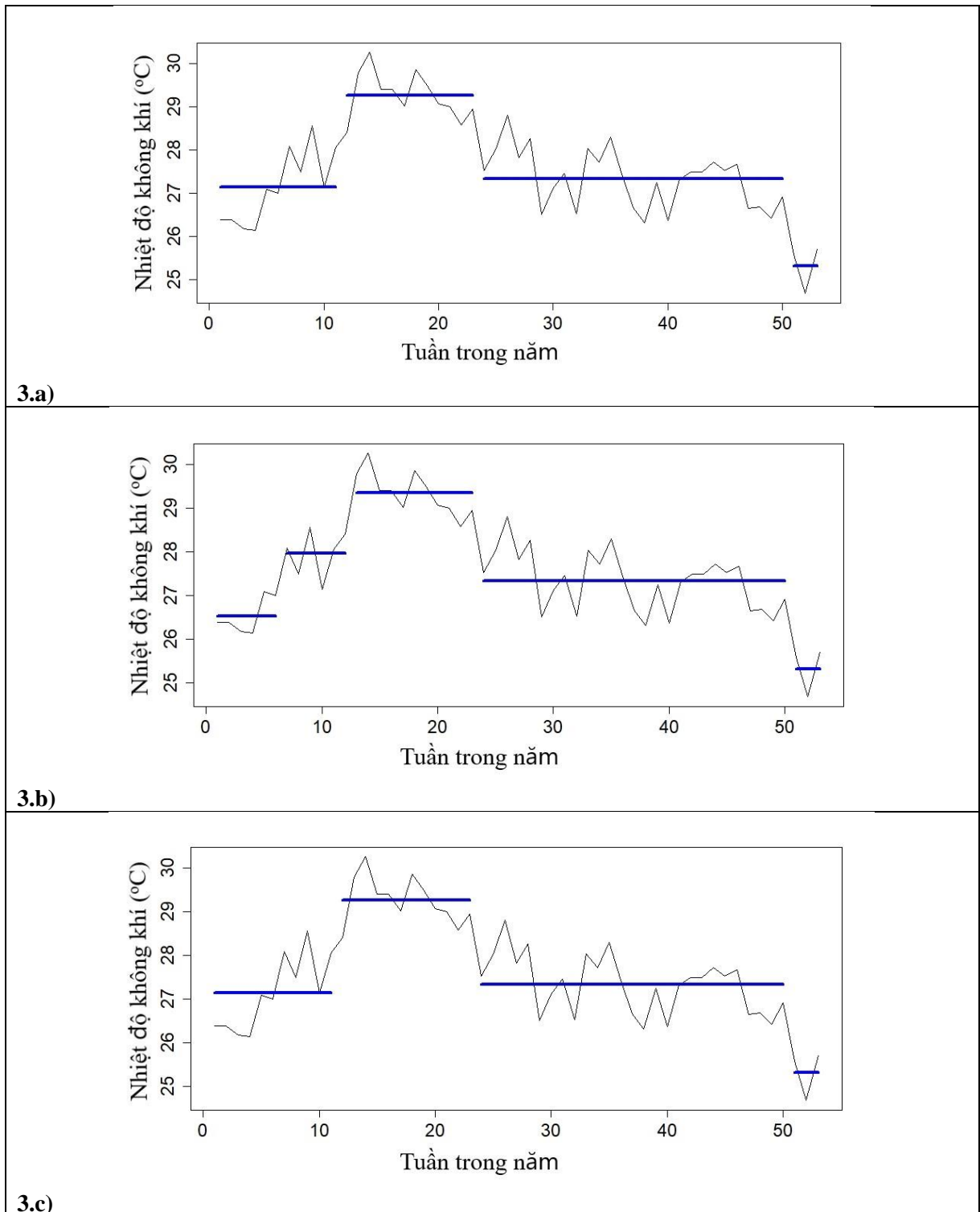


2.b)



2.c)

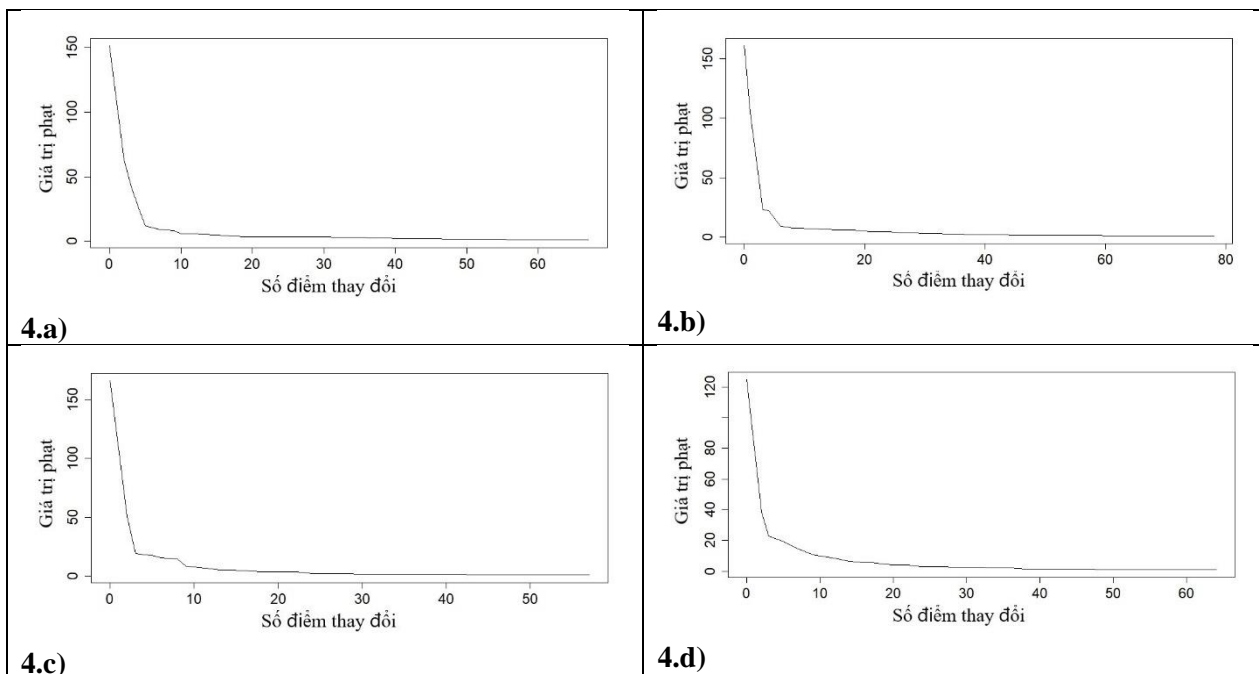
Hình 2. Kết quả các điểm thay đổi tìm được khi dùng phương pháp PELT dùng số liệu ngày năm 2013 cho trạm Nhà Bè với các hàm phạt khác nhau: 2.a) Hàm SIC; 2.b) Hàm AIC; 2.c) Hàm Hannan-Quinn.



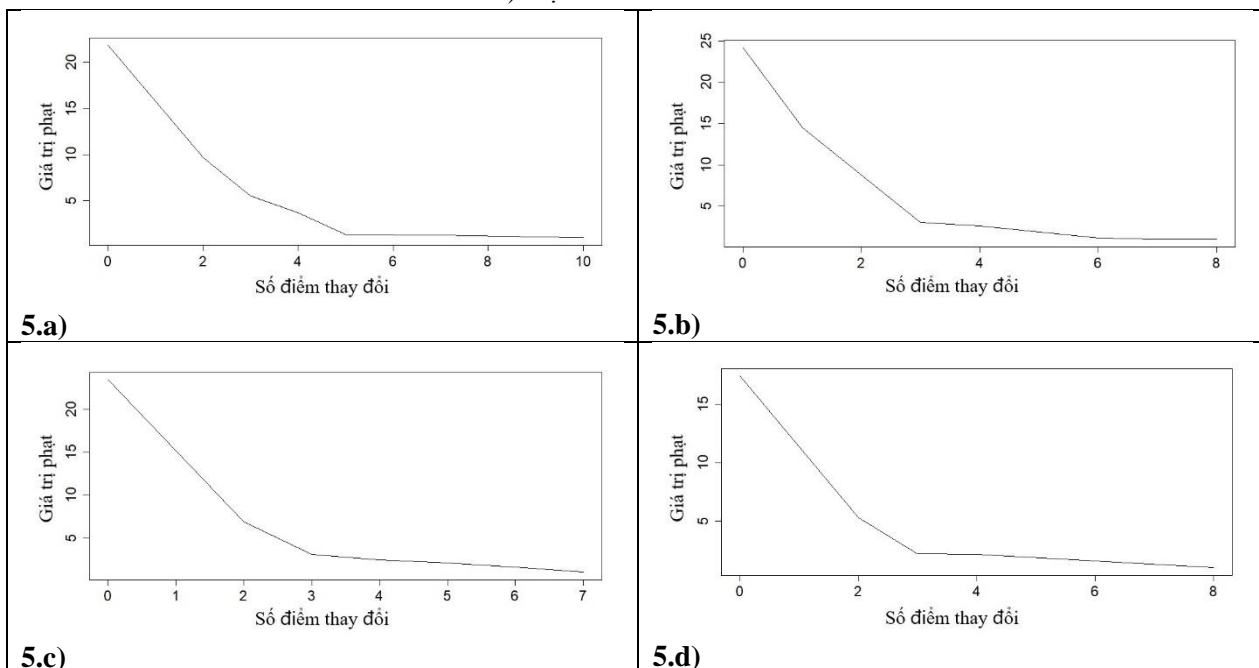
Hình 3. Kết quả các điểm thay đổi tìm được khi dùng phương pháp PELT dùng số liệu tuần năm 2013 cho trạm Nhà Bè với các hàm phạt khác nhau: 3.a) Hàm SIC; 3.b) Hàm AIC; 3.c) Hàm Hannan-Quinn.

Để kiểm tra nhận xét trên, kết quả số lượng điểm thay đổi tìm được khi dùng phương pháp PELT với tùy chọn “CROPS” cho giá trị phạt dùng số liệu ngày và số liệu tuần (ở cả 2 trạm, cho các năm 2013 và 2016) được thể hiện trong hình 4 và 5, tương ứng. Giá trị phạt trong tùy chọn “CROPS” được thay đổi trong khoảng  $1 \div 200$  để tính toán. Đồ thị ở các hình trên sau khi được xử lý bằng nguyên lý khuỷu tay cho thấy

sự trùng khớp giữa số lượng điểm thay đổi hợp lý là gần giống như kết quả số điểm thay đổi thu được từ hàm SIC. Điều này lần nữa xác nhận việc sử dụng hàm SIC trong đánh giá điểm thay đổi là phù hợp nhất trong số 3 hàm phạt đưa vào nghiên cứu. Bên cạnh đó, giá trị hàm phạt ở giao điểm của 2 đường thẳng theo nguyên lý khuỷu tay thường nằm trong khoảng  $8 \div 15$  và  $3 \div 8$  tương ứng đối với số liệu ngày và số liệu tuần.



Hình 4. Kết quả số lượng điểm thay đổi tìm được khi dùng phương pháp PELT và tùy chọn “CROPS” cho giá trị phạt dùng số liệu ngày: 4.a) Trạm Nhà Bè năm 2013; 4.b) Trạm Cần Thơ năm 2013; 4.c) Trạm Nhà Bè năm 2016; 4.d) Trạm Cần Thơ năm 2016.



Hình 5. Kết quả số lượng điểm thay đổi tìm được khi dùng phương pháp PELT và tùy chọn “CROPS” cho giá trị phạt dùng số liệu tuần: 5.a) Trạm Nhà Bè năm 2013; 5.b) Trạm Cần Thơ năm 2013; 5.c) Trạm Nhà Bè năm 2016; 5.d) Trạm Cần Thơ năm 2016.



### **Thảo luận**

Kết quả thu được cho ta thấy một số điểm sau:

- Do hàm SIC (hay BIC) cho giá trị phạt lớn hơn AIC nên số điểm thay đổi thu được nhỏ hơn khi sử dụng hàm phạt SIC cho cả 2 trạm nghiên cứu Nhà Bè và Cần Thơ. Bên cạnh đó, hàm HQIC cũng cho giá trị phạt lớn hơn AIC tuy nhiên lại nhỏ hơn giá trị hàm SIC nên số điểm thay đổi lớn hơn kết quả từ hàm SIC và nhỏ hơn hàm AIC. Những kết quả này tương thích với đặc tính các hàm SIC và HQIC là áp dụng mức phạt mạnh mẽ hơn cho việc mất số hạng tự do so với hàm AIC. Tuy nhiên, nếu so sánh hoạt động của các hàm AIC và SIC như Gayawan và Ipinyomi [22] mô tả trong việc lựa chọn mô hình phục vụ đánh giá mẫu hình mức sinh sản với kết quả thu được từ 2 hàm chỉ khác biệt khoảng 5 % thì sự khác biệt của chúng trong lĩnh vực đánh giá điểm thay đổi nhiệt độ lớn hơn tương đối đáng kể.

- Hàm phạt hợp lý sẽ cho ra số lượng điểm thay đổi gần nhất với kết quả khảo sát theo nguyên lý đồ thị khuỷu tay, tức khi tăng giá trị phạt lên cao hơn nữa thì số lượng điểm thay đổi tìm được sẽ giảm không đáng kể. Kết quả đánh giá sự phụ thuộc của số lượng điểm thay đổi theo giá trị hàm phạt (điền hình như cho các năm 2013, 2016 trong hình 4 và 5) sau khi áp dụng nguyên lý khuỷu tay và đem so sánh với số điểm thay đổi thu được từ mỗi hàm phạt đã chứng minh hàm SIC cho ra kết quả số lượng điểm thay đổi hợp lý nhất đối với cả 2 bộ số liệu khảo sát. Việc sử dụng hàm HQIC thường dẫn đến số lượng điểm thay đổi nhiều hơn mức hợp lý và hàm AIC cho kết quả quá nhiều điểm thay đổi, tức đã chia quá nhỏ các đoạn khác biệt cho cả 2 trạm nghiên cứu Nhà Bè và Cần Thơ. Thêm vào đó, việc so sánh (trực quan) tính logic của kết quả phân chia điểm thay đổi của từng hàm với đường cong thay đổi nhiệt độ theo thời gian trên đồ thị cũng cho thấy hàm SIC phù hợp với phần lớn trường hợp đưa vào tính toán.

- Sử dụng kết quả phân tích điểm thay đổi theo hàm SIC, ta thấy cả số lượng cũng như thời điểm xảy ra điểm thay đổi trong chuỗi số liệu nhiệt độ không khí đối với số liệu tuần tại 2 trạm Nhà Bè và Cần Thơ có độ trùng khớp cao (với sai lệch cho phép 1 tuần thì tỷ lệ số lượng điểm trùng khớp là 83.3 %); tuy nhiên, đối với số liệu ngày thì thời điểm thay đổi giữa 2 trạm có tỷ lệ độ trùng lặp lại ít hơn rất nhiều. Sự khác biệt này có thể do giá trị trung bình tuần được tính bởi trung bình cộng các ngày trong tuần đã làm dịu đi ảnh hưởng tương phản khác biệt ngày trong chuỗi số liệu nhiệt độ không khí ở các trạm. Điều này cho thấy việc chọn cấp thời gian (ngày hay tuần...) để sử dụng cho số liệu chuỗi đưa vào phân tích sẽ có ảnh hưởng đến kết quả công tác tìm điểm thay đổi. Đây là điểm quan trọng nên được tìm hiểu kỹ hơn trong tương lai cho các bộ số liệu và thông số khí tượng khác nhau.

### **4. KẾT LUẬN VÀ KIẾN NGHỊ**

Kết quả thu được từ phương pháp PELT và SegNeigh (để kiểm tra) là giống nhau trong đánh giá các điểm thay đổi theo tiêu chuẩn giá trị trung bình của nhiệt độ không khí, nhưng ảnh hưởng của các hàm phạt đến kết quả lại khác nhau đáng kể. Số lượng điểm thay đổi thu được khá khác nhau cho cùng bộ số liệu nhiệt độ (cùng năm, ở cùng trạm khí tượng) khi sử dụng các hàm phạt khác nhau. Phụ thuộc vào hàm phạt sử dụng, đối với bộ số liệu ngày và bộ số liệu tuần, số lượng điểm thay đổi mỗi năm tìm được dao động trong khoảng  $5 \div 26$  và  $2 \div 4$ , tương ứng. Trong đó, số điểm thay đổi tìm được tăng dần khi sử dụng lần lượt mỗi hàm SIC, HQIC và AIC cho nhiệt độ không khí tại các trạm khí tượng Nhà Bè và Cần Thơ.

Việc so sánh kết quả phân chia điểm thay đổi của từng hàm với đường cong thay đổi nhiệt độ theo thời gian trên đồ thị cùng kết quả áp dụng nguyên lý khuỷu tay trên đồ thị đều chứng minh hàm SIC cho ra kết quả số lượng điểm thay đổi hợp lý nhất cho cả 2 bộ số liệu nhiệt độ không khí ngày và tuần. Nghiên cứu cũng cho thấy việc chọn cấp thời gian (ngày hay tuần...) đối với số liệu đầu vào vào sẽ có ảnh hưởng đến kết quả công tác tìm điểm thay đổi. Khi sử dụng số liệu tuần, các điểm thay đổi trong năm tìm ra được có sự trùng khớp rất cao (trên 80%, với độ lệch cho phép 1 tuần) lúc so sánh hai trạm Nhà Bè và Cần Thơ dù khoảng cách đường thẳng giữa chúng là hơn 100 km. Tuy nhiên, kết quả những điểm thay đổi thu được lại khác biệt nhiều hơn đối với số liệu ngày cho 2 trạm trên.

Nghiên cứu này còn có điểm hạn chế khi số liệu khí tượng sử dụng chưa dài (5 năm), cũng như mới chỉ giới hạn cho 2 trạm khí tượng (Nhà Bè và Cần Thơ) tức là chịu ảnh hưởng đặc thù của khí hậu khu vực Nam Bộ. Bên cạnh đó, nghiên cứu chưa xét đến ảnh hưởng của một số yếu tố như đặc điểm phân bố số liệu và số lượng trạm quan trắc đưa vào khảo sát còn chưa nhiều. Tuy nhiên, kết quả thu được đã chứng minh tiềm năng phát hiện điểm thay đổi đối với yếu tố khí tượng điển hình ở đồng bằng Nam Bộ. Đối với nước ta là nước có nền nông nghiệp phát triển và sản lượng cây trồng phụ thuộc nhiều vào hiểu biết về sự biến đổi khí hậu của con người, điều này có ý nghĩa rất quan trọng. Để kết quả đánh giá mang tính tổng quát cao

hơn, các nghiên cứu sau có thể mở rộng thêm đến những vị trí nằm cách xa nhau trên lãnh thổ Việt Nam và xem xét những yếu tố khí tượng khác trong vai trò mục tiêu mô phỏng.

## LỜI CẢM ƠN

Tác giả xin cảm ơn Trường Đại học Công nghiệp thành phố Hồ Chí Minh, các đồng nghiệp ở Viện Khoa học Công nghệ và Quản lý Môi trường, đã tạo mọi điều kiện thuận lợi để cho nghiên cứu này có thể hoàn thành. Cảm ơn Đài khí tượng Thủy văn khu vực Nam Bộ đã cung cấp số liệu hiện trường cho nghiên cứu.

## TÀI LIỆU THAM KHẢO

- [1] Nguyễn Quang Hoan, Phạm Thị Trang, Hoàng Hồng Công, Nguyễn Thị Huyền, Dự báo thời tiết ứng dụng mạng nơron nhân tạo và thuật toán Bayes, *Tap chí Khoa học & Công nghệ*, Số 13, trang 39 - 43, 2017.
- [2] Trần Thị Vân, Hà Dương Xuân Bảo, Đinh Thị Kim Phượng, Nguyễn Thị Tuyết Mai và Đặng Thị Mai Nhung, Đặc điểm môi trường nhiệt và diễn biến đảo nhiệt đô thị bề mặt khu vực bắc thành phố Hồ Chí Minh, *Tap chí Khoa học Trường Đại học Cần Thơ*, 49(A), trang 11 - 20, 2017.
- [3] S. Aminikhanghahi, D. J. Cook, A Survey of Methods for Time Series Change Point Detection, *Knowledge and Information Systems*, 51(2), pp. 339 - 367, 2017.
- [4] S. Sharma, D. A. Swayne, C. Obimbo, Trend analysis and change point techniques: a survey, *Energy, Ecology and Environment*, 1(3), pp. 123 - 130, 2016.
- [5] F. Camci, Change point detection in time series data using support vectors, *International Journal of Pattern Recognition and Artificial Intelligence*, 24(1), pp. 73 - 95, 2010.
- [6] E. A. Khapalova, V. K. Jandhyala, S. B. Fotopoulos, Change-Point Analysis of Annual Mean Precipitation for Northern, Tropical and Southern Latitudes of the Globe in the Past Century, *Journal of Environmental Statistics*, 4(3), 2013.
- [7] S. Palaniswami, K. Muthiah, Change Point Detection and Trend Analysis of Rainfall and Temperature Series over the Vellar River Basin, *Polish Journal of Environmental Studies*, 27(4), pp. 1673 - 1681, 2018.
- [8] G. D. Wambui, G. A. Waititu, A. Wanjoya, The Power of the Pruned Exact Linear Time (PELT) Test in Multiple Changepoint Detection, *American Journal of Theoretical and Applied Statistics*, 4(6), pp. 581 - 586, 2015.
- [9] L. Xiong, S. Guo, Trend test and change-point detection for the annual discharge series of the Yangtze River at the Yichang hydrological station, *Hydrological Sciences*, 49(1), pp. 99 - 112, 2004.
- [10] A. N. Pettitt, A Non-Parametric Approach to the Change-Point Problem, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28 (2), pp. 126 - 135, 1979.
- [11] A. F. Militino, M. Moradi, M. D. Ugarte, On the Performances of Trend and Change-Point Detection Methods for Remote Sensing Data, *Remote Sensing*, 12, 2020.
- [12] C. Zhou, R. V. Nooijen, A. Kolechkina, M. Hrachowitz, Comparative analysis of nonparametric changepoint detectors commonly used in hydrology, *Hydrological sciences journal*, 64 (14), pp. 1690 - 1710, 2019.
- [13] R. K. Jaiswal, A. K. Lohani, H. L. Tiwari, Statistical Analysis for Change Detection and Trend Assessment in Climatological Parameters, *Environmental Processes*, 2, pp. 729 - 749, 2015.
- [14] I. E. Auger, C. E. Lawrence, Algorithms for the optimal identification of segment neighborhoods, *Bulletin of Mathematical Biology*, 51(1), pp. 39 - 54, 1989.
- [15] R. Killick, P. Fearnhead, I. A. Eckley, Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107 (500), pp. 1590 - 1598, 2012.

- [16] I. Akaike, A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, 19(6), pp. 716 - 723, 1974.
- [17] M. Costa, A. M. Goncalves, L. Teixeira, Change-point detection in environmental time series based on the informational approach, *Electronic Journal of Applied Statistical Analysis*, 9(2), pp. 267 - 296, 2016.
- [18] R. Killick, I. A. Eckley, Changepoint: An R Package for Changepoint Analysis, *Journal of Statistical Software*, 58(3), 2014.
- [19] M. Lavielle, Using penalized contrasts for the change-point problem, *Signal Processing*, 85(8), pp. 1501 - 1510, 2005.
- [20] R. Maidstone, T. Hocking, G. Rigai, P. Fearnhead, On optimal multiple changepoint algorithms for large data, *Statistics and Computing*, 27, pp. 519 - 533, 2017.
- [21] C. Truong, L. Oudre, N. Vayatis, Penalty Learning for Changepoint Detection, *25th European Signal Processing Conference (EUSIPCO)*, 2017.
- [22] E. Gayawan, R. A. Ipinyomi, A Comparison of Akaike, Schwarz and R Square Criteria for Model Selection Using Some Fertility Models, *Australian Journal of Basic and Applied Sciences*, 3(4), pp. 3524 - 3530, 2009
- [23] C. Beaulieu, J. Chen, J. L. Sarmiento, Change-point analysis as a tool to detect abrupt climate variations, *Philosophical Transactions of the Royal Society*, 370, pp. 1228–1249, 2012.

Ngày nhận bài: 07/05/2020

Ngày chấp nhận đăng: 04/09/2020