

NON-LINEAR WATER LEVEL FORECASTING OF DUNGUN RIVER USING
HYBRIDIZATION OF BACKPROPAGATION NEURAL NETWORK AND GENETIC
ALGORITHM

SITI HAJAR BINTI ARBAIN

A thesis submitted in fulfillment of the requirement for the award of the degree of Master of
Science (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

SEPTEMBER 2014

NON-LINEAR WATER LEVEL FORECASTING OF DUNGUN RIVER USING
HYBRIDIZATION OF BACKPROPAGATION NEURAL NETWORK AND
GENETIC ALGORITHM

SITI HAJAR BINTI ARBAIN

A thesis submitted in fulfillment of the requirement for the award of the degree of
Master of Science (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

SEPTEMBER 2014

Thanks to God, Allah *'azza wa jalla*

Then to my beloved family and friends

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, and lecturers. It is my greatest pleasure to take this opportunity to express my gratitude and thank you to all the people involve whether direct or indirect in my effort to successfully finish this project.

First of all, I would like to express my heartfelt thanks and deep sense of gratitude to my supervisor, Dr Antoni Wibowo for his guidance, patient, and encouragement as well as suggestions throughout the research. This project could never be successfully completed without his assistance and support. I am also very thankful to my co-supervisor, Prof. Madya Dr. Mohd Salihin Bin Ngadiman for his guidance and supportive in this research analysis. Without their continued support, this thesis would not be successfully finished.

I would like to extent my sincere appreciations to my lab research mates and friends for their help and support. The great comments and suggestions from all these individuals were invaluable to me. At it has been, and always will be, my greatest gratitude is to my family, who shared this burden fully and magnificently. Without their support, this project would have been difficult at best. Thank you

ABSTRACT

The Department of Irrigation and Drainage (DID) and Meteorological Malaysia Department (MMD) have identified that water level is one of the important indicators for flooding control. The aim of this study is to find the best regression model and to identify the dominant variables of water level in Dungun River. Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), Backpropagation Neural Network (BPNN) and Nonlinear Autoregressive Exogenous Model (NARX) are popular methods in time series forecasting. However, ARIMA and SARIMA produce linear models where the approximations of linear models for the complex real-world problems are not always satisfactory. Thus, Backpropagation Neural Network (BPNN) and Nonlinear Autoregressive Exogenous Model (NARX) can be implemented in the time series forecasting due to its nonlinear modelling capability. These four methods, however, cannot be used directly for water level prediction since the original data from DID and MMD contain missing data. In this thesis, two methods are employed to treat missing data which are pre-processing using Mean and preprocessing using Ordinary Linear Regression (OLR) substitutions. In addition, BPNN and NARX may be difficult to determine the optimal network architecture and weights design since the optimal weight are different in each learning process. Thus, it is difficult to get best model in prediction. Based on the limitation of BPNN and NARX, the hybridization of Single BPPN and Genetic Algorithms (S-BPNN-GA) and Multi BPNN and Genetic Algorithms (M-BPNN-GA) have been proposed in this study. Experiments indicate hybridization of M-BPNN-GA 5-6-1 using five predictor variables including monthly, rainfall, temperature, evaporation and humidity and give better results compared to the other methods.

ABSTRAK

Jabatan Pengairan dan Saliran (DID) dan Jabatan Meteorologi Malaysia (MMD) telah mengenal pasti bahawa paras air adalah salah satu petunjuk yang penting untuk pengawalan banjir. Tujuan kajian ini adalah untuk mencari model regresi yang terbaik dan untuk mengenal pasti pembolehubah dominan bagi paras air di Sungai Dungun. Autoregresi Bersepadu Purata Bergerak (ARIMA), ARIMA bermusim (SARIMA), *Backpropagation Neural Network* (BPNN) dan *Nonlinear Autoregressive Exogenous Model* (NARX) adalah kaedah yang popular dalam ramalan siri masa. Walau bagaimanapun, ARIMA dan SARIMA menghasilkan model linear di mana anggaran model linear kepada masalah dunia sebenar yang kompleks tidak sentiasa memuaskan. Oleh itu, kaedah BPNN dan NARX diaplikasikan dalam ramalan siri masa kerana keupayaan model tersebut untuk mengendalikan masalah tidak linear. Tetapi keempat-empat kaedah in tidak boleh digunakan secara langsung untuk meramal paras air menggunakan data dari DID dan MMD kerana mengandungi data yang tidak lengkap. Dalam tesis ini, dua kaedah digunakan untuk mengendalikan data yang tidak lengkap iaitu pra-pemprosesan menggunakan penggantian purata dan penggantian *Ordinary Linear Regression* (OLR). Di samping itu, BPNN dan NARX mungkin sukar untuk menentukan rangkaian seni bina dan reka bentuk pemberat yang optimum memandangkan setiap proses pembelajaran akan menghasilkan nilai yang berbeza. Akibatnya sukar untuk mendapatkan model terbaik dalam ramalan. Oleh kerana terdapat kekangan pada model BPNN dan NARX, maka model penghibridan BPPN dan Algoritma Genetik serta penghibridan Multi BPNN dan Algoritma Genetik dicadangkan dalam penyelidikan ini. Eksperimen menunjukkan penghibridan M-BPNN-GA 5-6-1 menggunakan lima pembolehubah peramal yang meliputi pembolehubah bulanan, kadar hujan, suhu, penyejatan dan kelembapan memberikan keputusan yang lebih baik berbanding dengan model yang lain.