

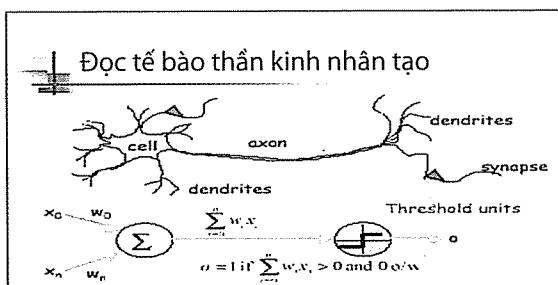
NGHIÊN CỨU ỨNG DỤNG MẠNG MÔ HÌNH THẦN KINH NHÂN TẠO DỰ BÁO DÒNG CHẢY THÁNG LƯU VỰC GHỀNH GÀ TRÊN SÔNG LÔ

TS. Lê Văn Đức - Trường Đại học Bách khoa, ĐHQG - TP. HCM

Bài báo giới thiệu kết quả nghiên cứu xây dựng phần mềm "NeuroForecaster"(NF), áp dụng mạng thần kinh nhân tạo lan truyền ngược (BPNN), và chương trình "ThuyVan Data System" (TVDATA) để xử lý và cung cấp nhập liệu, nhằm phục vụ việc mô hình hóa, mô phỏng và dự báo khí tượng thủy văn (KTTV) cho lưu vực sông. Hệ thống tích hợp NF-TVDATA đã được áp dụng thử nghiệm để mô phỏng dòng chảy tháng ở trạm thủy văn Ghềnh Gà thuộc lưu vực sông Lô, Việt Nam. Kết quả áp dụng đã chỉ ra sự phù hợp cao giữa dòng chảy mô phỏng tháng và số liệu thực đo. Bên cạnh đó, mối quan hệ giữa cấu trúc mô hình ANN với các biến KTTV trong lưu vực sông cũng đã được tìm thấy. Ngoài ra, hệ thống tích hợp này tỏ ra thân thiện, dễ thao tác, hỗ trợ tốt cho việc mô phỏng và dự báo KTTV lưu vực sông. Do đó, nó sẽ phục vụ tốt cho việc điều hành hồ chứa nhằm khắc phục hậu quả thiên tai, lũ lụt do ảnh hưởng của biến đổi khí hậu gây ra.

1. Giới thiệu mô hình ANN

Mô hình mô phỏng mạng lưới thần kinh nhân tạo là một mô hình toán học được sáng tạo dựa trên cơ sở những kiến thức về hoạt động tâm, sinh lý của não bộ con người (Hình 1). Một cách tổng quát, nó có thể được áp dụng đối với bất kỳ loại bài toán nào nhằm thiết lập mối quan hệ hàm ẩn giữa các biến nhập và biến xuất [1]. Nhờ những thành tựu nghiên cứu lý thuyết, khả năng tính toán và những tiện lợi trong thực hành, mô hình ANN đã được áp dụng rộng rãi trong nhiều lĩnh vực kỹ thuật khác nhau, đặc biệt trong lĩnh vực dự báo KTTV và quản lý nguồn nước [2, 3, 4].



Hình 1. Mạng Neuron thần kinh thực tế và mô hình [3]

Tuy nhiên, cho đến nay, các mô hình ANN đều gặp phải khó khăn khi xác định cấu trúc mạng phù hợp với từng loại bài toán khác nhau. Đó là, việc xác định số lượng nút nhập (biến nhập), số lượng tầng ẩn và số nút ẩn như thế nào cho hợp lý. Đây là quá trình thử nghiệm, kiểm tra, so sánh và chọn lựa cấu trúc ANN phù hợp, tốn kém thời gian và công sức.

Người đọc phản biện: ThS. Võ Văn Hòa

Ngoài ra, việc sử dụng số lượng mẫu huấn luyện và số lượng mẫu kiểm tra cần được chọn lựa thỏa đáng: vừa đủ dài để bao phủ miền trị của biến, đồng thời phải thích nghi với sự biến đổi của quy luật quan hệ. Chẳng hạn, trong trường hợp các yếu tố thủy văn, quy luật sẽ có thể biến đổi theo thời gian, vì sự phá rừng làm ảnh hưởng đến mặt đệm thực vật, hoặc vì sự thay đổi chế độ điều tiết hồ, hay do xây dựng hồ mới, hoặc vì ảnh hưởng của hiện tượng biến đổi khí hậu.

Để khắc phục những khó khăn, tổn phí thời gian và công sức khi sử dụng mô hình ANN, bài báo này trình bày kết quả xây dựng hệ thống NF-TVDATA nhằm hỗ trợ cho người sử dụng trong việc mô phỏng và dự báo KTTV lưu vực sông.

a. Sự hình thành và phát triển ANN

Mô hình ANN đã phát triển qua ba giai đoạn chính: ANN thế hệ thứ nhất với cấu trúc một tầng (1940-1960); ANN thế hệ thứ hai nhiều tầng (1986) với giải thuật lan truyền ngược (BPNN); ANN thế hệ thứ 3, được gọi là "Spiking Neuron Network" (MAASS, 1997) [5] nhằm khắc phục tốc độ tính toán chậm của các thế hệ ANN trước đây so với mạng thần kinh thật. Từ đó, sự phát triển mạng ANN nhằm vào những vấn đề như: đánh giá khả năng tổng quát hóa và sự giới hạn của mạng; tích hợp với các công nghệ khác như giải thuật di truyền (Genetic Algorithm, GA), các hệ mờ (Fuzzy sets) để

NGHIÊN CỨU & TRAO ĐỔI

nâng cao hiệu quả ứng dụng.

Mô hình ANN có thể được phân loại dựa trên nhiều tiêu chí khác nhau, nếu dựa vào đặc tính lan truyền tín hiệu, mạng ANN có thể được phân loại như sau:

Lan truyền thuận (Feedforward): gồm một tầng nhập, một hay nhiều tầng ẩn, một tầng xuất. Luật học thường dùng là giải thuật lan truyền ngược. Phương pháp học có giám sát. Loại ANN này thường được dùng trong các vấn đề tính toán, tìm hàm ẩn đa biến phức tạp và bài toán phân loại mẫu (Pattern Classification).

Lan truyền hồi tiếp (Feedback) hay mạng Hopfield : gồm một ma trận tầng neuron (neuron layer matrix). Luật học thường dùng là mô phỏng luyện kim (Simulated Annealing, SA). Phương pháp học không có giám sát. Nó thường được dùng trong bài toán tối ưu và tổ hợp mẫu (Pattern Association).

Lan truyền thuận kết hợp hồi tiếp (Feedforward/Feedback type) hay Kohonen Feature Map: tầng neuron gồm: một tầng nhập, và một tầng "map layer". Luật học được dùng là "Tự tổ chức" (Selforganization). Phương pháp học không có giám sát. Thường dùng trong các bài toán phân loại mẫu, tối ưu, và mô phỏng.

Đối với bài toán mô phỏng và dự báo KTTV, vấn đề chủ yếu là tìm mối quan hệ hàm ẩn đa biến giữa các yếu tố KTTV, nên chỉ cần mô hình mạng thần kinh nhân tạo với giải thuật lan truyền ngược là thỏa đáng. Tuy nhiên, cần phải cấu tạo chương trình như thế nào cho phù hợp để xử lý bộ dữ liệu lớn, tổ chức phối hợp dữ liệu thuận lợi và linh hoạt để có thể tìm kiếm một cấu trúc ANN thích hợp và nhanh chóng trong quá trình luyện mạng và kiểm tra mạng.

b. Đặc tính mô hình BPNN

Mô hình BPNN gồm một tầng nhập chứa một số nút nhập (biến độc lập), một tầng xuất chứa một hoặc nhiều nút xuất (biến phụ thuộc), ở bài toán áp dụng vào lĩnh vực mô phỏng và dự báo KTTV, thường chỉ cần một nút xuất. Mạng có thể có một hay nhiều tầng ẩn, với nhiều nút ẩn trên mỗi tầng. Số lượng tầng ẩn và số nút ẩn trên mỗi tầng sẽ phụ

thuộc vào từng loại bài toán và sẽ được điều chỉnh thông qua quá trình huấn luyện và kiểm tra mạng. Thường đối với bài toán trong lĩnh vực KTTV, chỉ cần không quá 2 tầng ẩn, với số nút tối đa không quá 10 nút trên mỗi tầng là thỏa đáng. Mô hình BPNN thường sử dụng hàm truyền (activation function) có dạng Logistic (Sigmoid) miền trị nằm giữa (0 - 1) hoặc hàm Hyperbol, Tang-Hyperbol có miền trị từ -1 đến +1.

Khi tính toán cập nhật trọng số trong giai đoạn điều chỉnh sai số, phương pháp moment có thể được sử dụng:

$$w_m = w_{m-1} + c_m \quad (1)$$

$$c_m = \mu.c_{m-1} - (1 - \mu).\varepsilon.d_m ; 0 \leq \mu < 1 \quad (2)$$

Ở đây, w_{m-1} và w_m lần lượt là trọng số ở bước tính m-1 và m; c_{m-1} và c_m lần lượt là biến thiên trọng số ở bước tính m-1, và m; d_m là đạo hàm riêng của hàm tổng sai số theo trọng số ở bước thứ m, ε gọi là tốc độ học có liên quan đến tốc độ hội tụ của quá trình học; u hệ số moment xét đến ảnh hưởng kết quả lần học trước, có liên quan đến tính ổn định của lời giải. Hoặc ta có thể dùng phương pháp thích nghi (Delta-Bar-Delta), như sau:

$$f_{m+1} = \theta.f_m + (1 - \theta).d_m \quad (3)$$

Với tốc độ học ε_m thay đổi theo từng bước tính, như sau:

$$\varepsilon_m = \begin{cases} \varepsilon_{m-1} + K & , d_m \cdot f_m > 0 \\ \varepsilon_{m-1} * \phi & , d_m \cdot f_m < 0 ; 0 < \phi < 1 \\ \varepsilon_{m-1} & , otherwise \end{cases} \quad (4)$$

Sau khi đạt được giá trị của ε_m thay vào (2), rồi (1), sẽ tính được trọng số ở bước tính m. Theo kinh nghiệm, để đảm bảo lời giải hội tụ và tốc độ huấn luyện nhanh, nên chọn các giá trị tham số như sau:

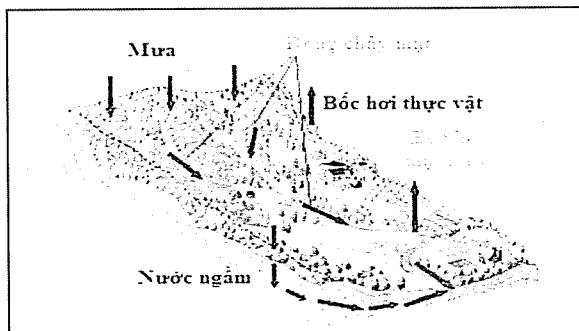
$$K = 0,1; \phi = 0,5; \mu = 0,9; \theta = 0,7$$

Tuy nhiên, giá trị thích hợp có thể tìm thấy trong quá trình huấn luyện và kiểm tra mạng.

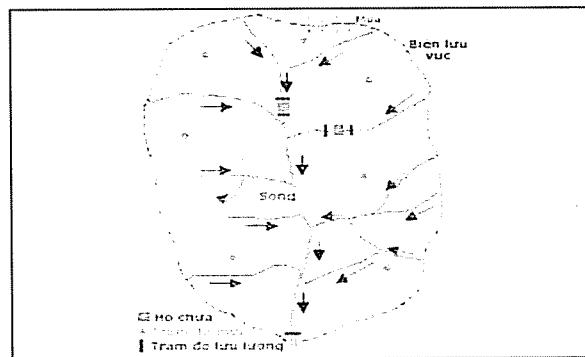
2. Mô hình lưu vực sông và ANN

Mưa trên lưu vực sông một phần bị bốc hơi thông qua thảm thực vật, bề mặt nước, một phần thẩm xuống đất tạo dòng nước ngầm, phần còn lại

chảy tràn mặt, rồi tập trung thành suối, sông, chảy vào hồ chứa tự nhiên hoặc nhân tạo, và ra biển. Dưới tác dụng điều tiết của hồ chứa, nước đổ xuống hạ lưu theo quy luật điều hành do con người chi phối. Một phần nước ngầm có thể thấm ra sông, suối (Hình 2).



Hình 2. Lưu vực sông



Hình 3. Mô hình lưu vực sông và ANN

Giả sử trong lưu vực có các trạm đo mưa, bốc hơi, nhiệt độ, các trạm thủy văn đo lưu lượng, mực nước (Hình 3). Gọi t là biến thời gian sử dụng trong mô hình (giờ, ngày, tuần, tháng, năm). Các biến KTTV sau cần được quan tâm:

Cường độ mưa đo được ở trạm i , ký hiệu $R_i(t)$, tác dụng trên diện tích A_i .

Cường độ bốc hơi đo được ở trạm j , ký hiệu là $EV_j(t)$, tác dụng trên diện tích A_j .

Lưu lượng ở trạm thủy văn k , ký hiệu là $Q_k(t)$, được tạo ra do diện tích thu nước A_k ở thượng lưu trạm, có thể bị ảnh hưởng bởi luật điều hành hồ chứa.

Cao trình mực nước ở trạm thủy văn n , ký hiệu $Z_n(t)$ được tạo ra do diện tích thu nước A_n ở thượng lưu trạm, mực nước ở trạm hạ lưu, và có thể bị ảnh

hưởng bởi luật điều hành hồ chứa.

Cường độ thấm $I_r(t)$ trên diện tích A_r

Lưu lượng thấm từ tầng nước ngầm vào suối hoặc sông, $QNN(t)$.

Phân tích bản chất vật lý của các yếu tố nêu trên, các mối quan hệ sau đây có thể tồn tại:

S Cường độ mưa tại một trạm có thể quan hệ với cường độ mưa của các trạm lân cận:

$$R_{\text{dich}}(t) = f(R_1(t), \dots, R_n(t)) \quad (5)$$

Lưu lượng ở một mặt cắt sông (Q_{dich}) phụ thuộc vào các yếu tố sau:

Lưu lượng ở cùng mặt cắt ở thời điểm trước để xét đến ảnh hưởng quán tính hình thành dòng chảy; ảnh hưởng của lưu lượng thấm từ tầng nước ngầm ra sông suối.

Lưu lượng ở mặt cắt thượng lưu của trạm đích ở cùng thời điểm (t), nếu khoảng cách gần so với trạm đích và bước thời gian tính lớn; hoặc ở các thời điểm trước: $(t-1); (t-2)$ tùy thuộc khoảng cách đến trạm đích gần hay xa và bước tính thời gian dài hay ngắn. Biến lưu lượng này thường có quan hệ rất chặt với lưu lượng ở trạm đích. Nếu đưa vào trong mô hình ANN, sẽ làm gia tăng độ chính xác kết quả mô phỏng một cách đáng kể.

Cường độ mưa ở các trạm trong lưu vực, thượng lưu trạm đích là nguyên nhân chủ yếu phát sinh ra dòng chảy mặt và hình thành lưu lượng của trạm đích.

Cường độ thấm, lưu lượng của nước ngầm chảy vào sông hay suối.

Bốc hơi thực vật, bốc hơi bề mặt, nhiệt độ, độ ẩm không khí, các yếu tố này có quan hệ lẫn nhau. Tuy nhiên, sự biến đổi tương đối của chúng trong khoảng thời gian mô phỏng, không thay đổi nhiều so với sự biến đổi của dòng chảy mặt, trừ những năm có thời tiết đặc biệt. Sự sai lệch này có thể xem như một yếu tố nhiễu, nếu thông tin của chúng không được đưa vào mô hình và chính là một trong những nguyên nhân gây ra sai số mô hình.

Diện tích, đặc điểm và tính chất của lớp phủ thực vật, bao gồm rừng, thảm cỏ, bụi rậm, cây công

và nông nghiệp trong lưu vực. Nếu như yếu tố này không thay đổi lớn theo mùa, thì tính chất hình thành dòng chảy không có biến động lớn trong thời gian mô phỏng, trừ những năm có chế độ KTTV đặc biệt. Tuy nhiên, trong trường hợp có biến động lớn như phá rừng trên diện tích rộng để xây đập, hồ chứa, hoặc làm rẫy, trồng trọt, chăn nuôi, thì đặc tính hình thành dòng chảy sẽ thay đổi rất đáng kể. Khi đó cấu trúc của mô hình ANN phải được cập nhật, dùng dữ liệu KTTV sau thời điểm bắt đầu có các yếu tố này tác động, để mô hình ANN có thể nắm bắt lại đặc tính mới của lưu vực sông.

Sự thay đổi cấu trúc dòng chảy do ảnh hưởng của con người như làm mới kênh mương để chuyển nước; xây dựng mới các trạm bơm để cấp nước cho các khu công, nông nghiệp. Chặn dòng chảy như xây đập, hồ chứa,... Khi các yếu tố này tác động thì mô hình ANN phải được huấn luyện và kiểm tra lại để cập nhập thông tin của các yếu tố biến động mới của lưu vực.

Tóm lại, trừ những yếu tố ảnh hưởng lớn đến sự hình thành dòng chảy của lưu vực, hàm mô tả lưu lượng đích có thể được diễn tả như sau:

$$Q_{\text{đích}}(t) = f(Q_{\text{đích}}(t-1); Q_1(t), Q_2(t-1), \dots, Z_1(t), Z_2(t-1), \dots, R_1(t), R_2(t-1), \dots) \quad (6)$$

Khi thượng lưu của trạm mô phỏng có dòng chảy do điều hành hồ chứa, thì nên đưa lưu lượng điều hành, hoặc lưu lượng tại một mặt cắt sau hồ chứa, ở thượng lưu của trạm đích làm biến nhập. Mô hình ANN cũng không khác các mô hình toán thủy lực khác ở điểm là khi có sự biến đổi dòng chảy, hoặc thay đổi điều kiện biên, thì những thông tin này cần phải được đưa vào trong mô hình ANN, huấn luyện và kiểm tra lại mạng, để cấu trúc mới của nó có thể nắm bắt các quy luật do các yếu tố biến động mới gây ra.

S Mực nước tại một mặt cắt sông $[Z_{\text{đích}}(t)]$, có thể quan hệ với lưu lượng $[Q_{\text{đích}}(t)]$ và mực nước ở thời điểm trước, $[Z_{\text{đích}}(t-1)]$ tại trạm đích, và ở các trạm lân cận thượng hoặc hạ lưu sông. Khi các yếu tố lưu lượng và mực nước ở các mặt cắt sông được đưa vào mô hình, và trên đoạn sông không có sự thay đổi lớn về cấu trúc hình thành dòng chảy, thường

kết quả sẽ đạt được độ chính xác khá cao không kém gì các chương trình mô phỏng thủy lực. Điểm giống nhau của hai mô hình này là cần được cung cấp điều kiện biên như nhau. Điểm khác biệt ở đây là mô hình thủy lực cần đưa các thông tin về hình dạng mặt cắt sông, hệ số nhám lòng dẫn, độ dốc và độ dài các đoạn sông trước; còn mô hình ANN thì phải đưa thông tin về dòng chảy đã chảy qua đoạn kênh như lưu lượng, mực nước ở dọc các mặt cắt kênh trong suốt quá trình điều chỉnh mô hình toán thủy lực. Như vậy, đối với mô hình thủy lực thì đặc tính vật lý, thủy lực của kênh được cung cấp từ đâu; còn đối với mô hình ANN, nó sẽ phát hiện và nhận biết sau quá trình huấn luyện, kiểm tra và sẽ lưu trữ trong bộ ma trận trọng số (W_{ij}) ở tầng ẩn, thường không được biết bởi người dùng.

Từ các nhận xét trên, ta có định hướng xây dựng hệ thống NeuroForecaster, gồm một chương trình ANN (NF), khai thác, trình bày kết quả và một chương trình TVDATA cung cấp giao diện cho việc nhập liệu, chọn lọc các tham số cho mô hình, chẳng hạn, chọn trạm đích, chọn các trạm nguồn, có thể là trạm lưu lượng, mực nước hoặc mưa,...chọn thời gian và thời đoạn để huấn luyện và kiểm tra. Chương trình này sẽ cung cấp các tập tin nhập cho chương trình NF.

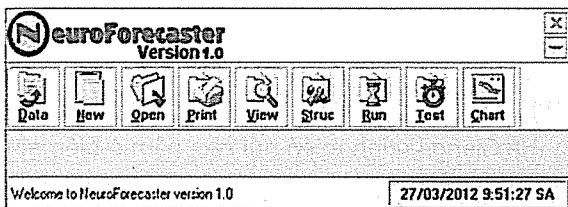
3. Phần mềm NEUROFORECASTER

a. Giới thiệu chung

Hệ thống phần mềm NeuroForecaster thực hiện chức năng mô phỏng và dự báo KTTV lưu vực sông gồm: chương trình tính toán ANN, NF và chương trình TVDATA dùng để thực hiện chức năng nhập liệu, cung cấp cho chương trình NF. Cả hai chương trình này đều được viết bằng ngôn ngữ Borland Delphi chạy trên nền hệ điều hành Windows XP/2000, Windows 7.

b. Chương trình NF

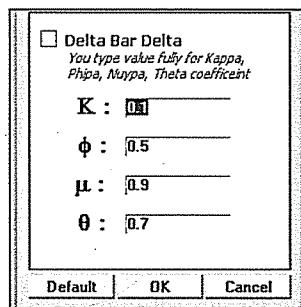
Giao diện chương trình NF được chỉ ra trong hình 4, gồm các chức năng chính sau:



Hình 4. Giao diện phần mềm NeuroForecaster

Chức năng tạo mới, thao tác và tổng kết dữ liệu: cung cấp dữ liệu đầu vào cho mạng ANN, cho phép người sử dụng dễ dàng lựa chọn bộ dữ liệu theo mong muốn, để thực hiện việc huấn luyện và kiểm tra mạng. Ngoài ra, nó còn cho phép thống kê, tạo báo cáo thống kê cho phép người sử dụng chọn phương án cuối cùng, được hỗ trợ bởi mô hình tối ưu đa mục tiêu AHP (Analytic Hierarchy Process).

Chức năng huấn luyện mạng: cho phép tiến hành huấn luyện mạng để tạo tập thông số mạng ("Tentaptin.WTS"), tạo tập tin sai số ("Tentaptin.ERR") và tập tin kết xuất sau khi huấn luyện ("Tentaptin.OUT"), cho phép người dùng chọn sai số tính và các tham số K, μ và θ (Hình 5) nếu dùng phương pháp huấn luyện thích nghi.



Hình 5. Nhập tham số huấn luyện mạng

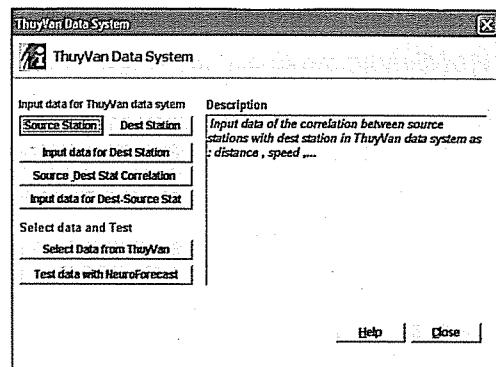
Chức năng kiểm tra mạng: cho phép tiến hành kiểm tra mạng, dựa trên bộ trọng số, và thông số mạng ("Tentaptin.NET") đạt được từ giai đoạn huấn luyện, và dữ liệu kiểm tra từ tập tin nhập (Tentaptin.TST), sẽ tạo ra tập tin kết xuất kiểm tra ("Tentaptin.OTS").

Chức năng vẽ biểu đồ: cho phép vẽ biểu đồ các đường cong sai số, đường kết xuất giai đoạn huấn luyện và kiểm tra. Biểu đồ so sánh số liệu tính toán so với giá trị thực đo trong giai đoạn huấn luyện được chỉ ra trong hình 11 và giai đoạn kiểm tra trong hình 12.

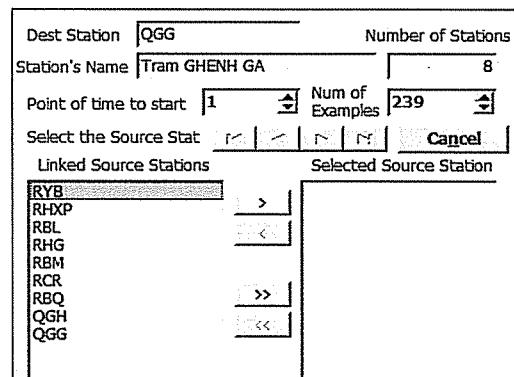
c. Chương trình TVDATA

Giao diện chương trình TVDATA được chỉ ra trong hình 6. TVDATA là một chương trình quản lý dữ liệu, cho phép người sử dụng nhập dữ liệu vào

các bảng trong tập tin cơ sở dữ liệu MS. ACCESS và cung cấp công cụ tạo lập mối quan hệ hàm giữa biến thủy văn ở "trạm dữ liệu đích" và các "trạm dữ liệu nguồn". Hình 7 chỉ ra sự liên kết đến các trạm KTTV trong cơ sở dữ liệu, và công cụ chọn lựa các trạm nguồn.



Hình 6. Giao diện chương trình TVDATA



Hình 7. Giao diện chọn lựa trạm nguồn

d. Chương trình tích hợp TVDATA và NF

Khi sử dụng tích hợp hai chương trình TVDATA và NF, tạo tập tin dữ liệu huấn luyện và kiểm tra từ chương trình TVDATA, chạy chương trình NF để huấn luyện và kiểm tra cho nhiều phương án khác nhau, hệ thống chương trình sẽ tạo ra các bộ dữ liệu chứa các tập tin với phần mở rộng lần lượt là: DTA, TST, Net, Lst, Out, Wts, Err, Ots và cho phép tạo lập một tập tin báo cáo tổng kết các phương án đã chạy, dựa theo các tiêu chí khác nhau, như:

- Số mẫu huấn luyện.
- Thời gian thực thi huấn luyện, thời gian thực thi kiểm tra rất nhanh nên có thể bỏ qua.
- Độ chính xác đạt được.
- Xếp hạng theo phương pháp phân tích tối ưu đa mục tiêu AHP.
- Nhờ bảng tổng hợp này, người dùng có thể dễ

NGHIÊN CỨU & TRAO ĐỔI

dùng so sánh và chọn lựa phương án tốt nhất.

Sau đây là phần trình bày áp dụng hệ thống NF để mô phỏng dòng chảy tháng ở trạm thủy văn Ghềnh Gà.

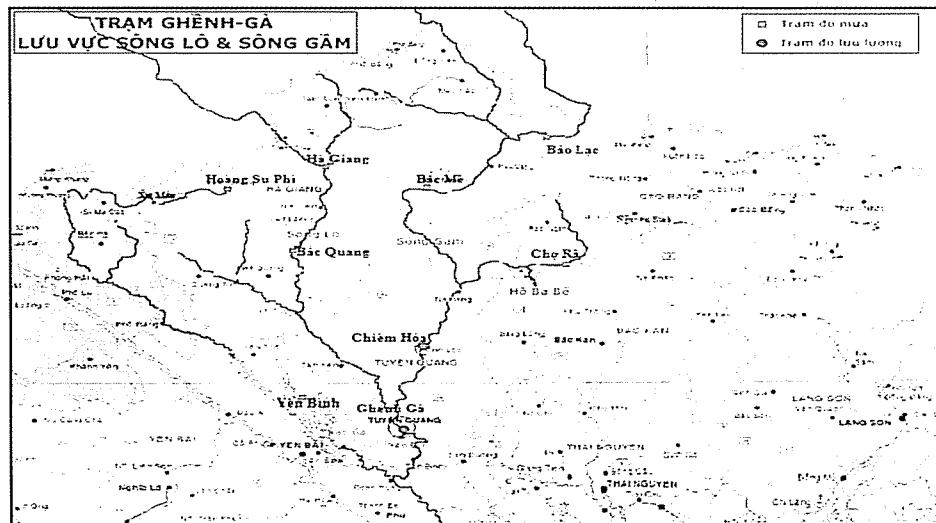
4. Ứng dụng hệ thống NF để mô phỏng dòng chảy tháng trạm Ghềnh Gà

a. Vị trí địa lý lưu vực Ghềnh Gà, sông Lô - Hà Giang

Trạm thủy văn Ghềnh Gà, thuộc sông Lô nằm ở hạ lưu nơi hai sông Gâm và Lô gặp nhau. Vì thế, nó tiếp nhận lưu lượng của cả hai sông này. Lưu vực

chính của trạm gồm các tỉnh Hà Giang (sông Lô), Tuyên Quang (sông Gâm), Cao Bằng, Bắc Cạn và một phần nằm trên lãnh thổ Trung Quốc.

Hà Giang là tỉnh miền núi cao, nằm ở biên giới thuộc vùng Đông Bắc Việt Nam, trong phạm vi tọa độ từ $22^{\circ}10' - 23^{\circ}13'N$; và từ $104^{\circ}24' - 105^{\circ}30'E$, phía bắc giáp tỉnh Vân Nam và Quảng Tây (Trung Quốc); phía nam giáp tỉnh Tuyên Quang; phía đông giáp Cao Bằng, phía tây giáp Yên Bái, Lào Cai. Diện tích tự nhiên toàn tỉnh là $7.884,37 \text{ km}^2$, chiếm khoảng 2,4% diện tích cả nước [6].



Hình 8. Bản đồ lưu vực trạm thủy văn Ghềnh Gà, sông Lô [7]

b. Đặc điểm khí hậu và thủy văn lưu vực Hà Giang

Nằm trong vùng nhiệt đới gió mùa và là miền núi cao, khí hậu Hà Giang về cơ bản mang những đặc điểm của vùng núi Việt Bắc – Hoàng Liên Sơn. Nhiệt độ trung bình cả năm khoảng $21,6^{\circ}C - 23,9^{\circ}C$. Chế độ mưa ở Hà Giang khá phong phú. Toàn tỉnh đạt bình quân lượng mưa hàng năm khoảng $2.300 - 2.400 \text{ mm}$, riêng Bắc Quang hơn 4.000 mm , là một trong số trung tâm mưa lớn nhất nước ta. Dao động lượng mưa giữa các vùng, các năm và các tháng trong năm khá lớn [6]. Các sông lớn có liên quan ở Hà Giang thuộc hệ thống sông Hồng, ở đây có mật độ sông - suối tương đối dày:

Sông Lô bắt nguồn từ Vân Nam (Trung Quốc), chảy qua Thanh Thuỷ, thị xã Hà Giang, Vị Xuyên, Bắc Quang về Tuyên Quang, Phú Thọ và gặp sông Hồng tại Việt Trì. Dòng chính sông Lô dài 470 km và diện tích lưu vực 39.000 km^2 , trong đó ở nước ta sông Lô

dài 275 km và diện tích lưu vực 22.600 km^2 (chiếm 60%). Trạm thủy văn Ghềnh Gà nằm trên sông Lô, tiếp nhận lưu lượng của cả hai sông Lô và sông Gâm, có diện tích lưu vực 29.600 km^2 [6, 7].

Sông Gâm bắt nguồn từ Nghiêm Sơn, Tây Trù (Trung Quốc) chảy qua Lũng Cú, Đồng Văn, Mèo Vạc, Yên Minh, Bắc Mê (Hà Giang), Na Hang, Chiêm Hóa (Tuyên Quang), về gần thị xã Tuyên Quang nhập vào sông Lô. Sông Gâm là phụ lưu cấp 1 lớn nhất của sông Lô. Phần diện tích lưu vực trên lãnh thổ Việt Nam là 9.649 km^2 , chiếm 63,23% tổng diện tích lưu vực sông Gâm [6, 7].

Sông ngòi ở Hà Giang có giá trị lớn về thủy điện, đồng thời là nguồn cung cấp nước quan trọng cho đời sống và sản xuất của tỉnh [6].

c. Mô phỏng lưu lượng tháng trạm Ghềnh Gà

Để ứng dụng thử nghiệm hệ thống NF, trạm thủy văn Ghềnh Gà, Hồng, Lô, được dùng làm trạm lưu

lượng đích (QGG), trạm lưu lượng Chiêm Hóa, Hồng Gấm (QGH) và bảy trạm đo mưa ở lưu vực gồm Hoàng Su Phì (RHXP), Bảo Lạc (RBL) (tỉnh Cao Bằng), Hà Giang (RHG), Bắc Mê - Yên Phú (RBM), Yên Bình (RYB) [tỉnh Yên Bái], Bắc Quang (RBQ) và Chợ Rã (RCR) (tỉnh Bắc Cạn) được dùng làm các trạm nguồn.

Phân bố dữ liệu mô phỏng:

Dữ liệu mưa tháng và lưu lượng tháng thực đo từ năm 1966 đến năm 1985, gồm 239 mẫu, được sử dụng cho giai đoạn huấn luyện để tìm quy luật hàm, hay cấu trúc của mạng ANN. Dữ liệu từ năm 1986 đến năm 1990, gồm 59 mẫu, với độ dài khoảng 25% so với số mẫu huấn luyện, được dùng làm dữ liệu kiểm tra độ chính xác của mô hình mô phỏng ANN (nguồn dữ liệu [4 & 3]).

Để huấn luyện tìm ra cấu trúc mạng ANN tốt nhất, phương pháp thay đổi tổ hợp nhập, thử nghiệm, kiểm tra và so sánh được sử dụng. Những yếu tố sau được thay đổi:

Thử lần lượt bỏ bớt các tổ hợp khác nhau của trạm lưu lượng nguồn gồm QGH, các trạm mưa RHXP, RBL, RHG, RBM, RYB, RBQ, RCR.

Bỏ không xét đến lưu lượng ở trạm Ghềnh Gà ở thời điểm trước [QGG(t-1)].

Dùng các tổ hợp dữ liệu ở tháng trước (t-1), (t-2),... của các trạm mưa trong lưu vực làm nhập liệu dựa trên hệ số hiệu chỉnh HesoHC của chương trình. Hệ số này dựa trên tỷ số khoảng cách dọc dòng chảy với vận tốc trung bình tính từ trạm nguồn đến trạm đích, chính là thời gian tập trung nước trung bình. Việc sử dụng luật "heuristic" này

hỗ trợ cho người dùng nhanh chóng tìm ra nhập liệu thích hợp. Và dù, nếu tại trạm nguồn RBL có HesoHC = 2 thì ta có thể lấy dữ liệu trong biểu thức quan hệ là:

$$QGG(t) = f[QGH(t); RBL(t); RBQ(t); RCR(t); RHG(t); RHXP(t); RYB(t); QGG(t-1)]$$

$$Hay: QGG(t) = f[QGH(t); RBL(t-1); RBQ(t); RCR(t); RHG(t); RHXP(t); RYB(t); QGG(t-1)].$$

$$Hay: QGG(t) = f[QGH(t); RBL(t-1), RBL(t); RBQ(t); RCR(t); RHG(t); RHXP(t); RYB(t); QGG(t-1)].$$

Thay đổi số lượng tầng ẩn, cũng như số nút trên mỗi tầng.

Thay đổi phương pháp học gồm: phương pháp Delta; thích nghi (Delta-Bar-Delta) và phương pháp học nhanh (Quick Propagation, QPROP) phối hợp nhau để tìm các lời giải khả dĩ tốt dựa trên các tiêu chí: thời gian huấn luyện nhanh, độ chính xác giai đoạn huấn luyện và kiểm tra cao; số lượng nút nhập (các trạm nguồn), số tầng ẩn và số lượng nút ẩn tối thiểu.

Sau quá trình kiểm tra, thử nghiệm, cuối cùng 8 phương án khả dĩ tốt, còn lại dùng để mô phỏng (Hình 9).

Kết quả thực hiện tám phương án này dựa trên hệ thống TVDATA và NF được lưu trữ trong tập tin MyReport.RPT và kết quả tính toán được chỉ ra trong Hình 10a và 10b:

Correlation						
$QGG(t) = f[RBL(t); RBM(t); RBQ(t); RCR(t); RHG(t); RHXP(t); RYB(t); QGG(t-1)]$						
$QGG(t) = f[QGH(t); RBL(t); RBQ(t); RCR(t); RHG(t); RHXP(t); RYB(t); QGG(t-1)]$						
$QGG(t) = f[QGH(t); RBL(t); RBQ(t); RCR(t); RHG(t); RHXP(t); RYB(t); QGG(t-1)]$						
$QGG(t) = f[QGH(t); RBL(t); RBM(t); RBQ(t); RCR(t); RHG(t); RHXP(t); RYB(t); QGG(t-1)]$						
$QGG(t) = f[RBL(t); RBM(t); RBQ(t); RCR(t); RHG(t); RHXP(t); RYB(t); QGG(t-1)]$						
$QGG(t) = f[QGH(t); RBM(t); RBQ(t); RCR(t); RHG(t); RHXP(t); RYB(t); QGG(t-1)]$						
$QGG(t) = f[QGH(t); RBL(t); RBQ(t); RCR(t); RHG(t); RHXP(t); RYB(t); QGG(t-1)]$						
$QGG(t) = f[QGH(t); RBL(t); RBM(t); RBQ(t); RCR(t); RHG(t); RHXP(t); RYB(t); QGG(t-1)]$						

Hình 9. Tám phương án chọn lọc để thực hiện và so sánh

No Examples of Training	Time for training(s)	Training Correct%
1 239	3	96.83199
2 239	3	98.14128
3 239	2	98.15720
4 179	2	98.12700
5 179	2	97.13207
6 179	2	98.07568
7 179	1	98.15218
8 239	3	98.18240

Hình 10 a. Sai số các phương án trong giai đoạn huấn luyện

No Examples of Testing	Time for Testing(s)	Testing Correct%
59	0	95.11726
59	0	97.15395
59	0	97.28170
119	0	97.12961
119	0	94.51209
119	0	97.02447
119	0	97.28776
59	0	97.33617

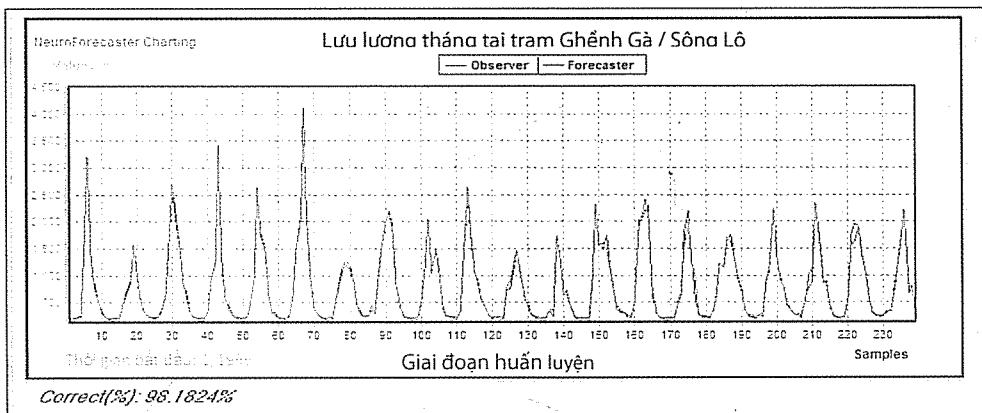
Hình 10 b. Sai số các phương án trong giai đoạn kiểm tra

NGHIÊN CỨU & TRAO ĐỔI

Ta có thể nhận thấy rằng phương án 8, với số mẫu huấn luyện là 239, số mẫu kiểm tra là 59, tuy có thời gian huấn luyện là 3 giây, lâu hơn các phương án khác không nhiều, nhưng lại cho ra độ chính xác trong giai đoạn huấn luyện là 98,18240% và trong giai đoạn kiểm tra là 97,33617% tốt nhất trong các phương án còn lại. Nên ta chọn phương án 8 là kết

qua cấu trúc ANN mô phỏng lưu lượng ở trạm Ghềnh Gà.

Biểu đồ so sánh kết quả tính trong giai đoạn huấn luyện so với số liệu thực đo chỉ ra trong hình 11, và biểu đồ so sánh kết quả tính toán mô phỏng trong giai đoạn kiểm tra được chỉ ra trong hình 12.

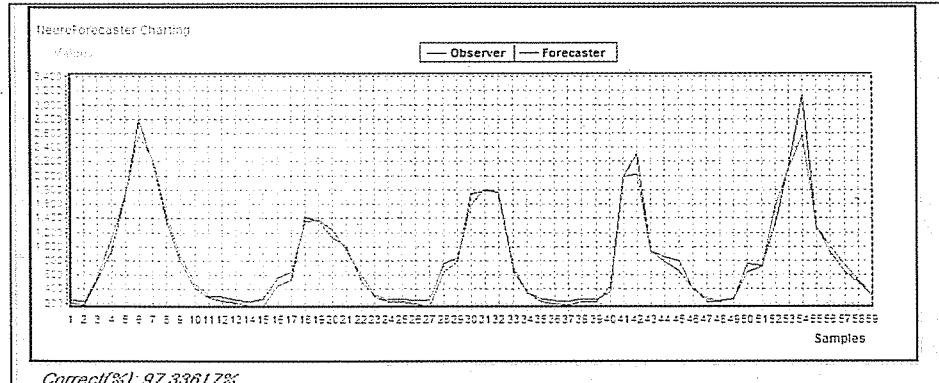


Hình 11. So sánh kết quả tính QGG(t) và giá trị thực đo trong giai đoạn huấn luyện (Phương án 8)

Tham số thống kê đánh giá kết quả mô phỏng:

Hệ số EI (Efficiency Index) được đề nghị bởi: NASH và SUTCLIFFE năm 1970:

$$EI = \frac{SST - SSE}{SST} \quad \text{với} \quad SST = \sum_{i=1}^T (Q_i - Q_{av})^2 \quad \text{và} \quad SSE = \sum_{i=1}^T (Q_i - F_i)^2$$



Hình 12. So sánh kết quả tính QGG(t) và giá trị thực đo trong giai đoạn kiểm tra (Phương án 8)

Với Q_i là lưu lượng thực đo ở thời điểm i , Q_{av} là lưu lượng thực đo trung bình trong khoảng thời gian T ; F_i là kết quả mô phỏng hay dự báo ở thời điểm i .

Hệ số RMSE (Root Mean Square Error) và RMSEM (Root Mean Square Error Mean):

$$RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^T (Q_i - F_i)^2} \quad \text{và}$$

$$RMSEM = \frac{RMSE}{Q_{av}}$$

Hệ số MAD (Mean Absolute Deviation) và RMSES (Root Mean Square Error over Standard Deviation):

$$MAD = \frac{1}{T} \sum_{i=1}^T |Q_i - F_i| \quad \text{và}$$

$$RMSES = \frac{RMSE}{S}$$

Với S độ lệch chuẩn (Standard Deviation).

Bảng đánh giá kết quả mô phỏng:

Kết quả giá trị EI trong hai giai đoạn huấn luyện (0,9824) và kiểm tra (0,9897) khá gần 1; giá trị RMSEM trong giai đoạn huấn luyện (0,1022) và

trong giai đoạn kiểm tra (0,1157); và RMSES trong giai đoạn huấn luyện (0,1004) và trong giai đoạn

kiểm tra (0,1323) nhỏ cho thấy rằng kết quả mô phỏng khá tốt.

Bảng 1. Đặc điểm thống kê của kết quả mô phỏng lưu lượng tháng ở trạm Ghềnh Gà, sông Lô

GIAI ĐOẠN	EI	RMSE (m ³ /s)	RMSEM	MAD (m ³ /s)	RMSES
Huấn luyện	0,9824	78,1702	0,1022	51,3402	0,1004
Kiểm tra	0,9897	90,9630	0,1157	56,4636	0,1323

5. Nhận xét về phương trình và kết quả áp dụng

Trong quá trình sử dụng chương trình và áp dụng đổi mới bài toán mô phỏng dòng chảy tháng tại trạm thủy văn Ghềnh Gà rút ra một số nhận xét sau:

Khi lưu lượng ở trạm đích thời điểm trước QGG(t-1) được loại bỏ khỏi danh sách trạm nguồn, sai số sẽ cao hơn trường hợp đại lượng này được giữ lại. Điều đó cho thấy, có một lưu lượng nền liên tục cung cấp cho trạm đích ở bước thời gian (t-1) ở khá xa trạm đích. Đó chính là phần diện tích lưu vực sông Lô ở lãnh thổ Trung Quốc (chiếm 40%) mà không có dữ liệu trạm đo mưa nào được đưa vào mô hình. Tuy các trạm đo mưa trong lãnh thổ Việt Nam, lân cận biên giới, cũng đại diện được phần nào, nhưng không hoàn hảo. Ngoài ra, còn một nguyên nhân khác: có thể tồn tại dòng nước ngầm thấm ra ngoài chuyển vào dòng chảy mặt trong lưu vực sông Lô, Gâm để duy trì lưu lượng nền. Điều này cũng có thể xảy ra, vì đây là lưu vực địa hình nhiều đồi núi, và thung lũng. Lưu lượng thấm này cũng là phần lưu lượng duy trì dòng chảy trong sông suối về mùa khô, không mưa.

Nhìn vào kết quả tính toán trong hai giai đoạn huấn luyện và kiểm tra, hình 11 và hình 12, ta có thể thấy rằng, sai số có xu hướng lớn ở đỉnh cao của các đường cong. Tức sai số lớn xảy ra chủ yếu vào mùa mưa, khi cường độ mưa lớn. Sự mô phỏng không nắm bắt được quy luật biến đổi này một cách hoàn hảo. Đây là điều hợp lý, vì ta đã không có đủ dữ liệu các trạm mưa ở phần lãnh thổ Trung Quốc để đưa vào mô hình mô phỏng, và do đó nó trở thành yếu

tố ngẫu nhiên mà mạng ANN không thể nắm bắt được vì thiếu thông tin. Qua đó chứng minh rằng, đặc điểm mưa lưu vực sông Lô ở phần lãnh thổ Trung Quốc có khác với quy luật mưa ở lãnh thổ Việt Nam.

Khi ta loại bỏ một trong bảy trạm mưa trong danh sách, thì ta nhận thấy độ chính xác giảm đi. Điều này chứng tỏ quy luật mưa giữa các trạm cũng khác nhau, không thể tạo tương quan cường độ mưa ở một trạm từ cường độ mưa ở các trạm còn lại. Kết luận này khá phù hợp vì như đã trình bày ở phần trên, đặc tính khí tượng của lưu vực: chế độ mưa ở lưu vực khá phong phanh, dao động khá lớn và khá khác nhau về mặt không gian cũng như thời gian (Xem mục 4).

Kết quả lưu lượng mô phỏng tốt nhất ở Ghềnh Gà QGG(t) là hàm số của lưu lượng tại trạm nguồn Chiêm Hóa, cùng thời gian QGH(t), và bảy trạm nguồn mưa, cũng ở cùng thời gian: RHXP(t), RBL(t), RHG(t), RBM(t), RYB(t), RBQ(t), RCR(t). Sở dĩ như thế vì các trạm này không xa trạm đích, và thời đoạn tính toán là tháng, khá dài. Tuy nhiên, nếu có dữ liệu chi tiết hơn, chẳng hạn 7 ngày, hoặc 1 ngày, thì kết quả mô phỏng tốt chắc chắn sẽ có bổ sung thêm các trạm nguồn ở các thời điểm quá khứ như (t-1), (t-2), (t-3)...tùy theo khoảng cách từ trạm này đến trạm đích. Và khi đó ta có thể đạt được mô hình dự báo, Ví dụ: $Q_{\text{đích}}(t) = f(Q_i(t-1), R_j(t-1), R_j(t-2) \dots)$; với i là chỉ số các trạm lưu lượng nguồn ở thượng lưu, và j là chỉ số của các trạm đo mưa trong lưu vực. Chẳng hạn, nếu t là ngày hoặc tuần, thì mô hình có thể giúp ích cho chế độ vận hành hồ chứa hợp lý, tránh đi trường hợp phải vận hành bị động, có thể gây hậu quả nghiêm trọng cho vùng dân cư ở hạ

NGHIÊN CỨU & TRAO ĐỔI

lưu hồ chứa.

Ngoài ra, nếu có được dữ liệu đo mực nước ở thượng và hạ lưu của trạm đích, kết hợp với dữ liệu lưu lượng tại trạm đích, ta có thể tìm được mối quan hệ hàm giữa yếu tố lưu lượng và mực nước. Nếu đạt được mối quan hệ tốt, thì mực nước tại các trạm đo này có thể được dùng để tìm ra lưu lượng trạm đích. Và như vậy, về mặt ứng dụng, kết quả này sẽ trở nên rất có ý nghĩa vì thiết lập trạm đo mực nước sẽ dễ dàng và ít tốn kém hơn nhiều so với việc thiết lập trạm đo lưu lượng.

Hệ thống chương trình NeuroForecaster tuy đã là công cụ giúp ích và hỗ trợ tốt cho người sử dụng, nhưng nó vẫn chưa hoàn hảo. Nếu như hệ thống này kết hợp được với hệ thống Công Nghệ Thông Tin Địa Lý (GIS) thì là một điều rất tốt. Vì hiện nay trên thế giới, cũng như ở nước ta, việc quản lý nguồn nước và lưu trữ dữ liệu của lưu vực sông đã được đưa vào cơ sở dữ liệu GIS. Ưu điểm của nó là hiển thị thông tin các lớp đối tượng không gian, kèm dữ liệu thuộc tính một cách tiện lợi thông qua bản đồ số. Ngoài ra, chính sách quản lý và điều hành nguồn nước ở nước ta cũng đã bắt đầu dựa trên nguyên tắc quản lý nguồn nước theo lưu vực sông. Biện pháp quản lý này nhằm đạt được chế độ quản lý nguồn nước tối ưu tổng thể, chất lượng

nước và môi trường bền vững, tránh hậu quả nghiêm trọng do thiên tai lũ lụt, hạn hán nhất là trong điều kiện mà ảnh hưởng của sự biến đổi khí hậu đang đe dọa nước ta cũng như trên toàn thế giới.

6. Kết luận

Qua quá trình sử dụng hệ thống NeuroForecaster đối với bài toán mô phỏng lưu lượng tháng ở trạm Ghềnh Gà, cho thấy NF là một công cụ tốt và tiện lợi, giúp người sử dụng tìm ra mô hình mô phỏng thích hợp mà không cần đòi hỏi một kiến thức nào về lập trình máy tính và tổ chức dữ liệu.

Kết quả áp dụng cho thấy có sự phù hợp cao giữa dòng chảy mô phỏng tháng và số liệu thực đo tại trạm thủy văn Ghềnh Gà. Nếu có dữ liệu chi tiết hơn theo tuần, hoặc theo ngày, hoặc dữ liệu mực nước ở thượng và hạ lưu trạm đích, thì kết quả mô phỏng sẽ có thể mang lại tính chất dự báo, mang lại hiệu quả cao hơn về mặt thực tế đối với chế độ quản lý nguồn nước tối ưu và bền vững theo lưu vực sông.

Nhu cầu tích hợp NeuroForecaster với GIS, mô hình thủy lực, thủy văn là các vấn đề nghiên cứu cần được quan tâm vì sẽ mang lại hiệu quả ứng dụng cao cho ngành KTTV và quản lý nguồn nước lưu vực sông.

Tài liệu tham khảo

1. Lê Văn Dực (2001), "Application of BackPropagation Neural Network Model to simulate the monthly runoffs at TriAn and PhuocHoa hydrological stations to fill up the missing of the historical data series", *Science & Technology Developement*, Vol 4, No.8, 8-9-2001, pp.57-64, VIETNAM NATIONAL UNIVERSITY-HO CHI MINH CITY.
2. Lê Văn Dực (2009), "Comparison of performance statistics on resulting monthly runoffs simulated by Artificial Neural Network, and other traditional models at Dong Nai river basin", *Tuyển Tập Công Trình, Hội Nghị Khoa Học Cơ Học Thủy Khí Toàn Quốc năm 2008, Phan Thiết*, trang 67 – 80, Nhà Xuất Bản Khoa Học Tự Nhiên và Công Nghệ, Hà Nội, Quý III, 2009.
3. Nguyễn Minh Đạo (2002)," Mạng Nơron với thuật giải lan truyền ngược kết hợp luật heuristic", *Luận án Cao học ngành CNTT* khoá 10, 2002, ĐHBK, ĐHQG TP.HCM.
4. NAM, L.H., PHIÊN,H.N. and DAS GUPTA, A. (1998); "Filtering and Forecasting of Monthly Streamflows by Backpropagation Neural Networks with an error Updating Method", *Water Resources Journal*, September 1998, pp.29-39.
5. W. MAASS (1997), "Networks of Spiking Neron: The Third Generation of Neural Network Model", *Institute for Theoretical Computer Science*, 4/8/1997, A-8010 Graz, Austria; http://mathsci.kaist.ac.kr/~nipl/am621/lecturenotes/spiking_neurons_2.pdf.
6. Cổng thông tin điện tử Tỉnh Hà Giang, <http://www.hagiang.gov.vn/>.
7. Cổng thông tin điện tử Chính Phủ Việt Nam, <http://gis.chinhphu.vn/>