

NGHIÊN CỨU DỰ BÁO LŨ TRUNG HẠN LƯU VỰC SÔNG CẢ

NCS. Hoàng Thanh Tùng,

Khoa Thủy văn và Tài Nguyên nước, Đại học Thủy lợi

PGS. TS. Vũ Minh Cát

Khoa Kỹ Thuật Bờ biển, Đại học Thủy lợi

GS. Roberto Ranzi

Khoa Xây dựng, Đại học Brescia, Italia

KS. Trương Tùng Hoa

Tóm tắt: Dự báo lũ trung hạn cho các lưu vực sông Miền Trung nói chung và cho lưu vực sông Cả nói riêng là vấn đề bức xúc hiện nay. Từ việc nghiên cứu tổng quan các phương pháp dự báo lũ trung hạn tiên tiến ở trong và ngoài nước, nghiên cứu đã lựa chọn các mô hình lai ghép giữa tất định và ngẫu nhiên như HEC-HMS lai ghép với ARIMA(p,d,q) và EANN (mô hình mạng nơ ron theo thuyết tiến hóa) lai ghép với ARIMA(p,d,q) để xây dựng phương án dự báo lũ trung hạn phù hợp cho lưu vực sông Cả. Việc lai ghép các mô hình trên cho phép tận dụng được những ưu điểm của các phương pháp dự báo trung hạn truyền thống với các phương pháp hiện đại. Kết quả áp dụng cho lưu vực sông Cả là khá tốt và hoàn toàn có thể mở rộng để áp dụng cho các lưu vực sông khác ở Miền Trung.

Các từ khóa: Sông Cả, dự báo, trung hạn, EANN, ARIMA, HEC-HMS

1. MỞ ĐẦU

Những năm gần đây ở miền Trung, thiên tai lũ lụt và hạn hán xảy ra thường xuyên hơn với mức độ trầm trọng hơn, đặc biệt là năm 2007 (có tới 5 trận lũ xảy ra liên tiếp trong vòng 1 tháng) gây thiệt hại nặng nề về người và của cho các tỉnh miền Trung, trong đó có nhiều huyện như Quế Phong, Quỳnh Châu, Nghĩa Đàn thuộc tỉnh Nghệ An thuộc lưu vực sông Cả; thêm vào đó việc ô nhiễm nước và thiếu nước trên các lưu vực sông miền Trung, ngoài các nguyên nhân khách quan do thời tiết, khí hậu, còn có những nguyên nhân chủ quan như khả năng dự báo mưa, lũ trung và dài hạn chưa tốt, sự phối hợp quản lý, vận hành các hồ chứa hiện có trên các lưu vực sông này là chưa hợp lý.

Để giảm nhẹ thiệt hại do thiên tai lũ lụt gây ra, bên cạnh các biện pháp công trình như xây dựng hồ chứa phòng lũ ở đầu nguồn, còn cần phải thực thi các biện pháp phi công trình như nghiên cứu xây dựng quy trình vận hành hợp lý các hồ chứa, nâng cao khả năng dự báo lũ, kéo dài thời gian dự báo lũ để có các biện pháp đối phó với lũ lụt hiệu quả. Bài báo này trình bày kết quả nghiên cứu dự báo lũ trung hạn cho lưu

vực sông Cả.

Giới thiệu tóm tắt khu vực nghiên cứu

Hệ thống sông Cả là một trong 9 hệ thống sông lớn của nước ta. Sông chính bắt nguồn từ nước bạn Lào, chảy qua hầu hết địa phận tỉnh Nghệ An, được gọi là sông Cả. Đến hạ lưu vùng Nam Đàn (tại Chợ Trảng) sông tiếp nhận phụ lưu sông La từ Hà Tĩnh chảy sang. Từ ngã ba này ra tới biển sông được gọi là sông Lam.

Lưu vực hệ thống sông Cả nằm ở vùng Bắc Trung bộ, có tọa độ địa lý từ 18⁰15' đến 20⁰10'30" vĩ độ Bắc; 103⁰45'20" đến 105⁰15'20" kinh độ Đông. Điểm đầu của lưu vực nằm ở tọa độ 20⁰10'30" độ vĩ Bắc; 103⁰45'20" kinh độ Đông. Cửa ra của lưu vực nằm ở tọa độ 18⁰45'27" độ vĩ Bắc; 105⁰46'40" kinh độ Đông. Điểm sông Cả chảy vào đất Việt Nam tại Biên giới Việt Lào trên dòng Nậm Mô có tọa độ: 19⁰24'59" độ vĩ Bắc; 104⁰04'12" kinh độ Đông. Lưu vực hệ thống sông Cả nằm trên hai quốc gia, phần thượng nguồn nằm trên đất tỉnh Phong Sa Ván và Sầm Nưa của nước Cộng hòa dân chủ nhân dân Lào. Ở Việt Nam, lưu vực sông nằm trên địa phận của 3 tỉnh Thanh Hoá, Nghệ An, Hà Tĩnh (Bản đồ lưu vực hệ thống sông Cả

được minh họa trong Hình 1).

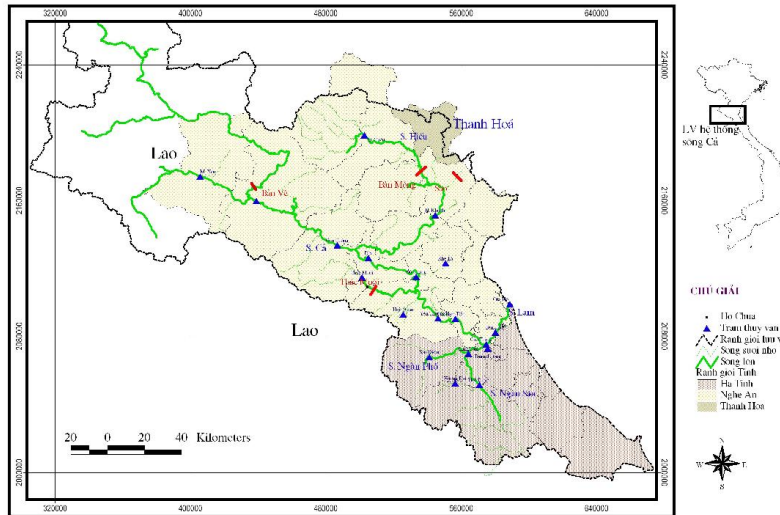
Trên đất tỉnh Thanh Hoá, lưu vực hệ thống sông Cả chiếm 1/2 diện tích huyện Như Xuân trên lưu vực sông Nhánh - sông Chàng. Trên tỉnh Nghệ An, lưu vực nằm trên đất huyện Quế Phong, Quỳnh Châu, Quỳnh Hợp, Nghĩa Đàn, Tân Kỳ (nhánh sông Hiếu). Kỳ Sơn, Tương Dương, Con Cuông, Anh Sơn, Đô Lương, Thanh Chương, Nam Đàn, Hưng Nguyên (dòng chính

sông Cả). Trên tỉnh Hà Tĩnh lưu vực nằm ở các huyện Hương Sơn, Hương Khê, Vũ Quang, Nghi Xuân

Theo tài liệu đặc trưng mạng lưới sông ngòi Việt Nam của tổng cục Thủy Văn xuất bản, diện tích tự nhiên toàn bộ lưu vực hệ thống sông Cả, tính từ thượng nguồn đến cửa sông là 27.200 km² và phân bố trên các địa dư hành chính như trong Bảng 1.

Bảng 1: Phân bố diện tích theo địa bàn hành chính

Lưu vực Hệ thống Sông Cả	Diện tích tự nhiên (km ²)	Diện tích lâm nghiệp (ha)	Diện tích nông nghiệp (ha)	Diện tích khác (ha)
Tổng lưu vực	27.200	1.798.830	449.266	471.910
Lào	9.470	681.840	66.290	198.870
Việt Nam	17.730	1.116.990	382.976	273.034
Thanh Hóa	441,21	32.400	1.500	10.221
Nghệ An	13860,79	884.410	331.734	168.935
Hà Tĩnh	3.428	200.180	49.742	92.878



Hình 1: Bản đồ lưu vực hệ thống sông Cả

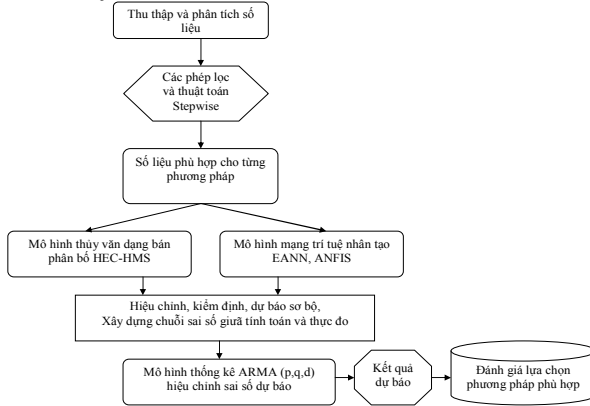
Trên lưu vực hệ thống sông Cả đã và đang xây dựng nhiều hồ chứa nước lớn như hồ sông Sào trên sông Sào, Bản Mông trên sông Hiếu, Bản Vẽ 1 và 2, Khe Bó trên sông Cả, Thác Muối trên sông Giăng. Đây đều là các hồ chứa đa mục tiêu như phòng lũ, phát điện, cấp nước cho lưu vực hệ thống sông Cả.

2. PHƯƠNG PHÁP NGHIÊN CỨU

Nghiên cứu dự báo lũ trung hạn trong nước và trên thế giới hiện nay vẫn là một đề tài hấp dẫn các nhà khoa học, đặc biệt là vấn đề làm thế

nào để nâng cao chất lượng và thời gian dự kiến. Hướng nghiên cứu trên thế giới trong những năm gần đây thường tập trung vào việc sử dụng phương pháp mạng Nơ ron nhân tạo (ANN) với nhiều thuật toán tối ưu khác nhau kết hợp với việc sử dụng các phương pháp thống kê truyền thống như mô hình ARIMA (p,d,q) và việc mở rộng áp dụng các mô hình thủy văn thông số phân bố dựa trên cơ sở tận dụng các thông tin từ viễn thám và GIS kết hợp với kết quả dự báo mưa từ các mô hình số trị dự báo

thời tiết như MM5, RAMS, HRM, BOLAM,... Chuyên đề nghiên cứu này cũng đi theo hướng mới này với phương pháp tiếp cận là sự lai ghép giữa các phương pháp truyền thống và hiện đại để tận dụng tối đa những ưu điểm của từng phương pháp. Phương pháp tiếp cận chung được đưa ra trong các sơ đồ tóm tắt ở Hình 2 dưới đây:



Hình 2: Sơ đồ nghiên cứu tóm tắt

Trong sơ đồ nghiên cứu trên, việc lựa chọn số liệu để phân tích là rất quan trọng, đặc biệt với phương pháp mạng nơ ron nhân tạo (ANN) khi ta có quá nhiều biến đầu vào mà không biết những số liệu nào có ảnh hưởng lớn đến biến đầu ra (biến cần dự báo). Nghiên cứu đã sử dụng thuật toán “stepwise” để thực hiện việc này. Thuật toán này cho phép lần lượt đưa vào và đưa ra từng biến để đánh giá mức độ ảnh hưởng/đóng góp của từng biến với biến cần dự báo. Khi mỗi biến đưa vào, mô hình sẽ phân tích và đánh giá các chỉ tiêu thống kê (ví dụ như P, T... hệ số tương quan bội R2) nếu chỉ tiêu đạt thì giữ lại, không đạt thì loại ra. Ngoài ra còn nhiều phương pháp lọc số liệu khác như “Goal Programming”, “Neuro wavelet” “Gamma Test”, ..vv. Như vậy sau khi áp dụng thuật toán này, ta đã lựa chọn được những biến đầu vào tốt nhất cho việc áp dụng mạng ANN.

Bên cạnh việc áp dụng thuật toán “stepwise” để lựa chọn số liệu, nghiên cứu đã sử dụng 2 mô hình lai ghép để dự báo dòng chảy sông Cã: đó là mô hình tất định bán phân bố (HEC-HMS) + mô hình ngẫu nhiên ARIMA (p,d,q) và mô hình mạng nơ ron nhân tạo (EANN) + mô hình ngẫu nhiên ARIMA (p,d,q). Trong đó mô hình HEC-

HMS và EANN là mô hình cho ra giá trị dự báo sơ bộ, còn mô hình ARIMA(p,d,q) cho phép hiệu chỉnh giá trị dự báo. Đề xuất này xuất phát từ bản chất các yếu tố thủy văn vừa mang tính ngẫu nhiên vừa mang tính tất định.

Mô hình HEC-HMS ở đây được áp dụng như một mô hình bán phân bố vì trong mô hình có các modul có thể sử dụng được mưa đầu vào dưới dạng lưới, thêm vào đó mô hình còn sử dụng các đầu vào dưới dạng ô lưới đó là mô hình số hoá độ cao DEM với độ phân giải 90m, các bản đồ sử dụng đất, thảm phủ thực vật dưới dạng ô lưới để từ đó xây dựng bản đồ chỉ số CN dưới dạng ô lưới – là một thông số trong mô hình tính toán tổn thất theo phương pháp SCS.

Mô hình mạng Nơ ron nhân tạo thuyết tiến hóa (EANN – Evolution Artificial Neural Network) được áp dụng ở đây cũng là một trong những dạng lai ghép giữa mạng quen thuộc BPNN được áp dụng nhiều trong thủy văn với thuật toán giải đoán gen GA (Genetic Algorithms).

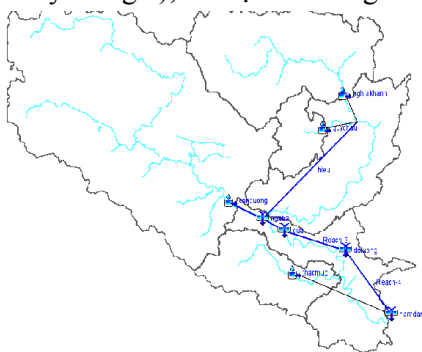
Mạng nơ ron lan truyền ngược (BPNN) là mạng có cấu trúc nhiều lớp trong đó có 1 lớp đầu vào, 1 lớp đầu ra và nhiều hơn 1 lớp ẩn. Trong các lớp có nhiều nút và chúng được liên kết với nhau bởi các trọng số. Với cấu trúc như thế này, chuỗi đầu ra tính toán sẽ được so sánh với chuỗi số liệu thực đo và được đánh giá thông qua các chỉ tiêu về sai số. Nếu không thoả mãn, các thông số của mô hình (thông số học và thông số mô men) sẽ tác động lên các trọng số liên kết và thay đổi chúng. Quá trình thay đổi này sẽ dừng lại khi nào các chỉ tiêu đánh giá được thoả mãn. Người ta gọi loại mạng nơ ron này là với thuật toán lan truyền ngược. Với thuật toán lan truyền ngược này, người sử dụng phải lựa chọn trước một cấu trúc mạng bao gồm số lớp và số nút trong từng lớp và cố gắng tìm ra bộ trọng số phù hợp cho cấu trúc này, còn nếu không tìm được sẽ phải thay đổi (lựa chọn lại) cấu trúc ban đầu (Chi tiết về BPNN có thể tham khảo trong Giáo trình Mô hình Toán Thủy văn - giảng dạy cao học, ĐHTL [2]). Nếu mạng có nhiều nút (nhiều biến đầu vào) thì quá trình này sẽ rất phức tạp và mất thời gian. Vì vậy nghiên cứu đã tích hợp vào mạng

BPNN thuật toán GA.

GA là thuật toán giải đoán gen, nó là một dạng của thuyết tiến hoá, nó chuyển quá trình tìm kiếm tối ưu thành một quá trình tiến hoá thông qua việc mã hoá các thông số của các hàm tối ưu thành các gen, là tập hợp của một số nơ ron. Quá trình tối ưu hoá được thực hiện thông qua việc dần dần chọn ra các gen trội, các gen đáp ứng được các chỉ tiêu đề ra, và dần dần qua từng thế hệ sẽ tìm được những gen trội nhất, hay cho ta một cấu trúc mạng phù hợp nhất.

Mục đích của việc tích hợp BPNN với GA là để khắc phục nhược điểm khó lựa chọn được mạng tối ưu khi số lượng các đầu vào là lớn (kích thước mạng lớn). Như vậy khi tích hợp vào, mô hình lai ghép sẽ có khả năng vừa thay đổi cấu trúc mạng tự động, vừa thay đổi trọng số kết nối giữa các phần tử của mạng để tìm ra một cấu trúc mạng, và trọng số kết nối phù hợp cho kết quả tính toán từ mô hình phù hợp nhất với số liệu thực đo. Đồng thời việc tích hợp này cho phép cập nhật số liệu liên tục, tạo điều kiện rất thuận lợi cho dự báo tác nghiệp sau này.

Nghiên cứu đã sử dụng phần mềm MiniTab 15, SPSS 11, và NeroSolution 5 vào ứng dụng nghiên cứu. MiniTab 15 và SPSS 11 là phần mềm phân tích thống kê chuyên dùng trong kỹ thuật và phân tích tài chính. Hai phần mềm này được sử dụng để lọc dữ liệu và để hỗ trợ tìm thông số của mô hình thống kê ARIMA (p,d,q). Phần mềm Neuro Solution 5 do tập đoàn Neuro Dimension Inc. phát triển; đây là phần mềm hoàn chỉnh hơn nhiều so với WinNN32 vì nó không chỉ phát triển theo hướng nơ ron thần kinh với thuật toán lan truyền ngược (BPNN) mà nó còn phát triển cả theo hướng mạng Fuzzy (Fuzzy Logic), thuật toán giải đoán Gen



(Genetic Algorithms - GA) và sử dụng nhiều thuật toán tối ưu khác. Phần mềm này tồn tại dưới 2 dạng: dạng thứ nhất như một phần mềm độc lập và dạng thứ hai là phần mở rộng được tích hợp trong Microsoft Excel (chỉ có từ phiên bản 4.0 trở lên). Với phần mềm này chúng tôi đã sử dụng thuật toán tối ưu giải đoán gen GA để hỗ trợ cho việc tìm ra mạng BPNN tốt nhất và vì vậy rút ngắn được thời gian chạy chương trình rất nhiều so với dùng Win NN32. Thêm vào đó với phần mở rộng được tích hợp trong Excel sẽ rất thuận tiện cho người dùng trong việc xử lý dữ liệu trước khi vào tính toán. Ngoài ra khác với WinNN32 việc sử dụng mạng tốt nhất trong dự báo cũng dễ dàng hơn nhiều với chức năng dự báo (Apply Production Dataset). Ngoài ra nghiên cứu còn sử dụng các phần mềm GIS như ArcGIS 9.3 và HEC-geo-HMS (một mô đun mở rộng của ArcView GIS) vào xây dựng hệ thống sông, lưu vực, các bản đồ chỉ số CN làm các thông số đầu vào cho mô hình thủy văn bán phân bố HEC-HMS.

3. KẾT QUẢ VÀ THẢO LUẬN

Nghiên cứu đã tiến hành áp dụng thử nghiệm từng mô hình (HEC-HMS, EANN, ARIMA) riêng rẽ rồi so sánh với việc thử nghiệm kết hợp các cặp mô hình nói trên (HEC-HMS + ARIMA và EANN + ARIMA). Kết quả đạt được cho thấy việc sử dụng lai ghép giữa mô hình tất định và ngẫu nhiên luôn cho kết quả tốt hơn việc sử dụng riêng rẽ từng mô hình, trong đó các mô hình tất định được sử dụng để đưa ra trị số dự báo sơ bộ, còn mô hình ngẫu nhiên được sử dụng để đưa ra giá trị hiệu chỉnh sai số dự báo. Các bảng 2, 3, 4 dưới đây trình bày kết quả áp dụng thử nghiệm mô hình lai ghép HEC-HMS với ARIMA.

Bảng 2: Bộ thông số của mô hình HEC-HMS cho các lưu vực bộ phận

Lưu vực	t_p (hr)	C_p	I_a (mm)	CN	S_{kt} (%)	Q_0 (m ³ /s)	RC	T_0 (m ³ /s)
Con Cuông	18	0,60	18	75	0	633	0,78	1340
Nghĩa Khánh	19	0,68	14	78	0	183	0,59	967
Quý Châu	6	0,59	14	79	0	93	0,59	450
Thác Muối	6	0,65	12	81	0	206	0,86	490

Hình 3: Sơ đồ lưu vực sông Cả trong HEC-HMS

Bảng 3: Các chỉ tiêu đánh giá kết quả dự báo thử nghiệm với thời gian $T_{dk} = 3$ ngày

	Dừa		Đô Lương		Nam Đàn	
	2005	2006	2005	2006	2005	2006
NASH	77	74	68	76	75	76
$\Delta Q(\%)$	15,4	20,1	20,4	21,2	20,3	24,0
σ'/σ	0,48	0,51	0,57	0,49	0,50	0,49
η	88	86	81	87	86	87
P(%)	80	77	71	74	76	73

P: mức bảo đảm phương án

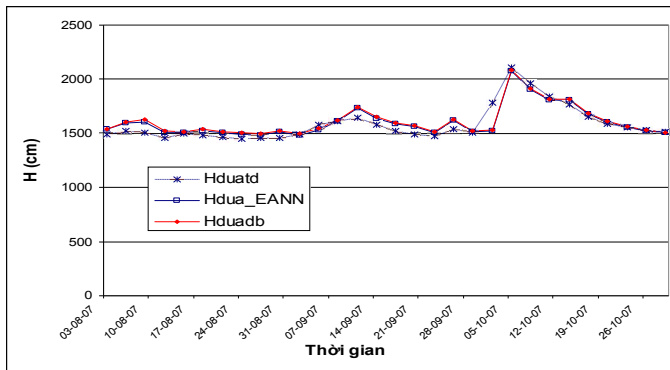
Kết quả đánh giá trong Bảng 3 và Bảng 4 cho thấy mô hình lai ghép HEC-HMS + ARIMA sử dụng mưa thực đo cho kết quả khá tốt khi dự báo 3 ngày nhưng không tốt khi dự báo 5 ngày. Tuy nhiên việc áp dụng mô hình lai ghép này trên thực tế còn chưa khả thi vì khi dự báo thì cần phải dùng mưa dự báo mà dự báo mưa ở nước ta hiện nay về lượng có độ chính xác chưa cao. Đây cũng chính là nguyên nhân mà rất nhiều nhà nghiên cứu trong và ngoài nước khi áp dụng kết hợp các mô hình thủy văn phân bố với các mô hình dự báo thời tiết

Bảng 4: Các chỉ tiêu đánh giá kết quả dự báo thử nghiệm với thời gian $T_{dk} = 5$ ngày

	Dừa		Đô Lương		Nam Đàn	
	2005	2006	2005	2006	2005	2006
NASH	77	58	61	75	69	60
$\Delta Q(\%)$	18,1	30,9	16,7	29,2	19,1	22,4
σ'/σ	0,47	0,65	0,62	0,49	0,55	0,63
η	87	76	77	87	83	77
P(%)	74	62	58	70	71	60

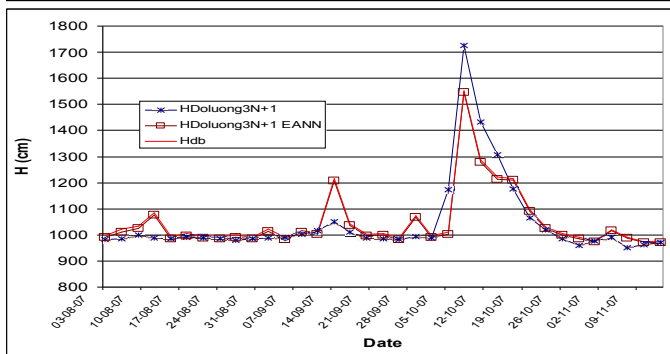
gặp phải. Vấn đề là ở chỗ làm thế nào nâng cao chất lượng dự báo mưa.

Việc thử nghiệm mô hình lai ghép giữa EANN + ARIMA cho kết quả dự báo khả thi hơn nhiều (xem hình 4), đều cho mức đảm bảo phương án khá cao (trên 85%) khi dự báo với thời gian dự kiến là 3 ngày. Đặc biệt phương pháp này có thể áp dụng vào trong thực tế vì các biến đưa vào đều là các số liệu thực đo đã có tại thời điểm dự báo mà không cần phải đưa vào mưa dự báo.



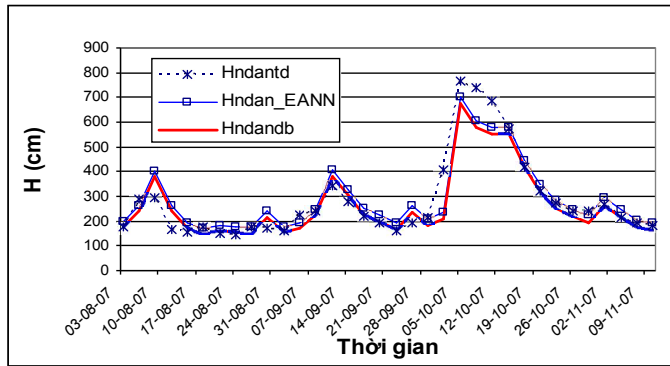
(a)

NASH	85
$\Delta Q(\%)$	15,1
σ'/σ	0,49
η	89



(b)

NASH	88
$\Delta Q(\%)$	14,7
σ'/σ	0,51
η	90



(c)

NASH	90
$\Delta Q(\%)$	17,
σ'/σ	0,5
η	91

Hình 4: Đường quá trình mực nước thực đo và dự báo tính toán với thời gian dự kiến 3 ngày bằng mô hình EANN-ARIMA (p,d,q) cho trận lũ năm 2007 tại trạm Dừa(a), Đô Lương(b), Nam Đàn(c)

4. KẾT LUẬN

Nghiên cứu dự báo lũ trung hạn lưu vực sông Cả đã tiến hành nghiên cứu tổng quan các phương pháp dự báo lũ trung hạn tiên tiến ở trong và ngoài nước từ đó lựa chọn hướng tiếp cận nghiên cứu phù hợp với xu thế phát triển hiện nay cho lưu vực sông Cả. Nghiên cứu đã lựa chọn các mô hình lai ghép giữa tất định và ngẫu nhiên như HEC-HMS lai ghép với ARIMA(p,d,q) và EANN (mô hình mạng nơ ron

theo thuyết tiến hóa) lai ghép với ARIMA (p,d,q) để xây dựng phương án dự báo lũ trung hạn phù hợp cho lưu vực sông Cả. Việc lai ghép các mô hình trên cho phép tận dụng được những ưu điểm của các phương pháp dự báo trung hạn truyền thống với các phương pháp hiện đại. Kết quả áp dụng cho lưu vực sông Cả là khá tốt và hoàn toàn có thể mở rộng để áp dụng cho các lưu vực sông khác ở Miền Trung

TÀI LIỆU THAM KHẢO

1. Hoàng Thanh Tùng, và nnk (8-2009). Chuyên đề nghiên cứu dự báo lũ trung hạn lưu vực Sông Cả - *Chuyên đề NCS*. Trường Đại học Thủy lợi.
2. Lê Văn Nghinh, Bùi Công Quang, Hoàng Thanh Tùng (6-2006). *Mô hình Toán Thủy văn – Giáo trình cho học viên cao học*. Trường Đại học Thủy lợi.

Abstract:

STUDY ON MID-TERM FLOOD FORECAST FOR THE CA RIVER BASIN

**Hoang Thanh Tung, Vu Minh Cat
Robeto Ranzi, Truong Tung Hoa**

Mid-term flood forecast for river basins in Central Provinces in general and for the Ca River Basin in particular has become an urgent matter. By overview studying on mid-term flood forecast in Vietnam and in the World, this research has selected the integrated hybrid modeling scheme between deterministic and stochastic models such as HEC-HMS with ARIMA(p,d,q) and EANN (Evolutional Artificial Neural Network) with ARIMA(p,d,q) models to build a suitable mid-term flood forecast for the Ca River. The integrated hybrid modeling scheme allows us to take advantages of both traditional and new modern models. Application results for the Ca River is very good, thus the integrated hybrid modeling scheme can be extendedly applied to other river basins in Central Provinces.