

HỘI CƠ HỌC VIỆT NAM

**Tuyển tập công trình khoa học
Hội nghị Cơ học toàn quốc lần thứ X**

Hà Nội, 08-09/12/2017

**Tập 2. Cơ học Thủy khí
Cơ học Tính toán
Cơ học Đất Đá và Môi trường rời**



NHÀ XUẤT BẢN KHOA HỌC TỰ NHIÊN VÀ CÔNG NGHỆ

MỤC LỤC

PHẦN I. CƠ HỌC THỦY KHÍ	[1]
Nguyễn Hải Anh	
Thuật toán lưới thích nghi cục bộ cho phương pháp phân tử xoáy trong tính toán dòng chảy rời	2
Nguyễn Quang Bình, Nguyễn Thế Hùng	
So sánh sơ đồ số trong tính toán dòng chảy qua hệ thống kè mò hàn.....	13
Nguyễn Chí Công, Lương Ngọc Lợi và Ngô Văn Hệ	
Nghiên cứu ảnh hưởng của tỷ số chiều dày tương đối đến đặc tính thủy động học biên dạng cánh NACA 66 sử dụng trong cánh chân vịt tàu thủy....	21
Phạm Thành Đồng, Đặng Ngọc Thanh, Phạm Vũ Uy	
Nghiên cứu khảo sát ảnh hưởng của tính nhót và tính nén môi trường đến đặc tính khí động của cánh quay trực thăng bằng phương pháp xoáy rời rạc	28
Nguyễn Trọng Đức, Bùi Đình Trí, Nguyễn Văn Thắng	
Nghiên cứu ảnh hưởng của dòng gió xung quanh tòa nhà cao tầng bằng phương pháp thủy động lực học số (CFD)	36
Tran Thu Ha, Nguyen Anh Son and Nguyen Van Tung	
Simulation correction in super cavity model with application of data assimilation	44
Dương Ngọc Hải, Nguyễn Quang Thái, A. A. Gubaidullin,	
Nguyễn Tất Thắng và N. G. Musakaev	
OpenFOAM và ứng dụng trong tính toán thủy động lực học dòng chảy có chuyển pha.....	53
Ngô Văn Hiền	
Mô hình điều khiển phối hợp cho AUS/M-AUVs bằng Automate lai và RealTime UML/MARTE.....	61
Ngô Văn Hiền, Phạm Gia Điem	
Mô hình phân tích, thiết kế và thực thi hệ thống điều khiển cho AUV/ASV với chuẩn MBSE/SysML	70
Dang Trung Hieu and Vu Cong Hoa	
Simulation of an Airfoil Flutterby Coupling Fluid Structure Depend on CFD Solver	81
Vũ Mạnh Hiếu, Nguyễn Trung Định, Lê Quốc Định	
Xác định đặc tính khí động của khí cầu	91
Nguyễn Ngọc Hòa	
Bài toán truyền sóng trong môi trường đàn hồi điền đầy chất lỏng bão hòa	98
Bùi Minh Hoàng, Nguyễn Thế Mịch, Nguyễn Thế Đức, Nguyễn Duy Chính	
Mô phỏng số cho dòng hai pha rắn - lỏng trong bơm ly tâm bằng mô hình DDPM.....	107
Phạm Chí Hùng, Đinh Thanh Phong, Phạm Huy Diệu,	
Phan Hoàng Phụng, Nguyễn Tiến Đạt và Nguyễn Tường Long	
Nghiên cứu dòng khí và hạt mài thổi qua vòi phun Laval	114
Vo Duy Hung	
An investigation of wind flow pattern across staycables at high wind speeds.....	123
Trần Đình Khôi, Trần Đức Cường	
Đặc tính chuyển động dọc của máy bay S-30 khi trang bị động cơ có vec tơ lực đẩy thay đổi	131
Nguyễn Chính Kiên	
Thử nghiệm ứng dụng mạng neuron nhân tạo trong dự báo thủy văn và thủy lực	136

Thử nghiệm ứng dụng mạng nơron nhân tạo trong dự báo thủy văn và thủy lực

Nguyễn Chính Kiên

Viện Cơ học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Email: nckien@imech.vast.vn

Tóm tắt: Bài báo trình bày những nghiên cứu bước đầu về mạng nơron nhân tạo và áp dụng trong bài toán dự báo thủy văn, thủy lực. Một chương trình được xây dựng bằng ngôn ngữ Fortran có kết hợp nhiều giải thuật (giải thuật di truyền, giải thuật lan truyền ngược sai số, thay đổi tốc độ học mang,...). Chương trình sử dụng phép biến đổi wavelet để chuẩn hóa dữ liệu đầu vào cũng như sử dụng các kỹ thuật tính toán song song OpenMP và GPU tăng tốc độ nhanh hơn hàng chục lần. Chương trình đã chạy thử nghiệm cho một số bài toán mẫu và các ví dụ cụ thể, cũng như sẵn sàng cho việc tác nghiệp dự báo.

Từ khóa: Mạng nơron nhân tạo, Dự báo, Thủy văn, Thủy lực.

1. Mở đầu

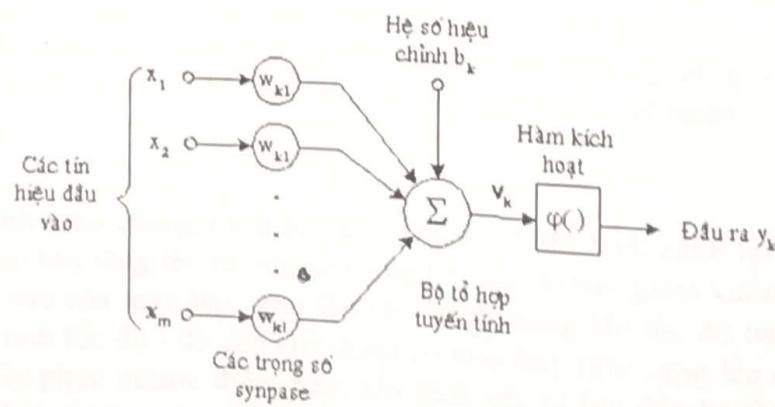
Một hướng tiếp cận mới để mô hình hoá các hiện tượng thủy văn, thủy lực là dựa trên các công nghệ học máy. Mô hình này dựa trên một cơ sở dữ liệu thực tế đủ lớn và áp dụng kỹ thuật ứng dụng mạng thần kinh nhân tạo đã và đang là một trong những lĩnh vực nghiên cứu phát triển mạnh mẽ. Gần đây, nhờ sự phát triển của công nghệ phần cứng cũng như dữ liệu thu thập, đã thúc đẩy việc nghiên cứu và đạt được các bước tiến mới trong lĩnh vực này. Ngoài ra, với ưu thế đưa ra các kết quả tính nhanh và tính được các trường hợp thiếu dữ liệu đầu vào, bài báo đã thử nghiệm áp dụng các lý thuyết mới (mạng hồi tiếp) cũng như các kỹ thuật tính toán song song tận dụng tối đa các lõi tính của CPU và GPU để xây dựng một chương trình tính cho một số bài toán dự báo thủy văn và thủy lực.

2. Mạng thần kinh nhân tạo và các thuật toán

2.1. Giới thiệu mạng nơron nhân tạo

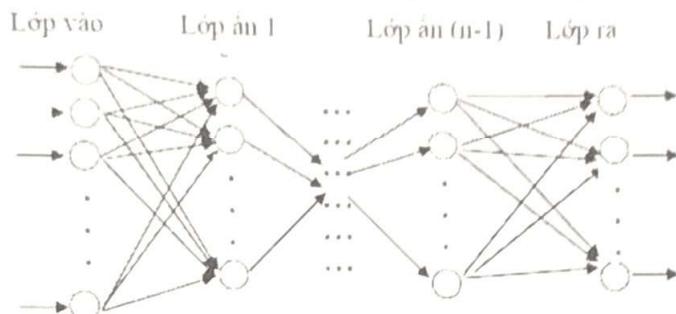
2.1.1. Khái niệm, mô hình [5]

Tương tự như cấu trúc nơron sinh học, nơron nhân tạo bao gồm các thành phần: các đầu vào (như các dây thần kinh vào) là các tín hiệu vào dưới dạng vec tơ m chiều; các liên kết biểu diễn bởi ma trận trọng số; hàm tổng là tổ hợp tuyến tính các tín hiệu vào với các trọng số; hàm kích hoạt giới hạn phạm vi của đầu ra sau khi nhận đầu vào là hàm tổng và ngưỡng – độ lệch; đầu ra là tín hiệu đầu ra của mỗi nơron.



Hình 1. Nơron nhân tạo

Trong bộ não của con người, các tế bào nơron liên kết với nhau chằng chịt và tạo nên một mạng lưới vô cùng phức tạp, tuy nhiên mạng nơron nhân tạo được chia thành các loại chính sau: Mạng có phản hồi (feedback network); Mạng nơron truyền thẳng một lớp (perceptron); Mạng có phản hồi (feedback network); Mạng truyền thẳng nhiều lớp; Mạng hồi quy,...[6]



Hình 2. Mạng MLP tổng quát

Một mạng nơron được huấn luyện sao cho với một tập các véc tơ đầu vào X , mạng có khả năng tạo ra tập các véc tơ đầu ra mong muốn Y của nó. Tập X được sử dụng cho huấn luyện mạng được gọi là tập huấn luyện (training set). Các phần tử x thuộc X được gọi là các mẫu huấn luyện (training example). Quá trình huấn luyện bản chất là sự thay đổi các trọng số liên kết của mạng. Trong quá trình này, các trọng số của mạng sẽ hội tụ dần tới các giá trị sao cho với mỗi véc tơ đầu vào x từ tập huấn luyện, mạng sẽ cho ra véc tơ đầu ra y như mong muốn. Có ba phương pháp học phổ biến là học có giám sát (supervised learning), học không giám sát (unsupervised learning) và học tăng cường (reinforcement learning). Sau khi học xong, nó có thể mô tả gần đúng mối quan hệ tương quan phức tạp giữa các yếu tố đầu vào và đầu ra của các quá trình cần nghiên cứu và khi đã học được thì việc kiểm tra độc lập thường cho kết quả tốt. Khi đó, mạng nơron nhân tạo có thể tính toán kết quả đầu ra tương ứng với bộ số liệu đầu vào mới.

2.1.2. Mạng nơron truyền thẳng nhiều lớp [1]

Mạng perceptron nhiều lớp (Multilayer Perceptron – MLP) còn được gọi là mạng truyền thẳng nhiều lớp là sự mở rộng của mô hình mạng perceptron với sự bổ sung thêm những lớp ẩn và các nơron trong các lớp ẩn này có hàm truyền (hàm kích hoạt) dạng phi tuyến. Mạng MLP có một lớp ẩn là mạng nơron nhân tạo được sử dụng phổ biến nhất, nó có thể xấp xỉ các hàm liên tục được định nghĩa trên một miền có giới hạn cũng như những hàm là tập hợp hữu hạn của các điểm rời rạc [4].

Cấu trúc của một mạng MLP tổng quát có thể mô tả như sau:

- Đầu vào là các véc tơ (x_1, x_2, \dots, x_p) trong không gian p chiều, đầu ra là các véc tơ (y_1, y_2, \dots, y_q) trong không gian q chiều. Đối với các bài toán phân loại, p chính là kích thước của mẫu đầu vào, q chính là số lớp cần phân loại.

- Mỗi nơron thuộc lớp sau liên kết với tất cả các nơron thuộc lớp liền trước nó.

- Đầu ra của nơron lớp trước là đầu vào của nơron thuộc lớp liền sau nó.

Hoạt động của mạng MLP như sau: tại lớp đầu vào các nơron nhận tín hiệu vào xử lý (tính tổng trọng số, gửi tới hàm truyền) rồi cho ra kết quả (là kết quả của hàm truyền); kết quả này sẽ được truyền tới các nơron thuộc lớp ẩn thứ nhất; các nơron tại đây tiếp nhận như là tín hiệu đầu vào, xử lý và gửi kết quả đến lớp ẩn thứ 2;...; quá trình tiếp tục cho đến khi các nơron thuộc lớp ra cho kết quả.

Trong bài báo này có nghiên cứu ứng dụng mạng hồi tiếp (recurrent neural network) là mô hình mạng được phát triển từ MLP.

2.2. Các thuật toán lan truyền ngược sai số

2.2.1. Thuật toán lan truyền ngược sai số (BP) [3]

Thuật toán sử dụng một tập các mẫu gồm các cặp đầu vào - đầu ra để luyện mạng. Với mỗi cặp đầu vào - đầu ra (x, d) thuật toán lan truyền ngược sai số thực hiện hai giai đoạn sau:

- Giai đoạn thứ nhất, mẫu đầu vào x được truyền từ lớp vào tới lớp ra và ta có kết quả đầu ra tính toán được là y .

- Giai đoạn tiếp theo, tín hiệu lỗi được tính toán từ sự khác nhau giữa đầu ra quan sát được d với đầu ra tính toán y sẽ được lan truyền ngược lại từ lớp ra đến các lớp trước để điều chỉnh các trọng số của mạng.

2.2.2. Thuật toán di truyền (GA) [2]

Các bước áp dụng thuật toán Di truyền vào việc học của mạng nơron như sau:

- Xây dựng hàm giá: Hàm giá này sẽ được sử dụng để tạo nên độ phù hợp của các cá thể và của cả quần thể trong GA. Trong nghiên cứu này sử dụng hàm sai số RMS tương tự như trong thuật toán Lan truyền ngược sai số.

- Mã hoá nhiễm sắc thể (NST): Mỗi cá thể trong GA sẽ thay mặt cho một bộ trọng số của mạng nơron.

- Thực hiện giải thuật Di truyền:

- + Chọn lọc: Gán 1 NST con bằng một NST có hàm giá nhỏ hơn của 2 NST bố mẹ ngẫu nhiên (hoặc là bố, hoặc là mẹ).
- + Lai ghép: Toán tử lai ghép này sẽ đưa một giá trị vào mỗi vị trí của NST con bằng cách lấy ngẫu nhiên một giá trị tại cùng vị trí của NST cha hoặc mẹ.
- + Đột biến: Một gen (trọng số) được lựa chọn ngẫu nhiên với một xác suất $p_{mutation}$ để tiến hành đột biến: cộng thêm một giá trị ngẫu nhiên quanh giá trị gốc ban đầu.

2.2.3. Kết hợp các thuật toán

Có nhiều cách để kết hợp thuật toán Di truyền và thuật toán Lan truyền ngược sai số nhưng cách đơn giản và khá hiệu quả là ta thực hiện lai ghép hai giải thuật nối tiếp nhau. Tập trọng số được mã hoá thành các nhiễm sắc thể và được tiến hoá nhờ thuật toán Di truyền. Kết thúc quá trình tiến hoá, bộ trọng số tốt nhất tương ứng với cá thể ưu việt nhất trong quần thể được lựa chọn làm những trọng số khởi tạo cho thuật toán Lan truyền ngược sai số.

Thử nghiệm bài toán XOR cho hai thuật toán riêng rẽ cũng như kết hợp với cùng một cấu trúc mạng và điều kiện dừng.

Bảng 1. So sánh khả năng hội tụ của mạng khi sử dụng hai phương pháp học GA và BP
với sai số dừng lặp khác nhau

Sai số dừng lặp	Số lần hội tụ trong 1.000.000 lần thử nghiệm		
	GA	BP	Kết hợp
0.05	83,75%	75,64%	
0.02	40,25%	56,75%	
0.001	0%	18,44%	60,45%

So sánh với việc sử dụng GA và BP riêng rẽ thì giải thuật kết hợp này cho kết quả tốt hơn rất nhiều.

3. Xây dựng chương trình

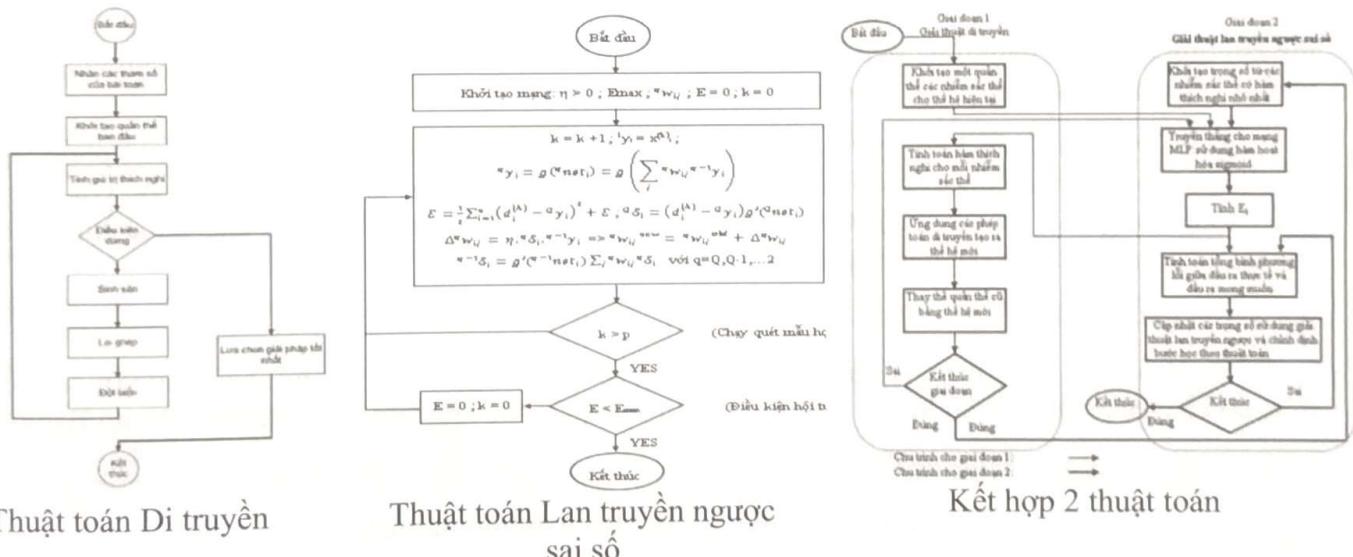
3.1. Giao diện và sơ đồ khối

Chương trình được xây dựng bằng ngôn ngữ Fortran bằng bộ công cụ lập trình Visual Fortran. Phiên bản được tối ưu trên nền hệ điều hành 64bit nhằm tận dụng hết sức mạnh của hệ điều hành khi làm việc với dữ liệu lớn. Chương trình cũng được xử lý song song nhằm nâng cao tốc độ tính toán, phục vụ các bài toán thực tế với kích thước dữ liệu lớn.



Hình 3. Giao diện của chương trình

Chương trình có thể tính độc lập cũng như kết hợp các thuật toán khác nhau. Tùy từng phương án tính toán người dùng lựa chọn mà các thuật toán theo các sơ đồ khối Hình 4 được áp dụng linh hoạt khác nhau. Nếu phương án tính có sử dụng giải thuật Lan truyền ngược sai số riêng rẽ hoặc kết hợp với thuật toán Di truyền hoặc đọc ma trận trọng số đã có để tính tiếp thì các modun này được sử dụng theo những cách khác nhau tương ứng với các đầu vào khác nhau.



Thuật toán Di truyền

Thuật toán Lan truyền ngược sai số

Kết hợp 2 thuật toán

3.2. Kỹ thuật song song

Để tạo điều kiện thuận lợi cho phát triển mã nguồn cũng như sử dụng, chúng tôi thực hiện 2 giải pháp song song cho phần mềm của mình là kỹ thuật OpenMP và kỹ thuật dùng card đồ họa GPU của Nvidia trên nền CUDA. Sự tiện lợi của hai giải pháp này là người lập trình xây dựng không cần phải có kiến thức về cài đặt, cấu hình chính xác mạng máy tính mà chỉ đóng gói kèm thư viện động hoặc cần cài các driver mặc định của nhà sản xuất phần cứng là sử dụng được. Khi được chuyển giao, người sử dụng cũng dễ dàng chạy trên các máy tính cá nhân nhanh chóng mà không cần kiến thức chuyên sâu về song song.

3.2.1. Kỹ thuật song song OpenMP

Kỹ thuật này được ứng dụng cho mọi máy tính phổ thông CPU đa lõi hiện có trên thị trường. Trong thử nghiệm này, chúng tôi sử dụng CPU Intel i7 3770 4 lõi 8 thread với tốc độ 3.9 Ghz. Đối với cả hai thuật toán, trong vòng lặp điều kiện dừng, thao tác huấn luyện mạng được chạy cho nhiều mẫu học. Mỗi mẫu học này được chương trình phân bổ trên các lõi CPU. Do đó, sau dòng lệnh vòng lặp điều kiện dừng và trước dòng lệnh vòng lặp huấn luyện theo các mẫu, bấy song song được đặt để chia nhỏ công việc cho các CPU.

3.2.2. Kỹ thuật song song GPU

Trong nghiên cứu này sử dụng thiết bị là 01 card NVIDIA GeForce GTX 1060 6GB với thông số được trình bày như Bảng 2. Đây là một cấu hình khuyến cáo tối thiểu của nhiều chuyên gia dành cho một hệ thống trí tuệ nhân tạo phục vụ công việc nghiên cứu, ứng dụng cơ bản,... cũng như tham gia sàn đấu kháng Kaggle dành cho các chương trình học máy.

Bảng 2. Thông số card NVIDIA GeForce GTX 1060 6GB

Thông số	Chi tiết
Tên mã thiết bị	GP106 (8/2016)
Số lõi tính / Xung nhịp tính	1280 / 1708 Mhz
Bộ nhớ	6144 MB / GDDR5 / 192 bit / 192.2 GB/s
Băng thông giao tiếp	PCIe 3.0 x16 / 16Gbps
Năng lực tính toán dấu phẩy động	4.375 TFLOPS

Đối với thuật toán Di truyền, chúng tôi chọn phương án song song dựa trên số lượng quần thể, mỗi cá thể trong quần thể sẽ được 1 đơn vị tính toán xử lý. Tuy nhiên, chỉ xử lý song song với các cá thể khi tính toán các hàm mục tiêu (hàm thích nghi), còn bỏ qua phần xử lý các phép toán di truyền chỉ chiếm một khối lượng nhỏ tính toán, chủ yếu là các phép gán giá trị mảng.

Đối với thuật toán Lan truyền ngược sai số, do chỉ còn 1 cá thể tối ưu nhất, các phép lặp chỉ thực hiện theo các mẫu học, do đó cần chia mỗi lần xử lý 1 mẫu học cho 1 đơn vị xử lý. Ưu điểm của phần song song này so với thuật toán Di truyền là song song được phần lớn các bước tính, dữ liệu truyền về bộ nhớ chung sau mỗi bước tính là nhỏ.

3.2.3. Kết quả thu được

Để đánh giá hiệu quả của phương pháp song song, chúng tôi lựa chọn 2 bài toán với các số liệu thực có cấu trúc mạng và số lượng mẫu học, số lượng quần thể khác nhau được mô tả trong Bảng 3.

Bảng 3. Thông số các bài toán thử nghiệm.

STT	Thông số	Bài toán 1	Bài toán 2
1	Số nơron lớp vào: lớp ẩn: lớp ra		
2	Số lượng trong quần thể di truyền	8:13:1	49:100:1
3	Số bước lặp di truyền	1.000	2.000
4	Số bước lặp phương pháp Lan truyền ngược sai số	100	50
5	Số lượng mẫu học	100.000	100.000
		480	1.000

Mỗi bài toán được chúng tôi chạy 100 lần để tính thời gian trung bình của mỗi phương án với sai số dừng nhỏ (mỗi phương án chạy hết số bước lặp), kết quả được thể hiện ở Bảng 4. dưới đây.

Bảng 4. Thời gian tính trung bình của các bài toán thử nghiệm

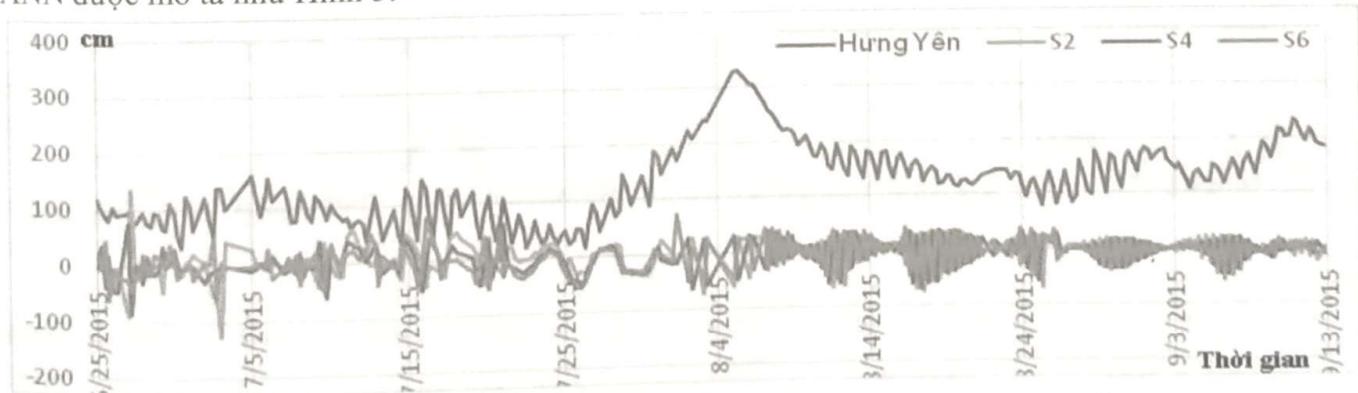
Phương án tính	Tính đơn nhân CPU	Tính đa nhân CPU (OpenMP)	Tính đa nhân GPU	Tỉ lệ thời gian
Bài toán 1	52s	19s	90s	1:0.36:1.73
Bài toán 2	1h:22m:53s (4973s)	0h:20m:24s (1224s)	0h:04m:34s (274s)	1:0.25:0.055

Nhận thấy: Khi kích thước các ma trận là nhỏ, số lượng mẫu học nhỏ, số lượng quần thể là nhỏ, việc song song trên GPU lại không mang lại hiệu quả, thậm chí còn gây ra sự chậm trễ tổng thể tính toán do phải truyền dữ liệu nhiều; tính đa lõi CPU bằng OpenMP đạt hiệu quả tốt nhưng cũng chưa đạt hết mức độ thiết kế. Khi số lượng nơron các lớp tăng, số lượng mẫu học tăng, số lượng quần thể tăng, việc tính toán song song cho bài toán mạng nơron nhân tạo càng hiệu quả do tính song song của các bài toán này lớn. Ngoài ra, phân tích biểu đồ đo hiệu năng card đồ họa cho thấy, kỹ thuật song song áp dụng chưa tận dụng hết 1280 lõi của card đồ họa nên chưa đạt hiệu quả cao nhất (chi đạt lớn nhất là 75%). Thuật toán Di truyền chưa được song song các phép toán di truyền, khối lượng truyền dữ liệu lại lớn, nên xảy ra hiện tượng ngắt quãng thời gian xử lý, cần phải tiếp tục cải tiến việc song song cho thuật toán này thêm nữa.

3.3. Kỹ thuật phân tích dữ liệu đầu vào

Một lý do chính trong việc nghiên cứu và sử dụng phép biến đổi wavelet là có khả năng thực hiện phân tích cục bộ, thể hiện được đặc tính của dữ liệu mà các kỹ thuật phân tích khác không có. Wavelet là dạng sóng có thời gian duy trì tới hạn với giá trị trung bình bằng không. Wavelet có thời gian giới hạn, bắt thường và bắt đổi xứng. Phương pháp wavelet có thể kết hợp với mô hình ANN cho các bài toán liên quan đến việc phân tích chuỗi số liệu rời rạc theo thời gian. Trước hết chuỗi số liệu sẽ được chuyển đổi thành bộ số liệu mới với các thành phần xấp xỉ và thành phần chi tiết dựa vào phép biến đổi wavelet rời rạc tại một tầng thích hợp nhất. Các thành phần này sau đó sẽ được sử dụng như là đầu vào cho mạng ANN.

Thử nghiệm với chuỗi số liệu mực nước đo được tại trạm thủy văn Hưng Yên mùa lũ năm 2015, sử dụng họ wavelet Daubechies phân tích 3 bậc thành các chuỗi số liệu làm tín hiệu đầu vào cho mạng ANN được mô tả như Hình 5.



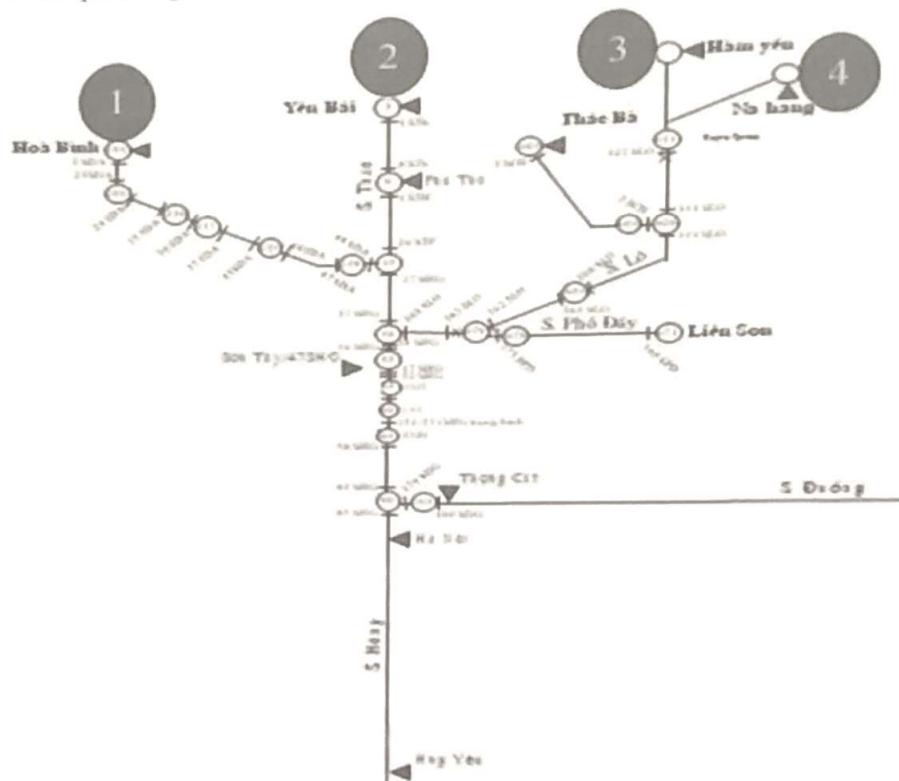
Hình 5. Phân tích chuỗi dữ liệu mực nước (cm) trạm Hưng Yên theo chuỗi thời gian

4. Thí nghiệm ứng dụng tính toán

4.1. Kết quả mô phỏng và dự báo thủy lực lưu vực đồng bằng Sông Hồng

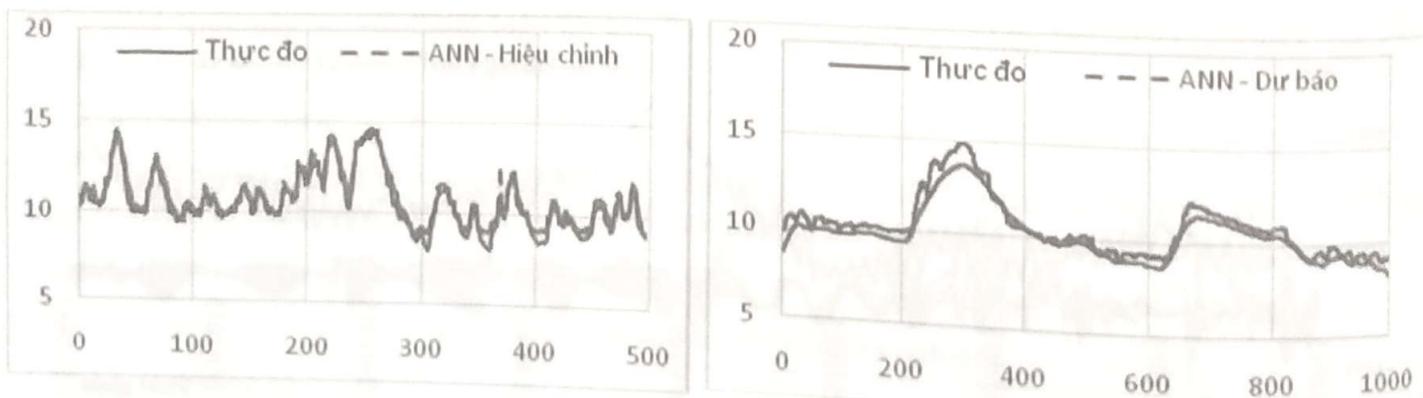
Hàng năm, từ 15 tháng 6 đến 15 tháng 9 cư dân vùng châu thổ Sông Hồng - Thái Bình luôn sống trong tình trạng bị uy hiếp bởi nạn lũ lụt. Để chủ động phòng tránh và hạn chế các tác hại, song song với việc phát triển kinh tế, chúng ta phải luôn luôn tìm kiếm các biện pháp hữu hiệu đối phó với lũ lụt. Tập thể nghiên cứu lũ lụt Viện Cơ học đã xây dựng và phát triển nhiều mô hình thủy lực nhằm mô phỏng và dự báo các yếu tố thủy văn thủy lực, từ đó phục vụ cho việc đề xuất, đánh giá và điều

hành phòng chống lụt bão. Trong nghiên cứu này sử dụng lại một phần số liệu được chọn lọc [7] của mô hình thủy lực đã phát triển để tính toán, thử “dự báo” lại mực nước và lưu lượng tại một số trạm trong hệ thống theo các phương án sau đây.



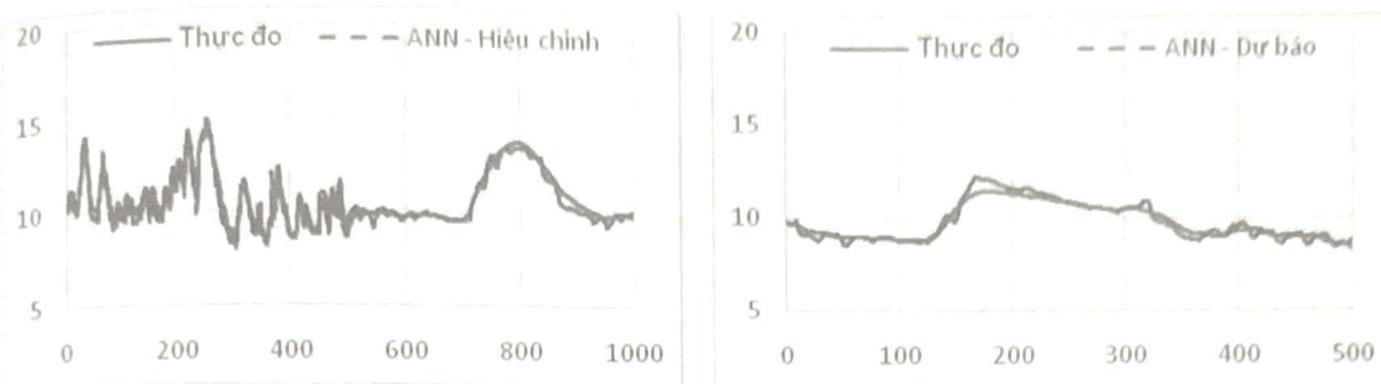
Hình 6. Sơ đồ mạng sông vùng nghiên cứu

Phương án 1: Lấy dữ liệu 4 nguồn tại Hòa Bình, Yên Bái, Hàm Yên, Na Hang các năm 2000, 2002, 2003 làm thành tập huấn luyện, dữ liệu năm 2004 dùng để làm tập kiểm định “dự báo lại”.



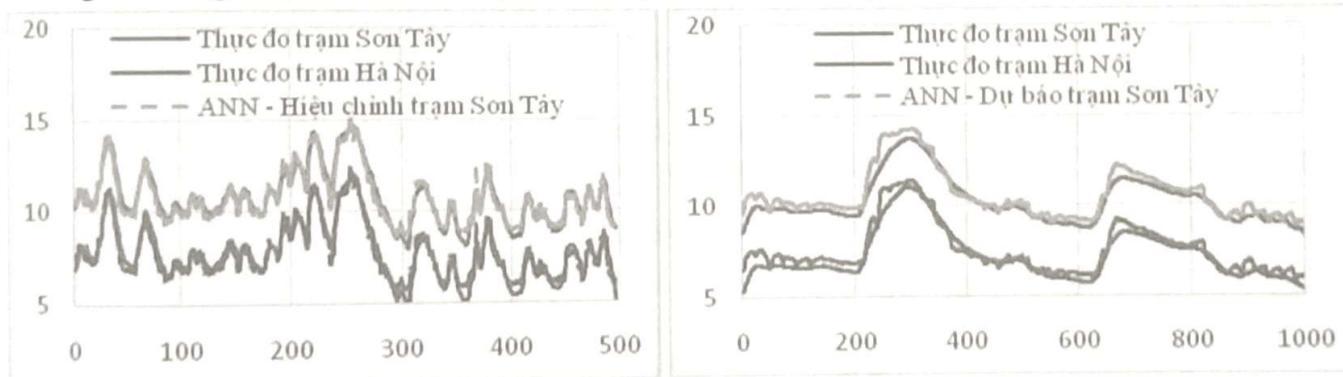
Nhận xét: với chỉ số đánh giá Nash-Sutcliffe (NSE) của phương án hiệu chỉnh và dự báo lần lượt là 0.933 và 0.826 đều đạt loại tốt cho thấy khả năng dự báo của mô hình là hiệu quả.

Phương án 2: Ngoài dữ liệu các năm 2000, 2002, 2003, lấy thêm nửa tập dữ liệu năm 2004 làm thành tập huấn luyện, dữ liệu còn lại của năm 2004 dùng để làm tập kiểm định “dự báo lại”.



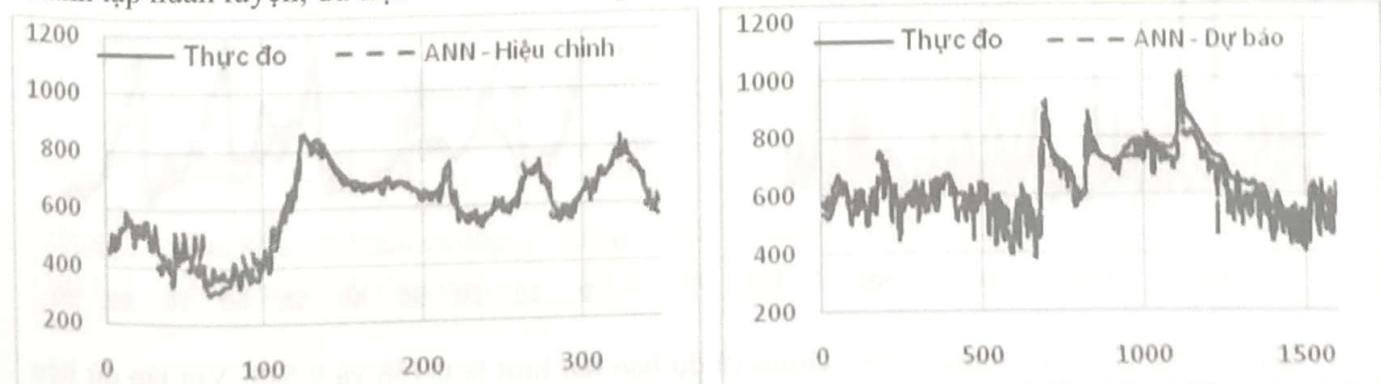
Nhận xét: NSE của phương án hiệu chỉnh và dự báo lần lượt là 0.922 và 0.909 cho thấy khi cập nhật thêm dữ liệu cho tập huấn luyện để “học thêm”, mạng ANN đã cho kết quả tốt hơn. Dữ liệu học thêm khá gần với dữ liệu dự báo nên phản ánh tốt hơn quan hệ giữa đầu vào và đầu ra của mạng, bổ sung các yếu tố có thể đã thay đổi (lòng dẫn, nhám, cấu trúc mạng sông,... của hệ thống sông năm 2004) mà các năm trước chưa có. Tuy nhiên vẫn có sự sai khác tại các đỉnh lũ, điều này lý giải bởi còn các yếu tố ảnh hưởng khác chưa được sử dụng làm tín hiệu đầu vào,... Do đó cần tiếp tục thêm vào mô hình các dữ liệu đầu vào khác cũng như tiếp tục cho chương trình học thêm các mẫu dữ liệu.

Phương án 3: Đầu vào của phương án này vẫn lấy như phương án 1, tuy nhiên kết quả đầu ra không chỉ là 1 giá trị mực nước ở trạm Sơn Tây mà còn thêm trạm Hà Nội (số nơron lớp ra là 2).



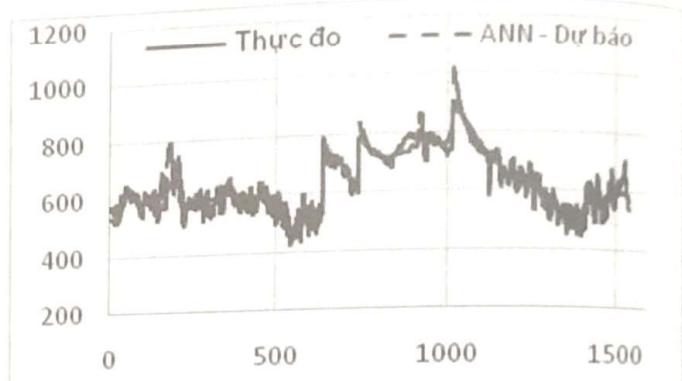
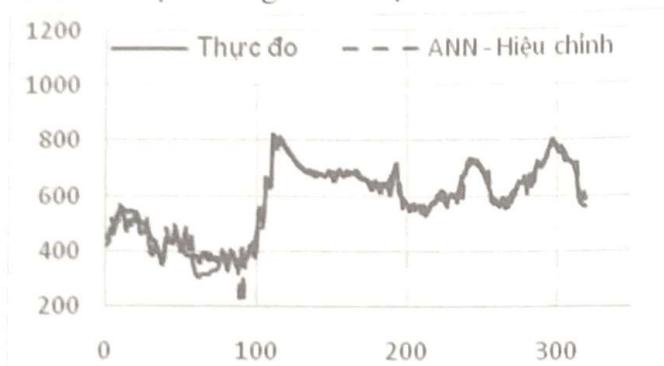
Nhận xét: Tại trạm Sơn Tây, NSE của phương án hiệu chỉnh và dự báo lần lượt là 0.942 và 0.864. Tại trạm Hà Nội, NSE của phương án hiệu chỉnh và dự báo lần lượt là 0.928 và 0.892. Qua phương án 3 thử nghiệm cho thấy, mạng ANN hoàn toàn có thể huấn luyện để cùng cho kết quả tốt với nhiều nơron lớp ra. Nhu cầu thực tế của bài toán dự báo đặt ra là cùng một lúc phải đưa ra được nhiều điểm trong mạng cũng như bức tranh đơn giản toàn bộ hệ thống tuy không đầy đủ như các mô hình số tính thủy văn, thủy lực.

Phương án 4: Lấy dữ liệu 4 nguồn tại Hòa Bình, Yên Bai, Hàm Yên, Na Hang năm 2015 làm thành tập huấn luyện, dữ liệu năm 2016 dùng để làm tập kiểm định “dự báo lại”.



Nhận xét: Tại trạm Sơn Tây, NSE của phương án hiệu chỉnh và dự báo lần lượt là 0.946 và 0.709. Năm 2015, mực nước Sơn Tây dưới 600cm là không nhiều, do đó khi mạng ANN học bộ dữ liệu này, mảng trọng số chưa thể hiện đúng quan hệ đầu vào – đầu ra khi mực nước thấp. Năm 2016, mực nước Sơn Tây chủ yếu lại nằm dưới mức 600cm, tuy nhiên, các kết quả “dự báo lại” vẫn nằm ở mức khá ($NSE > 0.7$). Thêm một lần nữa chứng tỏ, dù còn thiếu các yếu tố ảnh hưởng khác cũng như mẫu học còn ít nhưng mạng ANN có thể phản ánh được khá tốt quan hệ đầu vào – đầu ra của dữ liệu.

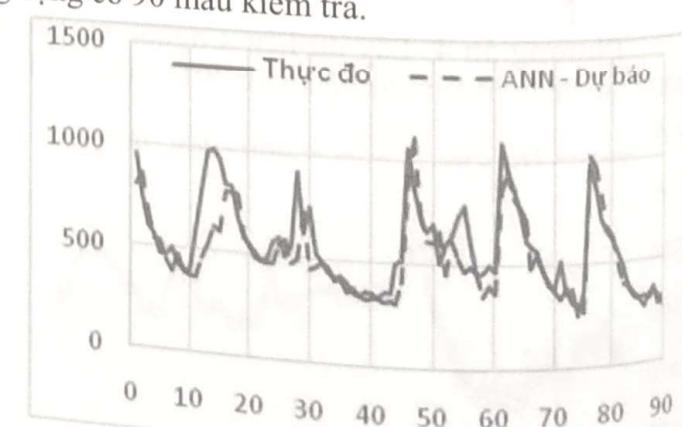
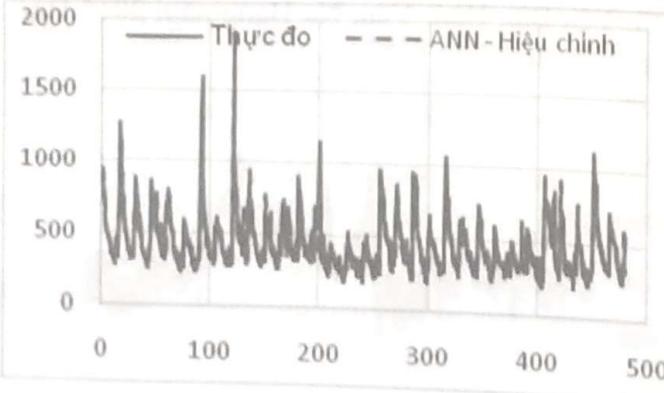
Phương án 5: Các số liệu đầu vào và ra tương tự như phương án 4 nhưng bổ sung thêm tín hiệu đầu vào trạm Hưng Yên ở hạ du.



Nhận xét: Tại trạm Sơn Tây, NSE của phương án hiệu chỉnh và dự báo lần lượt là 0.953 và 0.821. Đúng như nguyên tắc của mạng học máy, khi ta bổ sung thêm dữ liệu (tín hiệu đầu vào) có ảnh hưởng, mạng đã phản hồi với kết quả tốt hơn (các chỉ số NSE của các phép thử đều tốt hơn rõ rệt). Đây cũng chỉ ra một lợi thế của mô hình học máy: khi còn thiếu dữ liệu, hoặc dữ liệu không gian không đầy đủ, mạng ANN vẫn cho ra được kết quả dự báo dù độ chính xác chưa bằng được khi có đầy đủ dữ liệu vào nhưng cũng kết quả đạt sai số chấp nhận được.

4.2. Kết quả mô phỏng và dự báo thủy văn lưu lượng vào hồ

Bài toán dự báo lưu lượng dòng chảy đến hồ Hòa Bình phục vụ tốt mục tiêu điều tiết vận hành hồ. Nghiên cứu sử dụng số liệu trong [8] bao gồm các số liệu từ năm 1964 đến năm 2002 tại trạm đo Tạ Bú trên sông Đà, trạm đo lưu lượng gần hồ Hòa Bình nhất. Số liệu này được đo trong mùa cạn từ tháng 12 năm trước đến tháng 5 năm sau. Các giá trị được lấy làm đầu vào cho mô hình như sau: Lưu lượng nước trung bình hiện tại: $Q(t)$; Lưu lượng nước trước đó 10 ngày: $Q(t-10)$; Lưu lượng nước trước đó 20 ngày: $Q(t-20)$; Trung bình lượng mưa hiện tại: $X(t)$; Trung bình lượng mưa trước đó 10 ngày: $X(t-10)$; Trung bình lượng mưa trước đó 20 ngày: $X(t-20)$; Lưu lượng nước của ngày hiện tại: $Qng(t)$; Lượng mưa của ngày hiện tại: $Xng(t)$. Bộ dữ liệu được chia làm hai phần: Phần dữ liệu học (training set): Từ cuối năm 1964 đến đầu năm 1995 có tổng cộng 480 mẫu học; Phần dữ liệu kiểm tra (test set): Từ cuối năm 1995 đến đầu năm 2002 tổng cộng có 90 mẫu kiểm tra.

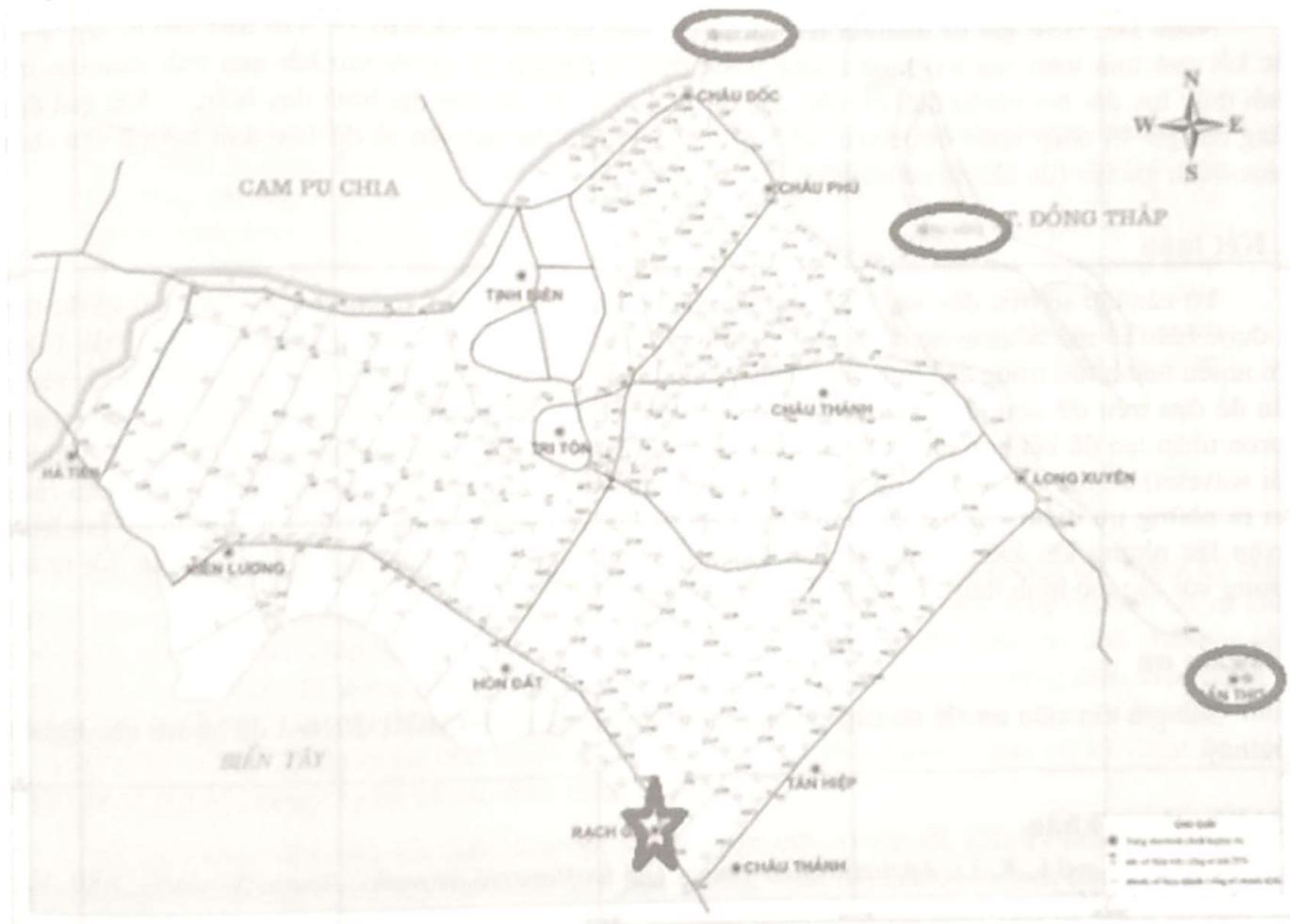


Nhận xét: NSE của phương án hiệu chỉnh và dự báo lần lượt là 0.758 và 0.579. Với tập dữ liệu thu được trải dài từ năm 1964 đến 2002, số lượng mẫu học không nhiều, các giá trị đầu vào cách nhau

khá rời rạc, mạng ANN cũng đã tạo được quan hệ đầu vào – đầu ra một cách tương đối thể hiện ở kết quả tính (NSE hiệu chỉnh chỉ đạt 0.75, NSE dự báo chỉ đạt 0.58).

4.3. Kết quả mô phỏng và dự báo độ mặn tại vùng Tứ Giác Long Xuyên

Tứ Giác Long Xuyên là 1 trong 3 khu vực nhiễm mặn đáng chú ý của đồng bằng sông Cửu Long, bị ảnh hưởng trực tiếp của nước mặn phía biển Tây. Vùng này có các kênh tiếp nước đều xuất phát từ miền nước ngọt của Sông Hậu, độ mặn ở đây được quyết định chủ yếu bởi khả năng tái nước của các kênh dẫn và lượng nước đã dùng trên dọc các tuyến kênh đó. Tuy nhiên các trạm đo mặn trong vùng rất ít và số liệu cũng không đồng bộ, đầy đủ.

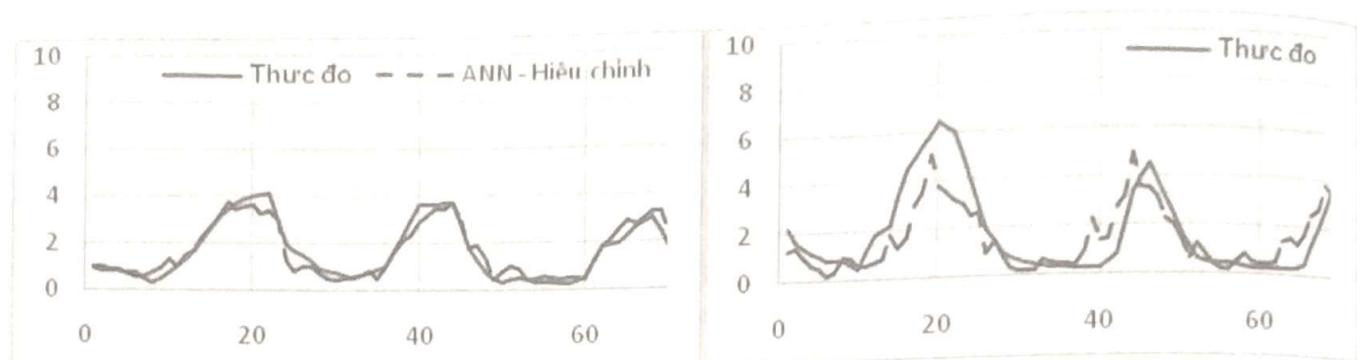


Hình 7. Mô hình tính tại vùng Tứ Giác Long Xuyên.

Yêu cầu của bài toán là tính giá trị mặn tại Rạch Giá qua các giá trị tín hiệu đầu vào tại các điểm như Châu Đốc, Vàm Nao, Cần Thơ, từ dao động mực nước biển xa (tính theo các hằng số điều hòa) cũng như từ giá trị mực nước tại chính Rạch Giá. Trong nghiên cứu này, sử dụng các kịch bản tính theo Bảng 5.

Bảng 5. Kịch bản và kết quả độ mặn tính toán

Năm	Mùa	Triều	Trạm	Phương án tính toán	NSE mô hình thủy lực kết nối 1-2D đã tính
2008	Mùa khô	Triều cường	Rạch Giá	Hiệu chỉnh	0.96
2011	Mùa khô	Triều cường	Rạch Giá	Kiểm định	0.61



Nhận xét: NSE giá trị mặn tại Rạch Giá lần lượt là 0.9094 và 0.6159. Với đầu vào ít, tuy nhiên các kết quả tính toán của mô hình mạng ANN đưa ra đạt giá trị tốt so với kết quả tính toán của mô hình thủy lực đòi hỏi nhiều dữ liệu như các dữ liệu mặt cắt, dữ liệu địa hình đáy biển,... Kết quả tính cũng đạt giá trị chấp nhận được (chỉ số NSE > 0.61) cho dù các yếu tố dữ liệu ảnh hưởng còn chưa được đánh giá hết (do không có dữ liệu).

5. Kết luận

Từ các tập số liệu đầu vào - đầu ra, cho dù có chứa các nhiễu, mạng nơron nhân tạo có thể tìm ra được hàm số mô tả quan hệ đó mà không đòi hỏi kiến thức về mối liên hệ giữa các yếu tố đó. Cùng với nhiều thành tựu trong các lĩnh vực khác, ANN hiện đang là mô hình hứa hẹn giải quyết được nhiều vấn đề dựa trên dữ liệu lớn và sự phát triển của các công cụ tính toán. Trong nghiên cứu này, mạng nơron nhân tạo đã kết hợp nhiều thuật toán và các kỹ thuật khác nhau (song song CPU, GPU và biến đổi wavelet) để thử nghiệm áp dụng cho bài toán thủy văn, thủy lực đạt kết quả tốt. Nghiên cứu cũng chỉ ra những ưu điểm nổi bật của mô hình mới, đó là yêu cầu về loại dữ liệu ít, tuy thời gian huấn luyện lâu nhưng khi kiểm định, dự báo lại rất nhanh, các chỉ số đánh giá NSE đạt mức tốt tương đương với các mô hình thủy văn, thủy lực đã xây dựng.

Lời cảm ơn

Tác giả xin cảm ơn Đề tài cấp Thành phố Hà Nội mã số 01C-07/1-2016-3 đã hỗ trợ cho nghiên cứu này.

Tài liệu tham khảo

- [1] Blum E. K. and L. K. Li. Approximation Theory and feedforward networks, *Neural Networks*, 1991, Vol. 4, pp. 511-515.
- [2] D. E. Goldberg, *Genetic Algorithm in Search, Optimization and Machine Learning*, Addison Wesley, Reading, MA, (1989).
- [3] Hecht-Nielsen, R. 1989. Theory of backpropagation neural network. In *Proceedings of the International Joint Conference on Neural Networks*, Washington DC., (June 1989), IEEE TAB Neural Network Committee, pp. I593-P605.
- [4] Kolmogorov A. N. *On the representation of continuous functions of many variables by superposition of continuous function of one variable and addition*. Dokl, Akad, Nauk SSSR, 114, 953-956, Trans. Am. Math-Soc. 2(28), 55-59, (1957).
- [5] Lê Minh Trung. *Giáo trình mạng nơron nhân tạo*, Nhà Xuất bản Thông kê, (1999).
- [6] Martin T. Hagan, *Neural Network Design*, PWS Publishing Company, (1996).
- [7] Oscar R. Dolling, Eduardo A. Varas, *Artificial neural networks for stream flow prediction*, Journal of Hydraulics research, 40(5), 547-554, (2002).
- [8] Phạm Thị Hoàng Nhụng, *Nghiên cứu ứng dụng các phương pháp học máy tiên tiến trong công tác dự báo vận hành hồ Hòa Bình*, Luận văn Thạc sỹ, (2007).

ISBN: 978-604-913-752-5



9 786049 137525

SÁCH KHÔNG BÁN