International Journal of Engineering & Technology, 7 (4.35) (2018) 182-185

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Monthly Rainfall Prediction Model of Peninsular Malaysia Using Clonal Selection Algorithm

N.S.Noor Rodi[1], M.A.Malek[2], A.R.Ismail[3]

[1]*College of Engineering, Universiti Tenaga Nasional, Malaysia*
[2]*The Institute of Energy, Policy, Research (IEPRe), Universiti Tenaga Nasional, Malaysia*
[3] *Department of Computer Science, KICT, Universiti Islam Antarabangsa, Malaysia*
**Corresponding author E-mail: syazwanirodi@gmail.com*

Abstract

Nowadays, various algorithms inspired by natural processes have been extensively applied in solving engineering problems. This study proposed Artificial Immune Systems (AIS), a computational approach inspired by the processes of human immune system, as an algorithm to predict future rainfall. This proposed algorithm is another alternative technique as compared to the commonly used Statistical, Stochastic and Artificial Neural Network techniques traditionally use in Hydrology. Rainfall prediction is pertinent in order to solve many hydrological problems. The proposed Clonal Selection Algorithm (CSA) is one of the main algorithms in AIS, which inspired on Clonal selection theory in the immune system of human body that includes selection, hyper mutation, and receptor editing processes. This study proposed algorithm is utilised to predict future monthly rainfall in Peninsular Malaysia. The collected data includes rainfall and other four (4) meteorological parameters from year 1988 to 2017 at four selected meteorological stations. The parameters used in this analysis are humidity, wind speed, temperature and pressure at monthly interval. Four (4) meteorological stations involved are Chuping (north), Subang Jaya(west), Senai (south) and Kota Bharu (west) represented peninsular Malaysia. Based on the results at testing stage, it is found that the trend and peaks of the hydrographs from generated data are approximately similar to the actual historical data. The highest similarity percentage obtained is 91%. The high values of similarity percentage obtained between simulated and actual rainfall data in this study, reinforced the hypothesis that CSA is suitable to be used for prediction of continuous time series data such as monthly rainfall data which highly variable in nature. As a conclusion, the results showed that the proposed Clonal Selection Algorithm is acceptable and stable at all stations.

Keywords: *artificial immune system; clonal selection algorithm; meteorological; Peninsular Malaysia; rainfall prediction*

1. Introduction

In Malaysia the weather is high climate, humid and rainy throughout the year. Temperatures are high and stable, with a slight decrease between November and January and a slight increase between March and August. Climate condition is a natural phenomenon occurs from the climate systems and affect the environment in all aspects. The rains in Malaysia are numerous and continuous throughout the year. Accurate information on rainfall is essential in planning and management of the country's water resources. In addition, rainfall events influence the traffic, sewer systems and various human activities especially in the urban areas. In Peninsular Malaysia, it is difficult to find an area with rainfall lower than 100 mm (4 inch) per month. Aside from being one of the main sources of life, water from rainfall can resolve dryness, replenish groundwater, clears the air thus maintain the water supply for domestics and commercials, industrial, and agricultural [1]. Thus, one of the greatest challenges in operational hydrology is rainfall prediction. Prediction is a challenging task that too for weather is even more complex. Nevertheless, mitigation measures should be navigated in order to overcome disasters as a result of continuous high intensity rainfall such as floods and landslides. Therefore, this study is significant important for management water resources and mitigation related flood disaster especially in Malaysia. Successful development of rainfall model will provide useful parameter for any hydrological model, which will eventually
lead to the development of the inundation model. The challenge would then be on the selection of the most reliable model [2].

2. Artificial Immune System (AIS) and Clonal Selection Algorithm (CSA)

Application areas that have been addressed by AIS can be summarized as [4]:

1. Learning (clustering, classification)
2. Anomaly detection (fault detection, computer and network security applications).
3. Optimization (continuous and combinatorial).

AIS is an algorithm and system inspired by the biological immune system .It is an adaptive systems inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving. The main function of a biological immune system is to defend human body from foreign elements known as antigens. Antigens can be parasites, bacteria or viruses. These antigens will then be eliminated or neutralized by the immune system [8].

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The immune system has many attractive features which can be implemented in machine learning algorithms for classification, clustering, and regression. In this study, the proposed prediction model uses Clonal selection mechanism to create specialized antibodies, ABs (immune memory cells with suitable shaped paratopes) for recognition. AGs (antigens) represent fragments of load time series where the actives Abs constructs the fragments.
Recognition abilities of ABs depend on their paratopes. Time series represent by patterns expressed in ABs and AGs simplifies the problem of predicted multiple seasonal nonstationary load time series by filtering of trend, annual, weekly cycles and unifying daily cycles. This leads to the simplification of relationship between input and output variables, and consequently, to increase the model accuracy [5]. Immune system has two types of immunity where are innate and adaptive, formed by two main lines of resistance in an immune system [6]. It is capable to virtually recognize any pathogen, foreign or molecules eliminated from the body [7]. This theory is illustrated in Figure 1.

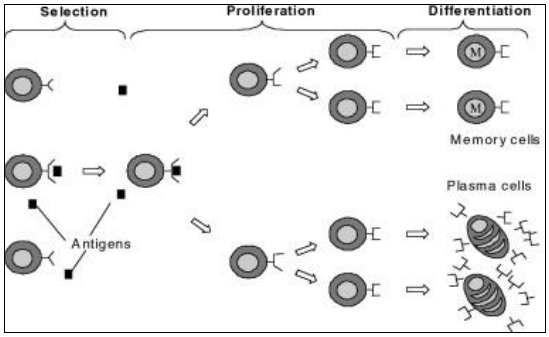


Fig. 1: Multiple layers in biological immune system [12]

An antibodies are available in two layers which antigens will meet an innate immune layer for the first response before it meets the adaptive immune layer for the second response. The area of AIS greatly influence applications in engineering with the development of computational intelligence in recent years. Other than Negative Selection theory and Immune Network, the fundamental concept of Clonal selection is used to explain the behaviors of modern immune system. The biological principles like clone generation, proliferation and maturation are combined into an artificial immune based algorithm termed as CSA [10]. CSA is inspired from Clonal selection principle which established the idea that only those cells that is capable to recognize an antigenic stimulus will proliferate and differentiate onto effector cells, thus being selected against those that do not [8]. It is a population of antibodies based on stochastic process. CSA is extensively used in artificial immune systems based on optimization method in pattern recognitions and multimodal optimization problems with binary representation of variables. The implementation of CSA in AIS involves the following steps [9].
• initialization of antibodies
• cloning and selection (proliferation and differentiation on the encounter of cells with antigens)
• maturation and diversification of antibody types by carrying out affinity maturation process through random genetic changes
• removal of differentiated immune cells which posses low affinity antigenic receptors.
Liliya mentioned that the forecasting model on the base of clonal selection algorithm was allow to increase the forecasting accuracy, significantly reduce the selection time of required analytical dependence and this algorithm was recommended for making short term forecasts [13].
In the process of optimization with AIS based CSA, affinity is the objective function evaluation and constraint satisfaction. Here the constraints are represented by antigens while the constraint satisfaction is attained by antibody – antigen affinity. In other words, higher affinity will give higher constraint satisfaction. Figure 2 shows the Clonal selection theory for human body systems.

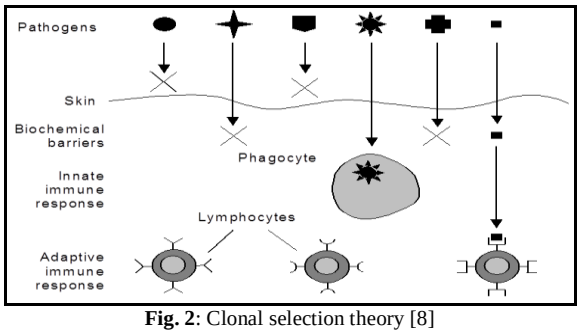


Fig. 2: Clonal selection theory [8]

3. Methodology

This study designs the structure of CSA using MATLAB 7.12 (R2017a). The data used were obtained from four (4) meteorological stations in Peninsular Malaysia managed by Malaysia Meteorological Department namely Chuping (Perlis), Subang (Selangor), Senai (Johor) and Kota Bharu (Kelantan). The stations that located in urban area were choose to minimize the other natural conflict issues and gaps. The meteorological parameters such as rainfall, wind speed, temperature and pressure were set as parameters or factors that influence the formation of rainfall called '*antigen*'.
These variables were selected because they are independent and influence precipitation. Temperature is a measure of the ability of the atmosphere and water to receive and transfer heat from other bodies. Humidity is the water vapor content of the air. As normal relative, if the humidity of the air below cloud is low, the evaporation can significantly reduce the amount of water in falling raindrops, therefor reduced rainfall.
The flowcharts for rainfall prediction methodology was explained in Figure 3.

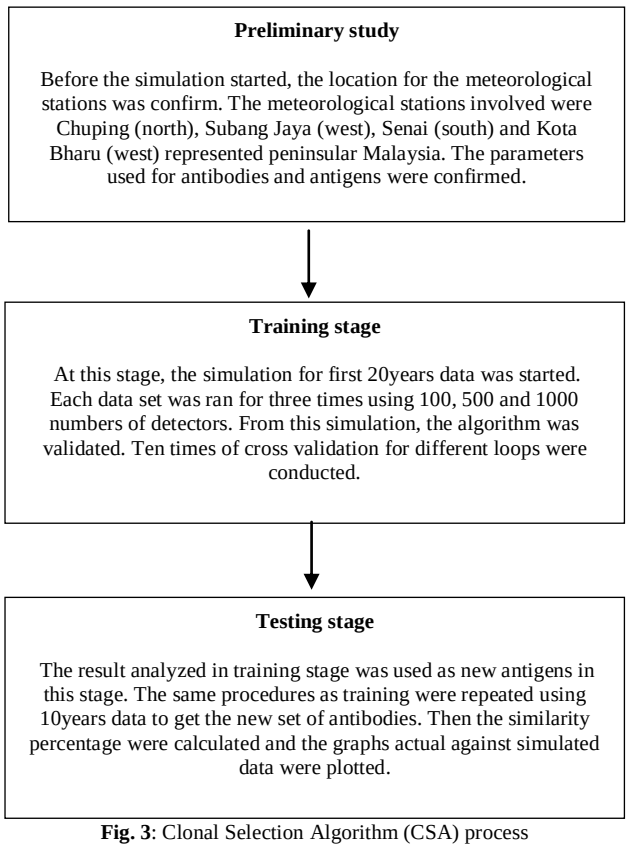


Fig. 3: Clonal Selection Algorithm (CSA) process

Rainfall data is use as '*antibodies*'. Rainfall data used in this model are divided into two stages, which are training stage and testing stage. Data from year 1988 until 2007 were used as training stage, while data from year 2008 until 2017 were used as testing set.

In figure 3, after preliminary study completed, the steps begin with training data input, testing and prediction processes. Historical rainfall data at four stations were selected as antibodies for the developed algorithm. To determine antigens, four meteorological parameters for each station were set as antigens which are humidity, wind speed, pressure and temperature. In the affinity measure, the number of clones generated per antibody depend on the affinity or fitness value. The aim of this study is to obtain higher number of clones generated for antibodies with larger fitness value and less number of clones generated for antibodies with smaller fitness value. The fitness value is calculated using *Euclidean* Distance (D) equation as shown in Equation 1.

$$D = \sqrt{\sum_{i=1}^{L}(Ab_i - Ag_i)^2} \quad (1)$$

Where,
Ab is antibodies
Ag is antigen

Here, cells which are capable to recognize an antigenic stimulus will proliferate and differentiate into effector cells, thus being selected against those that do not [8]. These selected Abs are then cloned. When the Abs was found to have low affinity value, cloned process were repeated. These iterations are named as mutation. The number of mutation is named as detectors. In this study, the number of detector used is 100 since earlier study [11] has proven that simulation conducted using 100, 500 and 1000 numbers of detector will produce the same results.

In this study, in order to generate synthetic future rainfall data, the pattern value of historical meteorological data for 10 years duration were calculated using the values of accuracy, shortest and f*inDetcs* obtained from the simulation of the proposed model. The values for accuracy are found using Equation 2.

$$Accuracy = finDetcs\text{-}Antibodies \quad (2)$$

The values for *finDetcs* are obtained from Equation 3:

$$FinDetcs = [shortest + rand\ (1,1)/10\ ] / [(v+(x-1)*5] \quad (3)$$

Each average value of accuracy, shortest and *finDetcs* are further multiplied with Forecasting Rates (FR) where its summation is the value obtain from the cloning process. Value of FR is obtained from Equation 4. The sum of average *finDects*, accuracy and shortest is multiplied with the value of FR.

$$FR = \frac{\sum Actual value - Simulated value}{No.of\ data} \quad (4)$$

Figure 4 shows the flowchart for clonal selection algorithm works.

4. Results and Discussion

After training process, FR for each station is obtained. In this study, it was found that the prediction factor obtained at each chosen meteorological station varied. This prediction factor was further used to build a group of antigens at testing stage. The prediction factors obtained at each station is shown in Table 1.

Table 1: Prediction factor

Meteorological Station	Prediction Factors
Chuping	0.04
Subang	0.03
Senai	0.04
Kota Bharu	0.05

Based on the results, it showed that the highest value of prediction factor is at Kota Bharu. These value will be used at both testing and prediction stage.

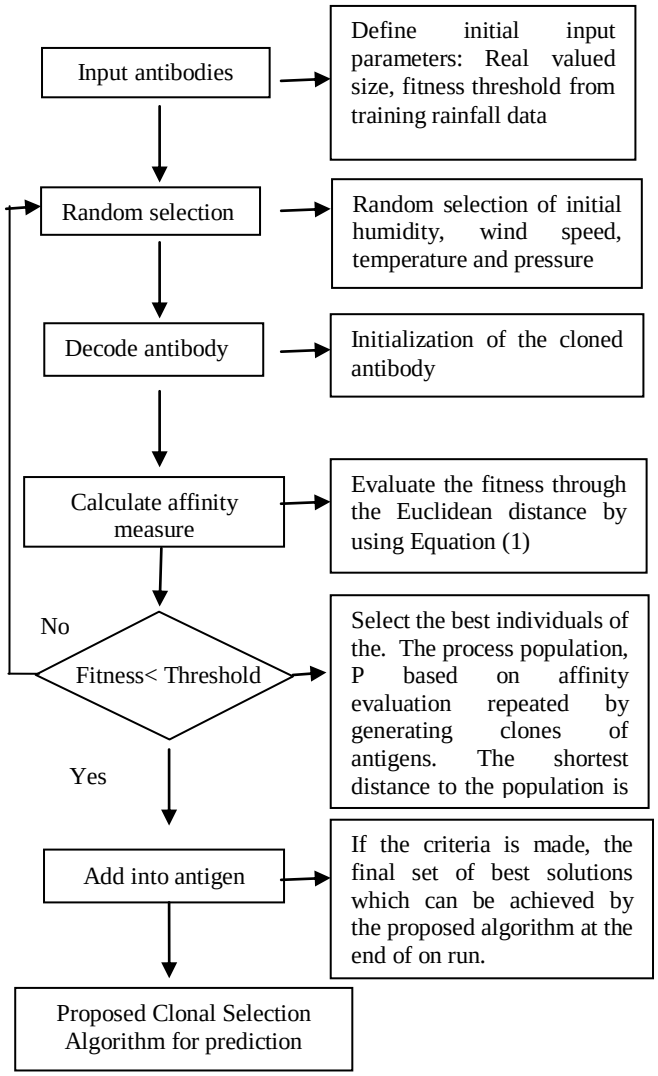


Fig. 4: Flowchart for rainfall prediction using clonal selection algorithm

Figure 4 a, b, c, d illustrate prediction of rainfall during testing stage within 120 months durations from year 2008 at Chuping, Subang, Senai and Kota Bharu stations. At all cases, it was found that the generated simulation rainfall data was similar from the actual data pattern.

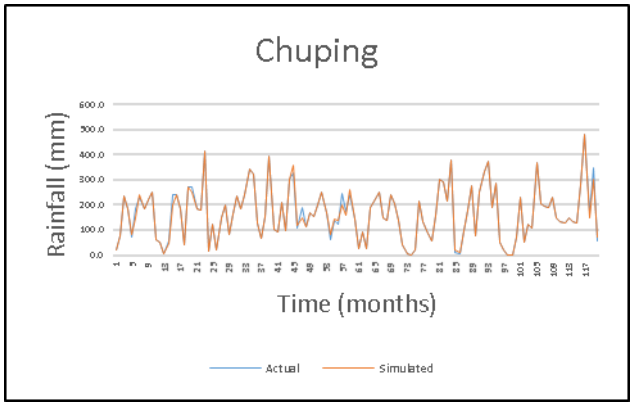


Fig. 4a: The comparison between actual and simulated rainfall at Chuping during testing stage

Figure 4a shows similar trend between actual and simulated rainfall data at Chuping station. Based on the simulation conducted, the similarity percentage between actual and predicted rainfall is found at 86.7 %. The highest value of simulated rainfall is 489mm.

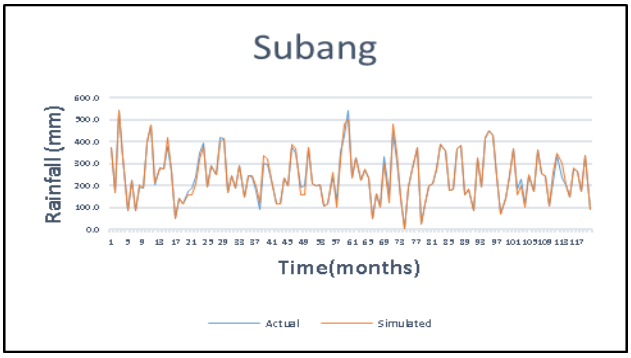


Fig. 4b: The comparison between actual and simulated rainfall at Subang at testing stage

Figure 4b shows almost similar trend between actual and simulated values for rainfall at Subang station. The similarity percentage between actual and predicted rainfall is 85.9 %. The highest value of simulated rainfall was obtained at 548mm.

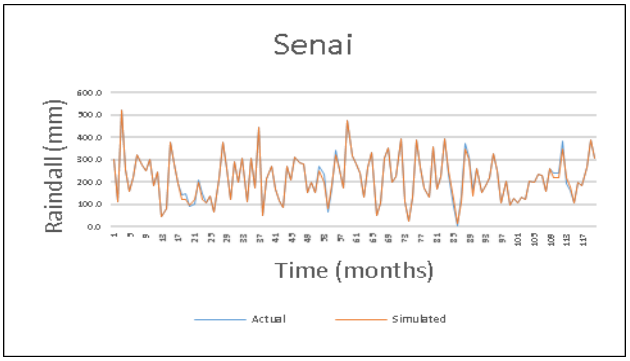


Fig. 4c: The comparison between actual and simulated rainfall at Senai at testing stage

Figure 4c shows similar trend between actual and simulated values for rainfall at Senai station. The similarity percentage between actual and predicted rainfall was found at 89 %. The highest value of simulated rainfall was 521mm.

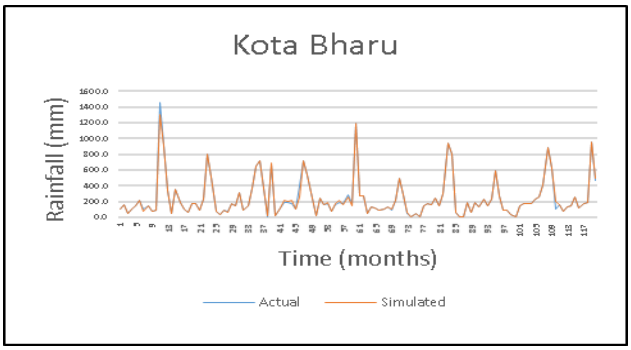


Fig. 4d: The comparison between actual and simulated rainfall at Chuping at testing stage

Figure 4d shows similar trend between actual and simulated values for rainfall at Chuping station. The similarity percentage between actual and predicted rainfall was found at 90.8 %. The highest value of simulated rainfall was 1321mm.

Based on simulations conducted in this study, the results showed that Subang station has the lowest similarity percentage at 85.9% while Chuping station in Kota Bharu showed the highest similarity percentage at 90.8%. When refer to forecasting rate value, this algorithm gives higher percentage of similarity when the forecasting rate is high. The highest value for forecasting rate for Kota Bharu can give higher similarity percentage. This result also influenced by the trend of data for Kota Bharu, which shows peak value of rainfall at November-December for each year. When the trends are repeating the same, the antibodies and antigens easily capture the trend and produce a solid bonding.

5. Conclusion

This research proposed an alternative method to predict future rainfall, as compared to the commonly used Statistical, Stochastic and Artificial Neural Network techniques traditionally use in Hydrology. The CSA method which adapts the biological immune system is used in the proposed prediction model of this study. The results obtained from the testing stage showed that the percentage of accuracy between actual and simulated rainfall data is in the range of 85.2% till 90.8%. This study also proves that an algorithm inspired by clonal selection theory in artificial immune system can be applied successfully for prediction of continuous time series data such as monthly rainfall data which highly variable in nature.

Acknowledgement

This study is financially supported by Yayasan Tenaga Nasional. The authors thanked Malaysia Meteorological Department (MetMsia) for access to data.

References

[1] Nur Farhana Hordria , Siti Sophiayati Yuhanizb , Kamilia Kamardin, (2015): Prediction Rainfall Distribution Using Artificial Neural Networks for Johor Rivers Open International. Journal of Informatics (OIJI) Vol 3 Issue 1

[2] Wardah Tahir, Noratina Alias, Lokman Hakim Ismail and Junaidah Ariffin, (2013): Data for Hydrological Use, International Journal of Integrated Engineering, Vol. 5 No. 1

[3] R. Nalinggam1 W. Ismail1 J.S. Mandeep, (2011): Comparison of Rain Attenuation Prediction Models with Ground Measurement Data for Penang, IET Microwaves, Antennas & Propagation

[4] Hart. E, and Timmis J., (2008): Application areas of AIS: The Past, The Present and The Future, Journal of Application Software Computer., Vol. 8, No. 1, pp. 191–201

[5] Grzegorz Dudek (2017) Artificial Immune System with Local Feature Selection for Short-Term Load Prediction, (2017), IEEE Transactions on Evolutionary Computation, Vol. 21, No. 1

[6] Timmis, Jonathan, (2002): Artificial Immune System; A Novel Data Analysis Inspired by the Immune Network Theory, Dissertation, University of Wales

[7] Anil S. Steven H., Stephanie F. (1998). Principles of a Computer Immune System, 1997 New Security Paradigm Workshop, pp. 75-82

[8] Timmis J. and de Castro L.N. (2002): Artificial Immune System: A New Computational Intelligence Approach.

[9] Muhammad Rozi Malim, Ahmad Tajudin Khader and Adli Mustafa, (2006): Artificial Immune Algorithms for University Timetabling, Faculty of Information Technology & Quantitative Sciences, UiTM, Shah Alam

[10] Balusu S. R. and Vaisakh. K., (2013): Adaptive Clonal Selection Algorithm for Solving OPF Problem with Emission Constraints, Annual IEEE India Conference (INDICON)

[11] N.S.Noor Rodi, A.R.Ismail, M.A.Malek (2014): A Clonal Selection Algorithm Model for Daily Rainfall Prediction. Water Science and Technology Journal

[12] Liu Cheng, Tian Yi- Mei, Wang Xiao- Hua, National(2011) Water Pollution Control and Management of Science and Technology Project of China

[13] Liliya A.Demidove (2014) Time Series Forecasting Models on the Base of Modified Clonal Selection Algorithm, IEEE