FORECASTING MODELS FOR DROUGHT INDEX USING EMPIRICAL WAVELET TRANSFORM AND STOCHASTIC RECONSTRUCTION APPROACH

MUHAMMAD AKRAM BIN SHAARI

UNIVERSITI TEKNOLOGI MALAYSIA

FORECASTING MODELS FOR DROUGHT INDEX USING EMPIRICAL WAVELET TRANSFORM AND STOCHASTIC RECONSTRUCTION APPROACH

MUHAMMAD AKRAM BIN SHAARI

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Philosophy

School of Computing
Faculty of Engineering
Universiti Teknologi Malaysia

OCTOBER 2021

ACKNOWLEDGEMENT

All praise to Allah, by His permission I was able to complete my Master of Philosophy. I would like to thank my supervisor, Dr. Ruhaidah Binti Samsudin for giving me guidance, support and encouragement throughout my research. I also would like to thank Dr. Ani Bin Shabri that helped me a lot in completing my research. I would also thank my family for providing the encouragement in completing my study. Lastly, I would like to express my gratitude to everyone that helped me during my study.

ABSTRACT

Droughts are natural disasters and extreme climate events with a large impact on different areas of the economy, agriculture, water resources, tourism, and ecosystems. Hence, the ability to forecast drought is important to manage water resources for agricultural and industrial uses. Traditionally, single models have been introduced to forecast the drought data; however, single models may not be suitable to capture the nonlinear nature of the data. Therefore, this study proposed the Empirical Wavelet Transform (EWT) and Stochastic Reconstruction based on Gaussian Process Regression (GPR) and ARIMA models. The study aims to reduce the computation complexity and enhance forecasting accuracy of decomposition ensemble model by incorporating intrinsic mode functions (IMFs) reconstruction method. The proposed model comprises four steps: (i) decomposing the complex data into several IMFs using the EWT method; (ii) reconstructing the decomposed IMFs through autocorrelation into stochastic and deterministic components; (iii) forecasting every reconstructed component using GPR and ARIMA models; (iv) ensemble all forecasted components for the final output. The Standard Precipitation Index (SPI) data from Arau, Perlis; and Gua Musang, Kelantan were employed in this study for the purpose of illustration and verification. The performance of the proposed model was then compared with the following models: ARIMA, GPR, EWT-ARIMA, and EWT-GPR. Based on percentage comparisons, for the Arau region, the EWT-Stochastic Reconstruction-GPR showed improvement in accuracy with reductions of RMSE over the following models: ARIMA (11.90%), GPR (12.71%), EWT-ARIMA (8.48%), EWT-GPR (1.54%) and EWT-Stochastic Reconstruction-ARIMA (3.34%). Similarly, for the Gua Musang region, EWT- Stochastic Reconstruction-GPR yielded reductions of RMSE by around 30.40%, 32.94%, 18.87%, 4.39%, and 20.24% compared to ARIMA, GPR, EWT-ARIMA, EWT-GPR, and EWT-Stochastic Reconstruction-ARIMA models respectively. The empirical results indicated that the EWT-Stochastic Reconstruction-GPR model is the best model for forecasting drought data, followed by EWT-GP, EWT-Stochastic Reconstruction-ARIMA, EWT-ARIMA, ARIMA, and GPR models. In conclusion, the proposed method of reconstruction of IMFs based on autocorrelation enhanced the forecasting accuracy of the EWT model.

ABSTRAK

Kemarau adalah bencana alam dan peristiwa iklim yang ekstrem yang memberikan impak besar kepada berbagai bidang ekonomi, pertanian, sumber air, pelancongan, dan ekosistem. Justeru itu, keupayaan untuk meramal kemarau adalah penting untuk menguruskan sumber air bagi kegunaan pertanian dan perindustrian. Secara tradisional, model tunggal telah diperkenalkan untuk meramalkan data kemarau. Walau bagaimanapun, model tunggal mungkin tidak sesuai untuk menghuraikan sifat data yang tidak linear. Oleh itu, *Empirical Wavelet Transform* (EWT) dan *Stochastic Reconstruction* berdasarkan *Gaussian Process Regression* (GPR) dan model ARIMA adalah dicadangkan di dalam kajian ini. Kajian ini bertujuan untuk mengurangkan kerumitan pengiraan dan meningkatkan ketepatan ramalan model himpunan penguraian dengan memasukkan kaedah pembinaan semula *Intrinsic Mode Function* (IMF). Model cadangan merangkumi empat langkah; (i) menguraikan data kompleks menjadi beberapa IMF menggunakan kaedah EWT; (ii) membina semula IMF yang terurai melalui autokorelasi menjadi komponen stokastik dan deterministik; (iii) meramalkan setiap komponen yang dibina semula menggunakan model GPR dan ARIMA; (iv) menyusun semua komponen yang diramalkan untuk hasil akhir. Data Indeks Pemendakan Standard (SPI) dari Arau, Perlis; dan Gua Musang, Kelantan digunakan sebagai sampel kajian untuk tujuan illustrasi dan pengesahan. Prestasi model yang dicadangkan kemudiannya dibandingkan dengan model ARIMA, GPR, EWT-ARIMA dan EWT-GPR. Berdasarkan perbandingan peratusan, untuk wilayah Arau, EWT-*Stochastic Reconstruction*-GPR menunjukkan peningkatan ketepatan dengan pengurangan RMSE berbanding dengan model ARIMA (11.90%), GPR (12.71%), EWT-ARIMA (8.48%), EWT-GPR (1.54%), dan EWT-*Stochastic Reconstruction*-ARIMA (3.34%). Begitu juga untuk wilayah Gua Musang, *EWT-Stochastic Reconstruction-GPR* menghasilkan pengurangan RMSE sekitar 30.40%, 32.94%, 18.87%, 4.39% dan 20.24% berbanding ARIMA, GPR, EWT-ARIMA, EWT-GPR dan *EWT-Stochastic Reconstruction-ARIMA* masing-masing. Hasil empirik menunjukkan bahawa model EWT-*Stochastic Reconstruction*-GPR adalah model terbaik untuk meramalkan data kemarau diikuti oleh model EWT-GP, model EWT-*Stochastic Reconstruction*-ARIMA, model EWT-ARIMA, model ARIMA dan model GPR. Kesimpulannya, kaedah cadangan pembinaan semula IMF berdasarkan autokorelasi telah meningkatkan ketepatan ramalan model EWT.

TABLE OF CONTENTS

	TITLE	PAGE
	DECLARATION	**iii**
	ACKNOWLEDGEMENT	**iv**
	ABSTRACT	**v**
	ABSTRAK	**vi**
	TABLE OF CONTENTS	**vii**
	LIST OF TABLES	**xi**
	LIST OF FIGURES	**xiii**
	LIST OF ABBREVIATIONS	**xv**
	LIST OF APPENDICES	**xvi**
CHAPTER 1	**INTRODUCTION**	**1**
1.1	Overview	1
1.2	Background of Study	2
1.3	Problem Statement	5
1.4	Objectives	6
1.5	Scopes	7
1.6	Significance of the Study	7
1.7	Thesis Organization	7
CHAPTER 2	**LITERATURE REVIEW**	**9**
2.1	Review on Existing Methods.	9
2.2	Drought Modelling	9
	2.2.1 Drought Indices	10
	2.2.2 Previous work on drought forecasting	12
2.3	ARIMA	14
2.4	Gaussian Process Regression	15
2.5	Data Decomposition – Empirical Wavelet Transform (EWT)	18

	2.5.1	Determination of Number of Modes	22
		2.5.1.1 Otsu's Method	23
		2.5.1.2 K-Means	25
		2.5.1.3 Half-Normal	26
2.6		Subseries Reconstruction	26
2.7		Summary	28
CHAPTER 3		**RESEARCH METHODOLOGY**	**31**
3.1		Introduction	31
3.2		Research Framework	31
3.3		Data Preparation	33
	3.3.1	Data Collection	33
	3.3.2	Data Splitting	33
3.4		Research Process	34
	3.4.1	Data Decomposition Method	34
		3.4.1.1 Empirical Wavelet Transform	35
	3.4.2	Model Development	37
		3.4.2.1 ARIMA Model Development	37
		3.4.2.2 EWT-ARIMA	40
		3.4.2.3 Gaussian Process Regression	41
	3.4.3	EWT-GPR	43
	3.4.4	Proposed Work: EWT-GPR Based on Stochastic Reconstruction	44
	3.4.5	Comparative Output Analysis	46
3.5		Summary	46
CHAPTER 4		**EXPERIMENTS AND RESULT ANALYSIS**	**47**
4.1		Introduction	47
4.2		SPI Calculation	47
	4.2.1	Case Study 1- Arau, Perlis	50
	4.2.2	Case Study 2- Gua Musang, Kelantan	51
4.3		Data Splitting	52
4.4		Single Model	54

	4.4.1	ARIMA	54
		4.4.1.1 Case Study 1- Arau	54
		4.4.1.2 Case Study 2 – Gua Musang	60
	4.4.2	GPR	65
		4.4.2.1 Case Study 1 – Arau	65
		4.4.2.2 Case Study 2 – Gua Musang	66
4.5	EWT Based Models		68
	4.5.1	EWT Decomposition	68
		4.5.1.1 Case Study 1 – Arau	69
		4.5.1.2 Case Study 2 – Gua Musang	71
	4.5.2	EWT-ARIMA	73
		4.5.2.1 Case Study 1 – Arau	73
		4.5.2.2 Case Study 2 – Gua Musang	75
	4.5.3	EWT-GPR	77
		4.5.3.1 Case Study 1 – Arau	78
		4.5.3.2 Case Study 2 – Gua Musang	80
4.6	EWT-Stochastic Reconstruction Based Models		82
	4.6.1	Determination of Stochastic and Deterministic IMFs	82
		4.6.1.1 Case Study 1 – Arau	83
		4.6.1.2 Case Study 2 – Gua Musang	85
	4.6.2	Stochastic Reconstruction - ARIMA	87
		4.6.2.1 Case Study 1 – Arau	87
		4.6.2.2 Case Study 2 – Gua Musang	89
	4.6.3	Stochastic Reconstruction – GPR	91
		4.6.3.1 Case Study 1 - Arau	91
		4.6.3.2 Case Study 2 – Gua Musang	93
4.7	Summary		95
CHAPTER 5	**COMPARISONS AND DISCUSSIONS**		**97**
5.1	Introduction		97
5.2	Comparison of Forecasting Results		97

5.2.1 Case Study 1 – Arau 97
5.2.2 Case Study 2 – Gua Musang 106
5.3 Discussions 110
5.3.1 Discussion on single models 111
5.3.2 Discussion on EWT-based Models 111
5.3.3 Discussion on EWT-Stochastic Reconstruction based Models 113
5.4 Summary 114

CHAPTER 6 CONCLUSIONS 115
6.1 Introduction 115
6.2 Contribution 115
6.3 Future Work 118

REFERENCES 119

LIST OF PUBLICATIONS 217

LIST OF TABLES

TABLE NO.	TITLE	PAGE
Table 2.1	Classifications of SPI	12
Table 2.2	Summary of reviewed GPR models	17
Table 2.3	Summary of reviewed EWT based models	20
Table 4.1	Statistical description of rainfall data for Arau and Gua Musang	48
Table 4.2	AIC of ARIMA models for SPI3	56
Table 4.3	ARIMA models obtained for each SPI in Arau	58
Table 4.4	ARIMA forecast accuracy for Arau	59
Table 4.5	AIC of ARIMA models for SPI3	62
Table 4.6	ARIMA models obtained for each SPI for Gua Musang	64
Table 4.7	ARIMA forecast accuracy for Gua Musang for each SPI	65
Table 4.8	Inputs used for GPR model for each SPI	66
Table 4.9	GPR forecast accuracy for Arau for each SPI	66
Table 4.10	Inputs used for GPR model for each SPI	67
Table 4.11	GPR forecast accuracy for Gua Musang for each SPI	67
Table 4.12	Number of IMFs produced by EWT	69
Table 4.13	Number of IMFs produced by EWT	71
Table 4.14	ARIMA models for each IMF of each SPI for Arau	74
Table 4.15	Forecast accuracy of EWT-ARIMA for Arau	75
Table 4.16	ARIMA models for each IMF of each SPI for Gua Musang	76
Table 4.17	Forecast accuracy of EWT-ARIMA for Gua Musang	77
Table 4.18	Input variables used for each IMF	79
Table 4.19	Forecast Accuracy of EWT-GPR for Arau	80
Table 4.20	Input variables used for each IMF	81
Table 4.21	Forecast Accuracy of EWT-GPR for Gua Musang	82
Table 4.22	Autocorrelation of lag 1 of Each IMFs	83

Table 4.23	Deterministic and Stochastic Components based on Threshold values for Arau	84
Table 4.24	Autocorrelation of Each IMFs	86
Table 4.25	Deterministic and Stochastic Components based on Threshold values for Gua Musang	86
Table 4.26	ARIMA model obtained for reconstructed stochastic series	88
Table 4.27	EWT-Stochastic Reconstruction-ARIMA forecasts accuracy for Arau	89
Table 4.28	EWT-Stochastic Reconstruction-ARIMA forecasts accuracy for Gua Musang	91
Table 4.29	Input variables used for the reconstructed stochastic series	92
Table 4.30	EWT-Stochastic Reconstruction-GPR forecasts accuracy for Arau	93
Table 4.31	Input variables used for the reconstructed stochastic series	94
Table 4.32	EWT-Stochastic Reconstruction-GPR forecasts accuracy for Gua Musang	95

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
Figure 2.1	Scale-space Representation of a Histogram	24
Figure 2.2	Histogram mode detection - All scale-space curves larger than T are meaningful modes	25
Figure 3.1	Research Framework	32
Figure 3.2	EWT-ARIMA development	40
Figure 3.3	Workflow of EWT-GPR	43
Figure 3.4	Workflow of EWT-Stochasic Reconstruction-GPR	44
Figure 4.1	SPI 3,6,9 and 12 for Arau Perlis	51
Figure 4.2	SPI 3,6,9 and 12 for Gua Musang, Kelantan	52
Figure 4.3	Data Splitting Arau and Gua Musang Region	53
Figure 4.4	SPI 3 based on rainfall data in Arau, Perlis.	54
Figure 4.5	ACF and PACF of SPI 3	56
Figure 4.6	Diagnostic Check for ARIMA (2,0,2) model using SPI3 data from Arau	57
Figure 4.7	SPI 3 for Gua Musang	60
Figure 4.8	PACF and ACF of SPI3 from Gua Musang	61
Figure 4.9	Diagnostic Check for ARIMA(1,0,2) model using SPI3 data from Gua Musang	63
Figure 4.10	Boundaries detected by Otsu's method for SPI 12	70
Figure 4.11	EWT decomposition on SPI 12	70
Figure 4.12	Boundaries detected by Otsu's method for SPI 9	72
Figure 4.13	EWT decomposition on SPI 9	72
Figure 4.14	PACF and ACF of IMF1 for SPI 9	74
Figure 4.15	ACF and PACF of IMF2 of SPI 9	76
Figure 4.16	PACF of IMF1 of SPI 6	78
Figure 4.17	PACF of IMF1 of SPI 6	81

Figure 4.18	The Dereterministic and scothastic IMFs for SPI 6, IMF1, IMF2, IMF5 and reconstructed stochastic IMF	85
Figure 4.19	ACF and PACF of Reconstructed Stochastic Component for SPI 6 using threshold value 0.5	89
Figure 4.20	ACF and PACF of Reconstructed Stochastic component of SPI 6 using threshold value 0.5	90
Figure 4.21	PACF of stochastic components for SPI 3	92
Figure 4.22	PACF plot for SPI3 Stochastic components	94
Figure 5.1	Graph of observed and predicted values for SPI 3 for Arau.	105

LIST OF ABBREVIATIONS

ACF	-	Autocorrelation Function
ANN	-	Artificial Neural Network
AI	-	Artificial Intelligence
AR	-	Autoregressive
ARIMA	-	Autoregressive Integrated Moving Average
DWT	-	Discrete Wavelet Transform
EWT	-	Empirical Wavelet Transform
EMD	-	Empirical Mode Decomposition
GPR	-	Gaussian Process Regression
MA	-	Moving Average
PACF	-	Partial Autocorrelation Functions
SPI	-	Standard Precipitation Index
SVM	-	Support Vector Machine
SVR	-	Support Vector Regression

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
Appendix A	Arau ACF and PACF	126
Appendix B	Gua Musang ACF and PACF	131
Appendix C	AIC of ARIMA models for Arau	136
Appendix D	AIC of ARIMA models for Gua Musang	140
Appendix E	Arau EWT Decomposition	143
Appendix F	Gua Musang EWT Decomposition	151
Appendix G	Arau ACF and PACF EWT	159
Appendix H	Gua Musang ACF and PACF EWT	187

CHAPTER 1

INTRODUCTION

1.1 Overview

Time series forecasting is a process of predicting the future observations where these observations have never been made before. Time Series is a sequence of data points that consists of successive measurement made over a time interval. Time series is commonly used in mathematics, finance, statistics, weather, engineering, and applied science. The process of making prediction relies on past and present data where the data are analysed. In most fields, forecasting can help in decision making and risk management as the result of the forecasting can be used to devise decision or plan. With regards to drought, an accurate forecasting of drought could enable the appropriate party to prepare for it to minimise the negative effects of the drought.

There are many forecasting models that have been developed for forecasting time series data. These techniques are normally based on statistical technique such as autoregressive integrated moving average (ARIMA), neural network, support vector machine (SVM) and gaussian process regression (GPR). The accuracy of a forecast can have an impact on the decision-making process, thus the research in improving the effectiveness of forecasting accuracy is still ongoing (A. K. Mishra & Singh, 2011).

This study is focused on the development of a hybrid model for forecasting drought. Drought is a natural hazard and is defined as deficiency in precipitation for an elongated period, which is usually a season or more that cause water shortages. World Meteorological Organization defined drought as a continuous, elongated shortage of precipitation. United Nations (UN) Convention to Combat Drought and Desertification defined drought as the naturally occurring phenomenon that happens when precipitation has been considerably lower than usual recorded levels, leading to serious hydrological imbalances which negatively impacts land resource production systems. The definition of drought varies per variable utilised to explain it. Thus, there are multiple categories for classifying drought.

American Meteorological Society (1997) classified droughts by four classification, namely meteorological, agricultural, hydrological, and socioeconomics. The deficit of precipitation leads to meteorological drought, soil moisture leads to agricultural drought, and stream flow leads to hydrological drought. Socio-economic drought is when water resource fails to meet the water demand. Therefore, drought can be associated with supply of and demand of water in economic good, in which the supply cannot match the demand for economic good (American Meteorological Society, 1997).

Droughts have negative impact on vegetation, animals, and people in the form of water shortages. Drought can be considered as a normal, periodical feature of a climate. Droughts occur across every climatic zone. Drought can be recognised by the drop of precipitation for an extended period of time over a timescale such as a season or a year (Mishra & Singh, 2010).

Drought can affect communities and environment in many ways. The strength of the drought influences the impact of the drought, which is considered by the period, or the area affected by the drought. When water supply runs low, the local government set a restriction on water, limiting the activities that can be performed in a community. However, to prepare for drought and issue warning to the masses, forecasting drought is required. The damage that are caused by drought highlights the importance of drought forecasting.

1.2 Background of Study

Forecasting drought can be done through the use of a physical or conceptual and data driven models. Physical models require many types of data as its input, thus they are result in complex models. For data driven models, they are accurate in various hydrology forecasting applications (Belayneh et al., 2014). Since they also have minimum information requirements and rapid development times, they are widely used in hydrological forecasting.

Stochastic models are among the frequently used models to forecast drought hydrologically. ARIMA & Seasonal ARIMA are among the most widely used models to forecast drought since they are simple yet effective (Mossad & Alazba, 2015). Stochastic models are good in forecasting linear time series, however they fall short in forecasting non-linear data (Hu & Wang, 2015). ARIMA is also not able to forecast time series with high amount of noise as ARIMA produced lower accuracy for lower SPI such as SPI 3 where the lower SPI contains more noise compared to higher SPI series (Mishra & Desai, 2005).

Using Artificial Intelligence (AI) to forecast drought is also a popular method that have been studied. AI methods have shown to provide great performance and accuracy in drought forecasting (Belayneh & Adamowski, 2012; Deo & Şahin, 2015; Soh et al., 2018). AI have been found to be flexible and adaptable in predicting the occurrence of drought that have varying durations, frequencies and intensities. For AI, the popular models include artificial neural network (ANN), support vector regression (SVR) and support vector machine (SVM). ANN is one of the non-linear methods used to forecast drought (Mishra & Desai, 2006). Using ANN provided a better forecast compared to ARIMA models. However, ANN requires estimation of a large number of parameters, thus it is complicated to choose the appropriate architecture for the model. Several researchers have successfully explored on Gaussian Process Regression (GPR) as a forecasting technique (Hu & Wang, 2015; A. Y. Sun et al., 2014). GPR is able to forecast non-linear timeseries accurately since Gaussian Processes (GP) are useful as priors over functions for doing non-linear regression. However, there are not many studies done on the application of GPR in forecasting timeseries with both linear and non-linear characteristics.

To further improve the accuracy of drought forecasting several researchers have used data decomposition technique to provide the models with simpler inputs. Empirical Mode Decomposition (EMD), discrete wavelet transform (DWT), and empirical wavelet transform (EWT) are among the methods that are commonly hybridised with forecasting models (Belayneh et al., 2014; Hu & Wang, 2015; Mishra & Singh, 2011). Numerous studies come to a conclusion that using these mentioned data decomposition methods improved the forecast accuracy in drought forecasting (Belayneh et al., 2014, 2016; Belayneh & Adamowski, 2012) and wind speed forecasting (Guo et al., 2012; Hu et al., 2015; Hu & Wang, 2015). EMD provides solution of non-linear and non-stationary data by decomposing the non-linear and non-stationary behaviour of the time series into a series of valuable independent time resolutions (Tang et al., 2012). EWT is conceptually similar to EMD thus it also provides similar solution to non-linear and non-stochastic data (Gilles, 2013). Several studies have found that by using EWT improves the forecast accuracy of several models, including GPR models (H. Liu, Wu, et al., 2018; W. Sun & Wang, 2018).

While data decomposition techniques such as EWT and EMD have been used as pre-processing technique, these models do not consider the differences of the data characteristics after decomposed. There is lack of work that are done to address the difference of the data produced. For EMD decomposed data, it was found that the residual series have small correlation between them and also zero forecasting result may be obtained (Shabri, 2016).

Yu et al. (2017) found that by extracting the trend from IMF1 produced by EMD, the forecast accuracy can be improved. Since IMF1 have the most disordered data and have little regularity, it is hard to accurately forecast. This data can also be described as stochastic data (Aamir & Shabri, 2018). However, it still has meaningful data as model that utilized the trend extracted from IMF1 performed better.

Another method to improve the input data is by clustering the input. Clustering technique can be useful to group the similarities of the IMFs from EMD and EWT and cluster them through the calculation of their dissimilarities matrices. Shabri (2016) proposed MEMD-ARIMA that implemented k-means clustering and silhouette analysis to cluster components resulting from EMD. The IMFs and residual from EMD were reconstructed into several components. The result shows that the proposed MEMD-ARIMA model have higher accuracy compared to ARIMA and EMD-ARIMA. Aamir and Shabri (2018) on the other hand classifies the components produced from Ensemble empirical mode decomposition into stochastic component and deterministic component by grouping them using autocorrelation function. In the study, the stochastic IMFs were reconstructed into a single component and deterministic IMFs were reconstructed into a single component. The result shows that the proposed EEMD-RSD-ARIMA model performed better than EEMD-ARIMA.

1.3 Problem Statement

Drought cannot be completely prevented. However, it can be predicted. The population growth and the expansion of industrial and agricultural sector causes the increase in the demand for water over time, and various parts of the world has been experiencing water scarcity (Mishra & Singh, 2010). In the recent years, an increase in the severity level of drought and floods has been experienced. Even though Malaysia is a humid country, drought regularly occurs in the country (Shaaban & Low, 2003). Droughts are able to reduce the level of agricultural output, which leads to the temporary shut-down of capital in the downstream manufacturing industries, therefore, drought is able to reduce the GDP of a region(Kilimani et al., 2015). Knowledge in droughts could bring effective planning and management of water resource.

Because of the impact of drought, predicting it will bring advantage to the community. With an accurate prediction, the impact of a drought can be reduced. People can prepare for the drought and reduce water usage beforehand, and government can prepare the infrastructure to combat the drought. Therefore, it is crucial to understand the work of drought mechanism and provide contribution to the existing forecasting model to improve the accuracy of the forecast and reduce the forecast complexity. However, the non-linear and stochastic characteristic of the drought data makes the forecasting process difficult.

The aim of this study is to develop a new forecasting model which is able to forecast SPI of drought data and also overcome the weakness of the existing models such as ARIMA, ANN, GPR and LSSVM. Previously, EWT has been used as a data pre-processing by researchers to deals with non-linear and non-stationary data. GPR has also been used as a forecasting model by researchers in the past as its development is less complex than ANN. Thus, this research preferred the hybridisation of EWT and GPR which have been shown in other domains to provide superior performance compared to single and stochastic models. Apart from that, the very few applications of EWT-GPR in drought forecasting has motivated this research. In this study, a modified model that is based on EWT-GPR is proposed where stochastic and deterministic influences is used to reconstruct the stochastic IMFs of SPI data from EWT into a single stochastic component. Since stochastic data is harder to forecast compared to deterministic ones, this will lead to better input variables for the GPR and thus able to improve its accuracy in forecasting SPI values. Therefore, the research question derived is:

> "How to design and develop a modified model consisting of EWT and Stochastic Reconstruction with ARIMA and GPR so that it can improve drought forecasting accuracy using SPI".

1.4 Objectives

Based on the statement of problems stated above, the following are the objectives proposed for this study: -

(1) To evaluate the capability of hybridising EWT with ARIMA and GPR in forecasting SPI for drought.

(2) To propose the EWT-Stochastic reconstruction-GPR model and EWT-Stochastic Reconstruction-ARIMA model to provide the best input for the EWT based forecasting model.

(3) To compare the performance of the proposed EWT-Stochastic Reconstruction-GPR and EWT-Stochastic Reconstruction-ARIMA with EWT-GPR, GPR, ARIMA, and EWT-ARIMA in drought forecasting using SPI.

1.5 Scopes

Based on the objectives above, the following are the scope of the research:

i) The data analysis uses daily rainfall data from Arau, Perlis and Gua Musang, Kelantan which were provided by Malaysia Meteorological Department.

ii) The research uses standard precipitation index (SPI) for the drought index with SPI 3, 6, 9 and 12.

iii) Two performance measure used to evaluate the performance of the proposed model, which are root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE)

1.6 Significance of the Study

The domain of this research is forecasting drought in Arau, Perlis and Gua Musang, Kelantan. This study proposes a hybrid forecasting model that hopes to increase the accuracy of the forecasts. Therefore, it is hoped that the result of this study will be beneficial to the government and the citizens of Malaysia. With a more accurate method to forecast drought, it is hopeful that the citizens can prepare better for drought, and the government can provide better infrastructure ahead of a drought.

1.7 Thesis Organization

This thesis consists of 6 chapters. Chapter 1 gives the reader an overview on the research areas and its importance. Chapter 2 discusses the previous works in similar area of interests in detail. The indices that are used to represent drought is studied briefly. The methods that have been used to forecasts drought are mentioned. Next, the models that are related with this research is thoroughly analysed. A review is also done on how wavelets technique has been used to improve the performance on the past works. Chapter 3 describes the research framework. The operational framework is described, and the step-by-step procedure is explained in this chapter. The design of the proposed model is also described. Chapter 4 shows the results of the experiments carried out in this study. Chapter 5 discusses and compares the forecasting result for all models used in this study and also the findings from the experiment conducted. Chapter 6 concludes the overall research.

REFERENCES

Aamir, M., & Shabri, A. (2018). Improving Crude Oil Price Forecasting Accuracy via Decomposition and Ensemble Model by Reconstructing the Stochastic and Deterministic Influences. *Advanced Science Letters*, *24*(6), 4337–4342.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

American Meteorological Society. (1997). Meteorological drought-policy statement. *Bull Am Meteorol Soc*, *78*, 847–849. Scopus.

Belayneh, A., & Adamowski, J. (2012). Standard Precipitation Index Drought Forecasting Using Neural Networks, Wavelet Neural Networks, and Support Vector Regression. *Applied Computational Intelligence and Soft Computing*, *2012*, e794061. https://doi.org/10.1155/2012/794061

Belayneh, A., Adamowski, J., & Khalil, B. (2016). Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods. *Sustainable Water Resources Management*, *2*(1), 87–101. https://doi.org/10.1007/s40899-015-0040-5

Belayneh, A., Adamowski, J., Khalil, B., & Ozga-Zielinski, B. (2014). Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology*, *508*, 418–429. https://doi.org/10.1016/j.jhydrol.2013.10.052

Box, G. E., & Jenkins, G. M. (1976). Time series analysis, control, and forecasting. *San Francisco, CA: Holden Day*, *3226*(3228), 10.

Chun, K. P., Wheater, H., & Onof, C. (2013). Prediction of the impact of climate change on drought: An evaluation of six UK catchments using two stochastic approaches. *Hydrological Processes*, *27*(11), 1600–1614. https://doi.org/10.1002/hyp.9259

Deo, R. C., & Şahin, M. (2015). Application of the Artificial Neural Network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using hydrometeorological parameters and climate indices in eastern

Australia. *Atmospheric Research*, *161–162*, 65–81. https://doi.org/10.1016/j.atmosres.2015.03.018

Deo, R. C., Tiwari, M. K., Adamowski, J. F., & Quilty, J. M. (2016). Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. *Stochastic Environmental Research and Risk Assessment*, 1–30. https://doi.org/10.1007/s00477-016-1265-z

Djerbouai, S., & Souag-Gamane, D. (2016). Drought Forecasting Using Neural Networks, Wavelet Neural Networks, and Stochastic Models: Case of the Algerois Basin in North Algeria. *Water Resources Management*, *30*(7), 2445–2464.

Edwards, D. C., & McKee, T. B. (1997). *Characteristics of 20th century drought in the United States at multiple time scales* [Report]. Colorado State University. Libraries. https://dspace.library.colostate.edu:443/handle/10217/170176

Fernandez-Manso, A., Quintano, C., & Fernandez-Manso, O. (2011). Forecast of NDVI in coniferous areas using temporal ARIMA analysis and climatic data at a regional scale. *International Journal of Remote Sensing*, *32*(6), 1595–1617. https://doi.org/10.1080/01431160903586765

Gilles, J. (2013). Empirical wavelet transform. *IEEE Transactions on Signal Processing*, *61*(16), 3999–4010.

Gilles, J., & Heal, K. (2014). A parameterless scale-space approach to find meaningful modes in histograms—Application to image and spectrum segmentation. *International Journal of Wavelets, Multiresolution and Information Processing*, *12*(06), 1450044. https://doi.org/10.1142/S0219691314500441

Guo, Z., Zhao, W., Lu, H., & Wang, J. (2012). Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. *Renewable Energy*, *37*(1), 241–249. https://doi.org/10.1016/j.renene.2011.06.023

Guttman Nathaniel B. (2007). Accepting the standardized precipitation index: A calculation algorithm. *JAWRA Journal of the American Water Resources Association*, *35*(2), 311–322. https://doi.org/10.1111/j.1752-1688.1999.tb03592.x

Han, P., Wang, P. X., Zhang, S. Y., & Zhu, D. H. (2010). Drought forecasting based on the remote sensing data using ARIMA models. *Mathematical and Computer Modelling*, *51*(11–12), 1398–1403. https://doi.org/10.1016/j.mcm.2009.10.031

Hong, Y.-Y., Liu, C.-Y., Chen, S.-J., Huang, W.-C., & Yu, T.-H. (2015). Short-term LMP forecasting using an artificial neural network incorporating empirical mode decomposition. *International Transactions on Electrical Energy Systems*, *25*(9), 1952–1964. https://doi.org/10.1002/etep.1949

Hu, J., & Wang, J. (2015). Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression. *Energy*, *93*, 1456–1466. https://doi.org/10.1016/j.energy.2015.10.041

Hu, J., Wang, J., & Ma, K. (2015). A hybrid technique for short-term wind speed prediction. *Energy*, *81*, 563–574. https://doi.org/10.1016/j.energy.2014.12.074

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: Mathematical, Physical and Engineering Sciences*, *454*(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193

Huang, Y. F., Ang, J. T., Tiong, Y. J., Mirzaei, M., & Amin, M. Z. M. (2016). Drought Forecasting Using SPI and EDI under RCP-8.5 Climate Change Scenarios for Langat River Basin, Malaysia. *Procedia Engineering*, *154*, 710–717. https://doi.org/10.1016/j.proeng.2016.07.573

J. Gilles. (2013). Empirical Wavelet Transform. *IEEE Transactions on Signal Processing*, *61*(16), 3999–4010. https://doi.org/10.1109/TSP.2013.2265222

Jalalkamali, A., Moradi, M., & Moradi, N. (2015). Application of several artificial intelligence models and ARIMAX model for forecasting drought using the Standardized Precipitation Index. *International Journal of Environmental Science and Technology*, *12*(4), 1201–1210. https://doi.org/10.1007/s13762-014-0717-6

Karthikeyan, L., & Nagesh Kumar, D. (2013). Predictability of nonstationary time series using wavelet and EMD based ARMA models. *Journal of Hydrology*, *502*, 103–119. https://doi.org/10.1016/j.jhydrol.2013.08.030

Kilimani, N., van Heerden, J., Bohlmann, H., & Roos, L. (2015). *Impact of drought on the Ugandan economy*. http://2015.essa.org.za/fullpaper/essa_3001.pdf

Kriechbaumer, T., Angus, A., Parsons, D., & Rivas Casado, M. (2014). An improved wavelet–ARIMA approach for forecasting metal prices. *Resources Policy*, *39*, 32–41. https://doi.org/10.1016/j.resourpol.2013.10.005

Liu, H., Mi, X., & Li, Y. (2018). Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. *Energy Conversion and Management*, *156*, 498–514. https://doi.org/10.1016/j.enconman.2017.11.053

Liu, H., Tian, H., & Li, Y. (2012). Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. *Applied Energy*, *98*, 415–424. https://doi.org/10.1016/j.apenergy.2012.04.001

Liu, H., Wu, H., & Li, Y. (2018). Smart wind speed forecasting using EWT decomposition, GWO evolutionary optimization, RELM learning and IEWT reconstruction. *Energy Conversion and Management*, *161*, 266–283. https://doi.org/10.1016/j.enconman.2018.02.006

Liu, W., & Chen, W. (2019). Recent Advancements in Empirical Wavelet Transform and Its Applications. *IEEE Access*, *7*, 103770–103780. https://doi.org/10.1109/ACCESS.2019.2930529

Maity, R., Suman, M., & Verma, N. K. (2016). Drought prediction using a wavelet based approach to model the temporal consequences of different types of droughts. *Journal of Hydrology*, *539*, 417–428. https://doi.org/10.1016/j.jhydrol.2016.05.042

McKee, T. B., Doesken, N. J., Kleist, J., & others. (1993). The relationship of drought frequency and duration to time scales. *Proceedings of the 8th Conference on Applied Climatology*, *17*, 179–183. http://ccc.atmos.colostate.edu/relationshipofdroughtfrequency.pdf

McKee, T., Doesken, N., & Kleist, J. (1995). *Drought monitoring with multiple time scales*. 233–236.

Min, L., & Zhi-min, H. (2013). Research and Prediction of Yellow Soil Moisture Content in Guizhou Province Based on ARIMA Model. In X. H. Liu, K. F. Zhang, & M. Z. Li (Eds.), *Material Design, Processing and Applications, Parts 1-4* (Vols. 690–693, pp. 3076–3081). Trans Tech Publications Ltd.

Mishra, A., Desai, V., & Singh, V. (2007). Drought forecasting using a hybrid stochastic and neural network model. *Journal of Hydrologic Engineering*, *12*(6), 626–638.

Mishra, A. K., & Desai, V. R. (2005). Drought Forecasting Using Stochastic Models. *Stochastic Environmental Research and Risk Assessment*, *19*(5), 326–339.

Mishra, A. K., & Desai, V. R. (2006). Drought forecasting using feed-forward recursive neural network. *Ecological Modelling*, *198*(1–2), 127–138. https://doi.org/10.1016/j.ecolmodel.2006.04.017

Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. *Journal of Hydrology*, *391*(1–2), 202–216. https://doi.org/10.1016/j.jhydrol.2010.07.012

Mishra, A. K., & Singh, V. P. (2011). Drought modeling – A review. *Journal of Hydrology*, *403*(1–2), 157–175. https://doi.org/10.1016/j.jhydrol.2011.03.049

Moreira, E. E. (2016). SPI drought class prediction using log-linear models applied to wet and dry seasons. *Physics and Chemistry of the Earth*, *94*, 136–145. https://doi.org/10.1016/j.pce.2015.10.019

Morid, S., Smakhtin, V., & Moghaddasi, M. (2006). Comparison of seven meteorological indices for drought monitoring in Iran. *International Journal of Climatology*, *26*(7), 971–985. https://doi.org/10.1002/joc.1264

Mossad, A., & Alazba, A. A. (2015). Drought Forecasting Using Stochastic Models in a Hyper-Arid Climate. *Atmosphere*, *6*(4), 410–430. https://doi.org/10.3390/atmos6040410

Nourani, V., Hosseini Baghanam, A., Adamowski, J., & Kisi, O. (2014). Applications of hybrid wavelet–Artificial Intelligence models in hydrology: A review. *Journal of Hydrology*, *514*, 358–377. https://doi.org/10.1016/j.jhydrol.2014.03.057

Ramana, R. V., Krishna, B., Kumar, S. R., & Pandey, N. G. (2013). Monthly Rainfall Prediction Using Wavelet Neural Network Analysis. *Water Resources Management*, *27*(10), 3697–3711. https://doi.org/10.1007/s11269-013-0374-4

Rashid, N. I. A., Samsudin, R., & Shabri, A. (2016). Exchange Rate Forecasting Using Modified Empirical Mode Decomposition and Least Squares Support Vector Machine. *Int. J. Advance Soft Compu. Appl*, *8*(3).

Rashid, N. I. A., Shabri, A., & Samsudin, R. (2017). COMPARISON BETWEEN MEMD-LSSVM AND MEMDARIMA IN FORECASTING EXCHANGE RATE. *Journal of Theoretical and Applied Information Technology*, *95*(2), 328.

Rios, R. A., & de Mello, R. F. (2013). Improving time series modeling by decomposing and analyzing stochastic and deterministic influences. *Signal Processing*, *93*(11), 3001–3013. https://doi.org/10.1016/j.sigpro.2013.04.017

Rohani, A., Taki, M., & Abdollahpour, M. (2018). A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I). *Renewable Energy*, *115*, 411–422. https://doi.org/10.1016/j.renene.2017.08.061

Schlüter, S., & Deuschle, C. (2010). *Using Wavelets for Time Series Forecasting – Does it Pay Off?* https://core.ac.uk/download/pdf/6677041.pdf

Shaaban, A. J., & Low, K. S. (2003). *Droughts in Malaysia: A look at its characteristics, impacts, related policies and management strategies*. 28–29.

Shabri, A. (2015). A hybrid model for stream flow forecasting using wavelet and least Squares support vector machines. *Jurnal Teknologi*, *73*(1).

Shabri, A. (2016). *A modified EMD-ARIMA based on clustering analysis for fishery landing forecasting*. *10*, 1719–1729. https://doi.org/10.12988/ams.2016.6389

Shafie-khah, M., Moghaddam, M. P., & Sheikh-El-Eslami, M. K. (2011). Price forecasting of day-ahead electricity markets using a hybrid forecast method. *Energy Conversion and Management*, *52*(5), 2165–2169. https://doi.org/10.1016/j.enconman.2010.10.047

Soh, Y. W., Koo, C. H., Huang, Y. F., & Fung, K. F. (2018). Application of artificial intelligence models for the prediction of standardized precipitation evapotranspiration index (SPEI) at Langat River Basin, Malaysia. *Computers and Electronics in Agriculture*, *144*, 164–173. https://doi.org/10.1016/j.compag.2017.12.002

Sun, A. Y., Wang, D., & Xu, X. (2014). Monthly streamflow forecasting using Gaussian Process Regression. *Journal of Hydrology*, *511*, 72–81. https://doi.org/10.1016/j.jhydrol.2014.01.023

Sun, W., & Wang, Y. (2018). Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. *Energy Conversion and Management*, *157*, 1–12. https://doi.org/10.1016/j.enconman.2017.11.067

Tan, M. L., Tan, K. C., Chua, V. P., & Chan, N. W. (2017). Evaluation of TRMM Product for Monitoring Drought in the Kelantan River Basin, Malaysia. *Water*, *9*(1), 57. https://doi.org/10.3390/w9010057

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. https://doi.org/10.1016/j.sigpro.2011.07.013

Wang, J., & Hu, J. (2015). A robust combination approach for short-term wind speed forecasting and analysis – Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy*, *93, Part 1*, 41–56. https://doi.org/10.1016/j.energy.2015.08.045

Wilhite, D. A. (2000). *Drought as a natural hazard: Concepts and definitions*.

Wu, H., Svoboda, M. D., Hayes, M. J., Wilhite, D. A., & Wen, F. (2007). Appropriate application of the standardized precipitation index in arid locations and dry seasons. *International Journal of Climatology*, *27*(1), 65–79. https://doi.org/10.1002/joc.1371

Wu, Q., Law, R., & Xu, X. (2012). A sparse Gaussian process regression model for tourism demand forecasting in Hong Kong. *Expert Systems with Applications*, *39*(5), 4769–4774. https://doi.org/10.1016/j.eswa.2011.09.159

Yu, C., Li, Y., & Zhang, M. (2017). Comparative study on three new hybrid models using Elman Neural Network and Empirical Mode Decomposition based technologies improved by Singular Spectrum Analysis for hour-ahead wind speed forecasting. *Energy Conversion and Management*, *147*, 75–85. https://doi.org/10.1016/j.enconman.2017.05.008

Yu, J., Chen, K., Mori, J., & Rashid, M. M. (2013). A Gaussian mixture copula model based localized Gaussian process regression approach for long-term wind speed prediction. *Energy*, *61*, 673–686. https://doi.org/10.1016/j.energy.2013.09.013

Yusof, F., Hui-Mean, F., Suhaila, S., & Ching-Yee, K. (2012). Trend analysis for drought event in peninsular Malaysia. *Jurnal Teknologi (Sciences and Engineering)*, *57*(SUPPL.1), 211–218. Scopus.

Zhang, C., Wei, H., Zhao, X., Liu, T., & Zhang, K. (2016). A Gaussian process regression based hybrid approach for short-term wind speed prediction. *Energy Conversion and Management*, *126*, 1084–1092. https://doi.org/10.1016/j.enconman.2016.08.086

LIST OF PUBLICATIONS

1. Shaari, M. A., Samsudin, R., & bin Shabri Ilman, A. (2018). Comparison of Drought Forecasting Using ARIMA and Empirical Wavelet Transform-ARIMA. In F. Saeed, N. Gazem, S. Patnaik, A. S. Saed Balaid, & F. Mohammed (Eds.), *Recent Trends in Information and Communication Technology* (pp. 449–458). Springer International Publishing.

2. Shaari, M. A., Samsudin, R., & Ilman, A. S. (2018). Forecasting Drought Using Modified Empirical Wavelet Transform-ARIMA with Fuzzy C-Means Clustering. *Indonesian Journal of Electrical Engineering and Computer Science, 11*(3), 1152–1161. https://doi.org/10.11591/ijeecs.v11.i3.pp1152-1161