PAPER • OPEN ACCESS

# Analysis of accuracy parameters of ANN backpropagation algorithm through training and testing of hydro-climatology data based on GUI MATLAB

To cite this article: Syaharuddin *et al* 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **413** 012008

IOP Conf. Series: Earth and Environmental Science **413** (2020) 012008 doi:10.1088/1755-1315/413/1/012008

# Analysis of accuracy parameters of ANN backpropagation algorithm through training and testing of hydro-climatology data based on GUI MATLAB

**Syaharuddin[1*], D Pramita[2], T Nusantara[3*], Subanji[4] and H R P Negara[5]**

[1,2] Department of Mathematics Education, Universitas Muhammadiyah Mataram, Indonesia

[3,4] Department of Mathematics Education, Universitas Negeri Malang, Indonesia

[5] Department of Mathematics Education, Universitas Islam Negeri Mataram, Indonesia

[*)]Email: syaharuddin.ntb@gmail.com

**Abstract.** The authors have developed a GUI Matlab to simplify the process of predicting Hydro-climatology data using ANN Back Propagation method. Five data for training, testing, and prediction were used. The data, i.e. rainfall, air humidity, duration of shine, temperature, and wind speed are taken from the last ten years with matrix input size *m x n*. Each data is trained 21 times using a combination of the activation functions (logsig, tansig, and purelin) and training methods (traingda, traingdx, and trainrp). The result of the training data was that the logsig function and trainrp on each layer are the best formulas in conducting training, testing, and predictions with an accuracy of 99.71%. This result is obtained from parameter settings including epochs of 1000, learning rate of 0.7, goal error of 0.0001, and training steps of 1.

## 1. Introduction

Forecasting is an activity to estimate what will happen in the future by using conditions or data in the past [1][2]. Forecasting is widely used in almost all agencies or government institutions to determine policies that must be taken based on previous data or facts. Therefore, various forecasting methods are used according to the types of data available. The output of forecasting can be in the form of predictive data in the future or a mathematical model constructed with the method so that it is easier to see the patterns that occur [3].

Forecasting is very important to do in preparation of and overcoming various problems that may occur in the future. It is also a characteristic of a forecasting method indicating that predictive results must be truly accurate. Forecasting with multiple data by weighting on each input network is highly recommended as an effort to reduce the results of improper forecasting [4].

Today many forecasting methods displaying a minimal error rate are known. The input data, however, is still single, meaning that it is unable to simulate multiple data. Therefore, the Artificial Neural Networks (ANN) Back Propagation has adopted the network with multiple inputs [5]. So the results of the predictions obtained are very good because each data is treated through training and testing data by weighting each neuron (network) before the predicted output is generated.

However, it is necessary to experiment with a combination of activation functions and training methods for training and testing various types of data. Therefore, the research team aimed to compile a combination of training methods and activation functions owned by ANN Back Propagation for hydro-climatology data simulations to obtain a truly reliable and accurate network in each experiment with other data in the future. In Matlab, an NNTools is available for forecasting. However, itstill has some shortcomings in terms of attributes or parameters of accuracy [6]. Due to that, the initial step conducted by the team in this research is developing a Matlab Graphical User Interface (GUI) with various attributes according to the ANN Back Propagation algorithm, to make it easier for the team or user to simulate data in large numbers of cases.

## 2. Method

This section focuses on two main discussions, namely data and accuracy parameters. Data used for training and testing are (1) hydrological data (rainfall), and (2) climatological data including wind speed, air humidity, duration of sunlight, and temperature. The data taken from 2008-2017 was sourced from the Central Statistics Agency of West Nusa Tenggara Province.

The accuracy parameters used in this forecasting consist of Mean Absolute Deviation (MAD), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE). The formulas are as follows:

$$MAD = \frac{\sum_{t=1}^{n}\left|X_t - F_t\right|}{n} \tag{1}$$

$$MSE = \frac{\sum_{t=1}^{n}(X_t - F_t)^2}{n} \tag{2}$$

$$MAPE = \left(\frac{1}{n}\right)\sum_{t=1}^{n}\left|\frac{X_t - F_t}{X_t}\right| \times 100\% \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(X_t - F_t)^2}{n}} \tag{4}$$

Where $\boldsymbol{X_t}$ is the actual data in period-t, $\boldsymbol{F_t}$ is the forecasting value in period-t, $\boldsymbol{n}$ is the amount of data, and $\boldsymbol{t}$ is the time series used [5] [7].

The process of training, testing, and predicting hydro-climatology data using ANN Back Propagation is presented further in the following flowchart. Based on Figure 1, it can be seen that training and testing are carried out 21 times using data from 2008-2016 to predict the 2017 data. In addition, the prediction phase of the 2008-2017 data training and testing was used to predict the 2018 and 2019 data.

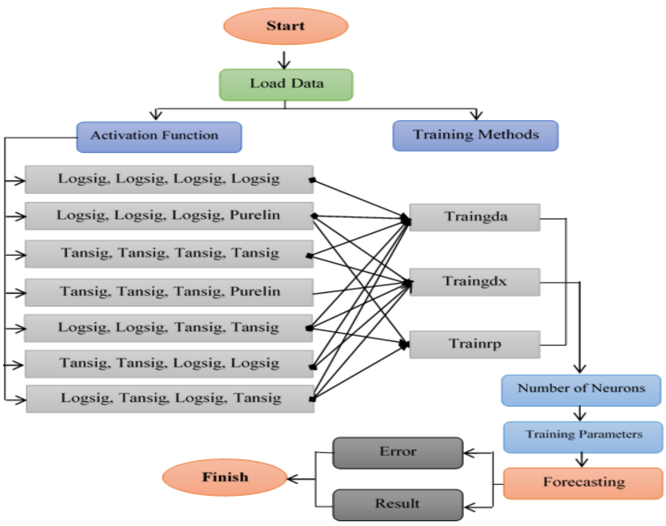


**Figure 1**. Flowchart of training, testing, and forecasting of ANN back propagation

## 3. Result and Discussion

### *3.1 Network Construction*

In the construction phase of the ANN Back Propagation network, we use the amount of data that becomes the input matrix measurement. In the training and testing phase, the data used is 9 (nine) years and each year consists of 12 (twelve) months, making the measurement of the input matrix data 9 x 12 = 108 data, while the prediction phase of the data is 10 (ten) years, making the data of the input matrix 10 x 12 = 120 data. Because we use 2 (two) screens hidden, the amount of data in the hidden 1 screen are 10 data and in the hidden 2 screen are 5 data. The ANN Back Propagation network obtained in this case is like Figure 2.

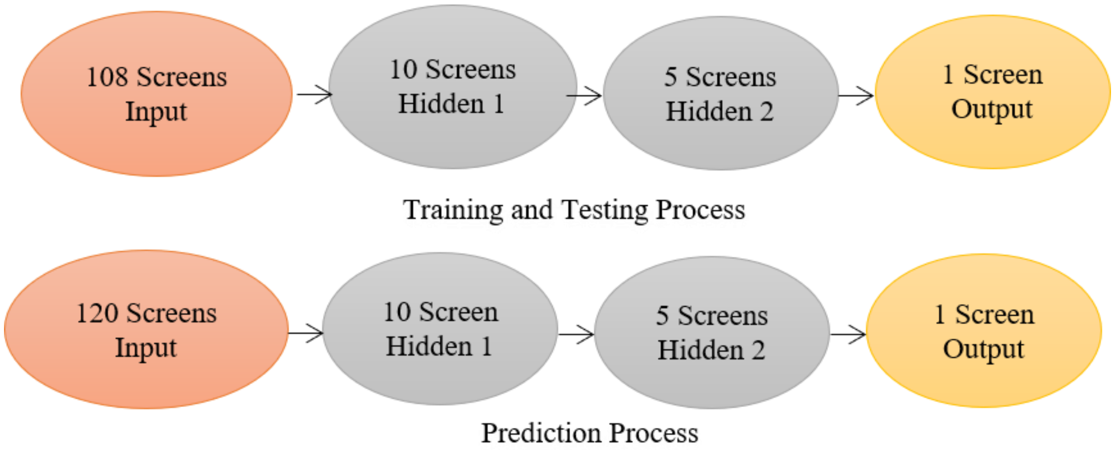


**Figure 2**. ANN back propagation training and testing network construction

### *3.2 Training & Testing*

The architecture design of ANN Back Propagation is done to determine the best architecture with certain parameter settings through training and testing of previously shared data. The architectural parameters used in this study are as follows:

| | |
|---|---|
| Number of Neurons: | |
| Layer Input | : 120 (prediction) and 108 (training & testing) |
| Layer Hidden 1 | : 10 |
| Layer Hidden 2 | : 5 |
| Layer Output | : 1 |
| Activation Function | : logsig, tansig, purelin |
| Algorithm Training | : traingda, traingdx, trainrp |
| Setting Parameter: | |
| Max. Epoch | : 1000 |
| Error (Goal) | : 0,0001 |
| Learning Rate (LR) | : 0.7 |
| Momentum | : 0.9 |
| Decrease ratio LR | : 0.7 |
| Increase ratio LR | : 1. 05 |

Based on the results of the training and testing of the five hydro-climatological data, each with 21 trials, the results with the lowest error rate are shown in Table 1.

IOP Conf. Series: Earth and Environmental Science **413** (2020) 012008 doi:10.1088/1755-1315/413/1/012008

**Table 1**. Result of training and testing

| Data | Experiment | MAD | MSE | RMSE | MAPE | R | Accuration |
|---|---|---|---|---|---|---|---|
| Rainfall | 3 | 3.83 | 39.24 | 6.26 | Inf | 0.99891 | 99.78% |
| Wind speed | 3 | 0.04 | 0.009 | 0.09 | 1.14 | 0.99854 | 99.70% |
| Humidity | 3 | 0.07 | 0.03 | 0.17 | 0.09 | 0.99862 | 99.72% |
| Duration of Sunlight | 3 | 0.45 | 0.52 | 0.72 | 0.63 | 0.99865 | 99.73% |
| Temperature | 3 | 0.03 | 0.002 | 0.05 | 0.09 | 0.99844 | 99.69% |

The results in Table 1 were obtained from trial no. 3 (three) of the 21 (twenty one) experiments carried out. All five data that of the training and testing produces the same results for the activation function and training method, and the best activation function of all screens is logsig, while the training method used is trainrp.

*3.3 Prediction*

The prediction of hydro-climatology data in 2018 was conducted after finding the right training method and activation function with the highest level of accuracy. The prediction results are presented in Table 2.

**Table 2**. Prediction result of hydro-climatological data in 2018

| Month | Rainfall | Wind speed | Air Humidity | Duration of Shine | Temperature |
|---|---|---|---|---|---|
| January | 217.66 | 3.38 | 82.84 | 54.7 | 30.65 |
| February | 143.68 | 3.66 | 80.31 | 39.85 | 30.90 |
| March | 165.03 | 3.94 | 81.79 | 52.96 | 30.92 |
| April | 279.19 | 3.60 | 86.85 | 53.15 | 31.97 |
| May | 196.42 | 3.40 | 83.44 | 64.46 | 32.52 |
| June | 192.78 | 3.19 | 83.14 | 58.45 | 31.70 |
| July | 83.19 | 3.23 | 77.92 | 89.38 | 31.27 |
| August | 97.45 | 3.46 | 80.09 | 88.21 | 31.61 |
| September | 208.51 | 3.31 | 85.87 | 86.72 | 31.92 |
| October | 147.17 | 4.06 | 82.81 | 86.17 | 31.79 |
| November | 328.88 | 3.43 | 84.53 | 80.86 | 31.87 |
| December | 224.90 | 3.52 | 77.74 | 84.46 | 31.81 |

Based on Table 2 above, it is obtained (1) average rainfall of 190.41 mm with the maximum volume occured in November and the minimum in July; (2) average wind speed of 3.515 knots with the maximum speed occured in October and the minimum in June; (3) average air humidity of 82.28% with the maximum humidity occurred in April and the minimum in July; (4) average duration of sunlight of 69.95% with the maximum occured in July and the minimum in February; and (5) average temperature of 31.58$^{o}$C with the maximum occured in May and the minimum in January. Further approaches of the actual data and forecast are available in Figure 3-7.

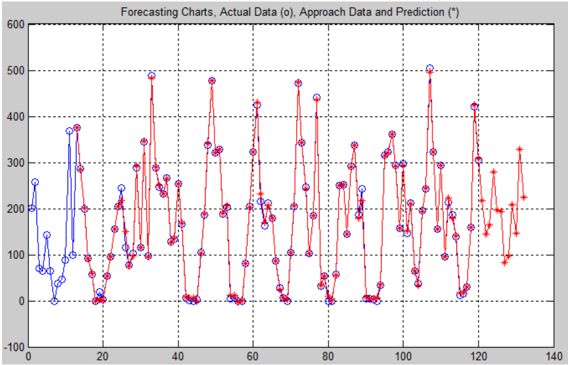


**Figure 3**. Approaches of actual data and forecasts for rainfall

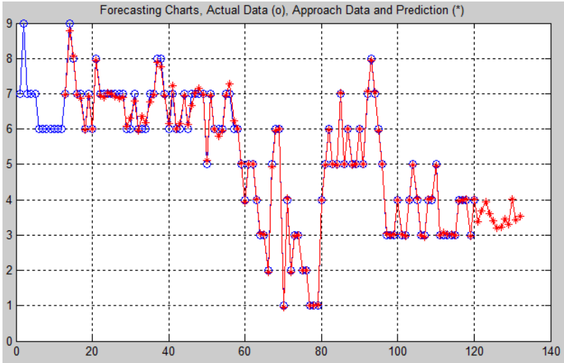


**Figure 4**. Approaches of actual data and forecasts for wind speed

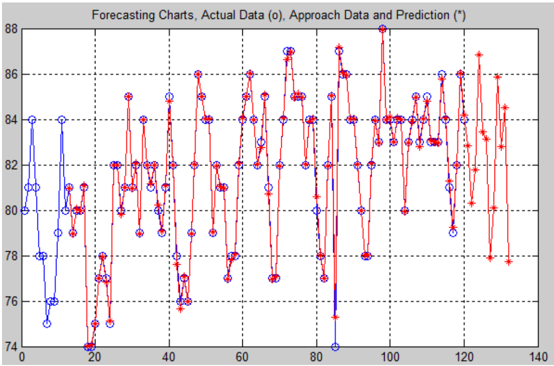


**Figure 5**. Approaches of actual data and forecasts for air humidity

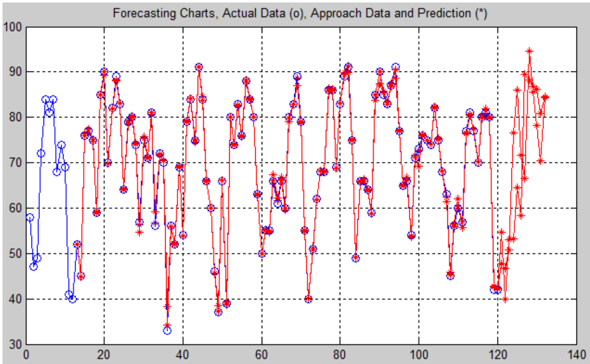


**Figure 6**. Approaches of actual data and forecasts for duration of sunlight

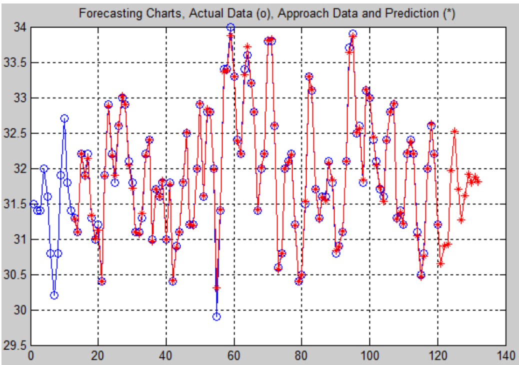


**Figure 7**. Approaches of actual data and forecasts for temperature

The output in the form of the actual data and forecast in Figure 3, Figure 4, Figure 5, and Figure 7 is obtained by setting the maximum epoch of 1000. However, the forecast for rainfall data is completed at 522 epoch and gradient is 0.110, the wind speed data is completed at 587 epoch and gradient is 0.247, the air humidity data is completed at 349 epoch and gradient is 0.000444, the duration of sunlight data is completed at 469 epoch and gradient is 0.346, the temperature data is completed at 450 epoch and gradient is 0.000960.

## 4. Conclusion

The activation function, training method, and the number of neurons for each network in the ANN Back Propagation greatly determine the outcome of the prediction. It can be seen from the 21 experiments that have been carried out with a combination formula between the three components. The activation function of logsig and the trainrp method are the most accurate combination in producing the smallest errors both MAD, MSE, RMSE, MAPE, and performance (R) values. It is evident from the results of the hydro-climatology data simulation that the average accuracy rate is 99.71%.

**Acknowledgement**
The team thank the RI Ministry of Research, Technology and Higher Education, the Research Partner Team from the State University of Malang, LPPM of Muhammadiyah University of Mataram, for their guidance and assistance in making it possible for carry out this research properly.

**References**

[1] Syaharuddin, et al, “Calculus Problem Solution and Simulation Using GUI of Matlab” *International Journal of Scientific & Technology Research*, vol. 6, no. 09, pp. 110-114, 2017.
[2] L. Sucipto and Syaharuddin, “Aplikasi GUI Matlab Dalam Peramalan Data IPM Provinsi NTB”, *Jurnal Register,* vol. 4, no. 2, pp. 114-121, 2018.
[3] A. Sudarsono, “Jaringan Syaraf Tiruan Untuk Memprediksi Laju Pertumbuhan Penduduk Menggunakan Metode Back Propagation”, *Jurnal Media Infotama*, vol. 12, no. 1, Februari 2016.
[4] N. Kourentzes, D. Barrow and F. Petropoulos, “Another Look at Forecast Selection and Combination: Evidence From Forecast Pooling”. *International Journal of Production Economics*, vol. 209, no. 3, pp. 226-235, 2019.
[5] M.I Irawan, et al, “Intelligent Irrigation Water Requirement System Based on Artificial Neural Networks and Profit Optimization for Planting Time Decision Making of Crops in Lombok Island”, *Journal of Theoretical and Applied Information Technology,* vol. 58, no. 3, pp. 657-671 2013.
[6] C. Reddy, S. Ravi, and Giriraja C. V, “Baby Monitoring Through MATLAB Graphical User Interface”, *International Journal Of Scientific & Technology Research*, vol. 3, no. 7, pp. 174-177, July 2014.
[7] P. Wallström and A. Segerstedt, "Evaluation of forecasting error measurements and techniques for intermittent demand", *International Journal of Production Economics,* Vol. 128, no. 2, pp. 625-636, December 2010.