EMPIRICAL MODE DECOMPOSITION WITH LEAST SQUARE SUPPORT VECTOR MACHINE MODEL FOR RIVER FLOW FORECASTING

SHUHAIDA BTE ISMAIL

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

MARCH 2016

ACKNOWLEDGEMENT

I am extremely grateful to my supervisor Dr. Ani Shabri, for his invaluable professional advice, guidance and financial support throughout the whole process. I am also tremendously grateful to all of the lab mate of C22-451-01 for their continuous support in many aspects of this research.

ABSTRACT

Accurate information on future river flow is a fundamental key for water resources planning, and management. Traditionally, single models have been introduced to predict the future value of river flow. However, single models may not be suitable to capture the nonlinear and non-stationary nature of the data. In this study, a three-step-prediction method based on Empirical Mode Decomposition (EMD), Kernel Principal Component Analysis (KPCA) and Least Square Support Vector Machine (LSSVM) model, referred to as EMD-KPCA-LSSVM is introduced. EMD is used to decompose the river flow data into several Intrinsic Mode Functions (IMFs) and residue. Then, KPCA is used to reduce the dimensionality of the dataset, which are then input into LSSVM for forecasting purposes. This study also presents comparison between the proposed model of EMD-KPCA-LSSVM with EMD-PCA-LSSVM, EMD-LSSVM, Benchmark EMD-LSSVM model proposed by previous researchers and few other benchmark models such as Single LSSVM and Support Vector Machine (SVM) model, EMD-SVM, PCA-LSSVM, and PCA-SVM. These models are ranked based on five statistical measures namely Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Correlation Coefficient (r), Correlation of Efficiency (CE) and Mean Absolute Percentage Error (MAPE). Then, the best ranked model is measured using Mean of Forecasting Error (MFE) to determine its under and over-predicted forecast rate. The results show that EMD-KPCA-LSSVM ranked first based on five measures for Muda, Selangor and Tualang Rivers. This model also indicates a small percentage of under-predicted values compared to the observed river flow values of 1.36%, 0.66%, 4.8% and 2.32% for Muda, Bernam, Selangor and Tualang Rivers, respectively. The study concludes by recommending the application of an EMD-based combined model particularly with kernel-based dimension reduction approach for river flow forecasting due to better prediction results and stability than those achieved from single models.

ABSTRAK

Maklumat yang tepat mengenai masa hadapan aliran sungai adalah kunci asas kepada perancangan dan pengurusan sumber air. Secara tradisi, model tunggal telah diperkenalkan untuk meramalkan nilai masa depan bagi aliran sungai. Walau bagaimanapun, model tunggal mungkin tidak sesuai untuk mengenalpasti ketaklelurusan dan ketakpegunan yang wujud dalam data. Dalam kajian ini, kaedah tiga langkah-ramalan berdasarkan Mod Impirikal Penguraian (EMD), Kernel Utama Analisis Komponen (KPCA) dan model Kuasa Dua Terkecil Mesin Sokongan Vector (LSSVM), yang disebut sebagai EMD-KPCA-LSSVM diperkenalkan. EMD digunakan untuk menguraikan data aliran sungai kepada beberapa Fungsi Intrinsik Mod (IMFs) dan reja. Kemudian, KPCA digunakan untuk mengurangkan kedimensian set data yang kemudiannya dimasukkan ke dalam LSSVM untuk tujuan peramalan. Kajian ini juga membandingkan antara model cadangan EMD-KPCA-LSSVM dengan EMD-PCA-LSSVM, EMD-LSSVM, model Penanda Aras EMD-LSSVM yang dicadangkan oleh penyelidik sebelum ini dan beberapa model penanda aras lain seperti model tunggal LSSVM dan Mesin Sokongan Vector (SVM), EMD-SVM, PCA-LSSVM dan PCA-SVM. Model ini dinilai berdasarkan lima ukuran statistik iaitu Ralat Mutlak Min (MAE), Ralat Punca Min Kuasa Dua (RMSE), Pekali Kolerasi (r), Kecekapan Korelasi (CE) dan Peratus Ralat Mutlak Min (MAPE). Kemudian, model terbaik kedudukannya diukur menggunakan Min Ramalan Ralat (MFE) untuk menentukan kadar terkurang dan terlebih ramalan. Keputusan menunjukkan bahawa EMD-KPCA-LSSVM menduduki tempat pertama berdasarkan lima ukuran bagi Sungai Muda, Sungai Selangor dan Sungai Tualang. Model ini juga menunjukkan peratusan yang kecil bagi nilai terkurang ramal berbanding nilai aliran sungai yang direkodkan masing-masing sebanyak 1.36%, 0.66%, 4.8% dan 2.32% bagi Sungai Muda, Bernam, Selangor dan Sungai Tualang. Kajian ini membuat kesimpulan dengan mengesyorkan penggunaan model berasaskan EMD terutamanya dengan pendekatan pengurangan dimensi berasakan Teras untuk ramalan aliran sungai kerana hasil ramalan yang lebih baik dan kestabilan yang dicapai berbanding dengan keputusan yang diperolehi daripada model tunggal.

TABLE OF CONTENTS

CHAPTER	TITLE		PAGE
	DECLARATION		ii
	ACKNOWLEDGEMENT		iii
	ABSTRACT		iv
	ABSTRAK		v
	TABLE OF CONTENTS		vi
	LIST OF TABLES		xi
	LIST OF FIGURES		xiii
	LIST OF ABBREVIATIONS		xvi
	LIST OF APPENDICES		xix
1	**INTRODUCTION**		**1**
	1.1	Background Study	1
	1.2	Problem Statement	2
	1.3	Research Goal	6
	1.4	Research Objective	7
	1.5	Research Scope	7
	1.6	Research Justification	8
	1.7	Significance of the Study	9
	1.8	Thesis Outline	9
	1.9	Research Framework	11

2	**LITERATURE REVIEW**		**12**
	2.1	Introduction	12
	2.2	Time Series Forecasting	13
	2.3	River Flow Forecasting	14
	2.4	Previous Study in Hydrology	15
		2.4.1 Regression Analysis	15
		2.4.2 Autoregressive Integrated Moving Average	16
		2.4.3 Application Using ARIMA Models	24
		2.4.4 Artificial Neural Network	26
		2.4.5 Application Using ANN	30
	2.5	Current Study in Hydrology	32
		2.5.1 Support Vector Machine	32
		2.5.1.1 SVM for Regression	33
		2.5.1.2 Applications Using SVM	36
		2.5.1.3 SVM in Hydrology	37
		2.5.2 Least Square Support Vector Machine	40
		2.5.2.1 LSSVM in Hydrology	41
		2.5.3 Data Pre-processing	44
		2.5.3.1 Principal Component Analysis (PCA)	44
		2.5.3.2 Kernel Principal Component Analysis (KPCA)	46
		2.5.4 Data Decomposition	49
		2.5.4.1 Empirical Mode Decomposition (EMD)	49
		2.5.4.2 EMD in Hydrology	51
	2.6	Combination Models	52
	2.7	Summary of Previous Researches	47
	2.8	Summary	64
3	**RESEARCH METHODOLOGY**		**65**
	3.1	Introduction	65

3.2	Data Source		66
3.3	Stationary Test		68
	3.3.1	ADF Algorithm	69
	3.3.2	ADF Test Regression Forms	70
3.4	Linearity Test		71
	3.4.1	BDS Algorithm	72
3.5	Empirical Mode Decomposition Technique		74
	3.5.1	Intrinsic mode function (IMF)	75
	3.5.2	Sifting Process	76
	3.5.3	Instantaneous Frequency (IF)	80
	3.5.4	Hilbert Spectrum (HS)	81
3.6	Principal Component Analysis Technique		82
	3.6.1	Deviation of Principal Component	83
	3.6.2	Variance of Principal Component	85
3.7	Kernel Principal Component Analysis Technique		87
	3.7.1	Derivation of Kernel Principal Component Analysis Technique	89
3.8	Least Square Support Vector Machine Model		95
	3.8.1	Karush-Kuhn Tucker	95
	3.8.2	Validation Technique	96
		3.8.2.1 Cross Validation	96
	3.8.3	LSSVM Algorithm	97
	3.8.4	Kernel Functions	100
3.9	Benchmark EMD-LSSVM Model		102
3.10	EMD-SVM / EMD-LSSVM Models		103
3.11	Proposed Model (EMD-KPCA-LSSVM)		105
3.12	Experiment Setup		108
	3.12.1	Setting for SVM	108
	3.12.2	Setting for LSSVM	108
	3.12.3	Data Pre-processing	109
3.13	Statistical Performances		110
3.14	Summary		114

4	**DATA COLLLECTION AND ANALYSIS**	**115**
	4.1 Introduction	115
	4.2 Testing the Stationary	115
	4.3 Testing the Linearity	118
	4.4 Data Splitting	119
	4.5 Input determination	122
	4.6 Results	124
	4.7 Monthly River Flow Using Single Model	124
	4.7.1 SVM Results	124
	4.7.2 LSSVM Results	126
	4.8 Monthly River Flow Using PCA-Based Models	127
	4.8.1 Data Pre-Processing Technique Using PCA	128
	4.8.2 PCA-SVM Results	130
	4.8.3 PCA-LSSVM Results	131
	4.9 Monthly River Flow Using EMD-Based Model	133
	4.9.1 Data Decomposition Using EMD	133
	4.9.2 Benchmark EMD-LSSVM Results	138
	4.9.3 IMF Identification Using R^2	139
	4.9.4 EMD-SVM Results	142
	4.9.5 EMD-LSSVM Results	144
	4.10 Monthly River Flow Using EMD-Based Model Combined with Feature Extraction Techniques	146
	4.10.1 EMD-PCA-LSSVM Model	147
	4.10.1.1 Stage 1 (Decomposition Using EMD)	147
	4.10.1.2 Stage 2 (Component Identification/Feature Extraction Using PCA)	147
	4.10.1.3 Stage 3 (Forecasting Using LSSVM)	150

4.10.2 EMD-KPCA-LSSVM Model 153
4.10.1.1 Stage 1 (Decomposition Using EMD) 153
4.10.1.2 Stage 2 (Component Identification/Feature Extraction Using KPCA) 153
4.10.1.3 Stage 3 (Forecasting Using LSSVM) 154
4.11 Summary 156

5 COMPARISON AND DISCUSSION 158

5.1 Introduction 158
5.2 Comparison 159
5.3 Discussions 169
5.3.1 Discussion on Single Models 170
5.3.2 Discussion on PCA-Based Models 171
5.3.3 Discussion on EMD-Based Models 172
5.3.4 Discussion on EMD-Based Models Combine with Feature Extraction Technique 174
5.2 Summary 177

6 CONCLUSIONS AND RECOMMENDATIONS 178

6.1 Introduction 178
6.2 Conclusions 178
6.3 Contributions 182
6.4 Suggestions for Future Research 183

REFERENCES 185

APPENDICES (A-D) 200 – 244

LIST OF TABLES

TABLE NO.	TITLE	PAGE
2.1	Previous Studies on hybrid or combination models	59
3.1	Length of Data for Various Station	66
3.2	Descriptions of the Catchment Areas	67
4.1	ADF Test Results for Muda, Bernam, Selangor and Tualang Rivers	117
4.2	BDS Test Results for Muda, Bernam, Selangor and Tualang Rivers	119
4.3	The Input Structure for Muda, Bernam, Selangor and Tualang Rivers	123
4.4	Single SVM Result for Muda, Bernam, Selangor and Tualang Rivers	125
4.5	Single LSSVM Result for Muda, Bernam, Selangor and Tualang Rivers	127
4.6	Descriptive Statistics of PCs for Muda, Bernam, Selangor and Tualang Rivers	129
4.7	PCA-SVM Result for Muda, Bernam, Selangor and Tualang Rivers	130
4.8	PCA-LSSVM Result for Muda, Bernam, Selangor and Tualang Rivers	132
4.9	Benchmark EMD-LSSVM Results for Muda, Bernam, Selangor and Tualang Rivers	138

4.10(a)	The R^2 Results For EMD-SVM/EMD-LSSVM for Muda River (2 – 8 Input Variables)	140
4.10(b)	The R^2 Results For EMD-SVM/EMD-LSSVM for Muda River (10 – 12 Input Variables)	141
4.11	EMD-SVM Results for Muda, Bernam, Selangor and Tualang Rivers	143
4.12	EMD-LSSVM Results for Muda, Bernam, Selangor and Tualang Rivers	145
4.13	Descriptive Statistics of EMD-PCA for Muda River	148
4.14	Descriptive Statistics of EMD-PCA for Bernam River	148
4.15	Descriptive Statistics of EMD-PCA for Selangor River	149
4.16	Descriptive Statistics of EMD-PCA for Tualang River	149
4.17	EMD-PCA-LSSVM Results for Muda and Bernam Rivers	151
4.18	EMD-PCA-LSSVM Results for Selangor and Tualang Rivers	152
4.19	EMD-KPCA-LSSVM Results for Muda, Bernam, Selangor and Tualang Rivers	155
5.1	Comparative Performances of All Models for Muda, Bernam, Selangor and Tualang Rivers	161

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
1.1	Research Framework	11
2.1	The Feed forward Network	27
2.2	Multilayer Perceptron Architecture	29
2.3	Nonlinear SVR function with Vapnik's ε-insensitive loss function	36
2.4	Flow Chart of PCA	45
3.1	Locations of the Case Study Sites	68
3.2(a)	The blue curve is the input signal $x(t)$, red circles represent the local maxima, and the green squares are local minima (Baseem, 2011)	77
3.2(b)	Upper and lower black lines envelopes represented cubic spline interpolation, and the red line is the mean envelope $m_1(t)$ (Baseem, 2011)	78
3.2(c)	The blue curve represents the input signal minus the mean envelop $h_1(t) = x(t) - m_1(t)$, and the black line is the envelopes (Baseem, 2011)	78
3.2(d)	The blue signal is the first IMF $h_1(t)$ since it meets the IMF requirements (Baseem, 2011)	78
3.2(e)	Blue curve is the input signal minus the first IMF (residual $r(t) = x(t) - h_1(t)$, to be considered as new input signal (Baseem, 2011)	79

3.2(f)	Blue curve is the second IMF $h_2(t)$ together with its upper, lower and mean envelopes (Baseem, 2011)	79
3.3	Benchmark EMD-LSSVM by Ding *et al*. (2010) and Lin *et al*. (2012)	102
3.4	The EMD-SVM and EMD/LSSVM Models	104
3.5	The Proposed Three-Steps-Forecasting EMD-KPCA-LSSVM Model	107
4.1	The ACF Graphs for Muda, Bernam, Selangor and Tualang Rivers	116
4.2	Monthly River Flow for Muda River	120
4.3	Monthly River Flow for Bernam River	120
4.4	Monthly River Flow for Selangor River	121
4.5	Monthly River Flow for Tualang River	122
4.6	The IMFs and Residue Components for Muda River	134
4.7	The IMFs and Residue Components for Bernam River	135
4.8	The IMFs and Residue Components for Selangor River	136
4.9	The IMFs and Residue Components for Tualang River	137
5.1	The 1:1 plot for EMD-KPCA-LSSVM Model for Muda River	159
5.2	Hydrograph for EMD-KPCA-LSSVM for Muda River	160
5.3	The 1:1 plot for EMD-LSSVM Model for Bernam River	162
5.4	Hydrograph for EMD-LSSVM for Bernam River	163
5.5	The 1:1 plot for EMD-KPCA-LSSVM Model for Selangor River	164
5.6	The 1:1 plot for Benchmark EMD-LSSVM Model for Selangor River	165
5.7	The 1:1 plot for LSSVM Model for Selangor River	166
5.8	Hydrograph for EMD-KPCA-LSSVM Model for Selangor River	166
5.9	Hydrograph for Benchmark EMD-LSSVM Model for Selangor River	167
5.10	Hydrograph for LSSVM Model for Selangor River	167

5.11	The 1:1 plot for EMD-KPCA-LSSVM Model for Tualang River	168
5.12	Hydrograph for EMD-KPCA-LSSVM for Tualang River	169

LIST OF ABBREVATIONS

ACF	Autocorrelation Function
AI	Artificial Intelligent
ADF	Augmented Dickey-Fuller
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
BDS	Brock, Dechert and Scheinken
BP	Back Propagation
BPN	Back Propagation Network
CE	Correlation of Efficiency
CANN	Cluster-based ANN
DF	Dickey-Fuller
DWT	Discrete Wavelet Transform
E	Coefficient of Efficiency
EL	Evaporation Loss
EMD	Empirical Mode Decomposition
ERM	Empirical Risk Minimization
FFT	Fast Fourier Transform
FNN	Feed Forward Neural Network
FOV	Flashover Voltage
GA	Genetic Algorithm
HHT	Hilbert Huang Transform
HS	Hilbert Spectrum
IID	Independent and Linearly Distributed
IKPCA	Improvement Kernel Principal Component Analysis

IF	Instantaneous Frequency
IMF	Intrinsic Mode Function
LSSVM	Least Square Support Vector Machine
KKT	Karush Kuhn Tucker
KNN	K-Nearest-Neighbors
KPCA	Kernel Principal Component Analysis
LR	Linear Regression
LSSVR	Least Square Support Vector Regression
LSSVR	Least Square Support Vector Machine
MA	Moving Average
MAE	Mean Absolute Error
MANN	Modular Artificial Neural Network
MSE	Mean Square Error (MSE),
MARE	Mean Absolute Relative Error
MAPE	Mean Absolute Percentage Error
MFE	Mean of Forecast Error
MLR	Multiple Linear Regression
NN	Neural Network
NLP	Nonlinear Prediction
NSOR	Non-stationary Oscillation Re-sampling
PANN	Periodic-ANN
PCA	Principal Component Analysis
PM	Penman–Monteith
r	Correlation Coefficient
R^2	Coefficient of Determination
RBF	Radial Basis Function
RMSE	Root Mean Square Error
RSM	Response Surface Methodology
SAR	Seasonal Autoregressive
SD	Seasonal Decomposition
SRM	Structural Risk Minimization
SSA	Singular Spectrum Analysis
SVM	Support Vector Machine

SVR	Support Vector Regression
SWE	Snow Water Equivalent
TANN	Threshold-based ANN

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A1	The R^2 Results for EMD-SVM / EMD-LSSVM For Bernam River (2 – 8 Input)	201
A2	The R^2 Results for EMD-SVM / EMD-LSSVM For Bernam River (10 – 12 Input)	202
A3	The R^2 Results for EMD-SVM / EMD-LSSVM For Selangor River (2 – 8 Input)	203
A4	The R^2 Results for EMD-SVM / EMD-LSSVM For Selangor River (10 – 12 Input)	204
A5	The R^2 Results for EMD-SVM / EMD-LSSVM For Tualang River (2 – 8 Input)	205
A6	The R^2 Results for EMD-SVM / EMD-LSSVM For Tualang River (10 – 12 Input)	206
B1	Descriptive Statistics of KPCs for Muda River	208
B2	Descriptive Statistics of KPCs for Bernam River	209
B3	Descriptive Statistics of KPCs for Selangor River	210
B4	Descriptive Statistics of KPCs for Tualang River	211
C1	The 1:1 Plot and Hydrograph for Muda Rives	213
C2	The 1:1 Plot and Hydrograph for Bernam Rives	221
C1	The 1:1 Plot and Hydrograph for Selangor Rives	229
C1	The 1:1 Plot and Hydrograph for Tualang Rives	235
D1	List of Published Papers.	243

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

In general, hydrology is the scientific study of the characteristics of water, its distribution, its surface and its impact on the soil and the atmosphere. Hydrological data such as flows and rainfall are the basic sets of information used in designing water resources systems. The essential information about the characteristics and volume of the river flow are very important especially during monsoon season. Knowing and analysing the statistical properties of hydrological records and data such as rainfall or river flow, enables hydrologists to estimate future hydrological phenomena, especially in predicting the future river flow. The flow is critical in many activities such as designing flood protections for urban areas and agricultural land. The quantity of water can be measured from the river for water supply or irrigation. River flow also plays a significant role in establishing some of the critical interactions that occur between physical, ecological, social or economic processes.

Therefore, continuous hydrological data, such as the rainfall-runoff or river flow data are necessary. With the help of the data, the pattern of the flow or the trend can be determined, thus the design and planning can be done accordingly. For instance, heavy river flow may cause some damage to the environment such as flood. Flood, also referred to as ‘deluge’, is a natural disaster that could damage properties and infrastructure, harm animals, plants, and even human lives. Flood occurs when the volume of water exceeds the capacity of the catchment area. Meanwhile, low river flow may also cause some major problems for water supply such as domestic consumption, transportation, industrial, as well as impeding the function of hydroelectric power plants.

River flow is a fundamental component of a water resource system. A reliable prediction of the river flow is always important for a thorough planning and smooth operation of the water resource system. Because of this, the ability to forecast the future river flow will be beneficial to water management and help in flood control. Moreover, reliable river flow prediction can prevent natural disasters such as floods, and optimize the management of water resources. The extent of damage caused by flood undeniably highlights the importance of river flow forecasting (Knight and Shamseldin, 2006). However, in order to issue flood warning as well as to manage the water resources properly, there is a need to enhance the prediction of future river flow.

1.2 Problem Statement

There is a variety of statistical modelling approaches developed to capture the properties of hydrological time series forecasting for a reliable prediction of water flow: such as the physically based distribution model known as ‘knowledge-driven modelling’ and empirical models, known as ‘data-driven-based modelling’. Knowledge-driven modelling is also useful for predicting other catchment variables such size, shape, slope, and storage characteristics of the catchment, as well as geomorphologic characteristics like topography, land use patterns, vegetation, and soil types that affect the infiltration. It is assumed that forecasting could be improved if the catchment characteristic variables which affect the flow are included (Jain and Kumar, 2007; Dibike and Solomatine, 2001; Shabri and Suhartono, 2012).

Although combining other variables may improve the prediction accuracy, in practice, for developing countries such as Malaysia, the information is often either difficult to obtain or unavailable. Moreover, these variables and many of the combinations in generating river flow, make predication a complex process. This difficulty is exacerbated by the complex nature of the data's multiple inputs and parameters, which are varied in space and time and often not clearly understood (Zhang and Govindaraju, 2000; Jain and Kumar, 2007).

On the contrary, data-driven model mathematically identifies the connection between the inputs and output without considering the internal physical mechanism of the catchment areas. The data-driven model uses historical data that are based on extracting and reusing the information that are implicitly contained in the hydrological data without directly taking into account any physical load that underlies the rainfall-runoff process (Samsudin *et al.*, 2011). In river flow forecasting, the data-driven model, which uses previous river flow time series data, becomes increasingly popular (Kisi, 2008; 2009; Wang *et al.*, 2009). Many researchers only use the historical river flow data for forecasting future river flow as it offers fast computing time and requires minimum information (Adamowski and Sun, 2010; Kisi, 2004; 2008; Wang *et al.*, 2009; Samsudin *et al.*, 2011).

Improving forecasting accuracy is fundamental yet it is one of the more difficult tasks faced by decision-makers in many areas. Computer science and statistics have improved the data-driven modelling approaches in discovering the patterns in water resources time series data. Using hybrid models has become a common practice to improve the forecasting accuracy. Several studies have showed that hybrid models can be an effective way to improve predictions compared to the models that were used individually (Zhang, 2003; Jain and Kumar, 2007).

Recently, many researchers believe in the idea of 'divide-and-conquer' or the 'decompose-and-ensemble' principle in constructing the forecasting model (Lin *et al.*, 2012; Yu *et al.*, 2008). An Empirical Mode Decomposition (EMD) offers the solutions for nonlinearity and nonstationary issues by decomposing the nonlinearity and nonstationary behaviour of the time series into a series of valuable independent time resolutions (Tang *et al.*, 2012). Meanwhile, linear Principal Component Analysis (PCA) is widely used as a data pre-processing technique and commonly used for dimensionality reductions (Lee *et al.*, 2004). PCA is a statistical technique that can linearly transform a set of correlated variables into a smaller set of uncorrelated variables named Principal Components (PCs) where the first few PCs represent most of the information in the original data set.

However, some researchers argued that PCA was not the best technique in dimensionality reductions as PCA only identifies the linear structure in a data set. Hence, Scholkopf *et al.* (1998) introduced a Kernel Principal Component Analysis (KPCA) to extract the nonlinear principal component features from the data. KPCA has been successfully applied in recent years as a promising technique in various areas, such as de-noising images and dimensionality reductions (Lee *et al.*, 2004). Researchers also believe that the combinations of two or more models are able to increase the prediction accuracies by applying the combination model as an alternative way to resolve the problem in the forecasting area. There are many types of combination models, which are very helpful in forecasting area and the number of combination models are increasing every day. On top of that, there are varieties of useful combination models in time series forecasting that can be used to predict future river flow.

The aim of this study is to develop a new forecasting model, which is able to forecast the monthly river flow data, and at the same time overcome the weakness of the existing models such as Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Network (ANN) and many others. The first combination model used in this study is adapted from Chen and Wei (2012). The first combination model measures the correlation between each of the IMFs produced by using EMD, with the original time series data, using Pearson product moment correlation and Kendall rank correlation. The meaningful signals are aggregated together as the new input for BPN forecasting stage. Using the idea from Chen and Wei (2012), this study uses LSSVM instead of BPN as the forecasting tool.

The second combination model is adopted from Ding *et al.* (2010) and Lin *et al.* (2010). They have proposed EMD-LSSVM for precipitation and the foreign exchange rate forecasting is referred as the Benchmark EMD-LSSVM. In the second combination model, EMD is used to decompose the data into several IMFs and residue. The IMFs are forecasted using LSSVM individually. Finally, the forecasted values are reconstructed as the sum of all components. The third combination of models used in this study is by applying the idea from Zhou *et al.* (2013), which combined the EMD with feature extraction techniques with SVM model for signal recognition. Using the idea from Zhou *et al.* (2013), this study aims to explore the application of combined models which used EMD, KPCA, and LSSVM, also referred as EMD-KPCA-LSSVM model and to test the capability and effectiveness of the proposed model with other models.

This study attempts to adopt a three-steps-prediction based on EMD-KPCA-LSSVM to forecast the monthly river flow in Malaysia. Since previous researchers have employed EMD-SVM, EMD-LSSVM referred to as Benchmark EMD-LSSVM and KPCA-LSSVM in their research, it is expected that the three-steps-prediction using EMD-KPCA-LSSVM will be able to further enhance the forecasting accuracy of river flow. Therefore, the research questions are as stated below:

1. How to design a three-steps-prediction architecture model based on EMD as the decomposition method with KPCA technique for dimensional reduction or feature extraction, and combined with LSSVM?

2. Will the proposed EMD-KPCA-LSSVM improve the prediction accuracy and at the same time outperform other models?

Thus, the following issues are considered in order to solve these problems:

i. As PCA is usually used for dimensional reductions, will KPCA outperform the PCA technique?

ii. As the Benchmark EMD-LSSVM is employed in other forecasting areas, can Benchmark EMD-LSSVM be employed in the river flow forecasting?

1.3 Research Goal

The goal of this research is to develop and propose a three-steps-prediction model that combines EMD and KPCA with the LSSVM referred as EMD-KPCA-LSSVM for river flow forecasting. The results of the proposed model are compared with other models and are examined to determine whether the proposed model of EMD-KPCA-LSSVM significantly outperforms the others. The proposed three-steps-prediction of EMD-KPCA-LSSVM is expected to be useful for river flow forecasting.

1.4 Research Objectives

In view of the aforementioned problems, this study intends to propose the three-steps-prediction model to predict the monthly river flow in Malaysia. Some of the specific objectives of the study are:

1. To explore the capability of combining EMD with LSSVM model for river flow forecasting.

2. To design and develop a model based on EMD-KPCA-LSSVM, which combines decomposition, data pre-processing, and forecasting techniques for river flow forecasting.

3. To evaluate the performance of the proposed model and compare it with other models which are SVM, LSSVM, PCA-SVM, PCA-LSSVM, EMD-SVM, EMD-LSSVM, Benchmark EMD-LSSVM, and EMD-PCA-LSSVM.

1.5 Research Scope

The scope of this research includes:

i. The research focused on proposing a new method for time series forecasting of EMD-KPCA-LSSVM, which combines the decomposition technique with

KPCA as the data pre-processing technique and LSSVM as a forecasting tool.

ii. Real time series data of monthly river flows are taken from JPS, Malaysia from four different rivers that are selected as the case studies.

iii. Radial basis function is selected as the kernel function for both SVM and LSSVM models.

iv. The new obtained datasets from PCA and KPCA are set within two-cut-off values, which are from 70% to 90%.

v. Several evaluation measures are used to verify the best models, which are mean absolute error (MAE), root mean square error (RMSE), correlation coefficient (*r*), mean absolute percentage error (MAPE), and Nash–Sutcliffe coefficient efficiency (CE). The model with smallest MAE, RMSE, and MAPE, and the largest values of *r* and CE is considered as the best model.

1.6 Research Justification

This research is expected to contribute towards the fulfilment of the need to produce an optimal architecture of the model, which is more flexible than before, as well as to improve the model's prediction accuracy. The obtained results are expected to demonstrate higher accuracy and superb predictive capability in comparison to some previous models available in the literature.

1.7 Significance of the Study

Predicting the future river flow is very important where heavy river flow can cause problems such as flooding and erosion, while low river flow is likely to restrict the supply of water for domestic use, industrial and hydroelectric power generation. The study reviews the effectiveness of the proposed model as an alternative tool in forecasting. This research attempts to study the suitability of the data decomposition technique of EMD and the data pre-processing technique using the KPCA model where the original data are decomposed into several signals. KPCA is used for dimensionality reduction, and the newly obtained data are used to forecast the future value of the river flow. As this study will provide the information of the future river flow value based on past time series data, it is required for the proposed model of EMD-KPCA-LSSVM to forecast the monthly river flow in Malaysia in order to produce a better result. This will provide a better understanding of the trend of the river flow in Malaysia.

1.8 Thesis Outline

This thesis consists of six chapters and each of the chapters were discussed accordingly. The first chapter presents the introduction of this study. It describes the background of the study followed by the problem statements, the research goal, objective and the scope of study. Chapter 1 ends with research justification and elaboration on the significance of the study.

The second chapter provides an overview and the literature study of each of the models used in this research, as well as its latest application in hydrology. The purpose of the literature review provided in this study is to review previous researches, which are related to the current study. This chapter also review the advantages of using combination or hybrid or conjunction models. Chapter 2 finishes with a table, summarizing the previous researches in hybrid models.

Chapter 3 presents the research methodology which describes the characteristics of the catchment area and its locations. This chapter also describes the application of ADF and BDS tests used in this research. Furthermore, Chapter 3 also describes the approaches employed in the forecasting area, which are EMD, PCA, KPCA, SVM, and LSSVM models. Other than that, it also describes the proposed three-steps-prediction model based on the EMD-KPCA-LSSVM model. The statistical performances used in the study are also described in detail.

Chapter 4 discussed the experiments and applications of the selected models, used in this study. In this chapter, explanation and selection of PCs and KPCs, justification on IMFs and residue selections after decomposition are discusses in details. The results of the Single SVM, Single LSSVM, PCA-based models which are PCA-SVM and PCA-LSSVM, EMD-based models which are EMD-SVM, EMD-LSSVM, and the three-steps-prediction model which are EMD-PCA-LSSVM and EMD-KPCA-LSSVM are presented and consequently, the best results of each model are selected.

Chapter 5 carries the comparisons and discussions for models used in all the case studies. The obtained results are also compared to EMD-LSSVM, proposed by Ding *et al.* (2010) and Lin *et al.* (2012) and which are referred as the Benchmark EMD-LSSM model. The performance of each-based models which are PCA-based and EMD-based models are also discusses.

Chapter 6 concludes the thesis by providing details conclusions drawn from the study and highlighting its contributions. This chapter also suggests recommendation for future research, in order to enhance the applicability and capability of the proposed model in either hydrological or forecasting area.

1.9 Research Framework

To achieve the set out objectives, this study was conducted by following the presented workflow in Figure 1.1.

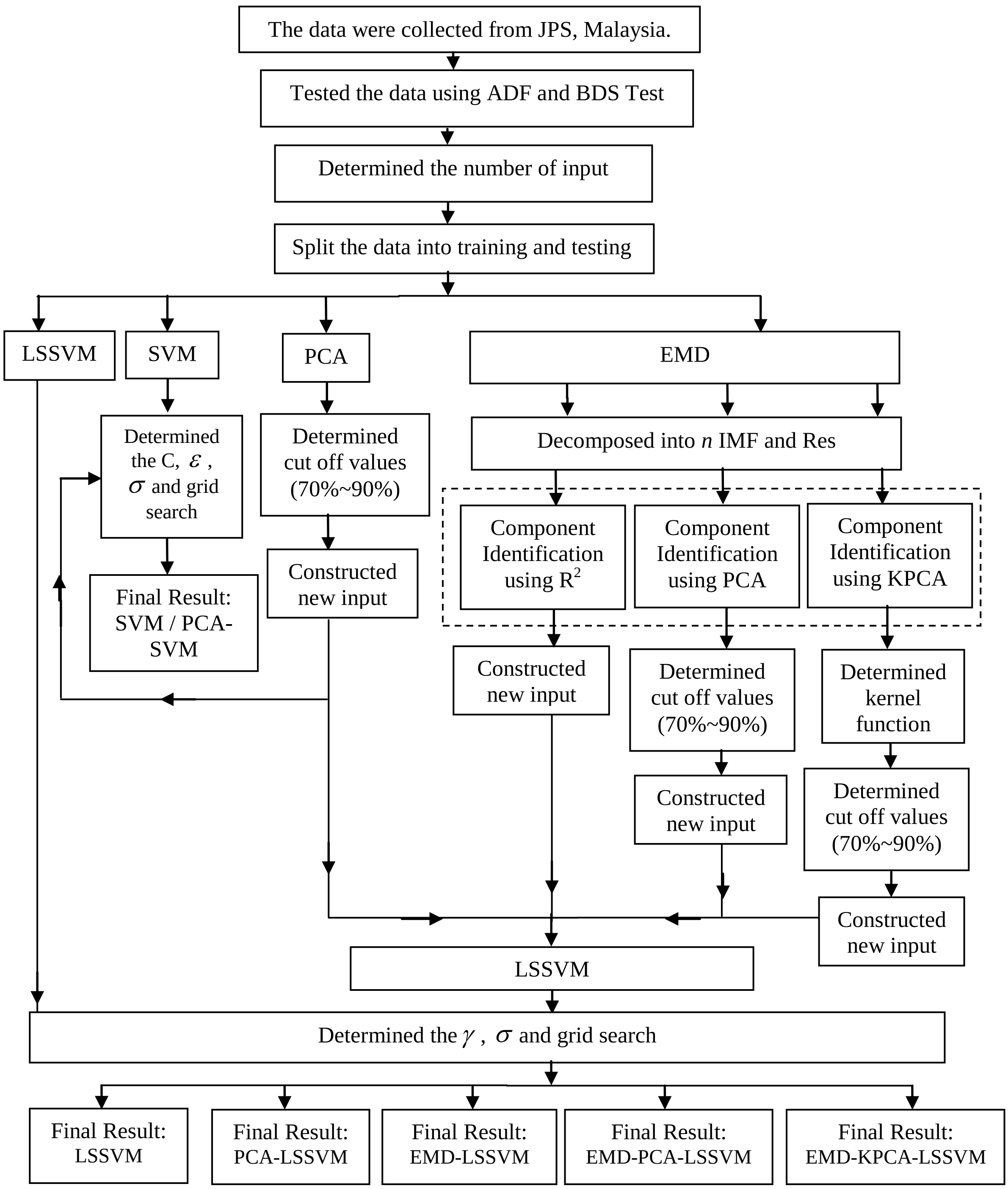

Figure 1.1 : Research Framework

REFERENCES

Adeloye, A. J., F. Lallemand, (2003). Regression models for within-year capacity adjustment in reservoir planning. *Hydrological Sciences Journal* 48(4): 539-552

Adeloye, A. J., and Munari, A. D., (2006). Artificial neural network based generalized storage–yield–reliability models using the Levenberg–Marquardt algorithm. *Journal of Hydrology*. 362: 215-230.

Ahmad, S., Khan, I. H., and Parida, B. P. (2001). Performance Of Stochastic Approaches For Forecasting River Water Quality. *Water Research* (35): 4261-4266.

An, X., Jiang, D., Zhao, M., Liu, C. (2012). Short-time prediction of wind power using EMD and chaotic theory. *Communications in Nonlinear Science and Numerical Simulation.* 17(2): 1036–1042.

Agarwal, A., Mishra, S. K., Ram, S. Singh, J. K. (2006). Simulation of runoff and sediment yield using artificial neural networks. *Biosystems Engineering*. 94(4): 597-613.

Afshin, M., Sadeghian, A. and Raahemifar, K. (2007). On efficient tuning of LS-SVM hyper-parameters in short-term load forecasting: A comparative study. *Proc. of the 2007 IEEE Power Engineering Society General Meeting (IEEE-PES).*

ASCE (2000). Artificial Neural Networks In Hydrology - II: Hydrologic Applications. *Journal of Hydrologic Engineering*. 5(2): 124-137

Asefa, T. and M. Kemblowski, (2002). Support vector machines approximation of flow and transport models in initial groundwater contamination network design. *Eos. Trans. AGU*. 83(47).

Asefa, T., Kemblowski, M., Mckee, M. and Khalil, A. (2006). Multi-time scale streamflow predictions: the support vector machines approach. *Journal of hydrology*. 318, 7-16.

Astel, A., Mazerski, J., Polkowska, Z., Namiesnik, J. (2004). Application of PCA and time series analysis in studies of precipitation in Tricity (Poland). *Advances in Environmental Research.* 8(3-4): 337-349.

Baseem, S. S. (2011). *De-noising seismic data by Empirical Mode Decomposition*. **Master Thesis**. Faculty of Mathematics and Natural Sciences. University of Oslo. Norway.

Batyrshin, I., Kacprzyk, J., Sheremetov, L., Zadeh, L. (2007). *Perception-based Data Mining and Decision Making in Economics and Finance*. Springer Berlin Heidelberg. 36: 119-134.

Birkinshaw S. J. , Parkin G., Rao, Z. (2008). A Hybrid Neural Networks And Numerical Models Approach For Predicting Groundwater Abstraction Impacts. *Journal of Hydroinformatics*. 127-137.

Bhagwat, P. P. and Maity, R. (2012). Multistep-Ahead River Flow Prediction Using LS-SVR at Daily Scale. *Journal of Water Resource and Protection.* 4: 528-539.

Bhagwat, P. P. and Maity, R. (2013). Hydroclimatic streamflow prediction using Least Square-Support Vector Regression. *Journal of Hydraulic Engineering.* 19(3): 320-328.

Bowerman, B. L. and R. T. O'Connell (1993). *Forecasting and time series : An Applied Approach*. California, Duxbury Press.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control*. New Jersey: John Wiley Publication.

Cao, L. J., Chua, K. S., Chong, W. K., Lee, H. P., Gu, Q. M. (2003). A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing*. 55(1-2): 321-336

Cimen, M. and Kisi, O. (2009). Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey. *Journal of Hydrology.* 378: 253-262.

Chang, J., Huang, S., Huang, Q., Chen, Y. (2014). Monthly streamflow prediction using modified EMD-based support vector machine. *Journal of Hydrology*. 511: 764-775.

Chau, K.-W., Wang, W.-C., Cheng, C.-T., Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology.* 374(3-4): 294-306.

Chen, Q., Xiaowei, C., Yun, W. (2008). The Combining Kernel Principal Component Analysis with Support Vector Machines for Time Series Prediction Model. *Second International Symposium on Intelligent Information Technology Application, 2008.* 2:90-94.

Chen, M.-C., and Wei, Y., (2012). Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C*. 21: 148-162

Chowdhury, A. R., Guhathakurta, K., Mukherjee, I. (2008). Empirical mode decomposition analysis of two different financial time series and their comparison. *Chaos, Solitons and Fractals*. 37: 1214–1227.

Choy, K. Y. and Chan, C. W., (2003). Modelling of river discharges and rainfall using radial basis function networks based on support vector regression. *International Journal of Systems Science*. 34(14-15): 763-773.

Chung, K. Y. C. (2010). *Facial Expression Recognition by Using Class Mean Gabor Responses with Kernel Principal Component Analysis*. **Master Thesis**. Russ Collage of Engineering and Technology, Ohio University.

Ding, Z., Zhang, J., Xie, G. (2010). LS-SVM Forecast Model of Precipitation and Runoff Based on EMD. *Sixth International Conference on Natural Computation (ICNC 2010).*

Deger, E. (2008). *Noise-Thresholding with Empirical Mode Decomposition for Low Distortion Speech Enhancement*. Master Thesis. Department of Information and Communication Engineering. University of Tokyo. Tokyo.

DeLurgio, S. A. (1998). *Forecasting principles and applications*. Boston: McGraw-Hill.

Deng, S. and Yeh, T.-H. (2010). Applying least squares support vector machines to the airframe wing-box structural design cost estimation. *Expert Systems with Applications.* 37(12): 8417-8423

Dibike, Y. B. and Solomatine, D. P. (2001). River Flow Forecasting Using Artificial Neural Networks. *Physics and Chemistry Of The Earth, Part B: Hydrology, Oceans And Atmosphere.*26: 1-7.

Dibike, Y. B., Slavco, V., Solomatine, D. P., Abbott, M. B. (2001). Model Induction with Support Vector Machines: Introduction and Applications. *Journal of Computing in Civil Engineering*. 15(3): 208-216.

Dolling, O. R. and Varas, E. A. (2002). Artificial neural networks for streamflow prediction. *Journal of Hydraulic Research.* 40(5): 547-554.

Elshorbagy, A., Corzo, G., Srinivasalu, S., Solomantine, D. P. (2010a). Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology. *Hydrology and Earth System Sciences*. 14(10): 1931-1941.

Elshorbagy, A., Corzo, G., Srinivasalu, S., Solomantine, D. P. (2010b). Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 2: Application. *Hydrology and Earth System Sciences.* 14(10): 1943-1961.

Fauset, L. (1994). *Fundamentals of Neural Networks Architectures, Algorithms And Applications*. New Jersey: Prentice Hall

Fauvel, M., Chanussot, J., Benediktsson, J. A. (2006). Kernel Principal Component Analysis for Feature Reduction in Hyperspectrale Images Analysis. *Proceedings of the 7th Nordic, Signal Processing Symposium, 2006. NORSIG 2006.* 238-241.

Flandrin, P. and P. Goncalves (2004). Empirical Mode Decompositions As Data-Driven Wavelet-Like Expansions. *International Journal of Wavelets, Multiresolution and Information Processing*. 2(4): 477-496.

Garen, D. C., (1992). Improved Techniques In Regression-Based Streamflow Volume Forecasting. *Journal Water Resources Planning Manage*. 118: 654-670.

Gestel, T. V., Suykens, J. A. K., (2001). Financial time series prediction using least squares support vector machines within the evidence framework. *Neural Networks, IEEE Transactions.* 12(4): 809-821.

Gencoglu, M. T. and Uyar, M. (2009). Prediction Of Flashover Voltage Of Insulators Using Least Squares Support Vector Machines. *Expert Systems with Applications.* 36(7): 10789-10798.

Goswami, M., K. M. O'Connor, et al. (2005). Assessing the performance of eight real-time updating models and procedures for the Brosna River. Hydrological Earth System Sciences. 9(4): 394–411.

Grayson, R. B., I. D. Moore, McMahon, T. A. (1992). Physically based hydrologic modelling: Is the concept realistic. *Water Resources Research*. 28: 2659-2666.

Gu, Y., Liu, Y., Zhang, Y. (2008). A Selective KPCA Algorithm Based on High-Order Statistics for Anomaly Detection in Hyperspectral Imagery. *IEEE Geoscience and Remote Sensing Letters.* 5(1): 43-47.

Guo, X., Sun, X., Ma, J. (2011). Prediction of daily crop reference evapotranspiration (ET_0) values through a least-squares support vector machine model. *Hydrology Research*. 42(4): 268-274.

Guhathakurta, P., Rajeevan, M., Thapliyan, V. (1999). Long Range Forecasting Indian Summer Monsoon Rainfall by a Hybrid Principal Component Neural Network Model. *Meteorology and ATM Ospheric Physics*. 71(3-4): 255-266.

Hanbay, D. (2009). An expert system based on least square support vector machines for diagnosis of valvular heart disease. *Expert Systems with Applications.* 36(4): 8368-8374.

Han, J. (2013). *Network Traffic Anomaly Detection Using EMD and Hilbert-Huang Transform*. Master Thesis. Department of Engineering and Technology. Western Carolina University.

Haykin, S. (1999). *Neural Networks - A Comprehensive Foundation.* 2nd. Prentice Hall.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J.M., Fernandez, L. (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research*. 34(3): 807–816.

He, K., Lai, K.K., Yen, J. (2010). A hybrid slantlet denoising least squares support vector regression model for exchange rate prediction. *Procedia Computer Science*. 1(1): 2397–2405.

Hipel, K.W., and McLeod, A.I. (1994). *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, Amsterdam.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*. (24): 417–441.

Hu, T. S., Lam, K.C., Ng, S.T. (2007). Rainfall-Runoff Modelling using Principal Component Analysis and Neural Network. *Nordic Hydrology*. 38(2): 235-248.

Hua, Z., Xu, X., Qi, Y. (2010). Forecasting demand of commodities after natural disasters. *Expert Systems with Applications*. 37: 4313-4317.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C., Liu, H. H. (1998). The Empirical Mode Decomposition and the Hilbert Spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A.* (454): 903-995.

Huang, N. E., Wu, M. L., Qu, W., Long, S. R., Shen, S. S. P. (2003). Applications of Hilbert–Huang transform to non-stationary financial time series analysis. *Appl. Stochastic Models Bus. Ind.* 19: 245-268.

Huang, W., Bing, X.B., Hilton, A. (2004). Forecasting Flow In Apalachicola River Using Neural Networks. *Hydrological Processes.* 18: 2545-2564.

Huang, N. E. and Wu, Z. (2008). A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*. 46(2): RG2006.

Hsu, S. -H., Hsieh, J. J., P.-A., Chih, T. -C., Hsu, K.-C. (2009). A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression. *Expert Systems with Applications.* 36(4): 7947-7951.

Jain, A. and A. M. Kumar (2007). Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing*. 7(2): 585-592.

Jauhir, H ., Zain, M. S., Mustafa, Z. Khamis, A. (2001). Ramalan Kepekatan Oksigen Terlarut Menggunakan Analisis Siri Masa di Sungai Langat, Hulu Langat Selangor. *Matematika* 17(2): 77-87.

Jolliffe, I.T. (2002). *Principal Components Analysis*. Second Edition. New York. Springer.

John, S. T. and Christianini, N. (2004). *Kernel Methods for Pattern Analysis.* Cambridge University Press.

Kang, Y.W., Li, J., Cao, G.Y., Tu, H.Y., Li, J. Yang, J. (2008). Dynamic temperature modeling 10 of an SOFC using least square support vector machines. *Journal of Power Sources.* 179: 683-692.

Karunanithi, N., W. J. Grenney, et al. (1994). Neural Networks for River Flow Prediction. *Journal of Computing in Civil Engineering.* **8**(2): 201-220.

Khan, M. S. and Coulibaly, P. (2006). Application of Support Vector Machine in Lake Water Level Prediction. *Journal of Hydrologic Engineering.* 11(3): 199-205.

Kisi, O. (2004). River flow modeling using artificial neural networks. *Journal of Hydrologic Engineering.* 9(1): 60-63.

Kisi, O. (2007). Streamflow forecasting using different artificial neural network algorithm. *Journal of Hydrologic Engineering ASCE.* 12(5): 532-539.

Kisi, O. (2008). River flow forecasting and estimation using different artificial neural network technique. *Hydrology Research*. 39(1): 27-40.

Kisi, O. and Cimen, M. (2011). A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *Journal of Hydrology*. 399: 132-140.

Kisi, O. (2012). Modeling discharge-suspended sediment relationship using least square support vector machine. *Journal of Hydrology*. 456–457: 110–120.

Kita, I., Yano, A., Nishiyama, S., Aqil, M. (2007). A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behaviour of runoff. *Journal of Hydrology.* 337: 22-34.

Keskin, M. E. and Taylan, D. (2009). Artificial Models For Interbasin Flow Prediction In Southern Turkey. *Journal of Hydrologic Engineering.* 14(7): 752-758.

Knight, D. W. and Shamseldin, A. Y. (2006). *River Basin Modelling for Flood Risk Mitigation*. London, UK. Taylor & Francis.

Kuai, K. Z. and Tsai, C. W. (2012). Identification of varying time scales in sediment transport using the Hilbert–Huang Transform method. *Journal of Hydrology.* 420-421: 245-254.

Kuligowski, R. J., and Barros, A. P., (1998). Localized Precipitation Forecasts from a Numerical Weather Prediction Model Using Artificial Neural Networks. *Weather and Forecasting*. 13: 1194-1204.

Kumar, N. D. and L. Karthikeyan (2013). Predictability of nonstationary time series using wavelet and EMD based ARMA models. *Journal of Hydrology.* 502: 103–119.

Legates, D. R. and G. J. McCabe Jr. (1999). Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. *Water Resources Research.* 35(1): 233–241.

Lee, J. M., Yoo, C. K., Choi, S. W., Peter A. V., Lee, I. B. (2004). Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science.* 59: 223-234.

Lee, T. and Ouarda, T. B. M. J. (2010). Long-term prediction of precipitation and hydrologic extremes with nonstationary oscillation processes. *Journal of Geophysical Research*. 115: D13107.

Lee, T. and Ouarda, T. B. M. J. (2012). Stochastic simulation of nonstationary oscillation hydroclimatic processes using empirical mode decomposition. *Water Resources Research*. 48(W02514).

Lehmann, A. and Rode, M. (2001). Long-Term Behaviour And Cross-Correlation Water Quality Analysis Of The River Elbe, Germany. *Water Research* 35(9): 2153-2160

Lin, G.-F. and Wu, M.-C. (2009). A Hybrid Neural Network Model for Typhoon-Rainfall Forecasting. *Journal of Hydrology*. 375(3-4), 450-458.

Lin, G. F., Chen, G. R., Huang, P. Y. and Chou, Y. C. (2009). Support vector machine-based models for hourly reservoir inflow forecasting during typhoon-warning periods. *Journal of Hydrology*. 3(32): 17-29.

Lin, J. Y., Cheng, C. T., Chau, K. K. (2006). Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*. 51(4): 599-612.

Lin, G. F., Chen, G. R., Huang, P. Y. (2010). Effective typhoon characteristics and their effects on hourly reservoir inflow forecasting. *Advances in Water Resources*. 33: 887-898.

Lin, T. Y., Lin, C. S., Chiu, S. H. (2012). Empirical mode decomposition–based least squares support vector regression for foreign exchange rate forecasting. *Economic Modelling*. 29: 2583-2590.

Liang, H., Bressler, S. L., Desimone, R., Fries, P. (2005). Empirical mode decomposition: a method for analyzing neural data. *Neurocomputing*. 65-66: 801-807.

Liong, S.Y., and Sivapragasam, C. (2002), Flood stage forecasting with support vector machines. *Journal of American Water Resources*. 38(1),173 -186.

Liu, C. (2004). Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 26(5): 572-581.

Liu, Z., Wang, X., Cui, L., Lian, X., Xu, J. (2009). Research on Water Bloom Prediction Based on Least Squares Support Vector Machine. *WRI World Congress on Computer Science and Information Engineering, 2009.*

Lorenz, D.W and Delin, G.N., (2007). A regression model to estimate regional ground water recharge. *Ground Water*. 45: 196–208.

Luchetta, A. and S. Manetti (2003). A real time hydrological forecasting system using a fuzzy clustering approach. *Computers and Geosciences.* 29(9): 1111-1117.

Luk, K. C., Ball, J. E., Sharma, A. (2001). An application of artificial neural networks for rainfall forecasting. *Mathematical and Computer Modelling*. 33(6-7): 683-693.

Mamat, M. and Samad, S. A. (2010). Performance of Radial Basis Function and Support Vector Machine in time series forecasting. *IEEE International Conference on Intelligent and Advanced Systems (ICIAS).*

Maier, H. R. and Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental modelling and software*. 15(1): 101-124.

Marhaba, T. F. and Bengraine, K. (2003). Using principal component analysis to monitor spatial and temporal changes in water quality. *Journal of Hazardous Materials.* B100: 179-195.

Maity, R., Bhagwat, P. P., Bhatnagar, A. (2010). Potential of support vector regression for prediction of monthly streamflow using endogenous property. *Hydrological Processes*. 24: 917–923.

McCuen, R.H. (1989). *Hydrologic Analysis and Design*. Printed Hall Inc. Englewood Cliff, New Jersey.

McMahon, T. A., Kiem, A. S., Peel, M. C., Jordan, P. W., Pegram, G. G. S. (2008). A New Approach to Stochastically Generating Six-Monthly Rainfall Sequences Based on Empirical Mode Decomposition. *Journal of Hyrdometeorological*. 9(6): 1377-1389.

Misra, D., Oommen, T., Agarwal, A., Mishra, S. K., Thompson, A. M. (2009). Application and analysis of support vector machine based simulation for runoff and sediment yield. *Biosystems Engineering*. 103: 527-535.

Mishra, S., Choubey, V., Pandey, S. K., Shukla, J. P. (2014). An Efficient Approach of Support Vector Machine for Runoff Forecasting. *International Journal of Scientific & Engineering Research*. 5(3): 158-166.

Modarres, R. (2007). Streamflow drought time series forecasting. *Stoch. Environ. Res. Risk Assess* 21: 223-233.

Moldovan, D. (2010). Testing the efficiency of the Romanian stock market. *Development, Energy, Environment, Economics*. 378-381.

Muhamad, J. R. and J. N. Hassan (2005). Khabur River flow using artificial neural networks. *Al-Rafidain Engineering.* 13(2): 33-42.

Ni, J., Zhang, C., Ren, L., Yang, S. X. (2012a). Abrupt Event Monitoring for Water Environment System Based on KPCA and SVM. *IEEE Transactions on Instrumentation and Measurement*. 61(4): 980-989.

Ni, J., Ma, H., Ren, L. (2012b). A Time-series Forecasting Approach Based on KPCA-LSSVM for Lake Water Pollution. *9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD2012).*

Nielsen, A. A. and Canty, M. J. (2008). Kernel principal component analysis for change detection. *Image and Signal Processing for Remote Sensing XIV*. 7109-7109.

Noakes, D. J., Mcleod, L. A., and Hipel, K. W. (1985). Forecasting monthly riverflow time series. *International Journal of Forecasting*: 179-190.

Noori, R., Abdoli, M. A. Ameri, A. Jalili, G. M. (2009). Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: A case study of Mashhad. *Environmental Progress and Sustainable Energy*. 28: (249-258).

Noori, R., Khakpour, A., Omidvar, B., Farokhnia, A. (2010). Comparison of ANN and Principal Component Analysis-Multivariate Linear Regression models for predicting the river flow based on developed discrepancy ratio statistic. *Expert Systems with Applications*. 37: 5850-5862.

Okkan, U. and Serbes, Z. A. (2012). Rainfall–runoff modeling using least squares support vector machines. *Environmetrics*. 23: 549-564.

Ouyang, Y. (2005). Evaluation of river water quality monitoring stations by principal component analysis. *Water Research*. 39: 2621-2635.

Page, R. M., Lischeid, G., Epting, J., Huggenberger, P. (2012). Principal component analysis of time series for identifying indicator variables for riverine groundwater extraction management. *Journal of Hydrology*. 432-433: 137-144.

Panagiotidis, T. (2002). Testing the assumption of Linearity. *Economics Bulletin*. 3(29): 1-9.

Parinet, B., Lhote, A., Legube, B. (2004). Principal component analysis: an appropriate tool for water quality evaluation and management—application to a tropical lake system. *Ecological Modelling*. 178: 295-311.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Phil Mag. (6), 2, 559-572.

Plummer, E. A. (2000). *Time series forecasting with feed-forward neural networks: Guidelines and limitations*, Master Thesis. University of Wyoming.

Pinheiro, E., Postolache, O., Girao, P. (2012). Empirical Mode Decomposition and Principal Component Analysis implementation in processing non-invasive cardiovascular signals. *Measurement*. 45(2): 175-181.

Roadknight, C. M., Balls, G. R., Mills, G.E, Palmer-Brown, D. (1997). Modeling complex environmental data. *IEEE Transactions on Neural Networks.* 8(4): 852-862.

Rong, G., Shao, J. D., Lee, J. M. (2009). Learning a data-dependent kernel function for KPCA-based nonlinear process monitoring. *Chemical Engineering Research and Design*. 87: 1471-1480.

Rood, S. B., G. M. Samuelson, (2005). Twentieth-century decline in streamflows from the hydrographic apex of North America. *Journal of Hydrology*. 306: 215–233.

Sahai, A.K., Soman, M.K., Satyan V. (2000). All India Summer Monsoon Rainfall Prediction Using an Artificial Neural Network. *Climate Dynamics*. 16: 291-302.

Samsudin, R., P. Saad, Shabri, A. (2011). River flow time series using least squares support vector machines. *Hydrology and Earth System Sciences.* 15: 1835-1852.

Sanders, N. R. and Manrodt, K. B. (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega.* 31(6): 511-522.

Samui, P. (2011). Application of Least Square Support Vector Machine (LSSVM) for Determination of Evaporation Losses in Reservoirs. *Scientific Research*. 3(4): 431-434.

Sarma, P., Durlofsky, L. J., Aziz, K. (2008). Kernel Principal Component Analysis for Efficient, Differentiable Parameterization of Multipoint Geostatistics. *Math Geoscience*. 40: 3-32.

Schreider, S. Y., Jakeman, A. J., Dyer, B. G., Francis, R. I. (1997). A Combined Deterministic And Self-Adaptive Stochastic Algorithm For Streamflow Forecasting With Application To Catchments Of The Upper Murray Basin, Australia. *Environmental Modelling and Software.* 12(1): 93-104.

Scholkopf, B., Smola, A.J., Muller, K. (1998). Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation.* 10: 1299-1319.

Scholkopf, B. and Smola, A.J. (2002). *Learning with Kernels.* MIT Press.

Seok, H. H., Dae, H. H., Joong, H. K. (2012). Forecasting Performance of LS-SVM for Nonlinear Hydrological Time Series. *Water Engineering*. 16(5): 870-882.

Sivapragasam, C., Liong, S. Y., Pasha, M. F. K. (2001). Rainfall and runoff forecasting with SSA–SVM approach. *Journal of Hydroinformatics*. 3. 141-152.

Shabri, A. (2001). Comparision Of Time Series Forecasting Methods Using Neural Networks And Box-Jenkins Model. *Matematika.* 17(1): 1-6.

Sharda, R. (1994). Neural Networks for the MS/OR analyst: An application bibliography. *Interfaces.* 24(2): 116-130.

Shabri, A. and Suhartono, (2012). Streamflow forecasting using least-squares support vector machines. *Hydrological Sciences Journal*. 57(7): 1275-1293.

Stathis, D. and Myronidis, D. (2009). Principal component analysis of precipitation in Thessaly Region (Central Greece). *Global NEST Journal*. 11(4): 467-476.

Sudheer, K. P., Gosain, A. K. Ramasastri, K. S. (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrological Processes.* 16: 1325-1330.

Suykens J.A.K. and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letter*. 9(3): 293-300.

Suykens, J. A. K., Gestel, T. V. (2005). *Least Square Support Vector Machine.* New Jersey: World Scientific.

Tan, C. N. W. and Wittig, G. E. (1993). A study of the parameters of a backpropagation stock price prediction model. *The First New Zealand International Two Stream Conference on Artificial Neural Networks and Expert Systems (ANNES.* University of Otago, Dunedin, New Zealand, IEEE Computer Society Press.

Tang, B., Dong, S., Song, T. (2012). Method for eliminating mode mixing of empirical mode decomposition based on the revised blind source separation. *Signal Processing*. 92(1): 248–258.

Tay, F. E. H., and Cao, L. J. (2001). Improved Financial Time Series Forecasting By Combining Support Vector Machines with Self-Organizing Feature Map. *Intelligent Data Analysis.* 5. 339–354.

Theodoridis, S. and Koutroumbas, K. (2009). *Pattern Recognition. 4th Edition.* Academic Press.

Twining, C. J. and Taylor, C. J. (2003). The use of kernel principal component analysis to model data distributions. *Pattern Recognition*. 36: 217-227.

Vapnik., V. (1995). *The Nature of Statistical Learning Theory*. New York: Springer.

Viessman, W. J., and Lewis, G. L.(1989). *Introduction To Hydrology*. Singapore: Harper and Row.

Wang, H. and Hu, D. (2005). Comparison of SVM and LSSVM for Regression. *International Conference on Neural Networks And Brain, 2005.* 1: 279-283.

Wang, W., Gelder, P. H. A. J. M. V., Vrijling, J. K., Ma, J. (2006). Forecasting Daily Streamflow Using Hybrid Ann Models. *Journal of Hydrology*. 324 (1-4):383-399.

Wang, W. (2006). *Stochasticity, nonlinearity and forecasting of streamflow processes.* IOS Press.

Wang, W.-C., K.-W. Chau, (2009a). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*. 374(3-4): 294-306.

Wang, S., Zhang, X., Yu, L., Lai, K. L. (2009b). Estimating the impact of extreme events on crude oil price: An EMD-based event analysis method. *Energy Economics*. 31: 768–778.

Wang, W., Zhang, R., Mac, Y. (2010). Approximations of the standard principal components analysis and kernel PCA. *Expert Systems with Applications.* 37: 6531–6537.

Wang, S., Yu, L., Tang, L., Wang, S. (2011). A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in China. *Energy*. 36: 6542-6554.

Wu, C. L., K. W. Chau, et al. (2008). River stage prediction based on a distributed support vector regression. *Journal of Hydrology*. 358(1-2): 96-111.

Wu, C. L., Chau, K. W., Li, Y. S. (2009). Methods to improve neural network performance in daily flows prediction. *Journal of Hydrology*. 372(1-4): 80-93.

Wu, C. L. (2010). *Hydrological Predictions Using Data-Driven Models Coupled with Data Preprocessing Techniques*. **PhD Thesis**. The Hong Kong Polytechnic University.

Wu, C. L., Chau, K. W., Fan, C. (2010a). Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *Journal of Hydrology*. 389(1-2): 146-167.

Wu, C. L. and Chau, K. W. (2010b). Data-Driven Models for Monthly Streamflow time series prediction. *Engineering Applications of Artificial Intelligence*. 23: 1350-1367.

Xu, Y., Yang, J. Y., Lu, J. F. (2005). An efficient kernel-based nonlinear regression method for two-class classification. Proceedings of the 2005 International Conference on Machine Learning and Cybernetics. 4442 4445.

Xu, Y., Zhang, D., Song, F., Yang, J. Y., Jing, Z., Li, M. (2007). A method for speeding up feature extraction based on KPCA. *Neurocomputing*. 70: 1056-1061.

Xu, W., Kuang, F., Zhang, S., Wang, Y., Liu, K. (2012). A Novel Approach of KPCA and SVM for Intrusion Detection. *Journal of Computational Information Systems*. 8(8): 3237-3244.

Ye, J. and T. Xiong (2007). SVM versus Least Squares SVM. The 11th International Conference on Artificial Intelligence and Statistics (AISTATS).640-647.

Yeh, C. C., Chen, C. F., Lai, M. C. (2012). Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems*. 26: 281–287.

Yu, P.-S., Chen, S.-T., Chang, I. F. (2006). Support vector regression for real-time flood stage forecasting. *Journal of Hydrology*. 328(3-4): 704-716.

Yu, X. and Liong, S.-Y. (2007). Forecasting of hydrologic time series with ridge regression in feature space. *Journal of Hydrology*. 332(3-4): 290-302

Yu, L., Wang, S. Y., Lai, K., (2008). An EMD-Based Neural Network Ensemble Learning Model for World Crude Oil Spot Price Forecasting. *Soft Computing Applications in Business.* 230: 261-271.

Yunrong, X. and Liangzhong, J. (2009). Water Quality Prediction Using LS-SVM And Particle Swarm Optimization. *Second International Workshop on Knowledge Discovery And Data Mining, 900-904.*

Zhang, Y. (1992). *Prediction Of Traffic Fluctuation In Telephone Networks With Neural Networks.* IEEE. 909-912.

Zhang, G., Eddy, P. B., Michael, Y. H. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting.* 14(1): 35-62.

Zhang, G. P. (2003). Time Series Forecasting Using A Hybrid ARIMA And Neural Network Model. *Neurocomputing,* 50: 159-175.

Zhang, Y., Li, Z., Zhou, H. (2010). Statistical analysis and adaptive technique for dynamical process monitoring. *Chemical Engineering Research and Design***.** 88(10): 1381-1392.

Zhao, Y., Dong, Z., Li, Q. (2012). Application Study of Least Squares Support Vector Machines in Streamflow Forecast. *Applied Mechanics and Materials*. 212-213: 436-440.

Zhou, W., S. Li, (2013). Feature extraction and recognition of ictal EEG using EMD and SVM. *Computers in Biology and Medicine***.** 43: 807-816.

Zhu, Y., and Day, R. L., (2009). Regression modelling of streamflow, baseflow, and runoff using geographic information systems. *Journal of Environmental Management.* 90: 946–953.

Zou, H. F., Xia, G. P., Yang, F. T., Wang, H. Y. (2007). An Investigation and Comparison of Artificial Neural Network And Time Series Models For Chinese Food Grain Price Forecasting. *Neurocomputing.* 70(16-18): 2913-2923.