FLOOD FORECASTING FOR MELAKA USING ARIMA AND NAR MODELLING METHODS

WONG WEI MING

MASTER OF SCIENCE IN ELECTRONIC ENGINEERING

2022

Faculty of Electronics and Computer Engineering

FLOOD FORECASTING FOR MELAKA USING ARIMA AND NAR MODELLING METHODS

Wong Wei Ming

Master of Science in Electronic Engineering

2022

FLOOD FORECASTING FOR MELAKA USING ARIMA AND NAR MODELLING METHODS

WONG WEI MING

A thesis submitted
in fulfillment of the requirements for the degree of
Master of Science in Electronic Engineering

Faculty of Electronics and Computer Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2022

DECLARATION

I declare that this thesis entitled "Flood Forecasting And Modelling Using ARIMA And NAR" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature	:	
Name	:	Wong Wei Ming
Date	:	15 July 2022

APPROVAL

I hereby declare that I have read this thesis and in my opinion, this thesis is sufficient in terms of scope and quality for the award of the degree of Master of Science in Electronic Engineering.

Signature : ....

Supervisor Name : Dr. Siva Kumar Subramaniam

Date : 15 July 2022

DEDICATION

I dedicate this thesis to my parents, Wong Yeut Oon and Ong Kim Eng

My wife Rose Lew Ai Fen and my son Leon Wong Ze Yun

They gave me inspiration and support

Dr Siva Kumar and Dr Farah Shahnaz that given me the drive

And discipline to tackle my research with enthusiast and determination

Without their love and support, this study would not have been made possible

ABSTRACT

Flooding is an annual occurring incident in Malaysia. Several states in Malaysia are strongly affected by the flooding including Melaka where the flash floods are a common occurence. A flash flood is challenging to forecast and requires a sophisticated algorithm and system compared to the seasonal flood. It is difficult to forecast the flash flood compared to the seasonal flood. In Melaka, flash flood occurs regularly and it can happen to rise and fall in pace. This is the reason that flash floods can cause more damage than the seasonal floods. This study aims to develop a flood monitoring system to provide real-time data for the flood forecast. The objective is to develop a flood forecast model by analysing the flood parameters on a specific geographical layout which is Pengkalan Rama Jetty, Melaka. Following this, the efficiency of the flood forecast model is evaluated to forecast the water level where two flood forecast models were studied in this research which are the Autoregressive Integrated Moving Average (ARIMA) and Nonlinear Autoregressive Neural Network (NAR). The water level data considered for both methods were taken from 1st July 2020 at 12:00 am until 30th July 2020 at 7.00pm . There was a total of 2782 data in this time-series. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were used to find the best ARIMA model. The second method using NAR as a flood forecast model. This research used the time series data in NAR training, validation and testing to forecast the flash flood. In this research, the model was set to forecast the water level in several hourly time period of 1, 3, 5, and 7 hours. The forecast accuracy were measured using the Pearson R and R-squared to find the most accurate model for this multiple time-step ahead. The model's accuracy was determined by comparing the original and forecasted time series using Pearson R, R-Squared, Root Mean Squared Error (RMSE), Mean Squared Error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The result of the flood forecast system were compared with 7 hours forecast ahead and it was found that the ARIMA (2, 1, 3) was the best model for the Pengkalan Rama, Jetty, with an AIC of 5653.7004 and a BIC of 5695.209. The model also produced a lead forecast of up to 7 hours for the time series. Meanwhile, the result showed that the NAR model outperforms ARIMA with the lowest value in terms of RMSE, MSE, MAE and MAPE which are 1.915715, 3.669963, 1.576785 and 1.785951 respectively. In terms of Pearson R and R-Squared, the NAR model achieved Pearson R value of 0.931505 and R-Squared was 86.77024% compared to ARIMA which achieved R's value of -0.73993 and R-Squared of 54.74961%. It can be concluded that the flood forecast model for 7 hours ahead of using NAR outperformed ARIMA and is suitable for use in the flood forecast system at Pengkalan Rama Jetty.

RAMALAN BANJIR DI NEGERI MELAKA MENGGUNAKAN KAEDAH PEMODELAN ARIMA DAN NAR

ABSTRAK

Banjir ialah peristiwa yang berlaku secara tahunan di Malaysia. Beberapa negeri di Malaysia yang sentiasa menghadapi banjir termasuk Melaka. Banjir kilat telahpun menjadi kejadian biasa di Melaka. Banjir kilat sukar untuk diramalkan dan memerlukan algoritma dan sistem yang canggih berbanding dengan banjir monsun. Ramalan banjir untuk banjir kilat adalah lebih sukar berbanding ramalan untuk banjir monsun. Di Melaka, banjir kilat ini sentiasa berlaku, dan ia boleh berlaku dengan cepat dan surut dengan laju. Inilah diantara sebab-sebab yang menyebabkan banjir kilat menyebabkan kerosakan yang tinggi berbanding banjir monsun.Kajian ini bertujuan untuk membina model ramalan banjir dengan menggunakan data masa nyata dengan menganalisis parameter banjir di tempat tertentu, iaitu Jeti Pengkalan Rama, Melaka. Dua model ramalan banjir yang dikaji dalam penyelidikan ini iaitu Autoregressive Integrated Moving Average (ARIMA) dan Nonlinear Autoregressive Neural Network (NAR). Data paras air yang digunakan ialah daripada 1 Julai 2020 pada 12:00 pagi sehingga 30 Julai 2020 pada 7.00 pagi. Jumlah data yang diperolehi ialah 2782 data. Akaike Information Information Criterion (AIC) dan Bayesian Information Criterion (BIC) digunakan untuk mencari model ARIMA yang terbaik. Kaedah kedua menggunakan NAR sebagai model ramalan banjir. Penyelidikan ini menggunakan data dalam latihan, pengesahan, dan pengujian untuk NAR dalam meramalkan banjir kilat. Kajian ini bertujuan untuk mencari model ramalan yang paling tepat untuk meramalkan paras air dalam beberapa jam kehadapan, iaitu 1, 3, 5, dan 7 jam. Ketepatan ramalan diukur dengan menggunakan Pearson R dan R-squared untuk mencari model yang paling tepat untuk ramalan beberapa jam kehadapan. Ketepatan model diuji dengan membandingkan data asal dengan data ramalan dengan menggunakan Pearson R, R-Squared, Root Mean Squared Error (RMSE), Mean Squared Error (MSE), mean absolute error (MAE) dan mean absolute percentage error (MAPE). Keputusan sistem ramalan banjir dibandingkan dengan ramalan 7 jam kehadapan dan mendapati bahawa ARIMA (2, 1, 3) adalah model terbaik untuk Jeti Pengkalan Rama, dengan AIC 5653.7004 dan BIC 5695.209. Model tersebut menghasilkan ramalan sehingga 7 jam kehadapan bagi siri masa. Sementara itu, keputusan menunjukkan bahawa model NAR mengatasi ARIMA dengan nilai terendah dari segi RMSE, MSE, MAE dan MAPE, iaitu dengan nilai masing-masing 1.915715,3.669963,1.576785 dan 1.785951. Daripada segi Pearson R dan R-Squared, model NAR mencapai nilai Pearson R ialah 0.931505 dan R-Squared ialah 86.77024% berbanding ARIMA, yang hanya dapat mencapai nilai Pearson R ialah 0.73993 dan R-Squared ialah 54.74961%. Ia dapat disimpulkan bahawa model ramalan banjir bagi 7 jam kehadapan menggunakan NAR lebih baik daripada ARIMA dan sesuai digunakan dalam sistem ramalan banjir di Jeti Pengkalan Rama.

ACKNOWLEDGEMENTS

I dedicate this research to all the flood victims in Malaysia, hoping that lives can be saved with a better future Flood Warning Systems implementation. I was only able to finish my master research with the aid of a large number of people. I owe a debt of gratitude to every one of you. I want to extend my appreciation to the Ministry of Higher Education (MOHE) gave me a scholarship and opportunity to future my study through Hadiah Latihan Persekutuan (HLP) and Universiti Teknikal Malaysia Melaka (UTeM) for providing the research platform.

First and most important, I would like to express my sincere appreciation to my main supervisor, Dr. Siva Kumar Subramaniam and my second supervisor, Dr. Farah Sahnaz Feroz. I thank him for his constructive comments and his never-ending support throughout my research. I genuinely thank him for his continuous contribution and valuable assistance in teaching me how to prepare and write a high-quality master thesis and allowing me to express my interest through this thesis. Second, I would like to thank all my university lecturers including as well as supported staffs in Centre for Graduate Studies and Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer (FKEKK) Universiti Teknikal Malaysia Melaka (UTeM). They taught me various subjects that may have contributed directly and indirectly to writing this thesis, especially the various courses and training conducted on writing thesis or papers and preparing and motivating oneself before and after embarking on this master journey. Finally, my special thanks go to Mr. Mazran Esro for all the help and support I received from him.

Thank you to all my friends, especially Yusry and Amierul, who had provided me with the assistance, support and inspiration to embark on my study. Last and most importantly, I want to thank my family members, family-in-laws, and my wife, Rose Lew Ai Fen. She has contributed in many significant ways in helping me complete this master thesis in terms of encouragement and support (financial and moral), motivation and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

IFOS	-	Internet of Things Flood Observation System
NAR	-	Nonlinear Autoregressive Neural Network
ARIMA	-	Autoregressive Integrated Moving Average
IoT	-	Internet of Things
GSM	-	Global System for Mobile Communications
NDMRC	-	Disaster Management and Relief Committee
cm	-	Centimeter
UFF	-	Urban Flash Flooding
SARIMA	-	Seasonal Autoregressive Integrated Moving Average
WSN	-	Wireless Sensor Network
AR	-	Autoregressive
MA	-	Moving Average
ADF	-	Augmented Dickey-Fuller
ARMA	-	Autoregressive Moving Average
AIC	-	Akaike Information Criterion
BIC	-	Bayesian Information Criterion
MAPE	-	Mean Absolute Percentage Error
ANN	-	Artificial Neural Network
RNN	-	Recurrent Neural Network
NARX	-	Nonlinear Autoregressive With Exogenous Input
MLP	-	Multilayer Perceptron
NARNET	-	Non-Linear Autoregressive Neural Network Networking
RMSE	-	Root Mean Square Error
MSE	-	Mean Square Error
MAE	-	Mean Absolute Error
MAPE	-	Mean Absolute Percentage Error
GESD	-	Generalized Extreme Studentized Deviate
GS	-	Gaussian Smoothing
MBMB	-	Majlis Bandaraya Melaka Bersejarah

PPSPM	-	Perbadanan Pembangunan Sungai dan Pantai Melaka
EEG	-	Electroencephalogram
ACF	-	Autocorrelation Function
PACF	-	Partial Autocorrelation Function
LM	-	Levenberg-Marquardt
Q-Q	-	Quartile to Quartile
RBFNN	-	Radial Basis Function Neural Network
ENN	-	Elman Neural Network
WiFi	-	Wireless Network Technology
LoRa	-	Long Range Low-power Wide Area Network

LIST OF APPENDICES

LIST OF PUBLICATIONS

Indexed Journal

Wong, W. M., Subramaniam, S. K., Feroz, F. S. and Subramaniam, I. D., 2020. Flood Prediction using ARIMA Model in Sungai Melaka. *International Journal of Advanced Trends in Computer Science and Engineering*, *9(4)*, pp. 5287–5295. (Scopus Indexed, Q4)

Non-Indexed Journal

Wong, W. M., Subramaniam, S. K., Feroz, F. S. and Rose, L. A. F., 2020. Flood Prediction Model In Malaysia: A Review Paper. *Ibu International Journal of Technical and Natural Sciences*, *1(2)*, pp. 22-29.

Wong, W. M., Lee. M. Y., Azman. A. S., and Rose. L. A. F., 2021. Development of Short-term Flood Forecast Using ARIMA. *International Journal Of Mathematical Models And Methods In Applied Sciences*, *15*, pp. 68-75.

Conference Proceedings

N/A

CHAPTER 1

INTRODUCTION

1.1 Background

Flood is a common incident anywhere around the world especially for countries located over tropical regions. For Malaysia, flood is becoming a major issue of late to the extent that a heavy downpour can easily cause flash floods . Floods can no longer be seen or viewed as an isolated event as they are closely linked to disease outbreaks, food insecurity and climate deterioration. However, floods still happens in Malaysia even though the Malaysian government had executed multiple plans with a large budget to tackle this issue such as the enlargement of drainage (Ibrahim, 2019).

In the past decades, hundreds of lives have been lost due to floods either directly or indirectly. Besides, flood presents the most widespread natural threat to life today compared to other natural hazards. Based on the 2014 & 2015 data from the Water Resources Management and Hydrology Division, 381 flood cases occurred in Malaysia (Ranhill, 2011). Generally, government buildings or schools were converted as an evacuation centre since these locations can act as a shelter for many flood victims. Importantly, these places also must have essential basics such as drinking water supply.

It is vital to forecast floods because continuous heavy rainfall that can cause floods is very challenging to predict and avoid especially in countries with high rainfall rates annually such as in Malaysia. Flood forecasting also became a discussion among world researchers to get the best prediction method for flood occurrence (Gude et al., 2020; Ruslan et al., 2018; Subramaniam et al., 2009).

Flood forecasting is among the most complex models but this research is required for risk reduction due to floods. Flood not only damage properties but can also cause serious injuries or lost of lives thus flood forecasting has become increasingly important. In general, the prediction model involves high accuracy and statistical modelling such as trend analysis. These models are essential to help the local authority to manage and identify the flood trends and patterns.

On 1st December 2019 in Melaka, Malaysia, continuous 2 hours of heavy rainfall caused a flash flood where 20 vehicles were sunk in the water and another 30 cars were stuck in the car park buildings and unable to move out while waiting for the water to recede (Zakinan, 2019). Figure 1.1 shows that flood victims were brought to an evacuation centre following a flood in their homes.

Figure 1.1: Victims at evacuation centres.

Meanwhile on the 7th of July 2019, in Melaka, Malaysia, a flash flood in Alor Gajah and Melaka Tengha resulted in the displacement of 1096 individuals thatwere caused by rainwater runoff from storms in Negeri Sembilan that overflowed into the state of Melaka.

According to Effendy Ali, head of the State Disaster Management Committee's secretariat, as of 5:00 p.m., there were 1021 victims from 229 households in Alor Gajah and another 75 from 18 families in Melaka Tengah (Hamid, 2019). On another instance, heavy rain in Melaka on the 17th of May 2021 caused flash floods in several areas which forced evacuations (Murali, 2021).

There are two types of floods in Malaysia which are the flash floods and seasonal floods that often occur. Flash floods typically happen in high population areas such as Kuala Lumpur or Melaka while seasonal flood occurs particularly due to the monsoon season and for a specified period of time or month. Seasonal floods mainly occur in the southern part of Peninsular Malaysia from November until February due to Northeast Monsoon (Suhaila et al., 2010).

The occurrence of flash floods has a strong relationship with the amount of rainfall and geography. Rainfall with high intensity in a short period and complex topography tend to have a high frequency of flash floods (Destro et al., 2018). Flash floods also occurred due to unplanned building development leading to a poor drainage system in the city. Whenever there is a heavy rain, rainwater could not adequately drain out of the town which then causes a flash flood. Besides, drastic land use without proper planning and development at river basins could increase the impermeable land surrounding the river basin thus increasing the volume and peak discharge of hydrograph generated by the river basin (Alaghmand et al., 2010).

1.2 Research Problem

There are various types of flood forecast models from researchers like fuzzy inference systems (Lohani et al., 2014), Multiple-Input Single-Output (MISO) Autoregressive with Exogenous Input (ARX), MISO Autoregressive Moving Average With

Exogenous Input (ARMAX) structure (Ruslan et al., 2017) and stochastic model or ARIMA (Hamidi Machekposhti et al., 2017). Each type of flood forecast model has different forecast accuracy and this certainly affects the forecast duration of the incoming flood. In the following chapters, opinions and suggestions to choose the best model will then be given to suit and serve the primary purpose of forecast flood by discussing and distinguishing different kinds of flood forecast models. Research Problem

Lately, the world has been facing various natural disasters and flooding is one of theunexpected, sudden disasters that could take place. Houses and buildings in low-laying areas are more susceptible to floods. Owing to environmental variability and geographical locations, it remains a question for the community in the low-lying area to monitor and control geo-hydrological events that may occur anytime. Flooding causes the population to immigrate from their usual residence to safe places due to damage in infrastructures, security and health risks, interrupted education and and more critically, high cost involved for cleaning and repair purposes. There are several water level monitoring nodes in Melaka. However, it is just for the water monitoring which were shown on the website of Department of Irrigation and Drainage (DID, 2021; Zainol, 2018). Hence, a flood forecasting system is important to help minimize the loss. There are various flood forecast models that have been developed by the researchers such as Backpropagation Neural Network (BPNN), Elman Neural Network (ENN), Autoregressive Integration Moving Average (ARIMA), Artifial Neural Network (ANN) and many more (Wong et al., 2020). Each flood forecast model has different accuracy and data to be forecast. There are flood forecast models that can forecast seasonal floods where some models have high accuracy levels. Moreover, different locations will need different flood forecast models thus each with a differing accuracy. .

The river basin land-use development condition leads to increased impermeability and increased volume and peak discharge of the hydrograph generated by the river basin (Alaghmand et al., 2010)There are a few ways to obtain the water level data where the data can be obtained from the InfoBanjir website (DID, 2021) or the Malaysian Meteorological Department. Unfortunately, the water level data obtained from these sources are for the upstream water level data. An analysis of the flood state and downstream water level data are important since the downstream areas are usually flood-prone. In order to obtain the downstream water level data, water monitoring system is necessary to be deployed in these areas. Where these data are important for the flood forecast model in order to analyse and forecast the flood.

Hence, developing a water monitoring system that provides accurate and reliable water level data are important. More importantly, the data from the water monitoring system must be user-friendly both in terms of monitoring and analysis . The advancement in the technology has created a network of connected devices via the Internet which is called the Internet of Things (IoT). This allow us to monitor various data IoT but technological limitations limit the usage of IoT especially the coverage (Jumhana et al., 2020).Therefore, it is extremely important to determine the optimal deployment location as this is one of the key parameters in the overall setup . Moreover, a site survey must be conducted to study the network coverage thus enabling us to decide the communication hardware to be used such as the Long Range Low-power Wide Area Network (LoRa) (Jumhana et al., 2020), Global System for Mobile communication (GSM) (Shah et al., 2018), Wireless Networking Technology (WiFi) (Faro et al., 2020) and etc. Besides, various servers can be used for the IoT, such as Thingboard, Blynk , and Thingspeak (Amir Alavi et al., 2018; Hasbullah et al., 2020) .Meanwhile, a crucial element that must be considered is the , choice of sensors which are the most crucial part of the system that allow researchers to get accurate and reliable