

# NGHIÊN CỨU, ỨNG DỤNG MẠNG NƠ RON NHÂN TẠO ĐỂ DỰ BÁO, CHỈNH BIÊN TÀI LIỆU MỨC NƯỚC SÔNG KHÔNG BỊ ẢNH HƯỞNG BỞI THỦY TRIỀU

Trần Cảnh Dương

Trường Đại học Tài nguyên và Môi trường Hà Nội

## Tóm tắt

*Tài liệu lưu lượng nước và mực nước rất quan trọng trong phục vụ kinh tế, xã hội và an ninh quốc phòng. Tuy nhiên, việc thu thập tài liệu này không đủ và không liên tục trong một khoảng thời gian dài nên cần phải có bước chỉnh biên tài liệu. Bài báo này đề xuất sử dụng mạng nơ ron nhân tạo nhiều lớp - MLP để mô hình hóa và tính toán, dự báo mực nước sông không bị ảnh hưởng bởi thủy triều với một sai số cho phép. Tác giả căn cứ dữ liệu quan trắc thực tế của mực nước một dòng sông, sử dụng mạng nơ ron MLP để dự báo mực nước sông sau một ngày và sau mười ngày. Kết quả dự báo được so sánh với kết quả thực tế đã được quan trắc từ đó đánh giá sai số. Sử dụng mạng nơ ron MLP để mô hình hóa và tính toán, dự báo mực nước sông không bị ảnh hưởng bởi thủy triều đã cho kết quả phù hợp với sai số cho phép. Khi sử dụng phương pháp dự báo này các cán bộ kỹ thuật cần dựa vào số liệu quan trắc cụ thể và căn cứ nhu cầu dự báo để điều chỉnh tham số đầu vào, tham số đầu ra của mạng nơ ron MLP một cách thích hợp. Như vậy, ta có thể ứng dụng phương pháp này để dự báo số liệu mực nước sông theo yêu cầu thực tế đồng thời hỗ trợ công tác chỉnh biên dữ liệu thủy văn.*

**Từ khóa:** Dữ liệu thủy văn; Mạng nơ ron nhân tạo nhiều lớp - MLP; Xử lý tín hiệu số; Giá trị cực trị đa biến; Tham số học thích nghi.

## Abstract

### ***Research and application of artificial neural networks to forecast and correct water level at rivers unaffected by tides***

*Data of water flow and water level are very important in socio-economic development and national security. However, such data may be insufficient or discontinuously collected over a long period of time; thus, in some cases, data estimation is required. This paper proposes to use Multi Layer Perceptron - MLP neural network to model and calculate, estimate water level of rivers unaffected by tides with an allowed error. Based on actual monitoring data of a river's water level, MLP neural network was used to forecast river water level after one day and after ten days. Forecast results were compared with actual observed results in order to evaluate errors. Using MLP neural network to model and calculate, estimated river water levels were acceptable with the allowed error. Using this forecasting method, technicians should base on specific monitoring data and forecasting needs to adjust input and output parameters of the MLP network appropriately. Thus, this method could be applied to forecast river water level data according to actual requirements and support for the correction of hydrological data.*

**Keywords:** Hydro-data; Multi Layer Perceptron - MLP neural network; Digital signals processing; Multivariate extreme values; Adaptive learning parameters.

## 1. Đặt vấn đề

Hiện nay, có nhiều phương pháp dự báo đã được đưa ra dựa trên mô hình vật lý và toán học với kết quả đáng ghi nhận. Theo Changhyun Choi và cộng sự (2019) [3] đã trình bày việc phát triển một mô hình dự đoán mực nước bằng cách sử dụng các mô hình học máy khác nhau, chẳng hạn như mạng nơ-ron nhân tạo (ANN), cây quyết định (DT), rừng ngẫu nhiên (RF) và máy vector hỗ trợ (SVM). Vùng đất ngập nước Upo, là vùng đất ngập nước nội địa lớn nhất ở Hàn Quốc, được chọn làm nghiên cứu. Dữ liệu đo mực nước hàng ngày từ năm 2009 đến năm 2015 được sử dụng làm biến phụ thuộc, trong khi dữ liệu khí tượng và dữ liệu đo mực nước thượng lưu được sử dụng như các biến độc lập. Đánh giá hiệu suất dự đoán bằng cách sử dụng RF làm mô hình cuối cùng tiết lộ giá trị 0,96 cho mỗi tương quan hệ số (CC); 0,92 đối với hiệu quả Nash - Sutcliffe (NSE); 0,09 đối với sai số trung bình căn bậc hai (RMSE) và 0,19 cho chỉ số bền bỉ (PI) [3]. Các kết quả nghiên cứu của các nhà khoa học Hàn Quốc đã cung cấp dữ liệu cơ bản để phát triển một phương pháp quản lý đất ngập nước, sử dụng mực nước dự báo của các khu vực trước đây chưa từng được đo đạc.

Tác giả Nguyễn Chính Kiên (2020) [6] đã nghiên cứu xây dựng mô hình thủy văn thông số tập trung trong dự báo lũ cho các lưu vực sông ở Việt Nam. Nội dung của đề tài đưa ra phương pháp giải bài toán ước tính thông số tối ưu trong mô hình thủy văn giúp giải quyết được bài toán hiệu quả, tìm ra bộ thông số phù hợp nhất cho từng vùng phục vụ cho quá trình tác nghiệp dự báo thủy văn một cách chính xác hơn. Đề tài đã giới thiệu sơ đồ

khối mô hình thủy văn Nielsen - Hansen và sơ đồ khối các phương pháp giải bài toán ước tính thông số tối ưu: Phương pháp ô vuông, Phương pháp Rosenbrock, Phương pháp Nelder - Mead, Phương pháp Hooke - Jeeves, Giải thuật di truyền, Phương pháp SCE. Tác giả đã xây dựng phần mềm dự báo lũ bằng mô hình thủy văn thông số tập trung có sử dụng phương pháp ước tính thông số tối ưu.

Công tác chinh biên hiện nay trong ngành thủy văn Việt Nam chủ yếu vẫn được thực hiện bằng phương pháp thủ công. Phần mềm HYDRODB 2010 (Hệ chương trình xử lý số liệu thủy văn) đã hỗ trợ một số công việc như tính toán mực nước trung bình ngày, lưu lượng nước, lưu lượng chất lơ lửng, tính toán các đặc trưng,... Tuy vậy, phần mềm này còn một số tồn tại như sử dụng phần mềm cần nhiều bước trung gian, giao diện chưa phù hợp, tốc độ tính toán chậm, vẫn còn nhiều bước thủ công để nhập dữ liệu.

Mỗi mô hình đều có những ưu và nhược điểm nhất định, vấn đề cần thiết đó là tìm kiếm phương pháp phù hợp, đáp ứng được các yêu cầu thực tế giải quyết bài toán dự báo thủy văn vẫn là nội dung nghiên cứu thời sự hiện nay. Ứng dụng mạng nơ-ron nhân tạo cho lĩnh vực thủy văn có một số ưu điểm nhất định. Bài báo này đề xuất sử dụng mạng nơ-ron nhân tạo nhiều lớp - MLP để mô hình hóa và tính toán, dự báo mực nước sông không bị ảnh hưởng bởi thủy triều với một sai số cho phép.

## 2. Phân tích khả năng ứng dụng mạng nơ-ron nhân tạo cho lĩnh vực thủy văn

Một trong những mục đích chính của ngành trí tuệ nhân tạo (AI- Artificial

## Nghiên cứu

Intelligence) là nghiên cứu và phát triển các mô hình mô phỏng những khả năng về tư duy, phân tích và xử lý thông tin của con người. Mạng nơ ron nhân tạo là một trong những chuyên ngành của trí tuệ nhân tạo được phát triển mạnh mẽ hiện nay với các ứng dụng rộng rãi nhiều vấn đề về kỹ thuật. Các mạng nơ ron có một đặc trưng khác so với các hệ thống tính toán tuyến tính và phi tuyến trước đây đó là mỗi mạng nơ ron được trang bị một hoặc nhiều thuật toán học cho mạng. Ta có thể sử dụng một số công cụ tính toán tối ưu trong quá trình tìm giải pháp cho các vấn đề kỹ thuật [4]. Giá trị cực tiểu của một hệ phương trình phi tuyến đa biến có thể được xác định bằng đồ thị hoặc các chùm lệnh [7]. Các thuật toán cho phép chúng ta xây dựng, lựa chọn cấu trúc cho mạng cũng như điều chỉnh thích nghi các thông số của mạng để phù hợp với yêu cầu của bài toán thực tế được ra cho mạng. Trí tuệ nhân tạo đang dần được thâm nhập vào mọi lĩnh vực của cuộc sống con người và ứng dụng vào khoa học công nghệ, trong đó có lĩnh vực khí tượng thủy văn. Tài liệu lưu lượng nước rất quan trọng trong phục vụ kinh tế, xã hội và an ninh quốc phòng. Tuy nhiên, việc thu thập tài liệu này không đủ và không liên tục nên cần phải có bước chỉnh biên tài liệu. Ta có thể sử dụng mạng nơ ron để hỗ trợ một số công việc như tính toán mực nước trung bình ngày, tính lưu lượng nước, lưu lượng chất lơ lửng, tính toán các đặc trưng,... Việc nghiên cứu áp dụng công nghệ trí tuệ nhân tạo (AI) vào công tác dự báo, chỉnh biên tài liệu lưu lượng nước nhằm đảm bảo bộ dữ liệu chuẩn, sát với thực tế trong nhiều năm với sai số nhỏ nhất, giảm bớt các thao tác thủ công bằng chuỗi thao tác một cách tự động.

### **3. Cấu trúc mạng MLP và sai số trong quá trình học của mạng MLP**

Tác giả sử dụng mạng MLP để dự báo mực nước bởi lẽ mạng này là một trong những mạng nơ ron kinh điển nhất, có một số ưu điểm và sử dụng thuận tiện đồng thời có số lượng lớn ứng dụng phổ biến trong các nghiên cứu trong nhiều lĩnh vực. Ngoài ra, cấu trúc mạng MLP được sử dụng rộng rãi trong việc tái tạo các ánh xạ đầu vào với đầu ra mà chúng được xác định từ bộ số liệu mẫu. Qua tìm hiểu tác giả thấy rằng, các hàm thể hiện quan hệ dữ liệu đầu vào với đầu ra của mực nước các dòng sông là các hàm phi tuyến bậc cao, mặt khác các số liệu quan trắc lại chứa nhiều nhiễu do đó để hạn chế các nhược điểm nêu trên thì việc chọn mô hình xấp xỉ bằng mạng nơ ron trở nên hữu hiệu, đặc biệt việc sử dụng mô hình mạng kinh điển MLP có nhiều thuận tiện và hiệu quả cao với sai số cho phép.

Mạng nơ ron nhân tạo nhiều lớp - MLP (Multi Layer Perceptron) là một mạng truyền thẳng với các khối cơ bản là các nơ ron McCulloch - Pits. Các nơ ron được sắp xếp thành các lớp (layer). Một lớp nối đến các kênh tín hiệu đầu vào (input layer), một lớp nối đến các kênh tín hiệu đầu ra (out layer) và có thể thêm một số lớp trung gian (lớp ẩn - hidden layer) [1]. Các nhà khoa học đã chứng minh rằng, chỉ cần sử dụng tối đa hai lớp ẩn là có thể mô hình hóa một hàm phi tuyến với độ chính xác tùy chọn. Mạng MLP được sử dụng để tạo các ánh xạ số liệu vào - ra trên cơ sở các bộ số liệu mẫu. Bộ số liệu mẫu là một tập hợp gồm  $p$  các cặp mẫu được cho dưới dạng véc tơ đầu vào - véc tơ đầu ra tương ứng với  $i = 1, 2, \dots, p$ , , trong đó  $N$  là số đầu vào và  $K$  là số đầu

ra của mạng nơ ron. Mạng MLP có thể có nhiều đầu ra đồng thời nên giá trị  $d_i$  là một véc tơ nhiều thành phần. Dựa theo bộ số liệu mẫu ta cần xác định một mạng MLP với các thông số cấu trúc và các trọng số ghép nối tương ứng theo cấu trúc đã chọn đảm bảo mạng MLP có thể xác định lại được ánh xạ từ đầu vào đến đầu ra của các số liệu mẫu [5].

Các giá trị đầu vào  $x_i$  được đưa vào đối tượng nghiên cứu sẽ cho ra kết quả  $d_i$  và giá trị  $x_i$  được đưa vào mạng MLP sẽ cho kết quả ở đầu ra là  $y_i$ . Như vậy, với cùng một đầu vào thì đầu ra của mạng MLP phải xấp xỉ bằng đầu ra của đối tượng nghiên cứu [5]. Sai số tổng cộng trên các mẫu tiến tới một giá trị cực tiểu:

$$E = \frac{1}{2} \sum_{i=1}^p \|MLP(x_i) - d_i\|^2 \rightarrow \min$$

Trong đó ký hiệu  $\|MLP(x_i) - d_i\|$  biểu diễn chuẩn khoảng cách.

Chuẩn khoảng cách giữa các véc tơ thường được dùng theo công thức Öclit:

$$x, c \in R^N : d(x, c) = \|x - c\| = \sqrt{\sum_{i=1}^N (x_i - c_i)^2}$$

ID	X0	X1	X2	X3	X4	X5	X6	X7	X8	Y	LABEL
0	250	261	299	295	242	216	210	209	199	187	30-9-10-2016
1	261	299	295	242	216	210	209	199	187	165	1-10-2016
2	299	295	242	216	210	209	199	187	165	156	2-10-2016
3	295	242	216	210	209	199	187	165	156	166	3-10-2016
.....											
88	142	131	131	133	126	123	127	129	111	106	27-12-2016
89	131	131	133	126	123	127	129	111	106	115	28-12-2016
90	131	133	126	123	127	129	111	106	115	131	29-12-2016
91	133	126	123	127	129	111	106	115	131	129	30-12-2016
92	126	123	127	129	111	106	115	131	129	129	31-12-2016

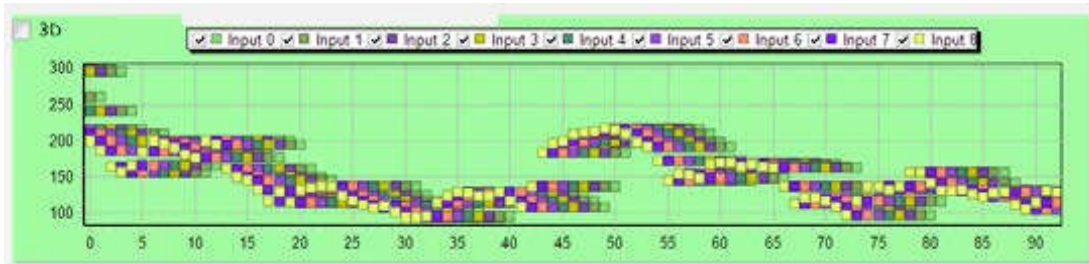
#### 4. Sử dụng mạng MLP để dự báo mực nước sông không bị ảnh hưởng bởi thủy triều

##### 4.1. Dự báo mực nước sông sau một ngày

Phần mềm MLP được sử dụng để mô phỏng việc dự báo dữ liệu mực nước sông [2]. Dựa trên số liệu đo thực tế ta lập bảng gồm 12 cột và 93 hàng. Cột ID chỉ ra định danh của dữ liệu một hàng (có giá trị từ 0 đến 93). Số ID cuối cho biết tổng các tập dữ liệu. Số liệu tập dữ liệu có thể điều chỉnh tùy theo số liệu đo thực tế có sẵn và theo yêu cầu dự báo cụ thể. Các cột Xi với i chạy từ 0 đến 7 biểu thị mực nước trước ngày hiện tại 8 ngày, 7 ngày,... cho đến 1 ngày. Cột X8 biểu thị mực nước ngày hiện tại, mà nó được xác định theo từng hàng và được chỉ ra theo số liệu của nhãn (LABEL) ở cột cuối cùng. Cột Y là số liệu thực tế đo được sau 1 ngày tính từ ngày hiện tại. Ta biểu thị một số dòng đầu và một số dòng cuối của bảng như sau:

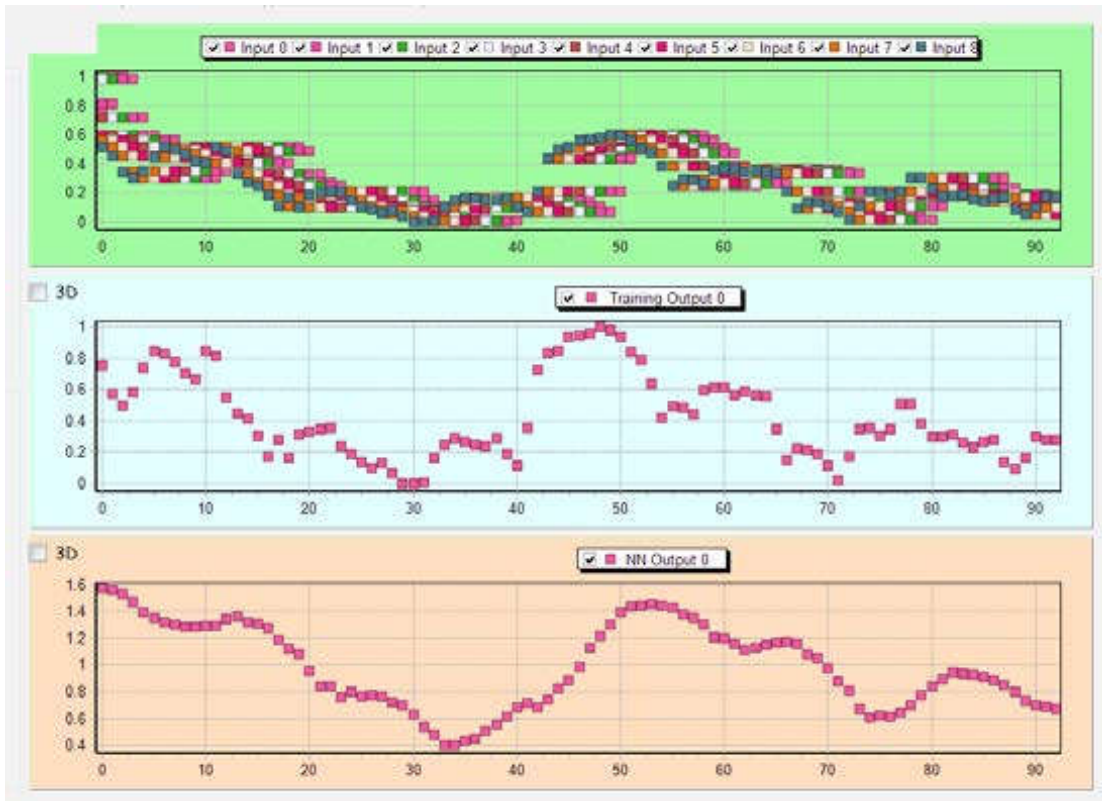
## Nghiên cứu

Số liệu mực nước tại hàng cuối tương ứng ngày 31-12-2016, trong đó cột X8 là mực nước ngày hiện tại có giá trị 129. Tất cả các dòng thể hiện số liệu đo thực tế, riêng giá trị 129 trong ô tương ứng cột Y tại hàng cuối ta giả định đó là mực nước ngày 1-1-2017 (bằng mực nước ngày 31-12-2016). Ta sẽ lấy giá trị tín hiệu ra của mạng nơ ron sau khi đào tạo tại ngày 1-1-2017 làm giá trị mực nước dự báo cho ngày 1-1-2017. Hình 1 biểu thị mức tín hiệu vào khi chưa được chuẩn hóa.



**Hình 1: Mức tín hiệu vào khi chưa được chuẩn hóa**

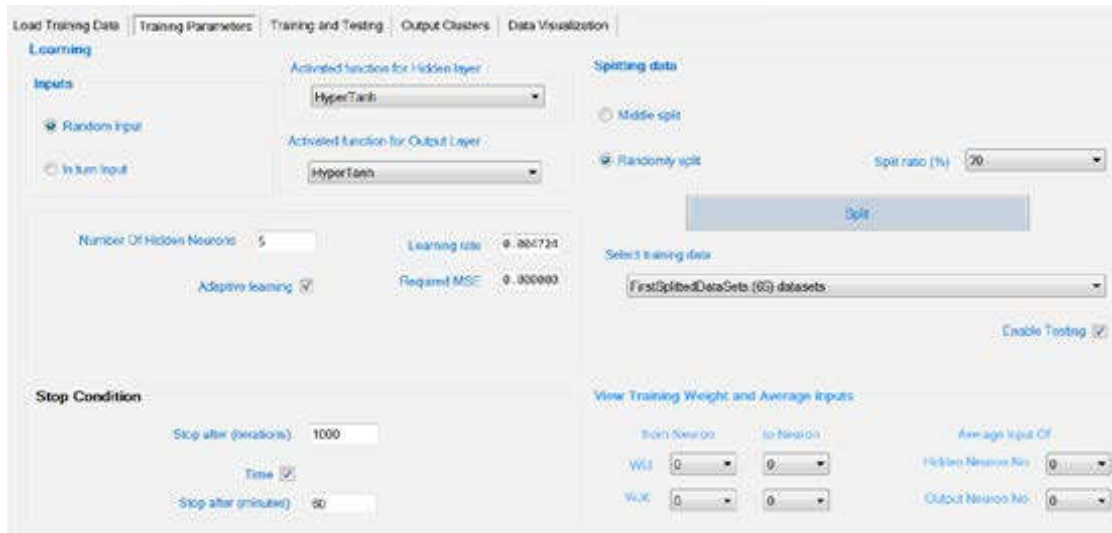
Sau khi chuẩn hóa dữ liệu đầu vào ta có kết quả được thể hiện ở Hình 2. Đồ thị biểu diễn bao gồm các tín hiệu vào theo thời gian; Tín hiệu đào tạo và tín hiệu kiểm tra; Tín hiệu ra của mạng nơ ron sau khi đào tạo.



**Hình 2: Đồ thị biểu diễn các tín hiệu theo thời gian sau khi tín hiệu đầu vào được chuẩn hóa**

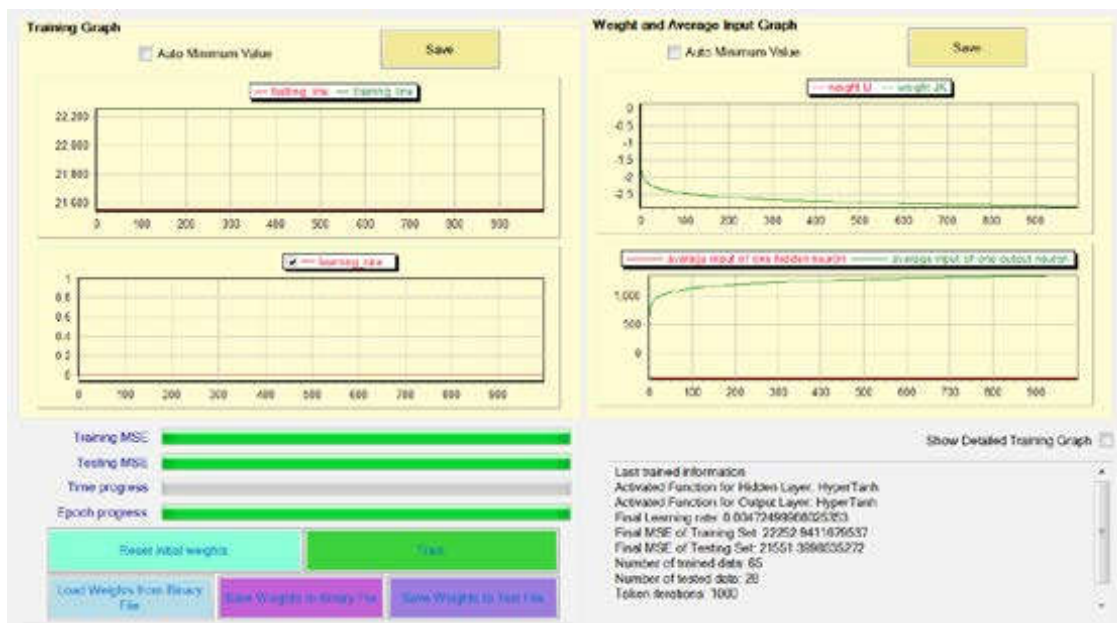
Khi cho trước một bộ số liệu mẫu của mực nước đầu vào, việc xác định một mạng MLP để tái tạo lại bộ số liệu bao gồm việc xác định các thông số về cấu trúc như số đầu vào, số đầu ra, số lớp ẩn, số nơ ron trên mỗi lớp, các hàm truyền đạt. Các hàm truyền

đạt của lớp ra phụ thuộc chủ yếu vào dải tín hiệu đích. Nếu đầu ra được giới hạn trong đoạn  $[0,1]$  thì ta sử dụng hàm  $\text{logsig}$ , nếu đầu ra được giới hạn trong đoạn  $[-1,1]$  thì ta sử dụng hàm  $\text{tansig}$ . Trường hợp đầu ra vượt ra ngoài đoạn  $[-1,1]$  thì ta có thể sử dụng hàm tuyến tính. Trong quá trình dự báo mực nước, cán bộ kỹ thuật có thể thay đổi tham số đào tạo để chọn bộ tham số thích hợp. Chẳng hạn, các tham số đào tạo mạng cần được lựa chọn một cách phù hợp như được chỉ ra ở Hình 3.



**Hình 3: Cửa sổ lựa chọn tham số đào tạo**

Các tham số có thể tham gia vào quá trình điều khiển thích nghi của mạng MLP là các trọng số ghép nối giữa các lớp. Hình 4 cho thấy cửa sổ xác nhận quá trình đào tạo thành công và đồ thị biểu diễn trọng số.



**Hình 4: Cửa sổ xác nhận quá trình đào tạo thành công và đồ thị biểu diễn trọng số**

Theo đồ thị của tín hiệu ra của mạng nơ ron sau khi đào tạo ta thấy mực nước dự báo ngày 1-1-2017 nhỏ hơn so với mực nước ngày 31-12-2016.

## Nghiên cứu

Ta có kết quả mô phỏng dữ liệu ra do mạng nơ ron đưa ra sau khi đào tạo cùng với dữ liệu đầu ra thực tế (Data with NN Outputs and Desired Outputs), chúng được biểu diễn trong một bảng. Số liệu một số dòng đầu bảng và một số dòng cuối bảng được thể hiện như sau:

```
Activated Function for Hidden Layer: HyperTanh;
Activated Function for Output Layer: HyperTanh;
Final Learning rate: 0.00472499966025353;
Final MSE of Training Set: 22252.9411679537;
Final MSE of Testing Set: 21551.3998635272;
Number of trained data: 65;
Number of tested data: 28;
Taken iterations: 1000;

No.,X0,X1,X2,X3,X4,X5,X6,X7,X8,Y_Desired,Y_Modeling,LABEL,Used as
0,0.760975609756098,0.814634146341463,1,0.980487804878049,0.721951219512195,0.595121951219512,0.565853658536585,0.560975609756098,0.51219512195122,0.75,1.57103173408962,30-9-10-2016,Training,
1,0.814634146341463,1,0.980487804878049,0.721951219512195,0.595121951219512,0.565853658536585,0.560975609756098,0.51219512195122,0.453658536585366,0.57258064516129,1.55565917465204,1-10-2016,Testing
2,1,0.980487804878049,0.721951219512195,0.595121951219512,0.565853658536585,0.560975609756098,0.51219512195122,0.453658536585366,0.346341463414634,0.5,1.52823669901192,2-10-2016,Training,
3,0.980487804878049,0.721951219512195,0.595121951219512,0.565853658536585,0.560975609756098,0.51219512195122,0.453658536585366,0.346341463414634,0.302439024390244,0.580645161290323,1.47074699743277,3-10-2016,Training,
4,0.721951219512195,0.595121951219512,0.565853658536585,0.560975609756098,0.51219512195122,0.453658536585366,0.346341463414634,0.302439024390244,0.351219512195122,0.741935483870968,1.39341586922342,4-10-2016,Training,
5,0.595121951219512,0.565853658536585,0.560975609756098,0.51219512195122,0.453658536585366,0.346341463414634,0.302439024390244,0.351219512195122,0.448780487804878,0.846774193548387,1.34834600551252,5-10-2016,Training,
6,0.565853658536585,0.560975609756098,0.51219512195122,0.453658536585366,0.346341463414634,0.302439024390244,0.351219512195122,0.448780487804878,0.51219512195122,0.82258064516129,1.31582176969683,6-10-2016,Training,
.....
85,0.209756097560976,0.307317073170732,0.307317073170732,0.234146341463415,0.180487804878049,0.180487804878049,0.190243902439024,0.15609756097561,0.141463414634146,0.266129032258065,0.90705058961642,24-12-2016,Training,
86,0.307317073170732,0.307317073170732,0.234146341463415,0.180487804878049,0.180487804878049,0.190243902439024,0.15609756097561,0.141463414634146,0.160975609756098,0.282258064516129,0.884361036383869,25-12-2016,Testing,
87,0.307317073170732,0.234146341463415,0.180487804878049,0.180487804878049,0.190243902439024,0.15609756097561,0.141463414634146,0.160975609756098,0.170731707317073,0.137096774193548,0.848338794431386,26-12-2016,Testing,
88,0.234146341463415,0.180487804878049,0.180487804878049,0.190243902439024,0.15609756097561,0.141463414634146,0.160975609756098,0.170731707317073,0.0829268292682927,0.0967741935483871,0.800016604665585,27-12-2016,Training,
89,0.180487804878049,0.180487804878049,0.190243902439024,0.15609756097561,0.141463414634146,0.160975609756098,0.170731707317073,0.0829268292682927,0.0585365853658537,0.169354838709677,0.735825764541826,28-12-2016,Training,
90,0.180487804878049,0.190243902439024,0.15609756097561,0.141463414634146,0.160975609756098,0.170731707317073,0.0829268292682927,0.0585365853658537,0.102439024390244,0.298387096774194,0.697954710034165,29-12-2016,Training,
91,0.190243902439024,0.15609756097561,0.141463414634146,0.160975609756098,0.170731707317073,0.0829268292682927,0.0585365853658537,0.102439024390244,0.180487804878049,0.282258064516129,0.693921676226953,30-12-2016,Training,
92,0.15609756097561,0.141463414634146,0.160975609756098,0.170731707317073,0.0829268292682927,0.0585365853658537,0.102439024390244,0.180487804878049,0.170731707317073,0.282258064516129,0.668641630493483,31-12-2016,Training,
```

Như vậy, hệ số Y (dữ liệu mực nước dự báo) của ngày 1-1-2017 là 0,668641630493483. Ta có thể làm tròn thành 0,669. Tương tự hệ số dữ liệu học máy của ngày 31-1-2016 là 0,693921676226953, mà nó được làm tròn thành 0,694. Ta có thể tham chiếu dữ liệu thực của mực nước ngày 31-1-2016 là 129. Từ đó ta tính được mực nước dự báo ngày 1-1-2017 như sau:

$$(129:0,694) \times 0,669 = 124,353$$

Theo thống kê số liệu mực nước đo thực tế ngày 1-1-2017 là 120 cm. Như vậy kết quả dự báo gần sát với thực tế.

**4.2. Dự báo mực nước sông sau mười ngày**

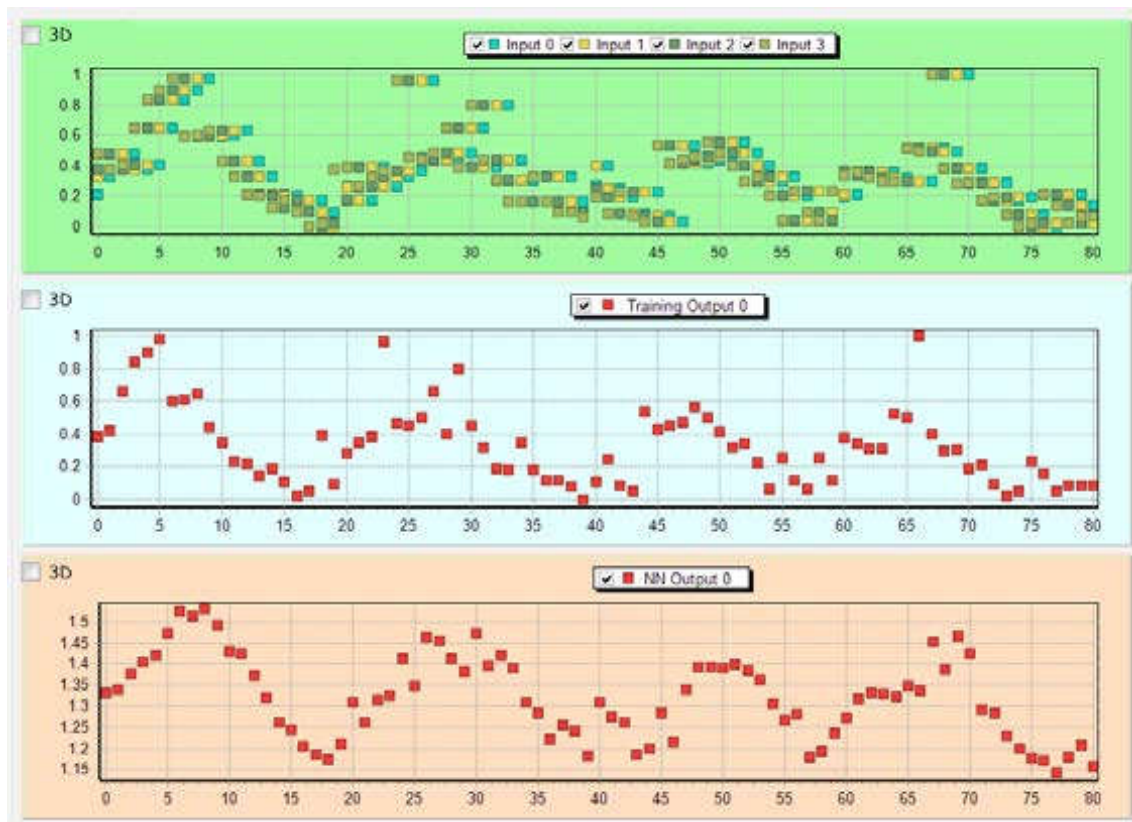
Ta lập bảng dữ liệu thực đo được trên thực tế.

IDT-30	T-20	T-10	T	T+10	ng-th-nam	
0	210	276	300	357	299	13-6-2013
1	276	300	357	299	318	23-6-2013
2	300	357	299	318	453	3-7-2013
3	357	299	318	453	556	13-7-2013

.....

78108	210	170	108	131	10-12-2016
79210	170	108	131	131	20-12-2016
80170	108	131	131	131	30-12-2016

Số liệu đầu ra sau 10 ngày của 20-12-2016 chính là số liệu thực ngày 30-12-2016. Số liệu ngày 30-12-2016 là thực. Tuy nhiên, số liệu sau 10 ngày của ngày 30-12-2016 chính là mực nước ngày 9-1-2017 là số liệu giả định (được xem bằng số liệu mực nước ngày 30-12-2016). Sau khi đào tạo, mạng nơ ron sẽ đưa ra số liệu dự báo mực nước ngày 9-1-2017 (sau 10 ngày so với ngày 30-12-2016) như được chỉ ra ở Hình 5.



**Hình 5: Số liệu dự báo mực nước ngày 9-1-2017 (sau 10 ngày so với ngày 30-12-2016)**

**Nghiên cứu**

```

Activated function for Hidden Layer: hyperfanh;
Activated function for Output Layer: Hyperfanh;
Final Learning rate: 0.005;
Final MSE of Training Set: 88495.9383701093;
Number of trained data: 81;
Taken iterations: 1000;

No. T-30,T-20,T-10,T+10 Desired,T+10 Modeling LABEL Used as
0,0.231573236889693,0.359522262314647,0.374321850650995,0.477396021699819,0.386925795053004,1.33088901747113,3-6-2013,Training,
1,0.330822242314647,0.374321850650995,0.477396021699819,0.372513562386998,0.4204946699646643,1.3403711046402,23-6-2013,Training,
2,0.374321850650995,0.477396021699819,0.372513562386998,0.406871609403255,0.859020600706714,1.37506631450189,3-7-2013,Training,
3,0.477396021699819,0.372513562386998,0.406871609403255,0.650994575045208,0.840689199293286,1.4039487742415,13-7-2013,Training,
4,0.372513562386998,0.406871609403255,0.650994575045208,0.837251356238698,0.895117540687161,1.47044404127464,2-8-2013,Training,
5,0.406871609403255,0.650994575045208,0.837251356238698,0.895117540687161,0.978798586572438,1.47044404127464,2-8-2013,Training,
6,0.650994575045208,0.837251356238698,0.895117540687161,0.978798586572438,0.602473498233216,1.5257831629376,12-8-2013,Training,
7,0.837251356238698,0.895117540687161,0.978798586572438,0.593128390596745,0.609540636042403,1.51282957018725,22-8-2013,Training,
8,0.895117540687161,0.978798586572438,0.593128390596745,0.600361663652803,0.641362756183746,1.53014174035515,1-9-2013,Training,
9,0.978798586572438,0.593128390596745,0.600361663652803,0.632911392405063,0.443462897526502,1.49174639732373,11-9-2013,Training,
10,0.593128390596745,0.600361663652803,0.632911392405063,0.430379746835443,0.34452296819788,1.42789427922016,21-9-2013,Training,
11,0.600361663652803,0.632911392405063,0.430379746835443,0.327013924050633,0.229681978798587,1.42260547547763,1-10-2013,Training,
12,0.632911392405063,0.430379746835443,0.327013924050633,0.211573236889693,0.219081272084806,1.3721839505717,11-10-2013,Training,
13,0.430379746835443,0.327013924050633,0.211573236889693,0.20072332305606,0.148409893929233,1.31888040476294,21-10-2013,Training,
14,0.327013924050633,0.211573236889693,0.20072332305606,0.128380596745027,0.190813720848057,1.25814422758864,31-10-2013,Training,
15,0.211573236889693,0.20072332305606,0.128380596745027,0.17190235081374,0.1113074204947,1.2442077120261,10-11-2013,Training,
16,0.20072332305606,0.128380596745027,0.17190235081374,0.0904159132007233,0.0229681978798587,1.20422556920082,20-11-2013,Training,
17,0.128380596745027,0.17190235081374,0.0904159132007233,0.0477031802120141,1.18335369940904,30-11-2013,Training,

.....

63,0.336347197106691,0.359855334538879,0.321880650994575,0.296564195298373,0.296564195298373,0.312720848056537,1.32658693088301,13-7-2016,Training,
64,0.359855334538879,0.321880650994575,0.296564195298373,0.296564195298373,0.522968197879859,1.32182908851859,23-7-2016,Training,
65,0.321880650994575,0.296564195298373,0.296564195298373,0.511754068716094,0.53533568904594,1.34662174149021,2-8-2016,Training,
66,0.296564195298373,0.296564195298373,0.511754068716094,0.491862567811935,1.1.3374497197254,12-8-2016,Training,
67,0.296564195298373,0.511754068716094,0.491862567811935,1.0.397526501766784,1.43150792041941,22-8-2016,Training,
68,0.511754068716094,0.491862567811935,1.0.383363471971067,0.295053003333369,1.3878363839949,1-9-2016,Training,
69,0.491862567811935,1.0.383363471971067,0.278481012658228,0.3056371024735,1.46484171630555,11-9-2016,Training,
70,1.0.383363471971067,0.278481012658228,0.289330922242315,0.1890093969576,1.42277660792238,21-9-2016,Training,
71,0.383363471971067,0.278481012658228,0.289330922242315,0.16998191681736,0.210247349823322,1.29267747607837,1-10-2016,Training,
72,0.278481012658228,0.289330922242315,0.16998191681736,0.191681735985533,0.097171448763251,1.28274885243046,11-10-2016,Training,
73,0.289330922242315,0.16998191681736,0.191681735985533,0.0759493670886076,0.0247349823321553,1.22990602627225,21-10-2016,Training,
74,0.16998191681736,0.191681735985533,0.0759493670886076,0.00180831826401447,0.049469964664311,1.1982947132384,31-10-2016,Training,
75,0.191681735985533,0.0759493670886076,0.00180831826401447,0.027124773960217,0.229681978798587,1.17494108735814,10-11-2016,Training,
76,0.0759493670886076,0.00180831826401447,0.027124773960217,0.211573236889693,0.15901060070671378,1.170019449631,20-11-2016,Training,
77,0.00180831826401447,0.027124773960217,0.211573236889693,0.139240506329114,0.049469964664311,1.14253298051535,30-11-2016,Training,
78,0.027124773960217,0.211573236889693,0.139240506329114,0.027124773960217,0.0901060070671378,1.1798336490743,10-12-2016,Training,
79,0.211573236889693,0.139240506329114,0.027124773960217,0.0687160940325497,0.0901060070671378,1.2066172611848,20-12-2016,Training,
80,0.139240506329114,0.027124773960217,0.0687160940325497,0.0687160940325497,0.0901060070671378,1.1575779488344,30-12-2016,Training,

.....

59,1.3878363839949,1-9-2016,Training,
5,1.46484171630555,11-9-2016,Training,
5,1.42277660792238,21-9-2016,Training,
3,210247349823322,1.29267747607837,1-10-2016,Training,
1,097171448763251,1.28274885243046,11-10-2016,Training,
0,0247349823321553,1.22990602627225,21-10-2016,Training,
87,0.049469964664311,1.1982947132384,31-10-2016,Training,
217,0.229681978798587,1.17494108735814,10-11-2016,Training,
593,0.159010600706714,1.170019449631,20-11-2016,Training,
14,0.049469964664311,1.14253298051535,30-11-2016,Training,
0,0901060070671378,1.1798336490743,10-12-2016,Training,
7,0.0901060070671378,1.2066172611848,20-12-2016,Training,
37,0.0901060070671378,1.1575779488344,30-12-2016,Training,

```

Như vậy, ta có hệ số T+10 (dữ liệu mực nước dự báo do máy đưa ra) của ngày 9-1-2017 là 1,15757779488344. Ta có thể làm tròn thành 1,158. Tương tự hệ số dữ liệu học máy của ngày 20-12-2016 đưa ra dự báo hệ số cho ngày 30-1-2016 là 1,2066172611848 mà nó được làm tròn thành 1,207. Ta có thể tham chiếu dữ liệu mực nước thực ngày 30-12-2016 là 131. Từ đó ta tính được mực nước ngày 9-1-2017 như sau:

$$(131:1,207) \times 1,158 = 125,682$$

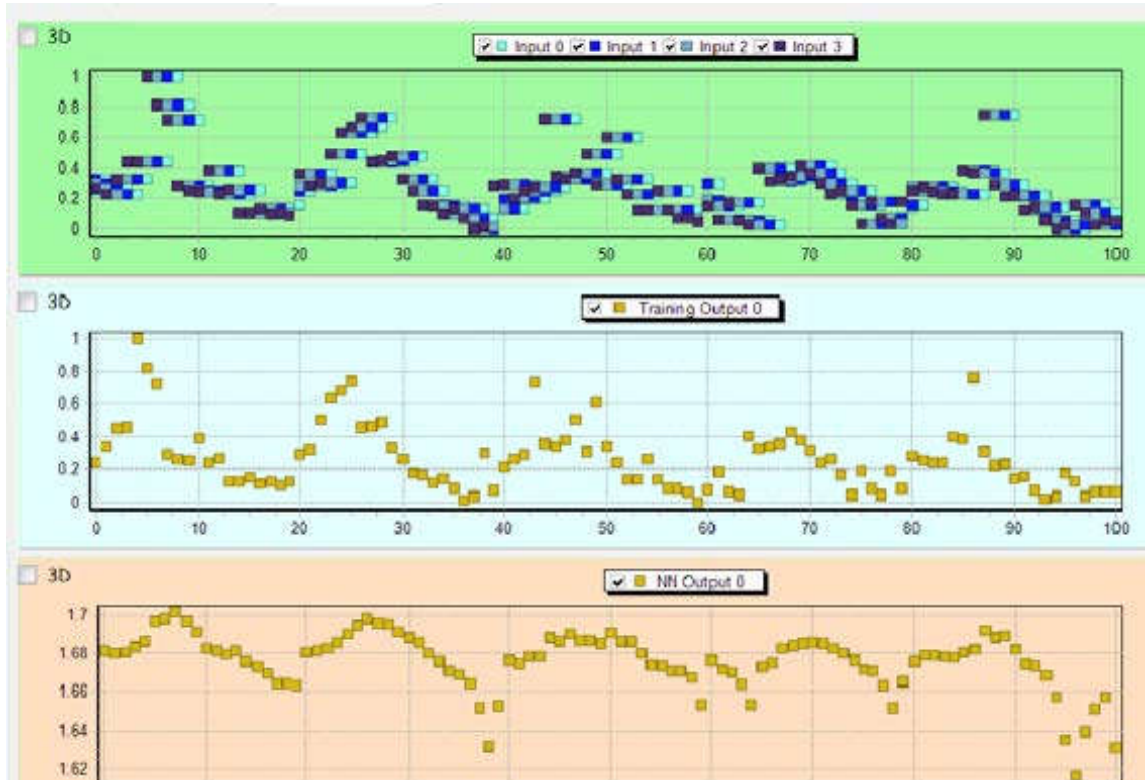
Lưu ý: Số liệu đo ngày 9-1-2017 mực nước là 181. Mực nước đo được ngày 8-1-2016 là 99. Ta có thể lấy mực nước trung bình 2 ngày gần nhau (8-1-2017 và 9-1-2017) là 140. Ta thấy kết quả dự báo gần đúng với mực nước đo được trên thực tế.

Ta thêm dữ liệu đầu vào đối với năm 2012.

IDT-30	T-20	T-10	T	T+10	ng-th-nam	
0	272	332	311	290	261	13-6-2012
1	332	311	290	261	332	23-6-2012
2	311	290	261	332	417	3-7-2012
3	290	261	332	417	419	13-7-2012

96135	94	108	210	170	20-11-2016
9794	108	210	170	108	30-11-2016
98108	210	170	108	131	10-12-2016
99210	170	108	131	131	20-12-2016
100	170	108	131	131	30-12-2016

Sau quá trình học máy, mạng đưa ra kết quả như được biểu diễn ở Hình 6.



**Hình 6: Đồ thị biểu diễn dữ liệu vào, dữ liệu ra đào tạo và dữ liệu mạng MLP sau khi thêm dữ liệu đầu vào năm 2012**

Activated Function for Hidden Layer: HyperTanh;  
 Activated Function for Output Layer: HyperTanh;  
 Final Learning rate: 0.0044999988079071;  
 Final MSE of Training Set: 98432.5262970911;  
 Number of trained data: 101;  
 Taken iterations: 1000;

No.,T-30,T-20,T-10,T,T+10\_Desired,T+10\_Modeling,LABEL,Used as  
 0,0,0.243869209809264,0.325613079019074,0.29700272479564,0.268392370572207,0.242302543507363,1.68118  
 761356151,13-6-2012,Training,  
 1,0.325613079019074,0.29700272479564,0.268392370572207,0.228882833787466,0.337349397590361,1.67995  
 509717862,23-6-2012,Training,  
 2,0.29700272479564,0.268392370572207,0.228882833787466,0.325613079019074,0.451137884872825,1.68076  
 030541103,3-7-2012,Training,  
 3,0.268392370572207,0.228882833787466,0.325613079019074,0.44141689373297,0.453815261044177,1.68324  
 641920231,13-7-2012,Training.

.....  
 95,0.14441416893733,0.0572207084468665,0.00136239782016349,0.0204359673024523,0.174029451137885,1.6  
 3525568630531,10-11-2016,Training,  
 96,0.0572207084468665,0.00136239782016349,0.0204359673024523,0.159400544959128,0.120481927710843,1.  
 61643616864593,20-11-2016,Training,  
 97,0.00136239782016349,0.0204359673024523,0.159400544959128,0.104904632152589,0.0374832663989291,1.  
 63922997915933,30-11-2016,Training,  
 98,0.0204359673024523,0.159400544959128,0.104904632152589,0.0204359673024523,0.0682730923694779,1.6  
 5088114626307,10-12-2016,Training,  
 99,0.159400544959128,0.104904632152589,0.0204359673024523,0.0517711171662125,0.0682730923694779,1.6  
 5734325270061,20-12-2016,Training,  
 100,0.104904632152589,0.0204359673024523,0.0517711171662125,0.0517711171662125,0.0682730923694779,1.  
 63114027813478,30-12-2016,Training,

## Nghiên cứu

Như vậy, ta có hệ số T+10 (dữ liệu mực nước dự báo do máy đưa ra) của ngày 9-1-2017 là 1,63114027813478. Ta có thể làm tròn thành 1,631. Tương tự hệ số dữ liệu học máy của ngày 20-12-2016 đưa ra dự báo hệ số cho ngày 30-1-2016 là 1,65734325270061 mà nó được làm tròn thành 1,657. Ta có thể tham chiếu dữ liệu mực nước thực ngày 30-12-2016 là 131. Từ đó ta tính được mực nước ngày 10-1-2017 như sau:

$$(131:1,657) \times 1,631 = 128,944$$

Lưu ý: Số liệu đo ngày 9-1-2017 mực nước là 181. Mực nước đo được ngày 8-1-2016 là 99. Ta có thể lấy mực nước trung bình 2 ngày gần nhau (8-1-2017 và 9-1-2017) là 140. Ta thấy kết quả dự báo gần đúng với mực nước đo được trên thực tế.

**Nhận xét:** Theo tập dữ liệu từ 13-6-2013 đến 30-12-2016 ta có mực nước dự báo ngày 9-1-2017 là 125,682 cm. Theo tập dữ liệu từ 13-6-2012 đến 30-12-2016 ta có mực nước dự báo ngày 9-1-2017 là 128,944 cm. Như vậy sai số mực ước cho hai trường hợp chỉ khoảng 3 cm. Số liệu này khi so sánh với số liệu mực nước đo thực tế ngày 9-1-2017 cũng không chênh lệch đáng kể.

Phương pháp dự báo mực nước nêu trên đã đạt được một số kết quả nhất định. Tuy vậy, về lý thuyết cũng như thực tế có rất nhiều phương pháp dự báo, mỗi phương pháp có những ưu, nhược điểm khác nhau. Phương pháp này muốn sử dụng hiệu quả cũng cần vận dụng vào thực tế một cách linh hoạt, có điều chỉnh các tham số mạng MLP một cách thích hợp.

## **5. Kết luận**

Sử dụng mạng nơ ron MLP để mô hình hóa và tính toán, dự báo mực nước sông không bị ảnh hưởng bởi thủy triều đã cho kết quả phù hợp với sai số cho phép.

Ta có thể ứng dụng phương pháp này để dự báo số liệu mực nước sông theo yêu cầu thực tế đồng thời hỗ trợ công tác chỉnh biên dữ liệu thủy văn. Khi sử dụng phương pháp dự báo này các cán bộ kỹ thuật cần dựa vào số liệu quan trắc cụ thể và căn cứ nhu cầu dự báo để điều chỉnh tham số đầu vào, tham số đầu ra của mạng MLP một cách thích hợp để được kết quả tối ưu.

## **TÀI LIỆU THAM KHẢO**

[1]. Boger B, Guyon I (1997). *Knowledge extraction from artificial neural network models*. IEEE Systems, Man, and Cybernetics Conference.

[2]. Cao Thang (2007). *Intructions for using Spice-MLP software*. Soft Intelligence Laboratory, Ritsumeikan University, Japan.

[3]. Changhyun Choi, Jungwook Kim, Heechan Han, Daegun Han, Hung Soo Kim (2019). *Development of Water Level Prediction Models Using Machine Learning in Wetlands: A Case Study of Upo Wetland in South Korea*. Institute of Water Resources System, Inha University, Michuhol-Gu, Incheon, Korea. Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523, USA.

[4]. Nguyễn Phùng Quang (2008). *Matlab & Simulink dành cho kỹ sư điều khiển tự động*. Nhà xuất bản Khoa học và Kỹ thuật.

[5]. Trần Hoài Linh (2019). *Mạng nơ ron và ứng dụng trong xử lý tín hiệu số*. Nhà xuất bản Bách Khoa, Hà Nội.

[6]. Nguyễn Chính Kiên (2020). *Nghiên cứu xây dựng mô hình thủy văn thông số tập trung trong dự báo lũ cho các lưu vực sông ở Việt Nam*. Đề tài cơ sở cấp Viện Cơ học - Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

[7]. Trần Cảnh Dương (2020). *Ước lượng thông số tài nguyên và môi trường theo thời gian đối với các địa điểm có dữ liệu tương quan bằng cách ứng dụng mạng nơ ron để xử lý tín hiệu số*. Tạp chí Khoa học Tài nguyên và Môi trường - số 30.

BBT nhận bài: 20/4/2021; Phản biện xong: 17/5/2021; Chấp nhận đăng: 29/6/2021