

**PERFORMANCE OF SHORT, MEDIUM, AND
LONG-TERM STREAMFLOW FORECASTING
USING MACHINE LEARNING**

WONG WAI KIT

UNIVERSITI TUNKU ABDUL RAHMAN

**PERFORMANCE OF SHORT, MEDIUM, AND LONG-TERM
STREAMFLOW FORECASTING USING MACHINE LEARNING**

WONG WAI KIT

**A project report submitted in partial fulfilment of the
requirements for the award of Bachelor of Civil
Engineering with Honours**

**Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman**

May 2023

DECLARATION

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at UTAR or other institutions.

Signature : Waikit

Name : WONG WAI KIT

ID No. : 1805294

Date : 27/4/2023

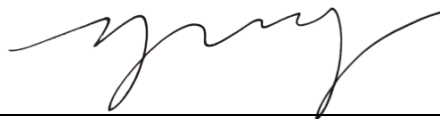
APPROVAL FOR SUBMISSION

I certify that this project report entitled “**PERFORMANCE OF SHORT, MEDIUM AND LONG TERM STREAMFLOW FORECASTING USING MACHINE LEARNING** ” was prepared by **WONG WAI KIT** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Civil Engineering with Honours at Universiti Tunku Abdul Rahman.

Approved by,

Signature

:



Supervisor

:

Prof. Ir. Dr. Huang Yuk Feng

Date

:

27 APRIL 2023

Signature

:

Co-Supervisor

:

Date

:

The copyright of this report belongs to the author under the terms of the copyright Act 1987 as qualified by Intellectual Property Policy of Universiti Tunku Abdul Rahman. Due acknowledgement shall always be made of the use of any material contained in, or derived from, this report.

© 2023, WONG WAI KIT. All right reserved.

ACKNOWLEDGEMENTS

The completion of this research marks an end to this knowledge-seeking venture. Allow me to express my highest gratitude to my research supervisor, Ir. Prof. Dr. Huang Yuk Feng, without whom I would have been “lost in woods” while embarking on this exploring journey. His profound confidence in my abilities is deeply appreciated. The completion of this research would not have been accomplished without his unwavering support and guidance.

Also, I could not express enough appreciation to Dr. Chong Kai Lun for his unparalleled advice and guidance. I am in abysmal debt for his insightful suggestion and comprehensive explanation, as well as his extensive technical support in Python modelling. Many thanks for the cherished memories along the endeavour.

ABSTRACT

Accurate streamflow forecasting is imperative for efficient water resources management. Yet, precise prediction of streamflow is difficult owing to the underlying complex relationships between variables, as well as the accompanying nonstationary and nonlinearity of the problem. Therefore, application of the hybrid ANN model has been given much consideration for its capabilities in delivering high predictive accuracy in hydrological forecasting. This study aims to assess the performance of short, medium and long term streamflow prediction by hybrid ANN model, after conducting thorough investigation on wavelet decomposition and developing the most accurate model. Historical streamflow data obtained from the Department of Irrigation and Drainage was separated into training and testing data set by a 80%:20% ratio, respectively. Data pre-processing was conducted by stationary wavelet transformation. In total, 12 different models with various combinations of 2 scenarios, 2 cases and 3 wavelets, including sym5, db5 and coif5 were used to predict streamflow. The improvement of hyperparameter tuning and hybrid model were verified, where tuning improved the models by a range of 0.93% to 68.17%, whereas hybrid models received 1082.78% to 1612.64% improvement as compared to the standalone model. By majority, S1 models have better performance than S2 models, while C1 models have better performance than C2 models. No wavelet was observed to exhibit apparent advantage. The best performing model was identified by the visualizations through the Taylor diagram and the Violin diagram and it was the S1*C1sym5 model that stood out. Prediction with this model had revealed that short term prediction is the most accurate, followed by the medium and long term predictions. By comparison then, the loss of accuracy in terms of (R^2 , RMSE, MAE) for the medium and long term respectively, are (95.18%, 1923.37%, 2070.52%) and (95.56%, 4811.91%, 3920.38%). Applications of different ANN learning algorithms such as grid search and random search; implementation of quantitative analysis to assess the similarity between wavelet and input time series; comparison of DWT and SWT; as well as combination of decomposed wavelets from different wavelet families; are but some of the recommended for future similar works.

TABLE OF CONTENTS

DECLARATION		i
APPROVAL FOR SUBMISSION		ii
ACKNOWLEDGEMENTS		iv
ABSTRACT		v
TABLE OF CONTENTS		vi
LIST OF TABLES		viii
LIST OF FIGURES		ix
LIST OF SYMBOLS / ABBREVIATIONS		xi
CHAPTER		
1	INTRODUCTION	1
1.1	General Introduction	1
1.2	Importance of the Study	3
1.3	Problem Statement	3
1.4	Aim and Objectives	4
1.5	Scope and Limitation of the Study	4
1.6	Contribution of the Study	5
1.7	Outline of the Report	5
2	LITERATURE REVIEW	7
2.1	Introduction	7
2.2	Neural Network	7
2.2.1	Artificial Neural Network (ANN)	7
2.2.2	Support Vector Machine (SVM)	13
2.2.3	Adaptive Neuro-Fuzzy Inference System (ANFIS)	17
2.3	Wavelet Transformation	20
2.3.1	Continuous Wavelet Transform	23
2.3.2	Discrete Wavelet Transform	24
2.3.3	Stationary Wavelet Transform	26

	2.3.4 Application of Wavelet Transform	28
	2.4 Coupled Signal Decomposition and Neural Networks	30
	2.5 Summary	34
3	METHODOLOGY AND WORK PLAN	35
	3.1 Introduction	35
	3.2 Study Area and Data Acquisition	36
	3.3 Stationary Wavelet Transform	38
	3.4 Development of Artificial Neural Network (ANN)	39
	3.5 Statistical Test	40
4	RESULTS AND DISCUSSION	42
	4.1 Introduction	42
	4.2 Hyperparameter Tuning Analysis	42
	4.3 Wavelet Analysis	48
	4.3.1 Comparison of Hybrid ANN and Standalone ANN	48
	4.3.2 Analysis of 2 Different Scenarios, 2 Different Cases (Univariate C1 or Multivariate C2) and 3 Different Wavelets	50
	4.3.3 Comparison of Family Wavelet	65
	4.4 Uncertainty Assessment	70
	4.5 Performance of Short, Medium and Long Term Forecasting	73
5	CONCLUSIONS AND RECOMMENDATIONS	75
	5.1 Conclusions	75
	5.2 Recommendations for future work	76
	REFERENCES	77

LIST OF TABLES

Table 3.1:	The List of Hydrological Observation Stations	38
Table 3.2:	List of Wavelets to be Tested	38
Table 4.1:	Performance of Untuned and Tuned Models	44
Table 4.2:	Hyperparameters of All Models	44
Table 4.3:	Performance of Standalone and Hybrid ANN models in both Scenarios	49
Table 4.4:	Performance of S1C1 and S2C1 cases with three different wavelets	51
Table 4.5:	Performance of S1C2 and S2C2, with three different wavelets	58
Table 4.6:	Comparison of C1S1 and C2S1, with three different wavelets	61
Table 4.7:	Comparison of C1S2 and C2S2, with three different wavelets	61
Table 4.8:	Type of Wavelet Families Incorporated (Patro and Kumar, 2016; Mandala et al., 2023; Santos et al., 2014)	66
Table 4.9:	Performance of Different Wavelets in C1 and C2 of S1	67
Table 4.10:	Performance of Different Wavelets in C1 and C2 of S2	67
Table 4.11:	Summary of Wavelet Properties (Upadhya et al., 2022)	69
Table 4.12:	Performance of S1*C1sym5 in Short, Medium and Long Term Forecasting	73

LIST OF FIGURES

Figure 2.1:	Typical Structure of ANN (Fathian et al., 2019)	8
Figure 2.2:	Typical Structure of BPNN (Zhang, Ma and Zhang, 2018)	9
Figure 2.3:	Basic SVM Classification (Olyaie, Abyaneh and Mehr, 2017)	13
Figure 2.4:	Network of Typical SVM (Liu et al., 2022)	15
Figure 2.5:	Difference of Boolean and Fuzzy Logic (Kaur and Mahajan, 2015)	18
Figure 2.6:	General Architecture of ANFIS (Al-qaness et al., 2022)	19
Figure 2.7:	Comparison of Fourier and Wavelet Transform (Nourani, Komasi and Mano, 2009)	21
Figure 2.8:	DWT by High Pass and Loss Pass Filter	25
Figure 3.1:	Flowchart of Streamflow Forecasting	36
Figure 3.2:	Location of Johor River Basin	37
Figure 3.3:	Network Architecture of ANN	40
Figure 4.1:	Performance of Untuned and Tuned Models in S1C1	46
Figure 4.2:	Performance of Untuned and Tuned Models in S1C2	46
Figure 4.3:	Performance of Untuned and Tuned Models in S2C1	47
Figure 4.4:	Performance of Untuned and Tuned Models in S2C2	47
Figure 4.5:	Performance of Standalone and Hybrid Models in S1	50
Figure 4.6:	Performance of Standalone and Hybrid Models in S2	50
Figure 4.7:	Decomposed Detail Series 1 of S1*C1sym5-5	52
Figure 4.8:	Decomposed Detail Series 2 of S1*C1sym5-5	52
Figure 4.9:	Decomposed Detail Series 3 of S1*C1sym5-5	53
Figure 4.10:	Decomposed Detail Series 4 of S1*C1sym5-5	53
Figure 4.11:	Decomposed Detail Series 5 of S1*C1sym5-5	54

Figure 4.12:	Decomposed Approximate Series of $S1*C1sym5-5$	54
Figure 4.13:	Partial Reconstruction by Adding A and D5 of $S1*C1sym5-5$	55
Figure 4.14:	Performance of combined $S1*C1sym5-5$	55
Figure 4.15:	Performance of S1C1 and S2C1, with sym5 Wavelet	56
Figure 4.16:	Performance of S1C1 and S2 C1, with db5 Wavelet	57
Figure 4.17:	Performance of S1C1 and S2C1, coif5 Wavelet	57
Figure 4.18:	Performance of S1C2 and S2C2, with sym5 Wavelet	59
Figure 4.19:	Performance of S1C2 and S2C2, with db5 Wavelet	59
Figure 4.20:	Performance of S1C2 and S2C2, coif5 Wavelet	60
Figure 4.21:	Performance of C1S1 and C2S1, with sym5 Wavelet	63
Figure 4.22:	Performance of C1S1 and C2S1, with db5 Wavelet	63
Figure 4.23:	Performance of C1S1 and C2S1, with coif5 Wavelet	64
Figure 4.24:	Performance of C1S2 and C2S2, with sym5 Wavelet	64
Figure 4.25:	Performance of C1S2 and C2S2, with db5 Wavelet	65
Figure 4.26:	Performance of C1S2 and C2S2, with coif5 Wavelet	65
Figure 4.27:	Input Streamflow Time Series	70
Figure 4.28:	Taylor Diagram of the Observed and 12 Studied Models	71
Figure 4.29:	Violin Diagram of the Observed and 12 Studied Models	72
Figure 4.30:	Performance of $S1*C1sym5$ in Different Horizons	74

LIST OF SYMBOLS / ABBREVIATIONS

a	scaling parameter of wavelet transform
b	shifting parameter of wavelet transform
$coif5$	Coiflet 5
$db5$	Daubechies 5
$\varphi(t)$	wavelet function
R^2	Coefficient of determination
$sym5$	Symlets 5
T	wavelet coefficient
$x(t)$	input time series
AI	artificial intelligence
ANFIS	adaptive neuro-fuzzy inference system
ANN	artificial neural network
ARIMA	autoregressive integrated moving average
BPNN	back propagation neural network
CDFt	cumulative distribution function transform
CEEMDAN	complete ensemble empirical mode decomposition with adaptive noise
CWT	continuous wavelet transform
DF	direct framework
DWT	discrete wavelet transform
ECG	electrocardiogram
FT	Fourier transform
LGP	linear programming
LMNN	Levenberg Marquardt neural network
LOD	level of decomposition
LSM	least squared method
MED	minimum entropy deconvolution
MF	multicomponent framework
MLR	multiple linear regression
MAE	mean absolute error
MSE	mean squared error

PSO	particle swarm optimization
QM	quantile mapping
ReLU	rectified linear unit
RMSE	root mean squared error
SSA	salp swarm algorithm
SVM	support vector machine
SWT	stationary wavelet transform
WNN	wavelet neural network
WT	wavelet transform

CHAPTER 1

INTRODUCTION

1.1 General Introduction

Accurate hydrological forecasting is challenging, not only because of the complicated underlying physical theory of the problem but also owing to the presence of nonlinearity, nonstationary, and uncertainty in the system (Feng et al., 2022). Regardless of the difficulty, precise hydrological prediction is of utmost significance to provide reliable information as the prerequisite data of water resources management. Amongst the hydrological forecasting problems, streamflow forecasting is frequently investigated to explore possible modifications for higher predictive accuracy. Indeed, the topic of streamflow forecasting is worth studying, considering its contribution to water resources planning. Depending on the forecast lead time, streamflow forecasting could be classified as short-term or long-term forecasting with different significance. Short-term forecasting facilitates decision making in domestic water reticulation systems and irrigation management, whereas long-term forecasting provides insight for effective agricultural strategies and reservoir operation, as well as floods and droughts management (Kambalimath and Deka, 2021; Li, Wang and Qiu, 2019; Poul, Shourian and Ebrahimi, 2019). Similar to other typical hydrological forecasting problems, streamflow forecasting is subjected to nonlinearity and nonstationary results by factors including but not limited to upstream flow conditions, climatic factors, and riverbed properties (Kambalimath and Deka, 2021).

The predictive models of hydrological forecasting have developed progressively from physical models to statistical models and eventually to data-driven models. The conventional physical model is developed by simulating the complex hydrological process with a set of related equations (Li, Wang, and Qiu, 2019). The application of a physical model requires an understanding of the underlying physical relationships of the parameters, which are often sophisticated, as well as a large data set for calibration. Despite the high computation difficulty, the predictive accuracy of the physical model is not guaranteed. Due to the complexity and scarcity of streamflow, low predictive

accuracy is typically found for streamflow forecasting (Feng et al., 2022). Some notable statistical methods, such as multiple linear regression and autoregression, are created to increase the prediction accuracy of streamflow forecasting issues, albeit the massive improvement observed compared to the physical model, these statistical models are incapable of handling nonlinear and nonstationary problems, which is almost certainly the case with streamflow forecasting (Adamowski and Sun, 2010). To date, data-driven models have become the prominent predictive models in hydrological forecasting. The capacity of data-driven models to detect the hidden link between parameters without having to grasp the complicated underlying relationship contributes to their efficacy in streamflow forecasting. Thus, data-driven models are capable of delivering highly accurate prediction results with minimum computation effort and data requirements, even in nonlinear problems (Zhou, Liu, and Duan, 2020).

Given its simplicity and efficacy, artificial intelligence in hydrological forecasting is fast gaining appeal. Amongst the variety of artificial intelligence approaches, the artificial neural network (ANN) and the support vector machine (SVM) are inarguably of the highest commonality and effectiveness. Despite being a member of the black box family, which is implicit and difficult to interpret, the accuracy of the artificial intelligence approach has no comparable alternatives (Poul, Shourian, and Ebrahimi, 2019). The ANN algorithm is sufficiently robust to identify the hidden association of parameters that resemble the function of the human brain. The SVM, employs a kernel function to avoid the computation of complex relationships between data and identify the mapping pattern between parameters. However, the effectiveness of these data-driven methods in forecasting nonlinear problems, limited predictive performances were observed in the presence of nonstationary time series (Leal, Costa, and Campos, 2019; Khan, Muhammad, and El-Shafie, 2020). In order to improve the performance of these models in nonstationary problems, the concept of data pre-processing is introduced to retrieve crucial information from the complex time series. Due to the ability to retrieve temporal and frequency information at various resolution levels, wavelet transformation is used as a signal decomposition. In fact, incorporating wavelet decomposition with neural networks has demonstrated significant improvement in many types of research

(Tayyab et al., 2017; Zhang, Zhang, and Singh, 2018; Xu, Chen, and Zhang, 2021).

1.2 Importance of the Study

Efficient water resources management is vital in both the economic and environmental aspects. Proper flood and drought management may help lessen the economic and environmental damage caused by natural catastrophes. However, proper planning of water reticulation and agricultural strategies could increase the cost-effectiveness of water distribution. These judgments would be impossible to make without the assistance of a sophisticated hydrological forecasting model. Therefore, developing an accurate yet user-friendly predictive model is of utmost importance to provide reliable analysis of the effect of water resources strategies. This study investigates the effect of several influential factors on the model accuracy to provide insight into developing an accurate forecasting model. Specifically, this study compares the effect of different mother wavelets, different levels of decomposition, different combinations of decomposed wavelets, as well as univariate and multivariate forecasting models, which are essential in developing the most accurate forecasting model.

1.3 Problem Statement

Integrated artificial intelligence (AI) models and signal decomposition methods are popular approaches in hydrological forecasting. The discrete wavelet transform (DWT) and its derivative, the stationary wavelet transform (SWT), are popular signal decomposition methods due to their ability to identify temporal and frequency information at various resolution levels (Chong et al., 2022; Chong, Lai and El-Shafie, 2019). Despite much research on univariate and multivariate forecasting model, a side-by-side comparison of both model has yet to be given. In addition, although extensive studies were conducted to investigate the performance of wavelet transformation, conclusive research on the selection of the mother wavelet and the level of decomposition for the best model performance is still lacking. Furthermore, the efficiency of wavelet transform in handling different forecasting lead times is also yet to be explored. Thus, it is desirable to compare the performance of wavelet transform in

different forecasting horizons. Nevertheless, even after multiple research have been conducted to exploit the problem, no unified outcome has been attained, as yet.

1.4 Aim and Objectives

The study aims to compare the performance of several wavelet transformation algorithms on short, medium, and long-term streamflow forecasting using wavelet-based artificial neural networks. Three specific objectives are outlined to be accomplished during the investigation, and they are:

- i. To conduct a thorough investigation of the performance of wavelet decomposition.
- ii. To develop the most accurate AI model with the optimized set of hyperparameters for the application of this study.
- iii. To assess the performance of the wavelet-incorporated model in different forecasting horizons.

1.5 Scope and Limitation of the Study

The scope of this study is focused on providing a thorough investigation of the performance of wavelet decomposition. This study considers the influence of different mother wavelets, different levels of decomposition and various combinations of decomposed wavelets. The comparison of univariate and multivariate models is also included in this study to investigate their superiority. While the best architecture suitable for all forecasting problems does not exist, the best-performing architectures of the ANN in this study are identified. The optimum architecture for the ANN is acquired by modifying the number of neurons and the hidden layer with the application of hyperparameter tuning. This study also includes the assessment of wavelet decomposition in different forecasting lead times. Moreover, statistical tests are to be performed to evaluate the performance of the models.

Nonetheless, certain limitations to this study should be considered in research. First of all, amongst the variety of AI models available, only ANN is considered. Although adding other AI models is possible, only the ANN was chosen to restrict the study's variables and focus on the changes in wavelet transformation. Only the SWT is applied in the wavelet transformation because

computing continuous wavelet transform (CWT) is redundant and time-consuming (Zhou et al., 2020). The comparison between the DWT and SWT has been excluded from this study and could be considered in future works. Lastly, it is certain that variation in the proportion of training to the testing data set will affect the model performance (Loh et al., 2021). Again, this impact will not be examined in this study, to limit the number of variable parameters that contribute to the model's performance.

1.6 Contribution of the Study

The study provides informative insights into the performance of wavelet transformation under the influences of multiple factors, prior to presenting convincing evidence about the improvements to the AI models by incorporating wavelet transformation as a data pre-processing technique. The knowledge serve as the prerequisites for suggesting the applicability of AI models for streamflow forecasting, as well as establishing a starting point for future comparison or extension work.

1.7 Outline of the Report

Chapter 1: Introduction

This chapter introduces the background to the application of AI models for streamflow forecasting. The significance of this study is explained through the narrative of the importance of study, aim and objectives, scope and limitation, as well as the contributions of this study.

Chapter 2: Literature Review

Ample credited recent publications were thoroughly reviewed in this chapter to provide the comprehensive and inspirational knowledge on the topics of AI modelling and wavelet transformation.

Chapter 3: Methodology and Work Plan

This chapter describes the procedure and methodologies adopted in this study. Justification of the methodologies is provided by referring to the findings of literature review. The meticulous plan for the progress flow of the study is illustrated in a flowchart.

Chapter 4: Results and Discussion

In this chapter, the results of the analyses are presented in an organized form to facilitate understanding to the readers. Necessary explanation and comparison were conducted to investigate the performance of the chosen AI models.

Chapter 5: Conclusions and Recommendations

This chapter summarises the gist of the study undertaken and concludes the findings of the study. Succinct yet precise conclusion, is produced for efficient review of all discussions. The recommendations for future extension work is also presented in this chapter.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Detailed literature reviews of the relevant topics are presented in this chapter. The main works include the Artificial Neural Networks (ANNs), the wavelet transformation (WT), and the coupling of ANN and WT. The fundamental concept and theories of the works are revealed. Recent related research is extensively examined, as are the noted gaps in the study. This literature review also validates the approach used in this investigation.

2.2 Neural Network

2.2.1 Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) is a subset of Artificial Intelligence (AI) developed to resemble the problem-solving and decision-making abilities of the human brain. By the implementation of ANN, the recognition of hidden patterns between the input and output data without the knowledge of the underlying complex relationship is made possible. Indeed, the application of ANN in forecasting has become prevalent as it aborts the necessity of understanding the relationship among endogenous variables. A typical ANN is made up of five major components: inputs, weights, a threshold or bias, an activation function, and an output. The illustration in Figure 2.1 effectively explains the notion of the ANN.

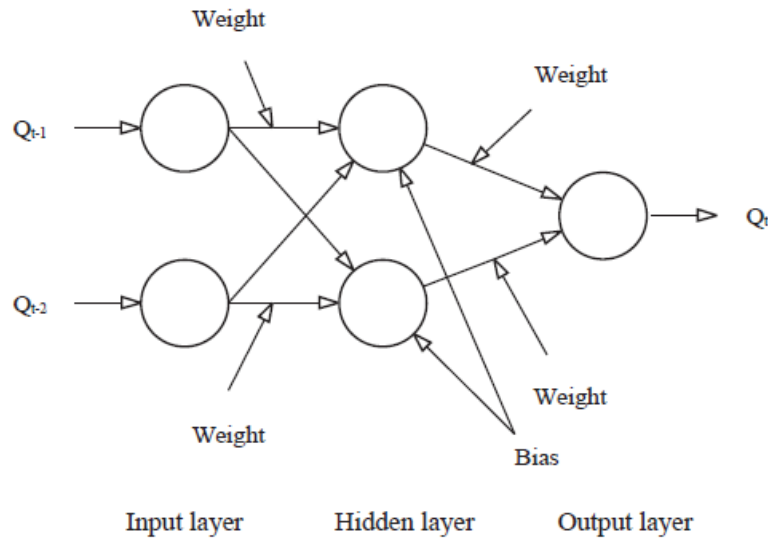


Figure 2.1: Typical Structure of ANN (Fathian et al., 2019)

The basic units of ANN are pronounced neurons and linked by connections known as synapses. These neurons and synapses form a layer of input, single or multiple hidden layers and an output layer. Each neuron is assigned with a specific weight and threshold, whereby if the weighted output of the neuron exceeds its threshold, the neuron is activated. Upon the activation of the neuron, the transmission of data to the next layer will take place. In contrary, no data will be passed to the next layer if the threshold is not achieved. Most of the neural networks leverage a sigmoid activation function to provide continuous activation value instead of discrete binary value (IBM, 2020). A typical ANN could be mathematically expressed as:

$$y = f\left(\sum_{i=1}^n w_i x_i + b_i\right) \quad (2.1)$$

where y indicates the output, f represents the activation function, while x_i , w_i and b_i denote the input, weight and threshold of the i th neuron respectively.

The forward propagation of the data from input to output is defined as feed-forward. Continuous adjustment of weight and threshold of each neuron is necessary to minimize the error of the output. The training of network can be accomplished by different algorithms such as back propagation, conjugate gradient and cascade correlation (Kişi, 2007). By the iterative training of the

algorithm, the parameters of the network ultimately converge at the local minimum, whereby the highest accuracy of the network is recorded. Apart from the parameters, the architecture of the network is also of major influence on the accuracy of the model (Zhang et al, 2018).

The Back Propagation Neural Network (BPNN) is essentially a combination of Artificial Neural Network (ANN) and Back Propagation (BP) training algorithm. For this reason, the structure of a BPNN shares a high similarity with the structure of a basic ANN, with the error propagation being the only major difference. Apart from the aforementioned five components of ANN, a back propagation of error occurs in BPNN during the training of the model. A typical structure of BPNN is illustrated in Figure 2.2.

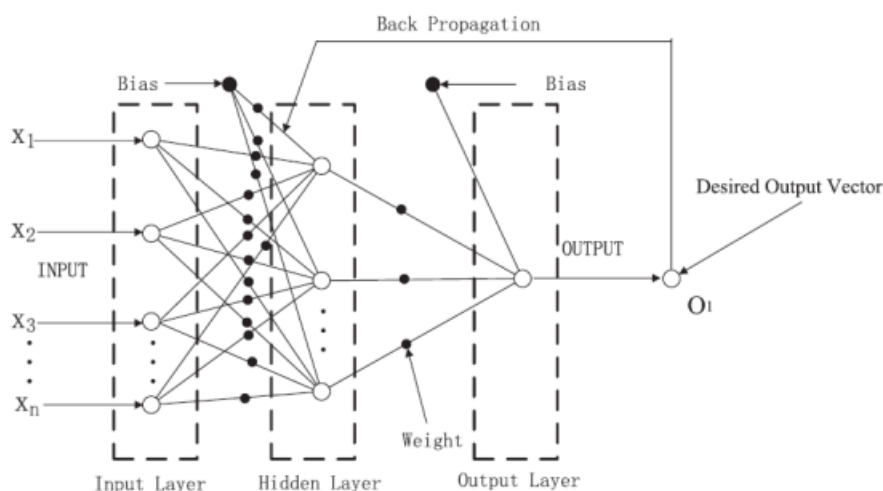


Figure 2.2: Typical Structure of BPNN (Zhang, Ma and Zhang, 2018)

Training of ANN is mandatory to improve the accuracy of the model before implementation. The deviation between the forecasted value and the reference value could be minimized by adjusting the weight and threshold of every neuron (Zhao et al., 2021). For BPNN algorithm, the training is performed repetitively in two stages, namely the forward propagation of input and back propagation of error. The input data are initially transmitted from the input layer to the hidden layers for processing and eventually to the output layer. The accuracy of the output is then assessed. In case of unsatisfactory accuracy, the error will be back propagated to the hidden and output layers to modify their weight and threshold. This modification is carried out in the steepest descent

direction, which contributes to the highest reduction of the performance function (Kişi, 2007). The training of model is completed when the error of the model falls in the desired range, or when the predefined iteration number is reached. The assessment of error could be performed by utilizing the definition below:

$$E = \frac{1}{P} \left(\sum_{p=1}^P E_p \right) \quad (2.2)$$

where E represents the global error, P is the total iteration of training and E_p denotes the error for the p th training, which is given as:

$$E_p = \frac{1}{2} \sum_{k=1}^N (O_k - D_k)^2 \quad (2.3)$$

where N depicts the number of output node, O_k and D_k represent the obtained output and the reference output of k th output node, respectively (Cobaner et al., 2010). Obviously, due to the cascading effect of the signal propagation, adjustment of weight and threshold of the neurons may significantly improve the predictive accuracy of the model (Chae et al., 2016).

The accuracy of a developed network to a problem is dependent on the architecture of the network, including the number of hidden layers and neurons. Most researchers utilized empirical or trial and error method to reach the optimum architecture of the network (Sharma, Singh and Sharma, 2021; Falamarzi et al., 2014; Adamowski and Sun, 2010). Higher number of neurons do not guarantee a high accuracy of the network, instead, fluctuation of accuracy is observed after the number of hidden nodes exceeds a certain limit (Yan, Li and Gao, 2014). Sharma, Singh and Sharma (2021) had investigated the accuracy of ANN with different architecture. Four ANN models with different network architecture had been developed by the authors to predict the river flow of Sot river catchment in India. The study had demonstrated the difference between performance of models with different architecture. Significant improvement of the model was observed, as the model architecture was

modified to reckon the input of lag-1 river flow data. Although the authors had concluded the best performing architecture in their study, however, the best architecture of ANN is case-specific, hence the conclusion might not be applicable for other hydrological forecasting problems. Therefore, the necessity of identifying the best architecture of ANN in every forecasting problem through empirical or trial and error methods remains existing. Another influential factor of the model performance, which is the training algorithm incorporated in the ANN had been examined by Kişi (2007). In the research, the comparison of four different ANN learning algorithms namely cascade correlation, conjugate gradient, back propagation and Levenberg-Marquardt was carried out. Statistical interpretation of the results had revealed that although every algorithm has different performance ranking in different forecasting application, the Levenberg-Marquardt had appeared to be the best overall model.

The performance of ANN in forecasting problems is often compared with other conventional statistical or data driven models, including but not limited to Least Squared based Method (LSM), Support Vector Machine (SVM) and ARIMA (Zhang, Zhang and Singh, 2018; Fashae et al., 2018; Zhang, Ma and Zhang, 2018). The performance comparison of ANN and SVM will be discussed in the later section, after the introduction to the algorithm of SVM. To date, numerous comparisons of ANN and ARIMA had been conducted to investigate the suitability of ANN as the superior alternative of ARIMA in forecasting problem involving nonlinear and nonstationary data (Toğa, Atalay and Toksari, 2021; Gui, Wu and Zhang, 2021; Nury, Hasan and Alam, 2017). In the research of Fashae et al. (2018), ARIMA was proven to have higher predictive accuracy than ANN, whereby the advantage is especially obvious when the input data are limited. Similar conclusions were drawn by Toğa, Atalay and Toksari (2021) and Gui, Wu and Zhang (2021), after comparing the performance of ARIMA and ANN in prediction. Despite the possible lower accuracy, ANN had displayed better performance and potential over ARIMA in problem with nonstationary time series data. With appropriate structure of ANN and increasing complexity of input data, there is a possibility for ANN to outperform ARIMA, as witnessed in the study of Zhang, Zhang and Singh (2018). The effectiveness of ANN in solving nonlinear and nonstationary

problem is also verified by Zhang, Ma and Zhang (2018), after recording an outstanding performance of ANN over LSM in predicting the nonlinear sea ice melting problem.

Several modifications of ANN had been proposed to enhance its performance. Falamarzi et al. (2014) had developed a wavelet neural network (WNN), which is essentially a neural network with wavelet activation function instead of sigmoid activation function as in conventional ANN. The developed WNN was used to predict evapotranspiration and was found to have higher accuracy than regular ANN. Researchers also noticed the improvement of model accuracy by incorporating wavelet decomposition to the ANN model (Falamarzi et al., 2014; Zhang, Zhang and Singh, 2018). The contribution of signal decomposition to the predictive accuracy shall be discussed in the later section. Another notable modification of ANN is the development of hybrid ANN models. ANN is often coupled with other algorithms to exploit their respective strengths, and their results are often encouraging (Fathian et al., 2019; Khan, Muhammad and El-Shafie, 2020; Babu and Reddy, 2014). In the work of Babu and Reddy (2014), the researchers developed a hybrid model of ARIMA and ANN to predict time series data. This combination is reasonable considering the ample recorded literatures which suggest higher accuracy of ARIMA over ANN in linear problem, and the better capability of ANN to reckon nonlinear and nonstationary problem. Certainly, the hybrid model outperformed both standalone model of ARIMA and ANN in the study. The hybrid model of ARIMA and ANN was further reinforced by Khan, Muhammad and El-Shafie (2020) by adopting discrete wavelet transform (DWT) as the signal decomposition technique. As a result, the developed Wavelet-ARIMA-ANN model outperformed the single ANN and Wavelet-ANN, implying the positive contribution of both wavelet and ARIMA combination with ANN. The superiority of hybrid ANN over standalone models were also observed in the study of Fathian et al. (2019).

In a nutshell, the utilization of ANN has made the prediction of complex hydrological problems without understanding the underlying relationship possible. The accuracy of ANN model is dependent on its architecture and the adopted training algorithm. Well-developed ANN could achieve higher forecasting accuracy than conventional statistical forecasting algorithm. Thus,

sound selection of training algorithm and network building is crucial for the performance of ANN. However, there is no conclusive standard in deciding the best ANN architecture, leaving the task to be done through time consuming trial and error method. To overcome its weaknesses and exploit its strength of being capable to perform well in nonlinear problems, hybrid models were developed by combining ANN with other algorithm such as ARIMA and wavelet activation function. Despite the encouraging performance of hybrid models, the best combining algorithm with ANN is yet to be identified, owing to the complexity resulted from numerous combinations of network architecture, training algorithm, as well as incorporating algorithm and type of input data.

2.2.2 Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a machine learning algorithm developed from statistical learning theory. The implementation of SVM is widespread in hydrological field, owing to its robustness in solving classification and regression problems. The knowledge of support vector classification is prerequisite to understand the algorithm of SVM regression. The SVM classifies the input data by identifying the hyperplane that differentiates the data most precisely. The hyperplane is an established boundary of the separated data. The nearest data to the hyperplane, known as the support vectors, will determine the location and the orientation of the hyperplane. Out of all possible separating boundaries of the data, SVM will select the hyperplane with the highest margin, which is the distance from the support vector to the hyperplane. A basic linearly separable SVM model is demonstrated in Figure 2.3:

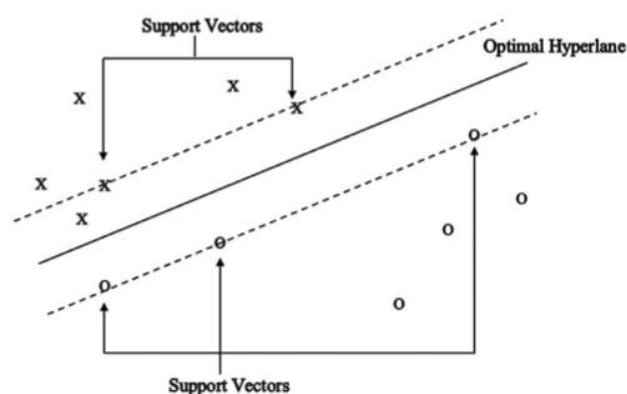


Figure 2.3: Basic SVM Classification (Olyaie, Abyaneh and Mehr, 2017)

In real hydrological cases, the data might not be separable by linear hyperplane. Thus, SVM employs a kernel function to map the input data from the original dimensional space to a higher dimensional space to obtain the hyperplane. The introduction of kernel function in SVM has simplified the identification of complex relationship between input and output data, by waiving the necessity of complicated mathematical transformation (Samantaray et al., 2022). Therefore, sound selection of kernel function is imperative for the high accuracy of SVM. Common mapping kernel functions include linear, polynomial and radial basis function (Raghavendra and Deka, 2014).

The capability of identifying complex input-output relationship has promoted the application of SVM in regression problem. Given a training set of data, SVM is capable of formulating the input-output relationship, which could then be utilized to forecast the output of a new input. The development of the SVM regression is subjected to the minimization of loss function, which is also known as the structural risk minimization principle (Yin et al., 2022). For a training data set denoted as:

$$\{x_i, y_i\}_{i=1}^n \quad (2.4)$$

where x_i is the i th input vector to be mapped onto the corresponding output y_i and n is the total number of data. The regression function inferred from the training data is:

$$g(x_i) = w_i \cdot \varphi(x_i) + b \quad (2.5)$$

where w_i represents the weight vector, φ is a nonlinear mapping function that maps the input vector to a higher dimensional space and b indicates a bias. There are several variations of loss functions of the SVM to be minimized. One of the most commonly proposed loss function is the ε -insensitive loss function outlined below:

$$L_\varepsilon(y_i, g(x_i)) = \begin{cases} 0, & |y_i - g(x_i)| \leq \varepsilon \\ |y_i - g(x_i)| - \varepsilon, & \text{otherwise} \end{cases} \quad (2.6)$$

where L_e is the loss and ε is the error tolerance (Olyaie, Abyaneh and Mehr, 2017). By solving the minimization problem of loss function subjected to multiple constraints, the regression function could eventually be evaluated as:

$$g(x_i) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (2.7)$$

where α_i and α_i^* are the Lagrange multiplier and $K(x, x_i)$ represents the kernel function, which is expressed as:

$$K(x_i, x_j) = \varphi(x_i) * \varphi(x_j) \quad (2.8)$$

The graphical illustration of a SVM network architecture is presented in Figure 2.4.

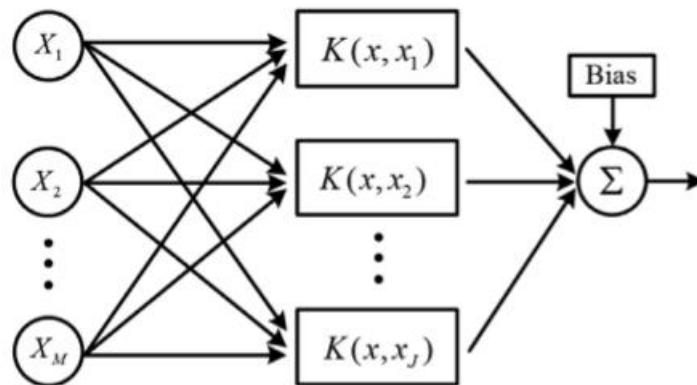


Figure 2.4: Network of Typical SVM (Liu et al., 2022)

The application of SVM in hydrological forecasting is always compared with other forecasting methods especially ANN. A predictive accuracy comparison of SVM, ANN and linear genetic programming (LGP) was performed in the study of Olyaie, Abyaneh and Mehr (2017). The authors estimated the dissolved oxygen (DO) concentration in Delaware River with the three aforementioned models. By analysing the results, the authors concluded that SVM has better predictive performance over ANN and LGP. The superior accuracy of SVM over ANN was also verified by other researchers (Liu et al.,

2022; Samantaray et al., 2022; Feng et al., 2022). Anyhow, the authors had emphasized on the importance of identifying more efficient optimization algorithm of SVM and ANN, as surging of predictive accuracy is possible with the presence of sound optimization algorithm. The variety of input data and size of training data set also contribute to the performance of the model. Indeed, with appropriate training algorithm and data pre-processing, ANN is capable of delivering higher accuracy over SVM, as witnessed in the study of Alquraish and Khadr (2021).

As an attempt to improve the performance of SVM, researchers had proposed hybrid SVM by combining conventional SVM and other algorithm. In the study of Yin et al. (2022), SVM was combined with quantile mapping (QM) and cumulative distribution function transform (CDFt) respectively to forecast precipitation. Although the conventional SVM had delivered satisfactory accuracy, both the combined models, namely SVM-QM and SVM-CDFt outperformed the conventional SVM by having higher accuracy and efficient computational time. In support of this, the SVM-SSA model developed by Samantaray et al. (2022) by combining SVM and Salp Swarm Algorithm (SSA) also demonstrated higher accuracy over the conventional SVM model. However, despite the improvements, some drawbacks of the hybrid model were also observed, which demanded further studies on the most appropriate incorporating algorithm.

The effect of input signal decomposition on the accuracy of SVM is investigated by numerous researchers (Feng et al., 2022; Kambalimath and Deka, 2021). Feng et al. (2022) employed the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) to decompose the nonlinear and nonstationary hydrological time series data before inputting the data to the SVM model. The authors concluded that CEEMDAN or other signal decomposition techniques are capable of reducing the impact of nonlinearity on runoff data, after observing the higher accuracy of decomposed model over the controlled model. Owing to the fact that only one signal decomposition technique is applied in this study, the effectiveness of CEEMDAN over other comparable methods remains debatable. Nevertheless, improvement of model accuracy is guaranteed with the presence of signal decomposition. Kambalimath and Deka (2021) had adopted a different signal decomposition technique, which

is the discrete wavelet transform (DWT). Four different mother wavelets including Haar, Daubechies, Coiflets and Symlets were utilized to decompose the input data. Significant improvement on the model accuracy was noticed by the authors after coupling the model with DWT. However, the optimum decomposition level of each wavelet, which is crucial for the performance of model is not considered by the authors in this study.

Summarizing the findings of the section and referring to the advantages and disadvantages of SVM outlined by Raghavendra. N and Deka (2014), the strengths and weaknesses of SVM are identified. SVM regression has high accuracy, often higher than ANN as suggested by researchers (Liu et al., 2022; Samantaray et al., 2022; Feng et al., 2022). The utilization of kernel function in SVM is twofold, whereby it allows the classification of nonlinearly separable data by mapping the data onto higher dimensional space and simplifies the complex input-output relationship. Since SVM could work with relatively smaller sample size, the problem of model overfitting could be avoided. The presence of outlier data will have minimal effect on SVM as the hyperplane is only dictated by the support vectors. Some other advantages of SVM include good generalization performance and free from local minimum. Nevertheless, the weaknesses of SVM are also noticed, mainly on the difficulty in selecting the most appropriate kernel function and hyperparameters. Prolonged training process of SVM due to the iterative process of kernel and hyperparameters selection, as well as the complexity induced from kernel mapping is another main drawback of the algorithm.

2.2.3 Adaptive Neuro-Fuzzy Inference System (ANFIS)

The Adaptive Neuro-Fuzzy Inference System (ANFIS) is a collaboration of ANN and Fuzzy Inference System (FIS) which adopts the principle of fuzzy logic. Before looking into the algorithm of ANFIS, the understanding of fuzzy logic is utmost important as the fundamental knowledge to the topic. Kaur and Mahajan (2015) provide a comprehensive introduction to the fuzzy logic by outlining its provision of intermediate values between absolute truth and absolute false. In traditional Boolean logic, true and false are defined with no transitional value to account for any intermediate. However, owing to the natural ambiguity and inaccuracy of the real world information, there are many

grey shades between the absolutes. Hence, the introduction of fuzzy logic is crucial to reckon the possible intermediate values, in pursuance of higher accuracy of the model. The difference of Boolean and fuzzy logic is depicted in Figure 2.5.

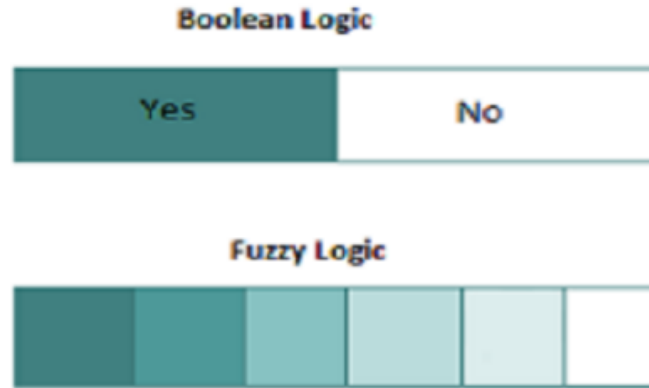


Figure 2.5: Difference of Boolean and Fuzzy Logic (Kaur and Mahajan, 2015)

The ANFIS employs the mapping ability of FIS to create a relationship between input and output. The model is then optimized by the self-learning capability of ANN (Ghenai et al., 2022). The FIS is capable of recognising hidden pattern between data and adaptively formulate a fuzzy rule base for prediction. Thus, the application of ANFIS in streamflow forecasting is advantageous as it neglects the underlying complicated technical issue. Most ANFIS adopt the Takagi-Sugeno Inference model, which relates the input and output with a series of IF-THEN rules (Al-qaness et al., 2022). These rules are constituted by the linear combination of the crisp numerical inputs. For a model with two inputs denoted as x and y , the rules are given as:

Rule 1: IF x is P_1 and y is Q_1 , then $f_1 = p_1x + q_1y + r_1$

Rule 2: IF x is P_2 and y is Q_2 , then $f_2 = p_2x + q_2y + r_2$

with P_i and Q_i being the fuzzy sets, f_i indicates the output of the fuzzy region, while p_i , q_i and r_i ($i = 1, 2$) denotes the model parameters subjected to improvement (Yaseen et al., 2017). The algorithm of ANFIS will be explained with the graphical illustration provided in Figure 2.6.

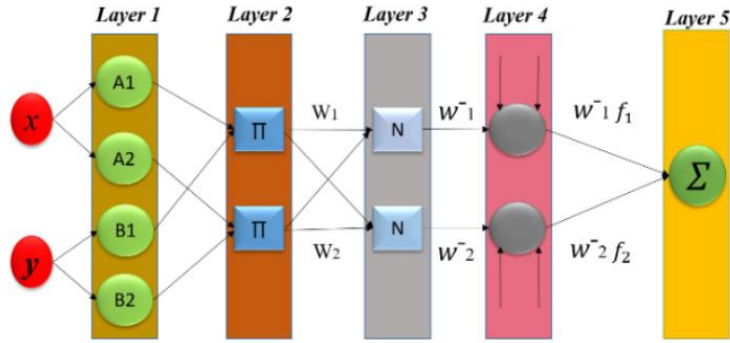


Figure 2.6: General Architecture of ANFIS (Al-qaness et al., 2022)

The Layer 1 of ANFIS is also known as the fuzzification layer. Conversion of crisp numerical input to linguistic output will occur within this layer. By adopting the membership function assigned to each neuron, the membership grades, which represent the degree of truth or falsehood of the inputs are determined. Examples of membership functions are continuous differentiable functions such as the Gaussian function, generalized bell-shaped function and the triangular or trapezoidal shaped function. However, the Gaussian function is a superior membership function over the others due to its smoothness and concise definition by two optimizable parameters (Gholami et al., 2017). Gaussian function based ANFIS is proven to have higher forecasting accuracy through numerous trial and error iterations (Zhou, Guo and Chang, 2019; Yaseen et al., 2017). The output of node i in the Layer 1, depicted as $O_{1,i}$ with a Gaussian membership function is given as:

$$O_{1,i} = \mu_{A_i}(x_1) \text{ for } i = 1, 2 \quad \text{or} \quad O_{1,i} = \mu_{B_i}(x_2) \text{ for } i = 1, 2 \quad (2.9)$$

$$\mu_{A_i}(x_1) = \exp\left(\frac{(x - c)^2}{2\sigma^2}\right) \quad (2.10)$$

where x_1 and x_2 are the crisp numerical input to the ANFIS, μ_{A_i} and μ_{B_i} represent the Gaussian membership function for fuzzy sets A_i and B_i respectively, c and σ are the conditional parameters of the function (Zhou, Guo and Chang, 2019).

The Layer 2 of the ANFIS determines the firing strength of the node. The firing strength is a quantitative measure of the degree of satisfactory of an input to the fuzzy rule. The inputs to the Layer 2 will undergo AND operation,

whereby their values are multiplied. Since the contribution of each input is multiplicative, this layer is also known as the product layer. The mathematical equation of the layer is presented below:

$$O_{2,i} = w_i = \mu_{A_i}(x_1) \times \mu_{B_i}(x_2) \text{ for } i = 1, 2 \quad (2.11)$$

where the output is also denoted as w_i , the weight function to the next layer.

Normalization of the firing strength will take place in Layer 3. The normalized weight of i th node, depicted as \bar{w}_i , is the proportion of the i th node weight in Layer 2 to the total node weight in Layer 2, which is mathematically equivalent to:

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2} \text{ for } i = 1, 2 \quad (2.12)$$

The Layer 4 of the ANFIS is the defuzzification layer with adaptive nodes. Parameters in an adaptive node is optimizable by training. The defuzzification of the input to Layer 4 is performed by a linear function with adaptive parameters, as given by:

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x_1 + q_i x_2 + r_i) \text{ for } i = 1, 2 \quad (2.13)$$

where p_i , q_i and r_i are consequent parameters subjected to improvement. The improvement of these parameters could be accomplished by the self-learning capability of the ANN (Ghenai et al., 2022).

Finally, all the incoming signals are summarized in Layer 5, giving one final output. The accuracy of the output will then be assessed and the model is trained. The summation of all the weighted signals is calculated by:

$$O_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \text{ for } i = 1, 2 \quad (2.14)$$

2.3 Wavelet Transformation

The utilization of wavelet transform (WT) as the superior alternative of Fourier transform in time series analysis has been given much consideration since the

past decade, for it could provide multi-resolution analysis in both time and frequency domain (Zhou, Liu and Duan, 2020; Sharma, 2016). Although the satisfactory application of Fourier analysis in stationary time series, it has no distinguishable ability in the time domain, since time information is lost during Fourier Transformation (Wang, Guo and Duan, 2013). For nonstationary time series, provision of varied time and frequency resolution is made possible by the introduction of wavelet analysis. For low frequency information, low time resolution (long time intervals) and high frequency resolution (small frequency intervals) are utilized to detect the gradual changes of frequency; whereas for high frequency information, high time resolution and low frequency resolution are utilized to record the abrupt frequency changes in rapid time frame (Adamowski and Sun, 2010). Nourani, Komasi and Mano (2009) depicted the comparison of Fourier transform and wavelet transform as Figure 2.7:

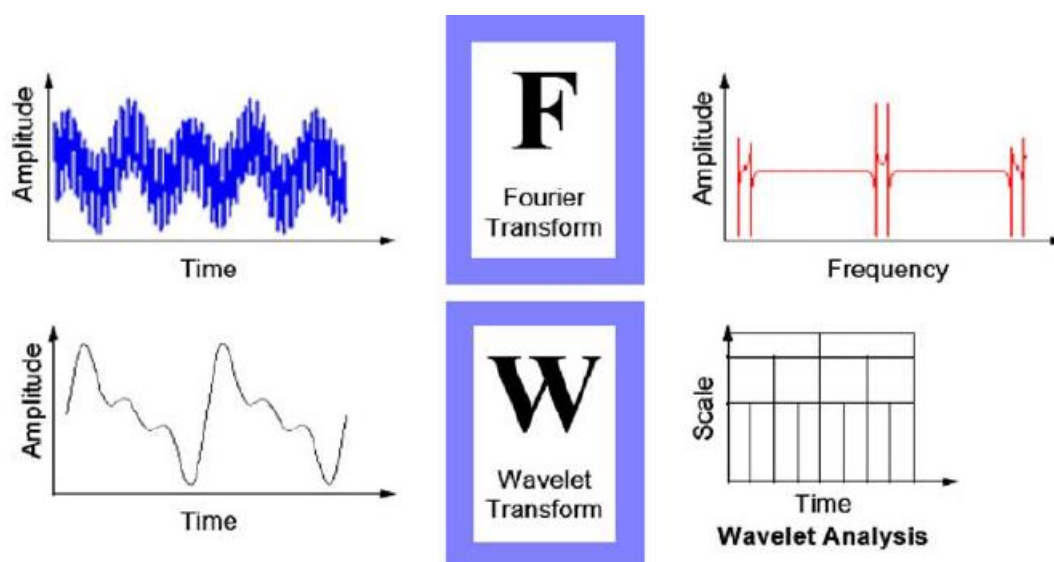


Figure 2.7: Comparison of Fourier and Wavelet Transform (Nourani, Komasi and Mano, 2009)

In Fourier transform, the basis function used to filter and decompose the time series is either trigonometric sine or cosine function, resulting in difficulty of reckoning time series with high difference of frequency (Falamarzi et al., 2014). Wavelet transform, allows the selection of the most suitable incorporating mother wavelet function by considering the characteristics of the time series to be investigated (Cheng et al., 2021; Falamarzi et al., 2014).

However, only functions with finite energy and fulfil the admissibility condition as follow could be selected as the mother wavelet (Sharma et al., 2016):

$$C_{\varphi} = 2\pi \int_{-\infty}^{+\infty} \frac{|\bar{\varphi}(\omega)|^2}{\omega} d\omega < \infty \quad (2.15)$$

where $\bar{\varphi}(\omega)$, the Fourier transforms of $\varphi(t)$ is calculated by the equation below:

$$\bar{\varphi}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \varphi(t) e^{-i\omega t} dt \quad (2.16)$$

The advantage of wavelet over Fourier transformation in analysing nonstationary time series is owing to its characteristics of compactly supported, zero mean and irregularly shaped. A compactly supported function with its finite length could effectively represent a localized data; the irregular shape of a wavelet function facilitates the detection of discontinuity or abrupt shape changes in the time series data (Chong, Lai and El-Shafie, 2019).

There are several mother wavelets available for different shape of time series, including but not limited to Morlet, Haar, Daubechies and Coiflet, with Morlet being the most frequently adopted in hydrological application, owing to its fair balance of temporal and frequency localization (Bilgili et al., 2021; Cheng et al., 2021; Chong et al., 2022). From these mother wavelets, a daughter wavelet could be derived to best fit the investigated time series and thus return the most accurate correlation between the wavelet and the time series. The basic derivation of daughter wavelet from its mother wavelet could be mathematically expressed as:

$$\varphi_{a,b}(n) = \frac{1}{\sqrt{a}} \varphi\left(\frac{n-b}{a}\right); \quad n, a, b \in R, a > 0 \quad (2.17)$$

whereby a is known as the scaling parameter which controls the dilation or shrinking of the wavelet; b is the shifting parameter which controls the position of the wavelet along the time series and R depicts the set of real number (Sharma

et al., 2016). The scaling and shifting parameters are responsible in determining the time and frequency resolution of the wavelet transformation. In case of a high scaling parameter, a , the daughter wavelet is stretched, thus having higher frequency resolution and lower time resolution. In contrary, for a low scaling parameter, the daughter wavelet is shrunken, delivering higher time resolution but lower frequency resolution. The shifting parameter is adjusted to translate the daughter wavelet across the time domain to retrieve localized data of the time series. Depending on the nature of these two parameters, there are three distinct transformations of mother wavelet to daughter wavelet, known as the continuous wavelet transform (CWT), discrete wavelet transform (DWT) and stationary wavelet transform (SWT).

2.3.1 Continuous Wavelet Transform

As suggested by the terminology, the continuous wavelet transform (CWT) utilizes the continuous real value for both the scaling and shifting parameters. Therefore, identification of global and localized features of the time series is feasible in CWT since detection of signals is performed in all scale (Zhou et al., 2020). The CWT of an input time series could be defined as the convolution of the wavelet function and the time series, which is mathematically expressed as:

$$T(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} \varphi^* \left(\frac{t-b}{a} \right) x(t) dt \quad (2.18)$$

where $T(a, b)$ depicts the wavelet coefficient corresponding to scaling factor a and shifting factor b ; $x(t)$ represents the input time series and φ^* indicates the complex conjugate of the wavelet function $\varphi(t)$. Since all frequency and time data of the time series are preserved in CWT, recovery of the series is feasible through the inverse wavelet transformation, given as:

$$x(t) = \frac{1}{c_\varphi} \int_{-\infty}^{+\infty} \int_0^\infty \frac{1}{\sqrt{a}} \varphi \left(\frac{t-b}{a} \right) T(a, b) \frac{da \cdot db}{a^2} \quad (2.19)$$

whereby c_φ is the function energy as calculated previously in Equation 2.15.

The major drawback of CWT is its data redundancy resulting from the continuous values of scaling and shifting parameters. The gigantic set of data necessitates long computation time and resources. Indeed, since practical time series are often sampled discretely, continuous values of scaling and shifting are usually not required to preserve the complete information of the original time series (Wang, Jin and Li, 2009). Therefore, to simplify the implementation of the wavelet transform, the discrete wavelet transform is introduced.

2.3.2 Discrete Wavelet Transform

In contrast to the continuous scaling and shifting parameter in CWT, the discrete wavelet transform (DWT) utilizes discrete parameters of the powers of 2, known as dyadic. Since only discrete dyadic scaling and shifting parameters are considered, the wavelet transformation from mother wavelet to daughter wavelet provided in Equation 2.17 could be rewritten as:

$$\varphi_{m,n}(t) = \frac{1}{\sqrt{2^m}} \varphi(2^{-m}t - n) \quad (2.20)$$

where m and n are integer parameters for scaling and shifting respectively. Subsequently, the discrete wavelet coefficient could be computed by the convolution of the discrete wavelet and time series as:

$$T_{m,n}(t) = 2^{-\frac{m}{2}} \sum_{t=0}^{N-1} \varphi(2^{-m}t - n)x_t \quad (2.21)$$

where $T_{m,n}$ is the discrete wavelet coefficient corresponding to scaling parameter $a = 2^m$ and shifting parameter $b = 2^m n$. This equation is applied to the finite time series, x_t from $t = 0$ until $t = N - 1$, where $N = 2^M$. Integer M governs the ranges of m and n , which are given as follows:

$$1 \leq m \leq M \quad (2.22)$$

$$0 \leq n \leq 2^{M-m} - 1 \quad (2.23)$$

In the process of DWT, the original time series will be iteratively decomposed by a high pass filter and a low pass filter at each decomposition level. The high pass filter produces detail coefficients while the low pass filter produces approximate coefficient, each having half the number of the input series. For every decomposition level, half of the coefficients, which are the detail coefficients will be retained, while the approximate coefficients will be passed down to the next level for recursive decomposition, until the desired level of decomposition. This process of halving coefficients in each level is known as decimation, which strengths and weaknesses are to be discussed later. Figure 2.8 presents a comprehensive graphic of DWT, illustrated by Pandey, Kesharwani and Singh (2015).

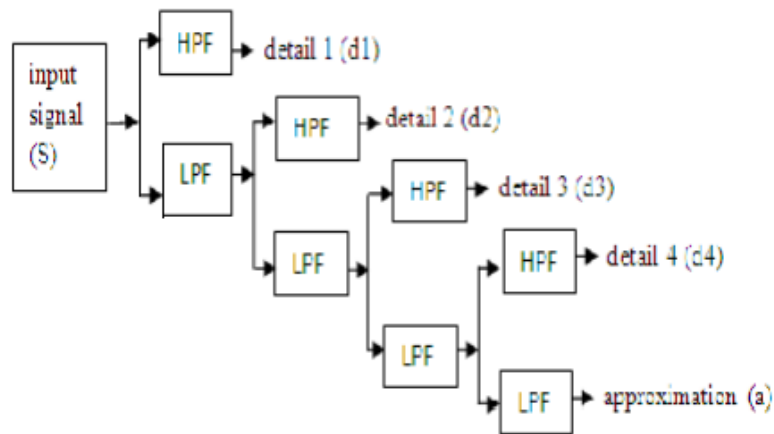


Figure 2.8: DWT by High Pass and Loss Pass Filter (Pandey, Kesharwani and Singh, 2015)

Similar to CWT, the reconstruction of the original time series is possible with DWT, by solving the inverse discrete transformation:

$$x_t = \bar{T}(t) + \sum_{m=1}^M W_m(t) \quad (2.24)$$

in which $\bar{T}(t)$ depicts the approximate coefficient at the last level M , and the latter expression represents the summation of all detail coefficients, $W_m(t)$,

from every level of decomposition, which could be calculated by (Nourani, Komasi and Mano, 2009):

$$W_m(t) = \sum_{n=0}^{2^{M-m}-1} T_{m,n} 2^{-\frac{m}{2}} \varphi(2^{-m}t - n) \quad (2.25)$$

The reduction of data generation in DWT simplifies its implementation as compared to CWT. Employment of dyadic discretization of parameters, together with decimation is helpful in reducing data redundancy. While this is advantageous for applications such as data and image compression, it could have adverse effect on the accuracy of prediction problem, as less training data is available to the forecasting model (Zhou, Liu and Duan, 2020). Maheswaran and Khosa (2012) also emphasized on the lack of shift invariant property in DWT, whereby the resulted incapability of reckoning newly added data will lower the forecasting ability of the model. The authors therefore concluded that the implementation of DWT is not suitable in problems involving singularity detection, nonparametric regression and forecasting. Quilty and Adamowski (2018), challenged the implementation of DWT in hydrological forecasting problems by considering the effect of boundary condition associated to wavelet decomposition. Three sources of boundary condition error were identified to be the utilization of future data, the inappropriate selection of decomposition level and wavelet filters, and also the careless separation of training and testing data. Ignorance of these boundary conditions often leads to optimistic forecasting accuracy of the model, which is in fact not practically achievable. Considering the drawback of DWT, the adoption of stationary wavelet transform is encouraged (Zhou, Liu and Duan, 2020; Quilty and Adamowski, 2018; Adamowski and Sun, 2010).

2.3.3 Stationary Wavelet Transform

The stationary wavelet transform (SWT), which is also known as the *à trous* (AT) wavelet transform, is the derivative of DWT to incorporate shift invariant property in the transformation. In contrast to DWT, no decimation is implemented in SWT, conserving the length of wavelet components at each

decomposition level. The resulting gaps from SWT are filled with redundant information, by the introduction of padding mode inside the filters (Laha et al., 2022). Despite the longer computation time and higher memory storage demanded by the redundant data, larger number of training data may facilitate the accurate calibration of the forecasting model.

The desirable shift invariant property is obtained by giving up the decimation of data. To validate the property, Maheswaran and Khosa (2020) had investigated the effect of adding new data into the training set of SWT. As a result, uniform and consistent transform was observed before and after the addition of new data, guaranteeing the forecasting ability of SWT as opposed to the shift variant DWT. The aforementioned boundary condition error of DWT resulted from the inclusion of future data is mitigated in SWT, since the computation of SWT requires only data of present time index (Zhou, Liu and Duan, 2020). For these reasons, it is rational to suggest that SWT is more applicable than DWT in forecasting problems.

The computation of approximate and detail coefficients of SWT could be accomplished by applying the following equation (Maheswaran and Khosa, 2012):

$$c_j(t) = \sum_{l=-\infty}^{\infty} h(l)c_j(t + 2^{j-1}l) \quad (2.26)$$

where c_j is the approximate coefficient; h represents a low pass filter storing the wavelet function. The selection of wavelet function dictates the coefficients of the filter. After computing the approximate coefficient, the detail coefficient d_j , could be identified by calculating the difference of two consecutive approximate coefficients, expressed as:

$$d_j(t) = c_{j-1}(t) - c_j(t) \quad (2.27)$$

It is also worth noticing that the input series is equal to the approximate coefficient at decomposition level 0, c_0 , therefore making the computation of d_1 possible. The reconstruction of the original series could be done by additive

reconstruction, since the length of the signals are conserved. The reconstruction equation of SWT is given as (Kumar et al., 2021):

$$c_0(t) = \sum_{j=1}^J c_j(t) + d_j(t) \quad (2.28)$$

2.3.4 Application of Wavelet Transform

The employment of wavelet transformation in forecasting model is prevalent to exploit its proven effectiveness in retaining time information. While the wavelet incorporation as signal decomposition method will be discussed in a designated section later, this section is focused in the application of wavelet transformation as the activation function of the forecasting model. Typically, conventional neural networks leverage a sigmoid function as the activation function of the neurons. Although the overall satisfactory performance, the sigmoid function may fail in recognizing abrupt shape changes of the input time series, thus resulting loss of useful information. However, in wavelet transformation, the most suitable incorporating mother wavelet could be selected from a variety of alternatives based on the input time series characteristics (Cheng et al., 2021).

Sharma et al. (2016) had developed a mixed wavelet neural network (WNN) by replacing the sigmoid activation function of conventional ANN with wavelet function. By taking into consideration the nonstationarity and rapid varying frequency of the input series, a combination of the Mexican hat and Morlet wavelets were chosen. This combination was claimed by the authors as effective in detecting and localizing data in multiscale scenario. Similar approach was adopted by Falamarzi et al. (2014) in predicting the evapotranspiration from temperature and wind speed data. The results of these two researches agreed well with each other, whereby the WNN had demonstrated inarguably superior performance than the conventional ANN.

The application of wavelet transformation (WT) in time series analysis has proven advantage over conventional signal processing method such as the Fourier Transform (FT). In the study of Chong, Lai and El-Shafie (2019), the authors compared the performance of wavelet and Fourier transform in analysing the river streamflow time series. Although both methods were capable

of detecting the trend of input series by identifying the periodic components, nonetheless, the time domain information of the input series was given up in the FT. Hence, the identification of occurrence time of a data is impossible after FT. Apart from the ability to preserve temporal information, the higher effectiveness of WT over FT was also verified by Leal, Costa and Campos (2019), who concluded that WT required shorter computation time to return higher performance than FT. The implementation of WT also allowed the further analysis of wavelet coherence. The wavelet coherence analysis was performed to investigate the relationship between time series data. The similarity in cyclic patterns of two input time series could be identified by wavelet coherence (Cheng et al., 2021). In fact, Chong et al. (2022) had conducted a wavelet coherence analysis to investigate the relationship between two time series, which were the standardized precipitation index (SPI) and the climatic indices. The wavelet coherence analysis was performed after the authors had identified the spatial and temporal patterns of droughts to reveal possible relationship between the time series, which FT is not capable of.

Another novel application of the wavelet transform is the signal denoising. Typically, the signal denoising by wavelet transformation is accomplished through three general steps, namely the transformation, filtering and reconstruction. The noise coefficient is first computed and filtered out by a threshold. The remaining coefficients which exceed the threshold are then combined to reconstruct the denoised series. Kumar et al. (2021) had applied SWT in denoising electrocardiogram (ECG). The presence of noise in the studied ECG was resulted by power line interference and baseline wandering. These noises were filtered out by the aforementioned procedures with SWT. Other denoising method such as empirical mode decomposition and Fourier decomposition as well as DWT were also applied for performance comparison. After statistically comparing the results, the authors concluded that the SWT based ECG outperformed the alternatives. Indeed, apart from the performance of the wavelet, the optimization of threshold also plays a crucial role in effectively denoising the signal. An optimization algorithm known as the Particle Swarm Optimization (PSO) was proposed by Laha et al. (2022). It was observed from the study that the implementation of PSO could appropriately adjust the threshold, thus delivering higher performance of signal denoising, as

compared to the conventional denoising method known as Minimum Entropy Deconvolution (MED).

2.4 Coupled Signal Decomposition and Neural Networks

Hydrological forecasting problems are often accompanied by nonlinear and nonstationary time series data, together with complex variables relationship resulting from their endogeneity. To alleviate the computation difficulty, ANN is utilized in identifying the hidden mapping patterns of the data without the requisite of understanding the underlying complex relationship. The application of ANN has proven effectiveness and potential in solving nonstationary forecasting problems (Zhang, Ma and Zhang, 2018). That said, despite the satisfactory performance of standalone ANN models, enhancement could be anticipated by the incorporation of signals decomposition methods. Wavelet transformation decomposes the input time series into sub-signals for multiresolution analysis, which is helpful in capturing useful information from the nonstationary time series at different resolution (Zhou, Liu and Duan, 2020). Therefore, it is rational to expect higher predictive accuracy from the combination of wavelet decomposition and neural network forecasting, especially in hydrological forecasting considering the nonlinearity and nonstationarity of the input time series. Indeed, the effectiveness of coupled signal decomposition and neural networks in hydrological forecasting had been investigated deliberately, with encouraging effect observed in most studies (Xu, Chen and Zhang, 2021; Li, Wang and Qiu, 2019; Poul, Shourian and Ebrahimi, 2019).

For verification of results and possible provision of supportive evidence, the literature review of the previous researches was performed in chronological order. The effectiveness of coupled wavelet decomposition and neural network was investigated by Adamowski and Sun (2010). The authors developed a coupled stationary wavelet transformation (SWT) and Levenberg Marquardt artificial neural network (LMNN) to forecast the short term streamflow of non-perennial river. The selection of LMNN was justified by the findings of Kişi (2007), who suggested that Levenberg Marquardt is the superior training algorithm for ANN. After comparing the performance of coupled model to standalone model, the authors concluded that signal decomposition had

significantly improved the performance of the ANN. This conclusion was then confirmed by many other researchers, guaranteeing the improvement of ANN accuracy by coupling of wavelet decomposition (Falamarzi et al., 2014; Poul, Shourian and Ebrahimi, 2019; Feng et al., 2022). Anyhow, the scope of this study was limited to the SWT and LMNN. Therefore, further investigation on the effect of different types of wavelet decomposition and different types of neural networks is demanded, in order to justify the reliability of the developed model.

Undoubtedly, the selection of incorporating neural network is of utmost importance in developing an effective coupled model. Considering its importance and complexity due to the numerous combinations with different wavelet transformation, the performance of different coupling neural networks is worth deliberate investigation (Zhang, Zhang and Singh, 2018; Poul, Shourian and Ebrahimi, 2019; Xu, Chen and Zhang, 2021). In the study of Poul, Shourian and Ebrahimi (2019), four different models including multi-linear regression (MLR), ANN, adaptive neuro-fuzzy inference system (ANFIS) and K-nearest neighbours (KNN) and their coupled models were compared. In terms of accuracy improvement, which was evaluated by Nash-Sutcliffe Efficiency (NSE) in this research, all the hybrid models exhibited huge improvement from their standalone models. The adoption of wavelet decomposition had significantly increased the NSE of MLR, ANN, ANFIS and KNN from 0.340, 0.404, 0.376 and 0.419 to 0.907, 0.930, 0.923 and 0.847 respectively. The upsurge of NSE had confirmed the effectiveness of wavelet decomposition in forecasting. Likewise, improvement of predictive accuracy was observed in other similar researches, assuring the positive effect of wavelet decomposition (Zhang, Zhang and Singh, 2018; Xu, Chen and Zhang, 2021). Amongst the models, ANN outperformed the others with the highest accuracy, consistent with the findings of Xu, Chen and Zhang (2021). Anyhow, the superiority of ANN is not deemed, owing to the uncertainties of input series, training algorithm, selection of activation functions and many other factors. Hence, the best model to be coupled with wavelet transformation remains debatable, as no obvious advantage of any model prevails to date.

Apart from the type of neural networks, the type of wavelet decomposition also contributed to the accuracy of the coupled model.

Researches were conducted to identify the applicability of CWT, DWT and SWT, as well as the performance of different mother wavelets (Maheswaran and Khosa, 2012; Tayyab et al., 2017; Kambalimath and Deka, 2021). Application of DWT is desirable over CWT to simplify the computation, by performing transformation with discretely sampled scaling and shifting coefficients. Tayyab et al. (2017) had developed a coupled ANN and DWT to forecast monthly streamflow. The authors observed improvement in the coupled models as compared to the standalone model and claimed that DWT could effectively enhance the forecasting ability of the model. However, the authors neglected the effect of different mother wavelets, which is highly influential to the performance of the model. In support of this, Kambalimath and Deka (2021) utilized different mother wavelets in DWT to investigate the corresponding performance. The authors successfully verified the improvement brought by DWT to the models, but failed to conclude the best incorporating mother wavelet, as no mother wavelets demonstrated consistent high accuracy for different variable predictions. This is reasonable as the suitability of mother wavelet is highly dependent on the characteristics of the input time series. In fact, Maheswaran and Khosa (2012) had reached a similar conclusion that no universal mother wavelet is suitable for all types of time series. Nevertheless, the authors had recommended the compact supported wavelets such as Haar wavelet for time series with short memory and transient features; and widely supported wavelet with higher vanishing moments like Daubechies 2 for time series with long memory and nonlinear features. These recommendations were made based on the observation of the study, which again, has limited generalized suitability. Another important factor ignored in previous researches is the level of decomposition. Although its importance is recognized by the authors, its effect and determining standard are yet to be investigated by the authors.

Despite the numerous researches conducted to investigate DWT and SWT separately, the study to compare these two transformations is still lacking. Being the modified redundant version of DWT, SWT is theoretically deemed to have better predictive accuracy, however, this statement is yet to be supported by sufficient research. In fact, Quilty and Adamowski (2018) had challenged the application of DWT in hydrological forecasting, on the ground that the induced

boundary condition errors were not resolved. Three sources of boundary condition errors in wavelet decomposition were outlined, which are the utilization of future data, the inappropriate selection of mother wavelets and decomposition level, as well as the improper partitioning of training and testing data. Most past application of wavelet transformation overlooked the effect of boundary condition errors and returned optimistic performance, which is not practical. To mitigate the error, the authors suggested a three steps best practice in wavelet decomposition. The best practice is begun with the selection of either maximal overlap discrete wavelet transform (MODWT) or SWT, followed by the selection of decomposition level and mother wavelets for both high and low pass filters. After that, the boundary condition must be corrected by removing the affected approximate and detail coefficient, which number are calculated by the given equation:

$$L_j = (2^{j-1})(L - 1) + 1 \quad (2.29)$$

where L_j is the number of approximate and detail coefficient affected, while J and L depict the decomposition level and length of the wavelet function. This idea is agreed by several researchers, as the adoption of this best practice was observed in the study of Zhou, Liu and Duan (2020) and Xu, Chen and Zhang (2021). Despite the convincing theoretical advantages of SWT over DWT, a side-by-side comparison of the two transformation is still of interest, as practicality might limit the performance of SWT below its potential, similar to the implementation of ANN over ARIMA.

The significance of direct framework (DF) and multicomponent framework (MF) was investigated in the research of Xu, Chen and Zhang (2021). These two frameworks are differentiated by the group of variables to be decomposed. For DF, only the input variable is decomposed, whereas for MF, both the input and output variables are decomposed. In MF, the decomposed subseries of output are predicted and the original output is reconstructed by inverse wavelet transformation. Since SWT was utilized in this study, the original output series was recovered by additive reconstruction of the subseries. The results of this study highly resembled a similar research previously

conducted by Zhou, Liu and Duan (2020). The authors of both studies agreed on the improvement brought by wavelet transformation, after comparing the performance of coupled and standalone models. Superiority of MF over DF was witnessed in both studies, and the advantage become more apparent as the forecasting lead time increased. As the lead time increased from 1 day to 5 days, the difference between root mean square error (RMSE) of MF and DF become more apparent (Xu, Chen and Zhang, 2021). The superiority of hybrid model over standalone model also become more prevalent with the increment of lead time, despite the reduction of the overall accuracy of all the models (Zhou, Lin and Duan, 2020). Anyhow, the study area of these two researches was limited. The importance of the ignored influential factors such as networks architecture and selection of mother wavelets were recognized by the authors. Thus, further investigation on the mentioned ignored factors is warranted.

2.5 Summary

Streamflow forecasting is critical in the water-related field because of the benefits it can bring to society and the environment. With the advancement of technology, AI-based models have been extensively used as an alternative to traditional physically-based hydrological models. The advantage of such AI-based models lies in their ability to produce comparable or better performance without relying on the underlying complex relationship of the hydrological systems. Nonetheless, as evidenced from the literature review, there are a few research gaps that needed to be considered, which are as follows:

- i. A lack of thorough investigation of the decomposed wavelets as input into the AI-based models. The pre-selection of decomposed wavelets can inhibit the potential of an AI-based model, considering that the discarded decomposed wavelets may contain meaningful information to be perceived by the model.
- ii. The efficiency of the wavelet transform in handling the forecasting horizons (time lead- 1 day up to 7 days) needs to be better understood.
- iii. A comparison of discrete and stationary wavelet transforms in the streamflow forecasting application using AI models is lacking to justify the better approach.

CHAPTER 3

METHODOLOGY AND WORK PLAN

3.1 Introduction

This chapter describes the study's procedure and methodologies. This study's primary works include wavelet decomposition, ANN construction, and statistical comparisons. The rationale and necessity of adopting the methods were revealed with justification given in the previous studies available in the literature review. Work was meticulously planned to guarantee the study's smooth flow and optimal resource allocation. The overall workflow of the study is illustrated in Figure 3.1.

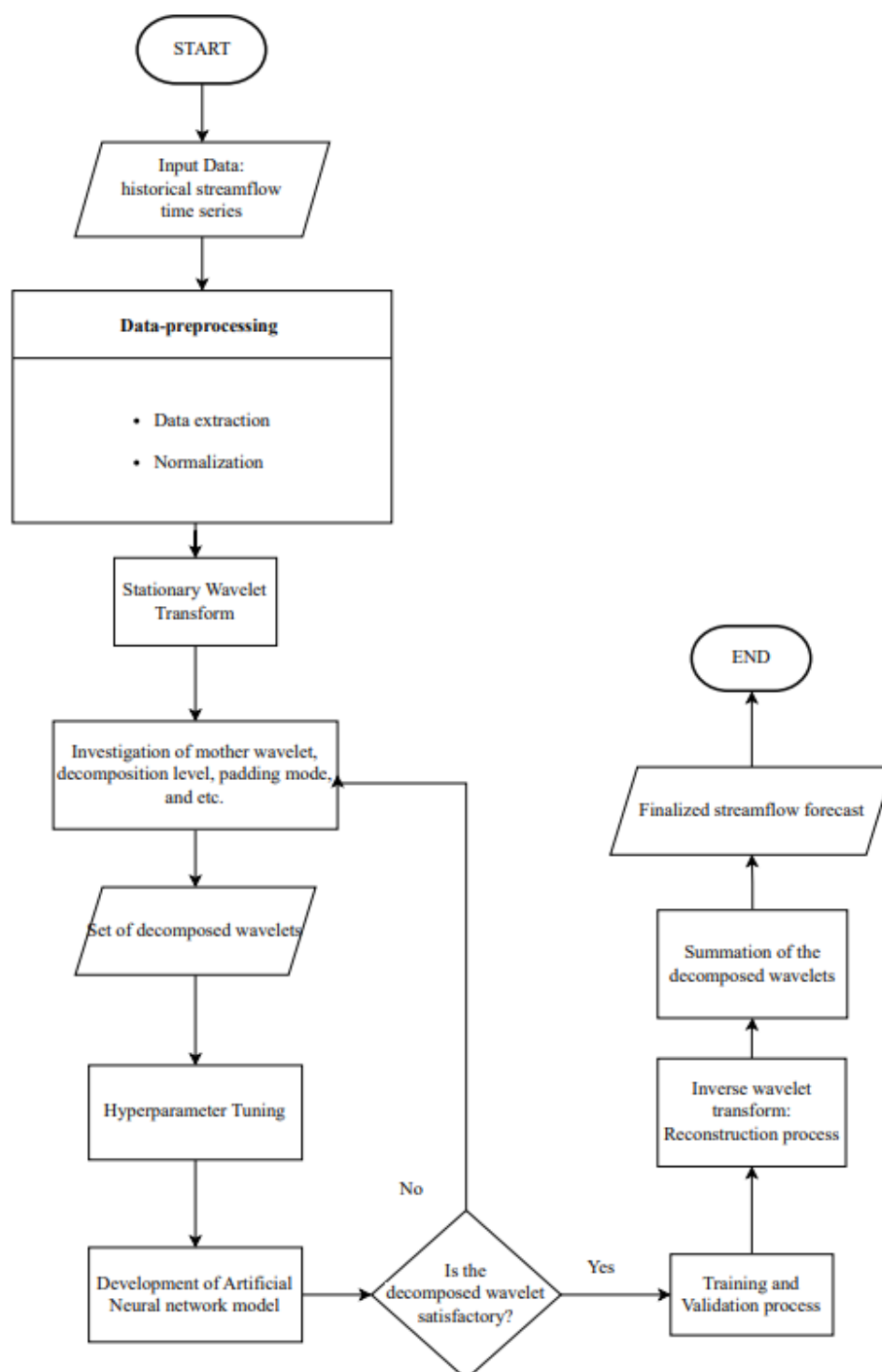


Figure 3.1: Flowchart of Streamflow Forecasting

3.2 Study Area and Data Acquisition

Johor is the only coastal state on Peninsular Malaysia with a east coast and a west coast as well, and it borders Singapore at the southern end of the peninsular, as shown in Figure 3.2. It has a 400-kilometer shoreline on both the east and west coastlines. It is known for its tropical rainforest environment, with the

South China Sea monsoon season blowing from November to February. The average annual rainfall is 1,788 mm, with an average temperature of 26.7°C and humidity levels of 84%. The main river in the Johor state is the Johor River, with an approximate length of 130 km and an area catchment of 2,600 km². The Johor River originates from Mount Gemuruh and flows in the north-south direction before emptying its flow into the Straits of Johor. Its main tributaries are the Linggui River in the north, the Sayong River in the northwest, the Tiram River in the southwest, and the Lebam River in the southeast. The temperature is estimated to be between 21 and 32 degrees Celsius.

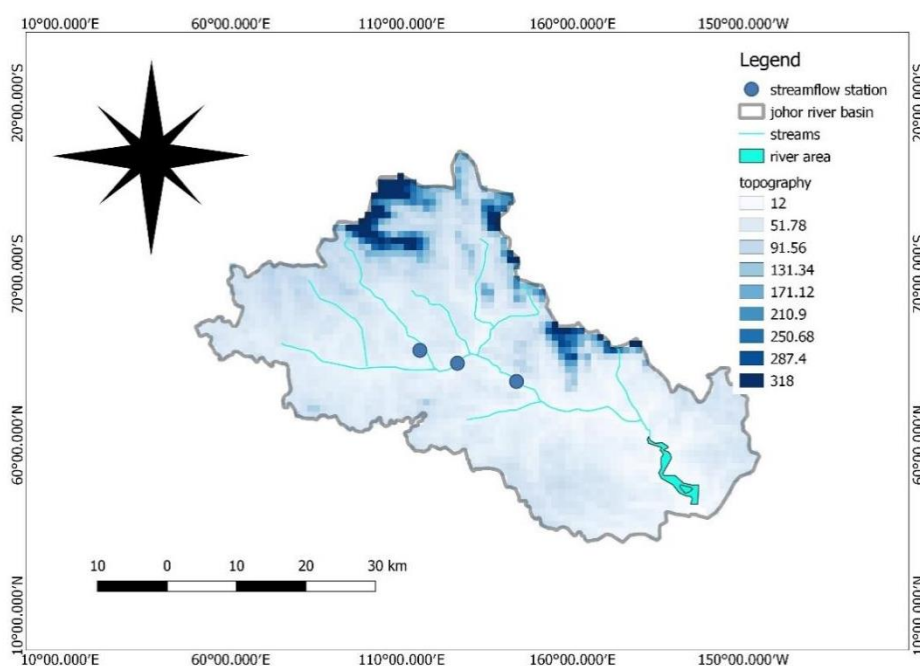


Figure 3.2: Location of Johor River Basin

In this study, the univariate streamflow time series will be utilized as the input of the forecasting. Thus, the historical streamflow dataset was obtained from an existing established hydrological station. The following streamflow station's data in Table 3.1 is sourced from Malaysia's Department of Irrigation and Drainage (DID), as shown.

Table 3.1: The List of Hydrological Observation Stations

List of hydrological observation stations				
Stn No	Station name	Location	Catchment area (km²)	Observation period
	Sg. Sayong at			
1836402	Jambatan Johor Tenggara	N 01048'15" E 103040'10"	624	1977-2008

3.3 Stationary Wavelet Transform

The SWT was utilized to decompose the input time series before input for model training, to increase the predictive accuracy of the forecasting model. Several studies advocated for the use of the SWT in hydrological forecasting as it may reduce boundary condition error (Quilty and Adamowski, 2018, 2021; Zhou, Liu, and Duan, 2020).

In SWT, two sets of coefficients, known as the detail coefficient and the approximation coefficient, were created from the high pass and loss pass filters, respectively. These coefficients were computed by convoluting input time series and wavelet functions within each filter. Consequently, careful selection of these wavelet functions is required to achieve high prediction accuracy via effective decomposition of the input signals. However, there is no convincing standard of wavelet selection presented to date. This study thus looked at the effects of using various mother wavelets as the wavelet (high pass) and scaling (low pass) functions. The list of mother wavelets to be investigated is provided in Table 3.2:

Table 3.2: List of Wavelets to be Tested

Number	Name of Wavelet
1	Symlets 5
2	Daubechies 5
3	Coiflet 5

Another significant factor influencing the performance of the wavelet transform is the decomposition level. Proper selection of the decomposition level is necessary to achieve a balance between computation complexity and predictive accuracy. Decomposition levels from one to five were applied in this study, with their performance analyzed by the statistical parameters including the coefficient of determination (R^2), mean absolute error (MAE) and root mean squared error (RMSE).

Poor decomposition level selection increases boundary condition error in the SWT, necessitating a three-step mitigation strategy proposed by Quilty and Adamowski (2018). The authors' three steps included (1) selecting the maximal overlap discrete wavelet transform (MODWT) or SWT for wavelet decomposition, (2) determining the decomposition level and wavelet functions for filters, and finally, (3) correcting boundary error by removing the affected approximate and detail coefficients, the number of which was determined by Equation 2.29.

The approximate coefficient was calculated based on the convolution of the input time series and the scaling function, as given by Maheswaran and Khosa (2012) in Equation 2.26. After computing the approximate coefficient for a particular level of decomposition, the corresponding detail coefficient was calculated by Equation 2.27. The reconstruction of the input time series was carried out by direct addition of the coefficients, as expressed in Equation 2.28.

3.4 Development of Artificial Neural Network (ANN)

An Artificial Neural Network of one input layer, a varied hidden layer, and one output layer was developed with Python to forecast the streamflow. The model's input is a univariant streamflow time series with varying lead times, which after transformation by the hidden layer, delivers the anticipated streamflow output. As previously noted in the literature review chapter, the model's prediction accuracy is dependent on the model's design and training technique. However, standards for selecting architecture are yet to be concluded, despite some empirical equations proposed to facilitate the selection. Thus, the architecture of the ANN, including the number of hidden layers and the number of neurons for each layer, was decided based on sequential optimization, termed Bayesian optimization. The performance of different architectures is statistically analyzed

through R^2 , MAE and RMSE, and the best-performing architecture will be adopted for further investigation. The training algorithm utilized to modify the parameters was the Levenberg-Marquardt algorithm, which has proven effective in the research of Kişi (2007) and Sharma, Singh, and Sharma (2021).

The number of neurons in the hidden layer should be properly designed to avoid the issue of underfitting or overfitting. Underfitting the model is undesirable due to its low prediction capabilities; although overfitting the model may produce high accuracy, it has poor generalization ability due to the model's overtraining. As a result, the input data were separated into 80%: 20% (training: testing) to ensure appropriate but not over-calibration of the model while conserving adequate data for validation (Loh et al., 2021). Since the effect of the activation function is out of the scope of the study, the developed models only leveraged the sigmoid function as the activation function. The architecture of the developed ANN is illustrated in Figure 3.3:

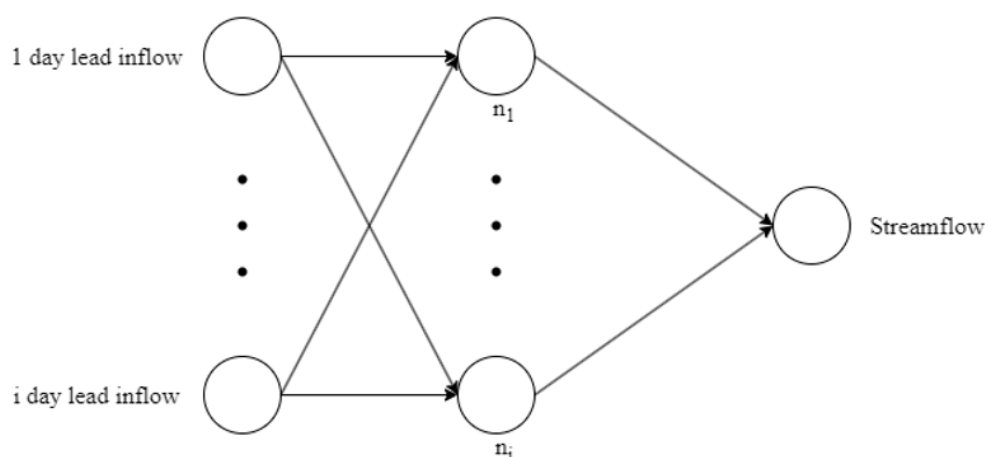


Figure 3.3: Network Architecture of ANN

The output of the developed ANN could be represented by Equation 2.1.

3.5 Statistical Test

The performance assessment of model is imperative in identifying the most effective incorporating wavelet transformation and neural network. With reliable evaluative parameters computed, the selection of alternatives could be justified by evidence. In this study, the statistical parameters MAE and RMSE

were utilized to measure the performance of the alternatives by calculating the error of forecasted outcome relative to the actual outcome.

As suggested by its terminology, the mean absolute error (MAE) is obtained by computing the absolute errors. The deviation between the expected value and the obtained value of the prediction is known as the error. Statistically, MAE suggests the distance of the data to the best-fitting regression line. That said, if the value of MAE is low, an accurate result is implied. The calculation of MAE was done by utilizing the formula expressed as:

$$MAE = \frac{1}{n} \sum (y_i - x_i) \quad (3.1)$$

where y_i , x_i , and n depict the expected outcome, obtained outcome and number of predictions, respectively. The RMSE is essentially the square root of the MSE computed to facilitate the interpretation of model performance, owing to its advantage of having the same unit as the forecasted parameter. The calculation of RMSE was performed by the equation:

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - x_i)^2} \quad (3.2)$$

The coefficient of determination, R^2 is a metric to evaluate the the prediction of a statistical model. The R^2 has a range a value from 0 to 1, with 1 being the most accurate while 0 being the least accurate. The coefficient of determination in this study was calculated by the equation:

$$R^2 = \frac{\sum(x_i - y_m)^2}{\sum(y_i - y_m)^2} \quad (3.3)$$

where y_m depicts the mean value of expected outcome (Kambalimath and Deka, 2021).

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

The results and discussion of the study are demonstrated in this chapter. Presentation of results, alongside the necessary comparison and explanation, are performed to investigate the performance of the AI models studied. In light of the numerous significant elements influencing the prediction performance under examination, the influence of each researched factor is analyzed separately in specified subsections to determine the relative contribution of each component to the prediction performance.

4.2 Hyperparameter Tuning Analysis

Two different scenarios and two different cases of streamflow modelling were taken into consideration in this study. Hence, a total of four combinations of scenarios and cases were investigated, denoted as:

- i. Scenario 1 Case 1 (S1C1),
- ii. Scenario 1 Case 2 (S1C2),
- iii. Scenario 2 Case 1 (S2C1) and,
- iv. Scenario 2 Case 2 (S2C2) respectively.

The difference between Scenario 1 (S1) and Scenario 2 (S2) arises from the sequence of streamflow prediction.

- i. In S1, the entire data set comprising 348 streamflow data was first decomposed by wavelet transformation into respective sub-signals. After being divided into training and testing data sets, the decomposed data were used in the ANN model training.
- ii. With S2, the data set was first divided into training and testing data. Wavelet decomposition of training data was executed, followed by ANN model training with the decomposed training data set. The decomposition of the testing data set took place after model training.

In this work, Case 1 (C1) and Case 2 (C2) illustrate univariate and multivariate streamflow forecasting, respectively.

- i. Several models were generated for the univariate C1 by recurrent training of the model at different decomposition levels. However, only one input variable was taken into account in the forecast.
- ii. In the case of multivariate C2, only one prediction model was created during the machine's first training. Contrary to univariate forecasting, C2 requires multiple inputs and is capable of identifying the relationship between these variables.

To begin the comparison, the effect of hyperparameter tuning was first investigated to justify further comparison in choosing the tuned version as a sample rather than the untuned version. As previously mentioned, a total of four combinations of scenarios and cases are possible in this study, whereby each of the four combinations has another three versions with different incorporating wavelets, namely:

- i. the *Symlets 5 (sym5)*,
- ii. *Daubechies 5 (db5)* and
- iii. *Coiflet 5 (coif5)* wavelet.

This study looks at a total of 12 distinct models. Nevertheless, as this subsection is primarily intended to validate the improvement of hyperparameter tuning, the wavelet variation was omitted by defaulting the wavelet to sym5, leaving just four scenario-case models to be compared to their tuned version in this section. For example, the comparison between S1C1sym5-5 (untuned S1C1 model of sym5 wavelet decomposed to the fifth level) and S1*C1sym5-5 (tuned version) was performed. The performances of all untuned and tuned models are presented in Table 4.1, while the corresponding hyperparameters are tabulated in Table 4.2:

Table 4.1: Performance of Untuned and Tuned Models

Model	LOD	Metrics					
		R ²		RMSE (m ³ /s)		MAE (m ³ /s)	
		Untuned	Tuned	Untuned	Tuned	Untuned	Tuned
S1C1	5	0.9598	0.9687	79.159	69.059	55.457	49.501
S1C2		0.3478	0.5158	324.799	274.407	222.005	196.436
S2C1		0.0886	0.1490	514.404	694.342	375.281	514.111
S2C2		0.1226	0.0802	679.264	521.807	527.600	385.285

Table 4.2: Hyperparameters of All Models

Model	Parameters					
	Learning Rate	Activation Function	Layers	Nodes	Steps	
S1C1	0.00417	ReLU	3	3	10	
S1C2	0.04699	ReLU	1	2	9	
S2C1	0.00081	ReLU	3	64	3	
S2C2	0.01862	sigmoid	1	8	3	
Untuned	0.01000	tanh	1	5	4	

Based on Table 4.1, out of four scenario-case models, three have improved tuned performance compared to the untuned version, with the exception of S2C2, in which a decline in R² metric from 0.1226 to 0.0802 is observed after tuning. Indeed, the performance of an ANN model relies on its architecture and the weightage of each node. The best set of parameters should be established after tuning. Thus, it is reasonable to expect a higher performance of the tuned model over the untuned model. The performance loss of S2C2 is unexpected and may be attributable to the sigmoid activation function in the model, considering that other models using the rectified linear unit (ReLU) had considerable R² improvements ranging from 0.93% to 68.17% compared to the untuned values. In fact, the application of conventional sigmoid as an activation function has been criticized in several research (Falamarzi et al., 2014; Adamowski and Sun, 2010). The incapability of the sigmoid function to identify the local minima of the input series might explain the inefficiency.

Despite the fact that interpreting R^2 values might give meaningful insight into model performance, it is not conclusive in assessing model accuracy. Therefore, apart from R^2 , other performance measures like the root mean squared error (RMSE) and mean absolute error (MAE) are examined since these metrics reflect the performance of an individual prediction rather than the overall performance of the regression model. Although frequently observed simultaneously, a high R^2 number may not necessarily promise a low RMSE or MAE value. In other words, a highly correlated regression model with a high R^2 value may also have a high individual prediction deviation, as measured by a high RMSE or MAE value. An example of the condition is witnessed in Table 4.1, whereby despite a higher R^2 value of S2*C1 (0.1490) than S2C1 (0.0886), a higher set of RMSE and MAE (694.342 m³/s , 514.111 m³/s) is observed as compared to those of S2C1 (514.404 m³/s, 375.281 m³/s).

The RMSE and MAE of each model were evaluated to validate the findings of the R^2 comparison. For S1C1 and S1C2, lower sets of RMSE and MAE values are witnessed in the tuned version. The inference made from a lower set of RMSE and MAE is that the model has better accuracy since the individual predicted value is less deviated from the observed value. This finding agrees with the finding of the R^2 comparison, showing the superior performance of a tuned model relative to its untuned version. However, for S2C1, the better performance of R^2 , a higher RMSE, and MAE set is witnessed in the tuned version. Although the simultaneous occurrence of high R^2 , RMSE, and MAE is possible, as explained previously, the findings inferred from these observations are indeed conflicting. The higher set of RMSE and MAE (694.342 m³/s, 514.111 m³/s) of the tuned version over the untuned version (514.404 m³/s, 375.281 m³/s) indicates that the performance of the untuned version is better. Anyhow, for this scenario-case model, it is assumed that the finding of R^2 controls, since the percentage of improvement in R^2 by tuning is higher (68.17%) than the percentage of increment in RMSE and MAE by tuning (34.98%, 36.99%). Lastly, for S2C2, the outcome of the R^2 comparison conflicts with the RMSE and MAE comparisons. Although the RMSE and MAE improved after turning in this scenario-case model, they contradict the conclusion of R^2 , which implies the tuning for this model is ineffective. Similarly, the finding of R^2 is accepted, considering the higher percentage difference of R^2 (34.58%) than the

percentage difference of RMSE and MAE (25.16% and 26.97%) after tuning. The comparisons of untuned and tuned models for each scenario-case are visualized in Figure 4.1, Figure 4.2, Figure 4.3, and Figure 4.4:

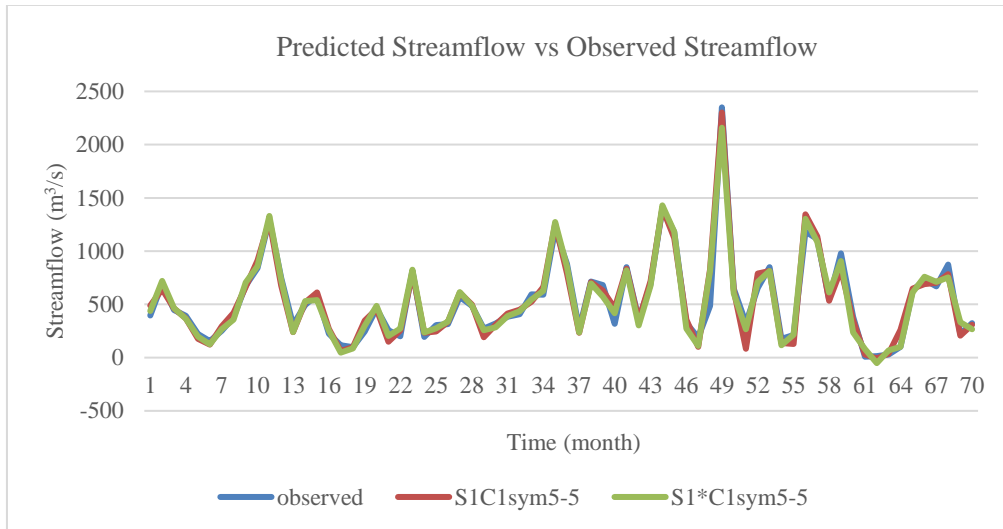


Figure 4.1: Performance of Untuned and Tuned Models in S1C1

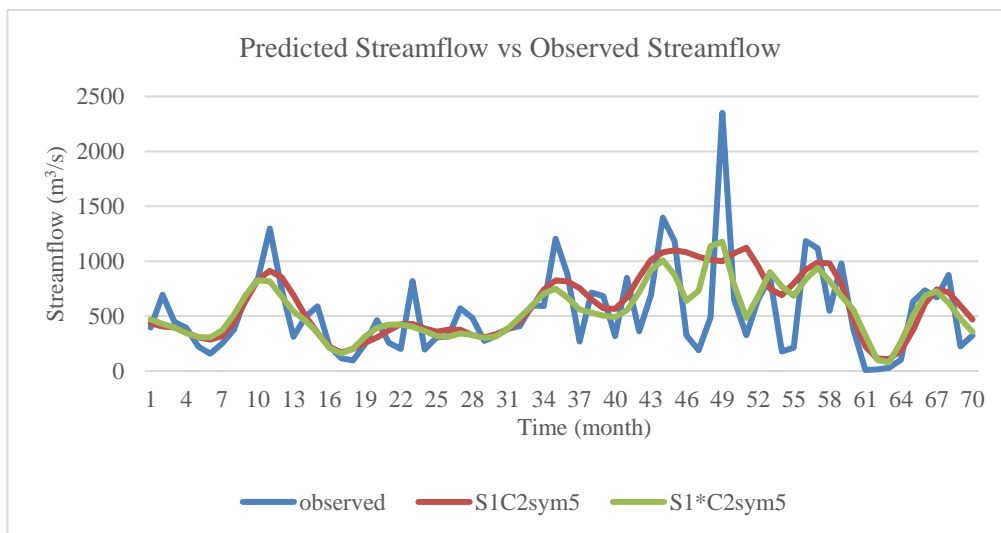


Figure 4.2: Performance of Untuned and Tuned Models in S1C2

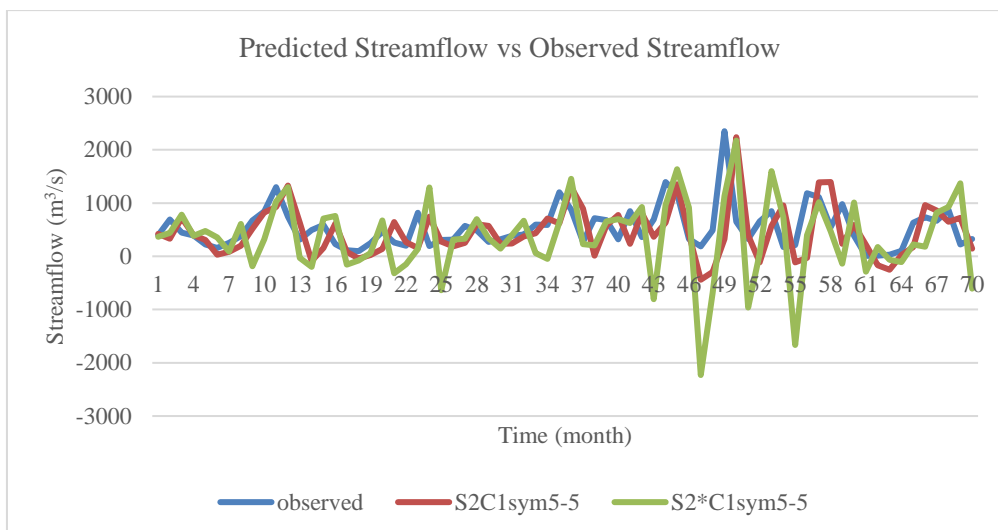


Figure 4.3: Performance of Untuned and Tuned Models in S2C1

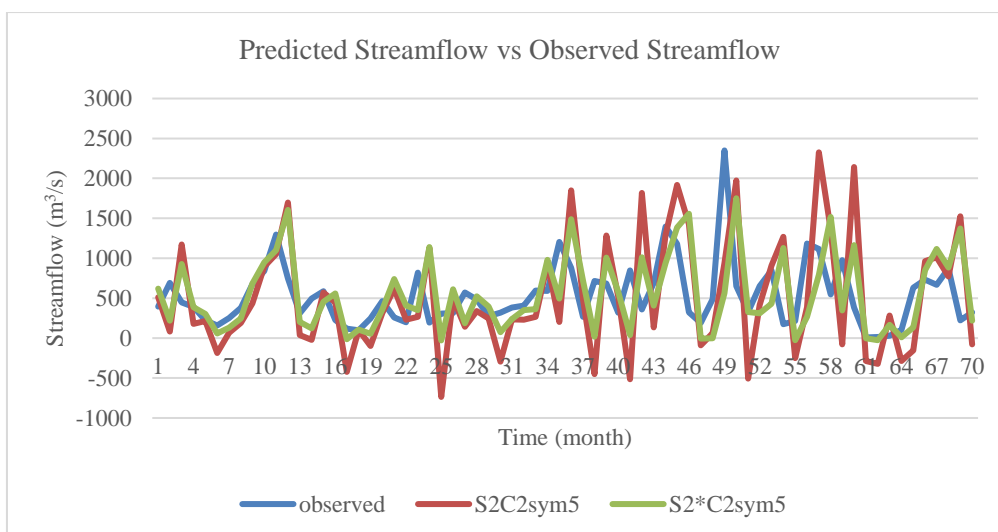


Figure 4.4: Performance of Untuned and Tuned Models in S2C2

Summarizing the findings for the comparison between untuned and tuned models, three out of four scenario-case models exhibit accuracy improvement in this study. The odd model (S2C2) with declined accuracy is suspected to have experienced ineffective tuning of the sigmoid activation function. Considering the general trend and by a majority, it is rational to recognize the improvement of hyperparameter tuning in streamflow forecasting. Henceforth, the comparison and analysis in the subsequent subsections are conducted with the tuned models by default.

In fact, it is realized that many different sets of hyperparameters could return the same results as the recorded set. Hence, the learning algorithm of the

model is worth further investigation. For example, since Bayesian optimization is adopted in this study, further comparison with grid search and random search optimization methods could be conducted. The Bayesian optimization was conducted using Gaussian processes with the assumption that the objective function of the ANN model followed a multivariate Gaussian. Therefore, a tree-based regression model could be further investigated.

4.3 Wavelet Analysis

4.3.1 Comparison of Hybrid ANN and Standalone ANN

As discussed previously in the chapter on literature review, the incorporation of wavelet decomposition with ANN often returns encouraging improvement (Xu, Chen, and Zhang, 2021; Li, Wang, and Qiu, 2019). The standalone ANN and hybrid ANN and wavelet models are compared in this work to first validate this hypothesis. In this comparison, since both standalone and hybrid models were tuned, the contribution of tuning to accuracy is assumed balanced so that any observed improvement is solely contributed by the incorporation of the wavelet decomposition. However, it is realized that the performance of parameter tuning in each model could be different, resulting in performance discrepancies. Anyhow, this plausible difference is not reckoned. Henceforth, the observed improvement is attributed to the wavelet decomposition. The effect of different mother wavelets will be discussed later in a designated section to eliminate the influential factors in this comparison. Also, only univariate forecasting was conducted for the standalone model, owing to the singular type of result. Therefore, the hybrid models discussed in this section are tuned univariate models with sym5 wavelet decomposition to the fifth level (S1*C1sym5-5 or S2*C1sym5-5). Due to the inclusion of wavelet, the wavelet transform is only relevant to hybrid models. The performance of the standalone and hybrid models is presented in Table 4.3:

Table 4.3: Performance of Standalone and Hybrid ANN models in both Scenarios

Scenario	LOD	Metrics					
		R ²		RMSE (m ³ /s)		MAE (m ³ /s)	
		Alone	Hybrid	Alone	Hybrid	Alone	Hybrid
1	5	0.0819	0.9687	385.812	69.059	276.925	49.501
2		0.0087	0.1490	470.265	694.342	341.847	514.111

Through comparison of Table 4.3, the significant improvement of hybrid models over standalone models is recognized. For S1, the R² of the standalone model increased from 0.0819 to 0.9687. By incorporating wavelet decomposition, the ANN model in this particular study experienced an improvement of over 1082.78% relative to the standalone value. Reduction of RMSE and MAE is also observed in S1, with respective percentages of reduction of 82.10% and 82.12% relative to the standalone values. The improvement of R² is also witnessed in S2 from 0.0087 to 0.1490. By calculation, this improvement is equivalent to around 1612.64% of the standalone value. Although an increment of RMSE and MAE is noticed in S2, this performance reduction is far not as significant as the improvement of R². The increment of RMSE and MAE in S2 are calculated as 47.65% and 50.39%. These results provide clear evidence and justification for the incorporation of wavelet decomposition, for it can enhance the accuracy of the standalone models by more than 1000% in terms of R². This finding is consistent with the previous research discussed in the literature review chapter (Xu, Chen, and Zhang, 2021; Li, Wang, and Qiu, 2019; Feng et al., 2022). The improvement is contributed by the capabilities of wavelet in solving forecasting problems with nonlinearity characteristics. Zhou, Liu, and Duan (2020) suggested that the multiresolution analysis enabled by wavelet decomposition is effective in solving nonlinearity, which is one of the weaknesses of ANN models. The performance of streamflow forecasting by standalone and hybrid models is demonstrated in Figure 4.5 and Figure 4.6:

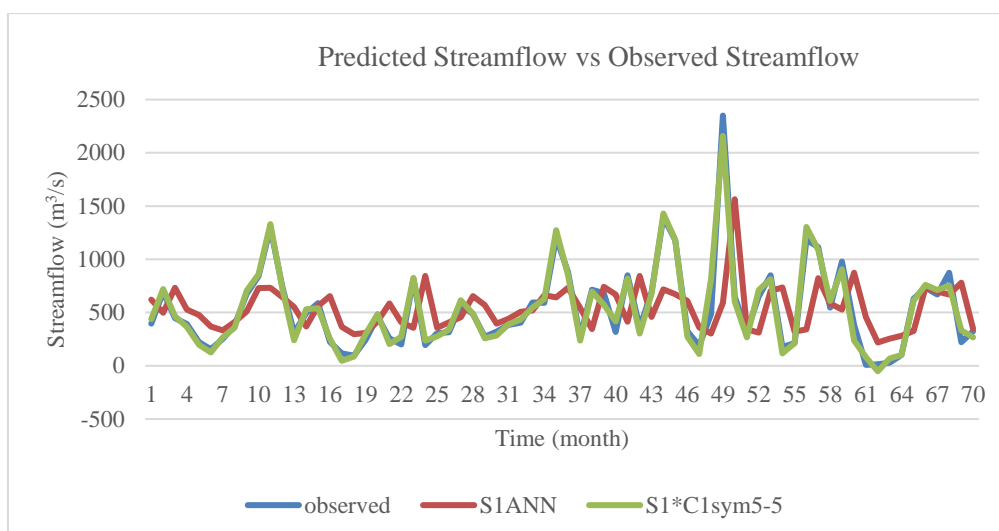


Figure 4.5: Performance of Standalone and Hybrid Models in S1

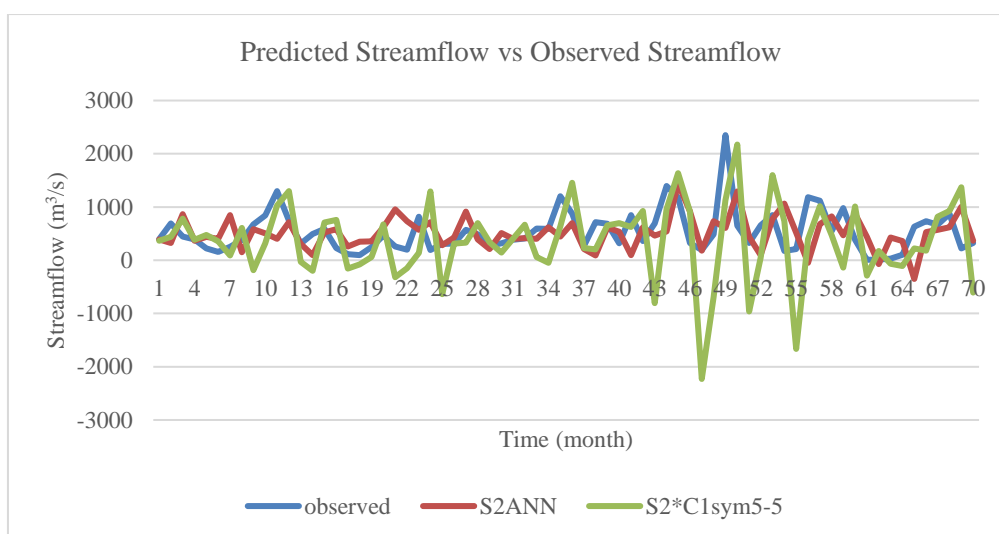


Figure 4.6: Performance of Standalone and Hybrid Models in S2

4.3.2 Analysis of 2 Different Scenarios, 2 Different Cases (Univariate C1 or Multivariate C2) and 3 Different Wavelets

4.3.2.1 Comparisons of S1C1 and S2C1 Cases with Three Different Wavelets

The original streamflow data set obtained from the Department of Irrigation and Drainage (DID) is comprised of 348 historical streamflow data. This set was subsequently divided into 278 training and 70 testing data according to an 80%:20% proportion, which is suggested by Loh et al. (2021) as an effective ratio of the dataset. The prediction accuracy of a model is reflected by its ability to reproduce the testing data set. For C1, the performance of S1 and S2 with

different wavelets at the level of decomposition (LOD) 5 is presented in Table 4.4:

Table 4.4: Performance of S1C1 and S2C1 cases with three different wavelets

Wavelet	LOD	Metrics					
		R ²		RMSE (m ³ /s)		MAE (m ³ /s)	
		S1	S2	S1	S2	S1	S2
sym5		0.9687	0.1490	69.059	694.342	49.501	514.111
db5	5	0.0339	0.1896	698.138	704.077	517.083	508.053
coif5		0.0115	0.1561	890.479	528.106	727.268	396.066

The results of C1 were obtained in the form of decomposed subseries. In order to compare the subseries with the reference original time series, reconstruction of the decomposed subseries was first performed. The reconstruction of the combined wavelet from decomposed wavelet was completed by direct summation of all the subseries, including the detail and approximate series. For example, to reconstruct the combined series for the S1*C1sym5-5 model (tuned Scenario 1 C1 model with sym5 wavelet decomposed to the level 5), the approximate series (A) and the detailed series from level 5 until level 1 (D5-D1) obtained from modeling were added. The reconstruction of decomposed wavelet is illustrated in Figure 4.7. Figure 4.8, Figure 4.9, Figure 4.10, Figure 4.11, Figure 4.12, and Figure 4.13. The combined series was then compared to the original time series, as shown in Figure 4.14:

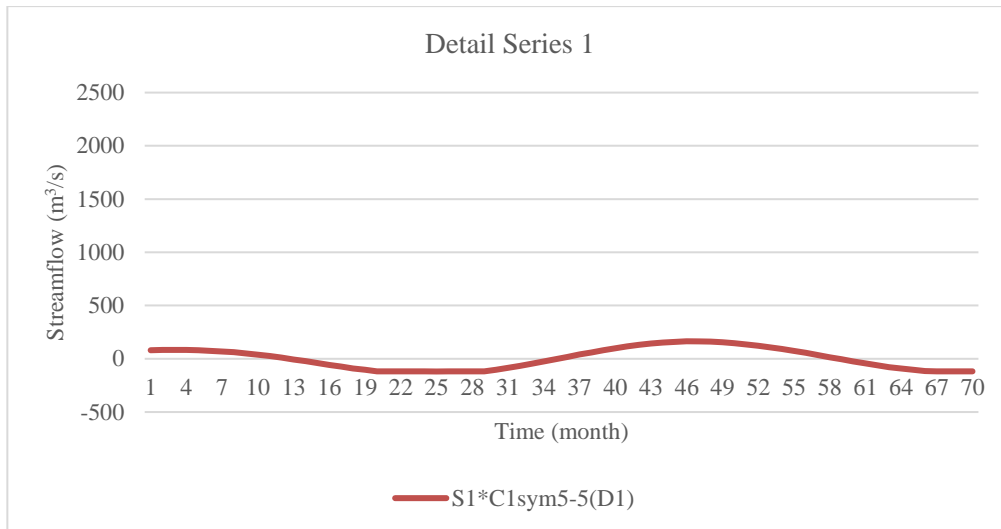


Figure 4.7: Decomposed Detail Series 1 of S1*C1sym5-5

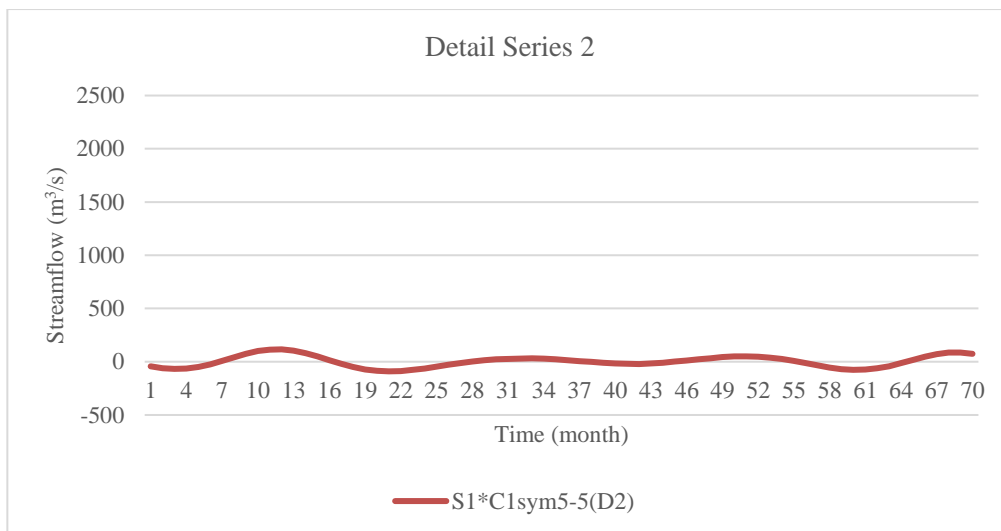


Figure 4.8: Decomposed Detail Series 2 of S1*C1sym5-5

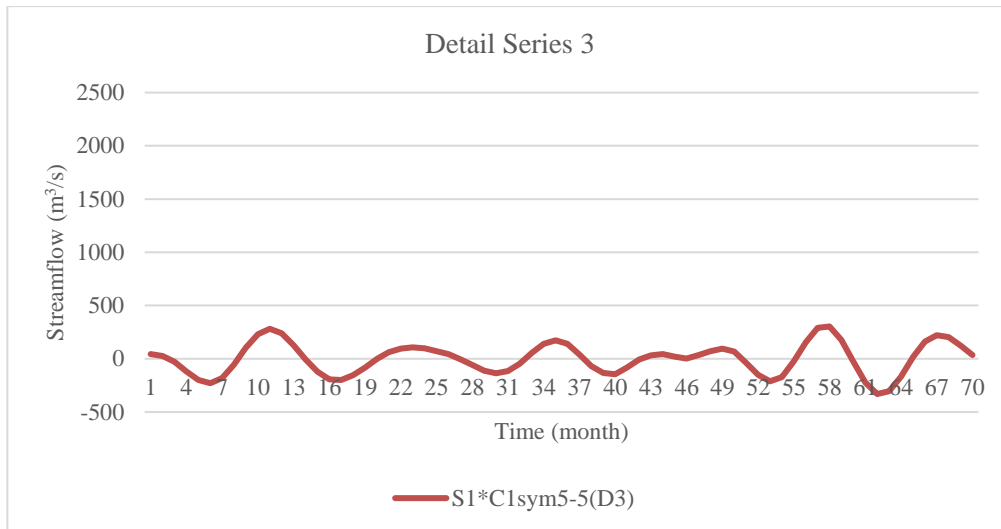


Figure 4.9: Decomposed Detail Series 3 of S1*C1sym5-5

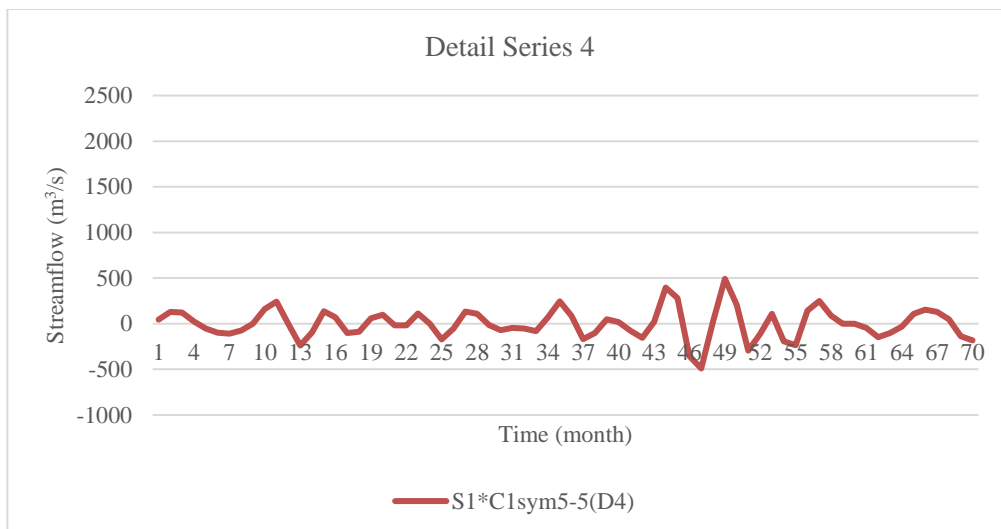


Figure 4.10: Decomposed Detail Series 4 of S1*C1sym5-5

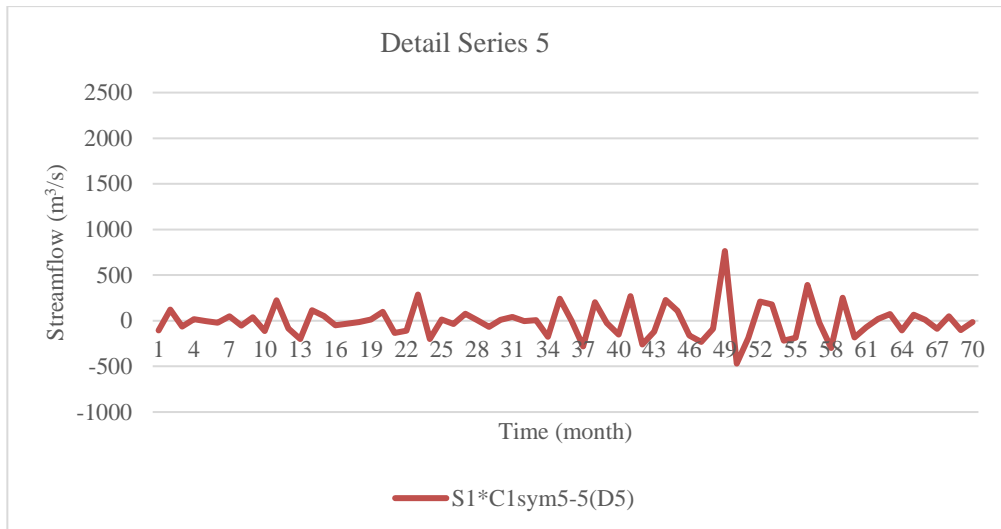


Figure 4.11: Decomposed Detail Series 5 of S1*C1sym5-5

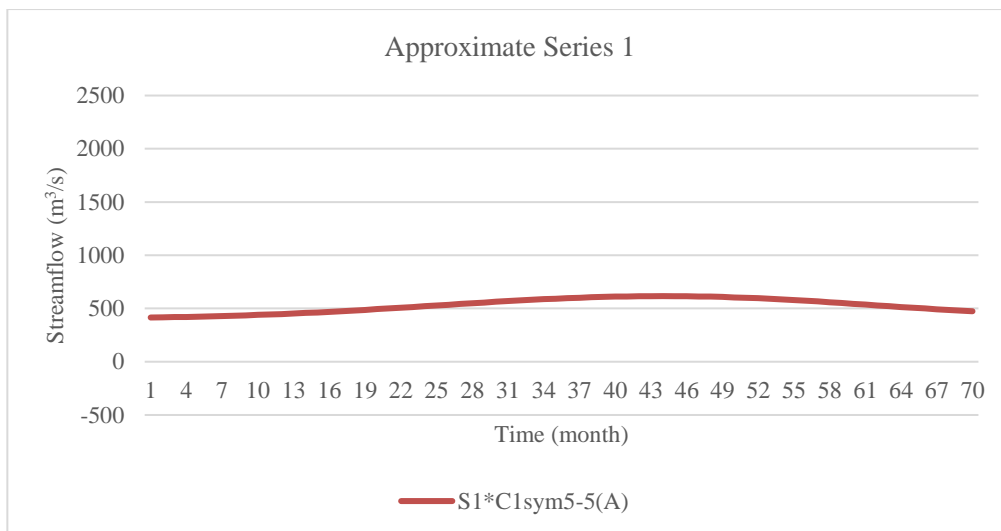


Figure 4.12: Decomposed Approximate Series of S1*C1sym5-5

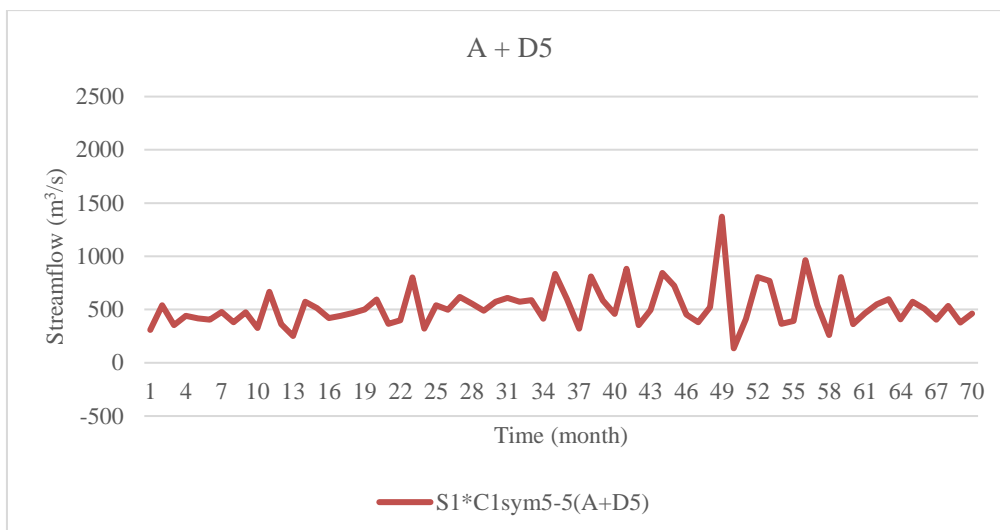


Figure 4.13: Partial Reconstruction by Adding A and D5 of S1*C1sym5-5

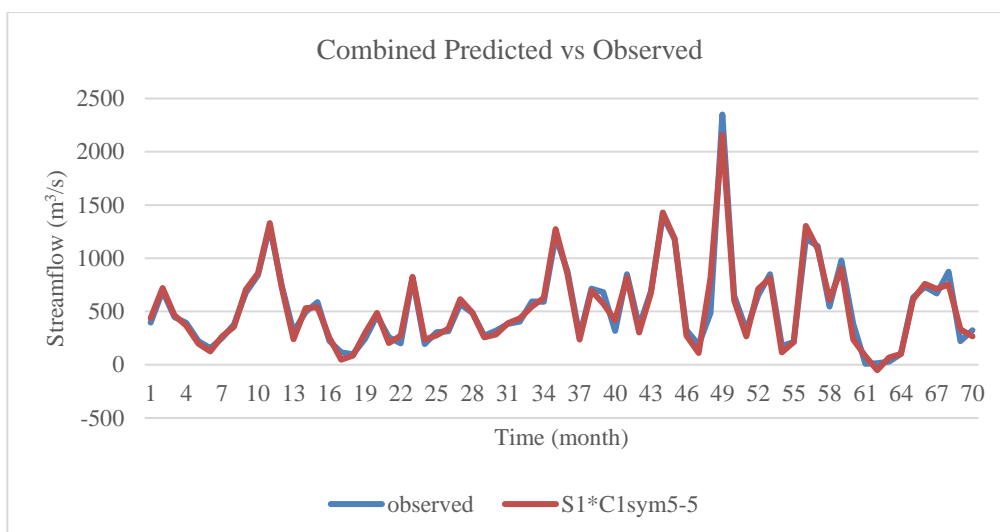


Figure 4.14: Performance of combined S1*C1sym5-5

From Table 4.4, it is observed that S1 of sym5 has a significantly higher value of R^2 as compared to S2. The recorded R^2 for S1 and S2 are 0.9687 and 0.1490, respectively, resulting in a difference of 0.8197. The high value of R^2 obtained in the S1 model suggests that a strong relationship exists between the predicted and observed data of the model, whereby most of the variation in predicted data could be explained by the observed data. In contrast, a different trend of R^2 is observed for the remaining models incorporating db5 and coif5. For db5 incorporated models, the recorded R^2 for S1 and S2 are 0.0339 and 0.1896, spanning a difference of 0.1557. Meanwhile, R^2 values of S1 and S2 models with coif5 as incorporating wavelet are 0.0115 and 0.1561, yielding a

difference of 0.1446. For these two wavelet families, the lower R^2 values of S1 have suggested a weaker regression performance of S1 compared to S2.

Table 4.4 shows that the S1 model has a lower RMSE and MAE than the S2 models for the sym5 wavelet. However, the differences of RMSE and MAE between S1 and S2 are not as significant as the R^2 value. For instance, the sym5 model displays an R^2 percentage difference of 550.13% relative to the S2 value but only displays an RMSE and MAE percentage difference of 90.05% and 90.37% relative to the S2 value. By looking into all three performance metrics, it may be argued that the difference in the performance of S1 and S2 is not as extensive as suggested by the R^2 alone. For db5, a conflicting result was obtained from the comparison of RMSE and MAE. A better-performing lower RMSE value was obtained in S1, which opposes the finding of R^2 . In any case, as previously described in the hyperparameter tuning subsection, it is realized that getting high R^2 and RMSE is possible. In terms of MAE for db5, S1 experienced a higher error of 517.083 m^3/s than 508.053 m^3/s of S2, which is in unison with the R^2 finding. Moving on to the coif5 wavelet, S2 has both RMSE and MAE lower than S1, indicating a better prediction of S2, similar to the finding suggested by R^2 . So far, no conclusive hypothesis can be drawn from the comparison because firm evidence of performance improvement is lacking. The performance comparison of S1 and S2 for each wavelet is visualized in Figure 4.15, Figure 4.16, and Figure 4.17:

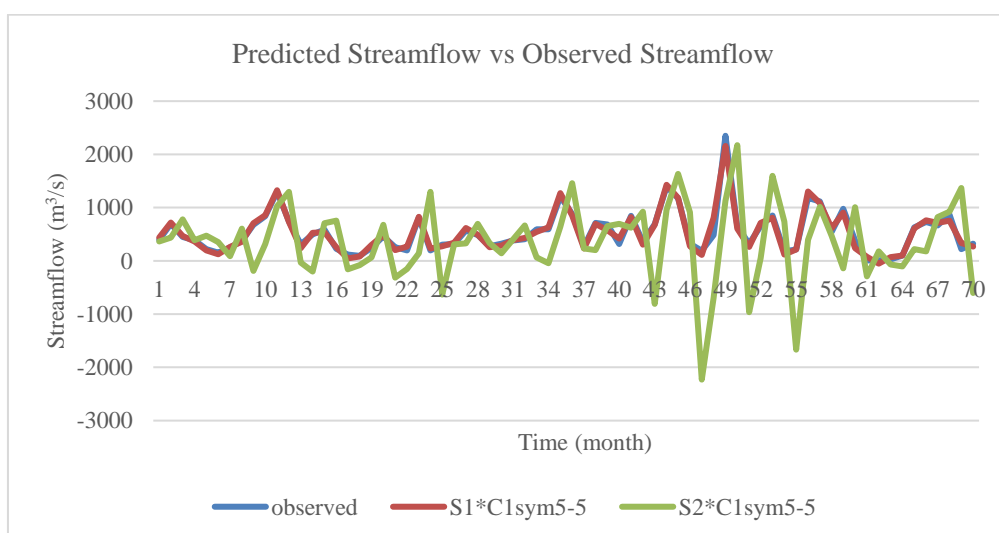


Figure 4.15: Performance of S1C1 and S2C1, with sym5 Wavelet

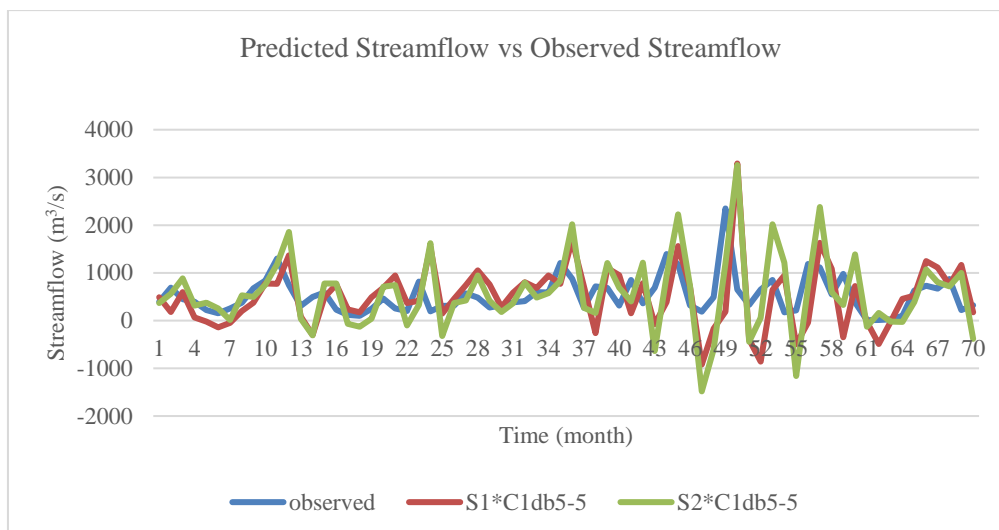


Figure 4.16: Performance of S1C1 and S2 C1, with db5 Wavelet

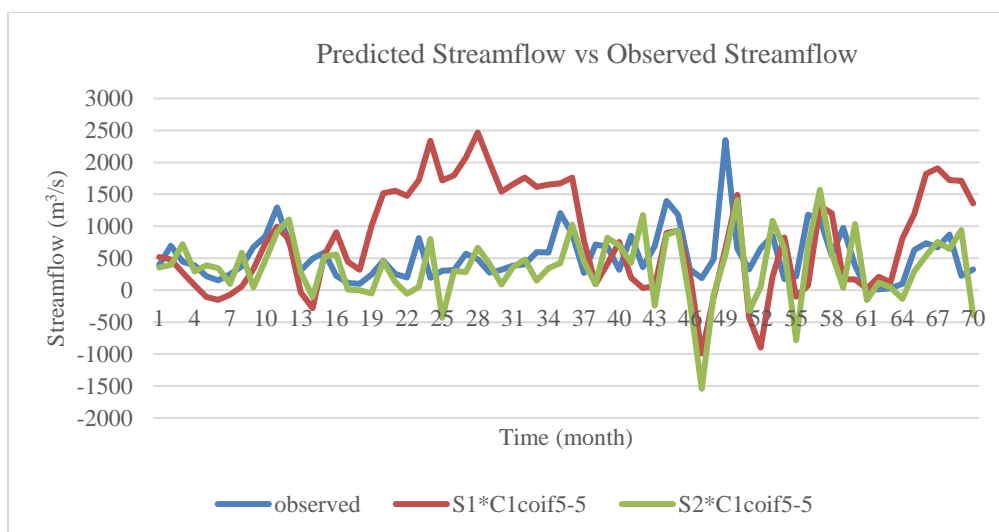


Figure 4.17: Performance of S1C1 and S2C1, coif5 Wavelet

4.3.2.2 Comparisons of S1C2 and S2C2, with three different wavelets

To continue the performance comparison between S1 and S2, the prediction accuracies of Case 2 models were compared to identify the plausible supportive or conflicting evidence to the findings in C1. In order to guarantee the performance difference is solely contributed by the difference in prediction scenarios, other influencing factors, namely the type of wavelet and level of decomposition, were fixed to ignore the corresponding influence on the model accuracy. The performance of S1 and S2 under the same wavelet type and level of decomposition is presented in Table 4.5:

Table 4.5: Performance of S1C2 and S2C2, with three different wavelets

Wavelet	LOD	Metrics					
		R ²		RMSE (m ³ /s)		MAE (m ³ /s)	
		S1	S2	S1	S2	S1	S2
sym5		0.5158	0.0802	274.407	521.807	196.436	385.285
db5	5	0.3421	0.1322	336.992	864.383	240.011	690.963
coif5		0.3513	0.0971	323.095	555.186	230.751	420.658

As observed from Table 4.5, S1 models of all three types of wavelets have achieved higher R² values than the corresponding S2 models. This result appears to confirm the previous finding from the performance comparison of S1 and S2 of sym5 in C1, which suggests that the S1 model has higher predictive accuracy than S2. For Case 2, the S1 models incorporated with sym5, denoted as S1*C2sym5-5 (tuned S1C2 model with sym5 wavelet decomposed to level 5), have returned an R² value of 0.5158, whereas the corresponding S2 model scores an R² value of 0.0802. The deviation of R² between these two models is 0.4356, which is equivalent to a percentage difference of 543.14% relative to the S2 value. Likewise, the S1 models of db5 and coif5 wavelets also demonstrated a better R² performance than the S2 models, showing a percentage difference relative to the S2 value of 158.77% and 261.79%, respectively. Since the R² value is higher in S1 than in S2, it is rational to deduce that a stronger relationship exists between the predicted and observed data in S1 than in S2.

The accuracy of individual prediction in each scenario was evaluated by comparing the RMSE and MAE of the models. By comparison, the S1 model has lower RMSE and MAE than the S2 model, indicating a better prediction performance of S1 since a smaller error is expected in each prediction. This statement holds for all three wavelets under investigation. For instance, the S1 model of the sym5 wavelet returns RMSE and MAE values of 274.407 m³/s and 196.436 m³/s, which is lower than the RMSE and MAE of S2, recorded as 521.807 m³/s and 385.285 m³/s, respectively. Therefore, it is reasonable to conclude that the predictive accuracy of S1 is higher than S2 in C2, as evidenced by the higher R² and lower RMSE and MAE shown in all comparisons. The performance comparison of S1 and S2 for each wavelet is visualized in Figure 4.18, Figure 4.19, and Figure 4.20.

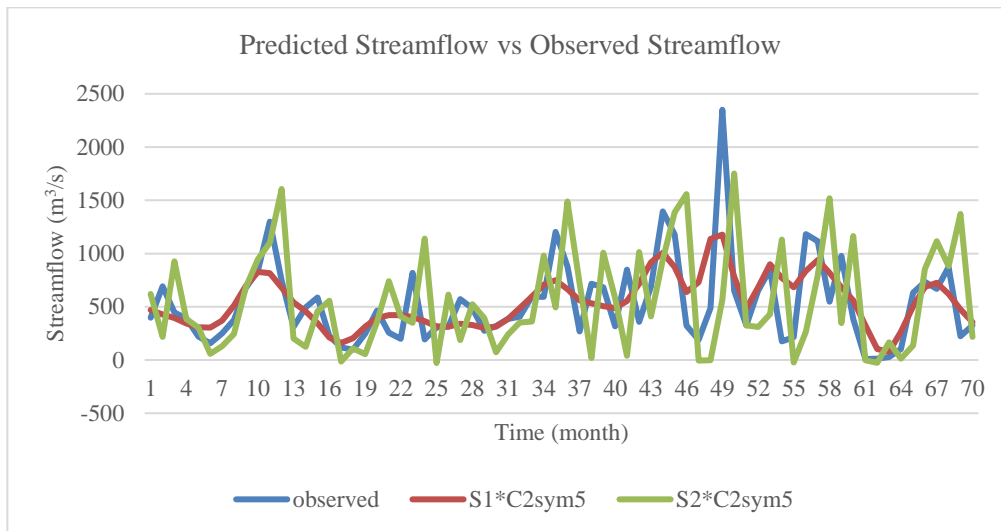


Figure 4.18: Performance of S1C2 and S2C2, with sym5 Wavelet

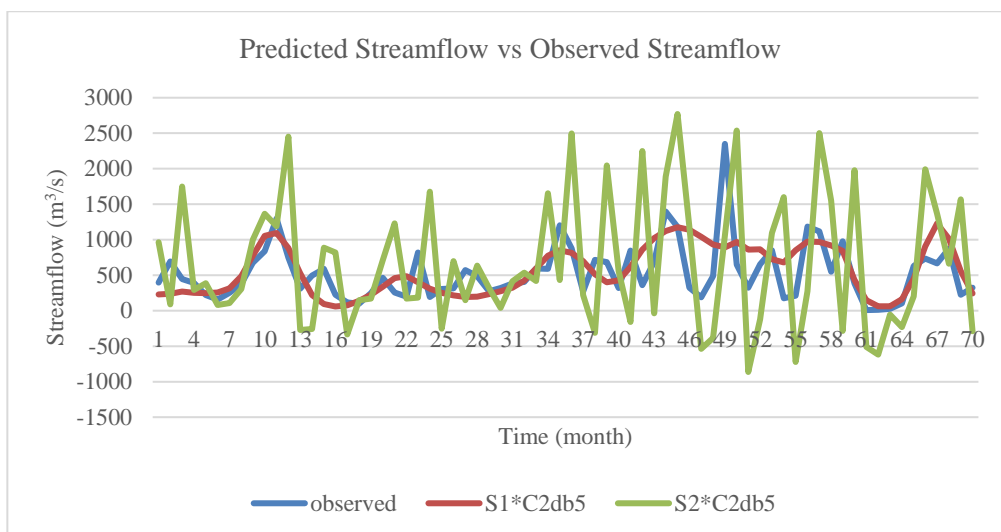


Figure 4.19: Performance of S1C2 and S2C2, with db5 Wavelet

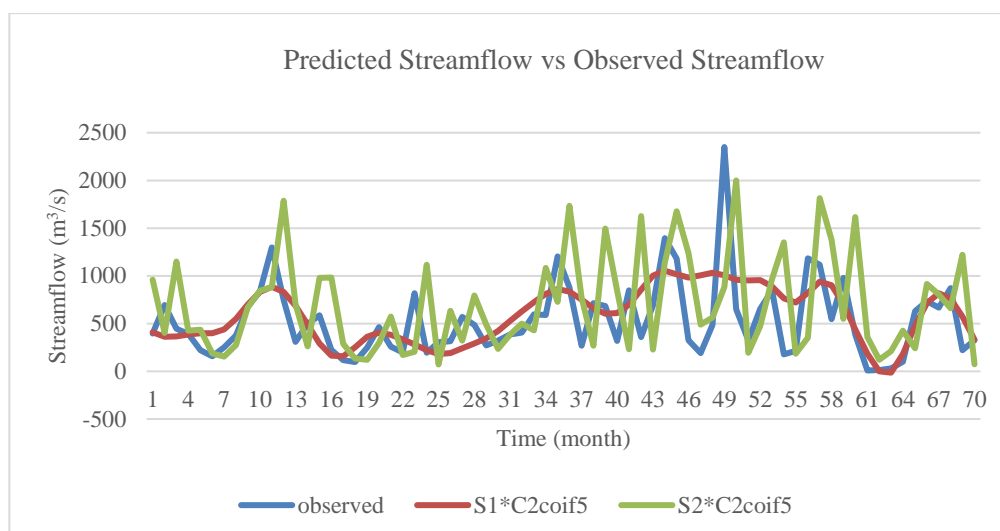


Figure 4.20: Performance of S1C2 and S2C2, coif5 Wavelet

Concluding the findings in both C1 and C2, four out of six comparisons suggested the better performance of S1 than S2. Thus, the hypothesis of S1 prediction having higher accuracy than S2 prediction is accepted. However, the practicality of S1 is questionable and worth discussing, especially when future data is used during training. The decomposition of the input time series took place before model training. The training data set is contaminated; Therefore, the pattern of the testing data might be initially captured, resulting in overfitting the model. In such a scenario, impractical high accuracy might be achieved since the model is predicting the training data set rather than new, unseen testing data.

4.3.2.3 Comparison of C1 (Univariate) and C2 (Multivariate)

Having identified the superior performance of S1 over S2, the performances of C1 and C2 are worth comparing to determine the effectiveness of univariate and multivariate streamflow forecasting. For the purpose of validation, the comparison between C1 and C2 is conducted in both scenarios to increase the credibility and reliability of the result. The comparison of C1 and C2, with S1 is tabulated in Table 4.6, while the comparison of C1 and C2, with S2 is presented in Table 4.7:

Table 4.6: Comparison of C1S1 and C2S1, with three different wavelets

Wavelet	LOD	Metrics					
		R ²		RMSE (m ³ /s)		MAE (m ³ /s)	
		C1	C2	C1	C2	C1	C2
sym5		0.9687	0.5158	69.059	274.407	49.501	196.436
db5	5	0.0339	0.3421	698.138	336.992	517.083	240.011
coif5		0.0115	0.3513	890.479	323.095	727.268	230.751

Table 4.7: Comparison of C1S2 and C2S2, with three different wavelets

Wavelet	LOD	Metrics					
		R ²		RMSE (m ³ /s)		MAE (m ³ /s)	
		C1	C2	C1	C2	C1	C2
sym5		0.1490	0.0802	694.342	521.807	514.111	385.285
db5	5	0.1896	0.1322	704.077	864.383	508.053	690.963
coif5		0.1561	0.0971	528.106	555.186	396.066	420.658

By looking in Table 4.6, two different trends are observed in the relationship of C1 and C2. Firstly, for the sym5 wavelet models, C1 had shown better accuracy than C2, as suggested by the set of higher R² but lower RMSE and MAE values (0.9687, 69.059 m³/s, 49.501 m³/s) as compared to C2 (0.5158, 274.407 m³/s, 196.436 m³/s). It implies the univariate forecasting model outperforms the multivariate forecasting model in terms of predictive accuracy. Secondly, an opposite trend is observed for the db5 and coif5 models, with C2 having better performance metrics than C1, particularly higher R² but lower RMSE and MAE values. In contrast to the previous finding, this observation suggests that the multivariate forecasting model performs better. The disagreement of trend is likely a result of the exceptionally high accuracy of the S1*C1sym5-5 model, in contrast to the db5 and coif5 models. This extreme value becomes the outlier of the trend, consequently leading to the observation of an unexpected trend. Similarly, the superior performance of multivariate forecasting models over univariate forecasting models has been recorded in several studies, showing consistency with the second observed trend (Aboagye-Sarfo et al., 2015; Dyar et al., 2016).

The performances of univariate C1 and multivariate C2 are also compared in S2. According to the R^2 presented in Table 4.6, it is realized that C1 of all three wavelets are more accurate than C2, showing a complete dominance of univariate over multivariate forecasting in this scenario. Nevertheless, the disagreement of R^2 , RMSE, and MAE in the sym5 model is realized and worth noting. By calculation, C1 is 85.79% better but 33.06% weaker and 33.44% weaker in terms of R^2 , RMSE, and MAE. Therefore, the metric R^2 controls the determination of the model accuracy since the deviation of RMSE and MAE are relatively small and negligible compared to the R^2 value. In fact, the superior performance of the univariate model has also been recorded in several research (Rao, Sinha, and Basu, 2013; Anowar and Eluru, 2018).

The comparisons performed in this sub-section are an effort to reckon the improvement of multivariate over univariate streamflow forecasting. Aboagye-Sarfo et al. (2015) observed the improvement of multivariate time series modeling over conventional univariate models. The author credited the improvement to the capability of the multivariate model in identifying the relationship between the input variables. This finding is subsequently verified by many other researchers, for instance, Dyar et al. (2016) and Felix et al. (2018). However, the ability to handle multiple variables may occasionally become a drawback to the model's efficiency, as Rao, Sinha, and Basu (2013) discovered. The authors explained the lower prediction accuracy of the multivariate model with the higher complexity of the model optimization due to the larger input data set. It is in unison with the findings of Anowar and Eluru (2018), which despite the overall high accuracy, a marginal deficiency of the multivariate model had been recorded in some of the results. The authors mentioned no denial of the improved model fitness of the multivariate model at the expense of extra complexity resulting from the involvement of multiple variables. Hence, improved C2 performance was anticipated from this study. However, among the six combinations, four combinations between C1 and C2 indicated that C1 performed better. A plausible explanation for the observation is due to the better model optimization of C1, as mentioned by Rao, Sinha, and Basu (2013), which resulted in better accuracy of univariate C1 over multivariate C2 models. The performance of C1 and C2 models with different incorporating wavelets in

different scenarios is visualized in Figure 4.21, Figure 4.22, Figure 4.23, Figure 4.24, Figure 4.25, and Figure 4.26:

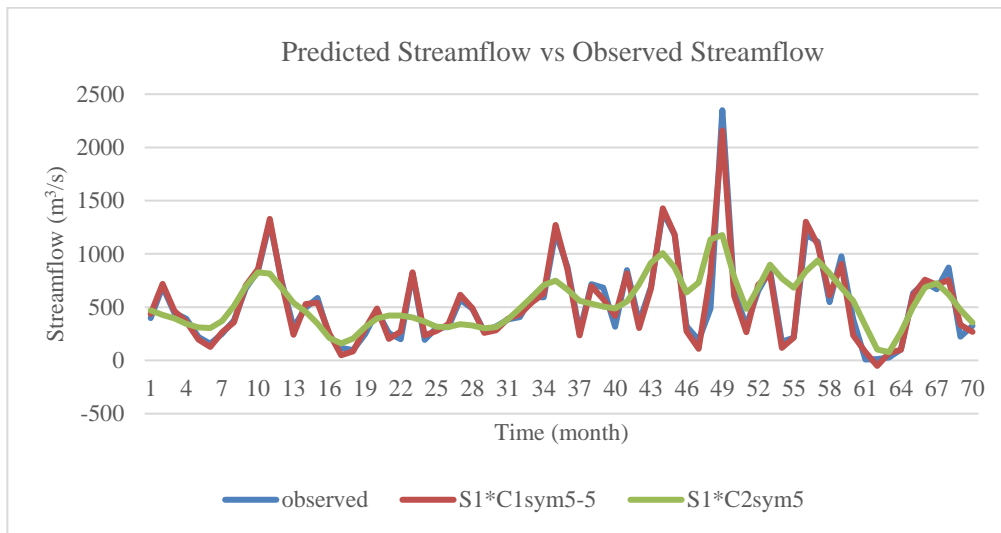


Figure 4.21: Performance of C1S1 and C2S1, with sym5 Wavelet

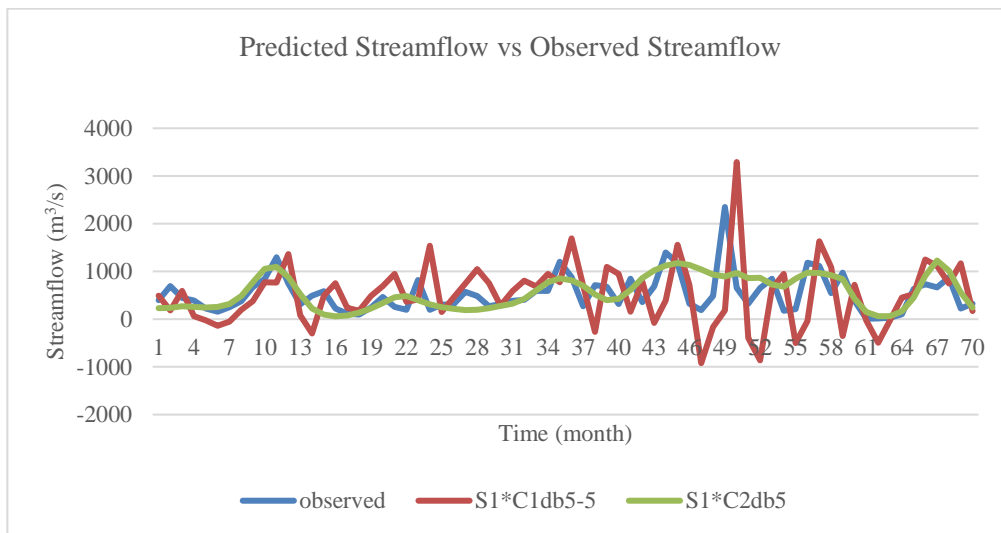


Figure 4.22: Performance of C1S1 and C2S1, with db5 Wavelet

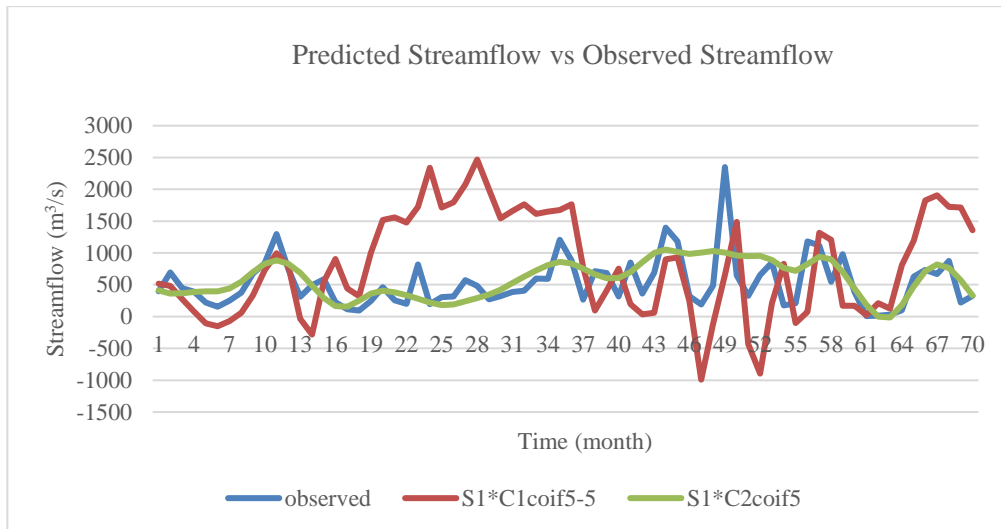


Figure 4.23: Performance of C1S1 and C2S1, with coif5 Wavelet

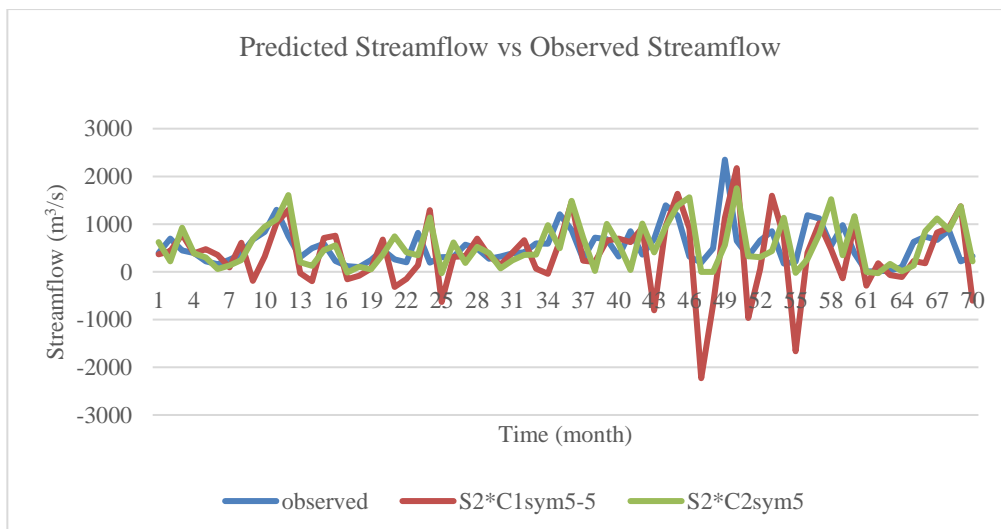


Figure 4.24: Performance of C1S2 and C2S2, with sym5 Wavelet

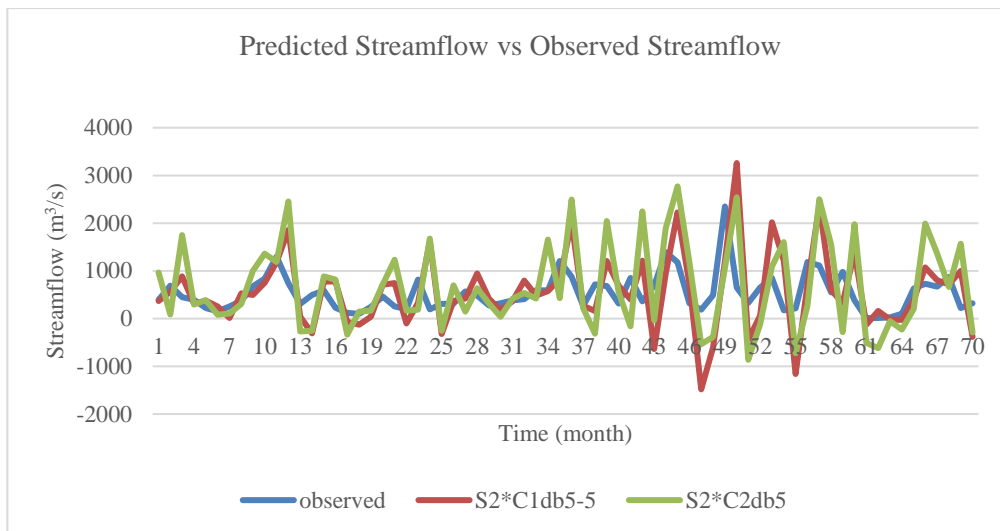


Figure 4.25: Performance of C1S2 and C2S2, with db5 Wavelet

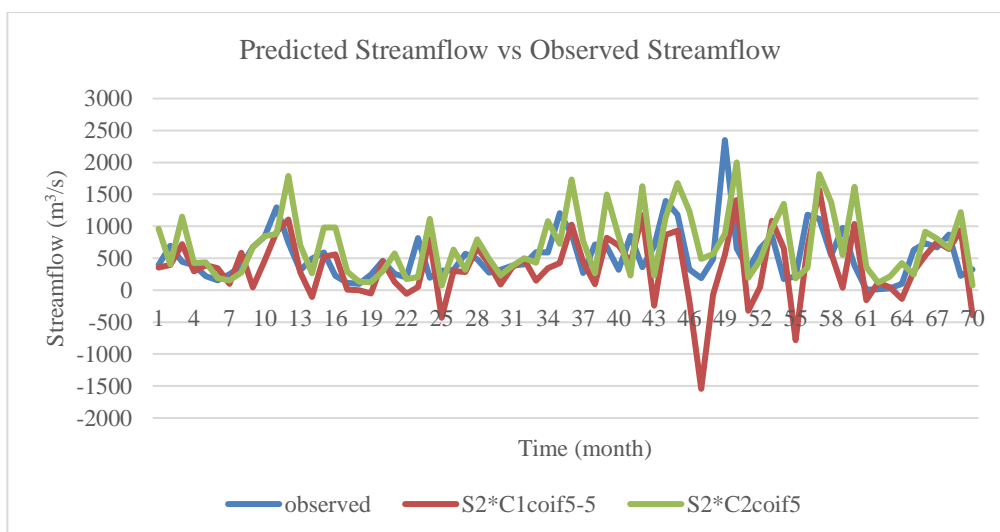


Figure 4.26: Performance of C1S2 and C2S2, with coif5 Wavelet

4.3.3 Comparison of Family Wavelet

Wavelet decomposition is incorporated in this study as a method of data pre-processing. Cheng et al. (2021) suggested that wavelet transformation relies on the suitability of the selected wavelet and the input time series to perform well. Therefore, the best-performing wavelet for any study is case-specific owing to the different characteristics of the input time series (Maheswaran and Khosa, 2012). In this particular study, three different wavelet families are used in this study to rank the suitability of each wavelet. The types of wavelets utilized in this study are shown in Table 4.8. The performances of each wavelet in both S1 and S2 are presented in Table 4.9 and Table 4.10, respectively:

Table 4.8: Type of Wavelet Families Incorporated (Patro and Kumar, 2016; Mandala et al., 2023; Santos et al., 2014)

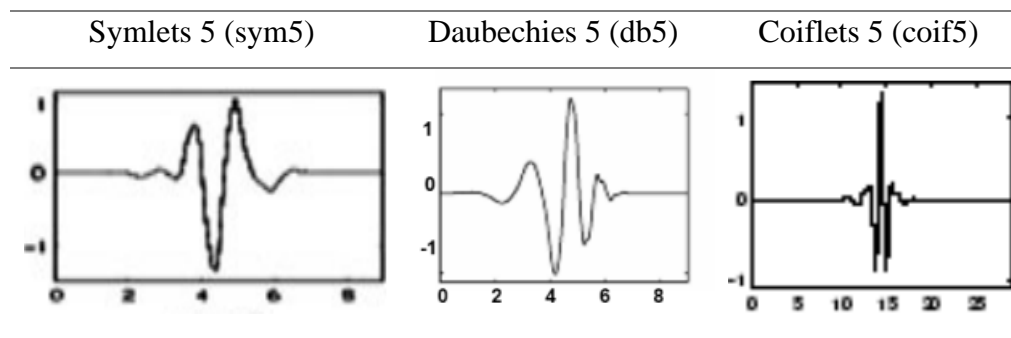


Table 4.9: Performance of Different Wavelets in C1 and C2 of S1

S1Case	LOD	Metrics								
		R ²			RMSE (m ³ /s)			MAE (m ³ /s)		
		sym5	db5	coif5	sym5	db5	coif5	sym5	db5	coif5
1	5	0.9687	0.0339	0.0115	69.059	698.138	890.479	49.501	517.083	727.268
2		0.5158	0.3421	0.3513	274.407	336.992	323.095	196.436	240.011	230.751

Table 4.10: Performance of Different Wavelets in C1 and C2 of S2

S2Case	LOD	Metrics								
		R ²			RMSE (m ³ /s)			MAE (m ³ /s)		
		sym5	db5	coif5	sym5	db5	coif5	sym5	db5	coif5
1	5	0.1490	0.1896	0.1561	694.342	704.077	528.106	514.111	508.053	396.066
2		0.0802	0.1322	0.0971	521.807	864.383	555.186	385.285	690.963	420.658

From Table 4.9, the best performance of sym5 amongst all three wavelets in S1 is recognized. Indeed, sym5 has the highest R^2 yet lowest RMSE and MAE in both C1 and C2. Therefore, the advantage of sym5 over db5 and coif5 in this scenario is certain. Meanwhile, the relationship between db5 and coif5 is ambiguous, considering the different observations from C1 and C2. The db5 model demonstrated higher accuracy in C1 but lower accuracy in C2 compared to the coif5 model in terms of all three metrics. As a matter of fact, the performances of db5 and coif5 are considerably similar, as suggested by the small difference in R^2 values. The difference in R^2 between these two wavelets is 0.0224 and 0.0092. In terms of RMSE and MAE, the differences range from 4.01% to 40.65% relative to the smaller value. Hence, the conclusion that db5 and coif5 have comparable performance is inferred.

For the wavelet performance in S2 presented in Table 4.10, the db5 models are found to be the best-performing models in both C1 and C2. However, this statement is only valid for the comparison of R^2 values. Despite achieving the highest R^2 value, the db5 models at the same time experience the highest RMSE and MAE. Since no similar trend is observed from the three metrics, a comparison of the three metrics was carried out to classify the performance of the models. Firstly, by comparison of R^2 , both cases deliver the same trend, where db5 has the highest accuracy, followed by coif5 and sym5 models. Secondly, by the comparison of RMSE, the model performance in descending order is coif5, sym5, db5 in C1 and sym5, coif5, db5 in C2. Thirdly, by the comparison of MAE, the model accuracy in descending order is coif5, db5, sym5 in C1 and sym5, coif5, db5 in C2. In fact, there is no apparent advantage of any wavelet over the others, owing to the relatively small difference of R^2 , RMSE, and MAE. Since no unanimous findings are observed, no conclusive statement is drawn on the best-performing wavelet in S2.

According to Rhif et al. (2022), the selection of a mother wavelet for high accuracy is dictated by two criteria, including the physical properties of the wavelet and the similarity between the wavelet and the input time series. It is reasonable as the improved performance of wavelet transformation over conventional Fourier transformation is attributed to the variety of wavelet shapes, which is helpful in localizing data and detecting data discontinuity (Chong, Lai, and El-Shafie, 2019). Thus, to identify the performance ranking of

sym5, db5, and coif5, the respective physical properties and resemblance to input time series are compared.

Rhif et al. (2022) outlined four properties of wavelet to be considered in discussing the accuracy of the mother wavelet, including orthogonality, compact support, symmetry, and vanishing moments. The application of an orthogonal wavelet is imperative to avoid loss of information during the reconstruction of decomposed wavelet. A compactly supported wavelet has better localization capacity in both the time and frequency domain; whereas a symmetrical wavelet is better at reducing non-linear phase delay, which in turn reduces the effect of border distortion in the signal. Wavelet with greater vanishing moments is better at processing the transient portions of a signal. (Upadhyaya et al., 2022). A summary of the discussed properties for the three incorporated wavelets is presented in Table 4.11:

Table 4.11: Summary of Wavelet Properties (Upadhyaya et al., 2022)

Properties	sym5	db5	coif5
Orthogonality	Yes	Yes	Yes
Compactly Supported	9	9	29
Orthogonal			
Symmetry	Near symmetric	Asymmetric	Near Symmetric
Vanishing Moments	5	5	10

Table 4.11 comparison reveals that the properties of the sym5 and db5 wavelets are similar, except for symmetry, where sym5 is more symmetric than db5. On the contrary, the orthogonal coif5 wavelet is relatively distinct. It is the most compactly supported and has the most vanishing moments. Anyhow, the comparison of wavelet properties alone is not conclusive on the performance of the mother wavelet; the wavelet's similarity to the input time series also influences performance. The input time series is illustrated in Figure 4.27 and compared to the mother wavelets shown in Table 4.8:

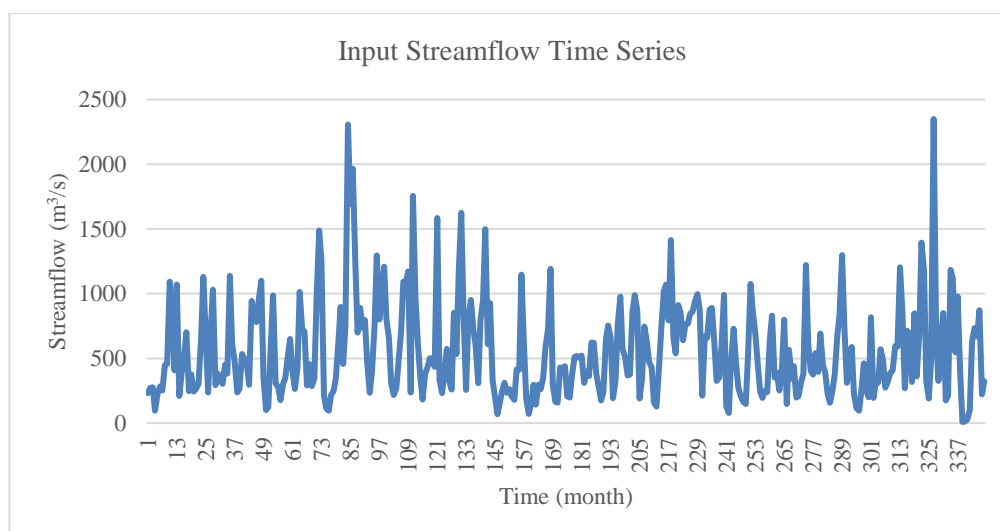


Figure 4.27: Input Streamflow Time Series

In contrast, to the similarity between the mother wavelet and the input time series, two qualitative and quantitative methods have been proposed (Moradi, 2022). The qualitative method is accomplished by visual inspection of the similarities between the time series and mother wavelet. However, due to the similarities between the three wavelets selected, visual inspection is bound to be inaccurate. Therefore, the performance ranking of the wavelets cannot be determined unless a quantitative analysis of the resemblance is performed. Anyhow, no relevant quantitative analysis is involved in the scope of this study; thus, the expected performance ranking of different wavelet families remains vague and is worth further investigation.

4.4 Uncertainty Assessment

In order to select the best-performing model from the 12 investigated models by considering multiple metrics such as the R^2 , standard deviation, and the distribution of the data, two visualization tools, namely the Taylor diagram and the Violin diagram, were employed. By plotting these two diagrams, visual aids are provided to assist in the determination of the best-performing model. The plotted Taylor diagram is presented in Figure 4.28:

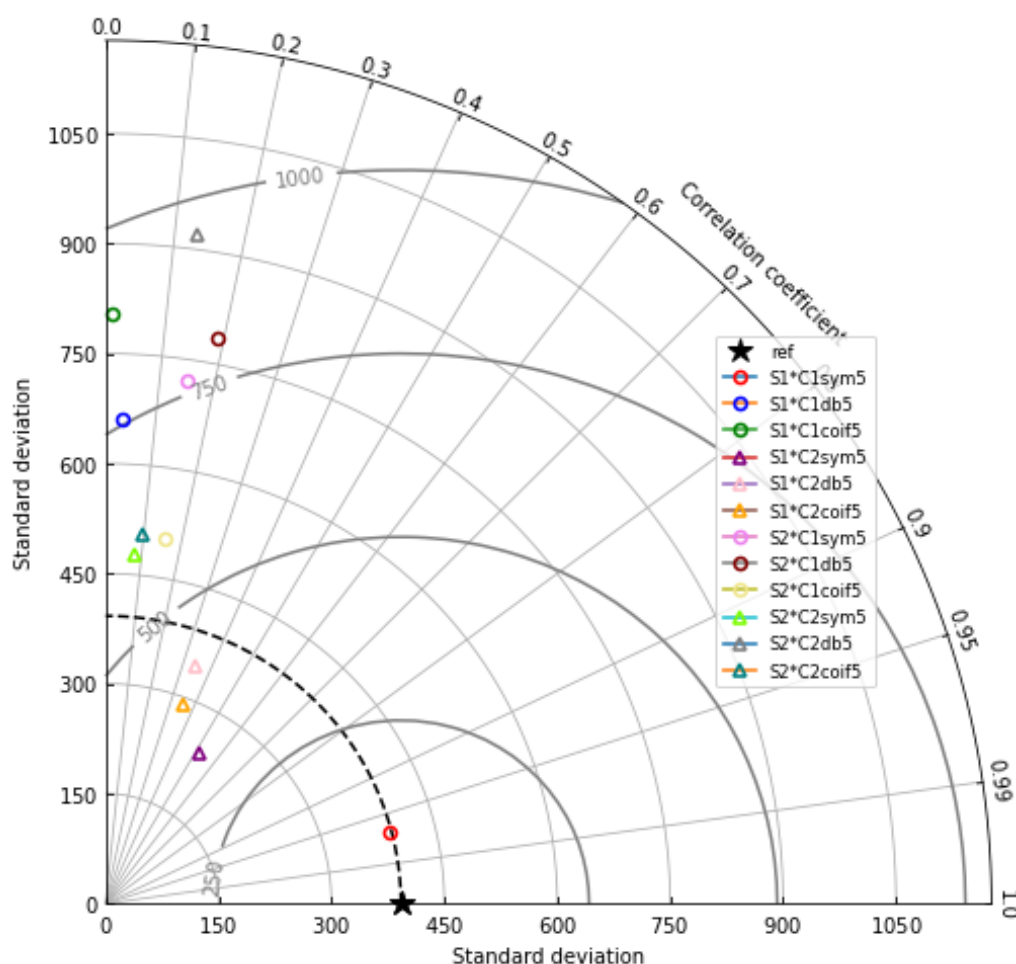


Figure 4.28: Taylor Diagram of the Observed and 12 Studied Models

A Taylor diagram is utilized to assess the resemblance of the predicted model to the observed model by taking into consideration the standard deviation and the coefficient of determination (R^2) values of the predicted model relative to the observed model. Therefore, it should be noted that the model's performance metrics should not be interpreted by the model's respective abscissa and ordinate. Instead, the metrics should be interpreted concerning the reference model. From Figure 4.28, the model with the highest resemblance to the reference model is S1*C1sym5. In other words, the predicted streamflow of the tuned S1C1 model incorporating sym5 wavelet has the most similar standard deviation and R^2 with the observed streamflow amongst all the studied models. Another informative observation from Figure 4.28 includes the worst models among S1*C1coif5 and S2*C2db5, with the lowest R^2 and the largest difference in standard deviation, respectively. The C2 models (triangular mark) are more consistent than the C1 models (circular mark). This is inferred from the

distribution of the models, where five out of six C2 models fall closely in the average zone; whilst the distribution of C1 models is relatively scattered. Before concluding the best-performing model, the comparison between predicted streamflow distribution and observed streamflow distribution is also necessary, to provide convincing evidence of the effectiveness of the model. To do so, a Violin diagram was plotted and presented in Figure 4.29:

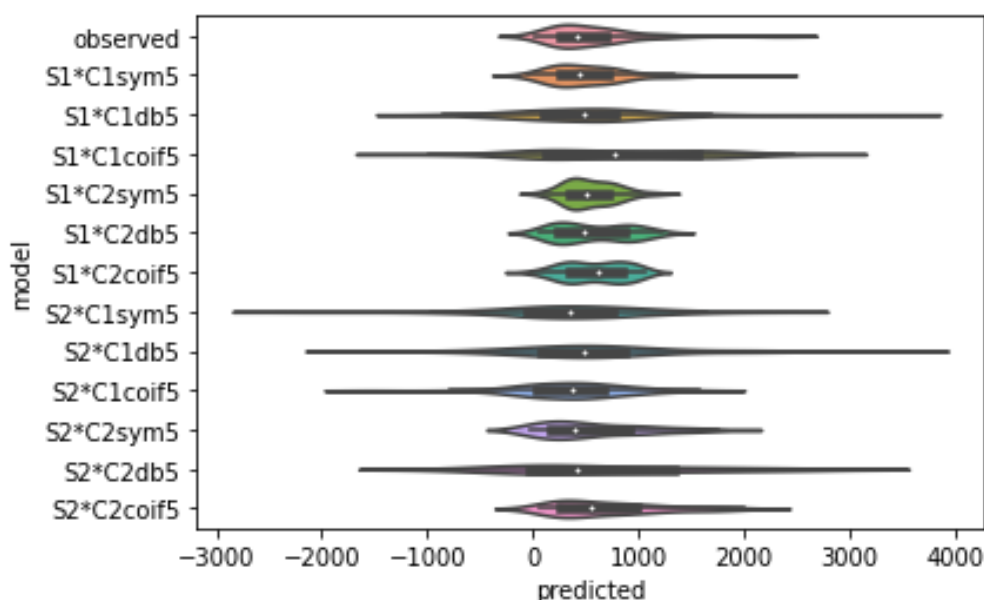


Figure 4.29: Violin Diagram of the Observed and 12 Studied Models

The Violin diagram in Figure 4.29 illustrates the distribution of the observed and predicted dataset. Several characteristics of the violin were compared to determine the most similar distribution of the dataset. First, the shape of the violin, which is essentially the distribution of the data was compared. Second, the box within the violin, representing the interquartile range or half number of the data set was observed. Lastly, the extent of the violin graph was contrasted to identify the difference between the range of datasets amongst models. Considering these three characteristics, the model with the highest similarity of dataset distribution is the S1*C1sym5 model, which is in unison with the previous finding from the Taylor diagram. Also, it is interesting to notice that except for S2*C2db5, C2 models generally have better resemblance in data distribution, in contrast with C1 models with a large range of data fluctuation. Hence, considering the unanimous findings from Taylor and Violin

diagram, as well as the justification from previous metrics comparison, the model S1*C1sym5 is identified as the best-performing model in this study.

4.5 Performance of Short, Medium and Long Term Forecasting

After the identification of the most accurate model in the study, the model (S1*C1sym5) is utilized to investigate the performance of the hybrid ANN model in predicting streamflow in different time frames. The forecasting horizons of 1 month, 6 months, and 12 months were selected to represent the short, medium, and long-term predictions, respectively. All previously discussed influential factors, such as the tuning performance, scenarios, cases, and incorporating wavelet, were made constant in this analysis using the selected best model to make predictions for all three forecasting horizons. The performances of the S1*C1sym5 model in different forecasting horizons are demonstrated in Table 4.12:

Table 4.12: Performance of S1*C1sym5 in Short, Medium and Long Term Forecasting

Term	Metrics		
	R ²	RMSE (m ³ /s)	MAE (m ³ /s)
Short	0.9687	69.059	49.501
Medium	0.0467	1397.316	1074.429
Long	0.0333	3392.118	1990.130

Through observation of Table 4.12, it is realized that the R² of the model had reduced significantly from 0.9687 to 0.0467 and further to 0.0333 for short, medium, and long-term forecasting. By setting short-term forecasting as the performance benchmark, the performance reduction in terms of R² is 95.18% and 95.56%. In other words, the model loses more than 95% accuracy during its application in medium and long-term prediction. This apparent reduction in R² is suggesting a sharp drop in the performance of the model in the longer prediction horizon. The model also experienced a soaring increment of RMSE, from 69.059 m³/s to 1397.316 m³/s and ultimately to 3392.118 m³/s, which are equivalent to over 1923.37% and 4811.91% of increment. Likewise, the MAE of the model had increased from 49.501 m³/s to 1074.429 m³/s and

1990.130 m³/s, resulting in a percentage increment of over 2070.52% and 3920.38%, relative to the original short-term forecasting. The findings of R², RMSE, and MAE provide firm evidence to infer that the performance of the S1*C1sym5 model in this study reduces significantly from short to medium and long-term prediction. A possible explanation for the performance reduction is the increased complexity of handling the data over a longer time horizon, which results in increased difficulty in identifying the hidden patterns of the input data by ANN. The performances of short, medium, and long-term forecasting are illustrated in Figure 4.30:

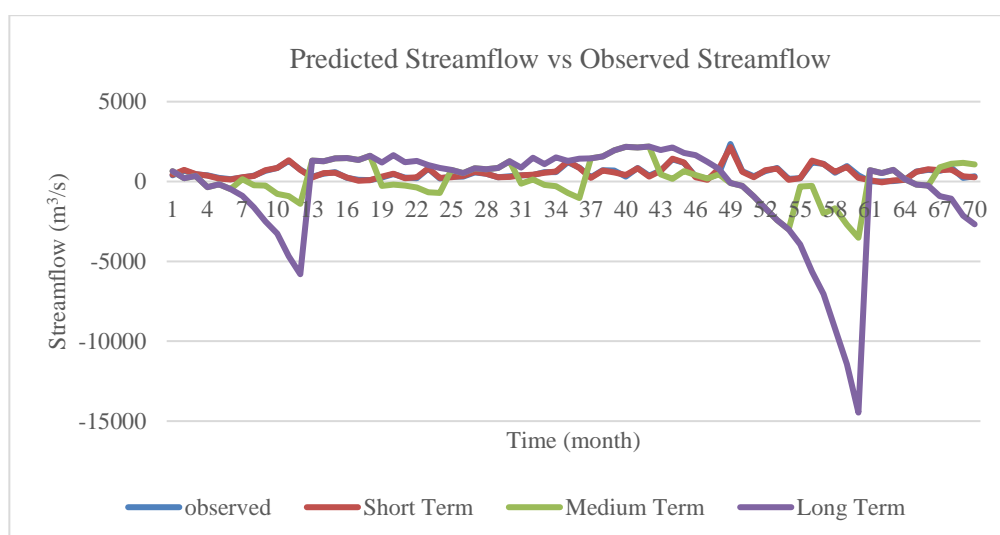


Figure 4.30: Performance of S1*C1sym5 in Different Horizons

Based on Figure 4.30, it is noticed that despite the near-perfect resemblance of the short-term prediction to the observed streamflow, the performance of its corresponding medium and long-term prediction is unsatisfactory. However, some similarities between the medium and long-term predictions are observed. For instance, the overall shape and occurrence of local minima share a certain degree of similarity.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

As a conclusive summary of the research, the wavelet transformation was thoroughly reviewed and incorporated in a hybrid ANN model to forecast streamflow over the short, medium, and long-term horizons. In total, 12 different models by the combination of 2 scenarios, 2 cases and 3 wavelets were investigated and examined.

The hyperparameter tuning performance of the model was first investigated by contrasting the tuned and untuned versions of a model. The results validate the advantageous performance of the tuned model, justifying the application of the tuned model in subsequent comparisons. Next, the comparison of standalone and hybrid ANN models was conducted to emphasize the improvement of wavelet incorporation. Performance enhancement in terms of R^2 of 1082.78% and 1612.64% were recorded for the S1 and S2 scenarios (data pre-processing conditions), respectively, by incorporating sym5 wavelet into the model.

The performance of S1 and S2 were compared for all three wavelets in both cases. In C1 (univariate model), S1 models were found to be more accurate when incorporated with sym5 wavelet but less accurate with db5 and coif5 wavelet. In C2 (multivariate), S1 models of all wavelets were found to have better performance than the corresponding S2 models. The percentages of R^2 improvement, relative to the values of S2, are 543.14%, 158.77%, and 261.79% for sym5, db5, and coif5, respectively. Four out of six comparisons between S1 and S2 have suggested the superiority of S1. A similar approach of comparison was utilized to compare the performance of C1 and C2 models for all three wavelets in both S1 and S2 scenarios. Similarly, a majority of four out of six comparisons have confirmed the better performance of univariate C1 over multivariate C2 in this study.

The three wavelets studied in this research were compared for all four scenario-case (S1C1, S1C2, S2C1, S2C2) models. However, no apparent

advantage of any wavelet prevailed in this study. The metrics of all 12 models were plotted into a Taylor diagram and also with a Violin diagram to facilitate the determination of the best-performing model. With the aid of these two diagrams, the best-performing model was identified as the S1*C1sym5. This model was subsequently applied to evaluate the performance of short, medium, and long-term forecasting. The results for 1-month, 6-month, and 12-month forecasting horizons revealed the best performance of short-term forecasting. Significant reduction of metrics for medium and long-term forecasting was observed as (95.18%, 1923.37%, 2070.52%) and (95.56% , 4811.91% , 3920.38%) in the form of (R^2 , RMSE, MAE). Summarizing, it is rational to declare the achievement of the aim and objectives of the research.

5.2 Recommendations for future work

The findings of this research have limited coverage on the topic of wavelet decomposition and hybrid ANN models. Therefore, several recommendations are outlined below to address the encountered limitations and the direction of future research:

- i. Application and comparison of different ANN learning algorithms, for instance, grid search and random search.
- ii. Implementation of quantitative analysis to evaluate the resemblance of wavelet to the input time series.
- iii. Performance evaluation of the combination of decomposed wavelets from multiple wavelet families.
- iv. Direct comparison of DWT and SWT to verify the theoretical advantage of SWT.

REFERENCES

- Aboagye-Sarfo, P. et al., 2015. A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in Western Australia. *Journal of Biomedical Informatics*, 57, pp.62–73.
- Adamowski, J. and Sun, K., 2010. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, 390(1-2), pp.85–91.
- Al-qaness, M.A.A. et al., 2022. Boosted ANFIS model using augmented marine predator algorithm with mutation operators for wind power forecasting. *Applied Energy*, 314, p.118851.
- Alquraish, M.M. and Khadr, M., 2021. Remote-Sensing-Based Streamflow Forecasting Using Artificial Neural Network and Support Vector Machine Models. *Remote Sensing*, 13(20), p.4147.
- Anowar, S. and Eluru, N., 2018. Univariate or multivariate analysis for better prediction accuracy? A case study of heterogeneity in vehicle ownership. *Transportmetrica A: Transport Science*, 14(8), pp.635–668.
- Babu, C.N. and Reddy, B.E., 2014. A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. *Applied Soft Computing*, 23, pp.27–38.
- Bilgili, F. et al., 2021. The role of hydropower energy in the level of CO₂ emissions: An application of continuous wavelet transform. *Renewable Energy*, 178, pp.283–294.
- Chae, Y.T., Horesh, R., Hwang, Y. and Lee, Y.M., 2016. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy and Buildings*, 111, pp.184–194.
- Cheng, V.Y.S. et al., 2021. Effects of hydrological forcing on short- and long-term water level fluctuations in Lake Huron-Michigan: A continuous wavelet analysis. *Journal of Hydrology*, 603, p.127164.

Chong, K.L. et al., 2022. Spatiotemporal variability analysis of standardized precipitation indexed droughts using wavelet transform. *Journal of Hydrology*, 605, p.127299.

Chong, K.L., Lai, S.H. and El-Shafie, A., 2019. Wavelet Transform Based Method for River Stream Flow Time Series Frequency Analysis and Assessment in Tropical Environment. *Water Resources Management*, 33(6), pp.2015–2032.

Cobaner, M., Seckin, G., Seckin, N. and Yurtal, R., 2010. Boundary shear stress analysis in smooth rectangular channels and ducts using neural networks. *Water and Environment Journal*, 24(2), pp.133–139.

Dyar, M.D. et al., 2016. Comparison of univariate and multivariate models for prediction of major and minor elements from laser-induced breakdown spectra with and without masking. *Spectrochimica Acta Part B*, 93(104).

Falamarzi, Y., Palizdan, N., Huang, Y.F. and Lee, T.S., 2014. Estimating evapotranspiration from temperature and wind speed data using artificial and wavelet neural networks (WNNs). *Agricultural Water Management*, 140, pp.26–36.

Fashae, O.A., Olusola, A.O., Ndubuisi, I. and Udombos, C.G., 2018. Comparing ANN and ARIMA model in predicting the discharge of River Opeki from 2010 to 2020. *River Research and Applications*, 35(2), pp.169–177.

Fathian, F., Mehdizadeh, S., Kozekalani Sales, A. and Safari, M.J.S., 2019. Hybrid models to improve the monthly river flow prediction: Integrating artificial intelligence and non-linear time series models. *Journal of Hydrology*, 575, pp.1200–1213.

Felix, L.B. et al., 2018. Comparison of univariate and multivariate magnitude-squared coherences in the detection of human 40-Hz auditory steady-state evoked responses. *Biomedical Signal Processing and Control*, 40, pp.234–239.

Feng, Z. et al., 2022. Hydrological time series forecasting via signal decomposition and twin support vector machine using cooperation search algorithm for parameter identification. *Journal of Hydrology*, 612, p.128213.

Ghenai, C. et al., 2022. Short-term building electrical load forecasting using adaptive neuro-fuzzy inference system (ANFIS). *Journal of Building Engineering*, 52, p.104323.

Gholami, A., Bonakdari, H., Ebtehaj, I. and Akhtari, A.A., 2017. Design of an adaptive neuro-fuzzy computing technique for predicting flow variables in a 90° sharp bend. *Journal of Hydroinformatics*, 19(4), pp.572–585.

Gui, H., Wu, Z. and Zhang, C., 2021. Comparative Study of Different Types of Hydrological Models Applied to Hydrological Simulation. *CLEAN – Soil, Air, Water*, 49(8), p.2000381.

IBM Cloud Education, 2020, *What are Neural Networks?* [Online]. Available at: <https://www.ibm.com/cloud/learn/neural-networks> [Accessed: 8 August 2022].

Kambalimath, S. and Deka, P.C., 2021. Performance enhancement of SVM model using discrete wavelet transform for daily streamflow forecasting. *Environmental Earth Sciences*, 80(3).

Kaur, J. and Mahajan, M., 2015. Hybrid of Fuzzy Logic and Random Walker Method for Medical Image Segmentation. *International Journal of Image, Graphics and Signal Processing*, 7(2), pp.23–29.

Khan, Md.M.H., Muhammad, N.S. and El-Shafie, A., 2020. Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting. *Journal of Hydrology*, 590, p.125380.

Khazaei Poul, A., Shourian, M. and Ebrahimi, H., 2019. A Comparative Study of MLR, KNN, ANN and ANFIS Models with Wavelet Transform in Monthly Stream Flow Prediction. *Water Resources Management*, 33(8), pp.2907–2923.

Kişi, Ö., 2007. Streamflow Forecasting Using Different Artificial Neural Network Algorithms. *Journal of Hydrologic Engineering*, 12(5), pp.532–539.

Kumar, A. et al., 2020. Stationary wavelet transform based ECG signal denoising method. *ISA Transactions*, 114.

Laha, S.K., Swarnakar, B., Kansabanik, S. and Ray, S., 2022. A novel signal denoising method using stationary wavelet transform and particle swarm optimization with application to rolling element bearing fault diagnosis. *Materials Today: Proceedings*.

Leal, M.M., Costa, F.B. and Loureiro Sousa Campos, J.T., 2019. Improved traditional directional protection by using the stationary wavelet transform. *International Journal of Electrical Power & Energy Systems*, 105, pp.59–69.

Li, F., Wang, Z. and Qiu, J., 2019. Long-term streamflow forecasting using artificial neural network based on preprocessing technique. *Journal of Forecasting*, 38(3), pp.192–206.

Liu, Y. et al., 2022. GNSS-derived PWV and meteorological data for short-term rainfall forecast based on support vector machine. *Advances in Space Research*, 70(4), pp.992–1003.

Loh, W.S. et al., 2021. Application of Machine Learning Model for the Prediction of Settling Velocity of Fine Sediments. *Mathematics*, 9(23), p.3141.

Maheswaran, R. and Khosa, R., 2012. Comparative study of different wavelets for hydrologic forecasting. *Computers & Geosciences*, 46, pp.284–295.

Mandala, S. et al., 2023. The Effects of Daubechies Wavelet Basis Function (DWBF) and Decomposition Level on the Performance of Artificial Intelligence-Based Atrial Fibrillation (AF) Detection Based on Electrocardiogram (ECG) Signals. *Applied Sciences*, 13(5), p.3036.

Moradi, M., 2022. Wavelet transform approach for denoising and decomposition of satellite-derived ocean color time-series: Selection of optimal mother wavelet. *Advances in Space Research*, 69, pp.2724–2744.

Nourani, V., Komasi, M. and Mano, A., 2009. A Multivariate ANN-Wavelet Approach for Rainfall–Runoff Modeling. *Water Resources Management*, 23(14), pp.2877–2894.

Nury, A.H., Hasan, K. and Alam, Md.J.B., 2017. Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in northeastern Bangladesh. *Journal of King Saud University - Science*, 29(1), pp.47–61.

Olyaie, E., Zare Abyaneh, H. and Danandeh Mehr, A., 2017. A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River. *Geoscience Frontiers*, 8(3), pp.517–527.

Pandey, S., Kesharwani, S. and Singh, D.K., 2015. Discrimination of Power Quality Disturbances Using Wavelet Transform. *International Journal of Science, Engineering and Technology Research*, 4(7).

Patro, K.K. and Kumar, P.R., 2016. A Novel Frequency-Time Based Approach for the Detection of Characteristic Waves in Electrocardiogram Signal. *Microelectronics, Electromagnetics and Telecommunications*.

Quilty, J. and Adamowski, J., 2021. A maximal overlap discrete wavelet packet transform integrated approach for rainfall forecasting – A case study in the Awash River Basin (Ethiopia). *Environmental Modelling & Software*, 144, p.105119.

Quilty, J. and Adamowski, J., 2018. Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. *Journal of Hydrology*, 563, pp.336–353.

Raghavendra. N, S. and Deka, P.C., 2014. Support vector machine applications in the field of hydrology: A review. *Applied Soft Computing*, 19, pp.372–386.

Rao, A.D., Sinha, M. and Basu, S., 2013. Bay of Bengal wave forecast based on genetic algorithm: A comparison of univariate and multivariate approaches. *Applied Mathematical Modelling*, 37(6), pp.4232–4244.

Rhif, M. et al., 2022. Optimal selection of wavelet transform parameters for spatio-temporal analysis based on non-stationary NDVI MODIS time series in Mediterranean region. *ISPRS Journal of Photogrammetry and Remote Sensing*, 193, pp.216–233.

Samantaray, S., Sawan Das, S., Sahoo, A. and Prakash Satapathy, D., 2022. Monthly runoff prediction at Baitarani river basin by support vector machine based on Salp swarm algorithm. *Ain Shams Engineering Journal*, 13(5), p.101732.

Santos, C.A.G., Freire, P.K.M.M., Silva, G.B.L. and Silva, R.M., 2014. Discrete wavelet transform coupled with ANN for daily discharge forecasting into Três Marias reservoir. *Proceedings of the International Association of Hydrological Sciences*, 364, pp.100–105.

Sharma, P., Singh, S. and Sharma, S.D., 2021. Artificial Neural Network Approach for Hydrologic River Flow Time Series Forecasting. *Agricultural Research*.

Sharma, V., Yang, D., Walsh, W. and Reindl, T., 2016. Short term solar irradiance forecasting using a mixed wavelet neural network. *Renewable Energy*, 90, pp.481–492.

Tayyab, M., Zhou, J., Adnan, R. and Zeng, X., 2017. Application of Artificial Intelligence Method Coupled with Discrete Wavelet Transform Method. *Procedia Computer Science*, 107, pp.212–217.

Toğa, G., Atalay, B. and Toksari, M.D., 2021. COVID-19 prevalence forecasting using Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN): Case of Turkey. *Journal of Infection and Public Health*, 14(7), pp.811–816.

Upadhya, M. et al., 2022. Mother wavelet selection method for voltage sag characterization and detection. *Electric Power Systems Research*, 211(108246).

Wang, G., Guo, L. and Duan, H., 2013. Wavelet Neural Network Using Multiple Wavelet Functions in Target Threat Assessment. *The Scientific World Journal*, 2013, pp.1–7.

Wang, W., Jin, J. and Li, Y., 2009. Prediction of Inflow at Three Gorges Dam in Yangtze River with Wavelet Network Model. *Water Resources Management*, 23(13), pp.2791–2803.

Xu, C., Chen, X. and Zhang, L., 2021. Predicting river dissolved oxygen time series based on stand-alone models and hybrid wavelet-based models. *Journal of Environmental Management*, 295, p.113085.

Yan, Y., Li, J. and Gao, D., 2014. Condition Parameter Modeling for Anomaly Detection in Wind Turbines. *Energies*, 7(5), pp.3104–3120. Available at: <https://www.mdpi.com/1996-1073/7/5/3104/htm> [Accessed: 17 April 2019].

Yaseen, Z.M. et al., 2017. Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model. *Journal of Hydrology*, 554, pp.263–276.

Yin, G. et al., 2022. A support vector machine-based method for improving real-time hourly precipitation forecast in Japan. *Journal of Hydrology*, 612, p.128125.

Zhang, N., Ma, Y. and Zhang, Q., 2018. Prediction of sea ice evolution in Liaodong Bay based on a back-propagation neural network model. *Cold Regions Science and Technology*, 145, pp.65–75.

Zhang, S.-Z., Chen, S. and Jiang, H., 2022. A back propagation neural network model for accurately predicting the removal efficiency of ammonia nitrogen in wastewater treatment plants using different biological processes. *Water Research*, 222, p.118908.

Zhang, Z., Zhang, Q. and Singh, V.P., 2018. Univariate streamflow forecasting using commonly used data-driven models: literature review and case study. *Hydrological Sciences Journal*, 63(7), pp.1091–1111.

Zhao, J., Tian, G., Qiu, Y. and Qu, H., 2021. Rapid quantification of active pharmaceutical ingredient for sugar-free Yangwei granules in commercial production using FT-NIR spectroscopy based on machine learning techniques. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 245, p.118878.

Zhou, F., Liu, B. and Duan, K., 2020. Coupling wavelet transform and artificial neural network for forecasting estuarine salinity. *Journal of Hydrology*, 588, p.125127.

Zhou, Y., Guo, S. and Chang, F.-J., 2019. Explore an evolutionary recurrent ANFIS for modelling multi-step-ahead flood forecasts. *Journal of Hydrology*, 570, pp.343–355.