KHẢ NĂNG ỨNG DỤNG MẠNG THẦN KINH NHÂN TẠO (ANN) VÀO DỰ BÁO LŨ SÔNG TRÀ KHÚC VÀ SÔNG VỆ

ThS. **Lê Xuân Cầu**, KS. **Nguyễn Văn Chương**
Trung tâm tư liệu KTTV

Tóm tắt: Mạng thần kinh nhân tạo (ANN) được dùng để thiết lập quan hệ mưa-mực nước (hoặc mưa-dòng chảy). ANN dùng để xây dựng các phương án dự báo lũ trên sông Cầu, sông Trà Khúc và sông Vệ. ANN dùng dự báo đường quá trình mực nước lũ và mực nước đỉnh lũ sông Cầu Trạm Gia Bảy và dự báo đỉnh lũ sông Trà Khúc Trạm Sơn Giang, sông Vệ Trạm An Chỉ. Kết quả dạy và kiểm tra độc lập ANN chỉ ra rằng ANN là công cụ hữu hiệu để dự báo lũ trên các sông.

1. Bài toán dự báo lũ trên các sông

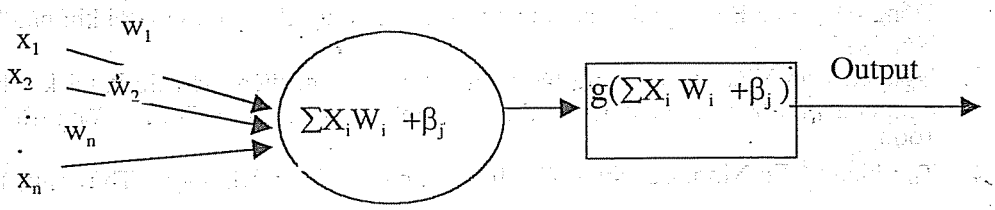
Dự báo *đường quá trình mực nước lũ (hoặc lưu lượng nước) và đỉnh lũ trên các sông là một bài toán quan trọng* trong dự báo thủy văn. Khác với công cụ dự báo lũ hiện có, ANN không những có thể được dùng dự báo dòng chảy theo mưa mà nó còn có thể dùng để dự báo mực nước lũ tại các trạm chỉ có tài liệu đo mực nước. Kết quả kiểm tra độc lập cho các con lũ thường tốt hơn so với các mô hình khác.

2. Khái niệm cơ bản về ANN

ANN là một quan hệ hàm bất kỳ giữa các biến đầu ra (outputs) và các biến đầu vào (inputs) của một hệ thống [1, 2, 3]. Trong thực tế xử lý thông tin ANN đã cho kết quả tốt khi đem ứng dụng trong một số lĩnh vực. Trong ANN dữ liệu được biến đổi ở mức biểu diễn số của máy tính. Các số trong máy tính được biểu diễn bởi các bít, bởi vậy các phép toán biến đổi dữ liệu sẽ thực hiện trên các bít riêng biệt. Người làm mô hình không có một ảnh hưởng trực tiếp nào tới các bít mang các thông tin. Như vậy, sau khi biến đổi dữ liệu mà nó mô tả ký hiệu của thế giới tự nhiên thành các dòng bít thì các ký hiệu nguyên thủy không còn có một liên quan nào tới các phép toán trên các bít nữa.

ANN là mô hình toán của một hệ tương tự não. Mô hình lấy neuron làm đơn vị xử lý cơ bản. Mỗi neuron được đặc trưng bởi:

1. Mức hoạt động (trạng thái phân cực của neuron).
2. Giá trị ra (Tốc độ phát hoả của neuron).
3. Tập nối các đầu vào của neuron (khớp nối các tế bào thần kinh).
4. Giá trị ngưỡng (Mức nghỉ bên trong của tế bào thần kinh).



H.1. Đơn vị xử lý cơ bản (neuron) của ANN

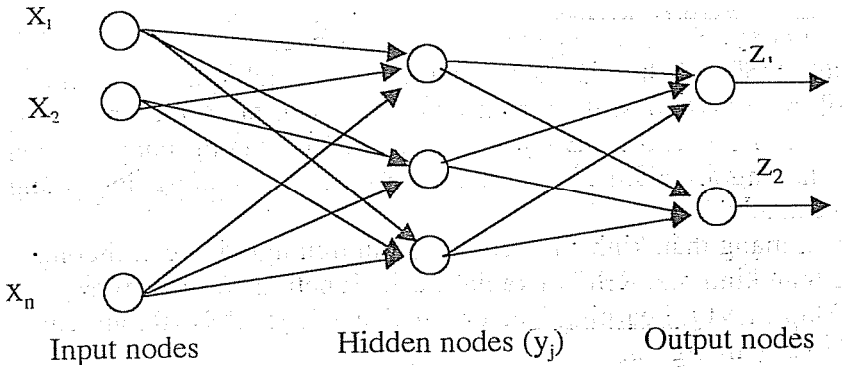
ở đây: $x_1, x_2, .., x_n$ là các đầu vào của neuron thần kinh.

$w_1, w_2, .., w_n$: các trọng số tương ứng với các đầu vào.

β_j: Giá trị ngưỡng của neuron.

g(.) là hàm chuyển đổi (thường là các hàm nén: hàm lôgíc).

Mạng thần kinh nhân tạo đa lớp gồm ít nhất 3 lớp các nút (nút là các neuron giả lập). Lớp đầu vào là lớp thụ động cho phép đưa vào ANN các ví dụ để học. Tiếp đến là lớp ẩn (lớp mà các nút của nó không có liên hệ trực tiếp với các lớp bên ngoài). Lớp ẩn tính các tổ hợp phi tuyến của các đầu vào. Lớp đầu ra tính các tổ hợp phi tuyến của đầu ra của các nút ẩn.



H.2. Mạng thần kinh nhân tạo đa lớp.
(ANN với 1 lớp ẩn)

Ở đây: $x_1, x_2, .., x_n$ là các đầu vào của ANN.
$z_1, z_2, .., z_n$ là các đầu ra của ANN.

Mạng thần kinh nhân tạo BPN (Error Backpropagation Network) (H. 2) thường hay được dùng hơn cả. Các tín hiệu đầu vào ANN sẽ được phân bố giữa các nút ẩn. Sau đó các nút ẩn sẽ biến đổi chúng thành các tín hiệu ra của nút ẩn, những tín hiệu này sẽ được truyền cho các đầu ra của ANN.

N_{inp} các đầu vào của ANN được ký hiệu:

x_i , i=1, .., N_{inp}

N_{hids} các đầu ra y_j của lớp ẩn sẽ là:

$$y_j = g\left(a_{0j} + \sum_{i=1}^{N_{inp}} a_{ij} x_i\right) \quad , \; j=1,.. , N_{hids}$$

N_{out} các giá trị ra của các đầu ra sẽ là:

$$z_k = g\left(b_{0k} + \sum_{j=1}^{N_{hids}} b_{jk} y_j\right) \quad , \; k=1,.. , N_{out}$$

Hàm chuyển đổi g(u) giới hạn giữa 0 và +1 là hàm lôgíc sau:

$$g(u) = \frac{1}{1+e^{-u}}$$

Như vậy ANN sẽ có tất cả $(N_{inp}+1).N_{hids} + (N_{hids}+1).N_{out}$ tham số (Các trọng số a_{ij} và b_{jk}). Các giá trị của các tham số a_{ij} và b_{jk} sẽ nhận được bởi quá trình dạy ANN sao cho sai số nhận được từ ANN nhỏ nhất.

Gọi các giá trị quan trắc thứ i của đầu ra **k** của ANN là OBS_{ki} và các giá trị tính toán thứ i của đầu ra **k** của ANN là MOD_{ki} khi đó sai số quân bình bình phương MSE của ANN sẽ là:

$$MSE = \frac{1}{N_{exams} \cdot N_{out}} \left(\sum_{k=1}^{N_{out}} \left(\sum_{i=1}^{N_{exams}} (OBS_{ki} - MOD_{ki})^2 \right) \right)$$

Ở đây: N_{examp} là số các ví dụ để ANN học.

Bởi vì ta biết MSE có thể biểu diễn bằng hàm giải tích qua các biến đầu vào và các tham số, thêm vào đó MSE là hàm khả vi cho nên ta có thể tìm giá trị của các trọng số a và b sao cho MSE đạt giá trị cực tiểu. Phương pháp gradien giảm (tìm trọng số theo hướng giảm của gradien hàm mục tiêu) thường được dùng để tìm các trọng số của ANN.

Ngày nay ngoài mạng thần kinh BPN có dạng như trên người ta còn thường hay dùng các mạng thần kinh như RBFN (Radial Basic Function Net) và SOM (Self-organization Map), LVQ (Learning Vector Quatizer) hoặc RNN (Recurrent Neural Network) để xử lý thông tin.

Định lý về sự tồn tại mạng thần kinh ánh xạ Kolmogorov như sau: Một hàm liên tục bất kỳ $f(x_1, x_2, \ldots, x_n)$ xác định trên khoảng I^n ($I=[0,1]$) có thể được biểu diễn dưới dạng:

$$f(x) = \sum_{j=1}^{2n+1} \chi_j \left(\sum_{i=1}^{n} \Psi_{ij}(x_i) \right)$$

*ở đây: χ_j, Ψ_{ij} là **các hàm liên tục một biến**. Ψ_{ij} là hàm đơn điệu ,không phụ thuộc vào hàm f.*

Hay nói một cách khác ANN có thể gần đúng một quan hệ phi tuyến bất kỳ. Một thuật toán lan truyền ngược sai số cho phép ANN có thể tìm trọng số một cách dễ dàng. Nhờ thuật toán này mà ANN đang ngày càng được áp dụng rộng rãi xử lý thông tin và giải các bài toán kỹ thuật.

3. Dự báo lũ bằng ANN

ANN dùng dự báo lũ có 4 nút ẩn. Tập số liệu mẫu đã có được chia làm hai tập số liệu con. Một tập mẫu để dạy ANN và tập mẫu còn lại để kiểm tra độc lập. Kết quả xây dựng phương án thể hiện rõ nét ở các đồ thị và các chỉ tiêu đánh giá dự báo.

a. Dự báo lũ sông Cầu

ANN được dùng để xây dựng 4 phương án dự báo lũ sông Cầu.

Sông Cầu thuộc hệ thống sông Thái Bình nằm khoảng 21°7 đến 22°20 độ vĩ bắc từ 105°27 đến 106°18 độ kinh đông. Phần lớn lưu vực nằm ở tỉnh Bắc Cạn và tỉnh Thái Nguyên. Lưu vực sông Cầu tính đến vị trí Trạm thủy văn Gia Bảy gồm: sông chính bắt nguồn từ núi Tam Tao. Phần thượng lưu sông (kể đến Bắc Cạn) khoảng 50 km chạy dọc theo hướng tây-đông, từ Bắc Cạn đến Thái Nguyên chạy theo hướng bắc nam. Sông chính chảy qua thị xã Bắc Cạn, thành phố Thái Nguyên và Khu công nghiệp gang thép Thái Nguyên.

Lưu vực sông Cầu có 2 tâm mưa lớn. Tâm mưa Chợ Đồn phía tây dãy Cứu Quốc. Tâm mưa Thái Nguyên do dãy Tam Đảo chắn phía tây. Lượng mưa từ 1700 đến 2300mm và theo qui luật lượng mưa giảm dần từ tây sang đông. Mùa mưa

thường xảy ra từ tháng 5, kết thúc vào tháng 10 và tập trung cao điểm vào tháng 7 - 8.

Các đặc trưng cơ bản của lưu vực Trạm Thủy văn Gia Bảy: Vị trí trạm 21°36' vĩ bắc — 105°50' kinh đông. Diện tích hứng nước 2830km^2. Độ dài sông 144km. Chiều dài lưu vực 117km. Chiều rộng lưu vực 24,2km. Độ dốc lưu vực 24,5%. Độ dốc sông 15%. Số liệu có được trên sông Cầu đã qua hiệu chỉnh, chất lượng đảm bảo.

- Số liệu mực nước mùa lũ từ năm 1991-1994 với thời gian 6h của 2 Trạm thủy văn Chợ Mới và Gia Bảy. Số liệu mực nước đỉnh lũ từ cấp báo động I trở lên (H ≥ 2450cm) đo được từ năm 1959 — 1997.

- Lượng mưa tương ứng với các thời đoạn của 7 trạm: Phương Viên, Bắc Cạn, Chợ Mới, Định Hoá, Phú Lương, Võ Nhai, Thái Nguyên.

Quan hệ mưa rào dòng chảy phi tuyến và rất phức tạp. Tuy nhiên, ta có thể gần đúng các quan hệ đó bằng ANN và lập các phương án dự báo lũ sông Cầu như sau:

- **Bài toán 1.** Dự báo mực nước lũ Trạm thủy văn Gia Bảy $H_{gb,t+\tau}$ theo $H_{gb,t}$ và lượng mưa trung bình lưu vực X:

$$H_{gb,t+\tau} = f(H_{gb,t}, X)$$

Có tất cả 1776 mẫu. Chọn 1408 mẫu để dạy và 368 mẫu để kiểm tra. Kết quả dạy ANN và kiểm tra độc lập thể hiện ở các chỉ tiêu đánh giá dự báo.

Sai số cho phép tính theo công thức:

$$\sigma_\Delta = \sqrt{\frac{\sum(\Delta H_i - \Delta\overline{H})^2}{n-1}} = \sqrt{\frac{880624,5}{1407}} = 25,0cm$$

Trong đó: $\Delta H_i = H_{(t+6)} - H_{(t)}$

$$\Delta\overline{H} = \frac{\sum \Delta H_i}{n}$$

Độ lệch của yếu tố dự báo $\sigma = \sqrt{\frac{\sum(H_{TD} - \overline{H})^2}{n-1}}$

Độ lệch của dự báo kiểm tra

$$\sigma_1' = \sqrt{\frac{\sum(H_{TD} - H_{oB})^2}{n-1}} = \sqrt{\frac{499005}{1407}} = 19cm$$

Tính tỷ số : $\frac{\sigma'}{\sigma} = \frac{19}{66,5} = 0,29$

Tính tỷ số tương quan:

$$\eta = \sqrt{1 - (\frac{\sigma'}{\sigma})^2} = \sqrt{1 - (0,286)^2} = 0,96$$

Mức đảm bảo: $P\% = \frac{m}{n}$

m : số điểm đúng

n : tổng số điểm tính toán

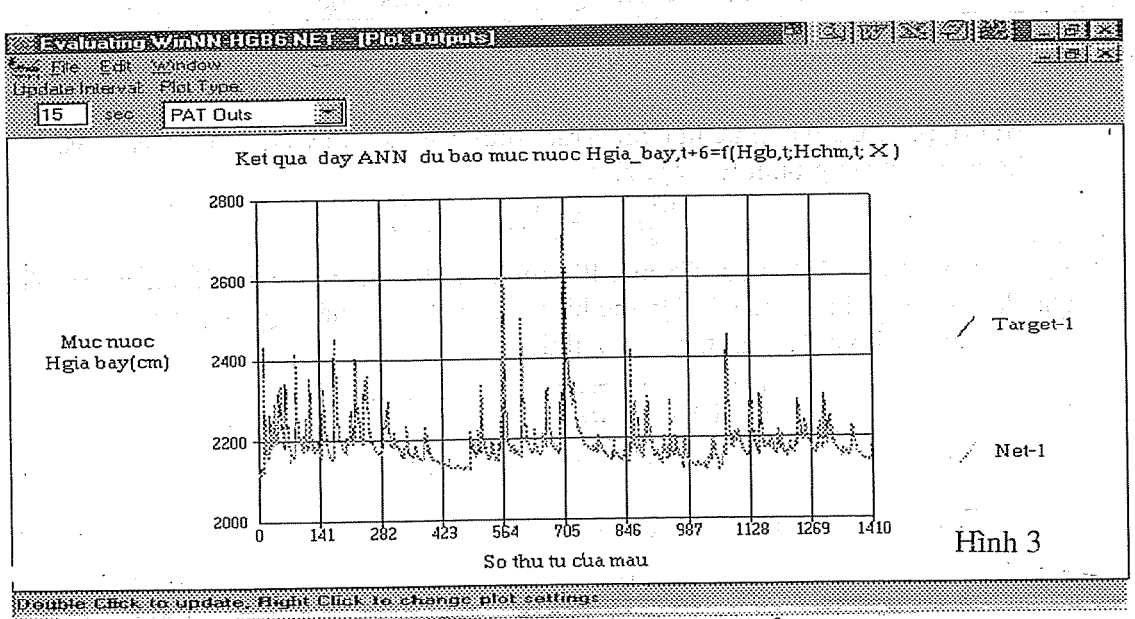
So sánh với các chỉ tiêu: $\frac{\sigma}{\sigma} \le 0{,}50 \quad ; \eta \ge 0{,}86, P \ge 82\%$ do đó chất lượng phương án dự báo thuộc loại tốt.

Số liệu để kiểm tra (độc lập) 368 mẫu. Mức đảm bảo đạt được là $P\% = \frac{310}{36} = 84\%$

- **Bài toán 2.** Dự báo mực nước Trạm thủy văn Gia Bảy $H_{gb,t+\tau}$ theo $H_{gb,t}$, mực mước Trạm thủy văn Chợ Mới $H_{chm,t}$ và lượng mưa trung bình lưu vực X:

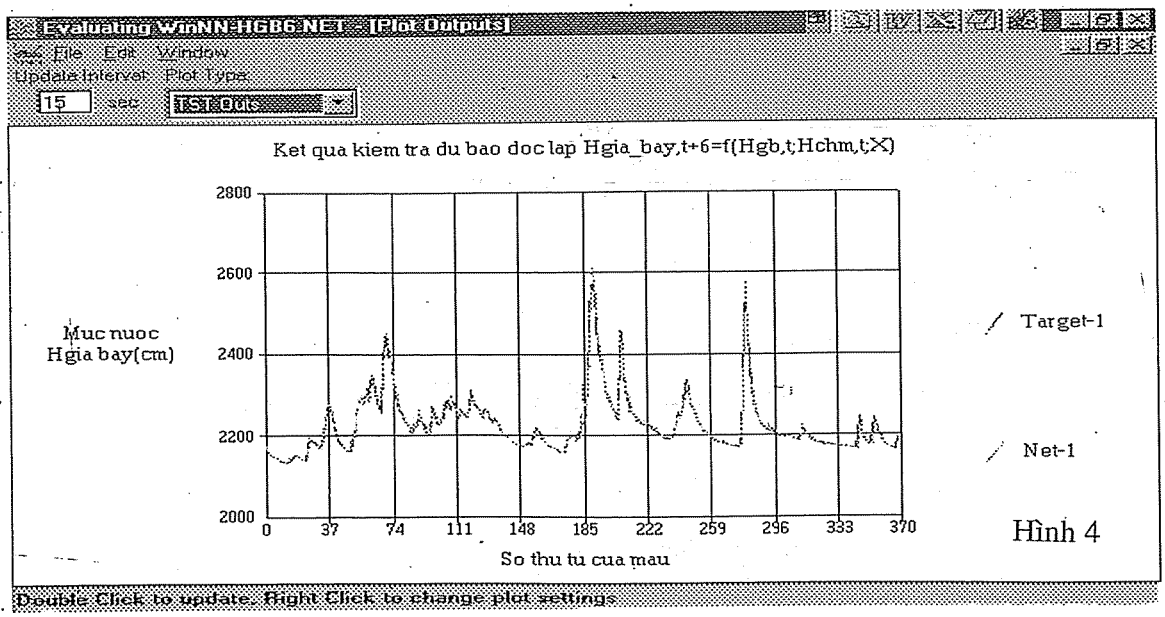
$$H_{gb,t+\tau} = f(H_{gb,t}, H_{chm,t}, X)$$

Có tất cả 1776 mẫu. Chọn 1408 mẫu để dạy và 368 mẫu để kiểm tra. Kết quả dạy ANN và kiểm tra độc lập thể hiện ở Hình 3 và Hình 4. Các đường thực đo và tính toán đi sát nhau. Mức bảo đảm cho dạy ANN: P=90% và kiểm tra độc lập tương ứng P= 86%.



Hình 3

- **Bài toán 3.** Dự báo mực nước đỉnh lũ Trạm thủy văn Gia Bảy $H_{gb,max}$ theo mực

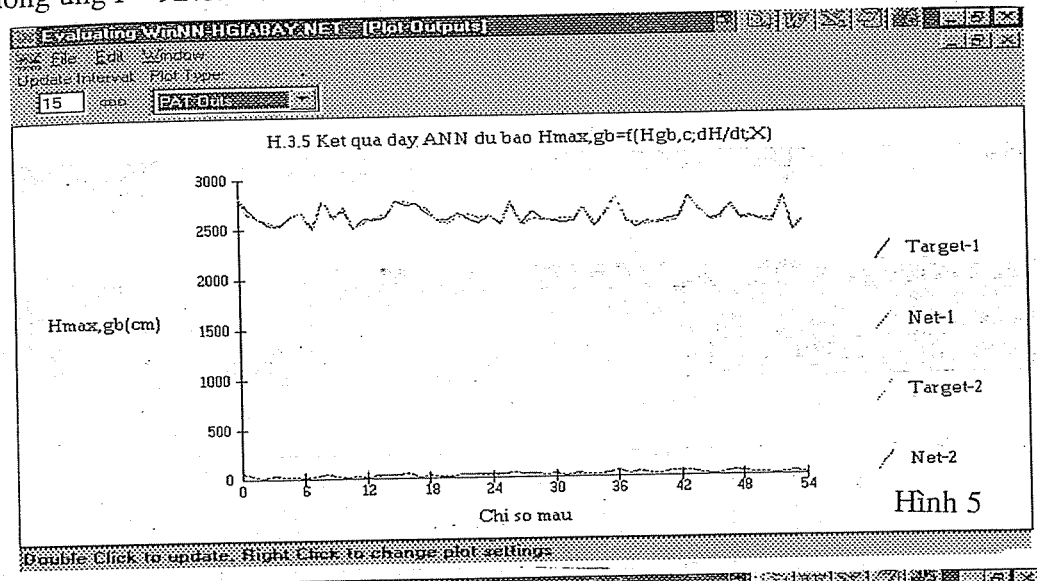


Hình 4

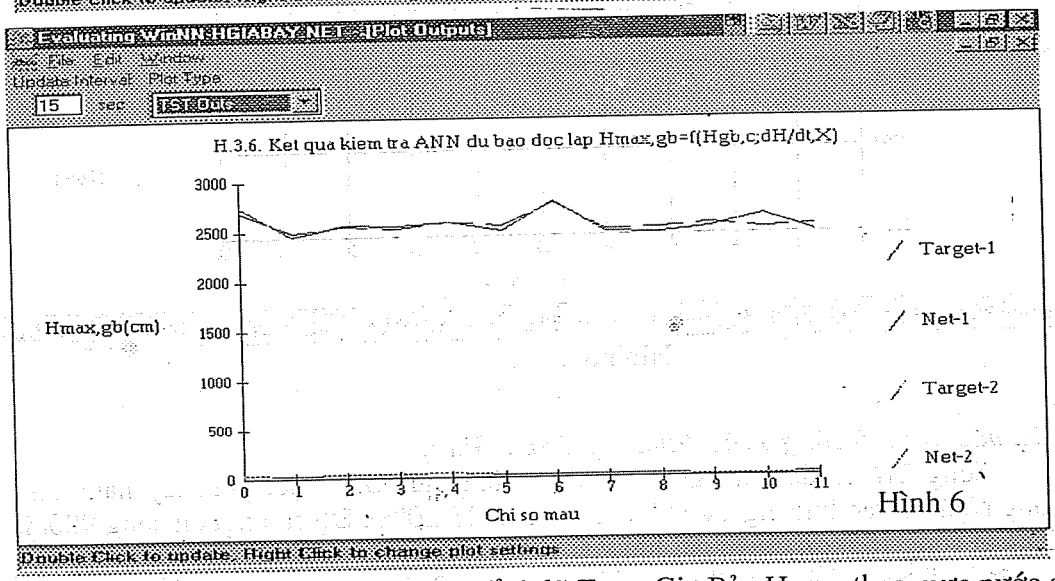
nước chân lũ Trạm thủy văn Gia Bảy $H_{gb,c}$. Cường suất mực nước và lượng mưa trung bình lưu vực X:

$$H_{gb,\max} = f(H_{gb,c}, dH/dt, X)$$

Có tất cả 67 mẫu, lấy 55 mẫu để dạy và 12 mẫu để kiểm tra độc lập. Kết quả dạy ANN và kiểm tra độc lập thể hiện ở H.5 và H.6. Các đường thực đo và tính toán đi sát nhau. Mức bảo đảm cho dạy ANN: P=96% và kiểm tra độc lập tương ứng P= 92%.



Hình 5

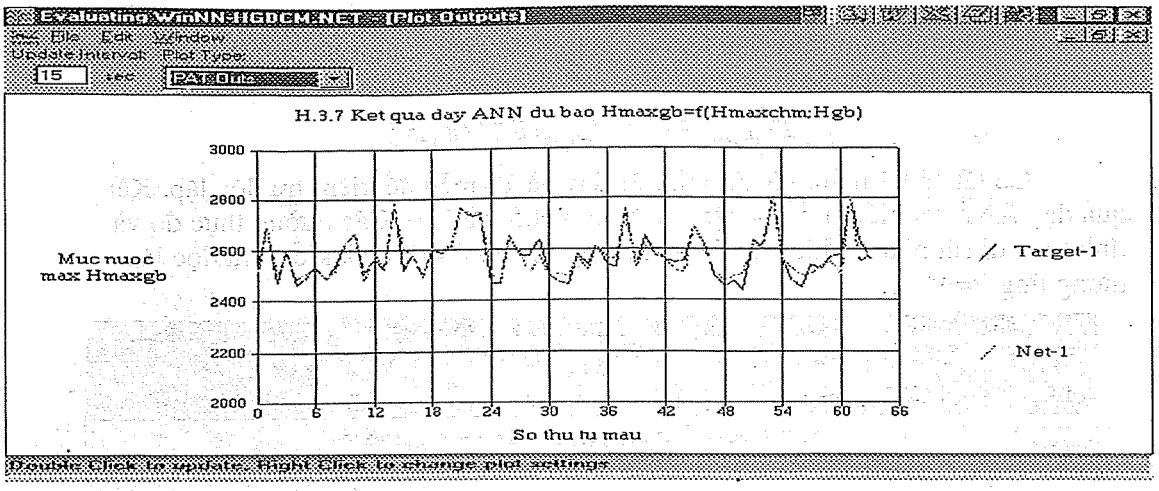


Hình 6

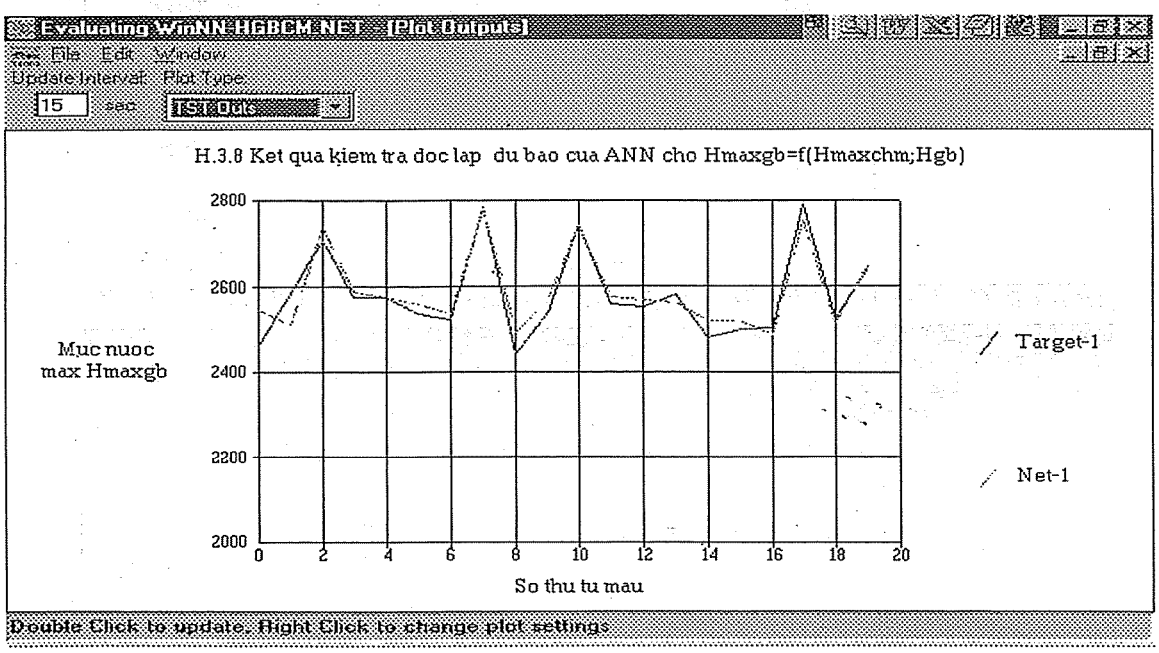
- **Bài toán 4.** Dự báo mực nước đỉnh lũ Trạm Gia Bảy $H_{gb,max}$ theo mực nước đỉnh lũ Trạm Chợ Mới $H_{chm,max}$ và mực nước Trạm Gia Bảy vào thời điểm $H_{chm,max}$.

$$H_{gb,\max} = f(H_{chm,\max}, H_{gb,t})$$

Có tất cả 84 mẫu. Chọn 64 mẫu để dạy và 20 mẫu để kiểm tra. Kết quả dạy ANN và kiểm tra độc lập thể hiện ở H.7 và H.8. Các đường thực đo và tính toán đi sát nhau. Mức bảo đảm cho dạy ANN: P=86% và kiểm tra độc lập tương ứng P= 85%.



Hình 7



Hình 8

b. Dự báo đỉnh lũ sông Trà Khúc_Trạm Sơn Giang

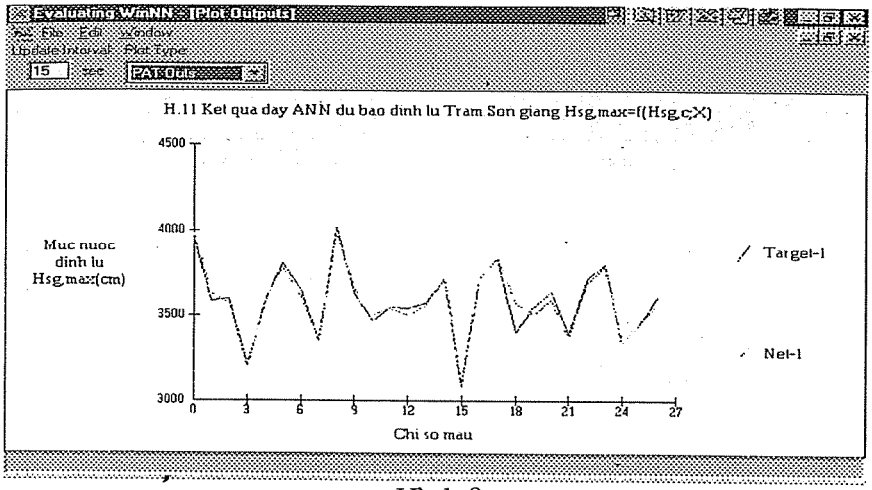
Sông Trà Khúc bắt nguồn từ vùng rừng núi Giá Vực, phía tây nam tỉnh Quảng Ngãi, ở vào khoảng $14^0$34'30"B và $108^0$25'20"Đ. Độ cao nguồn sông 900m, chiều dài sông 135km, chiều dài lưu vực 123km, diện tích lưu vực 3240km^2, độ dốc lưu vực 18,5 %, chiều rộng lưu vực 26,3km. Có 2 dạng lũ trên sông lũ đơn và lũ kép.

Tài liệu quan trắc có từ 1979 đến nay. Chất lượng tài liệu tốt [4].

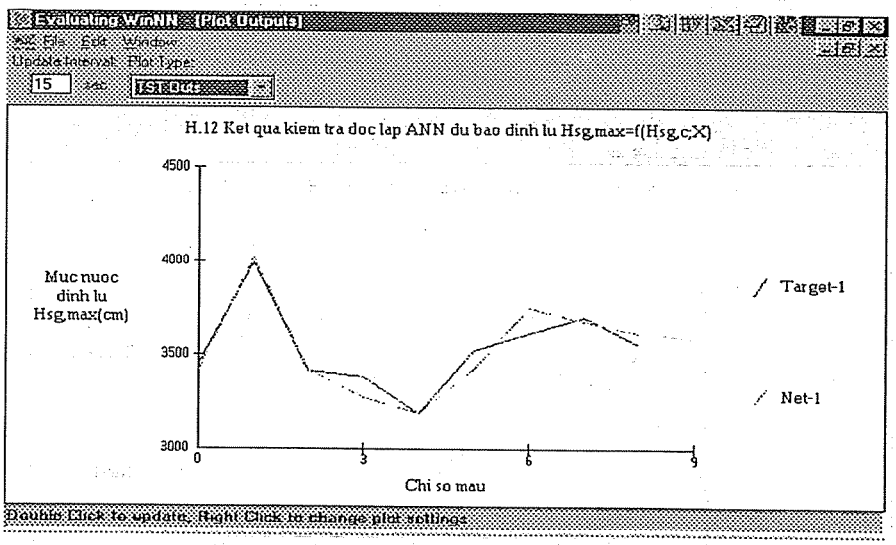
Tại Trạm Sơn Giang có tất cả 36 mẫu (từ 1995-1999) số liệu mực nước đỉnh lũ và mưa tương ứng tại hai Trạm Giá Vực và Sơn Giang. ANN được dùng để xây dựng phương án dự báo mực nước đỉnh lũ Trạm Sơn Giang $H_{sg,max}$ theo $H_{sg,c}$ và lượng mưa trung bình lưu vực X:

$$H_{sg,max} = f(H_{sg,c}, X)$$

Kết quả dạy ANN và kiểm tra độc lập thể hiện ở Hình 9 và Hình 10. Các đường mực nước thực đo và tính toán đi sát nhau.



Hình 9



Hình 10

c. Dự báo đỉnh lũ sông Vệ Trạm An Chỉ

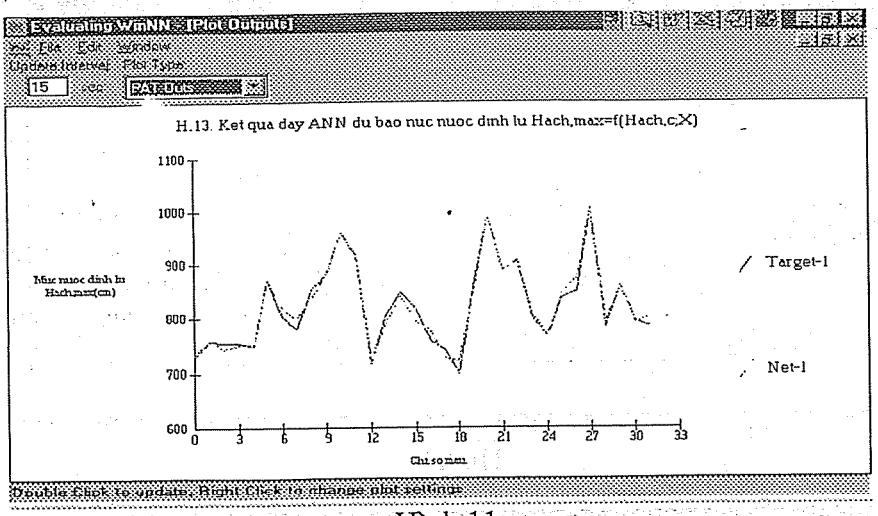
Sông Vệ bắt nguồn từ vùng rừng núi huyện Ba Tơ, đổ ra biển Đông tại Cổ Luỹ. Chiều dài sông 90km, chiều dài lưu vực 123km, diện tích lưu vực 1260km^2, độ dốc lưu vực 18,5 %, chiều rộng lưu vực 26,3km.

Tài liệu thủy văn từ Trạm An Chỉ và tài liệu mưa từ Trạm Ba Tơ có chất lượng tốt.

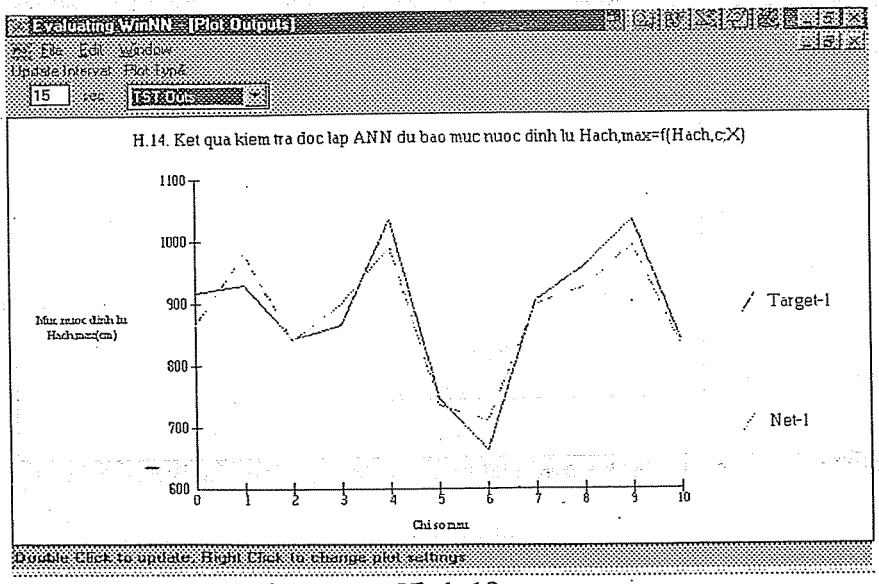
- Tại Trạm An Chỉ có tất cả 43 mẫu số liệu (từ 1990-1999) mực nước đỉnh lũ và mưa tương ứng tại Trạm Ba Tơ. ANN được dùng để xây dựng phương án dự báo mực nước đỉnh lũ Trạm An Chỉ $H_{ach,max}$ theo $H_{ach,c}$ và lượng mưa trung bình lưu vực X:

$$H_{ach,max} = f(H_{ach,c}, X)$$

Kết quả dạy ANN và kiểm tra độc lập thể hiện ở Hình 11 và Hình 12: Các đường thực đo và tính toán đi sát nhau.



Hình 11



Hình 12

4. Kết luận

Kết quả áp dụng ANN dự báo lũ trên các sông trên chỉ ra tính hiệu quả của ANN dự báo lũ trên một số sông miền Trung và ở miền Bắc. Mức bảo đảm dự báo thường tốt. Điều đó được chứng tỏ không những bằng các chỉ tiêu đánh giá dự báo thuỷ văn mà nó còn thể hiện bằng kiểm tra độc lập các kết quả. Để tăng độ chính xác dự báo ta có thể cho thêm thông tin về tâm mưa trên lưu vực sông. Tương tự như trên ANN có thể dùng để dự báo đường lưu lượng lũ hoặc mực nước lũ và mực nước đỉnh lũ cho các sông khác.

Tài liệu tham khảo

1. Lê Xuân Cầu. Ứng dụng của mạng thần kinh nhân tạo (ANN) xử lý dữ liệu khí tượng thuỷ văn. Tạp chí Khí tượng Thủy văn, 4(460), 1999.
2. Lê Xuân Cầu. Tìm số nút ẩn tối ưu của mạng thần kinh nhân tạo (ANN) bằng lý thuyết cực tiểu hoá mạo hiểm theo cấu trúc (SRM). Tạp chí Khí tượng Thủy văn, 8 (464), 1999.
3. Lê Xuân Cầu. Vấn đề ngoại suy mạng thần kinh nhân tạo. Tạp chí Khí tượng Thủy văn, 9(465), 1999.
4. Các phương án dự báo mực nước lũ sông Trà Khúc và sông Vệ. Trung tâm dự báo KTTV tỉnh Quảng Ngãi, 2000.