

Bài báo Khoa học

Nghiên cứu áp dụng thử nghiệm mô hình mạng Bayes trong đánh giá các yếu tố ảnh hưởng đến diễn biến cửa sông Đà Diễn

Hoàng Thu Thảo^{1*}, Trần Ngọc Anh^{1,2}, Trần Kiều Hương³

¹ Khoa Khí tượng Thủy văn và Hải dương học, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội; htthao@hus.edu.vn;

² Trung tâm Động lực học Thủy khí Môi trường, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội; tranngocanh@hus.edu.vn

³ Bộ Giáo dục và Đào tạo; tkhuong@moet.gov.vn

* Tác giả liên hệ: htthaovnu@gmail.com; Tel: +84-982981994

Ban Biên tập nhận bài: 26/1/2021; Ngày phản biện xong: 18/3/2021; Ngày đăng bài: 25/4/2021

Tóm tắt: Mô hình mạng Bayes (Bayesian Networks) là mô hình ứng dụng lý thuyết xác suất biểu diễn mối quan hệ nhân quả giữa các dữ liệu, đặc biệt là các dữ liệu không chắc chắn. Trên thế giới, mô hình mạng Bayes đã được ứng dụng trong nhiều lĩnh vực như y tế, ngôn ngữ học, sinh học, môi trường... và thủy văn và tài nguyên nước. Mục tiêu của bài báo này là ứng dụng thử nghiệm mô hình mạng Bayes trong đánh giá các yếu tố ảnh hưởng đến diễn biến cửa sông (tập trung phân tích biến thiên độ rộng cửa sông). Kết quả nghiên cứu cho thấy, năng lượng sóng có vai trò ảnh hưởng đáng kể đến biến thiên cửa sông Đà Diễn, đặc biệt là giai đoạn từ tháng 1 đến tháng 4 và giai đoạn từ tháng 10 đến tháng 12 hàng năm là các giai đoạn gió mùa Đông Bắc chiếm ưu thế. Giai đoạn từ tháng 5 đến tháng 9, cửa sông Đà Diễn có sự ổn định hơn và không thấy rõ được sự tác động vượt trội của yếu tố nào trong giai đoạn này. Các kết quả của nghiên cứu này phù hợp với các nhận định trước đó về cửa sông Đà Diễn và cho thấy khả năng ứng dụng linh hoạt của mô hình mạng Bayes vào nghiên cứu về đánh giá các yếu tố ảnh hưởng đến diễn biến cửa sông nói riêng và trong lĩnh vực khí tượng thủy văn nói chung.

Từ khóa: Mạng Bayes; Cửa sông Đà Diễn; Độ rộng cửa sông; Xác suất có điều kiện.

1. Mở đầu

Mạng Bayes (*Bayesian Network –BN*) là mô hình đồ họa xác suất để biểu diễn kiến thức về miền không chắc chắn trong đó mỗi nút của mạng tương ứng với một biến ngẫu nhiên và mỗi cạnh biểu thị xác suất có điều kiện cho các biến ngẫu nhiên tương ứng [1]. Mô hình này xuất hiện từ nghiên cứu trí tuệ nhân tạo và đã được áp dụng cho nhiều vấn đề, từ phân tích văn bản [2], đến các vấn đề trong chẩn đoán y tế [3] và đánh giá bằng chứng khoa học [4] và ngày càng được sử dụng nhiều hơn trong quản lý và lập mô hình giám sát và quản lý tài nguyên và môi trường [5–8].

Pearl phát biểu rằng mạng Bayes là mô hình đồ họa chứa thông tin về mối quan hệ xác suất nhân quả giữa các biến và thường được sử dụng để hỗ trợ việc ra quyết định [9]. Các mối quan hệ xác suất nhân quả trong mạng Bayes có thể được hình thành bởi dữ liệu sẵn có, sử dụng định luật Bayes hoặc được đề xuất bởi các chuyên gia. Cấu trúc phụ thuộc giữa các biến được biểu diễn bằng các nút (mô tả các biến) và các cạnh có hướng (mô tả các mối quan hệ điều kiện) dưới dạng một đồ thị xoay chiều có hướng (*Directed Acyclic Graph–DAG*). Các nút được coi là có quan hệ “cha mẹ–con cái” dựa trên sự phụ thuộc và hướng của các

cạnh liên kết trong mô hình mạng. Nếu nút trên mạng không có cạnh liên kết đến hay được gọi là không có nút “cha mẹ” ảnh hưởng (biến không phụ thuộc) thì nút đó có một phân phối xác suất ngẫu nhiên. Ngược lại, nếu nút “con” có sự phụ thuộc vào một hay nhiều nút “cha mẹ” thì nó có một phân phối xác suất có điều kiện cho mỗi tổ hợp các giá trị có thể của “cha mẹ”.

Có hai thành phần liên quan đến việc học một mạng Bayes: (i) học cấu trúc (*structure learning*), bao gồm việc xác định một DAG mô tả tốt nhất các mối quan hệ nhân quả giữa các nút trong mạng và (ii) học tham số (*parameter learning*), bao gồm việc tìm hiểu về các phân phối xác suất có điều kiện dựa vào dữ liệu của từng biến [10]. Định lý Bayes mô tả phân phối xác suất có điều kiện như sau:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Mạng Bayes sử dụng xác suất làm thước đo độ không chắc chắn của biến: Niềm tin về giá trị của các biến được biểu thị dưới dạng phân phối xác suất và độ không đảm bảo càng cao thì phân bố xác suất càng rộng. Khi thông tin tích lũy, kiến thức về giá trị thực của biến thường tăng lên, tức là độ không chắc chắn của giá trị giảm đi và phân phối xác suất ngày càng thu hẹp [11].

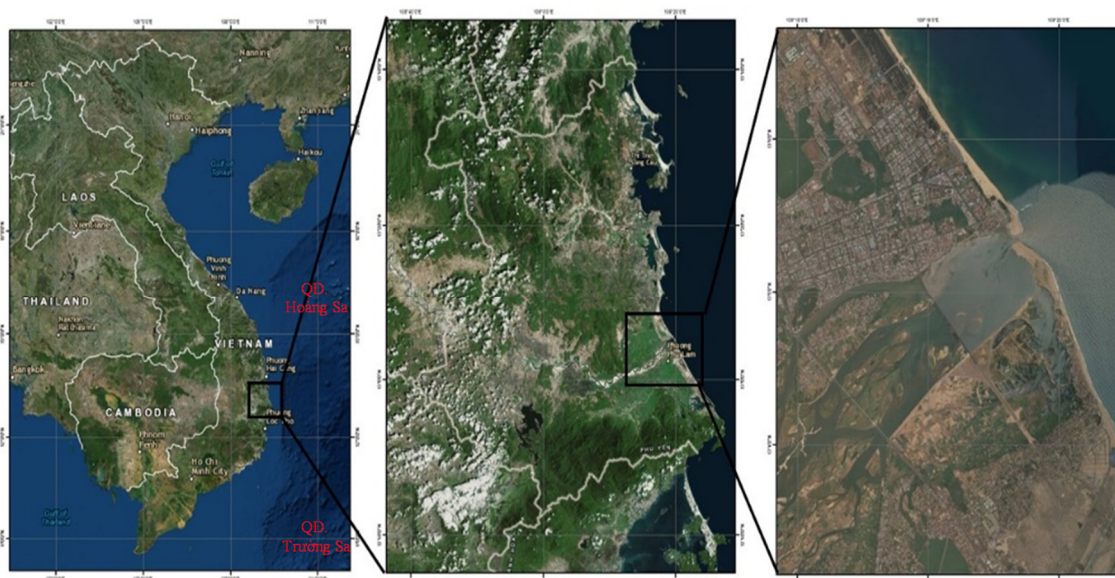
Trên thế giới, các nghiên cứu ứng dụng mạng Bayes trong lĩnh vực thủy văn và tài nguyên nước được phát triển với đa dạng các bài toán. Nhờ khả năng áp dụng linh hoạt phục vụ ra quyết định, mô hình mạng Bayes được áp dụng phổ biến vào hỗ trợ quản lý tổng hợp tài nguyên nước [12–14]. Bên cạnh đó, các nghiên cứu áp dụng mạng Bayes cho thấy khả năng kết hợp giữa các yếu tố thủy văn, thủy lực, kinh tế với các yếu tố xã hội thông qua các việc thu thập ý kiến chuyên gia [15–17]. Đây là một trong những ưu điểm của mô hình mạng Bayes khi có thể tận dụng và định lượng hóa ý kiến chuyên gia trong việc xây dựng mô hình. Các nghiên cứu ứng dụng mô hình mạng Bayes cho khu vực cửa sông cũng được công bố từ khá lâu, tuy nhiên chủ yếu tập trung nhiều ở lĩnh vực sinh thái vùng cửa sông và ven biển [18–20]. Hình thái cửa sông; các yếu tố thủy động lực và các hoạt động của con người được nghiên cứu như các yếu tố ảnh hưởng đến sinh thái khu vực cửa sông ven biển. Gần đây, các nghiên cứu ứng dụng mạng Bayes cho quản lý tổng hợp khu vực đường bờ biển dưới tác động phát triển kinh tế và biến đổi khí hậu đã được đẩy mạnh phát triển [21–23].

Sự phát triển của mạng Bayes đã và đang được nghiên cứu rộng rãi trên thế giới, tuy nhiên, đây vẫn còn là một phương pháp và cách tiếp cận mới mẻ tại Việt Nam. Một số nghiên cứu trong nước có hướng tiếp cận liên quan đến lý thuyết Bayes và mạng Bayes được phát triển trong lĩnh vực y tế và kinh tế, xã hội. Năm 2011, Giáo sư Nguyễn Văn Tuấn đã giới thiệu lý thuyết Bayes trong nghiên cứu y học lâm sàng [24]. Cũng áp dụng lý thuyết niềm tin Bayes, nhóm nghiên cứu của Nguyễn Ngọc Tuấn hướng tới đánh giá rủi ro trong lập lịch dự án phần mềm [25]. [26] đã kết hợp mô hình mạng Bayes với mô hình rủi ro trong kinh tế để ước lượng chi phí khám chữa bệnh ở Việt Nam. Gần đây, nhóm nghiên cứu [27] đã thực hiện đánh giá tác động của phát triển du lịch đến người dân tại đảo Lý Sơn. Nghiên cứu cho thấy một kết luận mới mẻ rằng người nông dân, những người trẻ tuổi, những người có trình độ học thức và có mức ảnh hưởng xã hội là những người có những đánh giá tiêu cực về tác động của sự phát triển du lịch.

Sông Ba, phần hạ lưu gọi là sông Đà Rằng, là hệ thống sông lớn nhất trong khu vực Nam Trung Bộ. Sông Ba có 3 phụ lưu chính là sông IaYun, sông Krông H’Năng và sông Hinh. Sông bắt đầu núi Ngọc Rô thuộc tỉnh Gia Lai, kéo dài 374 km và đổ ra biển tại cửa Đà Diễn, thành phố Tuy Hòa tỉnh Phú Yên. Tọa độ của cửa sông Đà Diễn là khoảng 13°5’23.65” vĩ độ Bắc, 109°19’40.79” kinh độ Đông (Hình 1) [28]. Vùng cửa sông nằm ở phía Nam thành phố Tuy Hòa, giáp với huyện Đông Hòa–tỉnh Phú Yên. Sông Ba có mùa lũ bắt đầu từ tháng 9 và kết thúc vào tháng 12, mùa kiệt bắt đầu từ tháng 1 đến tháng 8. Lưu lượng dòng chảy vào mùa lũ trên sông Ba chiếm 71,8% lưu lượng dòng chảy cả mùa. Tháng có lưu lượng lớn nhất

là tháng 11 với lưu lượng trung bình khoảng 862,3 m³/s, chiếm 25,8% tổng lượng dòng chảy cả năm [29].

Chế độ gió ở Phú Yên thể hiện hai mùa rõ rệt, từ tháng 10 đến tháng 4 là thời kỳ thịnh hành một trong ba hướng gió Bắc, Đông Bắc và Đông, từ tháng 5 đến tháng 9 là thời kỳ thịnh hành một trong ba hướng Tây, Tây Nam và Đông. Do ảnh hưởng của gió mùa, sóng tại khu vực nghiên cứu cũng có chế độ khác nhau rõ rệt giữa hai mùa. Từ tháng 1 đến tháng 4, hướng sóng thịnh hành là Đông Bắc. Độ cao trung bình của hướng sóng trên trong khoảng thời gian này là 1,0 m và cực đại là 4,0 m. Từ tháng 5 đến tháng 9, hướng sóng chủ đạo là Đông Nam với độ cao trung bình là 0,8 đến 1,0 m và lớn nhất là 3,5 m. Từ tháng 10 đến tháng 12, thịnh hành là hướng sóng Bắc và Đông Bắc với độ cao trung bình là 0,9 m và độ cao lớn nhất biến đổi từ 3,5 đến 4,0 m. Thủy triều tại khu vực này thuộc chế độ nhật triều không đều. Hàng tháng có từ 18 đến 22 ngày nhật triều. Thời kỳ triều cường thường xuất hiện nhật triều, khi triều kém thường xuất hiện bán nhật triều. Độ lớn triều trung bình là $1,50 \pm 0,20$ m [29].



Hình 1. Vị trí cửa sông Đà Diễn

Cửa sông Đà Diễn có những diễn biến phức tạp do chịu ảnh hưởng của chế độ thủy động lực phân mùa rõ rệt. Từ năm 2009, dưới tác động của các yếu tố ảnh hưởng khiến cửa sông bị thu hẹp đáng kể, chính quyền địa phương đã có các biện pháp tạm thời để giải quyết vấn đề này. Một trong số đó là việc nạo vét cát khu vực cửa sông. Tuy nhiên, việc nạo vét cát chưa cơ sở khoa học về khối lượng nạo vét, vị trí nạo vét và thời gian nạo vét đã khiến cửa sông có những biến động tiêu cực hơn sau trận lũ diễn ra vào tháng 11 năm 2017. Dải cát phía Nam cửa sông bị xói nghiêm trọng, chính quyền địa phương phải gấp rút làm kè đá để bảo vệ bờ, ngăn không cho hiện tượng xói ảnh hưởng vào sâu hơn (Hình 4). Đến năm 2018, từ kè đá bảo vệ bờ, chính quyền địa phương đã dần kéo dài kè đá để tái tạo lại dải cát phía Nam cửa sông.

Do có diễn biến phức tạp, cửa sông Đà Diễn là đối tượng nghiên cứu của nhiều nhóm nghiên cứu trong nước và quốc tế phục vụ công tác chinh trị và ổn định cửa sông, trong đó có thể kể đến các công trình: Rik Posthumus (2015) đã xây dựng mô hình nhận thức cơ bản về cơ chế vận động theo thời gian của cửa sông Đà Diễn [30]; [31] đã thực hiện đề xuất các cơ sở khoa học về đề xuất giải pháp ổn định cửa sông; hoặc nhóm nghiên cứu của Nguyễn Tiền Giang và các cộng sự [28, 32] đã hoàn thành đề tài nghiên cứu cấp nhà nước “Nghiên cứu cơ sở khoa học để xác định cơ chế bồi lấp, sạt lở và đề xuất các giải pháp ổn định các cửa sông Đà Diễn và Đà Nông tỉnh Phú Yên phục vụ phát triển bền vững cơ sở hạ tầng và

kinh tế xã hội” mã số ĐTĐL.CN.15/15... Các nghiên cứu này đều có sự liên kết, tham khảo, kế thừa và so sánh các kết quả nghiên cứu.

Mục tiêu của bài báo này là áp dụng thử nghiệm mô hình mạng Bayes trong phân tích các yếu tố tự nhiên ảnh hưởng đến sự biến động của cửa sông Đà Diễn trong giai đoạn từ năm 1988 đến 2018. Kết quả của mô hình được so sánh với các nghiên cứu trước đây nhằm khẳng định tính đúng đắn và khả năng áp dụng của mô hình trong bài toán này cũng như lĩnh vực khí tượng thủy văn nói chung.

2. Phương pháp và số liệu

2.1. Phương pháp nghiên cứu

Phần mềm mô hình Bayes Server được sử dụng để xây dựng mạng Bayes cho nghiên cứu này. Phần mềm Bayes Server được công ty chủ quản cùng tên (*Bayes Server*) có trụ sở tại Vương Quốc Anh phát triển. Công ty Bayes Server bắt đầu xây dựng các giải pháp Trí tuệ nhân tạo cho General Electric (GE) và Không quân Hoa Kỳ (USAF) và có hơn 15 năm kinh nghiệm cung cấp phần mềm Trí tuệ nhân tạo tùy chỉnh và có sẵn cho một số công ty tiên tiến nhất trên thế giới.

Phần mềm Bayes Server đã được phát triển để thử nghiệm cho cả mô hình mạng tĩnh (*Bayesian Networks-BNs*) và mô hình mạng động (*Dynamic Bayesian networks-DBNs*). Trong đó, nghiên cứu này tập trung thử nghiệm mô hình với mạng tĩnh và chuỗi số liệu rời rạc.

Trong mô hình này, mức độ tương quan hay độ liên kết giữa hai biến được thể hiện qua hệ số tương hỗ hay thông tin tương hỗ (*Mutual Information-MI*). Hệ số MI thể hiện là một phép đo định lượng về mức độ mà một biến ngẫu nhiên (A) cho chúng ta biết về hoặc dò tìm được một biến ngẫu nhiên khác (B). Như vậy, hệ số MI càng lớn thể hiện mức độ liên kết của hai biến càng cao. Hệ số MI được tính theo công thức như sau:

$$MI(A; B) = \sum_{b \in B} \sum_{a \in A} P(a, b) * \log \left(\frac{P(a, b)}{P(a) * P(b)} \right) \quad (2)$$

Trong đó $P(a)$ và $P(b)$ lần lượt là hàm phân phối xác suất biên của A và B; $P(a, b)$ là hàm phân phối xác suất chung của cả A và B.

Trong đó

$$P(a, b) = P(a).P(b) \quad (3)$$

Tương tự như trên, hệ số MI có điều kiện được tính như sau:

$$MI(A, B|C) = \sum_{c \in C} \sum_{b \in B} \sum_{a \in A} P(a, b, c) * \log \frac{P(c) * P(a, b, c)}{P(a, c) * P(b, c)} \quad (4)$$

Như vậy, hệ số MI có giá trị với cả hai chiều (từ A đến B hoặc từ B đến A). Giá trị lớn hơn của hệ số MI sẽ quyết định hướng của cạnh liên kết giữa hai nút A và B khi thực hiện học cấu trúc cho mạng Bayes.

2.2. Số liệu

Số liệu sóng, gió được thu thập từ dữ liệu của trung tâm dự báo hạn vừa châu Âu từ 1/1/1988 đến 31/12/2019. Trong đó, số liệu sóng bao gồm chiều cao sóng, chu kỳ sóng, hướng sóng; Số liệu gió tại độ cao 10m ($u10$, $v10$) đều được thu thập tại vị trí $13^{\circ}7'30''$ Bắc và $109^{\circ}22'30''$ Đông với độ phân giải là 32 km x 32km.

Số liệu thủy triều theo ngày tại trạm Hòn Chùa (tọa độ $13^{\circ}10'30''$ Bắc và $109^{\circ}18'36''$ Đông) từ 1/1/1988 đến 31/12/2016. Số liệu lưu lượng trung bình ngày tại trạm Củng Sơn (tọa độ $13^{\circ}02'35''$ Bắc và $108^{\circ}59'58''$ Đông) từ 1/1/1988 đến 31/12/2016. Số liệu độ rộng cửa

sông theo ngày từ năm 1988 đến năm 2018 được tính toán thông qua phân tích xử lý ảnh viễn thám Landsats với độ phân giải là 30m x 30m. Các số liệu kể trên được kế thừa từ nghiên cứu thuộc đề tài: “Nghiên cứu cơ sở khoa học để xác định cơ chế bồi lấp, sạt lở và đề xuất các giải pháp ổn định các cửa sông Đà Diễn và Đà Nông tỉnh Phú Yên phục vụ phát triển bền vững cơ sở hạ tầng và kinh tế xã hội” [28].

Cửa sông là nơi giao thoa giữa sông và biển do đó khu vực cửa sông là khu vực có diễn biến phức tạp và chịu ảnh hưởng bởi nhiều yếu tố, đặc biệt là các nhân tố tác động trực tiếp như các yếu tố sông, biển bao gồm sóng, gió và thủy triều [29]. Cũng theo nghiên cứu của Nguyễn Tiền Giang và các cộng sự, các yếu tố ảnh hưởng (các biến độc lập) cho phương pháp phân tích hồi quy được xác định bao gồm: Năng lượng gió; Hướng gió; Năng lượng sóng; Hướng sóng; Lãng trụ triều và Lưu lượng sông [32]. Tuy nhiên, có thể thấy, một số biến đã được thể hiện trong công thức tính toán cho các biến khác như hướng sóng và hướng gió lần lượt đều được sử dụng để tính toán năng lượng sóng và năng lượng gió; lãng trụ triều được tính thông qua biên độ triều; do đó, trong bài báo này, các biến trung gian đó được loại bỏ. Các yếu tố ảnh hưởng được thể hiện qua 6 biến đại diện như Bảng 1 và biến phụ thuộc (Yếu tố bị ảnh hưởng) được nghiên cứu là biến thiên độ rộng cửa sông (ký hiệu: deltaB). Trong đó, delta B được tính bởi công thức sau:

$$\text{deltaB} = B_{t+n} - B_t \quad (5)$$

Trong đó B_t là độ rộng cửa sông của ảnh tại ngày t ; B_{t+1} là độ rộng cửa sông tại ngày có ảnh kế tiếp $t + n$ và n được gọi là khoảng ảnh (khoảng thời gian giữa hai ảnh liên tiếp). Độ rộng cửa sông được đo tại một mặt cắt sao cho khoảng cách giữa hai dải cát bờ Nam và Bắc là hẹp nhất [28]. Do chỉ lựa chọn các ảnh vệ tinh Landsats đạt đủ điều kiện để thực hiện phân tích và tính toán độ rộng cửa sông nên khoảng ảnh n là không đồng đều [32].

Tương tự với đó, các dữ liệu lưu lượng sông, sóng và thủy triều đều được phân tích theo khoảng ảnh n . Theo đó:

- Lưu lượng ngày lớn nhất (Q_{\max}): là giá trị lưu lượng ngày lớn nhất trong khoảng ảnh n .
- Lưu lượng trung bình ngày (Q_{aver}): là giá trị lưu lượng trung bình ngày tính trong khoảng ảnh n .
- Biên độ triều (T): là giá trị biên độ triều trung bình tính trong khoảng ảnh n .
- WE và WiE: lần lượt là tổng năng lượng sóng và tổng năng lượng gió tính trong khoảng ảnh n .
- Giai đoạn (GD): là giai đoạn dài nhất trong khoảng ảnh n .

Như vậy, mỗi khoảng ảnh n sẽ tạo ra một giá trị của deltaB, GD, Q_{\max} , Q_{aver} , T , WE và WiE.

Bảng 1. Bảng các biến ảnh hưởng và phụ thuộc của mô hình.

TT	Ký hiệu	Tên biến	TT	Ký hiệu	Tên biến
1	Q_{\max}	Lưu lượng ngày lớn nhất	5	WE	Năng lượng sóng
2	Q_{aver}	Lưu lượng trung bình ngày	6	WiE	Năng lượng gió
3	T	Biên độ triều	7	deltaB	Biến thiên độ rộng cửa sông
4	GD	Giai đoạn			

Các dữ liệu này đều được chuẩn hóa về giá trị từ 0 đến 1. Sau khi được chuẩn hóa sẽ được đưa về các mức trạng thái như Hình 2.

- Các biến về năng lượng sóng, năng lượng gió được đưa về 3 mức trạng thái: “T”, “TB”, “C” thể hiện biến có giá trị so với các ngưỡng phân vị 25% và 75% của chuỗi số liệu nhiều năm.

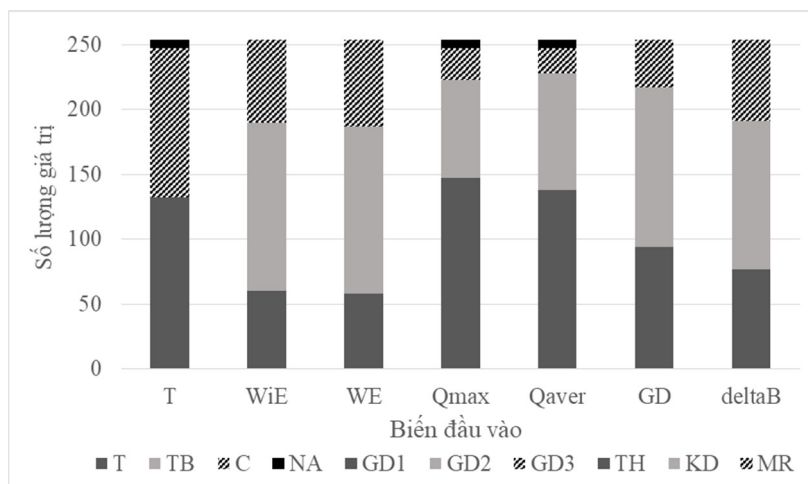
- Lưu lượng ngày lớn nhất được so sánh với cận dưới là giá trị trung bình lưu lượng ngày lớn nhất các tháng mùa kiệt nhiều năm và cận trên là giá trị trung bình lưu lượng ngày lớn nhất các tháng mùa lũ nhiều năm để chia thành ba mức trạng thái: thấp (“T”), trung bình (“TB”) và cao (“C”). Lưu lượng trung bình ngày cũng được chia làm ba mức trạng thái tương tự với cận dưới là giá trị lưu lượng ngày trung bình nhiều năm các tháng mùa kiệt và cận trên là giá trị lưu lượng ngày trung bình nhiều năm các tháng mùa lũ.

- Biên biên độ triều được chia thành 2 mức trạng thái là tương ứng với giá trị triều kém (“T”) và triều cường (“C”).

- Biểu thể hiện giai đoạn được đưa về ba trạng thái thể hiện ba giai đoạn: Giai đoạn 1 (“GD1”) từ tháng 1 đến tháng 4, Giai đoạn 2 (“GD2”) từ tháng 5 đến tháng 9 và Giai đoạn 3 (“GD3”) từ tháng 10 đến tháng 12 [28].

- Biến thiên độ rộng cửa sông (deltaB) cũng được chia thành 3 trạng thái cửa mở rộng ra (“MR”) khi biến thiên độ rộng lớn hơn 30m, cửa thu hẹp lại (“TH”) khi biến thiên độ rộng cửa nhỏ hơn -30 m và cửa không đổi (“KD”) khi biến thiên độ rộng cửa nằm trong khoảng từ -30 m đến 30 m. Ở đó, ± 30 m được coi là ngưỡng nằm trong độ sai số của ảnh Landsat với độ phân giải 30m x 30m.

- Trạng thái “NA” là trạng thái chung của các biến khi không có giá trị tại thời điểm đó.



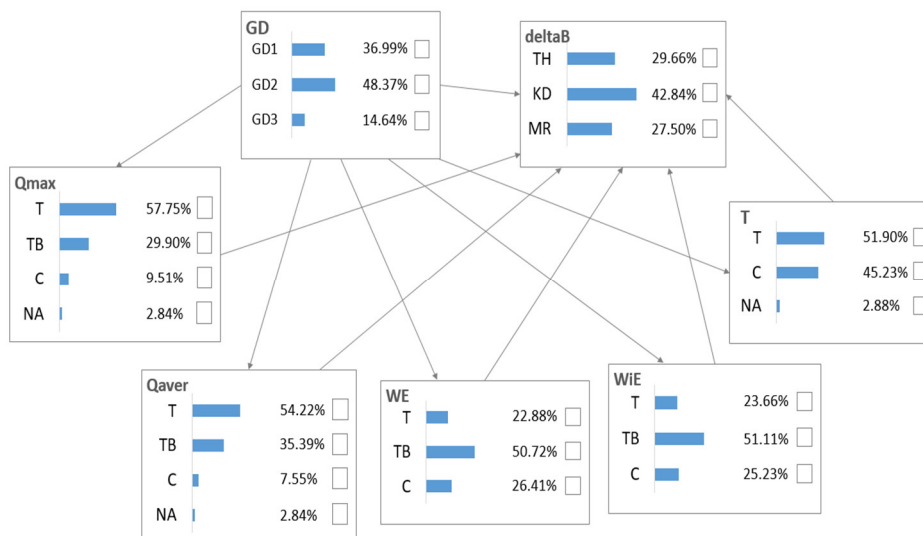
Hình 2. Mô tả các dữ liệu theo các mức trạng thái

Mỗi biến đều bao gồm 253 giá trị được chia thành nhiều nhất 4 trạng thái. Trong đó, có thể thấy, các biến đều có sự phân bố tương đối đều với từng mức trạng thái. Trạng thái “NA” không chiếm tỷ lệ quá lớn do đó ít có khả năng gây nhiễu cho mô hình. Nghiên cứu này tập trung vào chuỗi số liệu rời rạc và tiến hành rời rạc hóa các biến. Mặc dù, hầu hết các biến như lưu lượng sông, sóng, thủy triều đều là các biến liên tục và có hàm phân phối riêng nhưng các biến biến thiên độ rộng cửa sông (deltaB) và giai đoạn (GD) là các biến rời rạc do đó việc rời rạc hóa chuỗi số liệu trong nghiên cứu này là cần thiết để thống nhất được định dạng của tất cả các biến đầu vào.

3. Kết quả nghiên cứu

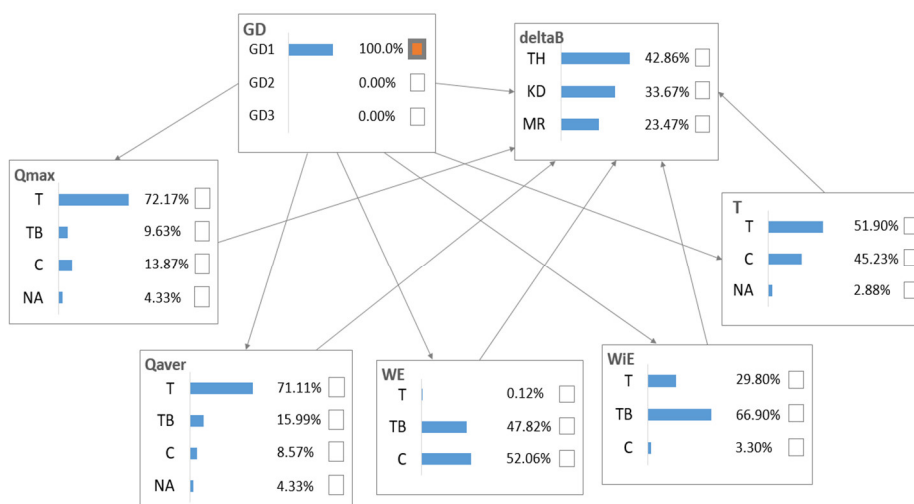
Các biến được đưa vào mạng Bayes trong mô hình Bayes Server và thực hiện tạo kết nối (links) giữa các biến ảnh hưởng: Qmax, Qaver, WE, WiE, T và GD đến biến phụ thuộc là deltaB. Ngoài ra, các biến Qmax, Qaver, WE, WiE và T cũng được coi là biến phụ thuộc của biến GD để xét sự thay đổi của các biến theo thời gian. Trong nghiên cứu này, cấu trúc mạng Bayes được định sẵn dựa trên các mối quan hệ giữa các yếu tố nên bước học cấu trúc mô hình được bỏ qua. Cho mạng học tham số dựa trên cấu trúc định sẵn, mạng Bayes cho bài toán được hình thành như Hình 3.

Ở mạng ban đầu, các biến thể hiện được xác suất xuất hiện của từng trạng thái. Ở đó, trạng thái “NA” được coi là một trạng thái của biến và được sử dụng như các trạng thái bình thường khác. Đây là một ưu điểm của mạng Bayes khi có thể tận dụng được tối đa số liệu sẵn có của nghiên cứu. Số liệu của các biến được kéo dài nhất có thể để tận dụng cho mô hình Bayes thay vì phải đưa toàn bộ số liệu về cùng một khoảng thời gian có số liệu như nhiều mô hình thủy văn trước đây.



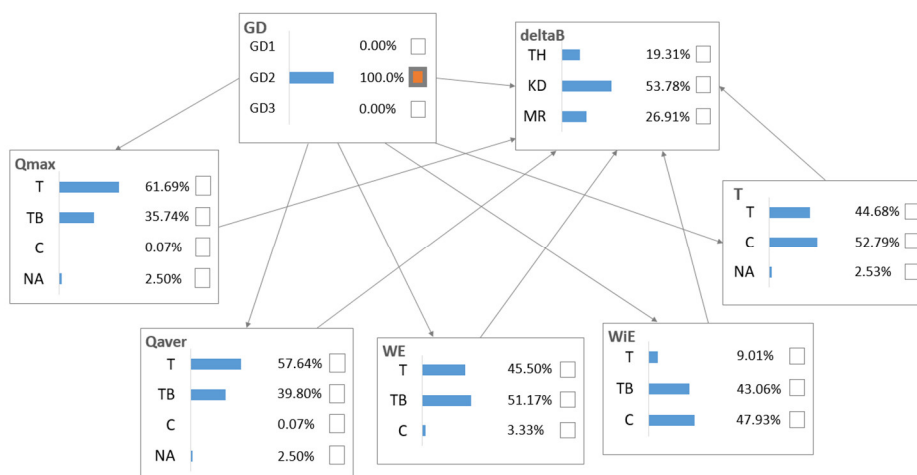
Hình 3. Kết quả ban đầu xây dựng mạng Bayes.

Thay đổi trạng thái của biến GD sẽ cho thấy các thay đổi theo thời gian của các biến. Ở giai đoạn 1 (từ tháng 1 đến tháng 4), độ biến thiên của sông thiên về trạng thái thu hẹp hoặc không đổi (xác suất hai trạng thái này chiếm hơn 70%). Cũng ở giai đoạn này, lưu lượng sông chủ yếu ở mức trạng thái thấp (“T”) do đây là giai đoạn kết thúc mùa lũ; triều kém cũng chiếm xác suất cao hơn. Trong khi đó, năng lượng sóng tập trung ở trạng thái trung bình và cao vì gió mùa Đông Bắc chiếm ưu thế trong giai đoạn này (Hình 4).



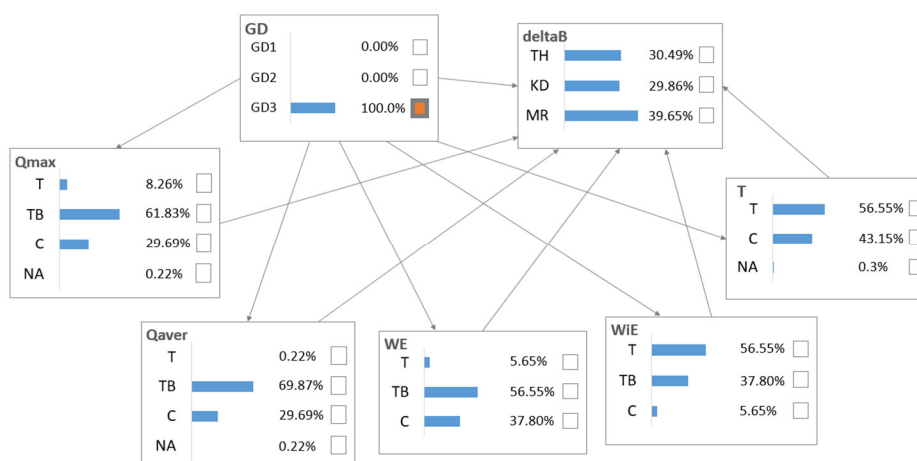
Hình 4. Kết quả mạng khi thay đổi trạng thái biến GD (GD1 = 100%).

Sang giai đoạn 2 (từ tháng 5 đến tháng 9), cửa sông cho thấy sự ổn định nhất với xác suất trạng thái không đổi là khoảng 53% (Hình 5). Ở giai đoạn này, yếu tố sông vẫn tập trung ở trạng thái thấp do đây là giai đoạn mùa kiệt. Năng lượng sóng cũng chuyển sang trạng thái thấp và trung bình khi gió mùa Tây Nam bắt đầu hoạt động mạnh. Tuy nhiên, năng lượng gió ở trạng thái cao và trung bình chiếm xác suất lớn.



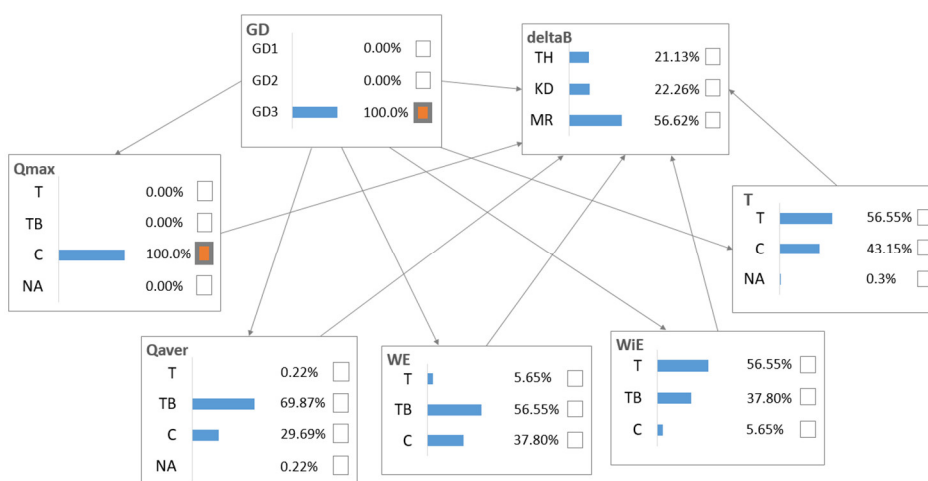
Hình 5. Kết quả mạng khi thay đổi trạng thái biến GD (GD2 = 100%).

Ở giai đoạn 3 (từ tháng 10 đến tháng 12), cửa sông không có xu thế biến động ổn định, xác suất xảy ra các trạng thái thu hẹp, mở rộng và không đổi gần như bằng nhau (khoảng 30%). Đây là thời kỳ mùa lũ nên lưu lượng sông tập trung ở mức trung bình và cao (Hình 6). Gió mùa Đông Bắc tác động mạnh khiến cho năng lượng sóng tăng lên trong giai đoạn này.



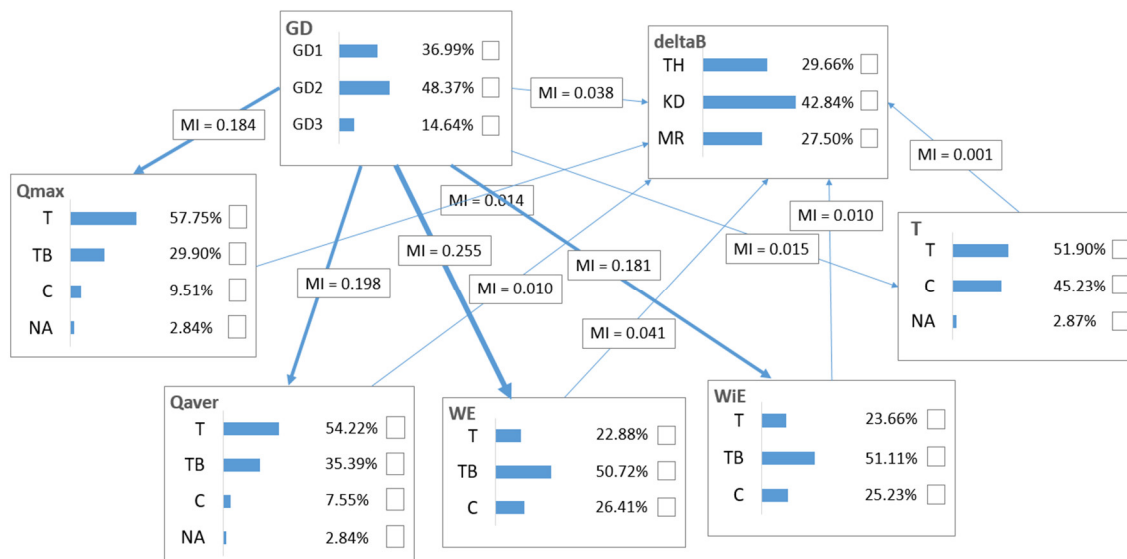
Hình 6. Kết quả mạng khi thay đổi trạng thái biến GD (GD3 = 100%).

Giai đoạn 3 là giai đoạn bất ổn nhất của cửa sông Đà Diễn, tuy nhiên, nếu lưu lượng sông đạt trạng thái cao thì xác suất cửa sông mở rộng là khoảng 60% (Hình 7).



Hình 7. Kết quả mạng khi thay đổi trạng thái biến GD và Qmax (GD3 = 100%, Qmax(C)=100%).

Tính mức độ ảnh hưởng của các biến trong mạng thông qua hệ số MI (Mutual Information), ta thấy rõ sự thay đổi của các biến rất phụ thuộc vào thời gian, do đó, việc phân chia giai đoạn để nghiên cứu cho khu vực cửa sông Đà Diễn là rất quan trọng. Các biến Qmax, Qaver, WE, WiE và T cũng cho thấy mức độ ảnh hưởng của chúng đến biến thiên độ rộng cửa sông (Bảng 2).



Hình 8. Kết quả mạng khi xét mức độ ảnh hưởng theo hệ số MI.

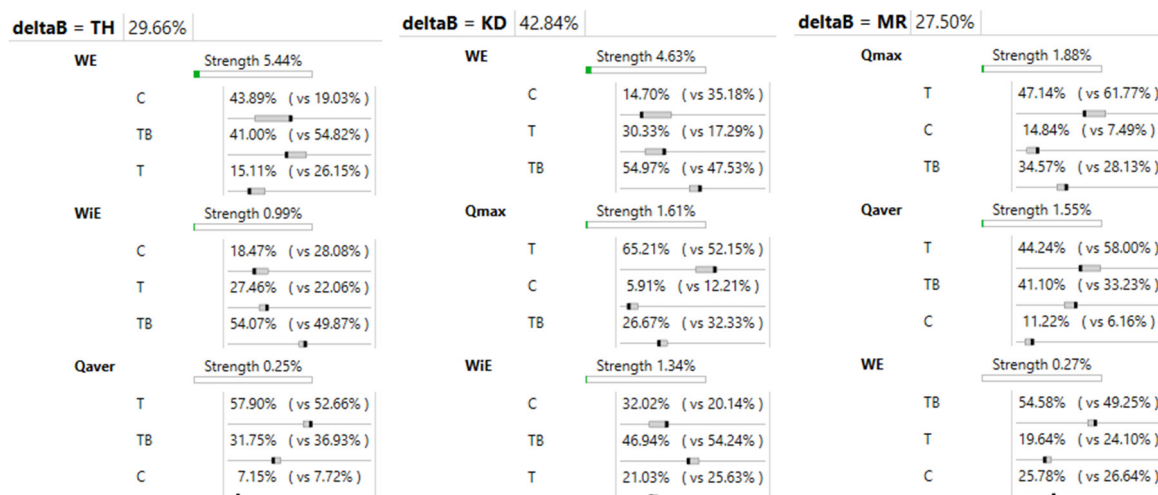
Xét trên toàn chuỗi số liệu, mức độ ảnh hưởng của các yếu tố gần như tương đương nhau trừ yếu tố thủy triều (T) có mức độ ảnh hưởng thấp nhất (Hình 8, Bảng 2). Xét riêng trong giai đoạn 1, hai yếu tố ảnh hưởng lớn nhất là năng lượng sóng và lưu lượng trung bình ngày. Sang đến giai đoạn 2, các yếu tố ảnh hưởng không thể hiện quá rõ, lớn nhất là sự ảnh hưởng của yếu tố lưu lượng ngày lớn nhất (Qmax).

Bảng 2. Bảng hệ số MI các yếu tố ảnh hưởng theo thời gian.

Giai đoạn	WE	WiE	Qaver	Qmax	T
Toàn bộ	0,015	0,01	0,01	0,014	0,001
GD1	0,023	0,00447	0,0163	0,00549	0,00756
GD2	0,00772	0,00777	0,0029	0,0171	0,00482
GD3	0,0501	0,024	0,0055	0,0271	0,0311

Ở giai đoạn 3, năng lượng sóng và thủy triều là hai yếu tố có mức ảnh hưởng lớn nhất. Bên cạnh đó, năng lượng gió và lưu lượng ngày lớn nhất cũng có đóng góp đáng kể trong giai đoạn này với mức độ ảnh hưởng tương đương nhau. Như vậy, có thể thấy, ở mùa gió Đông Bắc (giai đoạn 1 và 3), năng lượng sóng là yếu tố ảnh hưởng lớn đến biến thiên độ rộng cửa sông.

Xét riêng từng trạng thái của biến thiên độ rộng cửa sông, ta thấy năng lượng sóng là yếu tố ảnh hưởng quan trọng (Hình 9). Trạng thái thu hẹp cửa sông Đà Diễn phụ thuộc vào năng lượng sóng, năng lượng gió và lưu lượng trung bình. Cửa sông Đà Diễn ổn định (không đổi) phụ thuộc vào năng lượng sóng, lưu lượng ngày lớn nhất và năng lượng gió. Yếu tố chính ảnh hưởng đến sự mở rộng cửa sông là lưu lượng ngày lớn nhất, tiếp theo đó là lưu lượng trung bình ngày và năng lượng sóng.



Hình 9. Kết quả xét mức độ ảnh hưởng theo trạng thái biến phụ thuộc.

4. Kết luận

Nghiên cứu thử nghiệm mô hình mạng Bayes cho thấy, về mặt thời gian, cửa sông Đà Diễn có xu hướng diễn biến khác nhau theo từng giai đoạn. Giai đoạn 1, khi cửa sông chủ yếu ở trạng thái ổn định hoặc bị thu hẹp: đây là giai đoạn cửa sông chịu tác động của năng lượng sóng và lưu lượng trung bình ngày. Trạng thái cửa sông ổn định được thấy rõ hơn ở giai đoạn 2 và các yếu tố ảnh hưởng có tác động không rõ rệt nhất trong giai đoạn này. Ngược lại, giai đoạn 3 là giai đoạn cửa sông có trạng thái không theo xu hướng nhất định. Cả ba trạng thái thu hẹp, mở rộng và không đổi xuất hiện với xác suất ngang bằng nhau ở giai đoạn này. Yếu tố năng lượng sóng và thủy triều là hai yếu tố ảnh hưởng nhất ở giai đoạn 3. Kết quả về diễn biến và trạng thái của các giai đoạn này tương đồng với kết quả phân tích trong các nghiên cứu [28–29]. Các yếu tố ảnh hưởng được phân tích có phần khác nhau giữa nghiên cứu này với nghiên cứu [32] do sự lựa chọn biến có thay đổi ở nghiên cứu này. Tuy nhiên, cả hai nghiên cứu đều khẳng định năng lượng sóng là yếu tố được đánh giá là ảnh hưởng trong nhiều giai đoạn và có vai trò quan trọng trong diễn biến của cửa sông Đà Diễn.

Về mặt trạng thái, sự ảnh hưởng của các yếu tố khiến cửa sông có các trạng thái khác nhau, trong đó, năng lượng sóng là yếu tố có sự quyết định ở cả ba trạng thái. Lưu lượng ngày lớn nhất có sự ảnh hưởng đến trạng thái mở rộng cửa sông và cũng khiến cửa sông ổn định. Năng lượng gió lại có vai trò tác động đối với trạng thái cửa sông bị thu hẹp.

Mô hình mạng Bayes cho thấy các ưu điểm về sự tận dụng tối đa số liệu sẵn có trong nghiên cứu và cho thấy khả năng áp dụng linh hoạt đối với các dạng số liệu không đồng nhất khi có thể đưa vào mô hình chuỗi số liệu rời rạc với các trạng thái đa dạng hơn so với phương pháp hồi quy tuyến tính (chỉ nhận biến giá trị 0 hoặc 1). Nghiên cứu cho thấy sự phù hợp của mô hình mạng Bayes đối với nghiên cứu khu vực cửa sông nói riêng và lĩnh vực khí tượng thủy văn nói chung.

Trong nghiên cứu này, các yếu tố ảnh hưởng đến biến động cửa sông được xét đến đều là các yếu tố tự nhiên; tuy nhiên, có thể thấy, các yếu tố nhân sinh (các tác động của con người) thông qua hoạt động phát triển kinh tế, xây dựng công trình, hoạt động nạo vét, khai thác và sử dụng cát cũng là những yếu tố có thể tác động đến biến động của khu vực cửa sông. Do đó, ở các nghiên cứu tiếp theo cần xem xét thêm đến các yếu tố nhân sinh này.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: H.T.T.; Lựa chọn phương pháp nghiên cứu: H.T.T., T.N.A.; Xử lý số liệu: H.T.T.; Viết bản thảo bài báo: H.T.T., T.K.H.; Chỉnh sửa bài báo: T.N.A., H.T.T.

Lời cam đoan: Tập thể tác giả cam đoan bài báo này là công trình nghiên cứu của tập thể tác giả, chưa được công bố ở đâu, không được sao chép từ những nghiên cứu trước đây, không có sự tranh chấp lợi ích trong nhóm tác giả.

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội trong đề tài mã số TN.20.14. Bản quyền phần mềm Bayes Server sử dụng trong nghiên cứu được tài trợ bởi Trung tâm Động lực học Thủy Khí Môi trường. Nghiên cứu này cũng nhận được sự chia sẻ số liệu từ đề tài cấp Nhà nước, mã số ĐTĐL.CN.15/15. Tập thể nhóm tác giả xin chân thành cảm ơn các nguồn hỗ trợ quý báu này.

Tài liệu tham khảo

1. Xin, S.Y. Mathematical foundations, in *Introduction to Algorithms for Data Mining and Machine Learning*. Academic Press, 2019, pp. 19–43.
2. Andy, D.; Agogino, A.M. Text analysis for constructing design representations. *Artif. Intell. Eng.* **1997**, 11, 65–75.
3. Kahn Jr, C.E.; Roberts, L.M.; Shaffer, K.A.; Haddawy, P. Construction of a Bayesian network for mammographic diagnosis of breast cancer. *Comput. Biol. Med.* **1997**, 27, 19–29.
4. Paolo, G.; Taroni, F. Evaluation of scientific evidence using Bayesian networks. *Forensic Sci. Int.* **2002**, 125, 149–155.
5. Bruce, G.M.; Richard, S.H.; Martin, G.R.; Mary, M.R.; Micheal, J.W. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. *For. Ecol. Manage.* **2001**, 153, 29–42.
6. Borsuk, M.E.; Stow, C.A.; Reckhow, K.H. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecol. Modell.* **2004**, 173, 219–239.
7. Wooldridge, S.; Done, T. Learning to predict large-scale coral bleaching from past events: A Bayesian approach using remotely sensed data, in-situ data, and environmental proxies. *Coral Reefs* **2004**, 23, 96–108.
8. Bromley, J.; Jackson, N.A.; Clymer, O.J.; Giacomello, A.M.; Jensen, F.V. The use of Hugin® to develop Bayesian networks as an aid to integrated water resource planning. *Environ. Modell. Software* **2005**, 20, 231–242.
9. Pearl, J. Probabilistic Reasoning in Intelligent Systems. San Mateo, CA: Morgan Kaufmann, 1988.
10. Liu, S.; James McGree, J.; Ge, Z.; Xie, Y. Classification methods, in *Computational and Statistical Methods for Analysing Big Data with Applications*. Academic Press. 2016, pp. 7–28.
11. Sivia, D.S. Dealing with duff data. *Maximum Entropy and Bayesian Methods*, 1996, pp. 131–137.
12. Varis, O.; Rahaman, M.M.; Kajander, T. Fully connected Bayesian belief networks: a modeling procedure with a case study of the Ganges river basin. *Integr. Environ. Assess. Manage.* **2012**, 8, 491–502.
13. Thuc, D.P.; James, C.R.S.; Samantha, J.C.; Wade, L.H.; Oz Sahin. Applications of Bayesian belief networks in water resource management: A systematic review. *Environ. Modell. Software* **2016**, 85, 98–111.
14. Shi, H.; Lou, G.; Zheng, H.; Chen, C.; Bai, J.; Liu, T.; Ochege, F.U.; De Maeyer, P. Coupling the water-energy-food-ecology nexus into a Bayesian network for water resources analysis and management in the Syr Darya River basin. *J. Hydrol.* **2020**, 581, 124–387.
15. Cain, J. Planning improvements in natural resources management. UK: Centre for Ecology and Hydrology 2001, pp. 124.

16. Chan, T.U.; Hart, B.T.; Kennard, M.J.; Pusey, B.J.; Shenton, W.; Douglas, M.M.; Valentine, E.; Patel, S. Bayesian network models for environmental flow decision making in the Daly River, Northern Territory, Australia. *River Res. Appl.* **2012**, *28*, 283–301.
17. Varouchakis, E.A.; Palogos, I.; Karatzas, G.P. Application of Bayesian and cost benefit risk analysis in water resources management. *J. Hydrol.* **2016**, *534*, 390–396.
18. Tuấn, N.V. Giới thiệu phương pháp phân tích Bayes. *Tạp chí Thời sự y học* **2011**, *63*, 26–34.
19. McDonald, K.S.; Tighe, M.; Ryder, D.S. An ecological risk assessment for managing and predicting trophic shifts in estuarine ecosystems using a Bayesian network. *Environ. Modell. Software* **2016**, *85*, 202–216.
20. Madonna Vezzi; Colleen Downs; Victor Wepener and Gordon O'Brien. Application of the relative risk model for evaluation of ecological risk in selected river dominated estuaries in KwaZulu–Natal, South Africa. *Ocean and Coastal Management*, 2020, *185*, 105035.
21. Pelage, L.; Gonzalez, L.G.; Loc'h, F.L.; Ferreira, V.; Munaron, J.M.; Frédou, F.L.; Frédou, T. Importance of estuary morphology for ecological connectivity with their adjacent coast: A case study in Brazilian tropical estuaries. *Estuarine Coastal Shelf Sci.* **2021**, *251*, 107184.
22. Baldock, T.E.; Shabani, B.; Callaghan, D.P. Open access Bayesian Belief Networks for estimating the hydrodynamics and shoreline response behind fringing reefs subject to climate changes and reef degradation. *Environ. Modell. Software* **2019**, *119*, 327–340.
23. Furlan, E.; Slanzi, D.; Torresan, S.; Critto, A.; Marcomini, A. Multi-scenario analysis in the Adriatic Sea: A GIS-based Bayesian network to support maritime spatial planning. *Sci. Total Environ.* **2020**, *703*, 134972.
24. Sanuy, M.; Jiménez, J.A.; Plant, N. A Bayesian Network methodology for coastal hazard assessments on a regional scale: The BN-CRAF. *Coastal Eng.* **2020**, *157*, 103627.
25. Tuấn, N.N.; Hường, V.T.; Thắng, H.Q. Hướng tới mô hình mạng Bayes để đánh giá rủi ro trong lập lịch dự án phần mềm. Hội nghị Quốc gia lần thứ X về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR). Đà Nẵng, 2017.
26. Thắm, P.T.H. Áp dụng mô hình rủi ro nhóm và phương pháp Bayes trong ước lượng chi phí khám chữa bệnh ở Việt Nam. *Tạp chí Kinh tế và Phát triển* **2018**, *256*, 70–77.
27. Truong, Q.H.; Nguyen, A.T.; Trinh, Q.A.; Trinh, T.N.L.; Hens, L. Hierarchical Variance Analysis: A Quantitative Approach for Relevant Factor Exploration and Confirmation of Perceived Tourism Impacts. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2786.
28. Giang, N.T. Nghiên cứu cơ sở khoa học để xác định cơ chế bồi lấp, sạt lở và đề xuất các giải pháp ổn định các cửa sông Đà Diên và Đà Nông tỉnh Phú Yên phục vụ phát triển bền vững cơ sở hạ tầng và kinh tế xã hội. 2018: mã số: ĐTĐL.CN.15/15.
29. Thảo, H.T. Đánh giá ảnh hưởng của các yếu tố thủy động lực đến xu thế ổn định theo mùa vùng cửa sông Đà Diên. Đại học Khoa học Tự nhiên, 2019.
30. Posthumus, Rik, Conceptual model of the seasonal inlet closure in the Da Dien Estuary, in Civil Engineering. University of Twente 2015, pp. 49.
31. Hương, P.T. Nghiên cứu cơ sở khoa học cho việc đề xuất giải pháp ổn định cửa Đà Rằng, tỉnh Phú Yên. Đại học Thủy Lợi, 2013.
32. Pham, D.H.B.; Hoang, T.T.; Bui, Q.T.; Tran, N.A.; Nguyen, T.G. Application of Machine Learning Methods for the Prediction of River Mouth Morphological

Variation: A Comparative Analysis of the Da Dien Estuary, Vietnam. *J. Coastal Res.* **2019**, 35, 1024–1035.

Applying Bayesian Networks for assessing impact factors on Da Dien Estuary variation

Hoang Thu Thao^{1*}, Tran Ngoc Anh^{1,2}, Tran Kieu Huong³

¹ Faculty of Hydrology Meteorology Oceanography, University of Science, Vietnam National University, Hanoi; htthao@hus.edu.vn;

² Center for Environmental Fluid Dynamics, University of Science, Vietnam National University, Hanoi; tranngocanh@hus.edu.vn

³ Ministry of Education and Training; tkhuong@moet.gov.vn

Abstract: Bayesian Networks are probabilistic graphical model for presenting the causal relationship between variables, especially the uncertainty variables. There are many applications of Bayesian Network for several area as the medical, language, biology, environment and also hydrology. This paper applies the Bayesian Network for assessing factors affecting Da Dien Estuary (focus on the river mouth width evolution). The results show that wave energy is significant impact factor to change the estuary width, especially in period 1 and period 3). The Da Dien Estuary is quite stable from May to September (period 2) so it is difficult to detect the effect of hydrodynamic factors. In general, this study reinforces the flexible application possibilities of Bayesian network model in the field of hydrometeorology.

Keywords: Bayesian Network; Da Dien Estuary; Conditional Probability.