Tạp chí Khoa học Đại học Cần Thơ
Số chuyên đề: Công nghệ chiến lược - Động lực phát triển Đồng bằng sông Cửu Long (SDMD)
ISSN 1859-2333 | e-ISSN 2815-5599

DOI:10.22144/ctujos.2026.258

PHÂN TÍCH DỮ LIỆU TÔ-PÔ VÀ ĐIỀU KIỆN ÁP DỤNG: KHẢO SÁT TRÊN DỮ LIỆU KHÍ HẬU – SẢN XUẤT LÚA Ở ĐỒNG BẰNG SÔNG CỬU LONG GIAI ĐOẠN 1995 – 2024

Phạm Bích Như* và Dương Thị Bé Ba
Trường Khoa học Tự nhiên, Đại học Cần Thơ, Việt Nam
**Tác giả liên hệ (Corresponding author): pbnhu@ctu.edu.vn*

Thông tin chung (Article Information)

Nhận bài (Received): 16/08/2026
Sửa bài (Revised): 27/09/2026
Duyệt đăng (Accepted): 06/11/2026

Title: *Topological data analysis and its conditions of applicability: a case study on climate and rice production data of the Vietnamese Mekong Delta, 1995 – 2024*

Author(s): *Pham Bich Nhu* and Duong Thi Be Ba*

Affiliation(s): *College of Natural Sciences, Can Tho University, Viet Nam*

TÓM TẮT

Việc phân tích dữ liệu tô-pô (TDA) đo hình dạng dữ liệu được thực hiện bằng các bất biến ổn định với nhiễu. Nghiên cứu giới thiệu công cụ và xác định điều kiện áp dụng của nó. Mệnh đề 3 cùng hệ quả, là kết quả mới, chặn độ bền tô-pô bởi khoảng cách tới cấu trúc phụ thuộc tuyến tính hoàn hảo, cho tiêu chuẩn sàng lọc trước phân tích. Mô phỏng cho thấy chỉ báo tô-pô phân biệt được hai tổng thể trùng nhau đến mô-men bậc hai, điều mà mọi kiểm định dựa trên tương quan không thể làm, với công suất 100% từ cỡ mẫu 30 nhưng chỉ 20% tại cỡ mẫu 15. Kết quả khảo sát khí hậu và sản xuất lúa 13 tỉnh Đồng bằng sông Cửu Long 1995–2024 cho thấy mỗi đám mây 15 điểm, chỉ báo phát hiện thay đổi ở vụ hè thu–thu đông ($p \le 0,0086$) nhưng bỏ sót vụ đông xuân mà kiểm định quy ước phát hiện ở $p = 0,0002$. Trường hợp này là phản ví dụ cho nhận định TDA luôn nắm bắt nhiều hơn thống kê quy ước, đồng thời xác nhận ngưỡng cỡ mẫu trên.

Từ khóa: *Công suất kiểm định, điều kiện áp dụng, đồng điều bền vững, kiểm định hoán vị, phân tích dữ liệu tô-pô, sản xuất lúa*

ABSTRACT

Topological data analysis (TDA) measures the shape of data through invariants stable under noise. This study introduces the tool and establishes its conditions of applicability for Vietnamese Mekong Delta data. Three auxiliary propositions are proved, of which Proposition 3 and its corollary are new: topological persistence is bounded above by the distance to a perfectly linear dependence structure, yielding a screening criterion to be applied before analysis. A controlled simulation shows the topological indicator separates two populations identical up to second moments, a case in which every correlation-based test fails in principle, with 100% power from a sample size of 30 points but only 20% at size 15. Applied to climate and rice production data of 13 provinces of the Mekong Delta over 1995–2024, with 15 points per cloud, the indicator detects a structural change in the summer–autumn and autumn–winter seasons ($p \le 0,0086$) but misses the winter–spring season, which conventional tests detect at $p = 0,0002$. This case is both a counter-example to the claim that TDA always captures more than conventional statistics and a confirmation of the sample-size threshold above ($p \le 0.0086$).

Keywords: *Conditions of applicability, permutation test, persistent homology, rice production, statistical power, topological data analysis*

1. GIỚI THIỆU

Đồng bằng sông Cửu Long (ĐBSCL) là vựa lúa lớn nhất Việt Nam, đóng vai trò trọng yếu về an ninh lương thực và xuất khẩu nông sản, nhưng việc sản xuất lúa của vùng chịu tác động bất lợi ngày càng rõ của biến đổi khí hậu và xâm nhập mặn (Wassmann et al., 2004; Đan và ctv., 2015; Kontgis et al., 2019; Anh, 2021; Hương, 2026). Nhiều nghiên cứu đã tiếp cận vấn đề này từ các góc độ khác nhau: phân tích chuỗi giá trị và hệ thống canh tác lúa (Lộc & Son, 2011; Tuyến, 2013; Nhẫn và ctv., 2016), lựa chọn mô hình sản xuất của hộ trong điều kiện xâm nhập mặn (Khôi và ctv., 2022), đổi mới công nghệ ngành hàng lúa gạo (Anh, 2018) và giám sát mùa vụ, thực phủ bằng viễn thám (Sakamoto et al., 2006; Phụng & Nguyên, 2019; Diễm & Diễm, 2020), biến động mùa vụ theo vùng (Nguyên & Yến, 2021) và xu hướng hệ thống canh tác (Nhẫn và ctv., 2016; Nguyen et al., 2020). Tuy nhiên, phần lớn các nghiên cứu tập trung vào giá trị trung bình, xu thế hoặc phân bố không gian của từng yếu tố, trong khi tác động của khí hậu lên sản xuất lúa còn thể hiện qua sự biến đổi cấu trúc đồng biến thiên giữa nhiều yếu tố ít được khảo sát.

Các phương pháp thống kê thông dụng như so sánh trung bình hay hồi quy tập trung vào giá trị và xu thế nên không mô tả trực tiếp cấu trúc quan hệ giữa nhiều yếu tố. Việc phân tích dữ liệu tô-pô (Topological Data Analysis - TDA) được tiếp cận theo một hướng khác: thay vì đo giá trị, TDA đo hình dạng của tập dữ liệu thông qua các bất biến đại số, cụ thể là đồng điều bền vững (persistent homology) (Carlsson, 2009; Edelsbrunner & Harer, 2010; Chazal & Michel, 2021). Các đặc trưng thu được bất biến với phép biến dạng liên tục và ổn định với nhiễu theo nghĩa Lipschitz (Cohen-Steiner et al., 2007).

Trong lĩnh vực khí hậu và môi trường, đồng điều bền vững đã được vận dụng ở một số hướng. Muszynski et al. (2019) kết hợp đồng điều bền vững với học máy để nhận dạng các dòng khí quyển trong dữ liệu khí hậu toàn cầu, cho thấy bất biến tô-pô tách được những cấu trúc mà việc đặt ngưỡng trên giá trị đơn thuần không tách được; Tymochko et al. (2020) dùng giản đồ bền vững của ảnh mây hồng ngoại để định lượng chu kỳ ngày–đêm của bão nhiệt đới, Ver Hoef et al. (2023) tổng hợp thành một giới thiệu TDA cho khoa học môi trường và nhận định phần lớn ứng dụng hiện có dừng ở mức mô tả. Khác biệt then chốt về dữ liệu cần nêu ngay: cả ba làm việc với dữ liệu lưới hoặc ảnh, mỗi đám mây từ hàng nghìn điểm, trong khi thống kê cấp tỉnh theo năm chỉ cho 15 điểm – chênh ba bậc độ lớn. Đây chính là lý do phần mô phỏng xác định ngưỡng cỡ mẫu ở mục 2.6 là cần thiết.

Mục tiêu của nghiên cứu này là giới thiệu phân tích dữ liệu tô-pô như một công cụ phân tích và xác định điều kiện áp dụng của nó. Giá trị sử dụng của một công cụ như vậy không nằm ở chỗ nó tinh vi tới đâu, mà ở chỗ trả lời được câu hỏi: với loại dữ liệu mà người dùng thực có, khi nào công cụ cho kết quả tin cậy được và khi nào không? Nghiên cứu được thực hiện nhằm trả lời câu hỏi này bằng một con số cụ thể, theo hai nhánh bổ sung cho nhau: một mô phỏng có kiểm soát để xác định năng lực và ngưỡng cỡ mẫu của công cụ và một khảo sát trên dữ liệu thực nhằm kiểm chứng ngưỡng đó.

Bộ dữ liệu khảo sát được chọn là thống kê khí hậu và sản xuất lúa cấp tỉnh ĐBSCL-chính vì đây là nguồn dữ liệu mà các cơ quan quản lý và nghiên cứu trong vùng thực sự đang có và đang dùng: thống kê hành chính chính thức, cấp tỉnh, theo năm. Một kết luận về điều kiện áp dụng chỉ hữu ích nếu được rút ra từ chính loại dữ liệu đó. Kết quả cho thấy bộ dữ liệu này nằm ở phía bất lợi của điều kiện áp dụng: chỉ báo tô-pô bỏ sót một thay đổi mà kiểm định quy ước phát hiện rõ. Trường hợp đó mang hai ý nghĩa. Nó là phản ví dụ cho nhận định thường gặp trong tài liệu ứng dụng rằng TDA nắm bắt được cấu trúc mà thống kê quy ước không thấy, ở đây điều ngược lại xảy ra. Đồng thời, nó xác nhận đúng ngưỡng cỡ mẫu mà mô phỏng đã dự báo trước, nên không phải là một thất bại ngẫu nhiên mà là một quan sát có thể dùng làm quy tắc.

Hai rủi ro phương pháp luận sau đây được chính các tổng quan về TDA nêu ra và được nghiên cứu này kiểm soát từ bước thiết kế. Thứ nhất là rò rỉ nhãn: nếu biến dùng để phân nhóm cũng có mặt trong không gian đặc trưng thì khác biệt thu được mang tính định nghĩa chứ không phải phát hiện. Thứ hai là thiếu khung suy diễn chính quy: khoảng cách giữa hai giản đồ bền vững được so sánh trực tiếp mà không gắn với một kiểm định có mức ý nghĩa, dù công cụ suy diễn cho giản đồ bền vững đã được xây dựng (Fasy et al., 2014; Bubenik, 2015). Chazal and Michel (2021) xem việc thiếu nền tảng thống kê là một hạn chế chính của TDA ứng dụng và Ver Hoef et al. (2023) nhận định tương tự cho riêng khoa học môi trường.

Cơ sở toán học ở mục 2 - đồng điều bền vững, phép lọc Vietoris–Rips, khoảng cách bottleneck, định lý cấu trúc và định lý ổn định - là các kết quả đã công bố, được dẫn lại kèm nguồn; đây là cơ sở phương pháp, không phải đóng góp khoa học. Mệnh đề 1 và Mệnh đề 2 cũng là tính chất đã biết, được chứng minh lại vì dùng trực tiếp ở mục 3. Trên cơ sở đó, nghiên cứu có ba đóng góp. Thứ nhất, Mệnh đề 3 cùng Hệ quả 2 chặn độ bền của mọi đặc trưng bậc từ một trở lên bởi bốn lần khoảng cách Hausdorff tới đường thẳng gần nhất; theo hiểu biết của nhóm tác giả, đây là kết quả mới, nó được xem là một tiêu chuẩn sàng lọc trước phân tích. Thứ hai, mô phỏng có kiểm soát xác định ngưỡng cỡ mẫu tối thiểu để chỉ báo có công suất. Thứ ba, một thiết kế đối chứng song song trên dữ liệu thực, giữ nguyên cả kết quả bất lợi cho chỉ báo tô-pô, vì chính sự bất lợi đó mới định lượng được điều kiện áp dụng.

Phần còn lại của bài viết được tổ chức như sau. Mục 2 trình bày cơ sở toán học của phân tích dữ liệu tô-pô cùng ba mệnh đề bổ trợ, thiết kế thí nghiệm với cơ chế kiểm soát rò rỉ nhãn và kiểm định hoán vị, nghiên cứu mô phỏng xác định ngưỡng cỡ mẫu và mô tả dữ liệu. Mục 3 trình bày kết quả khảo sát, phản đối chứng với thống kê quy ước và những hạn chế. Mục 4 tóm tắt và nêu hướng tiếp theo.

2. CƠ SỞ TOÁN HỌC VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Phức đơn hình và đồng điều

Cho tập hữu hạn V. Một phức đơn hình K trên V là họ các tập con khác rỗng của V thỏa mãn điều kiện đóng theo mặt: nếu $\sigma \in K$ và $\emptyset \neq \tau \subseteq \sigma$ thì $\tau \in K$. Phần tử σ với $|\sigma| = k+1$ gọi là k-đơn hình.

Xét nhóm k-đơn hình

$$C_k = \{\sigma = [v_0, v_1, \cdots, v_k], v_i \in V\}.$$

Ta định nghĩa toán tử biên

$$\partial_k : C_k \to C_{k-1},$$

xác định trên đơn hình sinh bởi công thức

$$\partial_k[v_0, v_1, \cdots, v_k] = \sum_{i=1}^{k} (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, v_k].$$

Trong đó, $[v_0, \ldots, \hat{v}_i, \ldots, v_k] \in C_{k-1}(K)$ là $(k-1)$-đơn hình sinh ra từ k-đơn hình $[v_0, v_1, \cdots, v_k]$ bằng cách bỏ đi v_i.

Cố định trường hệ số $\mathbb{F}$. Ta định nghĩa tập hợp tất cả các tổ hợp tuyến tính của các k-đơn hình

$$C_k(K) = \{c = \sum_{i=1}^{n} \alpha_i \sigma_i, \alpha_i \in \mathbb{F}, \sigma_i \in C_k\},$$

ở đây, c sẽ được gọi là một k-dây chuyền.

Khi đó, $C_k(K)$ là một $\mathbb{F}$-không gian véctơ và mở rộng của toán tử biên ∂_k sẽ là một toán tử tuyến tính. Hơn nữa, ta có $\partial_{k-1} \circ \partial_k = 0$. Ý tưởng chính của chứng minh là mỗi cặp đỉnh bị bỏ trong k-đơn hình σ xuất hiện đúng hai lần với dấu ngược nhau nên tổng triệt tiêu (Edelsbrunner & Harer, 2010).

Như vậy, dãy các toán tử biên

$$\cdots \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \xrightarrow{\partial_{k-1}} C_{k-2} \to \cdots,$$

là một phức dây chuyền.

Từ đó, ta có

$$B_k = \mathrm{Im}\partial_{k+1} \subseteq Z_k = \mathrm{Ker}\ \partial_k,$$

và nhóm đồng điều thứ k là không gian thương

$$H_k(K, \mathbb{F}) = Z_k / B_k,$$

và số Betti

$$\beta_k = \dim_F H_k(K, \mathbb{F}).$$

Số Betti β_k có ý nghĩa hình học trực tiếp: β_0 đếm số thành phần liên thông, β_1 đếm số vòng độc lập, β_2 đếm số khoang rỗng.

Cho một tập dữ liệu X nằm trong không gian mêtric (M, d) và một tham số $\varepsilon \in R$. Tại mỗi ngưỡng ε_i, xây dựng một dãy phức đơn hình $\{K_{\varepsilon_i}\}$.

Giả sử có một dãy tham số tăng nghiêm ngặt

$$\varepsilon_1 < \varepsilon_2 < \cdots < \varepsilon_n,$$

ta định nghĩa dãy phức tăng theo quan hệ bao hàm tự nhiên

$$K_{\varepsilon_1} \subset K_{\varepsilon_2} \subset \cdots \subset K_{\varepsilon_n}.$$

Mỗi phức tại các tham số nhỏ sẽ là phức con của phức có tham số lớn hơn, tạo thành một chuỗi các phép nhúng $\iota_{i,i+1}$:

$$K_{\varepsilon_1} \overset{\iota_{1,2}}{\hookrightarrow} K_{\varepsilon_2} \overset{\iota_{2,3}}{\hookrightarrow} \ldots \overset{\iota_{n-1,n}}{\hookrightarrow} K_{\varepsilon_n}.$$

Chuỗi các phức đơn hình cùng với các phép nhúng ι này là một lọc của không gian, ký hiệu là F. Về mặt hình thức, lọc là một hàm tử thuộc phạm trù thứ tự toàn phần $(R, \leq)$ vào phạm trù các phức đơn hình.

Cố định một bậc $k \in N$, áp dụng hàm tử đồng điều $H_k(-;F)$ lên toàn bộ lọc F. Các phép nhúng $\iota_{i,j}: K_{\varepsilon_i} \hookrightarrow K_{\varepsilon_j}$ sẽ cảm sinh các ánh xạ tuyến tính tương ứng:

$$f_{i,j} = (H_k)_{i,j}: H_k(K_{\varepsilon_i}; \mathbb{F}) \to H_k\left(K_{\varepsilon_j}; \mathbb{F}\right).$$

Kết quả nhận được là một chuỗi các không gian véctơ được liên kết bởi các ánh xạ tuyến tính:

$$H_k(K_{\varepsilon_1}) \xrightarrow{f_{1,2}} H_k(K_{\varepsilon_2}) \xrightarrow{f_{2,3}} \dots \xrightarrow{f_{n-1,n}} H_k(K_{\varepsilon_n}).$$

Chuỗi này tạo thành một cấu trúc đại số được gọi là mô-đun bền vững (persistence module) M. Một mô-đun bền vững là một hàm tử từ tập sắp thứ tự $(R, \leq)$ vào phạm trù các $\mathbb{F}$-không gian véctơ. Cấu trúc M chứa toàn bộ thông tin về cách các lớp đồng điều lớn lên khi tham số ε thay đổi.

Số Betti tại một tham số cố định $\beta_k(K_{\varepsilon_i}) = \dim\left(H_k(K_{\varepsilon_i})\right)$ chỉ phản ánh cấu trúc tô-pô tức thời và rất nhạy cảm với nhiễu. Thay vào đó, cấu trúc hình học thực sự được định lượng thông qua nhóm đồng điều bền vững thứ k, định nghĩa bởi ảnh của các ánh xạ cảm sinh từ bước i đến bước j (với $i < j$):

$$H_k^{i,j} = \mathrm{Im}\left(f_{i,j}: H_k(K_{\varepsilon_i}) \to H_k\left(K_{\varepsilon_j}\right)\right).$$

Số chiều của không gian véctơ này là số Betti bền vững:

$$\beta_k^{i,j} = \dim\left(\mathrm{Im}(f_{i,j})\right).$$

Nếu một lớp đồng điều $[\alpha] \in H_k(K_{\varepsilon_i})$ thỏa mãn $f_{i,j}([\alpha]) \neq 0$ với $j - i$ đủ lớn thì nó đại diện cho một đặc trưng tô-pô mang tín hiệu. Ngược lại, nếu $f_{i,j}([\alpha]) = 0$ khi $j - i$ rất nhỏ, tức là $[\alpha] \in \mathrm{Ker}(f_{i,j})$ thì nó được gọi là nhiễu đồng điều. Như vậy, cấu xạ $f_{i,j}$ là công cụ phân loại tín hiệu và nhiễu.

Cho một tập điểm hữu hạn $X \subset R^d$ được trang bị một mêtric Euclid chuẩn $d(x,y) = \|x - y\|_2$. Đường kính của một tập con $\sigma \subseteq X$ được định nghĩa bởi

$$\mathrm{diam}(\sigma) = \max_{x,y \in \sigma} d(x,y).$$

2.2. Phép lọc Vietoris–Rips và định lý cấu trúc

Cho tập dữ liệu hữu hạn $X \subset R^d$ và tham số $\varepsilon \in R$. Ta định nghĩa phức Vietoris–Rips gồm các tập con có đường kính không vượt quá ε:

$$VR_\varepsilon(X) = \{\sigma \subseteq X: \mathrm{diam}(\sigma) \leq \varepsilon\}.$$

Từ đó, suy ra $\sigma = [v_0, v_1, \dots, v_k] \subset X$ là một k- đơn hình trong $VR_\varepsilon(X)$ khi và chỉ khi $d(v_i, v_j) \leq \varepsilon, \forall v_i, v_j \in \sigma$. Điều này có nghĩa là khoảng cách giữa mọi cặp đỉnh trong σ đều không vượt quá ε.

Rõ ràng, nếu $\varepsilon_1 \leq \varepsilon_2$ thì $VR_{\varepsilon_1}(X) \subset VR_{\varepsilon_2}(X)$. Do đó, ta thu được phép nhúng tự nhiên:

$$VR_{\varepsilon_1}(X) \hookrightarrow VR_{\varepsilon_2}(X).$$

Cho tham số ε chạy liên tục từ 0 đến $+\infty$, ta nhận được một lọc

$$F: \quad \emptyset = VR_0(X) \subseteq VR_{\varepsilon_1}(X) \subseteq VR_{\varepsilon_2}(X) \subseteq \dots \subseteq VR_\infty(X) = \Delta^{|X|-1}.$$

Trong đó, $\Delta^{|X|-1}$ là $\Delta^{|X|-1}$-đơn hình chứa mọi điểm trong X.

Việc áp dụng hàm tử đối đồng điều $H_k(-;F)$ lên lọc F này ta thu được một mô-đun bền vững.

Định lý 1 (*Định lý cấu trúc, Zomorodian & Carlsson, 2005*). Cho $\mathbb{F}$ là một trường và M là một mô-đun bền vững hữu hạn sinh trên $\mathbb{F}$. Khi đó, M phân tích được một cách duy nhất, sai khác bởi một đẳng cấu và một phép hoán vị các hạng tử, thành tổng trực tiếp hữu hạn các mô-đun khoảng:

$M \cong \bigoplus_{i=1}^{n} I[b_i, d_i)$.Trong đó, b_i, d_i là các chỉ số thỏa mãn $b_i < d_i \leq +\infty$.

Chứng minh. Ý tưởng là đồng nhất mô-đun bền vững với một môđun phân bậc hữu hạn sinh trên vành chính $\mathbb{F}[t]$, rồi áp dụng định lý cấu trúc cho môđun trên vành chính; tính duy nhất của các khoảng suy từ tính duy nhất của các nhân tử bất biến. Kết quả chứng minh được thể hiện đầy đủ trong nghiên cứu của Zomorodian and Carlsson (2005).

Đại lượng $d_i - b_i$ được gọi là độ bền, điểm càng xa đường chéo càng biểu thị đặc trưng thực, điểm sát đường chéo được xem là nhiễu. Giản đồ bền vững có thể được tóm tắt thành các đặc trưng số ổn định phục vụ phân tích thống kê (Bubenik, 2015); trong nghiên cứu này, độ phức tạp tô-pô được đo bằng tổng độ bền của các khoảng.

Trong thực hành, phép lọc Vietoris–Rips không chạy tới vô hạn mà được cắt tại một ngưỡng lọc (filtration threshold) - giá trị lớn nhất của tham số ε được xét khi xây dựng phép lọc. Để so sánh công bằng giữa các đám mây có thang khoảng cách khác nhau, ngưỡng lọc được đặt theo phân vị

của phân bố khoảng cách từng cặp điểm trong mỗi đám mây (nghiên cứu này dùng các phân vị 40%, 45% và 50%). Độ phức tạp tô-pô của một đám mây được đo bằng tổng độ bền của các khoảng sinh-diệt tính đến ngưỡng lọc; việc lặp lại phân tích ở ba ngưỡng nhằm kiểm tra tính bền vững của kết luận trước lựa chọn ngưỡng.

2.3. Khoảng cách bottleneck và định lý ổn định

Định nghĩa 1. Trong không gian các giản đồ bền vững, khoảng cách bottleneck giữa hai giản đồ D_1 và D_2 được xác định bởi:

$$d_B(D_1, D_2) = \inf_{\gamma} \sup_{x \in D_1 \cup \Delta} \|x - \gamma(x)\|_\infty.$$

Trong đó, inf được lấy trên mọi song ánh $\gamma: D_1 \cup \Delta \to D_2 \cup \Delta$, với $\Delta = \{(x, x) \mid x \in R\}$ là đường chéo chính và $\|\cdot\|_\infty$ là chuẩn cực đại trên không gian R^2.

Việc mở rộng miền xác định của song ánh γ bằng cách hợp với đường chéo Δ là một cơ chế hình thức nhằm kiểm soát nhiễu. Cụ thể, cơ chế này cho phép các điểm có độ bền nhỏ (nằm gần đường chéo) được ghép cặp trực tiếp với Δ. Khi đó, chi phí của phép ghép cặp tỷ lệ thuận với khoảng cách từ điểm đó đến đường chéo, giúp khoảng cách bottleneck tự động triệt tiêu ảnh hưởng của các đặc trưng tô-pô mang tính nhiễu cục bộ.

Cụ thể, chi phí ghép cặp (matching cost) của một song ánh là khoảng cách lớn nhất mà một điểm phải di chuyển khi ta ghép mỗi đặc trưng (mỗi điểm sinh – diệt) của giản đồ thứ nhất với một đặc trưng của giản đồ thứ hai, hoặc ghép với đường chéo; khoảng cách bottleneck là chi phí nhỏ nhất trên mọi cách ghép. Nói cách khác, hai giản đồ được xem là gần nhau nếu tồn tại một cách ghép các đặc trưng bền vững của chúng sao cho không đặc trưng nào phải dịch chuyển quá xa.

Định lý 2 (*Định lý ổn định, Cohen-Steiner et al., 2007*). Cho X là một không gian tô-pô tam giác hóa được và $f, g: X \to R$ là hai hàm thuần liên tục. Khi đó, khoảng cách bottleneck giữa hai giản đồ bền vững sinh bởi f và g bị chặn trên bởi khoảng cách chuẩn vô cùng giữa hai hàm:

$$d_B\big(\mathrm{dgm}(f), \mathrm{dgm}(g)\big) \le \|f - g\|_\infty.$$

Định lý này khẳng định ánh xạ chuyển từ không gian các hàm (trang bị chuẩn $\|\cdot\|_\infty$) sang không gian các giản đồ bền vững (trang bị khoảng cách bottleneck) là một ánh xạ liên tục Lipschitz với hằng số Lipschitz bằng 1.

Hệ quả 1. Xét hai tập đám mây điểm hữu hạn X, Y nhúng trong cùng một không gian mêtric (M, d). Gọi $\mathrm{VR}(X)$ và $\mathrm{VR}(Y)$ lần lượt là phép lọc Vietoris–Rips xây dựng trên X và Y. Khi đó, với mọi bậc đồng điều $k \ge 0$, sự phụ thuộc của giản đồ bền vững vào dữ liệu đầu vào thỏa mãn bất đẳng thức:

$$d_B\Big(\mathrm{dgm}_k\big(\mathrm{VR}(X)\big), \mathrm{dgm}_k\big(\mathrm{VR}(Y)\big)\Big) \le d_H(X, Y),$$

trong đó $d_H(X, Y)$ là khoảng cách Hausdorff giữa hai tập hợp X và Y.

Bất đẳng thức này cung cấp nền tảng độ đo vững cho các ứng dụng tính toán thực nghiệm. Nó đảm bảo rằng các thao tác như lấy mẫu con trên tập dữ liệu lớn, hoặc sự hiện diện của sai số quan trắc có biên độ nhỏ, sẽ được ràng buộc chặt chẽ và không làm thay đổi cục diện cấu trúc tô-pô được trích xuất.

2.4. Các kết quả bổ trợ dùng trong phân tích

Ba mệnh đề sau đây đóng vai trò là căn cứ toán học trực tiếp cho các thiết kế phân tích ở Mục 3.

Mệnh đề 1. Cho không gian mêtric hữu hạn (X, d) và hằng số $c > 0$. Ký hiệu $(X, c \cdot d)$ là không gian X được trang bị mêtric tỉ lệ $c \cdot d$. Khi đó, với mọi $\varepsilon \ge 0$, ta có đẳng thức về phức Vietoris–Rips:

$$\mathrm{VR}_{c\varepsilon}(X, c \cdot d) = \mathrm{VR}_{\varepsilon}(X, d).$$

Hơn nữa, giản đồ bền vững bậc k thỏa mãn $\mathrm{dgm}_k(X, c \cdot d) = c \cdot \mathrm{dgm}_k(X, d)$ và khoảng cách bottleneck giữa hai giản đồ D_1, D_2 bất kỳ có tính thuần nhất bậc một:

$$d_B(cD_1, cD_2) = c \cdot d_B(D_1, D_2).$$

Chứng minh. Lấy một đơn hình tùy ý $\sigma \subseteq X$. Ta có

$$\mathrm{diam}_{c \cdot d}(\sigma) = c \cdot \mathrm{diam}_d(\sigma).$$

Do đó, $\mathrm{diam}_{c \cdot d}(\sigma) \le c\varepsilon$ khi và chỉ khi $\mathrm{diam}_d(\sigma) \le \varepsilon$. Điều này chứng tỏ cấu trúc của hai phức đơn hình tại các ngưỡng tương ứng là hoàn toàn trùng khớp. Khi đó, mô-đun bền vững của $(X, c \cdot d)$ chính là ảnh của mô-đun bền vững của (X, d) qua phép đổi biến tham số lọc $\varepsilon \mapsto c\varepsilon$. Theo tính duy nhất của Định lý phân tích cấu trúc, mỗi mô-đun khoảng $[b, d)$ được ánh xạ thành $[cb, cd)$.

Khẳng định về khoảng cách bottleneck được suy ra trực tiếp từ tính thuần nhất của chuẩn $\|\cdot\|_\infty$ trong định nghĩa $d_B(D_1, D_2)$. □

Mệnh đề 2. Cho đám mây điểm quan trắc $W = \{w_1, w_2, \dots, w_n\} \subset R^d$ có trung bình μ và độ lệch chuẩn $\sigma > 0$. Gọi W_z là tập dữ liệu sau khi thực hiện chuẩn hóa z-score, tức là

$$W_z = \{(w_i - \mu)/\sigma \mid w_i \in W\}.$$

Khi đó, phép biến đổi $W \mapsto W_z$ bảo toàn hoàn toàn cấu trúc hình học sai khác một hệ số tỉ lệ và ta có:

$$\mathrm{dgm}_k\big(\mathrm{VR}(W_z)\big) = \frac{1}{\sigma} \cdot \mathrm{dgm}_k\big(\mathrm{VR}(W)\big).$$

Chứng minh. Phép biến đổi z-score có thể được phân tích thành hợp thành của hai phép biến đổi trên R^d.

Thứ nhất là phép tịnh tiến $x \mapsto x - \mu$ là một đẳng cự, hoàn toàn bảo toàn khoảng cách mêtric.

Thứ hai là phép vị tự tâm O tỉ số $1/\sigma$ là một phép đồng dạng (similarity transformation), khiến khoảng cách giữa mọi cặp điểm bị nhân với thừa số $1/\sigma$.

Kết hợp lại, mêtric trên W_z chính là mêtric trên W được nhân với hằng số $c = 1/\sigma$. Áp dụng Mệnh đề 1 với $c = 1/\sigma$, ta suy ra điều phải chứng minh. □

Sau khi áp dụng chuẩn hóa z-score cho từng đối tượng quan sát, các thông tin tuyệt đối về mức nền (trung bình μ) và biên độ dao động toàn cục (độ lệch chuẩn σ) đã bị triệt tiêu hoàn toàn. Do đó, những sai khác tô-pô còn tồn tại giữa các giản đồ bền vững hoàn toàn độc lập với phép biến đổi affine và chỉ xuất phát từ cấu trúc hình dạng nội tại của dữ liệu.

Tính chất toán học này là cơ sở vững chắc để khẳng định rằng: Các kết quả phân biệt và phân cụm thu được ở mục 3 không phải là hệ quả sinh ra do sự chênh lệch về giá trị trung bình hay phương sai giữa các nhóm mà là phản ánh sự khác biệt thực sự về cấu trúc đa chiều của không gian pha.

Mệnh đề 3. Cho $X \subset R$ là một tập điểm hữu hạn. Khi đó, với mọi ngưỡng $\varepsilon > 0$, mỗi thành phần liên thông của phức Vietoris–Rips $\mathrm{VR}(X, \varepsilon)$ đều co rút được. Suy ra, nhóm đồng điều $H_k\big(\mathrm{VR}(X, \varepsilon)\big) = 0$ với mọi $k \geq 1$ và do đó giản đồ bền vững bậc $k \geq 1$ của X là tập rỗng.

Chứng minh. Chứng minh được thực hiện qua hai bước.

Bước 1: Đồ thị Vietoris–Rips của X là đồ thị chordal. Đồng nhất mỗi điểm với tọa độ thực của nó trên đường thẳng. Giả sử phản chứng rằng tồn tại một chu trình cảm sinh có độ dài ít nhất bốn. Gọi v là đỉnh có tọa độ nhỏ nhất trong chu trình đó và u, w là hai đỉnh kề v trong chu trình; vì chu trình có độ dài ít nhất bốn nên u và w không kề nhau trên chu trình. Do v nhỏ nhất, ta có tọa độ của u và w đều không nhỏ hơn tọa độ của v, đồng thời khoảng cách từ mỗi đỉnh này tới v không vượt quá ε. Suy ra khoảng cách giữa u và w không vượt quá hiệu giữa tọa độ lớn nhất trong hai điểm và tọa độ của v, tức không vượt quá ε. Vậy u kề w, tạo thành một dây cung của chu trình. Điều này mâu thuẫn với giả thiết chu trình là cảm sinh.

Bước 2: Phức clique của một đồ thị chordal liên thông là co rút được. Ta quy nạp theo số đỉnh. Mọi đồ thị chordal đều có một đỉnh đơn hình v, tức là đỉnh mà tập lân cận $N(v)$ tạo thành một clique; hơn nữa, một đỉnh đơn hình không bao giờ là đỉnh cắt, nên đồ thị vẫn liên thông sau khi bỏ v. Ký hiệu $\mathrm{Cl}(G)$ là phức clique của G. Ta có $\mathrm{Cl}(G)$ là hợp của $\mathrm{Cl}(G - v)$ với hình nón đỉnh v trên đơn hình $N(v)$. Hình nón này co rút được, và giao của nó với $\mathrm{Cl}(G - v)$ đúng bằng đơn hình trên $N(v)$, cũng co rút được. Theo bổ đề dán, phép nhúng $\mathrm{Cl}(G - v)$ vào $\mathrm{Cl}(G)$ là một tương đương đồng luân. Vì đồ thị con cảm sinh của đồ thị chordal vẫn chordal, giả thiết quy nạp áp dụng được; trường hợp cơ sở một đỉnh là hiển nhiên. Vì phức Vietoris–Rips chính là phức clique của đồ thị Vietoris–Rips nên Mệnh đề 3 được chứng minh. □

Hệ quả 2. Gọi δ là khoảng cách Hausdorff từ X tới đường thẳng gần nó nhất. Khi đó, mọi đặc trưng trong giản đồ bền vững bậc $k \geq 1$ của X đều có độ bền không vượt quá 4δ.

Chứng minh. Gọi X' là hình chiếu vuông góc của X lên đường thẳng đó, khi đó khoảng cách Hausdorff giữa X và X' không vượt quá δ. Theo Hệ quả 1, khoảng cách bottleneck giữa hai giản đồ bền vững tương ứng không vượt quá 2δ. Theo Mệnh đề 3, giản đồ bậc $k \geq 1$ của X' là rỗng, nên mọi điểm của giản đồ bậc k của X phải cách đường chéo không quá 2δ theo chuẩn cực đại. Vì khoảng cách từ điểm (b, d) tới đường chéo bằng một nửa độ bền, ta thu được độ bền không vượt quá 4δ. □

Hai kết quả trên cho thấy tính chặt chẽ đối với chỉ báo được dùng trong nghiên cứu: độ bền tô-pô bị chặn trên bởi khoảng cách từ đám mây điểm tới

cấu trúc phụ thuộc tuyến tính hoàn hảo. Nói riêng, nếu bốn yếu tố khí hậu ràng buộc chặt theo một quan hệ hạng một thì chỉ báo tô-pô buộc phải nhỏ; ngược lại, chỉ báo tăng lên là bằng chứng cho việc ràng buộc đó lỏng đi. Cần lưu ý đây là trường hợp giới hạn cho quan hệ hạng một: nếu dữ liệu nằm trên một mặt phẳng thì đồng điều bậc một vẫn có thể khác không.

2.5. Thiết kế thí nghiệm

2.5.1. Ngăn ngừa rò rỉ dữ liệu

Nguyên tắc thiết kế cốt lõi trong phân tích là các biến mục tiêu được sử dụng để phân nhóm tuyệt đối không xuất hiện trong không gian đặc trưng. Trong cả hai ứng dụng thực nghiệm (phân cụm theo khoảng cách tới cửa sông và theo giai đoạn thời gian), nhãn phân nhóm đều được cách ly hoàn toàn khỏi véc-tơ dữ liệu được nhúng. Thiết kế này đảm bảo rằng các sai khác về cấu trúc tô-pô phát hiện được là những đặc tính bản chất của không gian pha, loại trừ khả năng quan sát thấy sự khác biệt chỉ như một hệ quả hiển nhiên (tautological) từ việc rò rỉ thông tin phân loại.

2.5.2. Kiểm định hoán vị

Để đánh giá ý nghĩa thống kê về sự khác biệt cấu trúc tô-pô giữa hai đám mây điểm A và B, phương pháp kiểm định hoán vị với thống kê kiểm định là khoảng cách bottleneck đã được sử dụng:

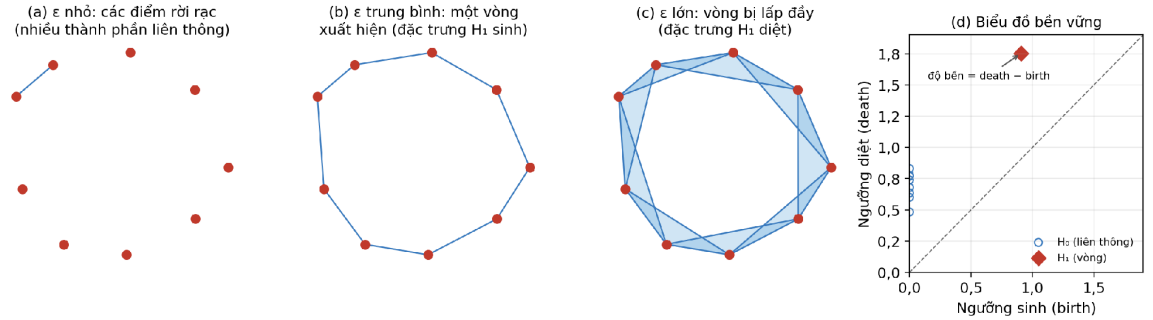
$$T = d_B\big(dgm(A), dgm(B)\big).$$

Dưới giả thuyết không (H_0) cho rằng hai mẫu được rút ra từ cùng một phân phối không gian cơ sở, phân phối không của thống kê T được mô phỏng bằng cách xáo trộn ngẫu nhiên nhãn nhóm của các quan trắc và tính toán lại giá trị T. Giá trị p (p-value) được ước lượng bằng tỉ lệ số lần hoán vị cho ra giá trị thống kê lớn hơn hoặc bằng giá trị T quan sát được trên tập dữ liệu gốc. Phương pháp tiếp cận phi tham số này không đòi hỏi bất kỳ giả định nào về dạng phân phối của dữ liệu, do đó đảm bảo tính vững và đặc biệt phù hợp với các thực nghiệm có cỡ mẫu nhỏ.

Cài đặt thuật toán và kiểm chứng

Toàn bộ quy trình tính toán được lập trình bằng ngôn ngữ Python, dựa trên hệ sinh thái mảng đa chiều NumPy. Các module được phát triển từ đầu bao gồm: xây dựng ma trận khoảng cách, thiết lập phức Vietoris–Rips, thuật toán rút gọn ma trận biên trên trường hữu hạn $\mathbb{F}_2$ (Otter et al., 2017), đối sánh, tính khoảng cách bottleneck và vòng lặp kiểm định hoán vị.

Để đánh giá tính chính xác, quy trình mã nguồn đã được kiểm chứng trên bộ dữ liệu mô phỏng một đường tròn có nhiễu. Kết quả trích xuất cấu trúc tô-pô cho thấy: lớp đồng điều bậc một (H_1) xuất hiện tại ngưỡng sinh $b \approx 0{,}404$ và bị lấp đầy tại ngưỡng diệt $d \approx 1{,}720$. Giá trị diệt thực nghiệm này rất gần với ngưỡng tới hạn lý thuyết đối với phức Vietoris–Rips của đường tròn đơn vị là $\sqrt{3} \approx 1{,}732$.



Hình 1. Minh họa phép lọc Vietoris–Rips và giản đồ bền vững trên một đám mây điểm dạng vòng

Hình 1 (a)-(c) lần lượt là các phức đơn hình khi ngưỡng lọc ε tăng dần, một vòng (đặc trưng H_1) sinh ra rồi bị lấp đầy; Hình 1(d) giản đồ bền vững tương ứng, mỗi điểm là một đặc trưng với hoành độ là ngưỡng sinh và tung độ là ngưỡng diệt, độ bền bằng hiệu ngưỡng diệt – ngưỡng sinh.

2.6. Nghiên cứu mô phỏng đánh giá năng lực phát hiện

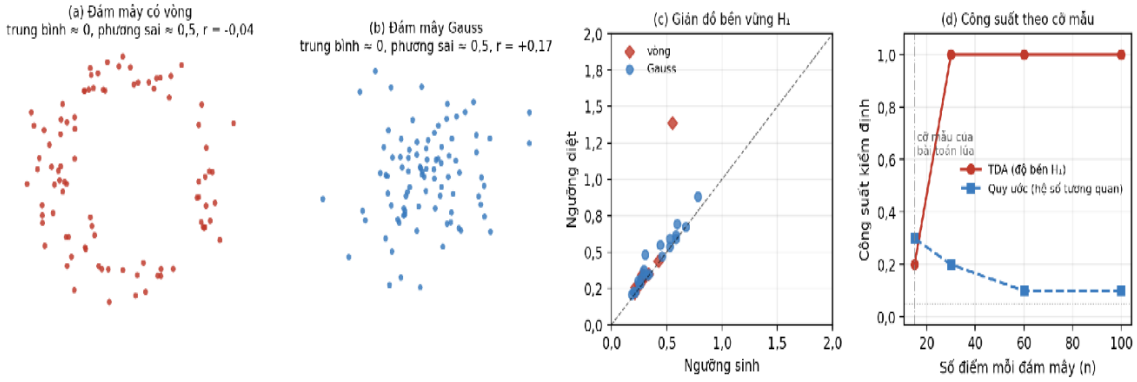
Vì bộ dữ liệu thực tế có cỡ mẫu hạn chế nên việc đánh giá năng lực phát hiện của chỉ báo tô-pô sẽ được thực nghiệm bằng một nghiên cứu mô phỏng có kiểm soát. Hai tổng thể được xây dựng sao cho trùng nhau về trung bình, phương sai và hệ số tương quan theo đúng cấu tạo, chỉ khác nhau về cấu trúc tô-pô: Tổng thể thứ nhất gồm các điểm

phân bố quanh một đường tròn có nhiễu, tổng thể thứ hai là phân phối chuẩn hai chiều có cùng ma trận hiệp phương sai. Theo cách xây dựng này, mọi thống kê chỉ dựa trên mô-men bậc một và bậc hai đều không thể phân biệt hai tổng thể, trong khi chúng khác nhau hoàn toàn về đồng điều bậc một.

Với mỗi cỡ mẫu n, việc mô phỏng sinh 15 đám mây cho mỗi tổng thể đã được thực hiện, độ bền cực đại H_1 được tính cho từng đám mây và hai nhóm được so sánh bằng kiểm định hoán vị; đồng thời, quy trình được lặp lại với thống kê quy ước là hệ số tương quan. Công suất được ước lượng qua 10 lần lặp độc lập ở mức ý nghĩa 0,05.

Hình 2(a) và Hình 2(b) hai đám mây có cùng trung bình, phương sai và hệ số tương quan nhưng khác cấu trúc tô-pô; Hình 2(c) giản đồ bền vững H_1 tách biệt rõ hai trường hợp; Hình 2(d) công suất kiểm định theo cỡ mẫu, đường đứt nét dọc đánh dấu cỡ mẫu của bài toán lúa. Kiểm định được thực hiện dựa trên hệ số tương quan có công suất dao động quanh mức nhiễu với mọi cỡ mẫu (30%, 20%, 10% và 10% lần lượt ứng với n=15, 30, 60 và 100), đúng như lý thuyết dự đoán vì hai tổng thể có cùng ma trận tương quan. Ngược lại, chỉ báo tô-pô đạt công suất tuyệt đối 100% từ n = 30 trở lên. Điều này cho thấy phân tích tô-pô nắm bắt được loại cấu trúc mà thống kê dựa trên mô-men bậc hai không thể tiếp cận, đúng như cơ sở lý thuyết ở mục 2.4.

Đáng chú ý, tại n = 15 công suất của chỉ báo tô-pô chỉ đạt 20%. Đây chính là cỡ mẫu của mỗi đám mây trong ứng dụng ở mục 3, khi mỗi tỉnh-giai đoạn chỉ có 15 quan trắc năm. Như vậy, ngưỡng cỡ mẫu để chỉ báo tô-pô phát huy hiệu lực nằm vào khoảng 30 điểm và kết quả mô phỏng này cung cấp căn cứ định lượng để diễn giải các hạn chế của phần ứng dụng.



Hình 2. Kết quả nghiên cứu mô phỏng

Kết quả này quyết định cách tổ chức phần thực nghiệm. Biết trước rằng ở cỡ mẫu mười lăm điểm công suất chỉ khoảng 20%, dùng chỉ báo tô-pô làm căn cứ duy nhất sẽ là thiết kế sai, vì khi đó kết luận "không có thay đổi" không phân biệt được với việc phép kiểm thiếu sức phát hiện. Vì vậy, phần thực nghiệm được tổ chức thành hai nhánh chạy song song trên cùng thiết kế chỉ báo tô-pô ở mục 3.1 và thống kê quy ước ở mục 3.2 và mọi kết luận về dữ liệu lúa đều dựa trên cả hai. Cách này trả lời được câu hỏi mà riêng nhánh tô-pô không trả lời được: khi hai chỉ báo lệch nhau, lệch đó do công suất thấp hay do chúng đo hai thứ khác nhau. Ở đây là do công suất thấp vì hai chỉ báo tương quan rõ rệt trên bốn mươi tám đám mây (mục 3.2). Đó cũng là lý do tên bài đặt hai cách tiếp cận ngang nhau.

2.7. Dữ liệu

Số liệu được lấy từ bộ dữ liệu công khai "Vietnamese Mekong Delta Rice Yield Dataset" trên nền tảng Kaggle (locduke2910, 2026); diện tích, sản lượng và các yếu tố khí hậu cho 13 tỉnh, thành ĐBSCL được tổng hợp theo đơn vị hành chính trước thời điểm sắp xếp năm 2025. Số liệu này cũng đã được đối chiếu với số liệu đã được công bố trong Niên giám Thống kê hằng năm của Tổng cục Thống kê (Tổng cục Thống kê, 2001-2024). Kết quả đối chiếu ngẫu nhiên cho thấy trùng khớp, ví dụ sản lượng lúa đông xuân tỉnh An Giang năm 2002 là 1.430,9 nghìn tấn ở cả hai nguồn.

Số liệu sản xuất lúa của 13 tỉnh, thành ĐBSCL trong 30 năm (1995 – 2024) theo hai nhóm mùa vụ (đông xuân, hè thu – thu đông), gồm 780 quan sát. Mỗi quan sát có diện tích, sản lượng và bốn yếu tố

khí hậu: tổng số giờ nắng, tổng lượng mưa, độ ẩm trung bình và nhiệt độ trung bình.

Sau khi loại 33 quan sát khuyết diện tích hoặc sản lượng, năng suất trung bình toàn vùng là 5,32 tấn/ha (vụ đông xuân 6,01, vụ hè thu – thu đông 4,65). Đây là dữ liệu quan trắc thực tế cấp tỉnh, phù hợp để khảo sát biến đổi cấu trúc đồng biến thiên của các yếu tố khí hậu theo thời gian. Toàn bộ mã nguồn tính toán được cài đặt bằng Python (NumPy) và sẵn sàng cung cấp để bảo đảm tính tái lập.

Bảng 1. Thống kê mô tả các biến trong mẫu phân tích

Biến	Đông xuân	Hè thu–thu đông
Tổng số giờ nắng (giờ)	1.572,75 ± 65,59	1.985,47 ± 101,31
Tổng lượng mưa (mm)	325,38 ± 167,99	1.717,23 ± 255,94
Độ ẩm trung bình (%)	77,20 ± 3,08	84,60 ± 1,72
Nhiệt độ trung bình (°C)	26,39 ± 0,54	27,11 ± 0,40
Diện tích (nghìn ha)	122,06 ± 85,99	160,58 ± 103,49
Sản lượng (nghìn tấn)	768,39 ± 573,60	773,18 ± 573,22

Ghi chú: 747 quan sát; 366 quan sát vụ đông xuân, 381 quan sát vụ hè thu – thu đông, trình bày dưới dạng trung bình ± độ lệch chuẩn.

3. KẾT QUẢ THẢO LUẬN

3.1. Biến đổi cấu trúc đồng biến thiên của các yếu tố khí hậu giai đoạn 1995 – 2024

Với mỗi tỉnh và mỗi mùa vụ, bốn yếu tố khí hậu tạo thành đám mây điểm trong không gian bốn chiều. Đám mây được chuẩn hóa z-score riêng cho từng tỉnh và từng giai đoạn; theo Mệnh đề 2, mọi thay đổi về trung bình và phương sai đều bị loại bỏ nên chỉ báo chỉ phản ánh hình dạng đồng biến thiên giữa bốn yếu tố. Hai giai đoạn 15 năm (1995 – 2009 và 2010 – 2024) được so sánh bằng kiểm định hoán vị với $B = 5.000$. Giai đoạn này biến nằm ngoài không gian đặc trưng nên không có rò rỉ nhãn.

Về mặt diễn giải, mỗi tỉnh – mùa vụ được biểu diễn bằng một đám mây điểm trong không gian bốn chiều, trong đó mỗi điểm ứng với một năm và bốn tọa độ là bốn yếu tố khí hậu đã chuẩn hóa. Hình dạng của đám mây phản ánh cách bốn yếu tố biến thiên đồng thời qua các năm: nếu chúng ràng buộc chặt theo một quan hệ đơn giản thì đám mây co về một cấu trúc thấp chiều (ít vòng, ít cụm); ngược lại nếu quan hệ giữa chúng trở nên phức tạp và đa chế độ thì đám mây sinh thêm các vòng và cụm, làm tăng độ phức tạp tô-pô. Do đó, đo độ phức tạp tô-pô của đám mây khí hậu chính là đo mức độ phức tạp của cấu trúc đồng biến thiên giữa các yếu tố khí hậu – thông tin mà trung bình hay phương sai của từng yếu tố riêng lẻ không thể hiện được.

Kết quả (Bảng 2, Hình 3) cho thấy vụ hè thu – thu đông có độ phức tạp tô-pô tăng có ý nghĩa giữa hai giai đoạn và kết luận bền vững ở cả ba ngưỡng lọc được thử (p lần lượt là 0,0002; 0,0012 và 0,0086). Ngược lại, vụ đông xuân không thay đổi ở bất kỳ ngưỡng nào ($p \geq 0{,}6485$). Với sáu phép kiểm, ngưỡng Bonferroni là 0,0083, kết quả của vụ hè thu vẫn có ý nghĩa ở hai ngưỡng đầu.

Bối cảnh khí hậu giúp diễn giải: ở vụ hè thu – thu đông, giữa hai giai đoạn, tổng số giờ nắng trung bình giảm từ 2.015,7 xuống 1.956,6 giờ, tổng lượng mưa tăng từ 1.648,7 lên 1.782,6 mm với độ lệch chuẩn tăng từ 228,2 lên 264,3 mm, nhiệt độ trung bình tăng $0{,}30°C$. Tuy nhiên, do phép chuẩn hóa đã loại bỏ các dịch chuyển về trung bình và phương sai (Mệnh đề 2) nên khác biệt tô-pô ghi nhận được không do các dịch chuyển này mà phản ánh sự tái cấu trúc quan hệ đồng biến thiên giữa bốn yếu tố khí hậu. Mục 3.2 đối chứng nhận định này bằng các kiểm định thống kê quy ước.

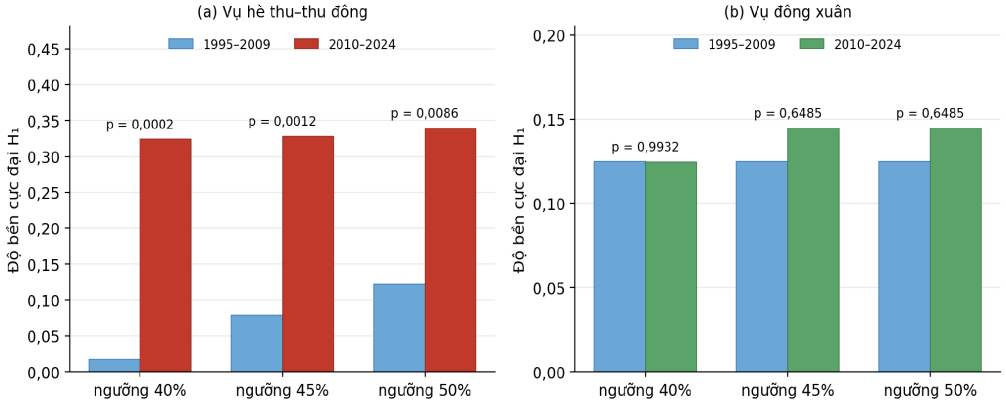
Hai phân tích bổ sung được thực hiện. Thứ nhất, nhóm tỉnh ven biển được so sánh với nhóm nội đồng không cho thấy sự khác biệt có ý nghĩa ở cả hai mùa vụ ($p = 0{,}586$ và 0,804), cho thấy chỉ báo không đơn thuần phản ánh vị trí địa lý. Thứ hai, sự tương quan giữa độ bền của không gian khí hậu và hệ số biến thiên năng suất cấp tỉnh là r = -0,497 ở vụ hè thu – thu đông ($n = 13$, $p \approx 0{,}08$) và $r = 0{,}160$ ở vụ đông xuân, xu hướng âm đáng chú ý nhưng chưa đạt mức ý nghĩa với cỡ mẫu 13 tỉnh.

Cần nhấn mạnh rằng sự thay đổi cấu trúc đồng biến thiên này mang ý nghĩa mô tả, không đồng nghĩa với khả năng dự báo năng suất. Trong một thử nghiệm dự báo ngoài mẫu (huấn luyện 1995 – 2016, kiểm tra 2017 – 2024) đối với phần bất

thường của năng suất (đã trừ xu thế theo thời gian), cả bốn yếu tố khí hậu lẫn các đặc trưng tô-pô đều không dự báo được năng suất (hệ số xác định R^2 xấp xỉ 0 hoặc âm) và việc thêm đặc trưng tô-pô không cải thiện một cách ổn định. Kết quả này phù hợp với thực tế nông học: Năng suất lúa ĐBSCL giai đoạn 1995 – 2024 chịu chi phối chủ yếu bởi tiến bộ công nghệ, giống và thủy lợi, tức phần xu thế đã được loại bỏ nên phần bất thường còn lại không được giải thích bởi biến thiên khí hậu ở quy mô năm – tỉnh. Do đó, đặc trưng tô-pô ở đây có giá trị phát hiện và cảnh báo thay đổi cấu trúc khí hậu, hơn là dự báo năng suất trực tiếp.

Bảng 2. Độ bền của không gian khí hậu theo giai đoạn, dữ liệu lúa ĐBSCL

Mùa vụ	Ngưỡng lọc	1995 – 2009	2010 – 2024	p
Hè thu–thu đông	40%	0,018	0,324	0,0002
Hè thu–thu đông	45%	0,079	0,328	0,0012
Hè thu–thu đông	50%	0,122	0,339	0,0086
Đông xuân	40%	0,125	0,124	0,9932
Đông xuân	45%	0,125	0,145	0,6485
Đông xuân	50%	0,125	0,145	0,6485



Hình 3. Độ bền của không gian khí hậu giữa hai giai đoạn

Ghi chú: (a) vụ hè thu – thu đông và (b) vụ đông xuân.

3.2. Đối chứng với kiểm định thống kê quy ước

Một tuyên bố được đưa ra là đặc trưng tô-pô nắm bắt được thông tin mà thống kê quy ước khó thấy cần được kiểm chứng chứ không thể mặc định. Chúng tôi lặp lại đúng thiết kế ở mục 3.1 cùng dữ liệu, cùng cách chuẩn hóa, cùng phân nhóm theo giai đoạn và cùng kiểm định hoán vị với B = 5.000, nhưng thay chỉ báo tô-pô bằng các thống kê quy ước tính trên ma trận tương quan của bốn yếu tố khí hậu: tổng bình phương hệ số tương quan ngoài đường chéo, $-\log(\det(R))$, trị riêng lớn nhất của R và khoảng cách Frobenius giữa ma trận tương quan trung bình của hai giai đoạn.

Bảng 3. Đối chứng giữa chỉ báo tô-pô và các thống kê quy ước về cấu trúc tương quan (giá trị p của kiểm định hoán vị, B = 5.000)

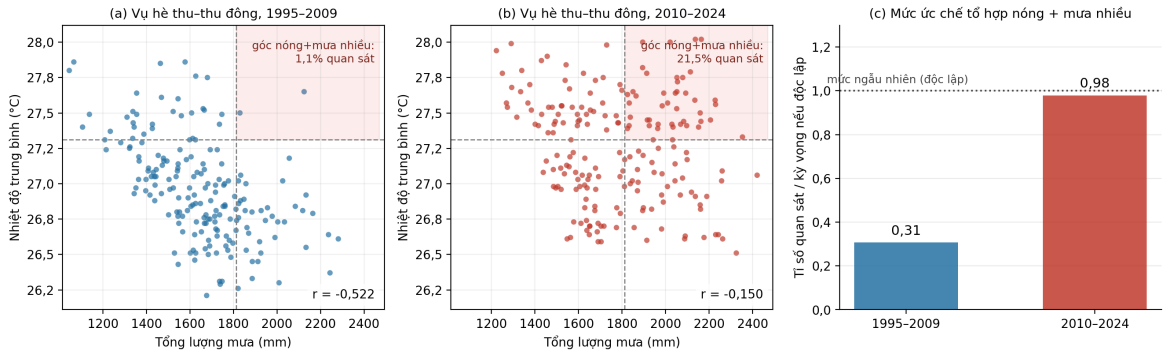
Mùa vụ	Độ bền tô-pô H_1 (Mục 3.1)	Tổng bình phương hệ số tương quan	$-\log(\det(R)$	$\lambda\max(R)$	Frobenius giữa hai R
Hè thu–thu đông	0,0002 – 0,0086	0,0002	0,0098	0,0002	0,0002
Đông xuân	0,6485 – 0,9932	0,0002	0,0002	0,0002	0,0002

Kết quả (Bảng 3) cho thấy các kiểm định quy ước phát hiện thay đổi cấu trúc ở *cả hai* mùa vụ với mức ý nghĩa cao (p = 0,0002); trong đó, vụ đông xuân mà phân tích tô-pô kết luận là không đổi lại thay đổi rất rõ. Như vậy, ở bộ dữ liệu này chỉ báo tô-pô không nhạy hơn mà kém nhạy hơn các kiểm định quy ước. Điều này phù hợp với hạn chế về cỡ mẫu đã nêu: mỗi đám mây chỉ gồm 15 điểm trong không gian bốn chiều, quá thưa để đồng điều bậc một ổn định.

Hai cách tiếp cận đo cùng một hiện tượng. Trên bốn mươi tám đám mây (tỉnh × mùa vụ × giai đoạn), hệ số tương quan giữa tổng bình phương hệ số tương quan và độ bền cực đại H_1 là $r = -0{,}394$: mức phụ thuộc tuyến tính càng yếu thì độ phức tạp tô-pô càng cao. Nói cách khác, "độ phức tạp tô-pô tăng" ở mục 3.1 chính là biểu hiện gián tiếp của việc các ràng buộc tuyến tính giữa bốn yếu tố khí hậu lỏng dần.

Ưu thế của phân tích quy ước là chỉ ra được quan hệ nào đã thay đổi. Ở vụ hè thu – thu đông, tương quan giữa tổng lượng mưa và nhiệt độ trung bình giảm mạnh về trị tuyệt đối, từ −0,589 xuống −0,159, tương quan giữa tổng lượng mưa và độ ẩm giảm từ +0,751 xuống +0,468. Ở vụ đông xuân, các thay đổi cùng chiều nhưng nhỏ hơn (mưa × nhiệt độ: −0,674 → −0,449). Nghĩa là ràng buộc "mưa nhiều thì mát và ẩm hơn" đã lỏng đi rõ rệt giữa hai giai đoạn, một kết quả cụ thể và diễn giải được về mặt khí hậu nông nghiệp mà riêng chỉ báo tô-pô không cung cấp.

Ràng buộc lỏng đi có một hệ quả đo được trực tiếp, không phải chỉ là cách diễn đạt khác của hệ số tương quan: nếu những tổ hợp thời tiết trước đây gần như không xảy ra thì nay xảy ra thường xuyên. Gọi một quan trắc tỉnh – năm của vụ hè thu – thu đông là "nóng" nếu nhiệt độ trung bình thuộc một phần ba cao nhất và "mưa nhiều" nếu tổng lượng mưa thuộc một phần ba cao nhất. Nếu hai yếu tố độc lập thì tổ hợp nóng và mưa nhiều phải chiếm khoảng 3,5% quan trắc ở giai đoạn 1995–2009; thực tế chỉ có 1,1%, bằng 0,31 lần mức kỳ vọng. Ở giai đoạn 2010 – 2024, tần suất quan sát là 21,5% so với kỳ vọng 22,0%, tức tỉ số 0,98: sự ức chế đã biến mất gần như hoàn toàn. Mức chênh 20,4 điểm phần trăm có ý nghĩa thống kê theo kiểm định hoán vị trên nhãn giai đoạn (B = 5.000, p = 0,0002). Ở vụ đông xuân, dịch chuyển cùng chiều nhưng yếu hơn: tỉ số tăng từ 0,15 lên 0,41 (p = 0,0244). Hình 4 trình bày kết quả này.



Hình 4. Tổ hợp nóng và mưa nhiều ở vụ hè thu – thu đông

Ghi chú: (a) giai đoạn 1995 – 2009, (b) giai đoạn 2010 – 2024 (đường gạch là phân vị 2/3 của từng yếu tố và vùng tô là góc nóng đi kèm mưa nhiều) (c) tỉ số giữa tần suất quan sát của góc này và tần suất kỳ vọng nếu hai yếu tố độc lập.

Kết quả này cho thấy sự phân công giữa hai công cụ chứ không phải sự thay thế. Chỉ báo tô-pô trả lời "cấu trúc đã đổi hay chưa" bằng một bất biến không cần giả định phân phối và không phụ thuộc thang đo; thống kê quy ước trả lời "đổi cái gì". Trên bộ dữ liệu này nhánh thứ hai cho nội dung diễn giải được vì hiện tượng nằm gọn trong mô-men bậc hai; khi hiện tượng không nằm ở đó thì nhánh thứ hai bất lực về nguyên tắc chứ không phải vì thiếu dữ liệu, như mô phỏng ở mục 2.6 cho thấy. Về mặt hình học, ràng buộc mưa – nhiệt ép đám mây điểm về gần một đường thẳng nên khi nó lỏng ra, đám mây rời khỏi đường thẳng và độ bền bậc một tăng; Hệ quả 2 là cầu nối hình thức giữa hai cách nói. Về nông học, góc nóng đi kèm mưa nhiều là góc bất lợi nhất cho vụ hè thu–thu đông: nền nhiệt cao làm tăng hô hấp đêm và rút ngắn giai đoạn vào hạt, mưa lớn tập trung gây ngập úng và đổ ngã, trong khi điều kiện nóng ẩm đồng thời thuận lợi cho bệnh hại lá và bẹ (Kontgis et al., 2019; Nguyen & Yen, 2021). Đây là giả thuyết nông học sinh ra từ phân tích khí hậu, chưa được kiểm chứng trong nghiên cứu này.

Từ đó có thể phát biểu chính xác đóng góp của mỗi cách tiếp cận, thay cho một tuyên bố chung về ưu thế. Về độ nhạy, chỉ báo tô-pô kém hơn và kém một cách có thể chỉ tên: nó bỏ sót vụ đông xuân mà kiểm định quy ước phát hiện ở p = 0,0002. Về nội dung kết luận, hai cách tiếp cận trùng nhau nơi hiệu ứng đủ mạnh ($r = -0{,}394$ trên bốn mươi tám đám mây). Khác biệt còn lại nằm ở phạm vi giả thiết: chỉ báo tô-pô không giả định phụ thuộc là tuyến tính và bất biến với phép đồng dạng. Trên bộ dữ liệu này ưu

điểm đó không đổi thành lợi thế phát hiện và Hệ quả 2 nói rõ vì sao: quan hệ giữa bốn yếu tố khí hậu ở quy mô năm – tỉnh về cơ bản tuyến tính nên đám mây nằm gần một đường thẳng và độ bền bậc một bị chặn nhỏ. Kết quả bất lợi ở đây vì vậy không phải bằng chứng chống lại phương pháp, mà là trường hợp lý thuyết đã dự báo và mô phỏng đã định lượng trước.

3.3. Hạn chế của nghiên cứu

Nghiên cứu có bốn hạn chế. Thứ nhất, dữ liệu tuy dài 30 năm nhưng chỉ có 13 đơn vị không gian (cấp tỉnh), làm hạn chế độ phân giải của giản đồ bền vững, việc mở rộng xuống cấp huyện sẽ tăng đáng kể số điểm và độ tin cậy. Thứ hai, việc nghiên cứu mô tả sự thay đổi cấu trúc chứ chưa thiết lập quan hệ nhân quả với biến đổi khí hậu, vì giống lúa, kỹ thuật canh tác và hệ thống thủy lợi cũng thay đổi trong cùng giai đoạn. Thứ ba, các đặc trưng tô-pô mô tả hình dạng nên không thay thế được các chỉ số định lượng truyền thống mà đóng vai trò bổ sung. Thứ tư, kết quả mục 3.2 cho thấy cỡ mẫu hiện tại (15 điểm cho mỗi đám mây) chỉ báo tô-pô kém nhạy hơn các kiểm định dựa trên ma trận tương quan và đã bỏ sót thay đổi ở vụ đông xuân; ngoài ra, những thay đổi về mạng lưới trạm quan trắc và cách tổng hợp số liệu trong 30 năm cũng là một cách giải thích thay thế chưa loại trừ được.

Về quy kết nguyên nhân cần phân biệt hai tầng. Bốn biến được phân tích đều là quan trắc khí hậu và không chịu ảnh hưởng của đê bao, thủy lợi hay cơ cấu giống, nên thay đổi cấu trúc ghi nhận được là thay đổi bên trong hệ khí hậu; nghiên cứu không quy kết nó cho canh tác. Điều không kết luận được là hệ quả của thay đổi đó đối với sản xuất lúa: trong cùng giai đoạn 1995 – 2024, hạ tầng thủy lợi, đê bao khép kín và giống lúa đều biến động mạnh nên có thể che lấp, bù trừ hoặc khuếch đại tác động khí hậu và thử nghiệm dự báo ở mục 3.1 cho thấy đúng tình trạng đó. Kết quả nghiên cứu cũng không quy kết thay đổi cấu trúc cho biến đổi khí hậu toàn cầu: ba mươi năm là quãng ngắn so với thang khí hậu và biến thiên nội tại nhiều thập niên của chế độ gió mùa là cách giải thích thay thế chưa loại trừ được.

3.4. Khả năng tự động hóa và tính tái lập

Quy trình được cài đặt thành các hàm Python độc lập trên NumPy, nhận vào bảng quan trắc dạng tỉnh–năm–yếu tố và trả ra giản đồ bền vững cùng giá trị p, không có bước nào cần can thiệp thủ công. Toàn bộ bốn mươi tám đám mây và 5.000 lần hoán vị cho mỗi phép kiểm hoàn tất trong vài phút trên một máy tính cá nhân nên quy trình có thể đóng gói thành tác vụ định kỳ chạy mỗi khi Niên giám Thống kê công bố số liệu năm mới. Hai điều kiện kỹ thuật bắt buộc: cố định trước ngưỡng lọc theo một quy tắc dựa trên phân vị khoảng cách để các lần chạy so sánh được với nhau, và kiểm soát sai số loại một cho kiểm định tuần tự. Theo mục 2.6, hệ tự động chỉ nên triển khai khi mỗi đám mây đạt ít nhất ba mươi điểm. Về tính tái lập, dữ liệu dùng trong bài là bộ công khai "Vietnamese Mekong Delta Rice Yield Dataset" trên Kaggle (locduke2910, 2026), địa chỉ https://www.kaggle.com/datasets/locduke2910/vietnamese-mekong-delta-rice-yield-dataset; số liệu đơn vị hành chính lấy từ Niên giám Thống kê tại https://www.nso.gov.vn/nien-giam/.

4. KẾT LUẬN

Nghiên cứu đã trình bày cơ sở toán học của phân tích dữ liệu tô-pô như nền phương pháp và chứng minh ba mệnh đề bổ trợ. Mệnh đề 1 và Mệnh đề 2 là tính chất đã biết, được chứng minh lại vì dùng trực tiếp ở mục 3; riêng Mệnh đề 3 cùng Hệ quả 2 là kết quả mới theo hiểu biết của chúng tôi: độ bền của mọi đặc trưng bậc từ một trở lên bị chặn bởi bốn lần khoảng cách Hausdorff từ đám mây điểm tới đường thẳng gần nó nhất. Hệ quả này cho chỉ báo tô-pô một cách đọc chặt chẽ là thước đo mức lệch khỏi phụ thuộc tuyến tính hoàn hảo và quan trọng hơn bài viết cung cấp một tiêu chuẩn sàng lọc tính được trước khi phân tích.

Hai nhánh thực nghiệm cho hai kết quả bổ sung nhau. Mô phỏng có kiểm soát dựng hai tổng thể trùng nhau đến mô-men bậc hai, trường hợp mọi kiểm định dựa trên tương quan không giải quyết được về nguyên tắc chứ không phải vì thiếu dữ liệu, chỉ báo tô-pô phân biệt được với công suất tiến tới một từ cỡ mẫu ba mươi điểm, nhưng chỉ khoảng 20% tại cỡ mẫu mười lăm. Kết quả khảo sát trên dữ liệu khí hậu và sản xuất lúa cấp tỉnh cho thấy nơi mỗi đám mây đúng mười lăm điểm, cho kết quả ngược lại: chỉ báo tô-pô phát hiện thay đổi cấu trúc ở vụ hè thu–thu đông ($p \leq 0{,}0086$) nhưng bỏ sót vụ đông xuân mà kiểm định trên ma trận tương quan phát hiện ở $p = 0{,}0002$. Trường hợp này có hai vai trò. Nó là phản ví dụ cho nhận định thường gặp rằng phân tích tô-pô nắm bắt được cấu trúc mà thống kê quy ước không thấy, ở đây điều ngược lại xảy ra. Đồng thời nó xác nhận đúng ngưỡng cỡ mẫu mà mô phỏng đã dự báo trước, nên là bằng chứng kiểm nghiệm chứ không phải một sự cố ngẫu nhiên.

Về ý nghĩa thực tiễn, kết quả không nói rằng khí hậu vụ hè thu – thu đông đã xấu đi theo một chỉ số trung bình nào, mà nói rằng cấu trúc phụ thuộc giữa các yếu tố đã đổi: tổ hợp nóng đi kèm mưa nhiều, từ chỗ bị ức chế xuống còn 0,31 lần mức ngẫu nhiên,

nay xuất hiện ở 0,98 lần mức ngẫu nhiên (mục 3.2). Hệ quả thao tác là các kịch bản rủi ro không còn suy ra được từ phân phối biên của từng yếu tố riêng lẻ: một giống chịu nóng tốt nhưng kém chịu ngập, hay một lịch gieo sạ tối ưu theo lượng mưa trung bình, có thể an toàn khi xét từng yếu tố mà vẫn gặp rủi ro ở tổ hợp đồng thời nay đã trở nên thường gặp. Điều này gợi ý đưa tần suất đồng xuất hiện của các tổ hợp bất lợi vào bộ chỉ tiêu theo dõi khí hậu nông nghiệp của vùng, bên cạnh các trung bình mùa vụ. Đây là hàm ý rút từ phân tích khí hậu, chưa phải khuyến nghị đã kiểm chứng trên số liệu năng suất.

Tóm lại, sản phẩm sử dụng được của nghiên cứu là một quy tắc kiểm tra trước khi áp dụng chỉ báo tô-pô cho dữ liệu thống kê vùng: mỗi đám mây cần từ ba mươi điểm trở lên và đám mây không được nằm quá gần một đường thẳng theo nghĩa Hệ quả 2. Cả hai điều kiện đều tính được từ dữ liệu trước khi chạy bất kỳ phép lọc nào nên chi phí kiểm tra là không đáng kể so với chi phí một phân tích sai hướng. Hướng tiếp theo vì vậy không phải thay phương pháp mà là đưa dữ liệu vượt ngưỡng, theo hai phương án. Phương án thứ nhất là mở rộng phạm vi ra cả nước, nâng số đơn vị không gian từ 13 tỉnh lên toàn bộ các tỉnh, thành. Phương án thứ hai là giữ phạm vi ĐBSCL nhưng hạ đơn vị phân tích xuống cấp huyện rồi cấp xã: theo Niên giám Thống kê, toàn vùng có 1.253 xã tính đến ngày 31 tháng 12 năm 2023 (Tổng cục Thống kê, 2001-2024), gấp khoảng chín mươi lần số đơn vị đang dùng. Phương án thứ hai hứa hẹn hơn vì vừa vượt xa ngưỡng cỡ mẫu vừa cho phép gộp các đơn vị trong cùng một huyện thành đám mây lớn hơn, nó đòi hỏi bổ sung dữ liệu viễn thám hoặc dữ liệu khí hậu dạng lưới vì thống kê hành chính không có số liệu khí hậu dưới cấp tỉnh.

TÀI LIỆU THAM KHẢO

Anh, Đ. T. (2018). Nhu cầu đổi mới công nghệ trong sản xuất và sau thu hoạch lúa gạo của Việt Nam. *Tạp chí Khoa học & Công nghệ Việt Nam*, *6*, 16-18. https://vjol.info.vn/khcn/vi/article/view/36225/29605/

Anh, N. H. (2021). Phát triển kinh tế nông nghiệp bền vững và thích ứng biến đổi khí hậu tại Đồng bằng sông Cửu Long. *Tạp chí Công Thương*, *25*, 122-129. https://tapchicongthuong.vn/phat-trien-kinh-te-nong-nghiep-ben-vung-va-thich-ung-bien-doi-khi-hau-tai-dong-bang-song-cuu-long-85863.htm

Bubenik, P. (2015). Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research, 16*(1), 77–102. https://www.jmlr.org/papers/volume16/bubenik15a/bubenik15a.pdf

Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*, *46*(2), 255–308. https://doi.org/10.1090/S0273-0979-09-01249-X

Chazal, F., & Michel, B. (2021). An introduction to topological data analysis: Fundamental and practical aspects for data scientists. *Frontiers in Artificial Intelligence*, *4*, 667963. https://doi.org/10.3389/frai.2021.667963

Cohen-Steiner, D., Edelsbrunner, H., & Harer, J. (2007). Stability of persistence diagrams. *Discrete & Computational Geometry, 37*(1), 103–120. https://doi.org/10.1007/s00454-006-1276-5

Đan, N. H., Thời, N. K., & Dung, B. T. N. (2015). Đánh giá tình hình sử dụng đất lúa ở Đồng bằng sông Cửu Long. *Tạp chí Khoa học và Phát triển*, *13*(8), 1435–1441. https://vie.vjas.vn/index.php/vjasvn/article/view/251

Diễm, P. K., & Diễm, N. K. (2020). Đánh giá sự thay đổi tổng sản lượng sơ cấp theo thời gian của các kiểu thực phủ khu vực Đồng bằng sông Cửu Long. *Tạp chí Khoa học Đại học Cần Thơ*, *56*(6), 57–68. https://doi.org/10.22144/ctu.jvn.2020.143

Edelsbrunner, H., & Harer, J. (2010). *Computational topology: an introduction*. American Mathematical Society. https://doi.org/10.1090/mbk/069

Fasy, B. T., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S., & Singh, A. (2014). Confidence sets for persistence diagrams. *The Annals of Statistics*, *42*(6), 2301–2339. https://doi.org/10.1214/14-AOS1252

Hương, T. T. T. (2026). Chuyển đổi cơ cấu cây trồng, vật nuôi trên đất trồng lúa: Những vấn đề đặt ra ở vùng Đồng bằng sông Cửu Long. *Tạp chí Khoa học Đại học Cửu Long*, *43*, 3–11. https://doi.org/10.65934/mkusj.2026.43.1042

Khôi, P. Đ., Khải, H. V., Danh, V. T., Trúc, N. T. T., & Mỹ, N. T. H. (2022). Lựa chọn mô hình sản xuất của hộ trồng lúa trong điều kiện xâm nhập mặn ở Đồng bằng sông Cửu Long. *Tạp chí Kinh tế và Phát triển*, *297*, 74–82. https://js.ktpt.edu.vn/index.php/jed/article/view/299

Kontgis, C., Schneider, A., Ozdogan, M., Kucharik, C., Tri, V. P. D., Duc, N. H., & Schatz, J. (2019). Climate change impacts on rice productivity in the Mekong River Delta. *Applied Geography*, *102*, 71–83. https://doi.org/10.1016/j.apgeog.2018.12.004

Lộc, V. T. T., & Son, N. P. (2011). Phần 1: Phân tích chuỗi giá trị lúa gạo vùng Đồng bằng sông Cửu Long. *Tạp chí Khoa học Đại học Cần Thơ*, (19a), 96–108. https://ctujsvn.ctu.edu.vn/index.php/ctujsvn/article/view/1048

locduke2910. (2026). *Vietnamese Mekong Delta rice yield dataset*. Kaggle. https://www.kaggle.com/datasets/locduke2910/vietnamese-mekong-delta-rice-yield-dataset

Muszynski, G., Kashinath, K., Kurlin, V., Wehner, M., & Prabhat. (2019). Topological data analysis and machine learning for recognizing atmospheric river patterns in large climate datasets. *Geoscientific Model Development*, *12*(2), 613–628. https://doi.org/10.5194/gmd-12-613-2019

Nguyen, V. H., & Yen, H. P. H. (2021). Seasonal variation and its impacts in rice-growing regions of the Mekong Delta. *International Journal of Climate Change Strategies and Management*, *13*(4–5), 483–491. https://doi.org/10.1108/IJCCSM-05-2020-0048

Nguyen, V. K., Nguyen, H. H., & Cramb, R. (2020). Trends in rice-based farming systems in the Mekong Delta. In R. Cramb (Eds.), *White gold: The commercialisation of rice farming in the Lower Mekong Basin* (347–373). Springer. https://doi.org/10.1007/978-981-15-0998-8_17

Nhẫn, P. P., Hoa, L. V., Qui, C. N., Huy, N. X., Hữu, T. P., Macdonald, B. C. T., & Tường, T. P. (2016). Increasing profitability and water use efficiency of triple rice crop production in the Mekong Delta, Vietnam. *The Journal of Agricultural Science*, *154*(6), 1015–1025. https://doi.org/10.1017/S0021859615000957

Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., & Harrington, H. A. (2017). A roadmap for the computation of persistent homology. *EPJ Data Science*, *6*(1), 17. https://doi.org/10.1140/epjds/s13688-017-0109-5

Phụng, H. P., & Nguyên, L. Đ. (2019). Giám sát mùa vụ lúa ở Đồng bằng sông Cửu Long và Đồng bằng sông Hồng sử dụng dữ liệu viễn thám Sentinel-1. *Kỷ yếu Hội nghị: Nghiên cứu cơ bản trong "Khoa học Trái đất và Môi trường"* (230–233). https://doi.org/10.15625/vap.2019.000123

Sakamoto, T., Van Nguyen, N., Ohno, H., Ishitsuka, N., & Yokozawa, M. (2006). Spatio–temporal distribution of rice phenology and cropping systems in the Mekong Delta with special reference to the seasonal water flow of the Mekong and Bassac rivers. *Remote Sensing of Environment*, *100*(1), 1–16. https://doi.org/10.1016/j.rse.2005.09.007

Tổng cục Thống kê. (2001-2024). *Niên giám Thống kê*. Nhà xuất bản Thống kê. https://www.nso.gov.vn/nien-giam/

Tuyến, N. Q. (2013). Hệ thống hóa các mô hình sản xuất lúa vùng sinh thái ngọt Đồng bằng sông Cửu Long. *Tạp chí Khoa học Đại học Cần Thơ*, (29), 60–69. https://ctujsvn.ctu.edu.vn/index.php/ctujsvn/article/view/1716

Tymochko, S., Munch, E., Dunion, J., Corbosiero, K., & Torn, R. (2020). Using persistent homology to quantify a diurnal cycle in hurricanes. *Pattern Recognition Letters*, *133*, 137–143. https://doi.org/10.1016/j.patrec.2020.02.022

Ver Hoef, L., Adams, H., King, E. J., & Ebert-Uphoff, I. (2023). A primer on topological data analysis to support image analysis tasks in environmental science. *Artificial Intelligence for the Earth Systems*, *2*(1), 1-18. https://doi.org/10.1175/AIES-D-22-0039.1

Wassmann, R., Hien, N. X., Hoanh, C. T., & Tuong, T. P. (2004). Sea level rise affecting the Vietnamese Mekong Delta: Water elevation in the flood season and implications for rice production. *Climatic Change*, *66*(1), 89–107. https://doi.org/10.1023/B:CLIM.0000043144.69736.b7

Zomorodian, A., & Carlsson, G. (2005). Computing persistent homology. *Discrete & Computational Geometry*, *33*(2), 249–274. https://doi.org/10.1007/s00454-004-1146-y