FLOOD FORECASTING USING [illegible] NETWORK (ANN) IN MARAN, PAHANG

NUR ATIYAH DINIE BINTI MAT ARIFIN

A report submitted in partial fulfillment of the
requirements for the award of the degree of
B. ENG (HONS.) CIVIL ENGINEERING

Faculty of Civil Engineering & Earth Resources
Universiti Malaysia Pahang

JUNE 2014

ABSTRACT

Maran is located at district of the same name between Temerloh and Kuantan, Pahang which is surrounded by remote forest and palm oil plantations. In December 2013, a worst flood occurred in Maran and had cause loss of lives and massive damages to the area. In this study, the data from Lubok Paku Station was used for analyzing data for flood forecasting. Lubok Paku Station in district Maran was one of the stations that are located along Sungai Pahang. In order to reduce the possibility for flood events to occur again in Maran, a reliable water level forecasting models is extremely important. Developing water level forecasting models is essential in water resources management and flood prediction. Accurate water level forecasting aids in achieve proficient and ideal utilization of water resources and minimize flooding damages. Conventional linear modelling forecasting model approach such as regression mostly provided relatively poor accuracy for forecasting peak inflow events of floods or drought. An artificial neural network or known as ANN is a computing model with a non-linear mathematical approach that has been proven in many forecasting studies. Improving the ANN computational approach could help produce accurate forecasting results. ANN also has the ability to map inputs and outputs pattern and copy different elements experienced in the data. In this study, ANN Modelling is used to analyze the water level data. Different model architectures are examined based on the length of the forecasting period, epochs and training data. The data training was conducted using six ANN architectures. The performance of data training and data validation were evaluated using the coefficient of efficiency (NSC) and root mean square (RMSE). The results showed a strong performance of forecasting accuracy which can provides a significant tool for the authorities to take proper actions to minimize the damage of flooding.

ABSTRAK

Maran terletak di daerah dengan nama yang sama di antara Temerloh dan Kuantan, Pahang yang dikelilingi oleh hutan pedalaman dan ladang kelapa sawit. Pada bulan Disember 2013, banjir paling buruk telah berlaku di Maran dan ia menyebabkan banyak kehilangan nyawa dan kerosakan besar-besaran di kawasan ini. Dalam kajian ini, data dari Stesen Lubok Paku telah digunakan untuk menganalisis data bagi ramalan banjir. Stesen Lubok Paku di daerah Maran merupakan salah satu stesen yang terletak di sepanjang Sungai Pahang. Untuk mengurangkan kemungkinan bagi kejadian banjir untuk berlaku lagi di Maran, satu model ramalan paras air yang boleh dipercayai adalah sangat penting dalam pengurusan sumber air dan ramalan banjir. Ramalan paras air yang tepat dapat membantu dalam mencapai penggunaan mahir dan sumber air ideal dan mengurangkan kerosakan akibat banjir. Pendekatan konvensional model linear ramalan kebanyakannya memberi keputusan yang kurang tepat untuk ramalan peristiwa aliran masuk air semasa puncak banjir atau kemarau. Rangkaian neural tiruan atau dikenali sebagai ANN adalah model pengkomputeran dengan pendekatan bukan linear matematik yang telah terbukti dalam banyak kajian ramalan. ANN juga mempunyai keupayaan untuk memetakan corak input dan output dan menyalin elemen yang berbeza dalam data. Dalam kajian ini, Permodelan ANN digunakan untuk menganalisis data paras air. Seni bina model yang berbeza diperiksa berdasarkan tempoh masa ramalan, epoch dan data latihan. Data latihan telah dijalankan menggunakan enam seni bina ANN. Prestasi latihan data dan pengesahan data telah dinilai menggunakan pekali kecekapan (NSC) dan punca min kuasa dua (RMSE). Keputusan menunjukkan prestasi ketepatan ramalan yang kukus di mana ia boleh dijadikan alat yang penting bagi pihak berkuasa untuk mengambil tindakan yang betul bagi mengurangkan kerosakan akibat banjir.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
ARMA	Auto-Regressive Moving Average Model
ATF	Activation Transfer Function
COD	Department of Irrigation and Drainage
FKASA	Faculty of Civil Engineering & Earth Resources
MLP-NN	Multi-Layer Perceptron Neural Network
MATLAB	Matrix Laboratory
NSC	Nash-Sutcliffe Coefficient of Efficiency
RMSE	Root Mean Square Error
UMP	University Malaysia Pahang

CHAPTER 1

INTRODUCTION

1.1 Introduction

Flood is an overflow or an accumulation of an expense of water that submerges land. It is also can be refer to the inflow of the tide or the backflow of the river, which occurs at the location where the rivers meet. From a geological perspective, floods are natural consequences of stream flow in a continually changing environment. The streams receive most of their water input from precipitation and the amount that falling in drainage basin varies from day to day. Based on the role of precipitation, the amount and time which precipitation takes place is not constant for each area.

Overall, the water cycle is a balanced system and the reason for the flood to occur is a large amount of precipitation, causing the river to overflow due to not efficient cross section of the river itself. As the amount of water increase, the stream must adjust to its velocity and cross section in order to form a balance.

Most town in Pahang state received heavy rainfall in monsoon period which results in flood in low area especially near the flood plain area. Heavy rainfall in such high area results in increasing water level in the river below their level.

1.2 Problem Statement

Anticipating flood before they occur allows for precautions to be taken and people to be warned, so that they can be prepared in advance for flooding conditions. For example, farmers can remove animals from low-lying areas and utility services can put in place emergency provisions to re-route services if needed. Emergency services can also make provisions to have enough resources available ahead of time to respond to emergencies as they occur.

Flood Forecasting had to be done to prevent the primary effects of flooding include loss of life, damage to buildings and other structures including sewerage systems, and roadways. Flood also frequently damage power transmission and sometimes power generation, which then has knock-on effects caused by the loss of power. This includes loss of drinking water treatment and water supply, which may result in loss of drinking water or severe water contamination. It may also cause the loss of sewage disposal facilities. Lack of clean water combined with human sewage in the flood waters raises the risk of waterborne disease and other disease depending upon the location of the flood.

Since flood occurrences seem to be getting more frequent in recent years, forecasting ability is needed. There is many forecasting modeling that we can use. One of the modeling is based on mathematics statistical method such ad Linear Regression Method and Auto-Regressive Moving Average Model (ARMA) but it is less accurate. So, for this research, we use Artificial Neural Network (ANN) modeling which is a simpler and workable method for flood forecasting.

1.3 Objectives

The main objectives of this study are:

i. To be able to forecast flood events in Maran, Pahang

ii. To identify the usage of Artificial Neural Network (ANN)

iii. To determine and evaluate whether flood forecasting can be done by single data which is water level data.

1.4 Scope of Study

The water level data will be taken from Station *Sg. Pahang di Lubok Paku*, Lubok Paku, Maran, Pahang. The catchment area is 25, 600 km^2. It is located in the latitude of N03° 30' 45" and longitude of E102° 45' 30".

In this research, the water level data will be analyzed by using Artificial Neural Networks (ANN). The method of ANN that will be used is Multi-Layer Perceptron (MLP) which will be run by Matrix Laboratory (MATLAB). By analyzed this data, we can predict the flood events and the relationship between flood and water level data.

1.5 Significance of Study

A flood prediction model can play a key role in providing relevant information of possible impending floods in populated locations. The development of such models can reduce the damage by decreasing the economic and environmental impacts of floods.

The significant for using ANN is because ANN is a simple and workable method. If ANN can provide sufficiently accurate forecasts, the lead time for flood warning can be extended and the flood emergency measures can be planned

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter explains about the components and the elements in this study for any related issue and the definition in a more detailed approach. The comments and facts from other researchers on this topic obtained from journals. A clearer understanding of the research objective, problem statement and the significant of the study are provided in this chapter.

2.2 Flood Forecasting

Floods are caused by many factors: heavy rainfall, highly accelerated snowmelt, severe winds over water, unusual high tides, tsunamis, or failure of dams, levees, retention ponds, or other structures that retained the water (Radhika, 2012). The extremely high or low rainfall of precipitation leading to flood or drought is an example of substantial weather risk.

Flood forecasting is an essential requirement for solving a wide range of scientific and/or management problems related to the design and operation of river systems (Tareghian & Kashefipour, 2007; R. Garcia-Bartual, 2002). There is no doubt that this forecasting can have a great impact on any human activity linked to a river. Flood forecasting is important in predicting its environmental impacts, such as the transport and concentration of pollutants (Atiya et al., 1991).

2.3 Artificial Neural Network (ANN)

There are many models used in flood forecasting, but some are not able to cope with dynamic changes inside the catchments. Some models are sensitive to the initial state of the catchments, some models are too complicated to calibrate and need to have a robust optimization tools (Alvisi et al. 2006; Toth and Brath, 2007) and some models require the understanding of the physical processes (Chauhan and Shrivasta, 2008).

The introduction to ANN to the field of water resource has opened various new approaches in modelling. The ANN are recognized as being universal approximates and are capable of extracting the underlying relationship between any input, or stimulus, and its subsequent output, or response. The ANN composed of layers with many nodes in each layer. Input data are fed into the first layer called the input layer, while the output are taken from the last layer, called the output layer, and the layers in between are hidden layers (Zhang et al. 1998).

ANN offers an important alternative to traditional methods of data analysis and modeling. While different types of neural networks exist, the one that is of interest at the moment of the feed-forward multilayer perceptron (MLP). The basic structure of an MLP is not complicated. It consists of a number of simple processing elements (neurons) arranged in different layers, joined together to form a network (Coulibaly et al. 2000; Joorabchi et al. 2007; Solaimani and Darvari 2008; Turan and Yurdusev 2009).

2.3 Water Cycle

Theoretically, flood is caused by an excess amount of rainfall in a long direction. The ideal percentage of a water cycle balance in order to balance the eco-system to avoid flood is clearly highlighted and as long as Malaysia is on the line, the flood can be avoided. However, there are several factors that causing the water balance system is disturbed. Since Malaysia received a few numbers of rainfalls annually, the water source condition is usually affected by weather conditions and natural water cycle (Strathler, 1997). The water in the earth surface is evaporated to the atmosphere as a vapour and the group of water vapour in the atmosphere produced clouds which finally condensed and precipitate to the earth as rainfall. This process is known as precipitation (Syukri, 2009). Figure 2.3 indicate the illustration of water cycle on the earth surface.

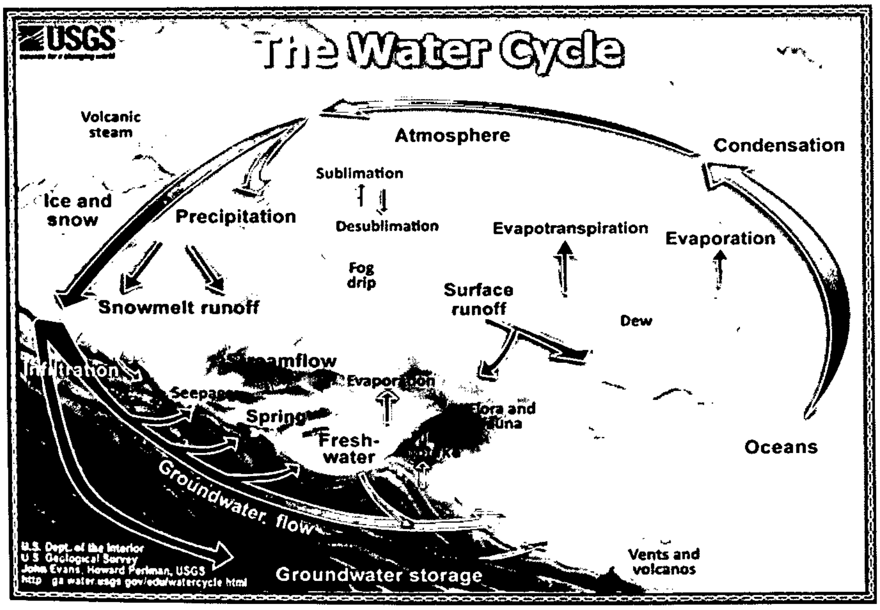

Figure 2.3 Natural Water Cycles (Strathler, 1997)

CHAPTER 3

METHODOLOGY

3.1 Introduction

This research will focus on the assessment on the effectiveness of Artificial Neural Network (ANN) to analyze the data for flood forecasting. In this chapter, the detail explanation about the software used for this research, location of study area and their performance evaluation. The study methodology includes steps that are involved from the beginning until the final stage to achieve the study objective.

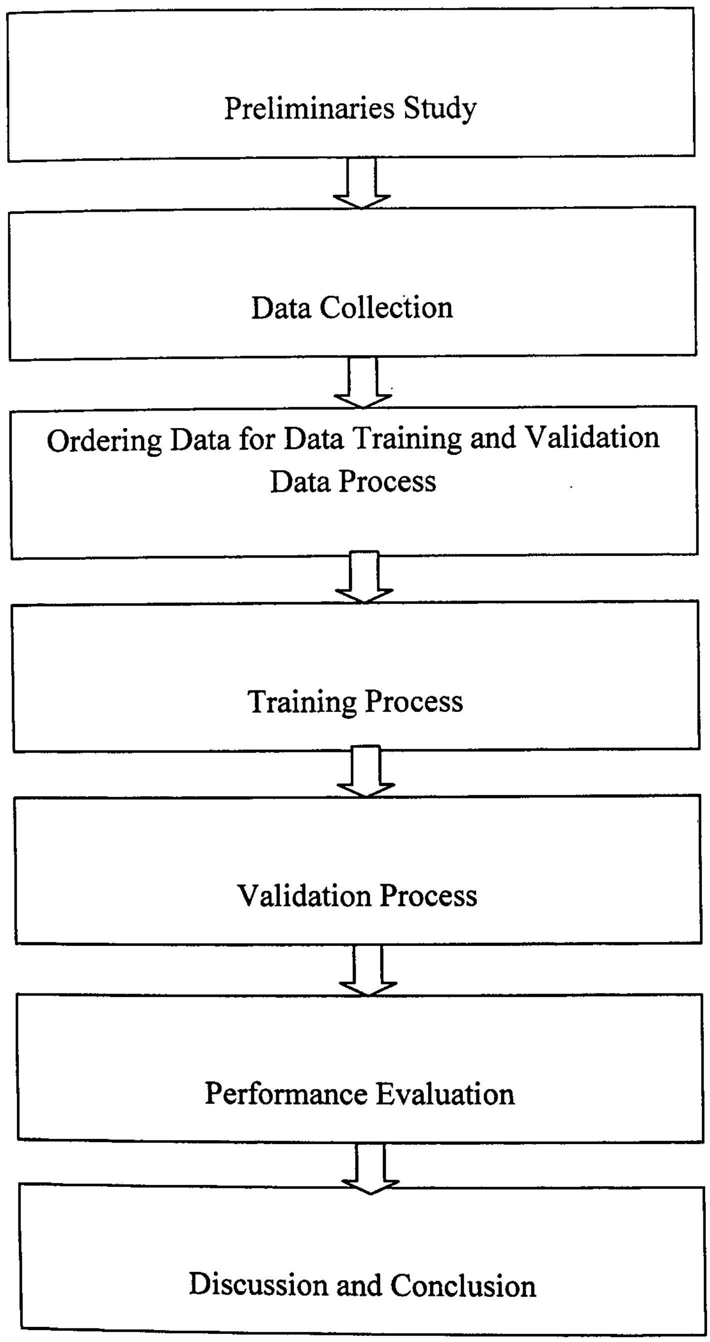

Figure 3.1 Research Methodology

3.2 Location of Study

The location in the present study was Lubok Paku station, which is located along Pahang River, the longest river in Peninsular Malaysia. The study area was shown in Figure 3.2. Water Level data was provided by the Department of Irrigation and Drainage, Ampang, Selangor. The data consists 10 years of past daily water level data from 2003-2013. The data were divided into two; data training and data forecasting. The data training is between 2003 to 2009 and the data forecasting is between 2010 to 2013.

Due to missing past records, the data training and data validation were divided into smaller groups of continuous datasets. Six water level datasets were extracted from 2003 to 2009 for data training and from 2010 to 2013 for data validation. The datasets in this study were normalized based on the following equation:

$$N = \frac{Oi - Omin}{Omax - Omin}$$

Where N is the normalized value, O is the observed value, O_{max} is the maximum observed value and O_{min} is the minimum observed value.

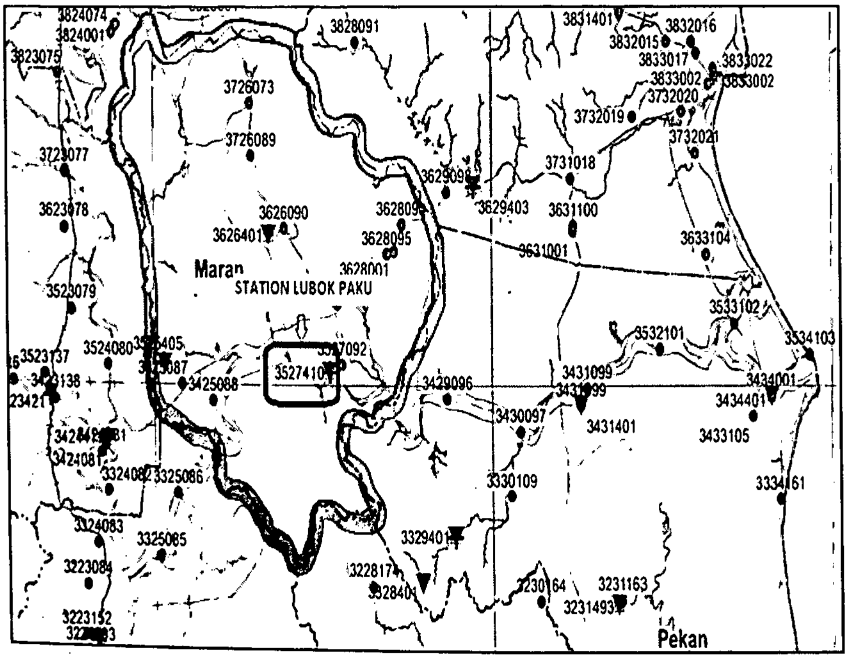

Figure 3.2 Location of Lubok Paku Station

3.3 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is a parallel computing mathematical solutions for solving linear and nonlinear problems. The modeling algorithm of ANN is based on imitation of how human brain performed. ANN is built based on interconnection between input layer, hidden layer and output layer. Data are assigned at the input layer, and the number of data input is depends on the problem of being solved.

The general ANN non-linear equation can be defined as follows:

$$y_1, \ldots y_k = f(x_1, \ldots x_m \ldots c_1, \ldots c_n)$$

where y is the forecasted output, k is the number of forecasted outputs, x is the input, m is the number of data input, c is a constant term and n is the number of constant terms. The constant term c is referred to as the weight in ANN modelling. y can be computed if m, the c values and n are known. The x values are data input. To identify the m, c values and n, it is necessary to understand the ANN modelling.

The modelling of ANN is built based on interconnection between input layer, hidden layer and output layer. Data are assigned at the input layer and the number of data input is depends on the problem of being solved. Before the ANN model can process the input data, the data are usually normalized or scaled to values between1 to -1 or 0 to 1. The processing of data starts from the input layer and propagates to hidden layer.

In the hidden layer, there are neurons that processed the data input and weight of the input to the neuron and convert the inputs into an output through activation function. The output of the activation function will become an input to next layer either another hidden layer or output layer. Sigmoid, linear, step and sign functions are among activation functions that are commonly used in ANN Modeling. An artificial neural neuron is shown on Figure 3.3

At final stage of the modeling, the computed output form the output layer is compared to an output of an actual data, and adjustments are made to neuron weights in a backward form. This process is called back propagation learning processes. Each adjustment to neuron weights and computation from an input to an output layer is called an epoch or iteration. The number of epoch depends on the number chosen. Attempt is made so that the computed output is to have a similar result of the actual output data. The whole process of computation and learning is called a training process. Once training is completed, the knowledge acquired is stored in the neuron weights. Based on this knowledge, forecasted data are assigned as input and computation process propagated from the input to the output layer through the hidden layer. The computed output is called forecasting result.

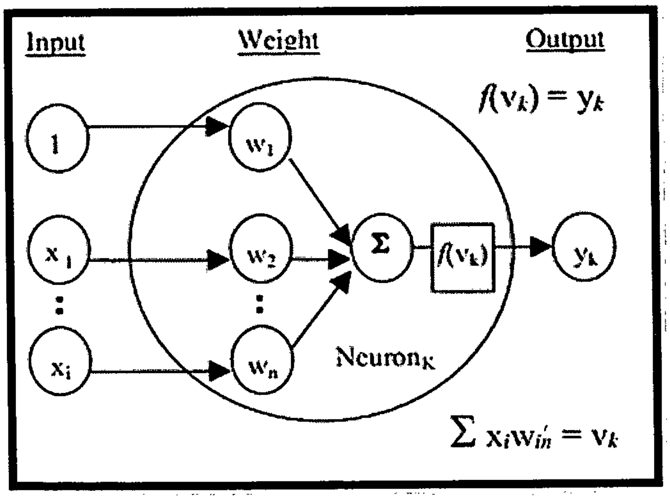

Figure 3.3 an Artificial Neural Network (Sulaiman M et al. 2011)

There are many models of ANN and the models are defined based on the network structure of the neurons, connection strengths and the processing performed at computing strengths and processing performed at computing elements or neurons. The most common of ANN Model is Multi-Layer Perceptron Neural Network.

3.3.1 Multi-Layer Perceptron Neural Network (MLP-NN)

MLP is known as a supervised feed forward back propagation learning ANN model. It comprises of an input layer, one or more hidden layers, and an output layer. The term regulated implies that the model requires output data to learn the pattern of data input. There are no cycles or loops in the network. MLP is the common type of ANN used in forecasting study.

The reason for the popularity of MLP is because it is simple, easy implementation and has demonstrated success in forecasting study. Figure 3.4 shows an example of the MLP Architecture. There are three layers in MLP, an input layer, hidden layers and an output layer. There can only be one input layer and output layer; however they can be more than one hidden layer. Each layer contains neurons. The total number of neurons is being determined by the user. The number of neurons in the input layer is equivalent to the number of data inputs selected in the study. The number of output neurons in the output layer is generally one, resulting in one forecasting output. The number of neurons in the hidden layer is subjected to user selection. Often, the number of input neurons and hidden neurons are determined based on trial and error. An additional dummy neuron known as a bias neuron that holds a value of 1 is added to the input layer and each hidden layer. The bias neuron acts as a threshold value so that the value of the forecasted output falls between 0 and 1. The neurons are interconnected between layers as shown in the figure. The direction of the link is from the input layer to the output layer, and these links represent the computational process within the ANN architecture.

The actual computational process occurs in the neuron, in which the activation transfer function assigned to the neuron is used to compute the incoming value and produce an output. In the input neuron, the activation transfer functions relay the incoming data input as output. In the hidden and output neurons, the activation transfer function computes the incoming value and produces an output. In the input neurons, the activation transfer function employed is a linear transfer function. In hidden and output neuron, the activation transfer function that is commonly used is a sigmoid function.

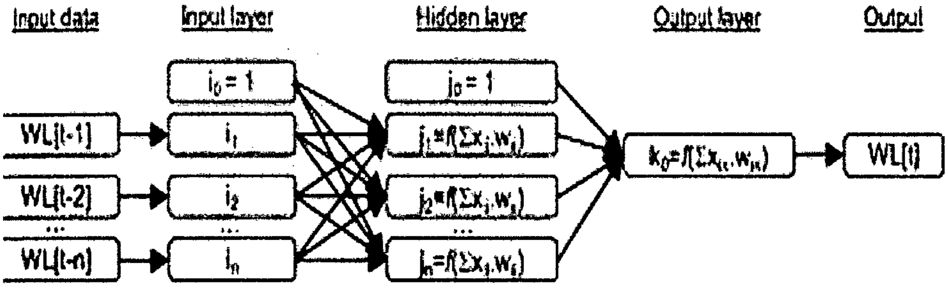

Figure 3.3.1 MLP-NN used in the study (Sulaiman et al. 2011)

Finally, the weight c values can be determined by data training. Data training can be accomplished when the number of neurons in each layer is defined, an activation transfer is assigned to the neurons in each layer, and data training are available. The data training process involves two parts of computation, feed-forward and back propagation. In the feed-forward computation, computing starts from the input layer and proceeds to the output layer based on the links and ATF in the ANN architecture. The output of the feed-forward computations is the forecasted value. In the data training process, performance id evaluated based on the model output value with respect to the observed value. If the measured performance does not achieve the target performance, a back-propagation takes place. The back-propagation computation is a process of adjusting the weights in the architecture based on the gradient descent method. The process of the weight adjustments starts from the output layer and proceeds backward toward the input layer. The process of feed-forward and back-propagation continues until the performance target is achieved.

Once the data training process is completed, the weights of the c values are determined. Then, the forecasting process proceeds based on the single feed-forward computation utilizing the input data as shown on the model architecture. The results are then evaluated by applying performance measures to the observed data versus the model output values.

3.3.2 Sigmoid Function

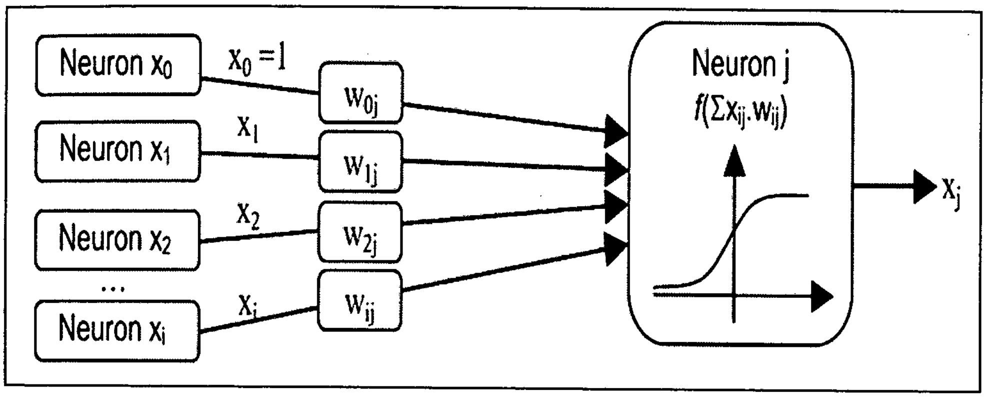

Figure 3.3.2 An active neuron with Sigmoid Function (Sulaiman et al. 2011)

Figure 3.5 shows a neuron with sigmoid function that computes incoming data and produces an output. The data received by the neuron is the summation of the output of the previous neuron based on its weight.

The activation transfer function (ATF) forces incoming values to range between 0 to 1 and 1 to -1 depending on the type of function used. The most commonly used ATF in the MLP is the sigmoid function. The sigmoid function is a differentiable function in which the gradient method can be applied to the ANN to adjust the ANN weights so that the model output and observed data reach a target performance value during the data training. The sigmoid function is defined as follows:

$$y = \frac{1}{1 + e^{-x}}$$