ROBUST DATA FUSION TECHNIQUES INTEGRATED MACHINE LEARNING MODELS FOR ESTIMATING REFERENCE EVAPOTRANSPIRATION

CHIA MIN YAN

DOCTOR OF PHILOSOPHY (ENGINEERING)

LEE KONG CHIAN
FACULTY OF ENGINEERING AND SCIENCE
UNIVERSITI TUNKU ABDUL RAHMAN
MAY 2022

ROBUST DATA FUSION TECHNIQUES INTEGRATED MACHINE LEARNING MODELS FOR ESTIMATING REFERENCE EVAPOTRANSPIRATION

By

CHIA MIN YAN

A thesis submitted to the Department of Civil Engineering,
Lee Kong Chian Faculty of Engineering and Science,
Universiti Tunku Abdul Rahman,
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy (Engineering)
May 2022

ABSTRACT

**ROBUST DATA FUSION TECHNIQUES INTEGRATED MACHINE
LEARNING MODELS FOR ESTIMATING REFERENCE
EVAPOTRANSPIRATION**

Chia Min Yan

Evapotranspiration (ET) is one of the most important hydrological processes as it has

prominent effects on the environment's energy balance and water budget. Accurate

estimation of the ET is vital for many national-level decisions making processes,

including water resources allocation, irrigation scheduling as well as crop management.

To date, many related hydrological works still endorse the Penman-Monteith (PM)

model as the standard for the computation of the reference evapotranspiration (ET_0) as

per the recommendation by the United Nations Food and Agricultural Organisation.

ET_0 is a value which can be converted to the actual crop ET (ET_c) with the inclusion

of a crop-dependent factor. However, despite the PM model being accepted as a

universal method for determining the ET_0, this method is often criticised due to the

high number of meteorological variables needed. Thus, many researchers had resorted

to the utilisation of machine learning models to overcome this pitfall of the PM model.

Nonetheless, based on the literature review performed, machine learning models are

data-hungry in nature, which increases the difficulty of training a model from scratch.

The data hunger of machine learning models can be classified into two categories,

namely the qualitative hunger (where machine learning models need for various

features for training) and quantitative hunger (need for a vast amount of historical data for training). This forms the major gap in the research field. The works presented in the thesis strive to solve the data hunger of machine learning models through the integration of data fusion techniques, with a minimalistic approach by using simple yet robust models. Besides, two scenarios (Scenario 2 and Scenario 3) were designed to evaluate the spatial robustness of the developed models, so that the local data dependency can be discounted. This study was performed at 12 meteorological stations, using meteorological data dated from 1st January 2000 to 31st December 2019, and which are distributed across Peninsular Malaysia, whereby about 19.7 % of its land is covered by oil palm plantations (a major contributor to the country's agricultural output). The multilayer perceptron (MLP), the support vector machine (SVM) and the adaptive neuro-fuzzy inference system (ANFIS) were used as the base models for obtaining optimum input combinations as well as benchmark performances at each of the stations. Three different data fusion techniques were investigated, including the data centric bootstrap aggregating, the model centric Bayesian modelling approach and the black-box based non-linear neural ensemble (NNE). Observations of the results of this study revealed that the solar radiation (R_s) is the most essential variable for estimating ET_0 in Peninsular Malaysia. The accuracy of the estimations using the MLP, SVM and ANFIS could be improved by the inclusion of different complementary variables, which vary depending on the geographical characteristics at the meteorological stations. The bootstrap aggregating failed in enhancing the performance of the MLP, SVM and ANFIS. The size of the dataset overwhelmed the problem's dimensionality, thus rendering the bootstrap aggregating to be ineffective. The Bayesian model averaging (BMA) enhanced the estimation of the ensembles of

the base MLP, SVM and ANFIS. This was done through the Bayesian weight assignments to combine the favourable traits of the individual models. However, the BMA algorithm was found to be rigid as it was results-oriented and might opt to omit some base models if their performance were significantly poorer than the others. This happened when the number of input meteorological variables was high, and the BMA was converted to the Bayesian model selection (BMS). Nevertheless, when the number of input meteorological variables was low, the BMA based ensemble (BMA-E) produced satisfactory performance. As for the NNE, a novel meta-learner based on the stochastic-enabled extreme learning machine integrated with whale optimisation algorithm (WOA-ELM) was developed and used in such an application for the first time. The results showed that the WOA-ELM based ensemble (WOA-ELM-E) improved the performance of the base models in general. This was attributed to the flexibility of its structure and its ability to "look" at the target value once more during its training phase. The WOA-ELM-E was found to be the best model at most of the meteorological stations. Furthermore, when the best local models were tested at external stations (Scenario 2), only the WOA-ELM-E could produce estimations with satisfactory accuracy. The best models of other variants such as the BMA-E, MLP and ANFIS could only produce acceptable accuracy if they were applied in regions with similar geographical characteristics. In other words, the WOA-ELM-E can be said to have good spatial robustness, especially the one trained at Station 48620 (Sitiawan). This, in turn, could nullify the need for local model development and local data collection, consequently overcoming the qualitative and quantitative hungers of the classical machine learning models. Another scenario (Scenario 3) was designed to study the usefulness of globally pooled data in enhancing the spatial robustness of the

WOA-ELM-E. The results showed that such an approach produced a hybrid model which had similar robustness as the one trained at Station 48620 (Sitiawan) and was considered to be effective. In conclusion, the output of the research works reported in this thesis includes the approach for developing a one-for-all model to estimate the ET_0 across Peninsular Malaysia accurately. This can be regarded as the major contribution as it could possibly eliminate the process of local data collection for the development or calibration of a local ET_0 estimating model. Subsequently, the proposal and implementation of water resources-related policies can be accelerated to improve the social welfare at a national level.

ACKNOWLEDGEMENTS

I would like to express my gratitude to the Universiti Tunku Abdul Rahman (UTAR) for providing me with this research opportunity. The financial and technical supports given are much appreciated. Millions thanks to my supervisor, Ir. Dr. Huang Yuk Feng, and my co-supervisor, Ir. Dr. Koo Chai Hoon for their guidance along the journey of this research. Without them going through the ups and downs with me, I probably would not have made it to the end. They were committed and dedicated persons. I will treasure our numerous fruitful discussions and meetings which gave me a lot of insights into the research world.

Next, I would like to thank Mr. Fung Kit Fai for being generous in sharing his knowledge and experience with me. As a novice in the field of the engineering application of artificial intelligence and machine learning, his sharing diverted me away from many unnecessary detours. To all the friends who had lent me their helping hands, I sincerely thank you and apologise if I cannot recall each one of you.

Finally, I am grateful to my parents and family that have given unconditional support to me throughout the three years of study. Words cannot express how important they are to me in the furtherance of my study without a hitch.

LEE KONG CHIAN FACULTY OF ENGINEERING AND SCIENCE

UNIVERSITI TUNKU ABDUL RAHMAN

Date: ___16/5/2022_____

SUBMISSION OF THESIS

It is hereby certified that **CHIA MIN YAN** (ID No: **19UED05172**) has completed this thesis entitled "ROBUST DATA FUSION TECHNIQUES INTEGRATED MACHINE LEARNING MODELS FOR ESTIMATING REFERENCE EVAPOTRANSPIRATION" under the supervision of Ir. Dr. Huang Yuk Feng (Supervisor) from the Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, and Ir. Dr. Koo Chai Hoon (Co-Supervisor) from the Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science.

I understand that University will upload softcopy of my thesis in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

_____

(Chia Min Yan)

APPROVAL SHEET

This dissertation/thesis entitled "**ROBUST DATA FUSION TECHNIQUES INTEGRATED MACHINE LEARNING MODELS FOR ESTIMATING REFERENCE EVAPOTRANSPIRATION**" was prepared by CHIA MIN YAN and submitted as partial fulfilment of the requirements for the degree of Doctor of Philosophy (Engineering) at Universiti Tunku Abdul Rahman.

Approved by:

_____

(Ir. Dr. Huang Yuk Feng)
Date:…16/5/2022……..
Associate Professor/Supervisor
Department of Civil Engineering
Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

_____

(Ir. Dr. Koo Chai Hoon)
Date:…16/5/2022……..
Associate Professor/Co-supervisor
Department of Civil Engineering
Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

DECLARATION

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

Name _____CHIA MIN YAN_____

Date _____16/5/2022_____

TABLE OF CONTENTS

Page

CHAPTER

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS / ABBREVIATIONS

A-A	Advection-Aridity
ABC	Artificial Bee Colony
ACO	Ant Colony Optimisation
ALEXI	Atmosphere-Land Exchange Inverse
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
BANFIS	Bagged ANFIS
BATS	Biosphere-Atmosphere Transfer Scheme
BBO-ANFIS	Biogeography-Based Optimisation-ANFIS
Bi-LSTM	Bi-Directional LSTM
BMA	Bayesian Model Averaging
BMA-E	Bayesian Model Averaging Ensemble
BMLP	Bagged MLP
BMS	Bayesian Model Selection
BPNN	Back-Propagation Neural Network
BSVM	Bagged SVM
CART	Classification and Regression Tree
CatBoost	Categorical Boosting
CLM	Community Land Model
CNN	Convolutional Neural Network
DisALEXI	Disaggregated Atmosphere-Land Exchange Inverse
DNN	Deep Neural Network
ELM	Extreme Learning Machine

ESTARFM	Enhanced STARFM
ET	Evapotranspiration
ET_0	Reference Evapotranspiration
ET_c	Crop Evapotranspiration
FFBP	Feed-Forward Back-Propagation
FIS	Fuzzy Inference System
GA-ANFIS	Genetic Algorithm-ANFIS
G-ANFIS	Grid Partitioning ANFIS
GA-SVM	Genetic Algorithm-SVM
GBDT	Gradient Boosting Decision Tree
GDP	Gross Domestic Product
GEP	Gene Expression Programming
GFF	Generalised Feed-Forward
GPI	Global Performance Indicator
GRNN	Generalised Regression Neural Network
GRU	Gated Recurrent Unit
GWO-SVM	Grey Wolf Optimisation-SVM
HS	Hargreaves-Samani
ISRO	Indian Space Research Organisation
KLIA	Kuala Lumpur International Airport
LR	Linear Regression
LSGD-ANFIS	Least Squares and Back-Propagation Gradient Descent-ANFIS
LS-SVM	Least-Square Support Vector Machine

LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MARS	Multivariate Adaptive Regression Spline
MBE	Mean Bias Error
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression
MMD	Malaysian Meteorological Department
MNLR	Multiple Non-Linear Regression
MSE	Mean Square Error
NASA	National Aeronautics and Space Administration
NNE	Non-Linear Neural Ensemble
PCA	Principal Component Analysis
PET	Potential Evapotranspiration
PM	Penman-Monteith
PNN	Probabilistic Neural Network
PSO-ANFIS	Particle Swarm Optimisation-ANFIS
PSO-ELM	Particle Swarm Optimisation-ELM
PSO-GBDT	Particle Swarm Optimisation-GBDT
PSO-SVM	Particle Swarm Optimisation-SVM
PT	Priestley-Taylor
PT-JPL	Priestley-Taylor-Jet Propulsion Laboratory
R^2	Coefficient of Determination
RBF	Radial Basis Function

RF	Random Forest
RMSE	Root Mean Square Error
RS-PM	Remote Sensing-Penman Monteith
SA	Simple Annealing
S-ANFIS	Subtractive Clustering ANFIS
SEBS	Surface Energy Balance System
SIM	Simple Hybrid ET
STARFM	Spatial and Temporal Adaptive Reflectance Fusion Model
SVM	Support Vector Machine
SVR	Support Vector Regression
TLBO-ANFIS	Teaching-Learning-Based Optimisation-ANFIS
TSS	Taylor's Skills Score
VIC	Variable Infiltration Capacity
WMO	World Meteorological Organisation
WOA	Whale Optimisation Algorithm
WOA-ELM-E	Ensemble Based on Extreme Learning Machine Integrated with Whale Optimisation Algorithm
XGBoost	Extreme Gradient Boosting
Af	Tropical Rainforest Climate
e_a	Actual Vapour Pressure
e_s	Mean Saturation Vapour Pressure
G	Soil Heat Flux
K_c	Crop Coefficient

R_a	Extraterrestrial Radiation
RH	Relative Humidity
RH_{max}	Maximum Relative Humidity
RH_{mean}	Mean Relative Humidity
RH_{min}	Minimum Relative Humidity
R_n	Net Radiation
R_s	Solar Radiation
T	Daily Mean Temperature at 2 m Height
T_{max}	Maximum Temperature
T_{mean}	Mean Temperature
T_{min}	Minimum Temperature
TR	Daily Temperature Range
u	Mean Wind Speed
u_2	Mean Wind Speed at 2 m Height
γ	Psychrometric Constant
Δ	Slope of Vapour Pressure Curve

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 General Introduction

The evapotranspiration (ET) is defined as the "loss of water from the ground, lake or pond, and vegetative surfaces to the atmosphere through vaporisation of liquid water" (Pokorny, 2019). In other words, the ET value is representing the amount of water lost from the Earth's surface and shall be replenished by other hydrological processes, such as the precipitation. In Malaysia, particularly Peninsular Malaysia, instruments for measuring and monitoring the trend of ET have been lacking for decades, despite the fact the nation has an enormous dependency on the agricultural outputs that are greatly affected by the supply of water. A recent report pointed out that the agricultural sector was the third largest contributor to Malaysia's gross domestic product (GDP) in the year 2020, with the least contraction amidst the COVID-19 pandemic (Mahidin, 2021). Oil palm plantation remains the top performer in the agricultural sector with the recent uprise of global crude palm oil price. For that reason, many decision-makers and researchers opt to solve the problems with ET estimation through secondary options, such as the empirical models to estimate the potential evapotranspiration (PET) and reference evapotranspiration (ET_0).

In a recent study conducted by Theng Hue, et al. (2022), the estimated ET_0 rate in the year 2018 of Peninsular Malaysia ranged between 1000 – 1250 mm/year. Owing to the global warming caused by anthropogenic activities, ET deficits over the East Asia would be on the rise, piling greater water stress as well as agricultural risk (Kim, Ha and Yeo, 2021). It is important to devise a clear, concise and capable strategy to tackle the aforementioned issue. With this respect, an accurate yet robust tool for estimating the ET needs to be developed so that the subsequent decisions can be made based on justifiable grounds.

The PET and ET_0 have been widely used in Peninsular Malaysia for the purpose of water budget allocation. These estimations are mainly done using empirical models or equations such as the Thornthwaite (PET), Priestley-Taylor (PET), Hargreaves-Samani (ET_0) and Penman-Monteith (ET_0) equations. However, to use these empirical models or equations effectively, a vast number of meteorological variables have to be collected beforehand, in turn posing financial burdens to the administrative agencies or institutions. In fact, for models like the Thornthwaite, Priestley-Taylor and Hargreaves-Samani equations, many historical data have to be recorded to calibrate the location-specific coefficient and parameters, hence limiting their applications, not to mention low accuracy of these models had been reported frequently (Liu, et al., 2017; Shiri, et al., 2014; Tabari, Grismer and Trajkovic, 2011).

The recent development of computing technologies enables the application of machine learning for the estimation of ET. Many successful

2

research have been reported over the past two decades, confirming the capabilities of different types of machine learning models. Some of the most notable research reported the improvement in the ET_0 estimation using methods such as the multilayer perceptron (MLP), support vector machine (SVM), extreme learning machine (ELM), and the adaptive neuro-fuzzy inference system (ANFIS) to estimate ET_0 at different areas of the world (Abdullah, et al., 2015; Kisi and Cimen, 2010; Kisi and Öztürk, 2007; Kumar, et al., 2002).

The machine learning models have been continuously showing promising performances in the estimation of ET_0, not to mention that many of the empirical models or equations were outperformed. Nevertheless, the locality characteristic still exists in most of the developed machine learning models reported in the literature. In other words, the application of the developed machine learning models is highly constrained to where they are originated from. This issue also affects the empirical models that require calibration before ET_0 estimation. Besides, the data-hungry nature of the machine learning models also restricted their widespread application, as data scarcity is a major problem faced by many developing nations (including Malaysia), not to mention the availability of complete sets of data.

The study reported in this thesis targets to resolve these issues via the development of a robust and reliable machine learning model for the estimation of daily ET_0 in Peninsular Malaysia. The problem statement, aim and objectives

as well as the potential contribution of the study are presented in the following sections.

1.2   Problem Statement

The machine learning models are characterised as a data-hungry approach to estimate the ET_0. This is the major issue that has been constraining the widespread of the application of machine learning models in estimating ET_0. The data hunger of machine learning models can be grouped into two categories, namely the qualitative hunger and the quantitative hunger, as shown in Figure 1.1.
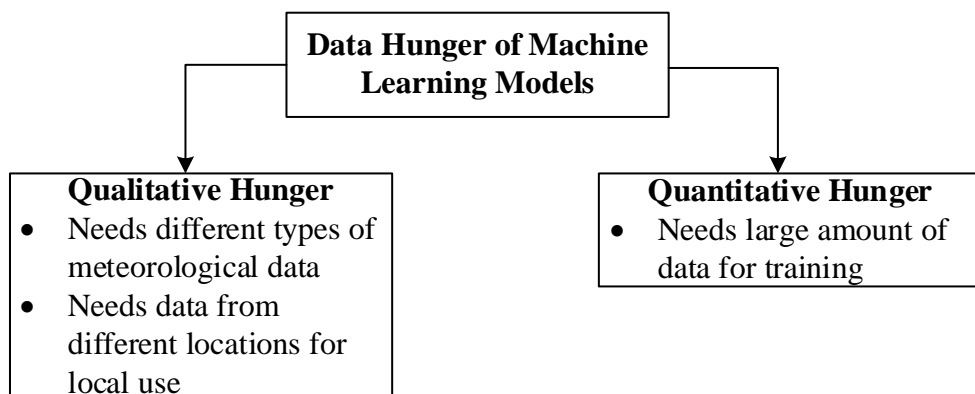
Figure 1.1: Data Hunger of Machine Learning Models

The qualitative hunger of machine learning models refers to the need for multiple types of input data so that they can produce estimations with satisfactory accuracies. For instance, to estimate ET_0, a machine learning model would require input meteorological variables such as temperature, humidity, solar radiation and so forth. The removal of essential meteorological variables

4

from the input could possibly render the model irrelevant. At the same time, the collection of many meteorological variables can be costly in the labour and fiscal terms. Hence, a model with minimal input meteorological variables is desirable.

Besides, qualitative hunger can also be reflected from the locality characteristic of the machine learning models. Many machine learning models can only be trained and tested on the spot which means that the spatial robustness of the models is rather poor. The issue that arises from this problem is that for each region of interest, local data collection infrastructures need to be installed and a local model must be trained from scratch. In other words, the pre-requisite of a workable machine learning model is the existence of various features in the input. This does not only incur a financial burden to the stakeholders, but also jeopardises the efficiency of ET_0 estimation. For example, if the decision-makers want to estimate the ET at a new site, firstly, they need to build a data collection system (which is costly) and wait for a period of time before a model can be developed.

This waiting period can be a few months, or even years before the data collected are sufficient for decent modelling. This is as a result of the quantitative hunger of machine learning models. Machine learning models need ample number of examples (training data) so that they can adjust their hyper-parameters to suit the application. It is impractical if data need to be collected for several years before the modelling process can proceed.

The qualitative hunger, coupled with the quantitative hunger together form the bottleneck of the wide application of machine learning models in the estimation of ET_0, and hence the major gap of this research field. Therefore, this research work aims to eliminate this bottleneck and bridge the gap through the development of a robust machine learning model using data fusion techniques that needs minimal input with high spatial transferability.

1.3   Aim and Objectives

The aim of this research is to develop a robust machine learning model for the estimation of ET_0 with minimum data requirement, considering the qualitative and quantitative requirements of data, in Peninsular Malaysia. The aim can be accomplished by pursuing the following objectives:

i)   To identify optimum input combinations to estimate ET_0 in Peninsular Malaysia for the MLP, SVM and ANFIS

ii)   To evaluate the ensemble hybridised using different data fusion techniques, namely bootstrap aggregating, Bayesian modelling approach and non-linear neural ensemble (NNE) on ground-observed meteorological data

iii)   To investigate the relationship between different stations by estimating local ET_0 using exogenous models for the reduction of local data dependency (Scenario 2)

iv)   To assess the spatial robustness of the hybrid model trained using pooled global dataset in ET_0 estimation (Scenario 3)

1.4 Scope of Study

The investigated study area is located in Peninsular Malaysia. The investigation was done based on the ground-observed meteorological data at 12 different meteorological stations that were managed by the Malaysian Meteorological Department (MMD). A total of 20 years of data duration (1st January 2000 to 31st December 2019) was retrieved from the MMD, amounting to 7305 data points at each station. To minimise the computational cost, data fusion techniques were integrated with the three selected elementary base models for improvisation. The three elementary base models selected are the MLP, SVM and ANFIS due to their suitability in regression analysis. The decision tree was not selected as its binary nature was more suitable for classification task rather than being used as a regressor. The bootstrap aggregating (data centric), Bayesian modelling approach (model centric) and NNE approaches were chosen as the data fusion techniques to be integrated to the base models. The justification of the design of the scope of this study will be outlined in detail in Chapter 3.

1.5 Contribution of the Study

A robust data fusion integrated machine learning model was developed at the end of this study. This model is capable of overcoming the bottleneck stated in the problem statement, which arose due to the qualitative and

quantitative data hunger. The machine learning model developed has the following characteristics:

i)   Simple structure with high computational efficiency

ii)   Requires a minimum number of input meteorological variables

iii)   Wide spatial applicability to resolve the issue with local data dependency

The novel output of this research is deemed to have a high contributory impact that can be elaborated from two aspects. Firstly, from the scientific point of view, this study investigated the feasibility of different data fusion techniques, each with a unique theoretical basis, in assisting simple machine learning models in the estimation of ET_0. These findings laid an important foundation for the integration of data fusion techniques to other machine learning models and can be adopted to new models with more sophisticated designs. Besides, the nature, behaviour and interaction of different data fusion techniques with the selected machine learning models are discussed in detail in this thesis.

From the national economic point of view, the development of a one-for-all machine learning model for the ET_0 estimation is a huge aiding tool for the decision-making process. The developed model can be applied across the whole Peninsular Malaysia, subsequently facilitating the estimation and mapping of ET_0 across the whole region immediately when the essential data are available in the future. The policy makers should use this information wisely in drawing appropriate strategies for water resources allocation, irrigation

scheduling, crop planning and so on. Proper execution of these plans can help to catalyse the growth of the agricultural sector. This can lead to the stimulation of the recovery as well as the resilience of the national economy, which unfortunately had been dwindling in recent years. The welfare of the society can be further improved to spur Malaysia to become a high-income nation in the nearest possible future.

1.6   Thesis Structure

There are five chapters in this thesis, including the current chapter. Chapter 1 briefly introduces the background, aim and objectives as well as the scope and contributions of this study. Chapter 2 focusses on discussing the findings after reviewing the publications relevant to the ET as well as the application of machine learning models for the estimation of ET. This includes the latest trend of study and their pros and cons. At the end of Chapter 2, the rationale and the novelty of this research work (research gap) are outlined. In Chapter 3, the detailed methodology of this research study is presented, including the development of base and hybridised models to determine their optimum inputs and structures, as well as the testing strategies (different training and testing scenarios) to verify the robustness of the models. Chapter 4 discusses the results obtained in terms of the performance of the base models using different input combinations, the effect of data fusion techniques and the spatial robustness of the models. A concise analysis of the results is provided in this chapter. Finally, in Chapter 5, a brief and conclusive remark is given,

together with recommendations for future works related to the foundation/basis

laid down by this research study.

CHAPTER 2

LITERATURE REVIEW

2.1 Evapotranspiration

Citing from a report published by the United Nations (2019), the world population stood at 7.7 billion. It was anticipated that this number would continue to increase and reach 8.5 billion, 9.7 billion and 10.9 billion in the year 2030, 2050 and 2100, respectively. Subsequently, agricultural activities will increase accordingly to sustain the food supply and thus become vital. Agricultural activity is being considered as the human activity that consumes the largest amount of water (Cascone, et al., 2019). Therefore, a good estimation of the components of the water cycle can assist in various water resources allocation activities, such as irrigation planning, agriculture scheduling and so on, in turn optimising the utilisation of water.

Evapotranspiration (ET) is the combinatory effects of the water evaporation from the surfaces of land and vegetation as well as the transpiration from the stomatal openings of plant leaves (Stanhill, 2005). It is a natural event that influences the hydrological cycle, and it consists of several non-linear processes that make it very complex to be understood (Jovic, et al., 2018). There

are several components that dictate the rate of ET, including the temperature, solar radiation, air humidity as well as wind speed (Granata, 2019).

ET is a physical process. Therefore, the rate of ET can be measured and represented by a numerical value. Lysimeters can be used to measure ET directly in the absence of assumptions (Holmes, 1984). Its operating mechanism involves measuring the water percolation rate through the soil (Pokorny, 2019). Non-weighable lysimeters are usually used for long-term monitoring, while the weighable lysimeters give measurements with finer temporal resolution (Wang and Dickinson, 2012). Lysimeters were said to be capable of providing ET measurement with maximum accuracy. In fact, studies related to the ET estimation often used the measurements of lysimeters as the calibration standard (Anapalli, et al., 2016; Liu, et al., 2017). Unfortunately, constructing and maintaining lysimeters involve a high fiscal burden and ecological footprints. The scarcity and availability of lysimeters also limit its coverage and hindered the easy measurement of ET at different places (Stanhill, 2005). Thus, the development of other tools with better convenience became more in demand to estimate ET with better accuracy and cheaper cost.

2.2   Reference Evapotranspiration

According to Pereira, et al. (2015), reference evapotranspiration (ET_0) estimates the amount or rate of water consumption based on the weather's primary effect. It should be noted that this measurement is only for the

surrounding conditions of the reference crop at the station that conforms to the requirements such as no shortage of water supply. In the coefficient-reference system, a crop coefficient (K_c) will be multiplied by the ET_0, resulting in crop evapotranspiration (ET_c) as shown in Equation (2.1):

$$ET_c = K_c \times ET_o \qquad (2.1)$$

Over the years, many initiatives had been taken to obtain accurate ET_0 whilst reducing the computational complexity. Some of the most notable empirical approaches and models will be discussed in detail in the coming subsections.

2.2.1 Temperature-Based Models

The temperature-based models used to calculate the ET_0 generally are modified or derived from the temperature-based model for the potential evapotranspiration (PET) imputation. In essence, the temperature-based models measure the atmospheric evaporation demand based on the temperature data obtained for a specific time scale. Among the temperature-based models, the Hargreaves-Samani (HS) model was developed and improvised in the 1980s (Hargreaves and Samani, 1985). The HS model is shown in Equation (2.2).

$$ET_0 = aR_a(T_{mean} + 17.8)(TR)^b \qquad (2.2)$$

where:

ET_0 = daily reference evapotranspiration (mm/day)

R_a = extraterrestrial radiation (MJm^{-2}day^{-1})

T_{mean} = mean temperature (°C)

TR = daily temperature range (°C)

*a and b are empirical parameters typically assumed 0.0023 and 0.5, respectively

Even though the R_s is needed for the imputation of ET_0 using the HS model, however, the R_s can be estimated from the number of sunshine hours. In other words, the only compulsory meteorological variable needed for the HS model is only the maximum, minimum and mean temperature (on a daily scale as of Equation (2.2)). The HS model has been cited on more than 4000 different occasions for ET_0 estimations.

However, the constants and coefficients (the a and b parameters) used in the HS model can be specific to different locations. Hence, calibrating these coefficients in the HS model to suit local requirements became a norm. Luo, et al. (2014) validated the utilisation of calibrated HS model in Guilin, Kaifeng, Ganyu and Yinchuan using forecasted temperature to predict ET_0. Although the prediction was sufficiently accurate, however, the HS model would fail (either overestimate or underestimate) in extreme conditions as the effects of wind speed and relative humidity were not accounted for.

In Veneto, Italy with a sub-humid climate, an investigation was conducted to compare the calibrated and non-calibrated HS models (Berti, et al., 2014). The unmodified HS model tended to overestimate the ET_0, leading to

excess water requirements reported. Calibration of the HS model managed to reduce the overestimation from 18.9 % to 2.6 %, thus indicating the significance of calibration to the empirical HS model.

Another interesting method to adjust the estimation of the HS model is to produce a linear relationship between the adjusted and the original HS model. This was done by Zanetti, et al. (2019) for a variety of thermal amplitude classes, climate types (tropical savanna, tropical monsoon, humid subtropical, oceanic and subtropical highland) and seasons (dry and rainy). In the study, combining thermal amplitude class and climate types during adjustment produced the lowest estimation error, which outperformed a simplified Penman-Monteith (PM) model with limited data.

From the literature review, due to the constraints of the model which only takes temperature and radiation effect into account, the performance of the HS model is weaker than the PM model that is mentioned earlier in Chapter 1. Hence, in general cases, it cannot be considered as a good replacement for the relatively complex PM model.

Beside the HS model, there are several temperature-based models which are less well known, not to mention their rare applications in real-life cases. These include the Xu and Singh model developed based on the first version of the HS model (Xu and Singh, 2000), Trajkovic model developed based on the Hargreaves model (Trajkovic, 2007) and so on. In general, to maintain the

performance of the temperature-based models for ET_0 estimation, the correlation between the temperature as well as the ET_0 should be as high as possible. Local calibration and modifications based on regional needs are definitely recommended.

2.2.2   Radiation-Based Models

The radiation from extraterrestrial space and the Sun provides the energy needed for the water to escape from the Earth's surface. Hence radiation-based models are also a dominant force in the field of ET_0 estimation. The first radiation-based model can be tracked back as early as the Ritchie model developed in the 1970s (Ritchie, 1972). The initial version of this model was used to estimate the rate of evaporation. By including the parameters which consider the hydraulic conductivity of plants, the ET_0 can be computed using Equation (2.3).

$$ET_0 = \frac{\Delta}{\Delta + \gamma} R_n \qquad (2.3)$$

where:

Δ = slope of vapour pressure curve (kPa/°C)

γ = psychrometric constant

R_n = net radiation (MJm^{-2}day^{-1})

The Ritchie model had been cited in almost 3000 instances. In one of the latest investigations, the Ritchie model was claimed to have the best performance among the empirical models when being applied to National

Aeronautics and Space Administration (NASA) and Indian Space Research Organisation (ISRO) data in India by using the ET_0 calculated from the PM model as the standard. Nevertheless, the performance of the Ritchie model still fell far behind modern techniques such as the ANFIS (Gonzalez del Cerro, et al., 2021). This indicated that the Ritchie model, apart from being a convenient alternative, cannot be regarded as the first-choice solution for ET_0 estimation.

Xu and Singh (2000) also proposed several radiation-based models for ET_0 estimation. However, the applications of these models are highly constrained to their geographical origin, making them less relevant in solving the current issues faced by many researchers and administrators in ET_0 estimation. Xiang, et al. (2020) had made a conclusive remark on the radiation-based models. Citing from their publication works, it was mentioned that the performance of the radiation-based model varies drastically, primarily affected by the geographical locations as well as the local climate. The authors recommended that these models shall be applied in regions with semi-arid and semi-humid climates to harvest better ET_0 estimation accuracy.

2.2.3 Combinatory Models

The combinatory models combine different aspects that have to be considered in the ET process. These include the energy balance as well as the aerodynamic conditions. To use the combinatory models for ET_0 calculation, various factors that would affect the rate of ET need to be included. This means

that the combinatory models would require a luxurious number of meteorological variables, which are only completely available in certain regions of the world.

The PM model is considered as one of the most popular models in estimating ET_0. Furthermore, the Food and Agriculture Organisation of the United Nations, in their publication "Crop evapotranspiration – Guidelines for computing crop water requirements – FAO Irrigation and Drainage Paper 56", in short FAO56, revised the computation of ET_0 and PET based on PM model (Allan, et al., 1998). This indirectly endorsed the PM model as the standard in estimating ET_0. The PM model was cited in numerous research studies as a standard of comparison (Güçlü, Subyani and Şen, 2017; Saggi and Jain, 2019; Shiri, et al., 2019). The complete PM model is shown as Equation (2.4):

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\left(\frac{900}{T + 273}\right)u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \qquad (2.4)$$

where:

R_n = net radiation (MJm^{-2}day^{-1})

G = soil heat flux (MJm^{-2}day^{-1})

T = daily mean temperature at 2 m height (°C)

u_2 = wind speed at 2 m height (m/s)

e_s = mean saturation vapour pressure (kPa)

e_a = actual vapour pressure (kPa)

Δ = slope of vapour pressure curve (kPa/°C)

γ = psychrometric constant

The PM model includes most of the input meteorological variables that are believed to explain ET_0 well. However, the degree of dependency of PM model on the input meteorological variables (at least eight according to the equation) also limits its applicability in the area with limited data. As mentioned earlier, some regions would face difficulties using the PM model when most of the meteorological variables involved in the PM model could not be obtained from direct measurement. Therefore, a list of supporting equations and assumptions is needed to complement the PM model (Valiantzas, 2013).

Valiantzas (2013) did a comprehensive study on the modification of the PM model so that the equation can take the measured meteorological data directly as input. This was done by developing an equivalent PM model algebraically followed by further simplifications. On top of that, the author also produced a series of equations to cater to the needs of regions where wind speed and radiation data were not available. The investigation results showed that for humid or arid regions, the simplified PM model achieved sufficient correlation with the original PM model, with the reported coefficient of correlation (R^2) as high as 0.99. Furthermore, when wind speed and radiation data were not being measured, the reduced PM model could still compute similar ET_0 results as the PM model. This study has successfully proved that the potential of developing a more convenient and accurate tool of estimating ET_0.

2.2.4 Comments on Conventional Acquisition of ET Value

The measurement of the ET (as well as the ET_0 and PET) is virtually impossible and impractical due to the technical limitations, coupled with the high fiscal commitment and ecological footprints incurred. Therefore, the need for a secondary approach to get such values becomes pertinent. Prior to the popularisation of soft computing techniques, the empirical models played important roles in estimating the ET_0 based on the meteorological variables available. However, the temperature-based and radiation-based empirical models suffered from instability and inconsistency in the estimation accuracy. Moreover, the locality characteristics of these empirical models limited their applications over a wide range of regions unless tedious modifications or calibrations were conducted.

Current research trend generally employs the combinatory models, especially the PM model as the standard to compare the accuracy of newly developed ET_0 estimation tools. Hence, in this study, the PM model will also be used as a standard measure. Although the PM model shows outstanding ability in ET_0 estimation, this however, is at the expense of the need for plenty of meteorological variables. A simpler approach is needed for the accurate estimation of ET_0 with minimal input and computational effort. This will be the subject of discussion in the next section.

2.3   Machine Learning for Evapotranspiration

As mentioned in the previous sections, ET can be measured using a lysimeter which is expensive to build and maintain, notwithstanding its ecological impact. Alternatively, ET_0 can be computed using energy balance if various climatic data are available. This approach reduces the need for a costly lysimeter; however, it requires higher computational effort as the data obtained is too raw to be used. On top of that, the use of empirical models involves assumptions, and sometimes the exclusion of specific variables can degrade the accuracy of estimation. Hence, researchers now shift their direction of study to generate an easy-to-use model, without much understanding to the physical properties of ET and its causal correlation with the climate. In this context, machine learning came into the vision of researchers as a potential candidate to fulfil the objective. Chia, et al. (2020), in their recent review, showed that there is an exponential increase in publications reporting the use of machine learning methods for ET-related studies. In this section, machine learning models commonly used in ET_0 will be discussed, alongside the evolution of the models across the timeline.

2.3.1   Artificial Neural Network

The most commonly used machine learning model is the artificial neural network (ANN). Specifically, the ANN used in ET_0 is subdivided into several variations, including the multilayer perceptron (MLP), radial basis function

network (RBF), generalised regression neural network (GRNN), extreme learning machine (ELM) and the back-propagation neural network (BPNN), each with different architectures and estimation functions. A general illustration of ANN is provided in Figure 2.1.

Input Layer

Hidden Layers

Output Layer
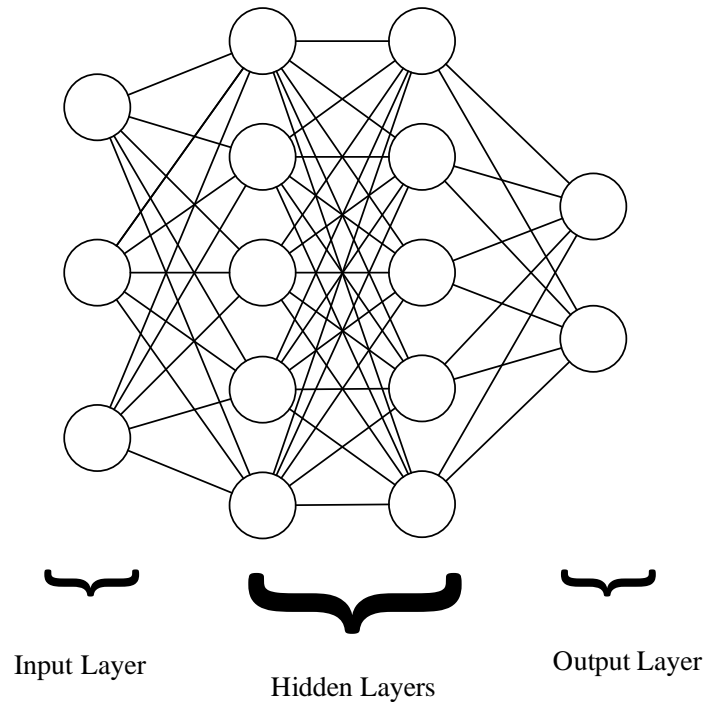
Figure 2.1: General Structure of the ANN

The application of the ANN is a simulation of biological neurons in the nervous system, where neurons are connected via synapses. In the ANN, the neurons are connected between layers through weights and biases. This intrinsically establishes the relationship between the input and output layers during the learning process (Abiodun, et al., 2018).

In recent years, the use of the ANN had been evolved by increasing the number of hidden layers. This was accompanied by an improvement in the types

of hidden neurons. Instead of using classical hidden neurons regulated by activation functions, advanced hidden neurons can exist in the form of long short-term memory (LSTM) cells, gated recurrent unit (GRU) and so on. This revolution sees a paradigmatic shift to the utilisation of the deep neural network (DNN) for various applications, including ET_0 estimation. As compared to the classical ANN, the DNN is believed to be relatively capable of learning more complex relationships due to its ability to store different states within the cells (Hu, et al., 2018). Furthermore, the emergence of big data and cloud computing also provides a more conducive environment for the application of the DNN. However, the computational (and time) cost of the DNN is much higher than the conventional ANN and thus unworthy when dealing with small sets of data (Nagappan, Gopalakrishnan and Alagappan, 2020). The simple and easy applications of the ANN have attracted the attention of numerous researchers to estimate ET_0 using ANN, and subsequently, their research has attained tremendous achievements. Table 2.1 summarises research using ANN in ET_0 study.

Table 2.1: Summary of Research using the ANN in ET$_0$ Study

Approach	Key Findings	Reference
MLP	• MLP was used to estimate mean monthly ET$_0$ in humid subtropical environment • MLP was trained using PM model as target and compared with calibrated HS model • When T_{max}, T_{min} and R_a were used as input meteorological variables, MLP six hidden nodes achieved the best performance	(Rahimikhoob, 2009)
BPNN	• Daily T_{max}, T_{min}, R_a, u, RH and sunshine hours were used to train the BPNN using PM model as reference in the area with semi-arid climate pattern • Different combinations of input meteorological variables were tested • Temperature-based BPNN performed better than HS model • Inclusion of u and RH data further enhanced the estimation accuracy • R_a and sunshine hours did not play a significant role in the improvement of model accuracy	(Traore, Wang and Kerh, 2010)

Table 2.1 (continued): Summary of Research using the ANN in ET$_0$ Study

Approach	Key Findings	Reference
RBF GRNN	• Daily T_{max}, T_{min}, T_{mean}, u and sunshine hours were used to train the networks to estimate ET$_0$ based on PM model in the temperate zone with mild Mediterranean climate • Different combinations of input meteorological variables were tested to investigate the best possible dataset • GRNN generally performed better than RBF • Removal of u caused a sudden drop in prediction accuracy • The performance of the models could be further improved by the addition of precipitation data	(Ladlani, et al., 2012)
MLP	• Daily T_{max}, T_{min}, T_{mean}, R_s, u, RH_{mean}, RH_{min} and sunshine hours were used to train the MLP using PM model as reference in the area with semi-arid climate pattern • The MLP was trained using different learning algorithms, including Levenberg-Marquardt, Delta-Bar-Delta, Step, Momentum, Conjugate Gradient and Quickprop • Levenberg-Marquardt, Delta-Bar-Delta and Conjugate Gradient algorithms with hyperbolic tangent transfer functions performed the best • The combinations of input meteorological variables with the highest prediction accuracy were T_{mean}, RH_{min}, u and sunshine hours, which well explained the properties of ET	(Tabari and Hosseinzadeh Talaee, 2012)

Table 2.1 (continued): Summary of Research using the ANN in ET$_0$ Study

Approach	Key Findings	Reference
Generalised ANN Multi Linear Regression (MLR)	• Daily T_{max}, T_{min}, R_s, u and RH were used to train the models using PM model as reference in humid, sub-humid, arid and semi-arid areas • One station was used to train in one climate region, while another was used for testing • Different combinations of input meteorological variables were tested • As the number of input meteorological variables decreased, the prediction accuracy decreased gradually • Generalised ANN performed better than MLR	(Wang, et al., 2013)
ELM Feed-Forward Back-Propagation (FFBP) Network	• Daily T_{max}, T_{min}, R_n, u and RH were used to train and test the models using PM model as target in arid and semi-arid regions • Different combinations of input meteorological variables were tested for both networks • ELM and FFBP estimated daily ET$_0$ with comparable accuracy • ELM was claimed to be more efficient • Reduction of input meteorological variables did not significantly affect the prediction accuracy	(Abdullah, et al., 2015)

Table 2.1 (continued): Summary of Research using the ANN in ET$_0$ Study

Approach	Key Findings	Reference
MLP Probabilistic Neural Network (PNN) Generalised Feed-Forward (GFF) Linear Regression (LR)	• Study was carried out to forecast daily ET$_0$ using forecasted temperature in humid subtropical environment • The four models were simulated with forecasted daily T_{max} and T_{min} • Performances of MLP, PNN and GFF were slightly better than LR • Errors were accumulated from the inaccuracy of forecasted temperature and short of climate data	(Luo, et al., 2015)
RBF MLP Support Vector Machine (SVM)	• Monthly T_{max}, T_{min}, e_a, u and sunshine hours were used to train the MLP using PM models as reference in moderate Mediterranean region • Optimisation of RBF network was modified, either by back-propagation or particle swarm optimisation • RBF performed better than MLP and SVM, but method of optimisation did not show a significant difference	(Petković, et al., 2015)
ELM	• Monthly T_{max}, T_{min}, e_a, u and sunshine hours were used to train ELM based on HS model, PT model as well as Turc model in the temperate region • The three models did not show notable differences; however, ELM could be employed as it reduced the complexity in calculating ET$_0$	(Gocic, et al., 2016)

Table 2.1 (continued): Summary of Research using the ANN in ET$_0$ Study

Approach	Key Findings	Reference
ELM MLP Genetic Programming SVM	• Monthly T_{max}, T_{min}, u, pan-evaporation rate, rainfall and sunshine hours were used to train the models using PM model as reference in humid subtropical region • Estimation of ELM was more accurate as compared to other models and also required lesser computational time • ELM using sigmoid transfer function performed better than its hard limit transfer function counterpart	(Kumar, et al., 2016)
ELM MLP Least-Square Support Vector Machine (LS-SVM)	• Weekly T_{max}, T_{min} and R_a were used to train the networks to estimate ET$_0$ based on HS model in arid region • Further investigation also included ET$_0$ value from other stations as input to the model • ELM was proved to have the best performance as it required less human intervention with good estimation efficiency • Inclusion of external ET$_0$ value further enhanced the accuracy of estimation	(Patil and Deka, 2016)

Table 2.1 (continued): Summary of Research using the ANN in ET_0 Study

Approach	Key Findings	Reference
MLP (back-propagation) PNN GFF LR	• Models were trained using daily T_{max}, T_{min}, R_s and R_a using PM model in humid subtropical environment • Forecasted climate data with different combinations were feed into the networks • It was found that MLP with T_{min}, T_{max} and R_s as input meteorological variables could produce the most accurate results, up to 15 days forecast horizon • T_{max} was claimed to be the most significant factor for ET_0 estimation	(Traore, Luo and Fipps, 2016)
MLP	• Daily T_{mean}, R_s, u and RH were used to train the MLP using PM model as reference in humid continental region • Different combinations of input meteorological variables were tested • MLP network with complete set of meteorological variables performed the best, followed by models that contained temperature and radiation data	(Antonopoulos and Antonopoulos, 2017)
ELM GRNN	• Daily T_{max}, T_{min}, and R_a were used to train the networks to estimate ET_0 based on PM model in warm humid region • ELM and GRNN outperformed HS model by reduction of deviation from actual ET_0 value	(Feng, et al., 2017b)

Table 2.1 (continued): Summary of Research using the ANN in ET$_0$ Study

Approach	Key Findings	Reference
Random Forest (RF) Extreme Gradient Boosting (XGBoost) MLP Convolutional Neural Network (CNN)	• Only temperature and humidity data were used to estimate ET$_0$ on an hourly basis • The MLP outperformed XGBoost and RF, while the CNN performed better in relation to the MLP • The authors recommended further studies on the deep learning approaches	(Ferreira and da Cunha, 2020)
RBF CNN	• T_{max}, T_{min} and u were used as the predictors after performing principal component analysis (PCA) • CNN performed better than RBF in ET$_0$ estimation, however, the time needed for training was longer	(Nagappan, Gopalakrishnan and Alagappan, 2020)
DNN	• DNN was designed from MLP with four hidden layers • It was found that input combinations that only included the temperature and radiation data were able to produce ET$_0$ estimation with accuracy close to that of complete dataset	(Sowmya, Santosh Kumar and Ambat, 2020)

As shown in Table 2.1, a considerable number of studies had been done on the utilisation of the ANN as an ET_0 estimation tool. However, the trend of the studies generally focused the following few aspects:

- Minimisation of mandatory inputs

- Generalisation of the ANN for wider applications

- Application of novel input features

- Improvisation of ANN's ability to forecast

It is believed that the four aspects stated above could revolutionise the prediction of ET_0, with a more general model without the need for much climate data. On top of that, a longer forecasting horizon acts as an important pre-requisite for a pro-active water management strategy. Unfortunately, using ANN alone seems to be insufficient in providing the solution. Hence, the coming subsections will be focussed on the discussion of other machine learning models that were employed in ET_0 estimation.

2.3.2   Support Vector Machine

The support vector machine (SVM) is said to be capable in both regression and classification tasks, making it to become of the most important algorithms in machine learning modelling (Vapnik, 1995). Cortes and Vapnik (1995) proposed the foundation of the SVM. The SVM transforms the data points into a feature space using a kernel function. In the feature space, the

relationship between inputs and outputs mapped, where problem complexity and accuracy can be optimised at the same time.

Since the ET_0 estimation is a regression problem, the support vector regression (SVR) is typically used. In the working mechanism of an SVR, a loss (or cost) function is used to define the allowable deviation and the function to estimate the targeted output (Raghavendra and Deka, 2014). The working principle of SVM is shown in Figure 2.2.
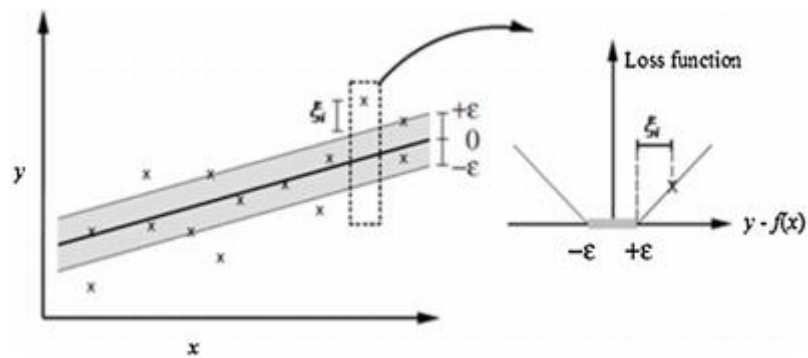
Figure 2.2: Working Principle of the SVM (Shrestha and Shukla, 2015)

The SVM had been popular in hydrology applications, including the ET_0 estimations (Raghavendra and Deka, 2014). The SVM has high robustness, capable of solving complex problems, less vulnerable to overfitting and could describe the model in a more compact manner (Zendehboudi, Baseer and Saidur, 2018). The network structure of the SVM is illustrated in Figure 2.3.
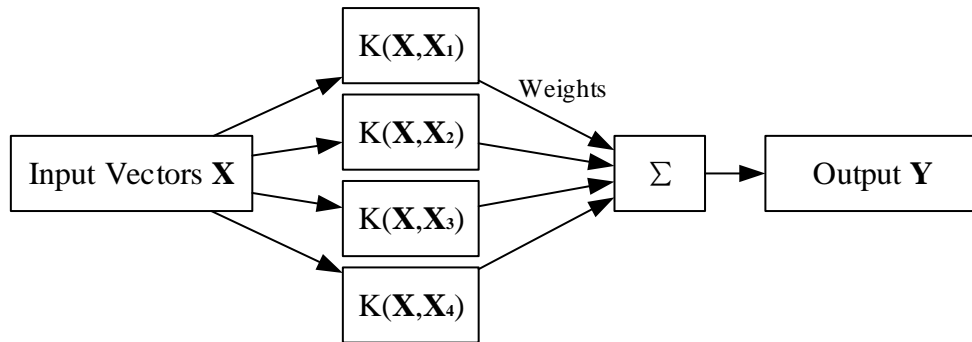
Figure 2.3: Network Structure of the SVM

Application of the SVM in ET_0 estimation has long become a practice in the field. This was encouraged by the ability of the SVM to learn complex relationship between input features and ET_0, followed by the deduction of accurate predictions. Table 2.2 summarises some research works using the SVM in ET_0 study.

Table 2.2: Summary of Research using the SVM in ET$_0$ Study

Approach	Key Findings	Reference
SVM Adaptive Neuro-Fuzzy Inference System (ANFIS) MLR Multiple Non-Linear Regression (MNLR)	• Mean monthly T_{max}, T_{min}, T_{mean}, RH, u and R_s were used to train SVM and ANFIS in cold mountainous areas based on PM model • Different combinations of input meteorological variables were tested • Optimum member function and kernel function were determined for ANFIS and SVM, respectively • It was found that for SVM, the best kernel function was RBF • SVM and ANFIS performed better than MLR, MNLR and other empirical models	(Tabari, et al., 2012)
LS-SVM MLP	• Daily T_{mean}, R_s, u and RH were used to train the models using PM model as reference at warm temperate region • Different combinations of meteorological variables were tested • When all inputs were available, least square SVM had the best performance • In the case where u and RH data were lacking, HS and PT models performed better than machine learning models	(Kisi, 2012)

Table 2.2 (continued): Summary of Research using the SVM in ET$_0$ Study

Approach	Key Findings	Reference
SVM BPNN Genetic Programming	• Monthly T_{max}, T_{min}, u, e_a and sunshine hours were used to train the models based on PM model at moderately continental area • Wavelet transformation and firefly algorithm were combined with SVM for the purpose of anticipating future ET$_0$ • Both SVMs performed better than BPNN and genetic programming, with the SVM using discrete wavelet transformed data achieving the highest estimation accuracy	(Gocić, et al., 2015)
SVM Multivariate Adaptive Regression Splines (MARS) Gene Expression Programming (GEP)	• Monthly data of T_{max}, T_{min}, T_{mean}, RH, u and R_s were used to train the models based on PM model at arid and semi-arid regions • Different combinations of meteorological variables (based on data types) were tested • The study revealed that irrespective of the data types, MARS would have the best performance, followed by RBF-based SVM, GEP and polynomial-based SVM	(Mehdizadeh, Behmanesh and Khalili, 2017)
LS- SVM	• Monthly T_{max} and T_{min} were used to train the model using HS model as reference at subtropical climate area • The objective of the study was to obtain information on evapotranspiration in the future generation	(Kundu, Khare and Mondal, 2017)

Table 2.2 (continued): Summary of Research using the SVM in ET$_0$ Study

Approach	Key Findings	Reference
SVM Categorical Boosting (CatBoost) RF	• Daily T_{max}, T_{min}, u, R$_s$ and RH data were used to train the models using PM model as reference at subtropical monsoon region • Different combinations of inputs were tested • RF was overfitted easily • SVM had the best performance • However, CatBoost consumes less computational cost as compared to SVM	(Huang, et al., 2019)
SVM MLP MLR	• The machine learning models were trained based on the ET$_0$ value calculated from the PM model • MLP had the best performance when the temperature, radiation and humidity data were fed to the model • However, when the humidity data was absent, the error rate of the MLP doubled • SVM was more robust towards the reduction of input meteorological variables	(Kaya, et al., 2021)

Based on the literature review, it can be inferred that the SVM is a reliable model for estimating ET_0. However, the performance of the SVM is strongly correlated to the kernel function as well as the quality of data inputs (Raghavendra and Deka, 2014), although some may find that the SVM is more robust towards the reduction of input meteorological variables than the ANN-based models. Since the construction of the model is purely dependent on the data provided, extrapolation of the SVM is likely to produce poor results. It means that the SVM has to be calibrated or trained using local data to ensure its reliability and relevance. On top of that, the SVM can be computationally expensive when modelling problems with higher non-linearity as well as dimensions, thus consuming a significant amount of time. Hence, there is still a wide option of research direction involving the optimisation of SVM in ET_0 estimation.

2.3.3   Fuzzy Logic

Fuzzy logic was introduced by Zadeh (1965) which allows describing data in the form of the "degree of likeliness". The "either A or B" description can be replaced with "partly A and partly B" by assigning a membership degree that typically ranges from 0 to 1. The application of fuzzy logic needs the experts to select a membership function that can describe the type of distribution of the data (Gaussian function is generally favoured). Besides, three essential elements should be supplied to the fuzzy inference system (FIS). These include a set of fuzzy rule base, the membership functions and a fuzzy rule application

mechanism (Kisi, 2013). The conceptual diagram of an FIS is illustrated in Figure 2.4.
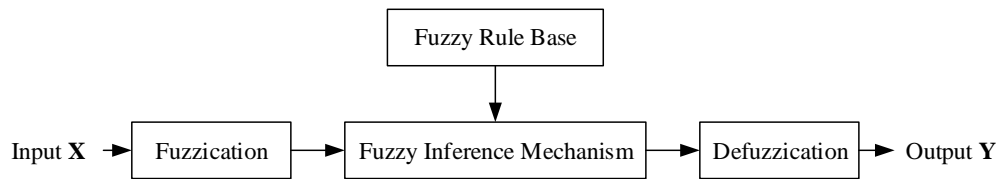
Figure 2.4: Conceptual Diagram of the FIS

Unlike the non-linear learning in the ANN and the kernel transformation of the SVM, fuzzy logic offers an alternative for the machine to learn the complex relationship between the input meteorological variables and the ET_0. Many recent works hybridised the fuzzy logic with the ANN to form the adaptive neuro-fuzzy inference system (ANFIS) to estimate ET_0 (will be discussed further in the following subsection).

2.3.4   Other Models

Apart from the three most popular basic machine learning models used to predict the behaviour of the ET, there are some other miscellaneous models which were investigated by the researchers. Of all the varieties, the tree-based models have a larger share among the researchers' preference. Feng, et al. (2017a) compared the performance of the RF and the GRNN in estimating ET_0 at an area with a warm and humid climate. The RF works to assemble multiple trees for regression or classification purposes to produce a more accurate outcome through output aggregation. In the study, it was found that the RF

performed slightly better than the GRNN, regardless of the input combinations and selection of the training datasets.

Besides, the M5 model tree was also applied to estimate ET_0 (Fan, et al., 2018; Granata, 2019; Pal and Deswal, 2009; Rahimikhoob, 2014). Unlike the classical decision tree, an M5 model tree is more suitable for regression analysis. It works by splitting the data into binary decision trees based on the splitting criterion (normally the standard deviation is used). The splitting possibility with minimised error will be chosen and the linear regression between independent and dependent variables will be stored in the leaves of the tree. Due to the highly split tree structure, overfitting tends to occur, and to overcome this problem, pruning is necessary to reduce the size of the tree, where portions that are insignificant to the tree performance will be removed.

The literature suggested that the M5 model tree could provide satisfactory, some even produced better predictions, as compared to the ANN and SVM. However, the M5 model tree only involves binary piecewise functions and linear correlations in its operation. This makes the computational complexity of the M5 model tree greatly reduced. This noble characteristic of the M5 model tree offers a potentially promising solution to the modelling of complex problems such as the ET_0 estimation or forecasting.

2.3.5 Evolution and Hybrid of Machine Learning Models

Over the past two decades, the basic machine learning models experienced different degrees of evolution. For example, the DNN and the deep belief network (DBN), which originated from the MLP network structure were applied in ET_0 estimation. As mentioned in Section 2.3.1, the DNN works in such a way that the hidden layers of the MLP are increased, while the DBN operates on top of this basis. DBN comprises a stack of restricted Boltzmann machine that is trained to generate suboptimal initial parameters. It was claimed that a model such as DBN would require a large amount of historical climate data in order to work well (Xu, et al., 2018).

The most pronounced characteristic of the current research culture is the utilisation of hybrid machine learning models. A hybrid model combines the traits of multiple basic models to create a new model for training. This can be done by combining the basic theory of more than one model (such as ANFIS), transformation of data (wavelet decomposition), metaheuristic approach (optimisation algorithm) or even ensemble models (bootstrap aggregation). It is believed that combining models can produce more accurate results by forming a decision committee. This statement was justified by previous studies (Alizadeh and Nikoo, 2018; Traore, Luo and Fipps, 2016). The advantages of using ensemble models in machine learning will be discussed further in Section 2.5. Table 2.3 summarises some of the research works involving the utilisation of hybrid machine learning models in the field of ET_0 estimation.

40

As can be seen in Table 2.3, many studies focused on studying the incorporation of data decomposition or metaheuristic methods to improve the performance of the base machine learning models. In fact, the recent trend of research shows a preference for the use of metaheuristic methods such as the swarm intelligence and other optimisation algorithms. However, from these publications, the selection of the metaheuristic methods used were mainly experimental based, often without the justification of their choice. This phenomenon is reasonable as there is a high number of optimisation algorithms available and it is of the researchers' interest to test out the best algorithm that is relevant to the ET study.

Table 2.3: Research Studies using the Hybrid Machine Learning Model in ET₀ Study

Approach	Key Findings	Reference
MLP (base model) G-ANFIS S-ANFIS	• Daily T_{mean}, RH, u and R_s were used to train the models based on PM model at moderate Mediterranean climate region • Different combinations of input meteorological variables were tested • G-ANFIS and S-ANFIS had better performance as compared to MLP and other empirical models • Besides T_{mean} and R_s, the meteorological variable that affected the predictions significantly is RH	(Cobaner, 2011)
G-ANFIS S-ANFIS	• Daily T_{mean}, R_s, u and RH were used to train the models based on PM model in Mediterranean region • Different combinations of input meteorological variables were tested • G-ANFIS and S-ANFIS generally performed better than empirical models • u was found to be the most important input meteorological variable • G-ANFIS achieved better accuracy even in the case of missing data and had higher computational efficiency	(Kisi and Zounemat-Kermani, 2014)

Table 2.3 (continued): Research Studies using the Hybrid Machine Learning Model in ET$_0$ Study

Approach	Key Findings	Reference
ANN (base model) Wavelet-ANN	• Original and decomposed daily T_{max}, T_{min} and u were used to train the models based on PM model in region with semi-arid climate • Decomposition of data reduced the accuracy of prediction • Wavelet ANN which used wavelet function as activation function, performed better than ANN and had better compatibility with the decomposed data	(Falamarzi, et al., 2014)
ANFIS Firefly-ANFIS	• Daily T_{max}, T_{min}, RH_{max}, R_s, u and e_a were used to train the machine learning models based on the PM model in area with tropical climate • Different combinations of input meteorological variables were tested • Accuracy of prediction increased with the number of input meteorological variables • ANFIS-Firefly achieved a better performance as compared to ANFIS	(Tao, et al., 2018)
ANN (base model) Wavelet ANN	• Original and decomposed daily T_{max}, T_{min} and R_a were used to train the models based on HS model in arid, semi-arid, humid and sub-humid regions • Wavelet ANN performed better than ANN with good generalisation over different areas	(Adamala, 2018)

Table 2.3 (continued): Research Studies using the Hybrid Machine Learning Model in ET$_0$ Study

Approach	Key Findings	Reference
ELM (base model) ANN (base model) Online Sequential ELM Wavelet ELM Wavelet ANN	• Decomposed daily T_{max}, T_{min}, R_s and RH were used to train the models based on PM model in semi-arid area • Different combinations of input meteorological variables were tested • Machine learning models integrated with wavelet function generally had better performance • Wavelet ELM and online sequential ELM had the best performance among all other models	(Kisi and Alizamir, 2018)
ANN (base model) RF ELM (base model) Particle Swarm Optimisation-ELM (PSO-ELM)	• The machine learning models with different input combinations were trained based on the ET$_0$ calculated from the PM model, and compared with empirical models with similar inputs • Radiation-based machine learning models had the best performance, followed by temperature and mass transfer models • The integration of PSO with the ELM avoided the latter from using non-optimal network structure, thus improving the generalisation ability as well as accuracy	(Zhu, et al., 2020)

Table 2.3 (continued): Research Studies using the Hybrid Machine Learning Model in ET$_0$ Study

Approach	Key Findings	Reference
Bi-LSTM ANN-Bi-LSTM	• T_{max}, T_{min}, u and sunshine duration were used to predict ET$_0$ up to seven days lead time • It was claimed that the ANN was able to account for the large number of outputs of the Bi-LSTM, subsequently improved the predictions • However, prediction accuracy declined as the lead time increased	(Yin, et al., 2020)
Biogeography-Based Optimisation-ANFIS (BBO-ANFIS) Firefly-ANFIS PSO-ANFIS Teaching-Learning-Based Optimisation-ANFIS (TLBO-ANFIS) Least Squares and Back-Propagation Gradient Descent-ANFIS (LSGD-ANFIS) Entropy Weight Ensemble Coefficient of Variance Ensemble Grey Relational Analysis Ensemble	• Different optimisation algorithms were used to tune the FIS parameters • The optimised ANFIS were assembled using three different weight assignment techniques • Firefly-ANFIS was found to be the most outstanding hybrid model for ET$_0$ estimation • The three ensembles produced very similar results	(Roy, et al., 2020)

Table 2.3 (continued): Research Studies using the Hybrid Machine Learning Model in ET$_0$ Study

Approach	Key Findings	Reference
ANN (base model) ANFIS PSO-ANFIS GA-ANFIS Classification and Regression Tree (CART)	• Monthly temperature, radiation, humidity and wind speed data were used to train the models based on ET$_0$ calculated from the PM model • Hybridised ANFIS (PSO-ANFIS and GA-ANFIS) were found to be more superior • Radiation data was the most critical input to ensure models' accuracy	(Alizamir, et al., 2020)
ANN (base model) Grey Wolf Optimisation-SVM (GWO-SVM) Genetic Algorithm-SVM (GA-SVM) PSO-SVM	• Monthly T_{max}, T_{min}, RH_{max}, R_s, and u were used to train the machine learning models based on the PM model • The machine learning models performed better than empirical models, even with lesser inputs	(Tikhamarine, et al., 2020)
SVM (base model) PSO-SVM Gradient Boosting Decision Tree (GBDT) PSO-GBDT	• Input combinations were selected based on factor analysis, path analysis, logistic regression analysis and stepwise linear regression analysis • The models were trained based on the ET$_0$ calculated from the PM model • Path analysis was found to be the most effective component features selector • PSO managed to improve the performance of SVM and FBDT	(Zhao, et al., 2021)

2.4   Data Requirement and Priority

It should be noted that the completeness of training data and meteorological variables poses a significant impact on the outcome of predictions. This has been clearly indicated in the literature review mentioned in the previous sections. In this section, a detailed analysis on the data requirement will be reviewed. It is believed that the different climatic data will have different weightages and effects in regions that have diverse climatic patterns. Hence, this literature review will be useful to prioritise data when limited resources are available for data collection.

Many past research works compared different input combinations of meteorological variables to select the best input combination for specific model(s). However, a comprehensive study and summary of the input meteorological variables' combinations according to regions or climates has not been discovered. In this literature review, a case study was done on the climate data requirements for arid and semi-arid regions. The reason for choosing arid and semi-arid regions is due to their frequent mention in the literature. Furthermore, these regions represent areas where water resource allocation strategy is imperative and in dire straits. Hussain, et al. (2019) mentioned in their publication that the arid and semi-arid regions are encountering bottlenecks in development due to the water shortage that threatens the food production and security as well as forestry sectors. Therefore, ET-related studies (patterns, trends and estimations approaches) are so ever popular in these

regions and provide important insights and references. Table 2.4 summarises the significant input meteorological variables for accurate machine learning predictions in arid and semi-arid regions.

Table 2.4: Significant Input Meteorological Variables for the Accurate ET$_0$ Estimation by Machine Learning Models in Arid and Semi-Arid Regions

Location	Climate	Significant Input Meteorological Variables	Reference
Shiyang River Basin, China	Arid/Semi-Arid	T_{max}, T_{min}	(Huo, et al., 2012)
Hamedan, Iran	Semi-Arid	T_{mean}	(Tabari, et al., 2012)
Redesdale, Australia	Semi-Arid	T_{max}, T_{min}	(Falamarzi, et al., 2014)
Iraq	Arid/Semi-Arid	T_{max}, T_{min}	(Abdullah, et al., 2015)
Ejina Basin, China	Arid	R_s	(Wen, et al., 2015)
Anatolia, Turkey	Semi-Arid	R_s	(Kisi and Alizamir, 2018)

From Table 2.4, temperature and solar radiation are two major input meteorological variables required by most machine learning models for precise output generation. Since the ET is a process that is strongly linked with the energy flux (Wang and Dickinson, 2012), which includes surface net radiation, sensible heat flux and ground heat flux, therefore, the two aforementioned meteorological variables can provide essential elements and information for the estimation of ET. This finding is also in agreement with the number of temperature and radiation-based empirical models to estimate ET$_0$ such as the HS and the Ritchie models. In other words, the temperature and radiation data

can be deemed as the mandatory meteorological variables that have to be collected for the estimation of ET_0 in arid and semi-arid areas.

Further study in other areas with different climate patterns confirmed the trend in the arid and semi-arid regions. It was realised that in humid areas, temperature and radiation data as well as sunshine duration are useful in inferencing accurate ET_0 using a machine learning model (Feng, et al., 2017a; Feng, et al., 2017b; Traore, Luo and Fipps, 2016). Sunshine hours are usually measured in hours, and it is highly correlated to temperature and radiation data. The finding is expected and does not deviate much from arid and semi-arid regions; hence the importance of temperature and radiation data is justified.

However, when the study is conducted in a region with a Mediterranean climate, wind speed emerges as one of the most influential meteorological variables (Ladlani, et al., 2012). It is believed that the presence of aerodynamic swirls creates turbulence known as eddies that will assist the movement of heat flux and water in the atmosphere (Wang and Dickinson, 2012). In this region where the temperature is relatively low, the significance of wind speed increases as it will exert a greater effect on the ET. Nevertheless, it should be noted that the temperature and radiation data are still of paramount importance and that the wind speed data only help to further increase the accuracy to a satisfactory level; in other words, acting as a complementary meteorological variable (Cobaner, 2011; Ladlani, et al., 2012).

From the literature, a deduction can be made such that only certain meteorological variables are compulsory and essential for the computation or estimation of the ET_0 using machine learning models. The absence of other meteorological variables can be tolerated as most of them only play the role of complementary variables during the estimation process. Hence, a desired model with low climatic data requirement and high accuracy can be realised based on this deduction, which is part of the research objective.

2.5   Data Fusion

According to Meng, et al. (2020), data fusion can be defined as the "technology that merges data to obtain more consistent, informative and accurate information than the original raw data that are mostly uncertain, imprecise, inconsistent, conflicting and alike". As mentioned in the previous section, the ensemble of models is favoured due to the perception that a "committee" could produce a better decision than a standalone model. The formation of the "committee", namely the ensemble learning model, can be done in several ways. It is thought that individual models can sometimes provide accurate results in certain cases, but not for all instances. Combining several models during ensemble learning can overcome this problem, subsequently improving the robustness of the model. The models can complement for each other's weaknesses (Xiao, 2019).

2.5.1 Bootstrap Aggregating

Bootstrap aggregating, also known as bagging was proposed by Breiman (1996). Bootstrap aggregating combines the resampling algorithm of bootstrapping and aggregates the predictions, resulting in a more reliable and unbiased outcome. This can be done by either using the method of simple averaging or the majority vote (depending on whether it is a regression or classification problem). In bootstrap aggregating, independent samples are drawn from the originally available dataset and produce several datasets with equivalent sizes (bags). The originally available dataset has now become an "apparent population". These bags of data would be used to train independent machine learning models, and the outputs of the individual models will be aggregated into a final estimation.

Bootstrap aggregating has been successfully applied in numerous studies in different fields. In the work of Szafranek (2019), bootstrap aggregating was used to forecast inflation. It was claimed that the application of bootstrap aggregation could prevent misspecification bias and overfitting. The results showed that higher prediction accuracy could be achieved when contrasted with the conventional methods by using the bagged samples. On top of that, the bagged MLP used in the research work also showed that the model was more sensitive and could produce meaningful estimations even when the inflation is low. However, the author stressed that bagging did not consider the

effect of the time dimension, resulting in the loss of inference when correlating the effect of time with inflation.

Dantas and Cyrino Oliveira (2018) suggested that the bootstrap aggregating performed resampling technique that resulted in lower correlation among the members of the ensemble. The purpose of having a lower correlation was to ensure that the prediction of each model is independent from one another. The authors used bootstrap aggregating alongside the clustering technique and the exponential smoothing in forecasting a time series. The results agreed with the hypothesis of the authors, where greater accuracy was achieved. However, the forecasting time horizon should be long enough for the ensemble to be useful.

Bootstrap aggregating was also used to predict the thermal comfort of occupants in buildings. It was compared with other traditional methods such as the standalone ANN and the SVM. It was found that the bagging technique that takes the results of several weak learners into account naturally could produce better results than the ANN and SVM. A similar finding was also discovered by Lee, Ahmad and Jeon (2018) and Luo (2018) where the resampling meta-algorithm had successfully enhanced the estimation accuracy.

Likewise, bootstrap aggregating was used as a tool to improve the accuracy of the ET_0 estimation. In the work done by Carter and Liang (2019). Ten machine learning models were compared to estimate the ET_0. It was reported that the bagged decision tree could tolerate with smaller tree size. This

in turn removed the need for the pruning process and became a preventive measure to avoid overfitting. The authors also opined that smaller tree sizes due to bootstrap aggregating provided a more remarkable ability to generalise over wider variations of climate pattern and geographical location.

It should be noted that the bootstrap aggregating involves the resampling of the original dataset and uses the newly generated datasets to train individual models. Then, the results of the individual model will be combined to produce the final decision (via simple averaging or majority vote). Thus, the bootstrap aggregating in fact consists of a data pre-processing algorithm and the ensemble technique. Its reliability in overcoming bias through variance reduction in the estimation is well-recognised and its integration to various models is easy and simple.

2.5.2   Bayesian Modelling Approaches

Model ensembles that utilise Bayes rules are known as Bayesian modelling. According to Höge, Guthke and Nowak (2019), Bayesian modelling can be classified into two main approaches, namely the "winner-takes-all", where the selection of the most relevant model is made, and "team-of-rivals" where the model averaging is done. The selection should be based on the goal and the nature of the problems. For instance, if one is convinced that a true model exists to explain a problem, he/she should consider Bayesian model selection (BMS) to identify the model, provided the true model is present among

the candidates. Otherwise, one can formulate several hypotheses and test the problem with different models. Should he/she does not want to miss out on any of the hypotheses, he/she shall opt for the Bayesian model averaging (BMA) to combine the characteristics or favourable traits of each model.

In the BMS, the "true" model is selected based on the degree of probability that it is true. The value of the probability is updated via Bayesian model evidence. However, the algorithm can be indecisive when two models have similar performances. Bayesian model evidence can be presented in different forms such as model weight ratios and Bayes factor (Kass and Raftery, 1995). Application of the Bayesian model evidence can be difficult when the problem is highly non-linear, requires complex computation or involves high dimensionality (Höge, Guthke and Nowak, 2019).

As for the BMA, the weight factors of the individual model are still maintained. Practically, the BMA is the intermediate phase of the BMS. In the case that one cannot identify the true model, BMA is used to estimate the final output based on the weightage of each model. It should be noted that both the BMA and the BMS aim to search for the final true model, however the approaches are different due to the limitation of data size and set of models (Höge, Guthke and Nowak, 2019). From these two main principles, the Bayesian modelling approach can be highly branched into many distinct algorithms. For instance, the Bayesian joint probability (Zhao, Wang and

Schepen, 2019) and the Bayesian regression (Khoshravesh, Sefidkouhi and Valipour, 2015) and been used on different occasions for ET_0 estimation.

Bayesian modelling approaches have long been applied in fine-tuning the performance of machine learning models. Chen, et al. (2015) performed the BMA on the machine learning models and the conventional empirical models. The ensemble was done using two distinct strategies: (i) ensemble of all models and (ii) ensemble of the best models. The results showed that the best model ensemble produced ET_0 estimation of the highest accuracy at both the regional and global scales. The findings of this study also proved that the accuracy of the ensemble was contributed by its constituent models. The ensemble of the best models would result in better performance, whereas when some of the poor models were included, the accuracy deteriorated.

Bayesian regression was also used to estimate the ET_0 (Khoshravesh, Sefidkouhi and Valipour, 2015). In the study, Bayesian regression was compared alongside with a multivariable fractional polynomial model and robust regression. When the temperature and radiation data were the only meteorological variables fed into the model, the multivariable fractional polynomial model outperformed the Bayesian regression, but the difference was insignificant. Therefore, Bayesian regression still has the potential to be fine-tuned when other meteorological variables are used.

Separately, Zhao, Wang and Schepen (2019) used Bayesian joint probability as their approach to forecast the ET_0 based on the Australian

Community Climate and Earth System Simulator. The authors claimed that the raw forecasting by the global climate model could produce highly biased results. Hence, the authors attempted to integrate the Yeo-Johnson transformation, bi-variate normal distribution and Schaake Shuffle to enhance the model's forecasting ability. The combination of the aforementioned methods represented a form of Bayesian joint probability. The ensemble model was able to forecast ET_0 up to two weeks in advance with satisfactory precision and accuracy.

Despite the fact that the Bayesian model approaches had been utilised several times for the ET_0 prediction and estimation, however, the reports on the details of the ensemble were very limited. For the case of the BMA, the weights of the individual model present in the ensemble remained unknown until He, et al. (2020) presented the results systematically using the Bayesian three-cornered hat method. In the study, the authors adopted different datasets retrieved from remote sensing satellites and land surface models. The influence of the individual datasets on the resultant ET for different seasons and land covers, varied accordingly. It was found that Bayesian-based ensemble could produce better estimation of ET than the simple averaging. The influence of other models on the ET_0 estimation is still not clearly understood. Further investigations shall be carried out following this direction to produce contributing discoveries to the scientific community.

2.5.3   Boosting Algorithm

Boosting is a technique in which prediction accuracy is enhanced by averaging outputs of a few weak learners (Hassan, et al., 2017). Unlike the BMA, the boosting algorithm works stepwise, whereby learners are added one at a time to minimise the cost function. The first learner searches for a solution with optimum loss. Then, the following learners will be included into the ensemble and the residuals of their predecessors are reduced. Numerous variants of boosting algorithms had been proposed, each with their novel characteristics. The most well-known boosting methods include the gradient boosting (Friedman, 2001), adaptive boosting (Freund and Schapire, 1997), XGBoost (Chen and Guestrin, 2016) and CatBoost (Prokhorenkova, et al., 2018).

The application of the boosting algorithm in ET_0 estimation can be found in several studies. Fan, et al. (2018) applied gradient boosting to the decision tree to estimate ET_0. At the same time, the authors also compared the XGBoost and GBDT with the SVM, ELM, M5Tree and RF. Two sets of input meteorological variables, namely (i) the complete set (T_{max}, T_{min}, u, RH and R_s) and (ii) the temperature and radiation-based input set were partitioned to perform the ET_0 estimation. All the models exhibited similar performance. The author opined that the SVM and ELM provided better accuracy and stability whereas the tree-based models offered greater computational efficiency, particularly the XGBoost which showed comparable accuracy with the SVM and ELM.

Another investigation compared the CatBoost model with the RF and SVM (Huang, et al., 2019). The advantage of CatBoost over the RF was that instead of randomly generating a set of predictors, CatBoost generates one predictor after another, taking into account the error of the previous tree. The authors concluded that although the CatBoost did not show significant improvement in terms of accuracy and stability as compared to the SVM, the CatBoost was still highly recommended to minimise computational cost and time.

Based on the findings from the literature, it can be inferred that boosting algorithm could not stand alone as a model itself. This is because it cannot produce estimation with higher accuracy and precision. However, due to its nature that performs a greedy search, it can help to improve integration efficiency and overcome the problem of overfitting. These advantages of boosting algorithm should be considered when constructing a hybrid model.

2.5.4 Ensemble Model for Evapotranspiration

Multiple attempts had been done over the past few years to estimate ET using an ensemble. Zhu, et al. (2016) provided one of the examples. The authors collected data from the ground as well as remote sensing sources and fed the data into four different empirical models: PM model, two-layers Shuttleworth-Wallace model, modified PT model and Advection-Aridity (A-A) model. Data gaps were filled with the mean diurnal variation method. At the end of the

estimation, simple averaging and the BMA were used to ensemble the four models to produce a final output. The results showed that the BMA gave predictions with higher accuracy. However, at the validation stage, the accuracy degraded. The authors suggested that this phenomenon could be due to the inappropriate assumptions of BMA which were incompatible with the ET process. Recently, Nourani, Elkiran and Abdullahi (2019) applied various data fusion strategies, including simple averaging, weighted averaging and non-linear neural ensemble (NNE) using the ANN. Due to the non-linearity of the ET process, the neural ensemble produced the best result among the ensembles. Their work also ascertained that data assimilation techniques could generate a similar effect when applied to both machine learning models and empirical models.

Ensemble learning is a common approach when dealing with remote sensing data and using land surface models. This is due to the high uncertainty in each individual land surface models. The wide variations of land surface models adopting different assumptions and variables during their formulations and increases the uncertainties. Hence, multiple land surface models must be used simultaneously to backup each other so that the performance can be improved. In the work of Liu, et al. (2016), multiple remote sensing datasets from Princeton, Institute of Tibetan Plateau Research and Qian were collected. These datasets became the input of four land surface models, namely Biosphere-Atmosphere Transfer Scheme (BATS), Variable Infiltration Capacity (VIC), Community Land Model (CLM) 3.0 and CLM 3.5. The outputs were compared

with observed data from FLUXNET. The authors claimed that after performing

data assimilation on the four land surface models, the resultant ensemble model

had promising outputs of higher spatial and temporal resolution in a longer time

scale. The authors also suggested that the utilisation of the Bayesian modelling

approach could further reduce the uncertainties.

Nowadays, the application of data fusion on remote sensing, multi-

sensor and land surface model is prominent. It was known that the application

of data fusion could produce ET estimation at different regional scales. Apart

from the previously mentioned literature, several other research works were also

done on the same basis using different ensemble models. Table 2.5 summarises

the data fusion methods and the models involved in remote sensing-based ET_0

estimations.

Table 2.5: Data Fusion Methods and the Models Involved in Remote Sensing-Based ET$_0$ Estimations

Data Fusion Methods	Models	Data Sources	Reference
STARFM	ALEXI DisALEXI	Landsat MODIS	(Cammalleri, et al., 2013; Cammalleri, et al., 2014; Knipper, et al., 2018; Semmens, et al., 2016)
STARFM	SEBS	ASTER MODIS	(Li, et al., 2017)
Simple Taylor Skill	RS-PM Shuttleworth-Wallace PT-JPL Modified PT SIM	Landsat	(Yao, et al., 2017)
Ensemble Kalman Filter	Distribution Time Variant Gain Model	MODIS	(Zou, et al., 2017)
ESTARFM	SEBS PM model	Landsat MODIS	(Ma, et al., 2018)

* ALEXI: Atmospheric – Land Exchange Inverse, DisALEXI: Disaggregated Atmospheric-Land Exchange Inverse, SEBS: Surface Energy Balance, RS-PM: Remote Sensing Penman-Monteith, PT-JPL: Priestley-Taylor-Jet Propulsion Laboratory, SIM: Simple Hybrid ET, STARFM: Spatial and Temporal Adaptive Reflectance Fusion Model, ESTARFM: Enhanced STARFM

2.5.5 Metaheuristic Approach

The development of machine learning models requires fine-tuning of the hyper-parameters so that the resultant models can perform even better in terms

of accuracy and generalisation ability. This is due to the fact that many training algorithms (such as the back-propagation) only involve the adjustment of weights and biases which are case-specific. The fine-tuning process can be time-consuming, especially when the search space (possible range of solution) is huge. Taking the ANN as an example, the numbers of neurons in the hidden layers range from one to infinity, not to mention that a similar case could also happen when determining the number of hidden layers. Grid search and trial-and-error methods appear to be impractical when dealing with extremely large search space. Therefore, in recent years, the rapid diversification of the metaheuristic approaches has been observed. In essence, the metaheuristic approach refers to the formulation of objective function(s) to represent the problems of interest. Subsequently, a set of algorithm is used to minimise or maximise the objective function(s) in order to obtain an optimum solution with respect to the constraints and conditions provided (Yang, Bekdaş and Nigdeli, 2016).

The metaheuristic approach can be classified into several categories: evolutionary computing, swarm intelligence and iterative-based algorithms (Goh and Lee, 2019). Evolutionary computing involves the design of a symbolic regression to fit the input dataset through techniques such as genetic programming and genetic algorithms. According to Hebbalaguppae Krishnashetty, et al. (2021), genetic programming supports simultaneous searching to perform fine-tuning and represent the problems as a linear program. The evolution involves the detection of feature importance (analogous to fittest

genes) in the dataset and the cross-mutation that occurs in each iteration. Jing, et al. (2019) performed an extensive review on the implementation of evolutionary computing in ET-related research studies. They discovered that the lack of variability has caused evolutionary computing to lag other metaheuristic approaches, such as swarm intelligence and iterative-based algorithms.

The swarm intelligence, on the other hand, approaches the optimisation problems based on the perception that the intelligence can be obtained via the interactions among the individuals within a population or swarm (Janga Reddy and Nagesh Kumar, 2020). This is the imitation of the civilisation of the human society, where the development and growth become more rapid as the interactions among individuals (exchange of information) increase. Since the development of swarm intelligence is generally nature-inspired, hence many of the swarm-based optimisation algorithms attempt to imitate the behaviour of different organisms, such as the PSO, ant colony optimisation (ACO), artificial bee colony (ABC) and so on (Dorigo and Blum, 2005; Karaboga and Basturk, 2007; Kennedy and Eberhart, 1995). In recent years, the integration of swarm-based optimisation algorithms in machine learning for ET estimating applications has been increasing drastically. This, on the one hand, is due to the efficiency of the swarm intelligence in tackling various optimisation problems (Pham, et al., 2021), at the same time, the variation and diversity of swarm intelligence are still increasing on a year-on-year basis. Many new swarm-based optimisation algorithms had been introduced in the past few years (Rostami, et al., 2021), not to mention the improvements and modifications to existing

algorithms (Chia, et al., 2022). Despite swarm intelligence being a popular method among the researchers, Tang, Liu and Pan (2021) still reported the limitations of swarm intelligence in their publication. It was claimed that due to the deployment of high number of search agents in a large search space, the computational cost of swarm intelligence can be incredibly high, which would be unnecessary for simple problems. Besides, old algorithms are more likely to converge prematurely. This issue is being addressed by many developers of swarm-based optimisation algorithms during the recent extension of the swarm intelligence class (Mirjalili and Lewis, 2016).

Unlike the evolutionary computing and swarm intelligence, studies conducted on iterative-based algorithms are lesser. This could be due to the recent trend that focusses on the studies related to swarm intelligence. Essentially, the iterative-based optimisation works on the improvement of the objective function(s) via the neighbourhood search technique (Şen, Dönmez and Yıldırım, 2020). Some examples of the iterative-based algorithm include the simple annealing (SA) and black hole algorithm (Hatamlou, 2013; Kirkpatrick, Gelatt and Vecchi, 1983). Iterative-based algorithm for machine learning application, particularly those related to ET estimation is still very limited and could potentially be a promising hybridisation technique in the near future.

The use of metaheuristic approaches involves the improvement of the objective function(s), which is the mathematical representation of the problems of interest. Over the years, the mean square error (MSE) has been used

extensively as the objective function for machine learning-related applications, in which the hyper-parameters of machine learning models shall be tuned to minimise the error of predictions. Nevertheless, modern metaheuristic approaches can handle more than one objective function concurrently, resulting in the so-called multi-objective optimisation. This method has been deployed in many hydrological studies and yielded positive results. For instance, Yadav, Chatterjee and Equeenuddin (2021) used the genetic algorithm to optimise the ANN for suspended sediment yield modelling with the error variance and bias acting as competing for objective functions. The selection of objective functions had a significant impact on the performance of the ANN in which the overfitting and bias could be resolved. The successful adoption of multi-objective optimisation using metaheuristic approaches in hydrology implies that similar findings can also be replicated for ET-related problems. This can be one of the research directions for future studies.

The metaheuristics approach (mostly single-objective optimisation) has been applied for ET modelling in the past few years. The optimisation algorithms, regardless of the class were used to tune the hyper-parameters of the machine learning models. Some of the important findings have been summarised in Table 2.3 of Section 2.3.

2.6 Summary and Research Rationale

Chapter 2 provides a detailed discussion on ET and its observation/estimation methods, application of machine learning models in estimating ET_0, data requirement of machine learning models for ET_0 estimations as well as study of data fusion or ensemble models to compute ET_0. Individual empirical and machine learning models for ET_0 estimation were studied and reported based on previous research works. Furthermore, a comprehensive analysis on the types of datasets is presented. A case study was performed in arid and semi-arid regions to identify the priority ranking of input meteorological variables. Apparently, it was established that solar radiation and temperature emerged as the two most important factors for the accurate and precise ET_0 estimation. Numerous ways for integrating data fusion techniques on base machine learning models were also studied and discussed.

Careful identification of the research gap or rationale is crucial to continue this research work. In this context, several gaps were deemed to be filled. Firstly, past studies have shown that areas with different climate patterns could have different priority rankings of meteorological variables. However, a comprehensive study in equatorial climate has been lacking. Malaysia, which largely depends on agricultural production becomes an interesting area of study due to its equatorial Monsoon climate. Therefore, it is imperative to determine the most influential meteorological factor on ET_0 in such a region as an effort

to cut down the number of meteorological variables that have to be monitored for ET_0 estimation.

Machine learning models' accuracy in ET_0 estimation would be impaired in the case of limited input meteorological variables, which is undesirable. Therefore, the second research gap identified is the lack of a robust machine learning model that is resilient towards the reduction of input meteorological variables. In view of the situation, data fusion or ensemble model is proposed to solve the encountered problem by resampling or combining effects of multiple models. In fact, data fusion has not been widely used to assemble machine learning models with ground observation data. This gap could be filled with this study where different techniques of ensemble models are compared to enhance the performance of base machine learning models.

Thirdly, the literature review reveals that machine learning models need to be trained locally for local use. This means that a spatially robust model for ET_0 estimation is still absent. Hence, a model with broad spatial applicability besides having lower data requirements would be advantageous. This means that data need not to be collected for a long period before proceeding with the modelling work. The outlined research gaps align well with the problem statement mentioned in Section 1.2, in which the qualitative and quantitative hungers of the machine learning models need to be addressed and resolved.

CHAPTER 3

METHODOLOGY

3.1   Research Flow Chart

The aim of this research work is to develop a robust machine learning model for the estimation of ET_0 with minimum data requirements for Peninsular Malaysia. Meteorological data (from 1st January 2000 to 31st December 2019) were obtained from the Malaysian Meteorological Department (MMD). After performing suitable data pre-processing (cleaning of corrupted data, data normalisation), the meteorological data were used to train three base machine learning models, namely the MLP, the SVM and the ANFIS. The k-fold cross-validation was applied to ensure that the models were not biased and overfitted. The optimum input combinations were selected as an effort to reduce the qualitative hunger of the machine learning models (Objective (i)).

Then, data fusion techniques were employed to improve the performance of the base machine learning models. The data centric bootstrap aggregating was used to resample the original dataset and the bagged datasets were used to train the base models, resulting in the bagged MLP, SVM and ANFIS. On the other hand, the model centric BMA and black-box NNE combined the ET_0 estimations of the base models in an attempt to produce better

estimations. The performance of the base and hybrid models was assessed using several performance evaluation metrics, such as the MAE, RMSE, MAPE, R^2, MBE and GPI. Up to this point, all the base and hybrid models were trained and tested locally (Objective (ii)). This scenario was named as Scenario 1.

Subsequently, the best models selected at each station were used as exogenous models and tested elsewhere. This was known as Scenario 2. This step was necessary to confirm the spatial robustness of the developed models by checking their performance at other stations (Objective (iii)). At the end of this stage, a dominant machine learning model was selected. This dominant model was trained using pooled global data (Scenario 3) and tested across the whole Peninsular Malaysia (Objective (iv)). Figure 3.1 shows the workflow of the overall investigation, including methods performed in this present study. The analysis of this study is done using the MATLAB R2016a platform. All the codes used in this study are either self-developed or modified from the open-source codes.

Data Acquisition and Pre-Processing
- data cleaning
- calculation of ET_0 using PM equation
- data normalisation

Data Ready for Modelling

Objective (i)

Development of MLP, SVM and ANFIS
- identify optimum input combinations

Bootstrap Aggregating

Bayesian Model Averaging (BMA)

Non-Linear Neural Ensemble (NNE)

Objective (ii)

BMLP **BSVM** **BANFIS**

BMA-E

WOA-ELM-E

Base MLP, SVM and ANFIS

Selection of Best Local Models

Objective (iii)

Hybrid BMLP, BSVM, BANFIS, BMA-E and WOA-ELM-E

Assess Robustness of Local Models at External Stations

Most Robust Model

Objective (iv)

Assess Robustness of Dominant Model at all Stations using Pooled Global Data

Scenario 1

Scenario 2

Scenario 3

Figure 3.1: Overall Workflow of the Research Study

70

3.2 Meteorological Station

A total of 12 meteorological stations, which represent around 132265 km^2 of Peninsular Malaysia were selected to be included in this study. Despite the fact that including more stations in the investigation can refine the quality of the results, however, this study is limited by the amount of data acquired from the administration. Nevertheless, the ratio between the area of Peninsular Malaysia and the number of meteorological stations in this study fulfilled the minimum requirement by the World Meteorological Organisation (WMO) (Chacon-Hurtado, Alfonso and Solomatine, 2017). As reported by Fatchurrachman, et al. (2022), Peninsular Malaysia has a tropical rainforest climate (*Af*) according to the Köppen climate classification. Table 3.1 shows the geographical characteristics of all the meteorological stations and their exact locations are shown in Figure 3.2. Throughout this thesis, the stations will be mentioned in the form of "Station *Station_ID* (*Station Name*)". For example, the first station in Table 3.1 will be Station 48600 (Pulau Langkawi). The description of the data obtained is presented in Section 3.3.

Table 3.1: Geographical Characteristics of the Selected Stations

Station ID	Station Name	Latitude	Longitude	Elevation (m)
48600	Pulau Langkawi	6°20' N	99°44' E	6.4
48601	Bayan Lepas	5°18' N	100°16' E	2.5
48603	Alor Setar	6°12' N	100°24' E	3.9
48615	Kota Bharu	6°10' N	102°18' E	4.4
48620	Sitiawan	4°13' N	100°42' E	6.8
48623	Lubok Merbau	4°48' N	100°54' E	77.5
48625	Ipoh	4°34' N	101°06' E	40.1
48632	Cameron Highlands	4°28' N	101°22' E	1545.0
48647	Subang	3°08' N	101°33' E	16.6
48649	Muadzam Shah	3°03' N	103°05' E	33.3
48650	KLIA	2°44' N	101°42' E	16.1
48657	Kuantan	3°46' N	103°13' E	15.2
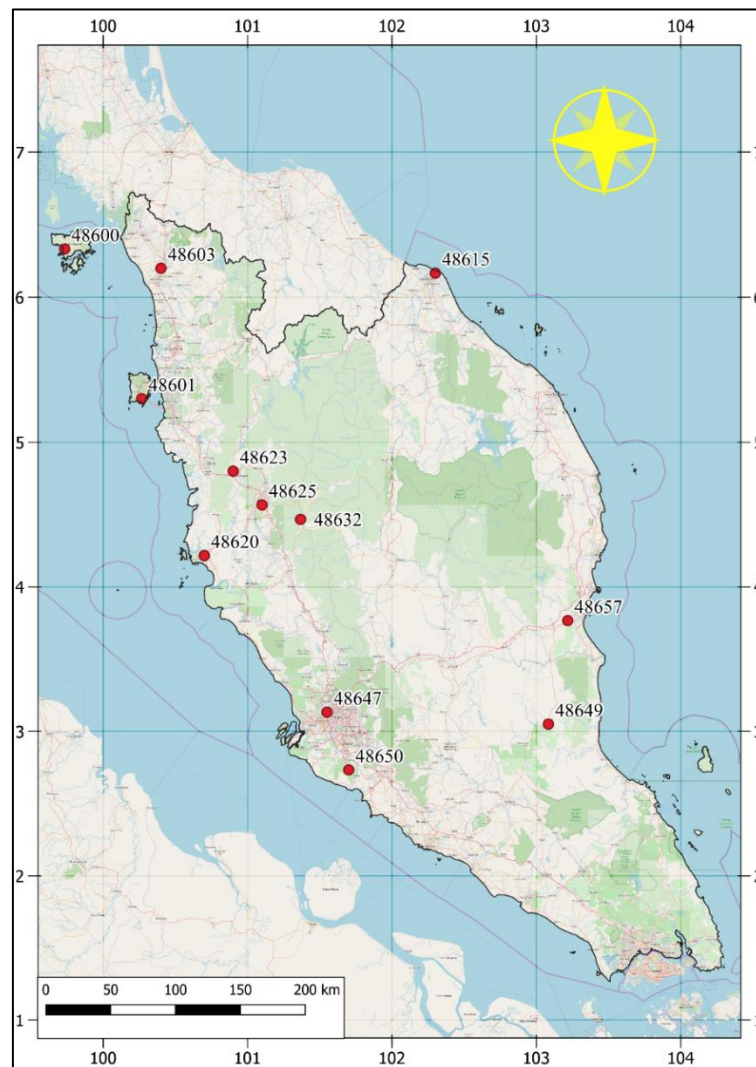
Figure 3.2: Locations of the Selected Stations

3.3 Data Pre-Processing

Data pre-processing is crucial to ensure that the data used to train the machine learning models can meet the necessary requirements. For instance, corrupted or absurd data points need to be removed or replaced; data have to be normalised to remove the scaling effects; and so on. Otherwise, the "garbage in, garbage out" situation may happen during the training process. This section outlines the pre-processing steps performed in this study.

3.3.1 Data Acquisition

Meteorological data between 1st January 2000 and 31st December 2019 were acquired from the Malaysian Meteorological Department (MMD). The dataset comprises of daily maximum, minimum and mean temperature (T_{max}, T_{min} and T_{mean}), daily mean relative humidity (RH), daily mean wind speed (u) and daily solar radiation (R_s), as tabulated in Table 3.2.

Table 3.2: Details of Data Obtained from MMD

Data	Time Step	Unit
Maximum Temperature (T_{max})	Daily	°C
Minimum Temperature (T_{min})	Daily	°C
Mean Temperature (T_{mean})	Daily	°C
Mean Relative Humidity (RH)	Daily	%
Mean Wind Speed (u)	Daily	m/s
Solar Radiation (R_s)	Daily	MJ/m^2

3.3.2 Penman-Monteith Model

To train the machine learning models, a standard target must be fed into the model. This is also known as supervised learning. In the case of ET_0 prediction, the PM model is deemed to be the standard target as recommended by the Food and Agriculture Organisation of the United Nations in FAO56 (Allan, et al., 1998). The complete equation of the PM model is shown by Equation (2.4). The inputs needed for ET_0 calculation were computed using the meteorological data obtained from the MMD. The results were compared with ET_0 values obtained from literature to ensure the correctness of computation steps.

3.3.3 Normalisation of Data

In the present study, normalisation of data was performed to rescale all the input meteorological variables to achieve efficient, unbiased and accurate estimation via the removal of the scaling effects. The purpose of rescaling was to bind the input meteorological variables of different ranges between 0 and 1, corresponding to the minimum and maximum values, respectively. Typically, the normalisation can be determined using Equation (3.1).

$$X_{norm} = \frac{X - x_{min}}{x_{max} - x_{min}} \hspace{3cm} (3.1)$$

where:

X_{norm} = normalised data vector

X = original data vector

x_{max} = maximum value in $\boldsymbol{X_{norm}}$

x_{min} = minimum value in $\boldsymbol{X_{norm}}$

3.3.4   K-Fold Cross-Validation

To produce results with lower bias or variance, the k-fold cross-validation was performed to ensure the generalisability of the model in this present study. This can be done by partitioning the original dataset into k equal parts. The model was trained with (k – 1) parts of the data while it was tested using the remaining unseen partition. The k-fold cross-validation allowed all the data points in the dataset to have the opportunity to become the training data as well as the testing data. As such, the models developed were trained and tested with data that had higher diversity and could cater different scenarios.

3.3.5   Input Combinations

A total of six meteorological variables, including T_{max}, T_{min}, T_{mean}, RH, u and R_s were obtained from the MMD. Hence, 63 possible input combinations were formed from the six meteorological variables. All the possible input combinations were studied to provide exhaustive preliminary results and avoid the possibility of missing out on any potential candidates for optimum input combinations. The nomenclatures and details of the 63 possible combinations are summarised in Figure 3.3.
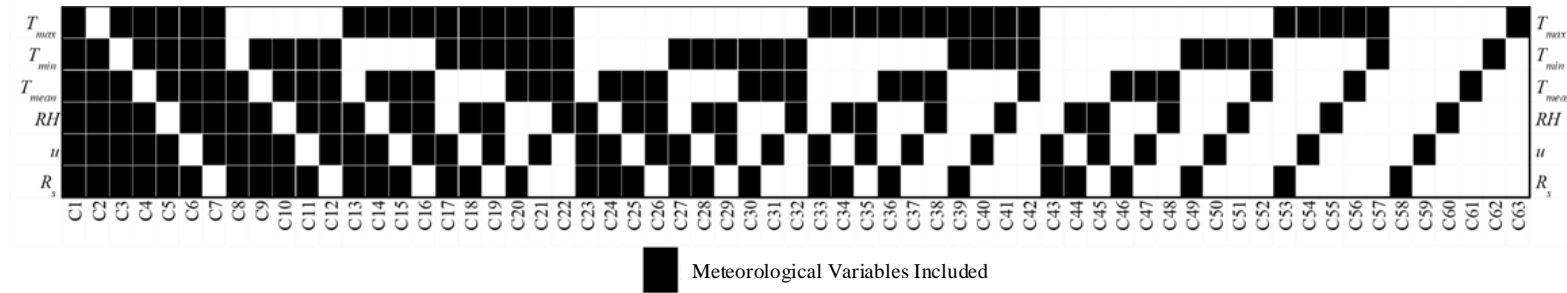
Figure 3.3: 63 Input Combinations from the Six Meteorological Variables

3.4   Base Model Training

Three base models were studied in this research work, namely the multilayer perceptron (MLP), the support vector machine (SVM) and the adaptive neuro-fuzzy inference system (ANFIS). These models were selected because they were the most common base machine learning models used for ET_0 estimation. Base models were selected as they required minimal computational cost, which could subsequently improve the efficiency of any hybrid models built on these foundations. Although tree-based models are also simple, however, they were not included as their binary splitting nature was more suitable for the classification analysis. On top of that, performing regression analysis with tree-based models on a large amount of data would result in massive trees that could overfit and require pruning, which could be tedious. Hence this study would only proceed with the MLP, the SVM and the ANFIS. The technical details of these machine learning models are provided in the following subsections.

3.4.1   Multilayer Perceptron

The MLP consists of several layers of neurons, including the input layer, the hidden layers and the output layer. The layers of neurons are connected to one another via artificial synapses known as weights. Each neuron (except for the neurons in the input layer) accepts signals from its precedent neurons. They are then excited by an activation function and will decide whether to transmit

signal to its subsequent neurons or not. The general mathematical representation of the MLP is shown in Equation (3.2).

$$Y = f\left(\sum_{i=0}^{n} w_{ij}X + b_i\right) \qquad (3.2)$$

where:

Y = output vector

X = input vector

w_{ij} = weight connecting i^{th} input to j^{th} neuron of the hidden layer

b_i = i^{th} bias term

f = activation function

The hyper-parameters of the MLP have to be fine-tuned before performing the simulation. The tuning process was carried out using the trial-and-error method. A general structure of MLP was adopted for all the stations and input combinations with the following heuristics:

- Single hidden layer to simplify the network structure
- There was always one extra neuron in the hidden layer as compared to the input layer
- Activation function for the MLP was the sigmoid function due to the range of output normalised between 0 and 1
- The Levenberg-Marquardt algorithm was used as the training algorithm

3.4.2 Support Vector Machine

The SVM works by transforming the available data into a feature space using suitable kernel functions. Subsequently, the relationship between the inputs and the output can be mapped within the feature space through high-dimensional regression analysis. In general, Cortes and Vapnik (1995) provided the mathematical expression of the SVM and can be expressed in Equation (3.3).

$$Y = w\varphi(X) + b \tag{3.3}$$

where:

Y = output vector

w = weight vector

X = input vector

b = bias term

φ = kernel function

Similar to the MLP, the hyper-parameters used for the training of the SVMs were determined by the trial-and-error method. Specifically, the two hyper-parameters tuned in this investigation were the box constraint as well as the epsilon (allowable error) of the SVM. By specifying the initialisation of the two hyper-parameters, the other hyper-parameters of the SVM were optimised by sequential minimal optimisation (SMO). The optimum box constraint should be able to cover at least 50 % of the target data and was set to be as wide as the normalised value of the interquartile range of the target ET_0. On the other hand, the epsilon was set to be 10 times narrower than the box constraint. Both values

were obtained through a trial-and-error method. The radial basis function (RBF) was selected as the kernel function in this study, which was suggested by numerous literatures that performed a similar investigation (Fan, et al., 2018; Ferreira, et al., 2019). The RBF function is shown in Equation (3.4).

$$K(x_n, x_i) = exp(-\gamma||x_n - x_i||^2 + C_1) \qquad (3.4)$$

where:

x_n = nth term of the input vector

x_i = ith term of the input vector

γ, C_1 = tuneable hyper-parameters

With the kernel functions selected, the SVM can be optimised by tuning the weights and biases in order to minimise the loss function shown in Equation (3.5) and Equation (3.6).

$$R = 0.5||w||^2 + C_2 \sum_{i=1}^{n} L_\varepsilon(y_{estimated}, y_{actual}) \qquad (3.5)$$

$$L_\varepsilon(y_{estimated}, y_{actual}) = \begin{cases} 0, if \ |y_{estimated} - y_{actal}| < \varepsilon \\ |y_{estimated} - y_{actal}|, otherwise \end{cases} \qquad (3.6)$$

where:

R = loss function

$0.5||w||^2$ = regularisation term

C_2 = penalty parameter

ε = margin of SVM

L_ε = ε-insensitive error function

$y_{estimated}$ = estimated value of ET_0

y_{actual} = actual value of ET_0

3.4.3 Adaptive Neuro-Fuzzy Inference System

The ANFIS adopts the ANN methodology for the tuning of the fuzzy logic. The ANFIS is similar to the MLP, where the layers are connected by weights. In this study, the subtractive clustering method was embedded into the ANFIS for optimisation purposes. In comparison with the grid partition method, the subtractive clustering method is capable of handling the data by computing suitable clusters to minimise the complexity of the problem as a whole. Consequently, the modelling efficiency of subtractive clustering ANFIS can be improved. The Sugeno fuzzy rule was used in this study, which can be expressed in the following form.

Rule 1: If x is A_1 and y is B_1, then $f_1 = p_1x + q_1y + r_1$

Rule 2: If x is A_2 and y is B_2, then $f_1 = p_2x + q_2y + r_2$

The values of p_i, q_i and r_i can be determined by black-box operation during the training process. In other words, the purpose of training the ANFIS is to explore the correct membership functions for the fuzzification of the inputs and then transform the inputs into targeted output via the optimised fuzzy rules.

Since the Sugeno type ANFIS integrated by subtractive clustering method was used, the number of membership functions need not to be specified as it was determined by the algorithm after the computation of clusters. The membership functions and fuzzy rules were optimised through back-

propagation. However, the Gaussian membership function was chosen for the ANFIS to resemble the structure of the SVM.

3.5 Data Fusion

According to Meng, et al. (2020), data fusion can be defined as "technology that merges data to obtain more consistent, informative and accurate information than the original raw data that are mostly uncertain, imprecise, inconsistent, conflicting and alike". Data fusion is the currently trending hybridisation technique to improve the estimating performance of the base models. It incorporates algorithms to complement the weaknesses of the base models so that their performances can be boosted. This investigation applied three data fusion techniques to the base models. The data fusion techniques selected were aimed at different possible pitfalls of the base models. The three techniques are the bootstrap aggregating (or bagging), the Bayesian model averaging (BMA) and the non-linear neural ensemble (NNE).

3.5.1 Bootstrap Aggregating

The bootstrap aggregating, commonly known as bagging, is an algorithm that involves the resampling of data for training multiple base models. The bootstrap aggregating was chosen as the data centric approach as this data fusion technique would alter the structure of the original dataset before it is used

to train the individual models. As such, the effect of data structure can be evaluated. At the end of the training, the estimations of the base models are aggregated by simple averaging to reduce the bias of the training sets. Strictly speaking, the bootstrap aggregating alters and diversifies the structure of the original training data so that the aggregated model can be exposed to different scenarios. In other words, the models developed using this data fusion technique are data centric models. For the purpose of this study, ten bags of data were resampled from the kth training set to produce ten identical models learning using different bags of data. The ten models were aggregated by simple averaging. The mathematical expression of the bootstrap aggregating is provided as shown in Equation (3.7) (Breiman, 1996).

$$y = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad (3.7)$$

where:

y = aggregated estimation

y_i = estimation of ith model

n = number of bags/models

3.5.2   Bayesian Model Averaging

Each model has its way of performing estimation and that would result in different accuracies as the intrinsic theory varies. The difference in performance can be viewed as the correctness or the appropriateness of the model for a particular estimation task. The BMA takes advantage of such characteristics and combines the estimations of different models by assigning

weights to the models based on their correctness. The BMA believes that there exists a true model that can explain the phenomenon (estimation of ET_0 for this study). If this perfectly true model does not exist, it can be created by computing the weighted average of the outputs produced by different models. The weights assigned by the BMA algorithm are determined by calculating the probability distribution function of the models. In other words, this data fusion technique aims to improve the estimation by looking for a better model instead of altering the data structure as seen in the bootstrap aggregating. This is also known as the model centric approach. To obtain the weights of the models participating in the end hybrid model, firstly, the posterior probabilities of the models for each combination can be calculated using Equation (3.8), as proposed by Kass and Raftery (1995).

$$P(M_i|y) = \frac{P(y|M_i)P(M_i)}{\sum_{j=1}^{2^n} P(y|M_j)P(M_j)} \qquad (3.8)$$

where:

$P(M_i/y)$ = posterior model probability of ith model

$P(M_i)$ = prior model probability of ith model

$P(y/M_i)$ = marginal likelihood of ith model

$P(y/M_j)$ = marginal likelihood of model with j[th] combination

$P(M_j)$ = prior model probability with j[th] input combination

n = total number of meteorological variables

The weights of the constituent models can be obtained by normalising the posterior model probabilities of the models that are subjected to the same

condition (same input) as shown in Equation (3.9). The sum of the weights of the models with the same training condition should be unity. Subsequently, the averaged ET_0 estimation can be obtained using Equation (3.10).

$$w_i = \frac{P(M_i|y)}{\sum_{i=1}^{n} P(M_i|y)} \tag{3.9}$$

$$y = \sum_{i=1}^{n} w_i y_i \tag{3.10}$$

where:

w_i = weight of ith model

n = number of models with the same training condition (same input combination)

y = estimation after data fusion

y_i = estimation of ith model

The BMA can be converted to the Bayesian model selection (BMS) by naively selecting the model with the highest posterior probability.

3.5.3 Non-Linear Neural Ensemble

The third data fusion approach used in this study is the NNE. The NNE does not have any statistical basis unlike the bootstrap aggregating and the BMA. The individual machine learning models are hybridised via black-box operation through a neural network, in contrast to the conventional statistical way of performing data fusion. In a preliminary investigation of this research work, it was found that the stochastic enabled ELM using the whale optimisation algorithm (WOA) had the best performance when it comes to ET_0 estimation,

where the accuracy of the estimation and the time cost were optimal (Chia, Huang and Koo, 2021b). Therefore, the ELM integrated with WOA (WOA-ELM) was utilised to hybridise the MLP, SVM and ANFIS to develop an ensemble. The mathematical expression of the ELM, as proposed by Huang, Zhu and Siew (2006), is shown in Equation (3.11).

$$Y = h(X)\beta \qquad (3.11)$$

where:

Y = output vector

$h(X)$ = sum of output from each hidden neurons fed with input vector

β = bias vector

An ELM consists of only one layer of hidden neurons, with all the hyper-parameters initialised at random. The absence of stochastic training in the algorithm of the ELM increases the risk of the model converging to various local optima instead of the desired global optimum. Hence, the WOA algorithm was used to complement this disadvantage and provide a continuous improvement mechanism for the base ELM. This can help to converge the model to the global optimum by increasing the iteration steps in the optimisation algorithm. The reason for selecting the WOA from all the swarm intelligence was because only three parameters (logarithmic spiral constant, number of iterations, distance between whales) need to be adjusted throughout the whole optimisation process. This number was considered low as compared to other swarm-based optimisation algorithms. Besides, the WOA had been tested on 29 different test functions to prove its stability, making it one of the most tested

optimisation algorithms for engineering applications. The detailed comparison of WOA with other optimisation algorithms can be found in a review by Johnvictor, et al. (2020). Equation (3.12) to Equation (3.14) proposed by Mirjalili and Lewis (2016) show the steps to search and update the position of the global optimum (target or prey).

$$X_{t+1} = X_{rand} + A|CX_{rand} - X_t| \tag{3.12}$$

$$A = 2ar - a \tag{3.13}$$

$$C = 2r \tag{3.14}$$

where:

X = position vector of the search agents (whales)

A, C = vectors of coefficient

r = randomised vector between 0 and 1

a = shrinking vector from 2 to 0 linearly throughout the iteration process

In order to capture the prey, the search agents would shrink their bubble net, and this will only happen when $|A|$ is sufficiently small. The best position of the global optimum will be updated using Equation (3.15).

$$X_{t+1} = \begin{cases} X_t^* - A|CX_t^* - X_t|, when\ p < 0.5 \\ |X_t^* - X_t|e^{bL} \cos(2\pi L) + X_t^*, otherwise \end{cases} \tag{3.15}$$

where:

X_t^* = best position vector at iteration t

b = shape parameter of the bubble net

L = spiral coefficient

p = randomised value between 0 and 1

The random parameter p was used to imitate the natural behaviour of the whales in their marine habitat. The piecewise function shown in Equation (3.15) assigns a 50:50 chance for the whales to attack their prey. If the whales are not exploiting for the prey (approaching the global optimum), then they will continue to encircle the prey and wait for the next chance (exploration phase). By incorporating the WOA to the ELM, it helps to optimise the hyper-parameters within the ELM model so that they are better suited for the hybridisation of multiple models.

To improve the position of the prey after each iteration, a fitness function has to be evaluated to assess the goodness of the current position. Numerous fitness functions have been developed to assess the fitness from different perspectives, including the mean square error (MSE) and Taylor's skills score (TSS). In this investigation, a fitness function that evaluates the positions from multiple aspects was selected as shown in Equation (3.16). Equation (3.16) was designed such that it had the advantage of a single-objective optimisation (efficient computation) and also multi-objective optimisation (involves competing aspects of a model). It incorporated the competitive nature of different metrics to achieve the optimum balance between accuracy and generalisability.

$$\text{Fitness} = (\text{MAE} + \text{RMSE}) \times (1 - R^2) \qquad (3.16)$$

where:

MAE = mean absolute error

RMSE = root mean square error

R^2 = coefficient of determination

The role of the optimisation algorithm is to minimise the fitness function. By reaching that objective, as shown in Equation (3.16), the mean absolute error (MAE) and root mean square error (RMSE) would converge to the minimum point, whereas the coefficient of determination (R^2) would approach a maximum value of 1. The sum of MAE and RMSE are taken instead of their product due to the fact that both of these metrics measure the deviation of estimated ET_0 from the actual ET_0 and exist in the same dimension as each other (Chia, Huang and Koo, 2021b). The working mechanism of the WOA-ELM is shown in Figure 3.4.
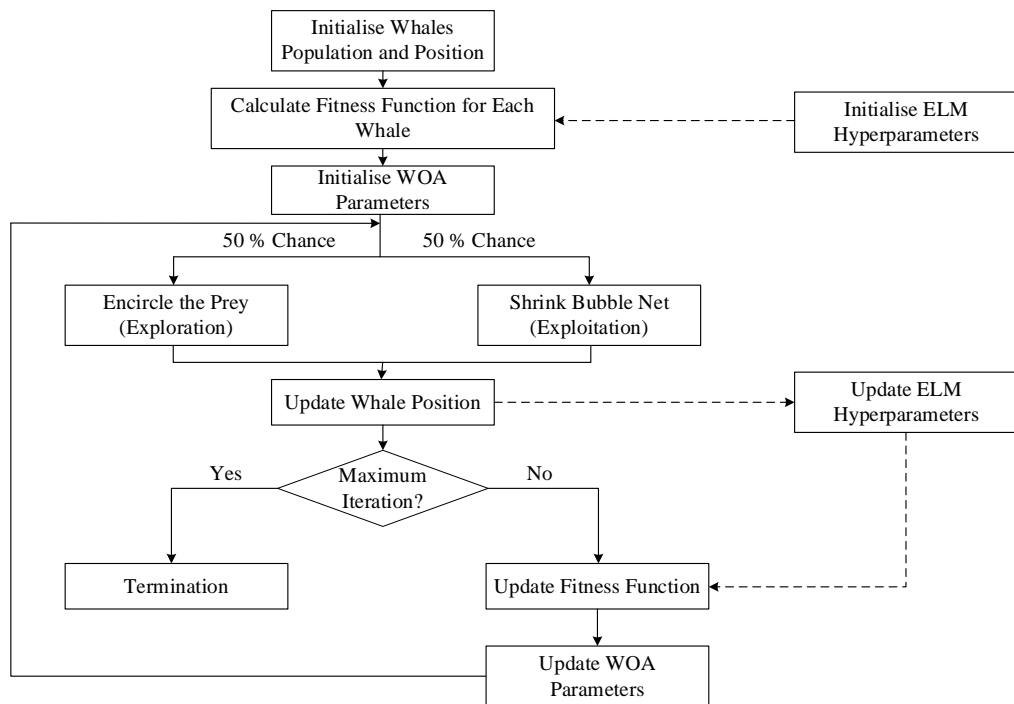
Figure 3.4: Mechanism of the WOA-ELM

3.6   Training Scenario

Various training scenarios were designed to assess the robustness of the developed models under different circumstances. Specifically, the developed model should be less data-hungry and adapt well when exposed to different data or deployed in different areas. The three proposed simulating scenarios adopted are explained in this section. Note that the k-fold cross-validation was performed in each of the scenarios. Hence, the training and testing set ratio was set to be 9:1 for each fold, where all data points had the opportunity to be the testing data (in rotation).

3.6.1   Scenario 1: Training and Testing with Local Data

The MLP, SVM and ANFIS were trained using local data in accordance with the k-fold cross-validation described in Section 3.3.4. All the available data in one station were partitioned into k (10) folds with one of the partitions being used for testing purposes, while the others were used for training, in rotation. Subsequently, the data fusion techniques were also applied accordingly to the data from respective stations to train and test the hybrid models locally.

3.6.2   Scenario 2: Estimation of ET$_0$ using Exogenous Models

To assess the robustness of the models in a comparatively global condition, trained and tested models at all the stations were deployed at different

stations. For example, the model trained at Station 48600 (Pulau Langkawi) was used to estimate the ET_0 at Station 48601 (Bayan Lepas) and other stations. Due to the large number of possible models, only the best models selected at one station were used for further testing. The models were selected from a list of candidates which include different base and hybrid models trained with various combinations of input meteorological variables. Six models were chosen at every station, representing the six different number of input meteorological variables.

The rationale of designing Scenario 2 was one of the efforts to counter the lack of data or infrastructures in certain regions. The readiness of the exogenously trained model to be used at another station would be desirable for eliminating the training phase at a new area (station) and the model is deemed to be less data hungry.

3.6.3 Scenario 3: Model with Pooled Global Data

As an initiative to combat the lack of data and infrastructures, Scenario 3 was designed with the purpose of increasing the amount of data by forming a pool of global data. In other words, the models will not be trained independently at every station with their own local data. In fact, the model was trained once (for each variant and input combination) using the globally pooled data. The global data pool was formed by combining data collected from different stations into a single dataset. The data were randomly shuffled to remove possible

91

patterns in the data at different stations. The models trained using the global

data pool were tested at different stations to assess their robustness.

3.7   Performance Evaluation

To assess the performance of the machine learning models (base and

hybrid), several performance evaluation metrics were used to grade the models

from different aspects. The details and purpose of the selected metrics are

described in the following subsections.

3.7.1   Mean Absolute Error

The MAE measures the differences between estimated ET_0 and actual

ET_0. MAE which has a smaller magnitude indicates that the estimated ET_0

values are close to the actual values of ET_0, and vice versa. The calculation of

MAE is shown in Equation (3.17).

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_{estimated,i} - y_{actual,i}| \qquad (3.17)$$

where:

$y_{estimated,i}$ = i^{th} estimated value of ET_0

$y_{actual,i}$ = i^{th} actual value of ET_0

N = number of observations

3.7.2 Root Mean Square Error

The RMSE was used to detect large errors. As compared to the MAE, the RMSE assigns higher weightage to errors with larger magnitude by using the sum of square errors. Ideally, the RMSE should be as small as possible and close to the value of MAE. Difference between the RMSE and the MAE implies that there are many large errors within the estimation. The RMSE can be calculated using Equation (3.18).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{estimated,i} - y_{actual,i})^2} \qquad (3.18)$$

3.7.3 Mean Absolute Percentage Error

The magnitude of MAE shows the exact Euclidean distance of the estimated ET_0 from the actual ET_0. The magnitude of MAE is often governed by the scale of the subject of interest (actual data with a high magnitude would scale up the MAE). Hence, the mean absolute percentage error (MAPE) is introduced to normalise the scaling effect. The MAPE converts the MAE into relative error by percentage through the normalisation against the actual value as shown in Equation (3.19).

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_{estimated,i} - y_{actual,i}|}{y_{actual,i}} \times 100\ \% \qquad (3.19)$$

3.7.4 Mean Bias Error

The mean bias error (MBE) has similar attributes as the MAE. However, the MBE takes the direction of error (either underestimate or overestimate) into consideration. In other words, by calculating the MBE metrics, one should have a clearer view of the generalised error of the model, whether the model tends to estimate a higher ET_0 or vice versa. An MBE value of zero does not mean that there is no error, instead, it means that the underestimation is as much as overestimation, which implies that the model is not biased to any of the directions. The MBE can be computed using Equation (3.20)

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(y_{estimated,i} - y_{actual,i}) \tag{3.20}$$

3.7.5 Coefficient of Determination

The R^2 indicates the goodness-of-fit of the model. It ranges from 0 to 1, where the value indicates the proportion of the data explainable by or well fitted to a specific model. The R^2 value closer to 1 is more desirable, and it can be calculated based on Equation (3.21).

$$R^2 = \left(\frac{\sum_{i=1}^{N}(y_{actual,i}-\bar{y}_{actual})(y_{estimated,i}-\bar{y}_{estimated})}{\sqrt{\sum_{i=1}^{N}(y_{actual,i}-\bar{y}_{actual})^2 \sum_{i=1}^{N}(y_{estimated,i}-\bar{y}_{estimated})^2}}\right)^2 \tag{3.21}$$

where:

$\bar{y}_{actual}$ = mean value of actual ET_0

$\bar{y}_{estimated}$ = mean value of estimated ET_0

3.7.6 Global Performance Indicator

It is important to have a performance evaluation metric that can compare different models comprehensively by considering all the aspects. The global performance indicator (GPI) introduced by Despotovic, et al. (2015) allows for such comparison. The advantages of the GPI as compared to other performance evaluation metrics are that it provides a comprehensive comparison that encompasses different aspects of the models and using the GPI eases the comparison with the median model. During the calculation of the GPI, the performance evaluation metrics of the models of interest were normalised from 0 to 1 using an equation like Equation (3.1). Then, the scaled performances of the models were compared with the median performance to calculate the GPI score. The GPI score can be determined using Equation (3.22).

$$GPI = \sum_{i=1}^{N} \alpha_j (\bar{y}_j - y_{ij}) \qquad (3.22)$$

where:

N = number of models

$\bar{y}_j$ = median value of the jth scaled performance metric

y_{ij} = j^{th} scaled performance metric of ith model

α_j = coefficient for j^{th} scaled performance metric (equals to 1 except for R^2)

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Performance of Base Models at Different Stations

The base MLP, SVM and ANFIS were trained and tested using the various input combinations at different stations under Scenario 1. In the following subsections, the effects of input meteorological variables, comparison of the base models as well as the selection of optimum input combinations are discussed.

4.1.1 Effect of Input Meteorological Variables

Six meteorological variables collected from the MMD were used to create a total of 63 input combinations for model training. The MLP, SVM and ANFIS were trained according to the k-fold cross-validation algorithm. The performance of the MLP is summarised in Figure 4.1. The actual values are available in Appendix A (Table A1 – A12).

Figure 4.1(a) to Figure 4.1 (c) show the MAE, RMSE and MAPE of the estimations of the MLP at different stations, respectively. A darker tone in the heat maps indicates a higher error and vice versa. The performance of the MLP

using C1 to C63 can be examined by analysing Figure 4.1(a) to Figure 4.1(c) column-by-column. The MAE of the MLP's estimations ranged from 0.0236 mm/day to 0.8457 mm/day whereas the RMSE ranged between 0.0289 mm/day and 1.0645 mm/day. Low MAE and RMSE values are found to be concentrated at the left-hand side of the heat maps, which correspond to the input combinations that consist of higher number of meteorological variables. The results are reasonable such that by providing more meteorological variables to the MLP, the model will be able to fetch more information from the inputs given. That is to say, to develop a good machine learning model for ET_0 estimation at different stations in Peninsular Malaysia, more meteorological variables have to be collected.

(a)

MAE (mm/day)

(b)

RMSE (mm/day)

(c)

MAPE (%)

(d)

R²
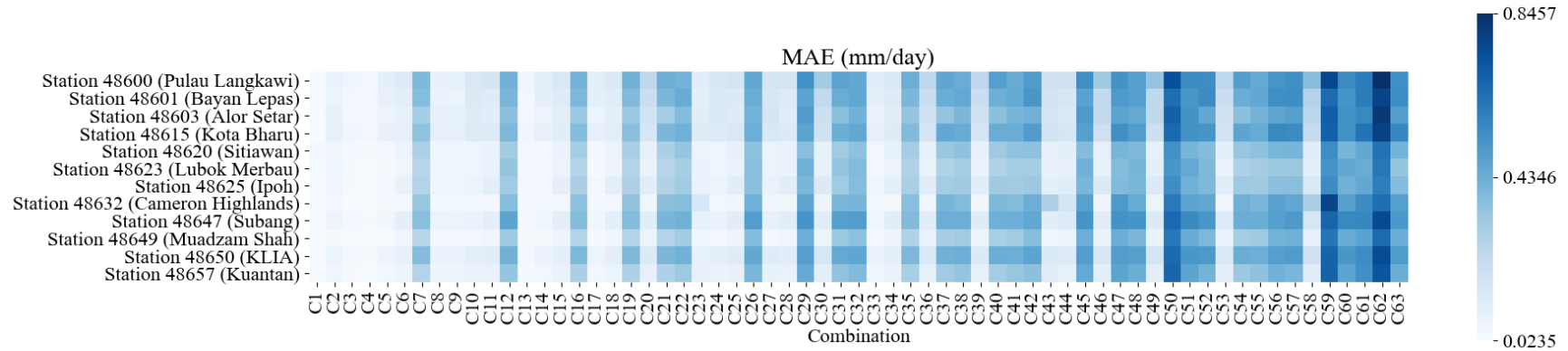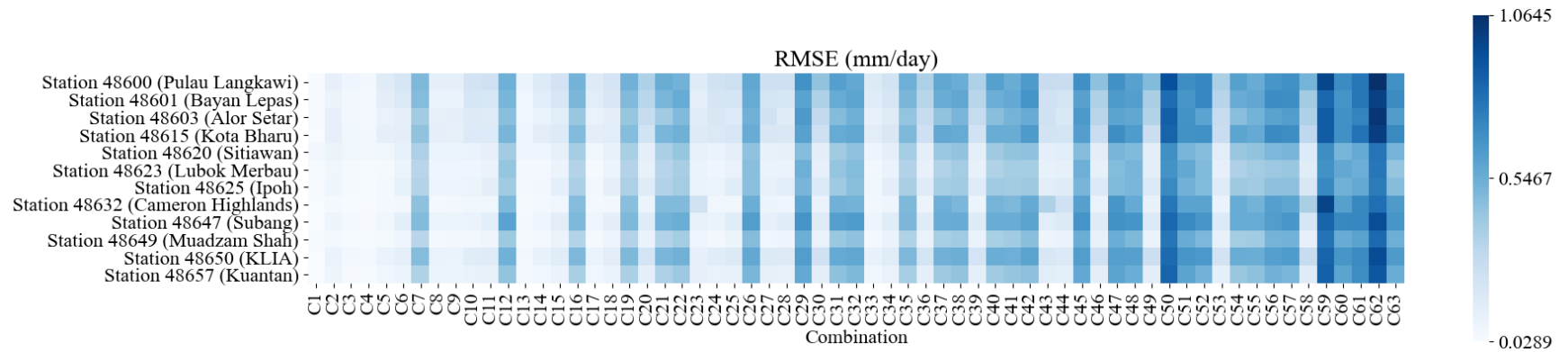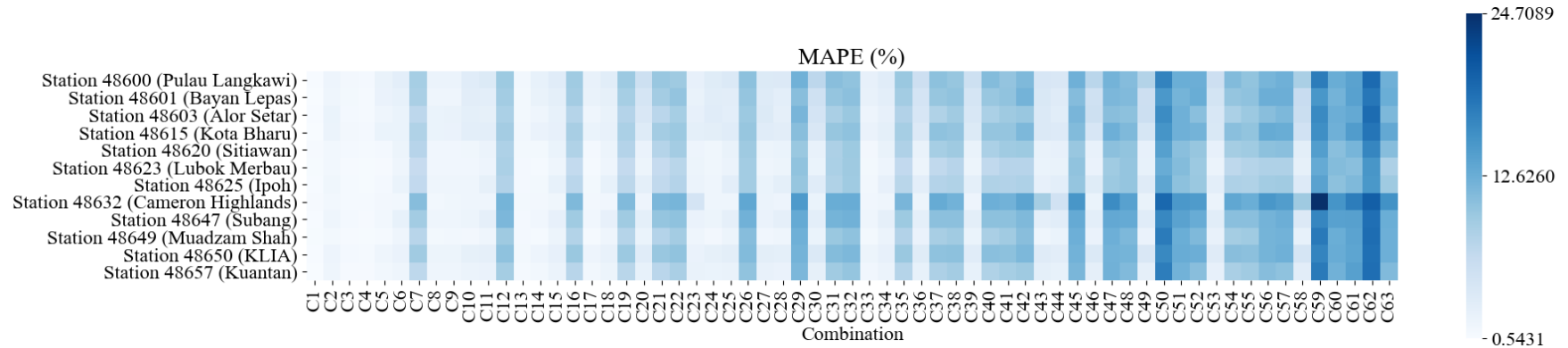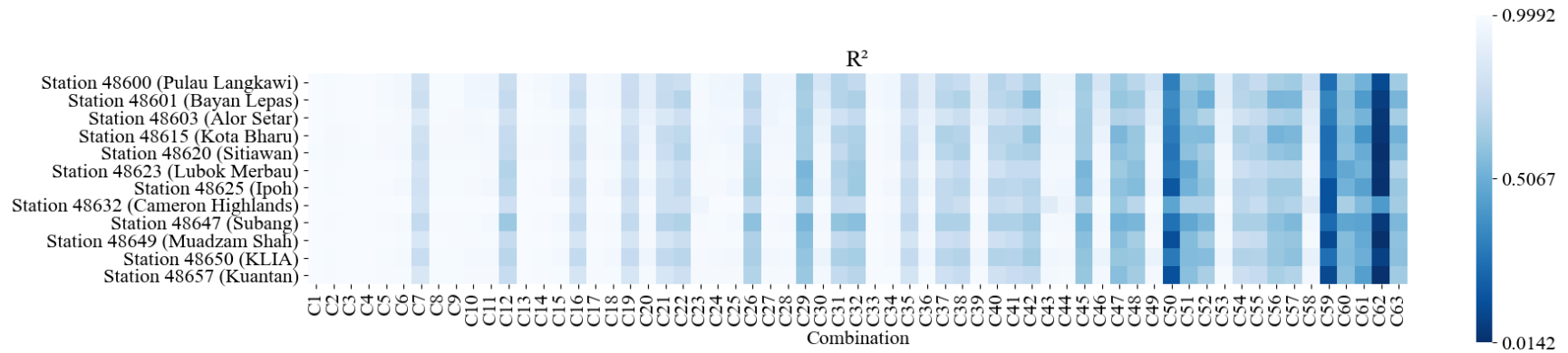
(e)
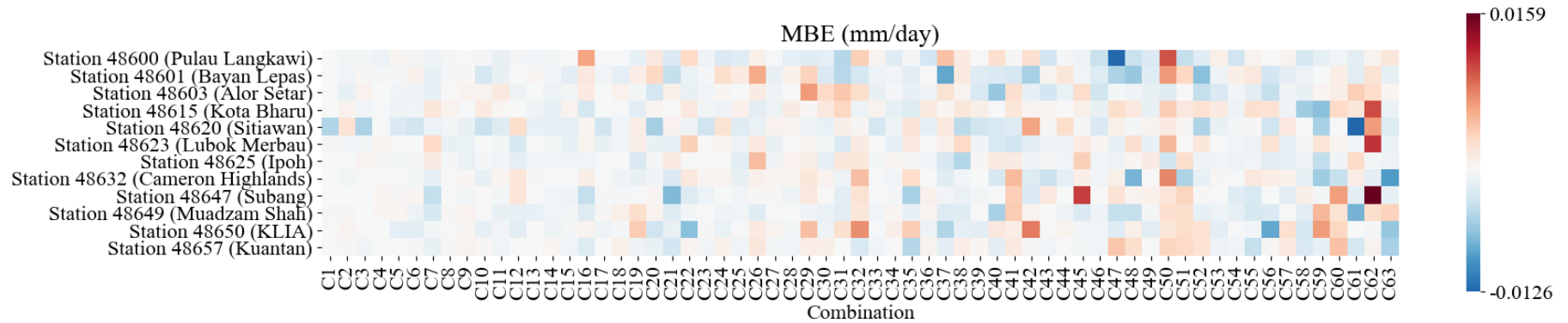
Figure 4.1: (a) MAE, (b) RMSE, (c) MAPE, (d) R^2 and (e) MBE of MLP Estimation at Different Stations with Different Input Combinations

In general, the error distribution in Figure 4.1(a) is similar to Figure 4.1(b) whereby Station 48600 (Pulau Langkawi), Station 48601 (Bayan Lepas), Station 48603 (Alor Setar) and Station 48615 (Kota Bharu) registered the highest MAE and RMSE when the number of input meteorological variables were lesser (C50 onwards). However, when the errors were rectified by calculating the MAPE, Station 48647 (Subang), Station 48649 (Muadzam Shah), Station 48650 (KLIA) and Station 48657 (Kuantan) had the highest MAPE among the 12 stations as shown in Figure 4.1(c) (C50 onwards as well). In other words, the MLP could estimate better in the northern regions than the stations located in the central Peninsular Malaysia.

In terms of generalisability, the MLP performed well in which the model could achieve R^2 values of at least 0.60. Nevertheless, there are some exceptions observed. When trained using the input combinations C50, C59 and C62, the MLP had rather poor performance. This could be explained by the absence of key meteorological variables, which will be discussed later. From the aspect of the MBE, the MLP registered a maximum underestimation of -0.0126 mm/day and could overestimate up to 0.0159 mm/day. These values correspond to a bias of -0.33 % and 0.39 % at their respective stations, which can be considered insignificant. It is interesting to note that the occurrence of underestimation (blue) and overestimation (red) increased as the number of meteorological variables fed as inputs was reduced. On top of that, estimation bias consistently appeared in the MLP estimation of ET_0 at Station 48620 (Sitiawan).

101

By observing Figure 4.1(a) to Figure 4.1(d), it can be seen that dark columns periodically appear irrespective of the stations tested. For instance, dark bands (poor performance) appeared at C7, followed by C12, C16, C19, C21, C22 and so on. When referring Figure 3.3, the mutual characteristic of these input combinations is the absence of R_s as one of the input meteorological variables. In other words, R_s can be claimed as the key meteorological variable in the estimation of ET_0 in Peninsular Malaysia by using the MLP. This also explains the low R^2 value attained for C50, C59 and C62. In fact, C50 is the union set of C59 and C62, which consist of only u and T_{min}, respectively. Further deduction can be made to imply that the u and T_{min} are the two least important features for ET_0 estimation in Peninsular Malaysia.

The finding of the key and least essential meteorological variables for ET_0 estimation shall be supported by scientific theory as a step forward to reduce the opacity of the black-box operation of machine learning based estimation. Ndiaye, et al. (2017) analysed the sensitivity of the ET_0 towards the change in meteorological variables in the region of Burkina Faso, which had a similar climate pattern to that of Peninsular Malaysia. The authors reported that the R_s was the most influential meteorological variable on the ET_0. On top of that, RH would become important during dry seasons which corresponded to low vapour pressure. In Kenya, it was argued that the R_s alone represented multiple scenarios that could probably affect the ET_0 (Odongo, et al., 2019). The authors stated that low R_s could be due to the increase in the cloud coverage as well as the aerosols. In other words, that would be well associated with low

surface temperature and higher RH. This situation had made R_s the key meteorological variables that dictate the regime of the ET_0.

Besides looking at areas with similar climates to Peninsular Malaysia, Pour, et al. (2020) performed a thorough analysis to study the relationship of different meteorological variables in Peninsular Malaysia itself. The trend analysis showed that the R_s correlated well with the ET_0 where both exhibited close to identical trends throughout the study period, whereas the RH had a reversed trend as compared to the ET_0. Moreover, the results for sensitivity of ET_0 towards the other meteorological variables were actually the opposite of that reported in Burkina Faso (Ndiaye, et al., 2017). Pour, et al. (2020) discovered that in Peninsular Malaysia, the R_s and the RH were least influential towards the ET_0, which contradicted their findings in the trend analysis. However, the authors did not provide any further explanation on this matter.

The results in this study point to the fact that R_s is the key or essential meteorological variable for estimating ET_0 in Peninsular Malaysia using the MLP. This finding is in agreement with all the cited research works, except for the sensitivity analysis of Pour, et al. (2020). This discrepancy could be due to the high associative relationship between the time series data of ET_0 and R_s in the study by Pour, et al. (2020). Furthermore, despite having similar seasonal trends, the fluctuation of R_s in the study is very low, as compared to the ET_0. Therefore, the change in the trend of the ET_0 would be mainly driven by the anomaly that occurred in other meteorological variables. The significant

deviations in other meteorological variables would result in a coercive change in the values of ET_0 that could not be accurately estimated by the MLP fed with only R_s.

Besides matching many empirical sensitivity studies, the discovery of the R_s to be the key features for ET_0 estimation also aligns well with the nature of the ET process. Essentially, incoming radiation from the Sun is the sole energy input to drive water depletion from the Earth's surface (Cascone, et al., 2019). Moreover, the R_s in Peninsular Malaysia is less prone to seasonal variation due to the geographical characteristic close to the Equator (Pour, et al., 2020). In fact, the ET is also strongly affected by other environmental conditions such as temperature and humidity. Nevertheless, this study revealed that the MLP was able to perform good estimations of ET_0 at various locations with only the R_s as input. This is because the other environmental conditions (temperature and humidity) depend on the R_s. In other words, the change R_s value alone can actually be translated to the change in temperature or humidity. High R_s would correspond to high temperature, which would in turn decrease the value of RH. The MLP can be improved by increasing the number of "complementary" meteorological variables in the training data as an effort to provide more explanatory features on the environmental conditions.

Of all the six meteorological variables used in this research work, only the u is independent of the R_s. However, in Peninsular Malaysia, the average

value of u is relatively low, and that in turn reduces its contribution towards the ET.

The study on the effect of input meteorological variables provides a clear picture that can enhance prioritisation during the data collection process. The R_s should be collected as it is the key meteorological variable for ET_0 estimation, followed by other complementary meteorological variables in Peninsular Malaysia. It is noteworthy that the complementary meteorological variables needed to further enhance the ET_0 estimations were different across the whole study area. In other words, for different stations, the optimum input combinations could be different. Discussion on this finding will be presented in this thesis in Section 4.1.3.

4.1.2 Comparison of Base Models

Besides training the MLP, the scope of this study also encompasses other base models, namely the SVM and the ANFIS. The training methodology was identical for all the models investigated. The performances of the SVM and the ANFIS are provided in Figure 4.2 and Figure 4.3, respectively. The actual value of MAE, RMSE, MAPE, R^2 and MBE are available in Appendix A (Table A13 – A24 for SVM, Table A25 – A36 for ANFIS).

(a)

MAE (mm/day)

(b)

RMSE (mm/day)

106
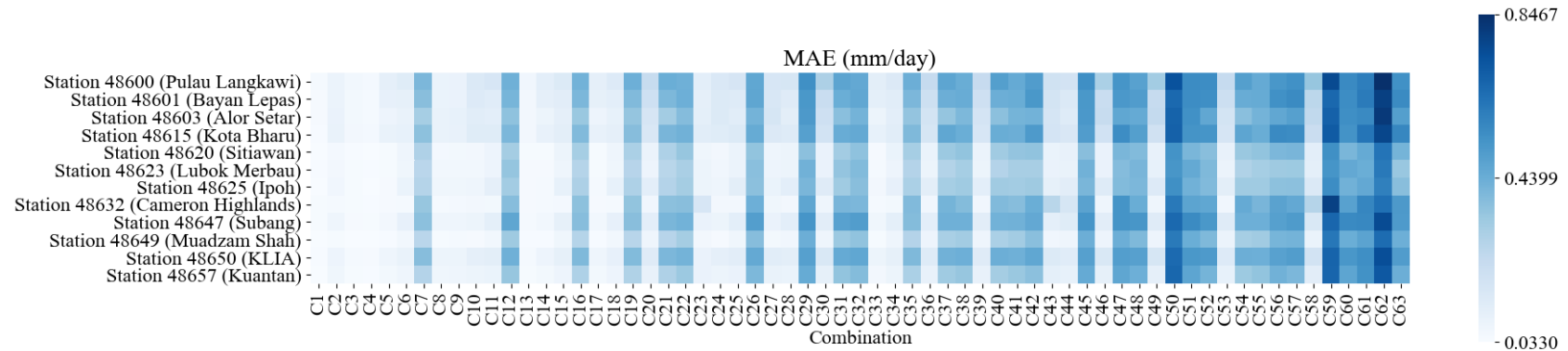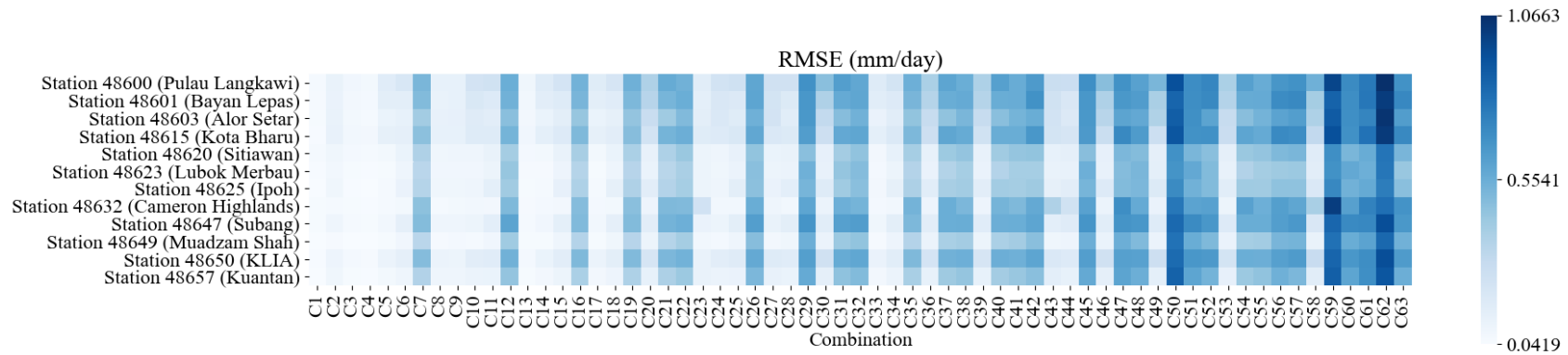
(c)
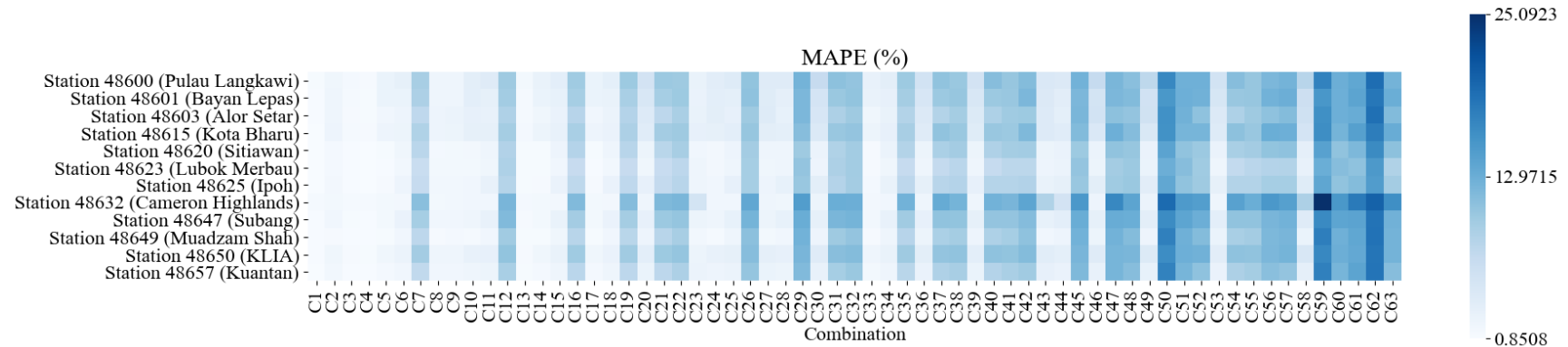
(d)
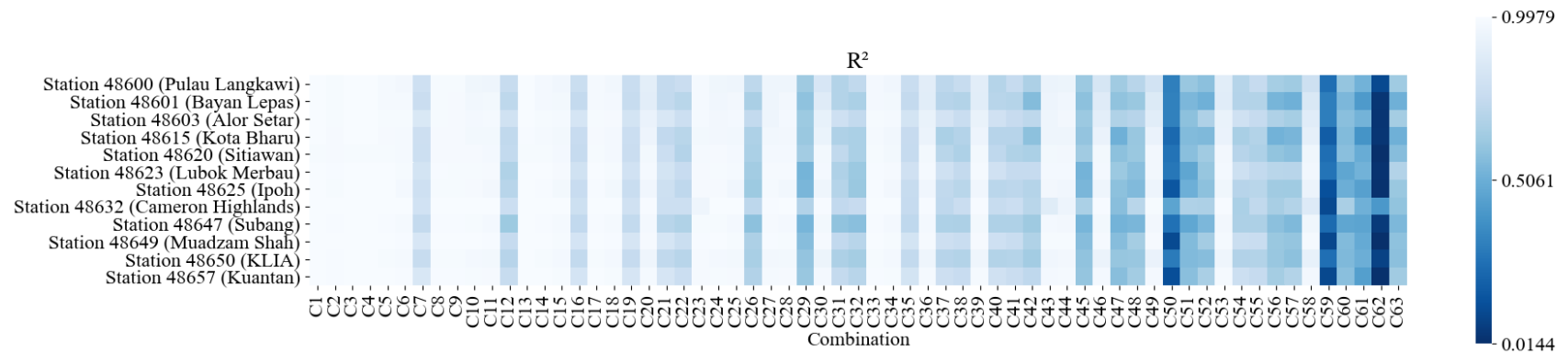
(e)
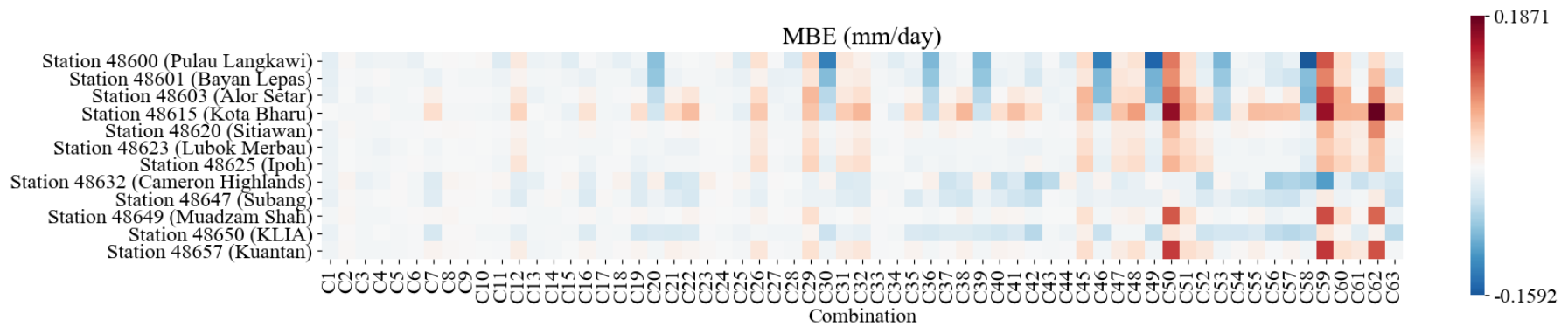
Figure 4.2: (a) MAE, (b) RMSE, (c) MAPE, (d) R^2 and (e) MBE of SVM Estimation at Different Stations with Different Input Combinations

(a)

MAE (mm/day)

(b)

RMSE (mm/day)

(c)

(d)

(e)

MBE (mm/day)
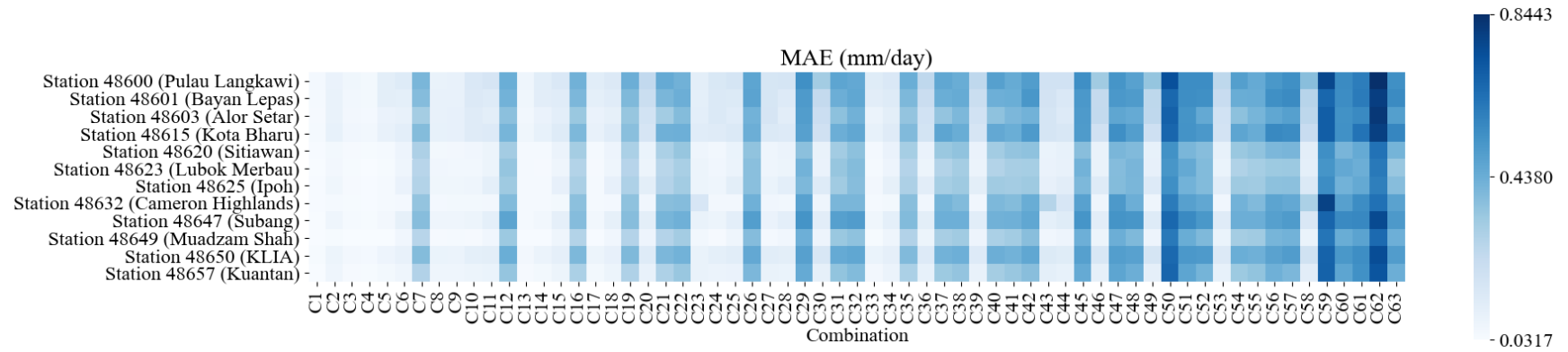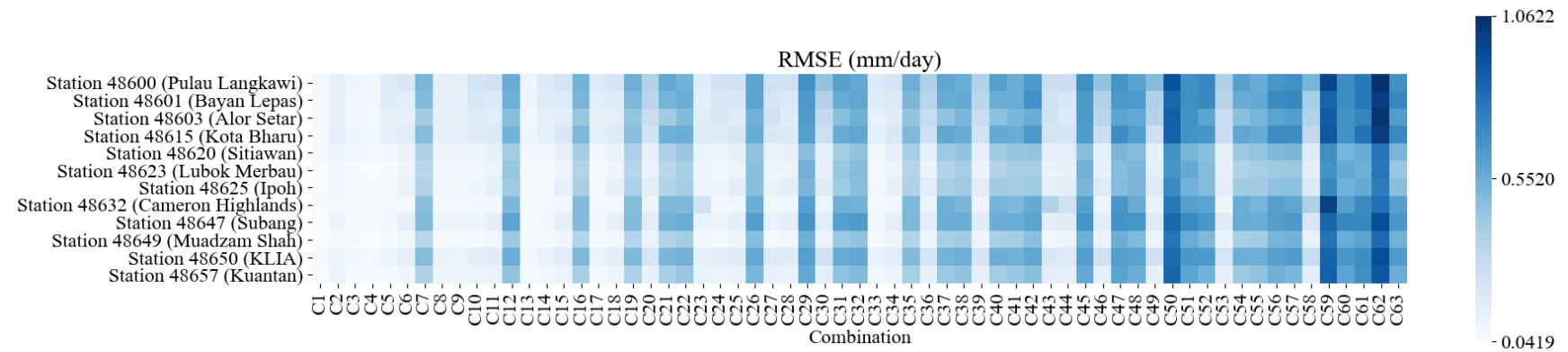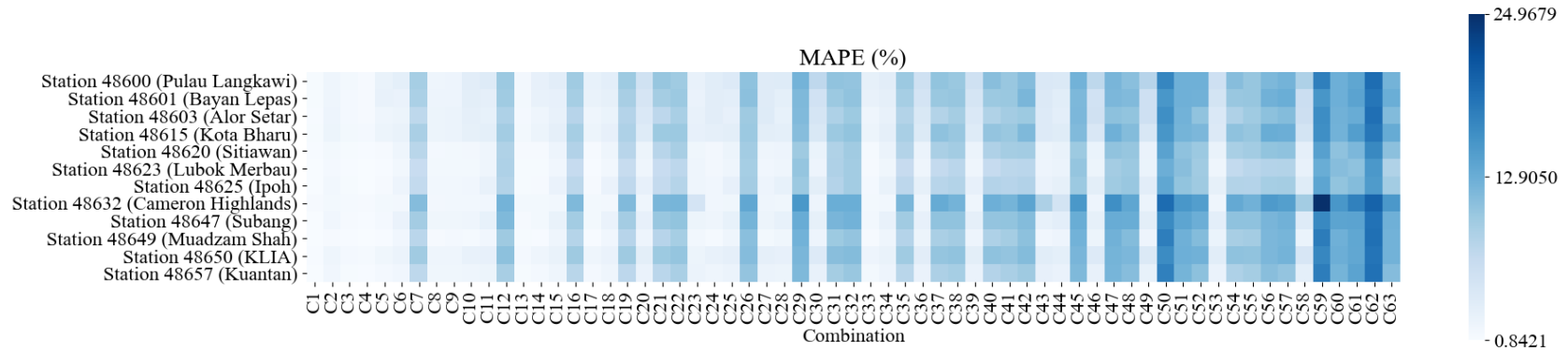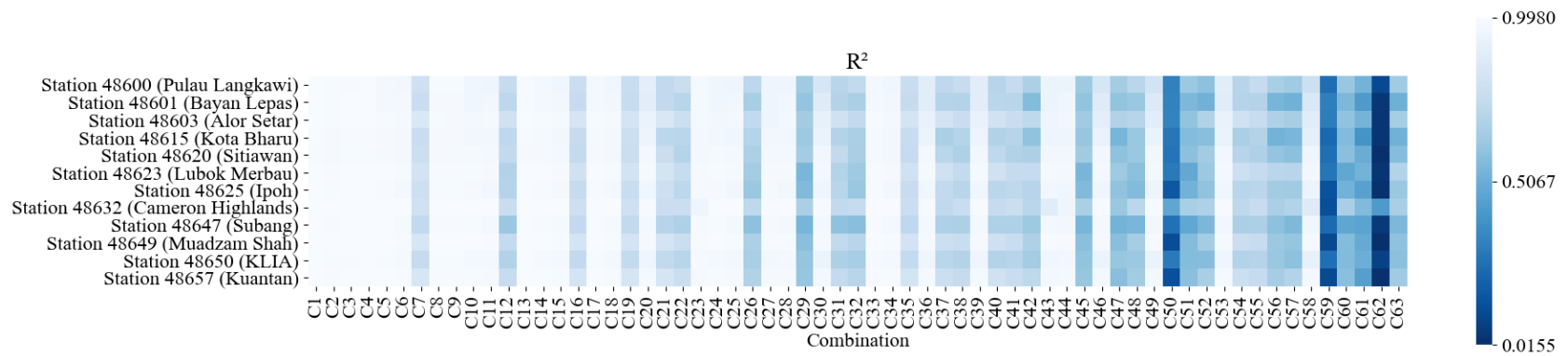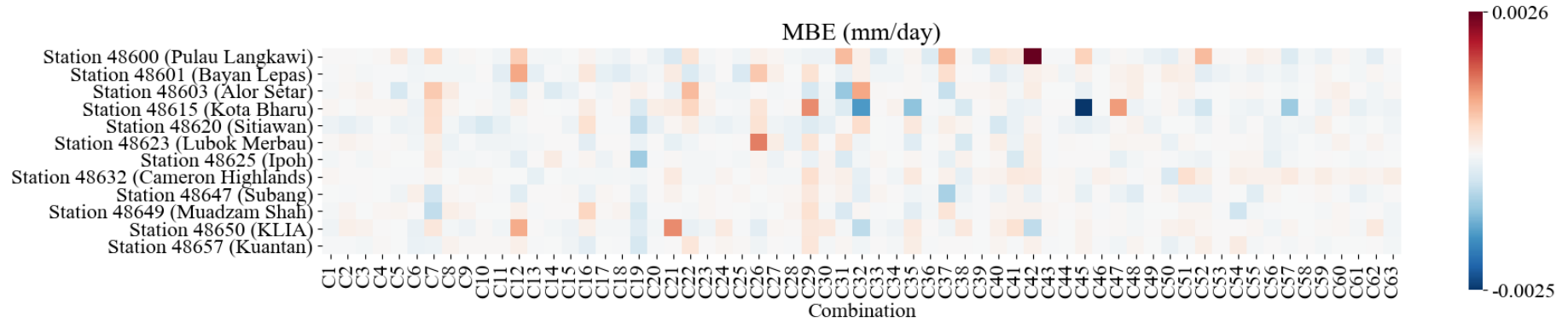
Figure 4.3: (a) MAE, (b) RMSE, (c) MAPE, (d) R^2 and (e) MBE of ANFIS Estimation at Different Stations with Different Input Combinations

The SVM and ANFIS exhibited similar traits to the MLP discussed previously. Darker bands appeared in the heat map of MAE, RMSE, MAPE and R^2 periodically when the essential R_s was missing from the input combinations. However, by comparing the performances of the models numerically, significant differences are discovered.

In comparison with the MLP, the SVM, generally estimated less accurately particularly for the cases with higher number of input meteorological variables. This phenomenon can be observed from Figure 4.2(c), where the MAPE of SVM is as high as 25 %. This was contributed by the high MAE as well as RMSE of the SVM estimated ET_0. For instance, the MAE and RMSE of the ET_0 estimated at Station 48603 (Alor Setar) by the MLP using C1 as input combination were 0.0277 mm/day and 0.0357 mm/day, respectively, whereas the corresponding values of the SVM were 0.0437 mm/day and 0.0538 mm/day. Despite having higher estimation errors with C1, the performance of the SVM did not deteriorate much as the number of input meteorological variables decreased. In fact, when using C44 as the input combination, the MLP registered MAE and RMSE of 0.1383 mm/day and 0.1835 mm/day, in which the performance declined significantly as compared to the C1 with all meteorological variables as input (MAPE increased by 2.5835 %). On the contrary, the MAE and RMSE of the SVM with the same condition are 0.1376 mm/day and 0.1835 mm/day. The resultant decrement in accuracy (MAPE increased by 2.1954 %) was not as bad as the MLP. In fact, the accuracy of the SVM was higher than the MLP in this case (i.e., C44). This suggests that the

SVM has better generalisability and resilience towards the reduction of input meteorological variables. A similar phenomenon also can be observed from the lower variance of the R^2 values at different stations when ET_0 was estimated using the SVM with different input combinations.

The performance of the SVM can be explained by examining the working mechanism of the model itself. The SVM transforms data into a feature space by using the kernel function (the RBF function for this study). Estimation of the SVM is accomplished by selecting a support vector structure that can minimise the risk or loss function. In other words, the SVM performs regression analysis by viewing the feature space as a big picture in contrast to the point-by-point adjustment of the MLP. The high generalisability and robustness of SVM allow it to be used for many cases, but at the same time sacrifices the possibility of achieving very high accuracy (which could also be overfitting).

Nevertheless, a major pitfall appears to be existing in the SVM. Figure 4.2(e) shows that the SVM had the tendency to overestimate ET_0 when the input meteorological variables were lesser. Besides, when comparing Figure 4.2(e) with Figure 4.1(e), it can be observed that the MBEs of the SVM's estimations (ranged between -0.15 mm/day to 0.2 mm/day) were higher than those of the MLP (ranged between -0.015 mm/day to 0.015 mm/day). This could be attributed to the relatively lower accuracy of the SVM, which consequentially affects the MBE of the SVM.

The performance of the ANFIS is illustrated by the heat maps shown in Figure 4.3. The ANFIS shares similar structure with the MLP, whereby they are composed of neural networks. The distinct feature of the ANFIS from the MLP is that the neural network of the ANFIS is used to tune the fuzzy rules and the membership functions instead of the weights and biases. The utilisation of the fuzzy rules in the ANFIS allows the description of the data in terms of likelihood.

Like the SVM, the MAE and RMSE of the ET_0 estimation by the ANFIS at Station 48603 (Alor Setar) using C1 as the input combination were 0.0438 mm/day and 0.0586 mm/day, respectively, which were relatively higher than those of the MLP. This corresponds to the MAPE of 1.0241 %. On the contrary, when C44 was used as the input combination, the MAPE only reduced by 2.1995 % (2.5835 % for MLP) with the MAE and RMSE, standing at 0.1379 mm/day and 0.1834 mm/day, respectively. Yet again, the MLP was outperformed by another model when input meteorological variables were reduced.

It is interesting to note that the ANFIS can successfully overcome the problem of bias error. The MBE of the ANFIS estimated ET_0 at all stations mostly lied between -0.002 mm/day and 0.002 mm/day with very few exceptions. Furthermore, the MBE of the ANFIS's estimations appeared to be unaffected by the input combinations. This observation could justify the versatility of the ANFIS towards various input combinations that renders its robustness and generalisability.

4.1.3 Selection of Input Combinations

As illustrated in Figure 4.1, Figure 4.2 and Figure 4.3, there were 63 input combinations tested at each station for the estimation of ET_0. The use of different input combinations resulted in the variance in the output accuracy, even though the same machine learning model was employed. For screening purposes, this study utilised the BMS algorithm to select the best input combinations in different scenarios.

Specifically, C1 will be selected for the scenario when all the six meteorological variables are available for making ET_0 estimation. When only five meteorological variables are available, the best input combination will be selected from C2 to C7 and so on. The selection was made based on the posterior model probability depicted in Equation (3.8). The best input combinations for each scenario at different stations are summarised in Table 4.1.

According to Table 4.1, at all of the studied stations, it was agreed that C1, C4 and C58 were the best input combinations when six, five and one meteorological variables were fed into the models, respectively. The C4 combination removed the T_{mean} from the input dataset, whereas C58 only used the key R_s as the sole meteorological variable. When only four meteorological variables were fed into the MLP, SVM and ANFIS, C13 was selected as the best input combination except for Station 48623 (Lubok Merbau) and Station

48649 (Muadzam Shah) that opted for C17. The combination C13 removed T_{min} from C4, whereas the C17 combination excluded the RH from C4.

On the other hand, for the MLP, SVM and ANFIS trained with only three meteorological variables, 83.33 % of the studied stations favoured C33, which was a result of the exclusion of RH from C13. This echoed the phenomenon observed when four meteorological variables were used to estimate ET_0, where the T_{min} and RH were selectively removed from the input combination. However, Station 48600 (Pulau Langkawi) and Station 48601 (Bayan Lepas) preferred to use C23 as the input combination, whereby all the temperature variables were omitted.

As for the input combinations that consist of only two meteorological variables, the discrepancy among the stations widens. There were three different input combinations selected at this stage, namely C43 (41.67 %), C44 (50 %) and C53 (8.33 %). Station 48632 (Cameron Highlands), located at higher altitudes, stood out amongst the other stations and used C53 for a two-variable estimation of ET_0. The C53 combination consists of T_{max} and R_s, which could be essential for the ET_0 process at highland due to the lower average temperature throughout the year. The stations located in the northern region and east coast of Peninsular Malaysia, namely Station 48600 (Pulau Langkawi), Station 48601 (Bayan Lepas), Station 48603 (Alor Setar), Station 48615 (Kota Bharu), Station 48649 (Muadzam Shah) and Station 48657 (Kuantan) preferred to use C44. The C44 includes RH and R_s whereby the RH is another important driving force of

the ET process, and it also appeared in C1, C4 and C13. The other stations which are not located at the northern region and east coast of Peninsular Malaysia, including Station 48620 (Sitiawan), Station 48623 (Lubok Merbau), Station 48625 (Ipoh), Station 48647 (Subang) and Station 48650 (KLIA) used C43 that was composed of R_s and u. It was explained in the previous sub-section that the R_s, being the key meteorological variable could encompass the effect of many other meteorological variables except for the u variable. Hence for the two-variable estimation, it is reasonable to complement the R_s using u.

Table 4.1: Best Input Combination using Different Number of Input Meteorological Variables at Different Stations

Stations	Number of Meteorological Variables					
	6	5	4	3	2	1
Station 48600 (Pulau Langkawi)	C1	C4	C13	C23	C44	C58
Station 48601 (Bayan Lepas)	C1	C4	C13	C23	C44	C58
Station 48603 (Alor Setar)	C1	C4	C13	C33	C44	C58
Station 48615 (Kota Bharu)	C1	C4	C13	C33	C44	C58
Station 48620 (Sitiawan)	C1	C4	C13	C33	C43	C58
Station 48623 (Lubok Merbau)	C1	C4	C17	C33	C43	C58
Station 48625 (Ipoh)	C1	C4	C13	C33	C43	C58
Station 48632 (Cameron Highlands)	C1	C4	C13	C33	C53	C58
Station 48647 (Subang)	C1	C4	C13	C33	C43	C58
Station 48649 (Muadzam Shah)	C1	C4	C17	C33	C43	C58
Station 48650 (KLIA)	C1	C4	C13	C33	C44	C58
Station 48657 (Kuantan)	C1	C4	C13	C33	C44	C58

By observing the sets of best input combinations, the 12 studied stations can be grouped accordingly into five different clusters. Stations that preferred the same set of input combinations were grouped into the same cluster. For example, both Station 48600 (Pulau Langkawi) and Station 48601 (Bayan Lepas) had the best input combinations of C1, C4, C13, C23, C44 and C58. Therefore, the two stations were being classified into the same cluster. The

clustering done in this context did not involve soft computing algorithms, but merely based on the observation of the preferred sets of input combinations of the 12 stations. This allows a closer look at the common characteristics of the stations and deduces the relationship among the stations. Table 4.2 shows the studied stations arranged in their respective clusters.

Table 4.2: Clustering of Stations Based on Best Input Combinations

Cluster	Station	Remark
Cluster 1	Station 48603 (Alor Setar) Station 48615 (Kota Bharu) Station 48650 (KLIA) Station 48657 (Kuantan)	Meteorological Stations in Airport
Cluster 2	Station 48620 (Sitiawan) Station 48625 (Ipoh) Station 48647 (Subang)	Urban Area
Cluster 3	Station 48600 (Pulau Langkawi) Station 48601 (Bayan Lepas)	Island
Cluster 4	Station 48623 (Lubok Merbau) Station 48649 (Muadzam Shah)	Thick Vegetation
Cluster 5	Station 48632 (Cameron Highlands)	Highland

Although the stations can be grouped into clusters, it is important to note that there are large intersections between the clusters. For example, Cluster 1 and Cluster 2 only differed when ET_0 was estimated using five meteorological variables whereas Cluster 2 and Cluster 3 have only two disagreements in selecting the best input combinations. Clustering provides a better understanding on the relationship between the stations but not the ultimate guideline.

Undeniably, applying the BMS for input combinations screening has successfully eliminated 57 less promising input combinations at each station.

This effort would improve the computational efficiency where the researchers and the authorities need not look at all the meteorological variables when estimating ET_0 in the future. Although there was partial disagreement among the studied stations when two, three, and four meteorological variables were used for training, one can still generalise the preferred input combination based on the number of votes each combination gets.

At this point, modelling works (training and testing) that involved the base MLP, SVM, and ANFIS have reached an end. The following sections will discuss the hybridisation of the base models to improve the model performance from different aspects. Also, the modelling work would not cover all the possible 63 input combinations, but only stress on the best input combinations obtained from the BMS screening (as shown in Table 4.1) as the unselected input combinations would produce less accurate estimations.

4.2   Data Fusion I: Bootstrap Aggregating

The data centric bootstrap aggregating was applied to the base MLP, SVM and ANFIS to observe its effects on the models' performance by altering the structure of the training data. The findings of this part of the study are described in the following subsections.

4.2.1 Bootstrap Aggregating with MLP

The MLP with a similar structure was trained using bootstrapped dataset to investigate the effects of the data fusion technique. However, instead of performing model training using all the input combinations, only the best input combinations were selected based on the clustering analysis that was done previously. The improvement of the bagged MLP (BMLP) model is summarised as in Figure 4.4 to Figure 4.8 (Cluster 1 to Cluster 5) whereas the actual values of the performance metrics are compiled in Appendix B (Table B1 – B5). The stations were grouped into clusters to investigate the effect of geographical characteristics on the response towards the integration of various data fusion techniques. The change in MBE was not included as the MBE only reflects the direction of bias estimation and would be less meaningful for measuring the improvement of the BMLP.

Unfortunately, the bootstrap aggregating did not consistently improve the performance of the MLP. As shown in Figure 4.4 to Figure 4.8, it can be observed that irrespective of the clusters (except for Cluster 2), most of the stations experienced an increment in the MAE, RMSE and MAPE when the trained BMLP was used to estimate ET_0. The error metrics increased by at most 2 % for all stations, except for Station 48601 (Bayan Lepas). The maximum increase in error for Station 48601 (Bayan Lepas) could reach as high as approximately 9 %.

On the other hand, the values of the R^2 were less affected by the use of the BMLP. In most of the cases, the R^2 values were close to those of the MLP. This could be due to the ability of the BMLP to maintain its generalisability as a high number of examples was shown to the model during the training process. The large number of training data fed into the BMLP helped to prevent the model from being "used to" and becoming biased to a small portion of the dataset. However, higher deviations of the R^2 values were witnessed when the number of meteorological variables was reduced. A significant drop in the R^2 value occurred when the BMLP was fed with only the R_s variable. Although the drop in the R^2 was within an acceptable range in most of the stations, the BMLP trained in Station 48601 (Bayan Lepas) once again stood out where the reduction of R^2 was as much as 7.32 %. Furthermore, the R^2 of the BMLP in Station 48601 (Bayan Lepas) was also reduced by 8.26 % when fed with two meteorological variables, namely R_s and RH.

As for Cluster 2, the BMLP slightly improved the base MLP and this effect was only obvious in Station 48620 (Sitiawan) whereby a maximum reduction of MAE, RMSE and MAPE of about 3 % was achieved. However, the positive effect gradually faded away as the numbers of input meteorological variables were reduced.
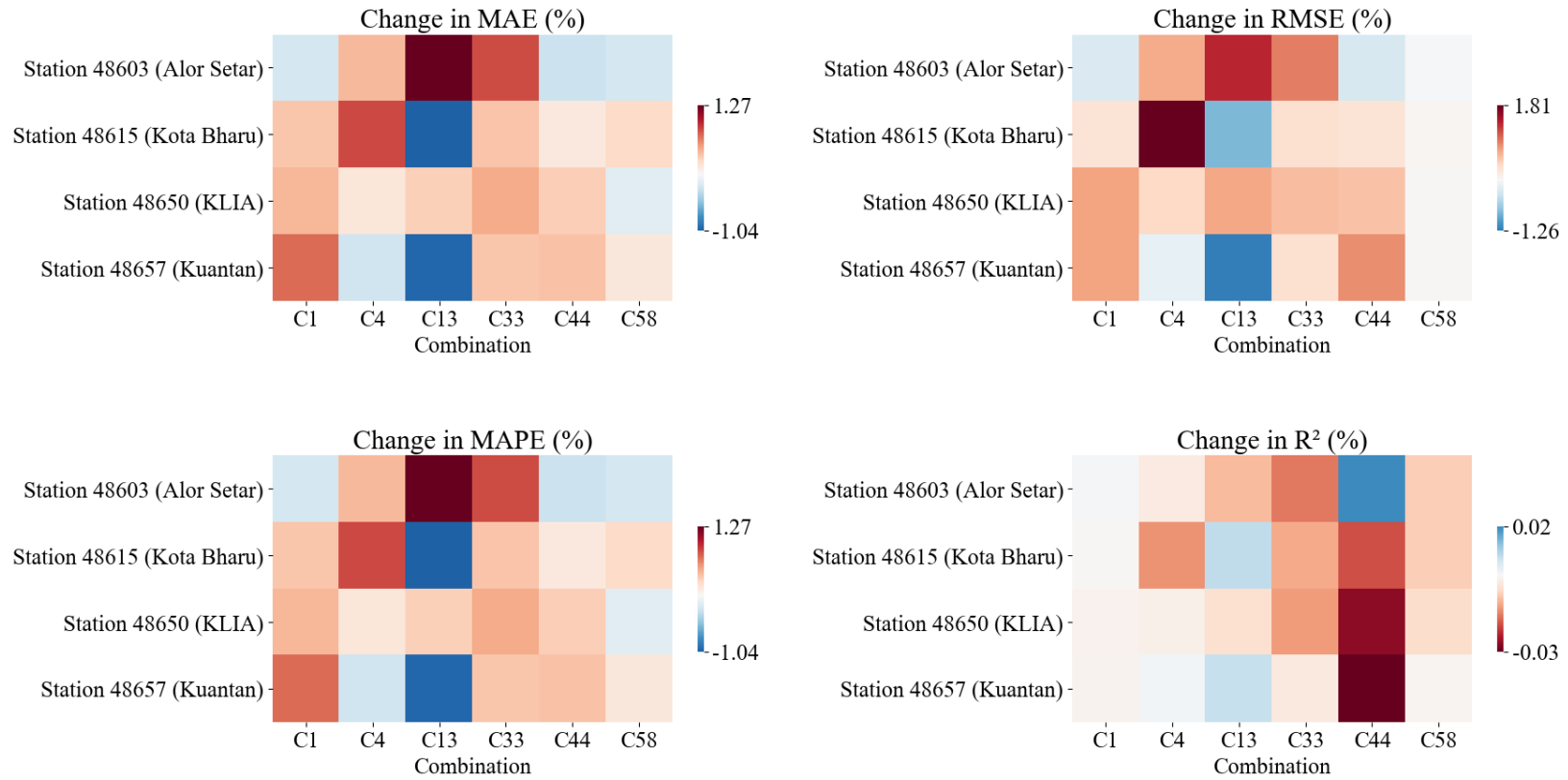
Figure 4.4: Changes in MAE, RMSE, MAPE and R^2 of BMLP (in %) based on MLP for Stations in Cluster 1
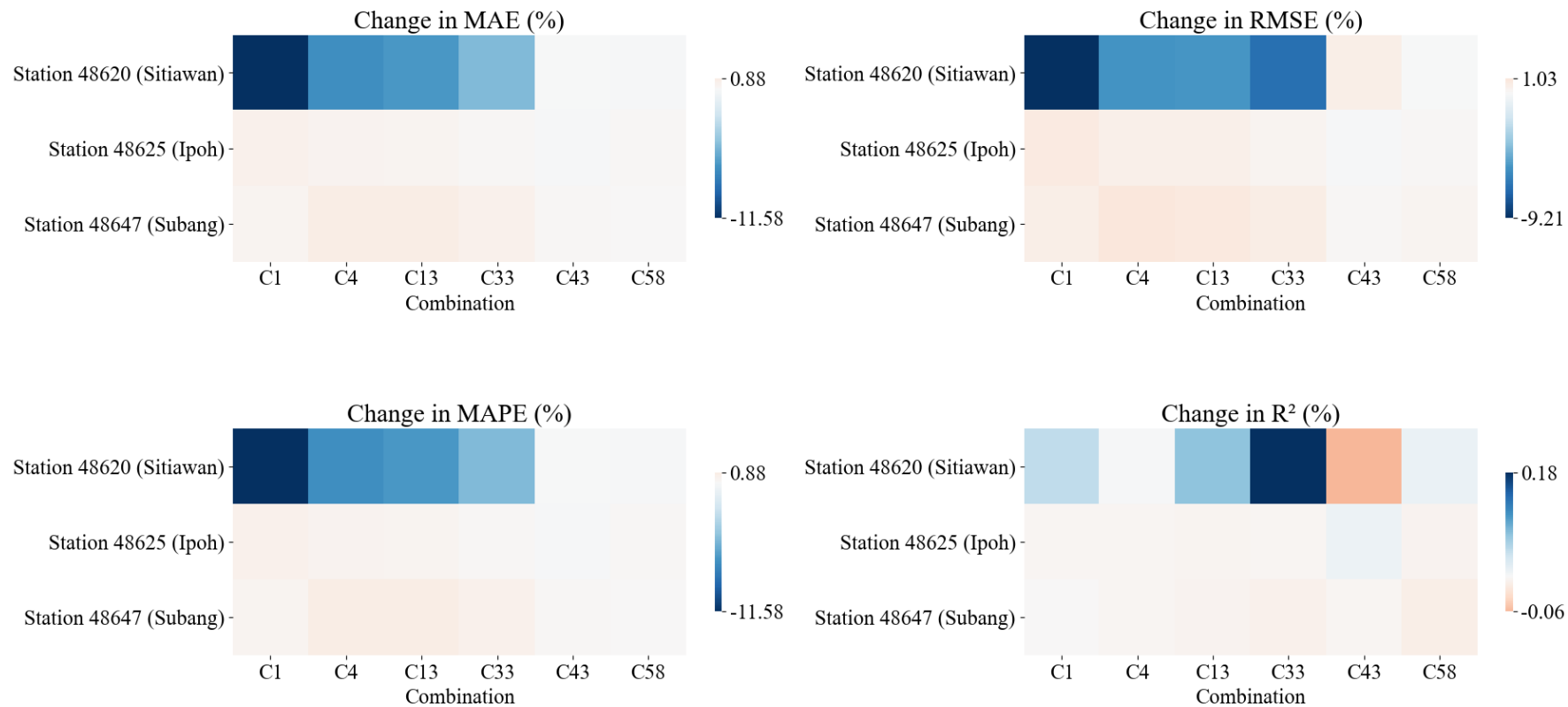
Figure 4.5: Changes in MAE, RMSE, MAPE and R^2 of BMLP (in %) based on MLP for Stations in Cluster 2
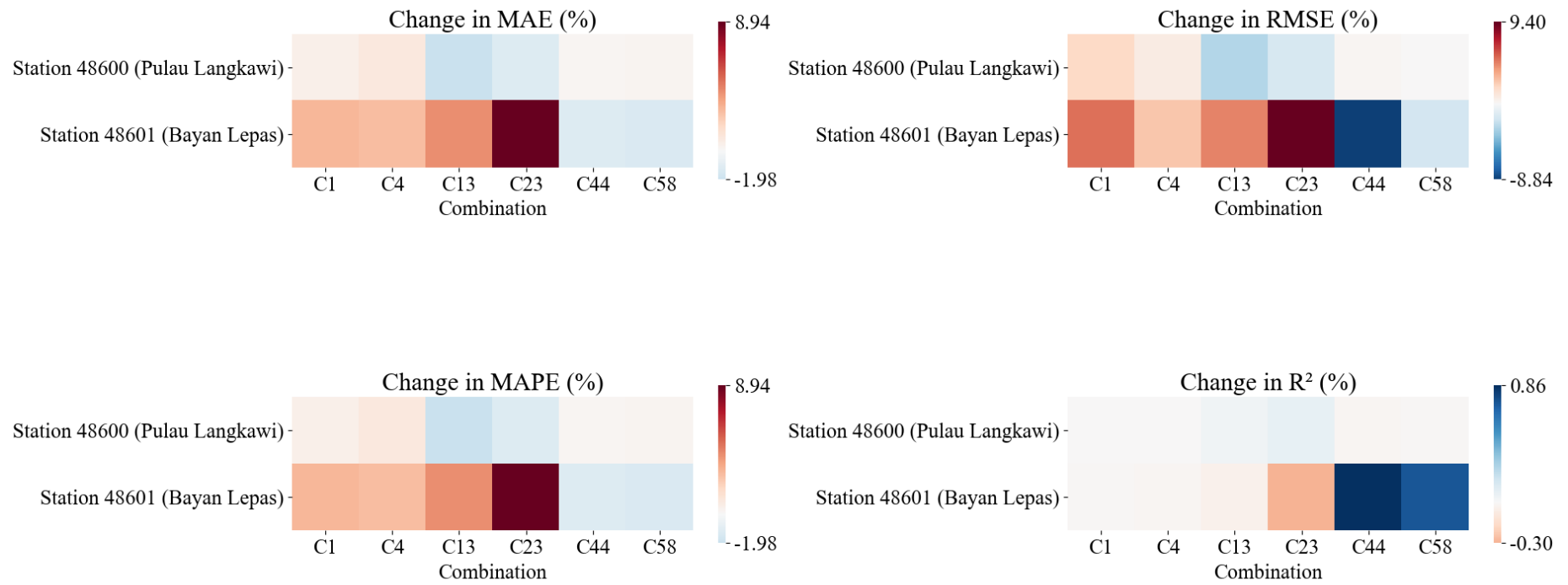
Figure 4.6: Changes in MAE, RMSE, MAPE and R^2 of BMLP (in %) based on MLP for Stations in Cluster 3
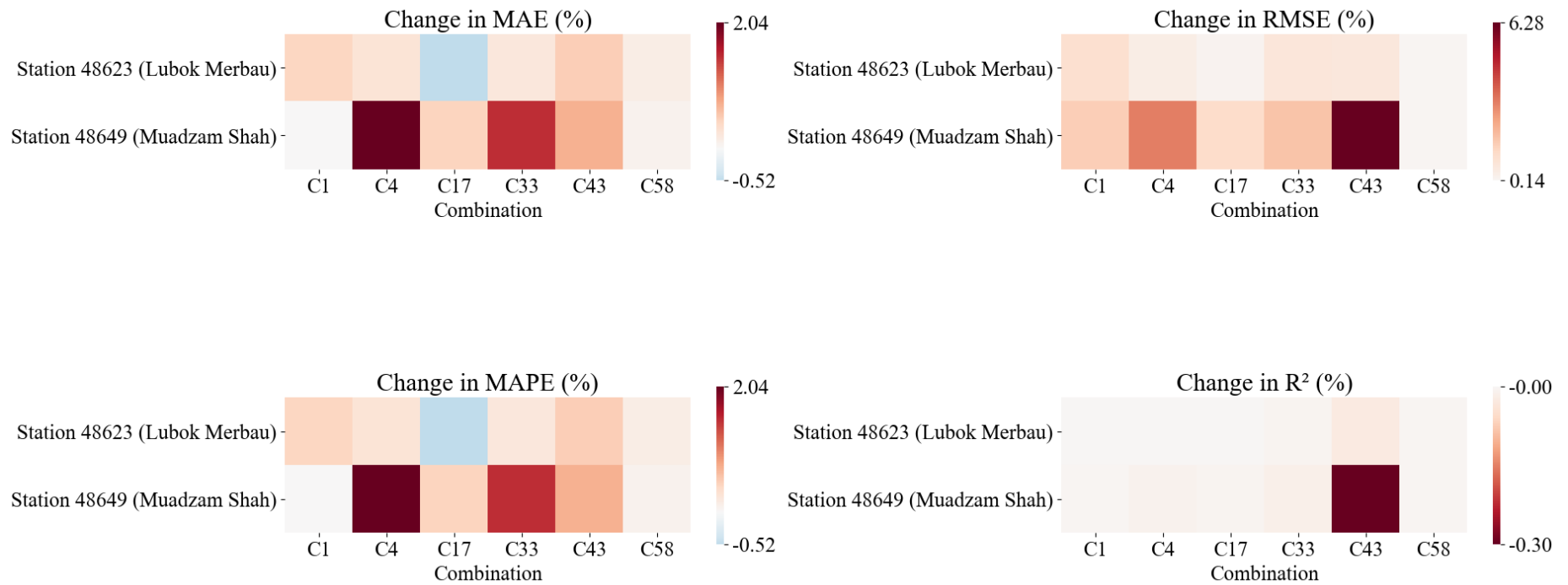
Figure 4.7: Changes in MAE, RMSE, MAPE and R^2 of BMLP (in %) based on MLP for Stations in Cluster 4
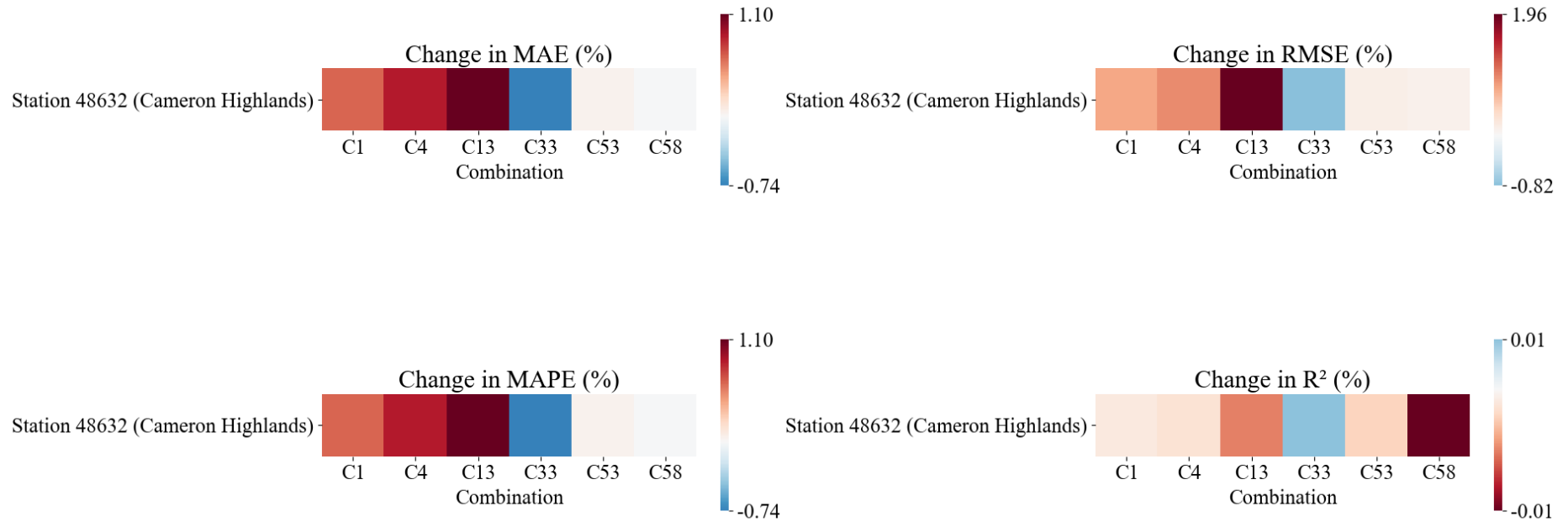
Figure 4.8: Changes in MAE, RMSE, MAPE and R^2 of BMLP (in %) based on MLP for Stations in Cluster 5

To sum up, the bootstrap aggregating did not establish itself as a universal solution to improve the ET_0 estimation accuracy in Peninsular Malaysia. On the contrary, the BMLP exhibited a weaker performance than the base MLP. However, the inability of bootstrap aggregating to show its effect could be due to the excellent performance of the base MLP. The effectiveness of the bootstrap aggregating is yet to be confirmed using other poorer models such as the SVM and the ANFIS.

4.2.2   Bootstrap Aggregating with SVM

Similar to the BMLP, the same bootstrapped dataset was used to train and test the SVM developed in the previous stage. The bagged SVM (BSVM) models evaluated at different stations were compared with the base SVM, and the improvement caused by the bootstrap aggregating hybridisation approach is shown in Figure 4.9 to Figure 4.13. The exact values of the performance metrics can be found in Appendix B (Table B6 – B10).

As shown in the figures, the integration of the bootstrap aggregating to develop BSVM returned discouraging results. There was a clear pattern in the improvement (or deterioration) in the performance of the BSVM as compared to the base SVM. Generally, the BSVM had poorer performance than the base SVM at all the investigated stations, except for Station 48620 (Sitiawan) and Station 48632 (Cameron Highlands) whereby the R^2 was slightly improved

(approximate 0.06 % to 0.07 %). Nevertheless, any attempts to prevent the error metrics from decreasing would result in a decrement in the R^2 values. From Table B6 to Table B10, it can be seen that lower MAE, RMSE and MAPE were normally associated with lower R^2. This could be attributed to the overfitting of the BSVM when lesser meteorological variables were fed as input (C43, C44, C53 and C58), which was similar to the finding of other published work (Logue and Manandhar, 2018). The maximum reduction in R^2 happened in Station 48649 (Muadzam Shah), which was about 1 % when C58 was used as the input combination.

On the other hand, attempts to maintain the generalisability of the BSVM (as shown in the high R^2 value) would instead deteriorate the accuracy of the models. The MAE, RMSE and MAPE of the BSVM trained using these input combinations experienced a maximum increase of about 2.06 %, 2.58 % and 2.06 %, respectively. The increase in the error metrics suggested that the accuracy of the BSVM had been reduced. However, the changes in the R^2 values were negligible. This reflects that the BSVM could not enhance ET_0 estimations in terms of accuracy and generalisability simultaneously.

Although the BSVM also experienced some degree of performance deterioration as compared to the base SVM, it was not as bad as that of the BMLP. This could be caused by the nature of the SVM which has high generalisability, and consequentially enhances its resilience towards overfitting or less accurate estimation. The comparison between the BSVM and SVM

showed observable but insignificant differences. Nevertheless, these results showed that the bootstrap aggregating data fusion technique failed to boost the performance of the model. The reasoning for this discovery is worthy of further in-depth discussion.

The poor performance of the BSVM confirmed that the data centric approach (bootstrap aggregating) is somehow not suitable for the hybridisation of base models to achieve better ET_0 estimation. It can be inferred that the current structure of the available dataset is sufficiently good without any processing or augmentation of data required. Performing bootstrap aggregating on the current dataset bears the risk of increasing the number of outliers, which could hamper or inhibit the development of a good and robust model. This opinion was suggested by Grandvalet (2004). It was stated in the study that the bootstrap aggregating equalises the influences of all observations, whereby good and bad data points would have equal weightages. This results in the deterioration effect, if the outliers in the original dataset have higher influences in the bootstrapped dataset.
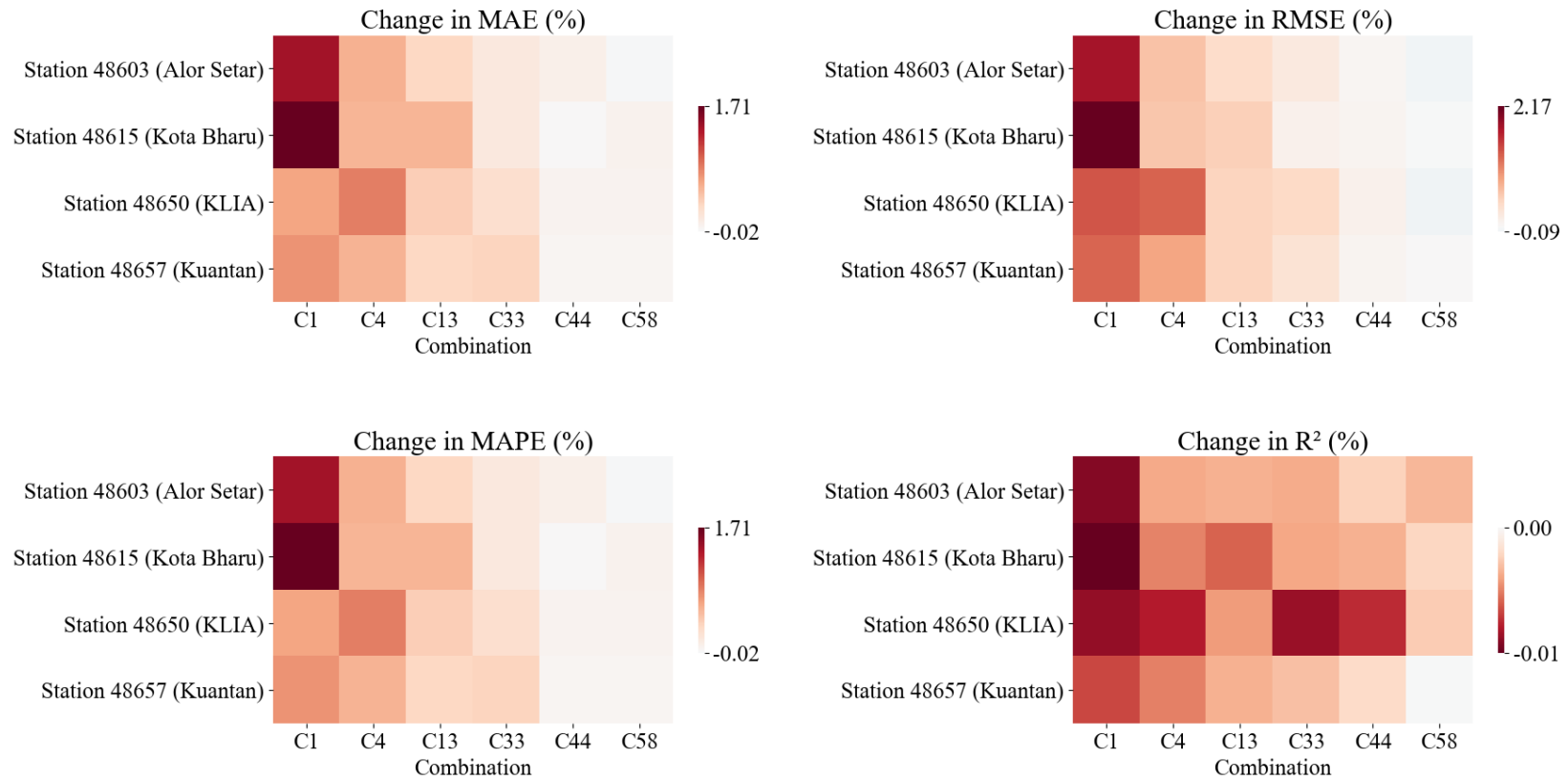
Figure 4.9: Changes in MAE, RMSE, MAPE and R^2 of BSVM (in %) based on SVM for Stations in Cluster 1

130
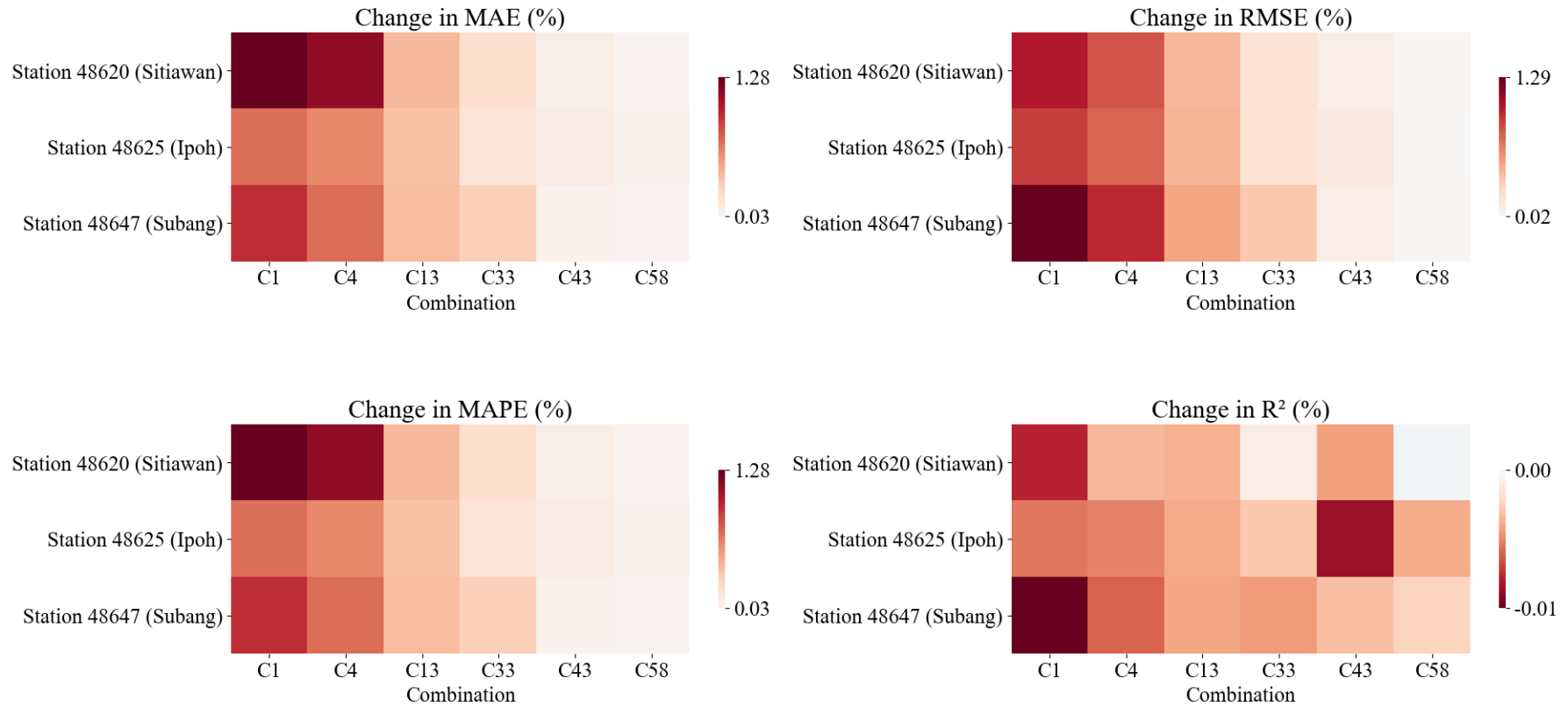
Figure 4.10: Changes in MAE, RMSE, MAPE and R^2 of BSVM (in %) based on SVM for Stations in Cluster 2
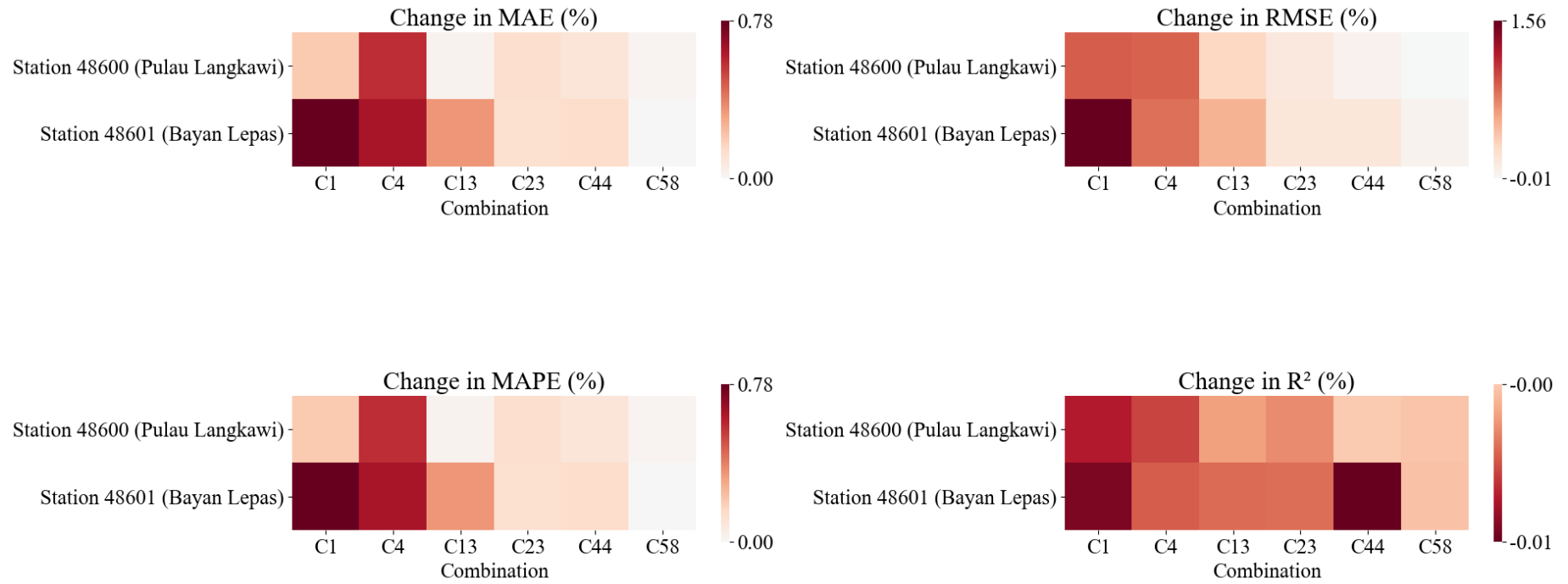
Figure 4.11: Changes in MAE, RMSE, MAPE and R^2 of BSVM (in %) based on SVM for Stations in Cluster 3
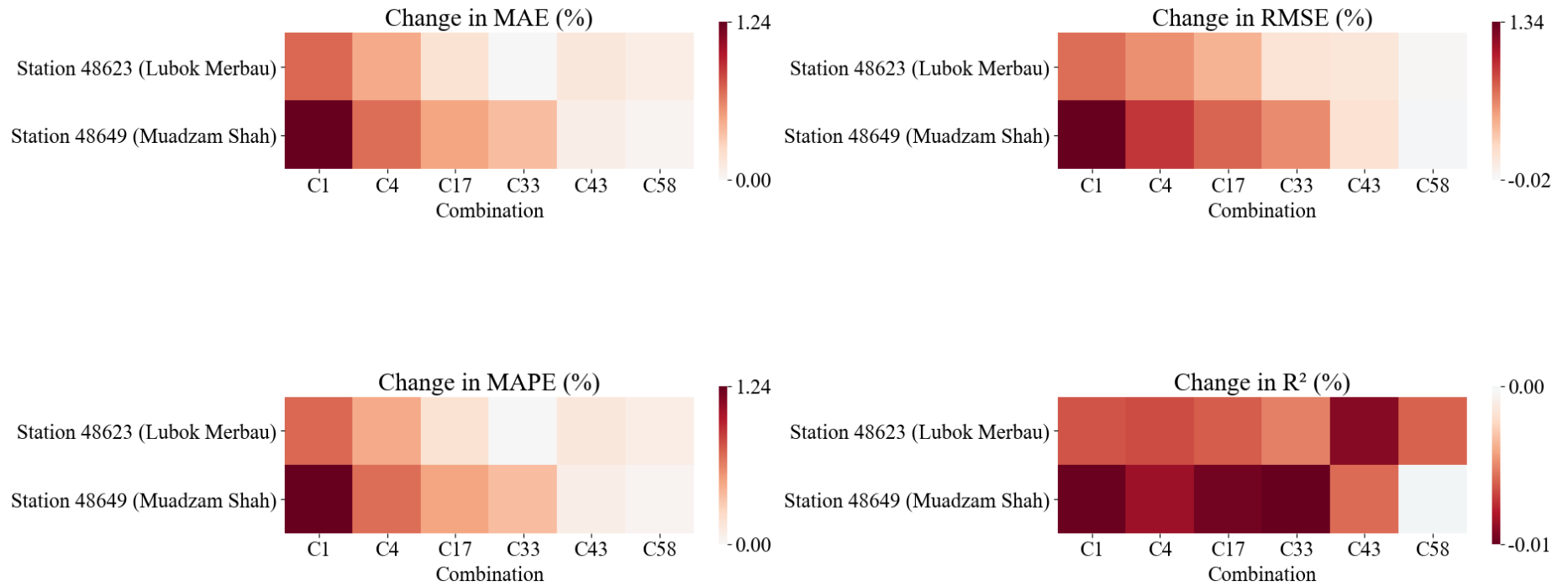
Figure 4.12: Changes in MAE, RMSE, MAPE and R^2 of BSVM (in %) based on SVM for Stations in Cluster 4
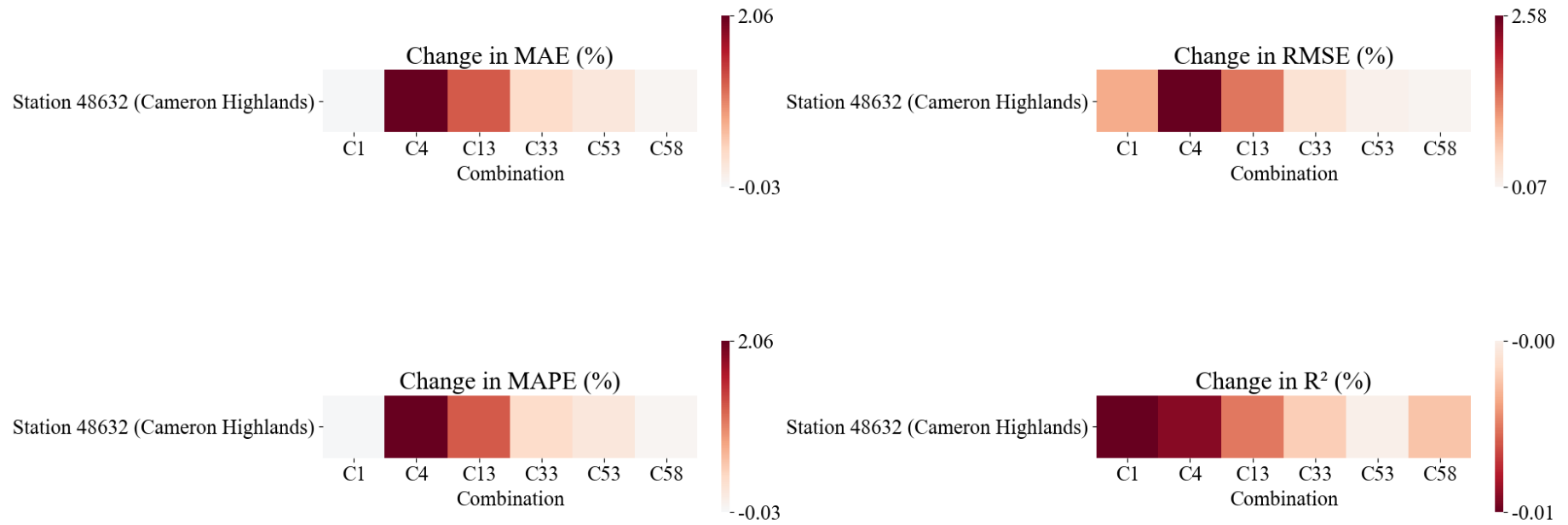
Figure 4.13: Changes in MAE, RMSE, MAPE and R² of BSVM (in %) based on SVM for Stations in Cluster 5

Besides, Skurichina, Kuncheva and Duin (2002) claimed that the bootstrap aggregating can only be useful if the size of the dataset is in proportion with the problems' dimensionalities. It means that lesser data would be required for simple problems and vice versa. The SVM, which is well-known for its good generalisability had its R^2 reduced when lesser meteorological variables were fed as inputs. This is in agreement with the claim of Skurichina, Kuncheva and Duin (2002), whereby when the dimensionality of the problem is reduced (lower number of input meteorological variables), lesser data are required (bootstrap aggregating not required). This conclusion had been reached by other studies (Belayneh, et al., 2016). In the case of the current study, the number of samples within the original dataset had already overwhelmed the problem's dimensionality (maximum of six) and therefore bootstrap aggregating seemed to be a redundant data fusion technique.

4.2.3   Bootstrap Aggregating with ANFIS

Despite the pitfalls experienced by the bootstrap aggregating on the MLP and the SVM, it was still desirable to investigate how the bootstrap aggregating would behave when it is being used as the data fusion technique to hybridise the ANFIS. The improvement of bagged ANFIS (BANFIS) models as compared to the base ANFIS are illustrated in Figure 4.14 to Figure 4.18, for different clusters. The actual values are available in Appendix B (Table B11 – Table B15).

135

Surprisingly, the BANFIS achieved significant improvement, which contrasted with the situation of the BMLP and BSVM. For Cluster 1 and Cluster 5, improvements in MAE, RMSE, MAPE and R^2 were observed for various input combinations. The reduction in MAE, RMSE and MAPE could be as high as 2.5 % coupled with maintained R^2 values. This phenomenon, however, did not appear in the case of Cluster 2, Cluster 3 and Cluster 4.

For the improvement shown in Cluster 1 and Cluster 5, the ET_0 estimation using the BANFIS was improved in terms of accuracy or model generalisability except when C1 was used as the input combination. This could be due to the fact that the performance of ANFIS trained using C4, C13, C33, C44, C53 and C58 had relatively poor performance which made them have greater room for improvement when bootstrap aggregating was integrated. On the contrary, the performance of ANFIS fed with C1 had comparably better performance, which resulted in the deterioration in model accuracy and generalisability after the integration of the bootstrap aggregating data fusion technique.
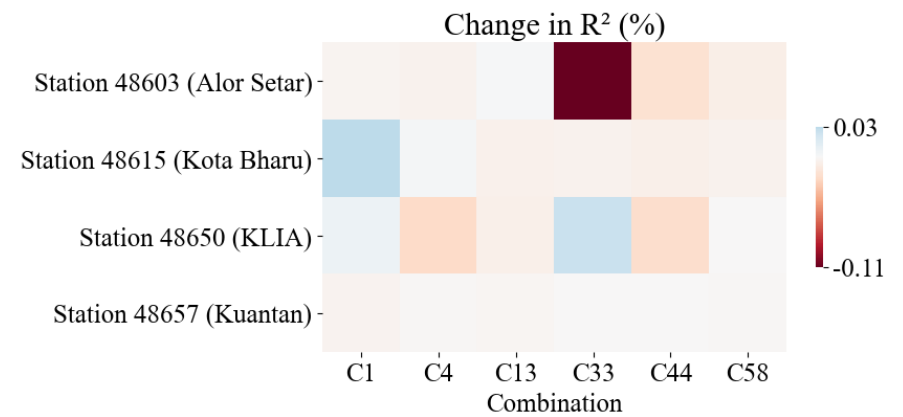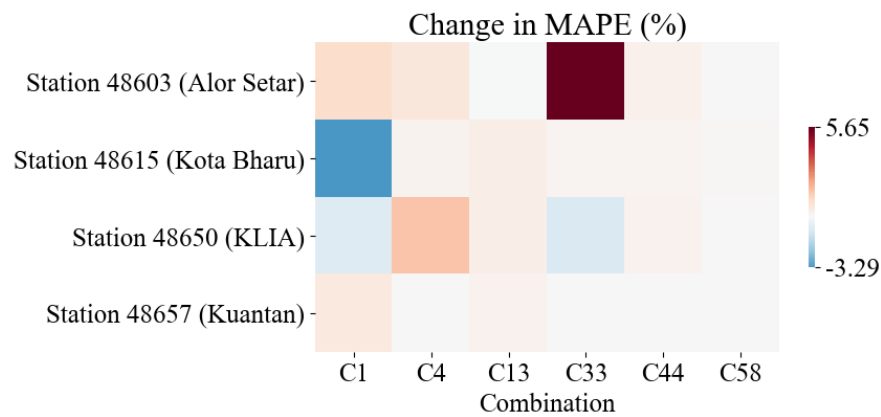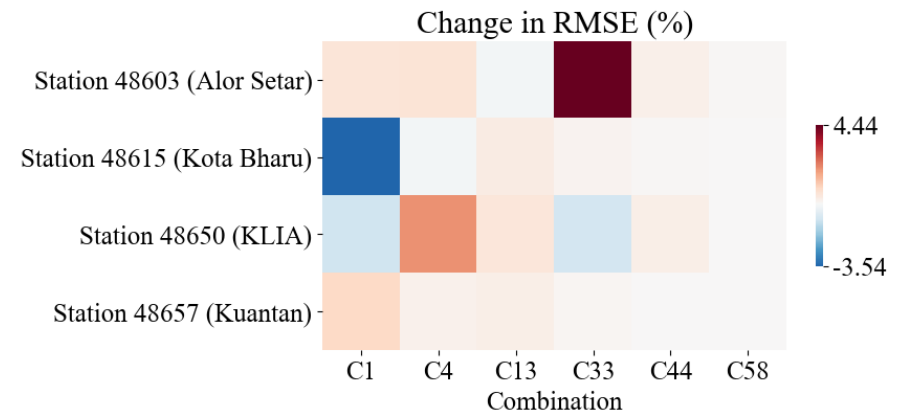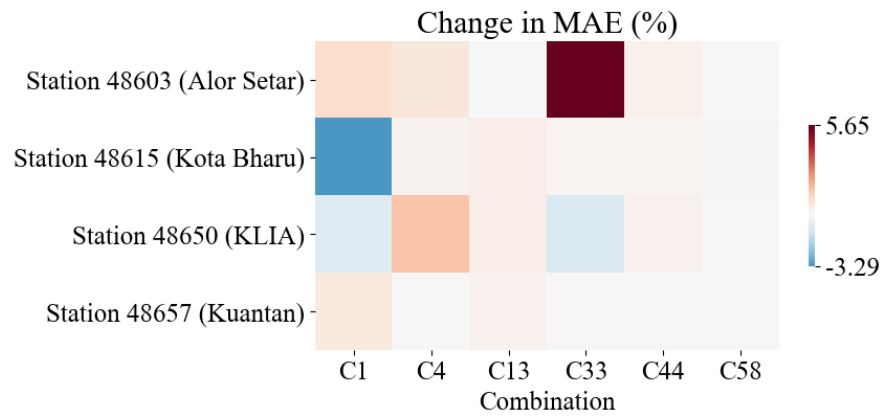
Figure 4.14: Changes in MAE, RMSE, MAPE and R² of BANFIS (in %) based on ANFIS for Stations in Cluster 1
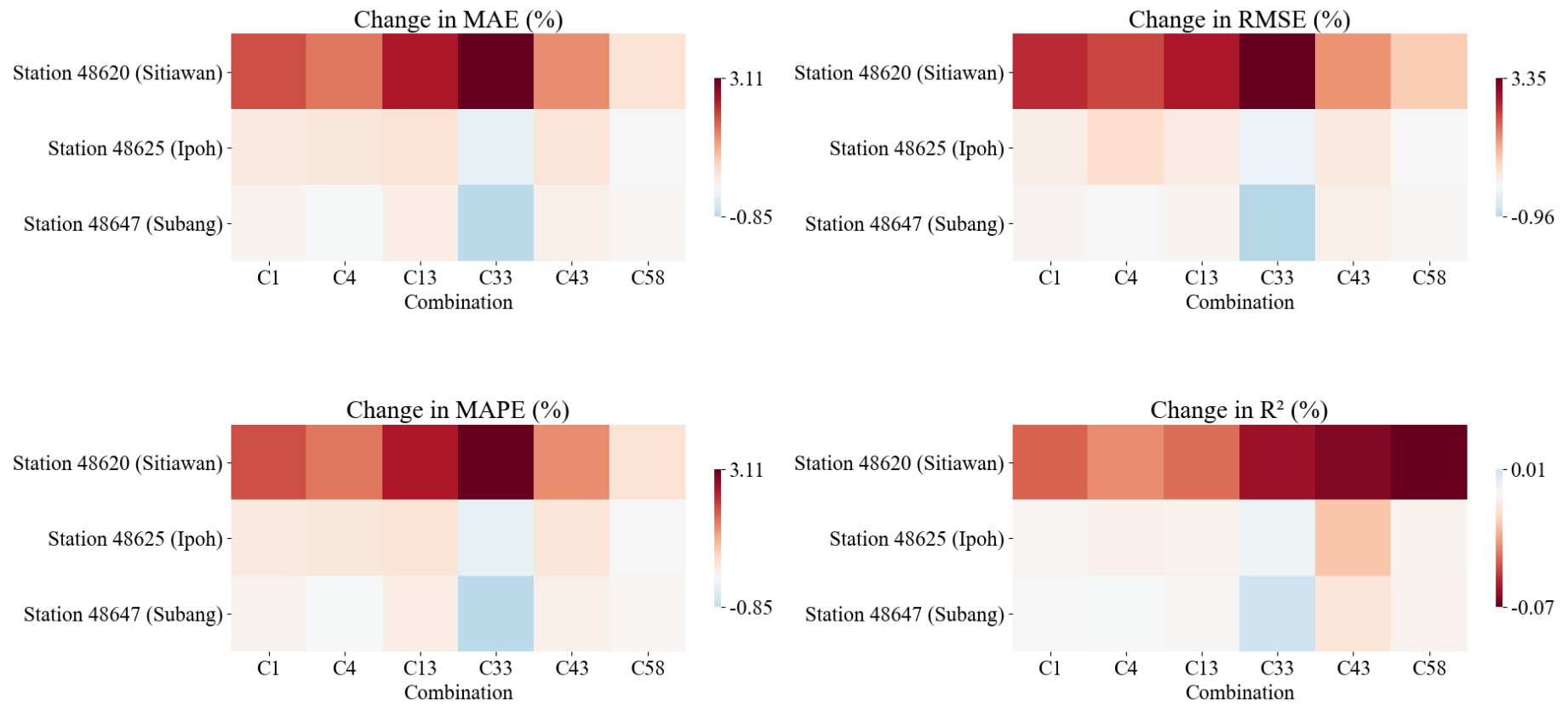
Figure 4.15: Changes in MAE, RMSE, MAPE and R^2 of BANFIS (in %) based on ANFIS for Stations in Cluster 2
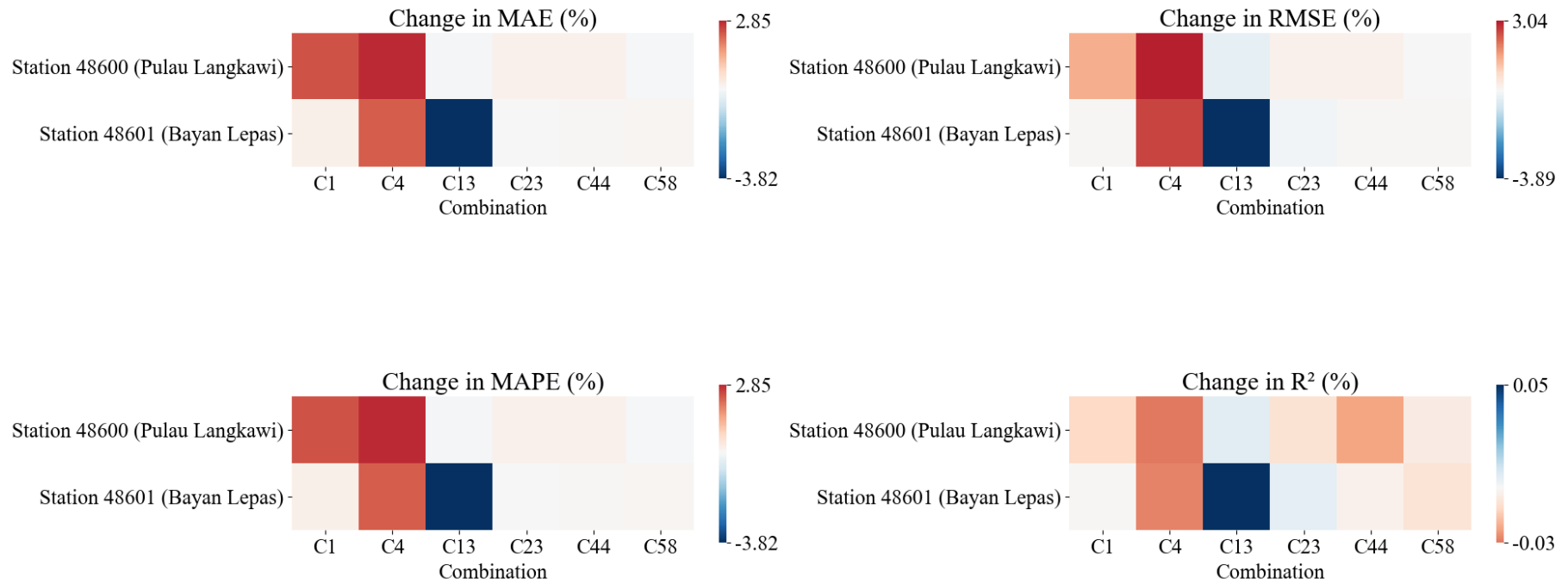
Figure 4.16: Changes in MAE, RMSE, MAPE and R² of BANFIS (in %) based on ANFIS for Stations in Cluster 3
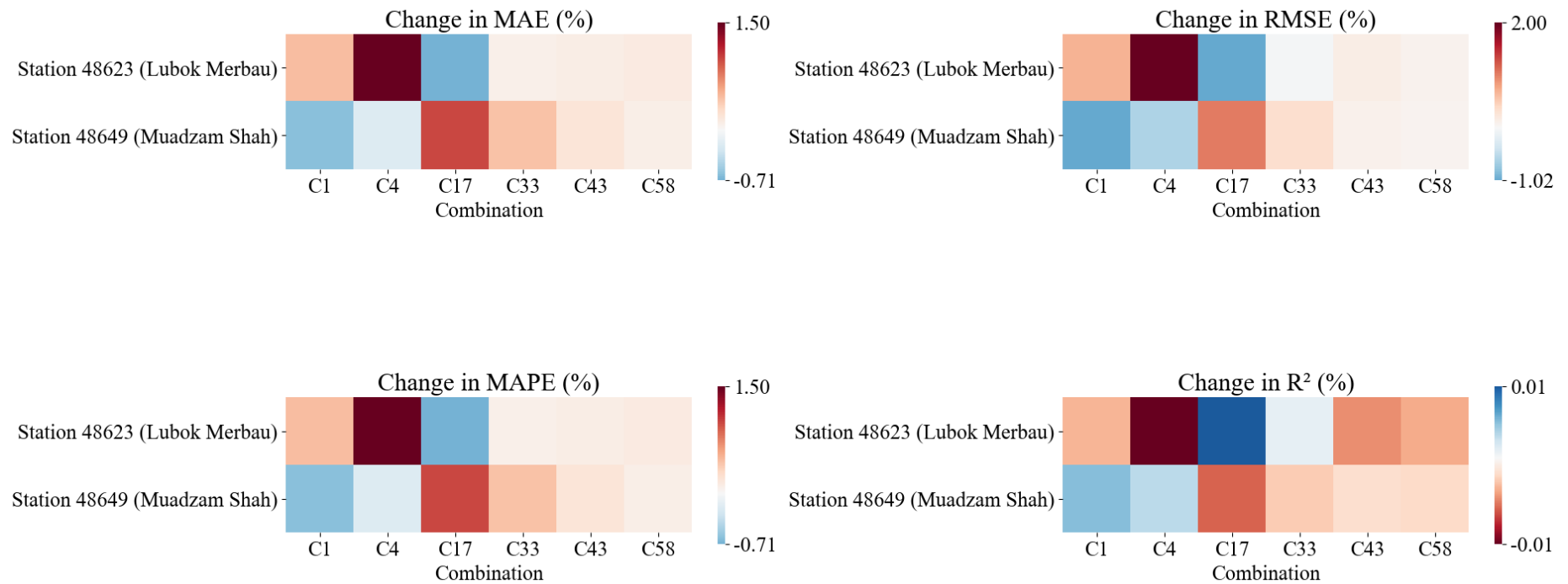
Figure 4.17: Changes in MAE, RMSE, MAPE and R^2 of BANFIS (in %) based on ANFIS for Stations in Cluster 4
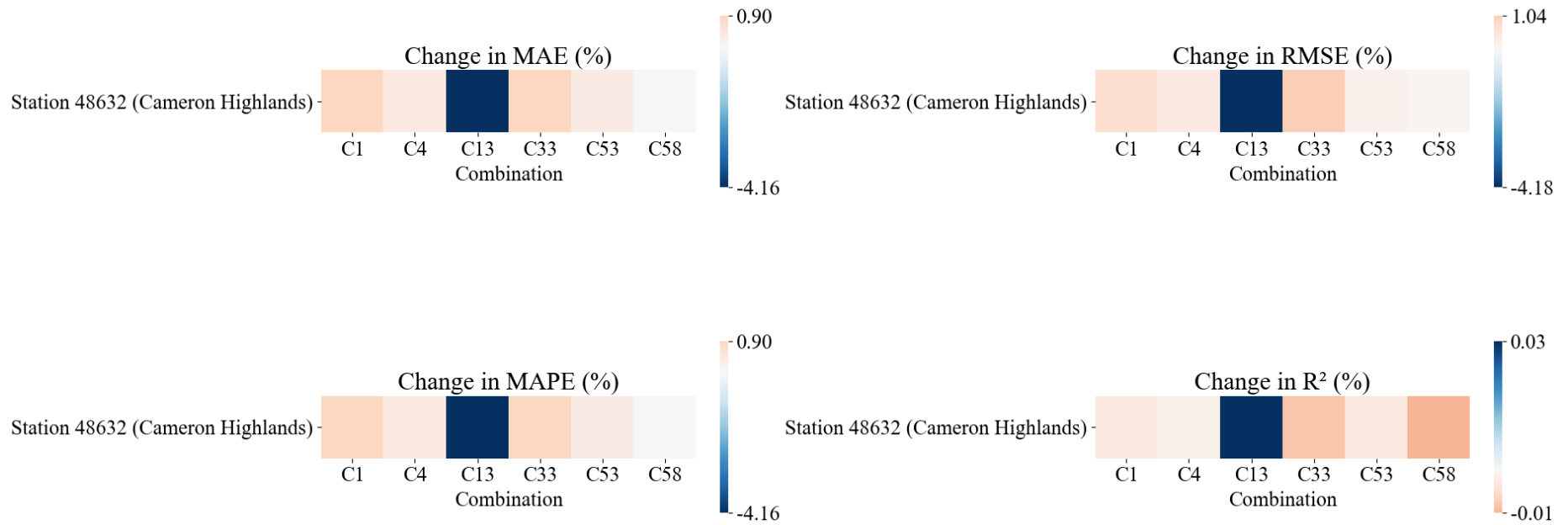
Figure 4.18: Changes in MAE, RMSE, MAPE and R² of BANFIS (in %) based on ANFIS for Stations in Cluster 5

For Cluster 3 (Figure 4.16) and Cluster 4 (Figure 4.17), it is observable that the bootstrap aggregating had multiple effects on the ANFIS, depending on the selection of input combinations. Whereas for Cluster 2 (Figure 4.15), the BANFIS generally had poorer performance than the base ANFIS. This finding, in fact, was parallel to those found in the BMLP and BSVM. Overall speaking, the bootstrap aggregating data fusion technique had heterogeneous impact on the ANFIS, as shown by the comparison between the BANFIS and base ANFIS at stations of different clusters.

4.2.4 Summary

The integration of the bootstrap aggregating data fusion technique to each of the base MLP, SVM and ANFIS had caused some impacts. Overall, the BMLP, BSVM and BANFIS had performed slightly worse than their corresponding base model. Nonetheless, the degree of accuracy and generalisability of the models were still satisfactory.

Although the BMLP and the BANFIS had significantly improved performance at certain meteorological stations, this inference that the data centric approach was not an effective data fusion technique in the context of ET_0 estimation remained intact. This deduction is further supported by past literature which suggests that the bootstrap aggregating can only be helpful when there is an appropriate ratio between the size of dataset and the dimensionality of the problem. In fact, naively bootstrapping the sample bears

the risk of assigning higher influences to the "bad" leverages, which could be detrimental to the model's accuracy.

To sum up, the bootstrap aggregating is not a suitable data fusion technique if a better ET_0 estimating model is desired. The structure of the available data obtained from the authority is sufficiently good, without any alteration, modification or augmentation performed on it.

4.3   Data Fusion II: Bayesian Model Averaging

Besides the data centric bootstrap aggregating, another data fusion technique, the BMA was also integrated into the base models. Unlike the bootstrap aggregating that altered the data structure, the BMA aggregates the participating models and produces a better estimation of the ET_0. The details of the findings are reported in the subsections below.

4.3.1   Bayesian Weight

The posterior probability of the MLP, SVM and ANFIS fed with different input combinations can be obtained using Equation (3.8). Subsequently, the weight of each model can be calculated by comparing the relative magnitude of their posterior probabilities (Equation (3.9) and Equation (3.10)). The Bayesian weights of the base MLP, SVM and ANFIS at different meteorological stations using different numbers of input meteorological

variables are tabulated in Table 4.3, where the sum of the Bayesian weights for different models using a particular input combination is unity.

From Table 4.3, it can be seen that the Bayesian weight assignments for the MLP, SVM and ANFIS are unique at different meteorological stations. In fact, the clustering effect (as shown in the selection of input combination) was not evident in the case of the Bayesian weight assignment. For instance, the distribution of Bayesian weights across the MLP, SVM and ANFIS were different for Station 48603 (Alor Setar), Station 48615 (Kota Bharu), Station 48650 (KLIA) and Station 48657 (Kuantan), even though all of these stations were from Cluster 1.

It is important to note that the MLP was usually given the highest weight. This coincides with the fact that the MLP outperformed the other two models in the comparison among the base models. However, as the number of input meteorological variables reduced, the dominance of the MLP decreased accordingly. The rarefaction left behind by the shrinkage of the MLP's dominance was filled by the other two base models, whereby the ANFIS seemed to be favoured over the SVM. The Bayesian weight of the ANFIS gradually increased as the number of input meteorological variables was reduced. This is because the performance of the ANFIS was at par with the MLP when lesser meteorological variables were used due to it being more resilient towards the reduction of input meteorological variables.

As opposed to the case of the ANFIS, the SVM did not receive much attention from the model centric BMA approach. This could be due to the nature of SVM, which tends to take a balance between accuracy and generalisability. This, in turn, results in the SVM being a model without any "outstanding traits" to be justified with a higher Bayesian weight. Nevertheless, an exception occurred at Station 48620 (Sitiawan) where the SVM was given high Bayesian weights, regardless of the number of input meteorological variables. This could be traced back to the exceptionally excellent performance of the SVM at the aforementioned station (refer to Appendix A).

It is worth mentioning that if the Bayesian weight of unity is given to a particular model, it means that the BMA approach is converted into the BMS approach. The Bayesian modelling approach identifies the model as the one and only "true" model from all the model candidates. This had happened several times for the MLP, especially when the number of input meteorological variables was relatively high. Nonetheless, the qualification of "true" model can be changed if the Bayesian modelling approach is shown with different sets of candidates, and so are the Bayesian weights. Hence, the results tabulated in Table 4.3 may not be reproducible if models other than the MLP, SVM and ANFIS are included in the candidature set.

Table 4.3: Bayesian Weights for MLP, SVM and ANFIS at Different Meteorological Stations Using Different Number of Input Meteorological Variables

Station	Model	Number of Input Meteorological Variables					
		6	5	4	3	2	1
Station 48600 (Pulau Langkawi)	MLP	1.0000	1.0000	0.0064	-	0.6777	0.4742
	SVM	-	-	0.9936	1.0000	0.0045	-
	ANFIS	-	-	-	-	0.3178	0.5258
Station 48601 (Bayan Lepas)	MLP	1.0000	1.0000	1.0000	0.9989	0.9973	0.4713
	SVM	-	-	-	0.0011	0.0027	-
	ANFIS	-	-	-	-	-	0.5287
Station 48603 (Alor Setar)	MLP	1.0000	1.0000	1.0000	1.0000	0.3057	0.4440
	SVM	-	-	-	-	0.3027	-
	ANFIS	-	-	-	-	0.3916	0.5560
Station 48615 (Kota Bharu)	MLP	1.0000	1.0000	0.9998	0.4450	0.8105	0.3658
	SVM	-	-	0.0002	0.2430	0.1448	0.0001
	ANFIS	-	-	-	0.3120	0.0447	0.6341
Station 48620 (Sitiawan)	MLP	0.3781	-	-	-	-	-
	SVM	0.6219	1.0000	1.0000	1.0000	0.9891	0.9999
	ANFIS	-	-	-	-	0.0109	0.0001
Station 48623 (Lubok Merbau)	MLP	1.0000	1.0000	1.0000	1.0000	0.9683	0.7355
	SVM	-	-	-	-	-	-
	ANFIS	-	-	-	-	0.0317	0.2645
Station 48625 (Ipoh)	MLP	1.0000	1.0000	1.0000	0.9989	0.0011	0.4576
	SVM	-	-	-	0.0011	-	-
	ANFIS	-	-	-	-	0.9989	0.6570
Station 48632 (Cameron Highlands)	MLP	1.0000	1.0000	1.0000	1.0000	0.0891	0.8123
	SVM	-	-	-	-	-	-
	ANFIS	-	-	-	-	0.9109	0.1877
Station 48647 (Subang)	MLP	1.0000	1.0000	1.0000	1.0000	0.2873	0.6294
	SVM	-	-	-	-	0.0060	0.0128
	ANFIS	-	-	-	-	0.7067	0.3578
Station 48649 (Muadzam Shah)	MLP	1.0000	1.0000	1.0000	1.0000	0.3143	0.3501
	SVM	-	-	-	-	0.0001	0.0476
	ANFIS	-	-	-	-	0.6856	0.6023
Station 48650 (KLIA)	MLP	1.0000	1.0000	1.0000	1.0000	0.9399	0.3430
	SVM	-	-	-	-	0.0001	-
	ANFIS	-	-	-	-	0.0601	0.6570
Station 48657 (Kuantan)	MLP	1.0000	1.0000	1.0000	0.9996	0.8692	0.5659
	SVM	-	-	-	-	0.0495	0.0373
	ANFIS	-	-	-	0.0004	0.0813	0.3968

Inter-model ensemble can be built by using the assigned Bayesian weights. In this study, the inter-model ensemble developed using the model centric BMA approach will be abbreviated as BMA-E, where the "BMA" represents the method, while "E" is short for ensemble.

4.3.2   Inter-Model Ensemble using BMA

The performance of BMA-E was studied at different stations, using their preferred input combinations according to the clustering effect. The performance of the BMA-E, in terms of MAE, RMSE, MAPE, R^2 and MBE for the stations in different clusters, are summarised in Figure 4.19 to Figure 4.23. The exact values of the performance evaluation metrics can be referred to Appendix C.
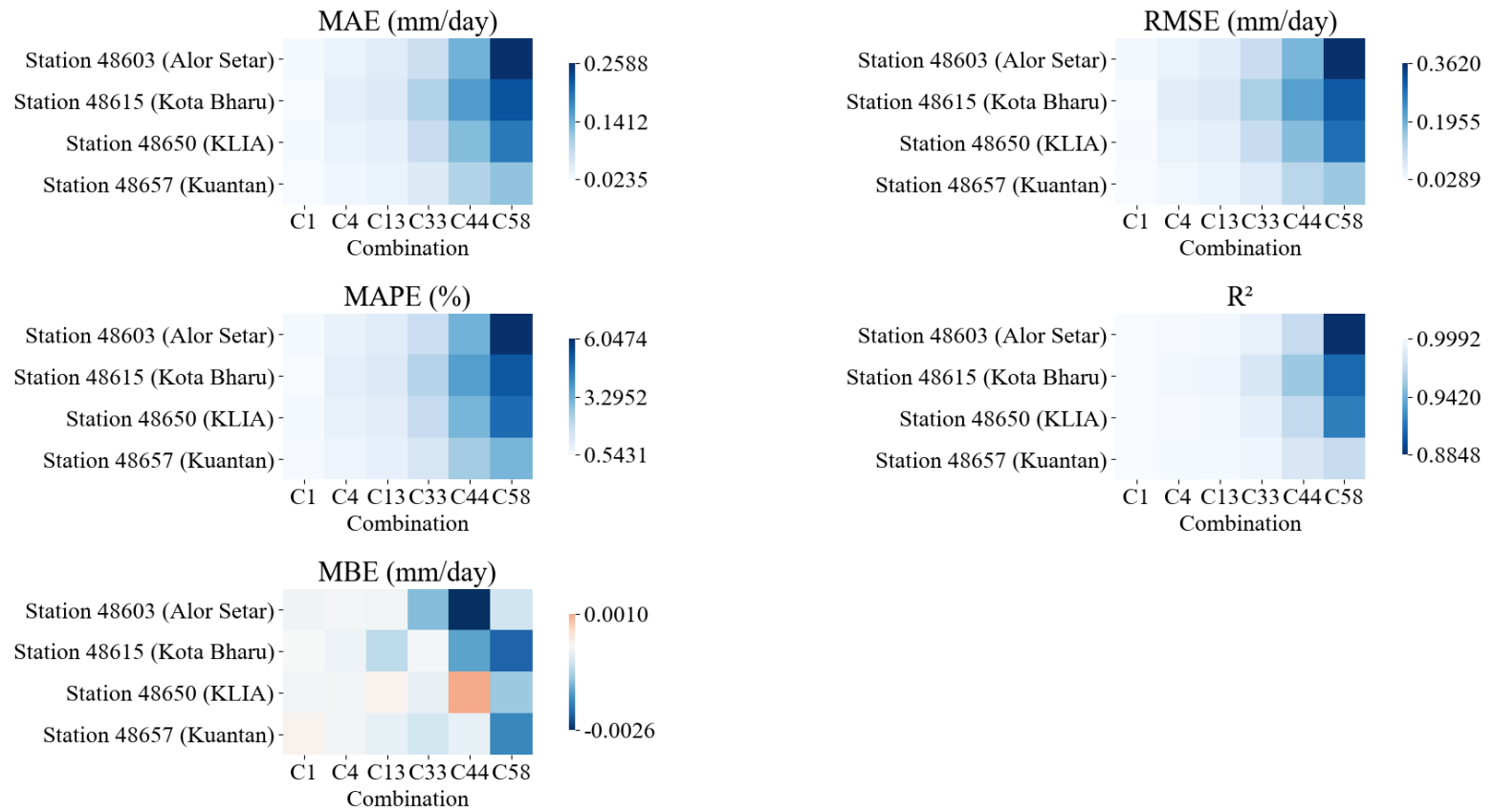
Figure 4.19: Performance of BMA-E in ET₀ Estimation using Different Input Combinations for Stations in Cluster 1

148
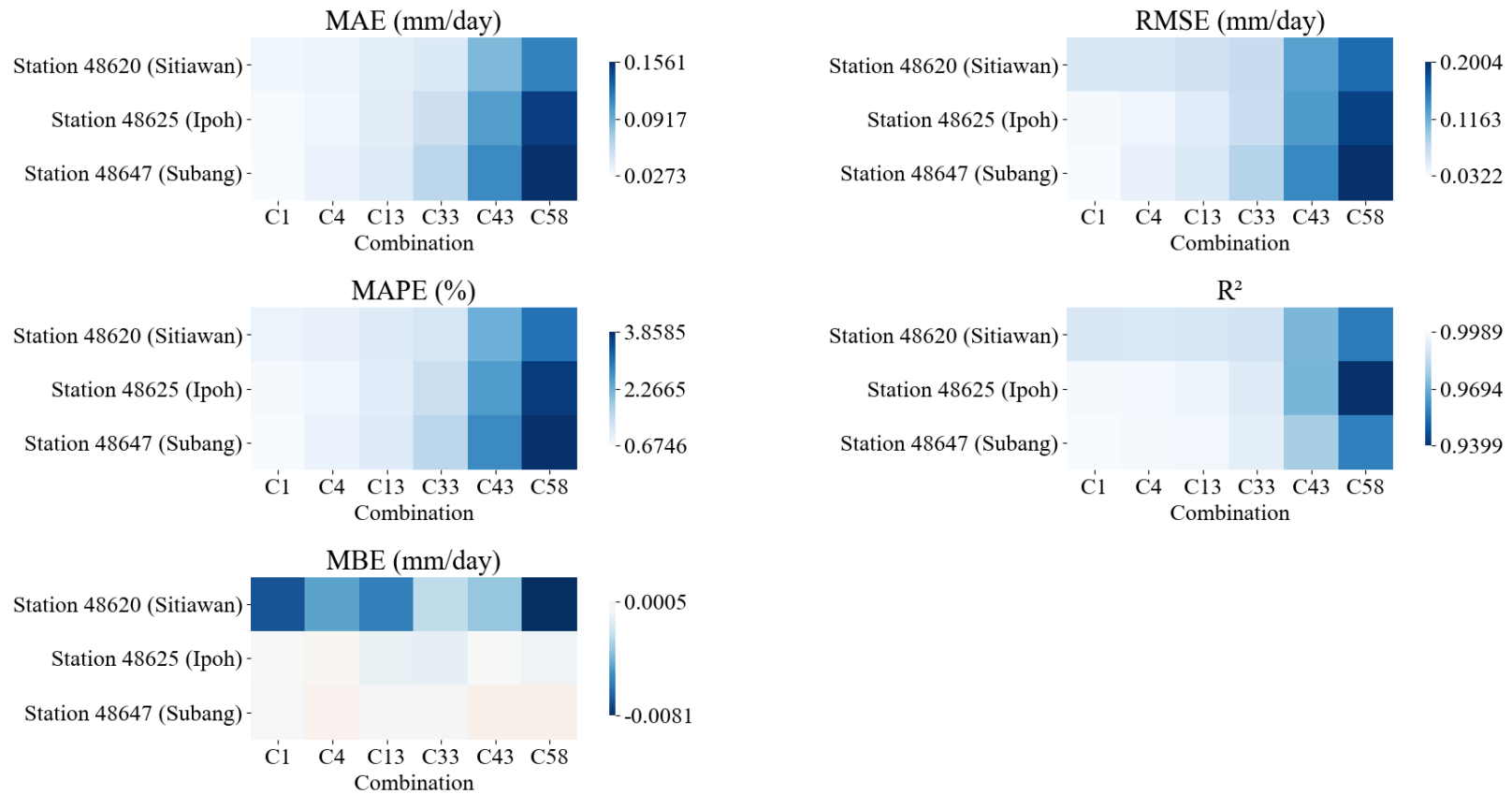
Figure 4.20: Performance of BMA-E in ET₀ Estimation using Different Input Combinations for Stations in Cluster 2

149
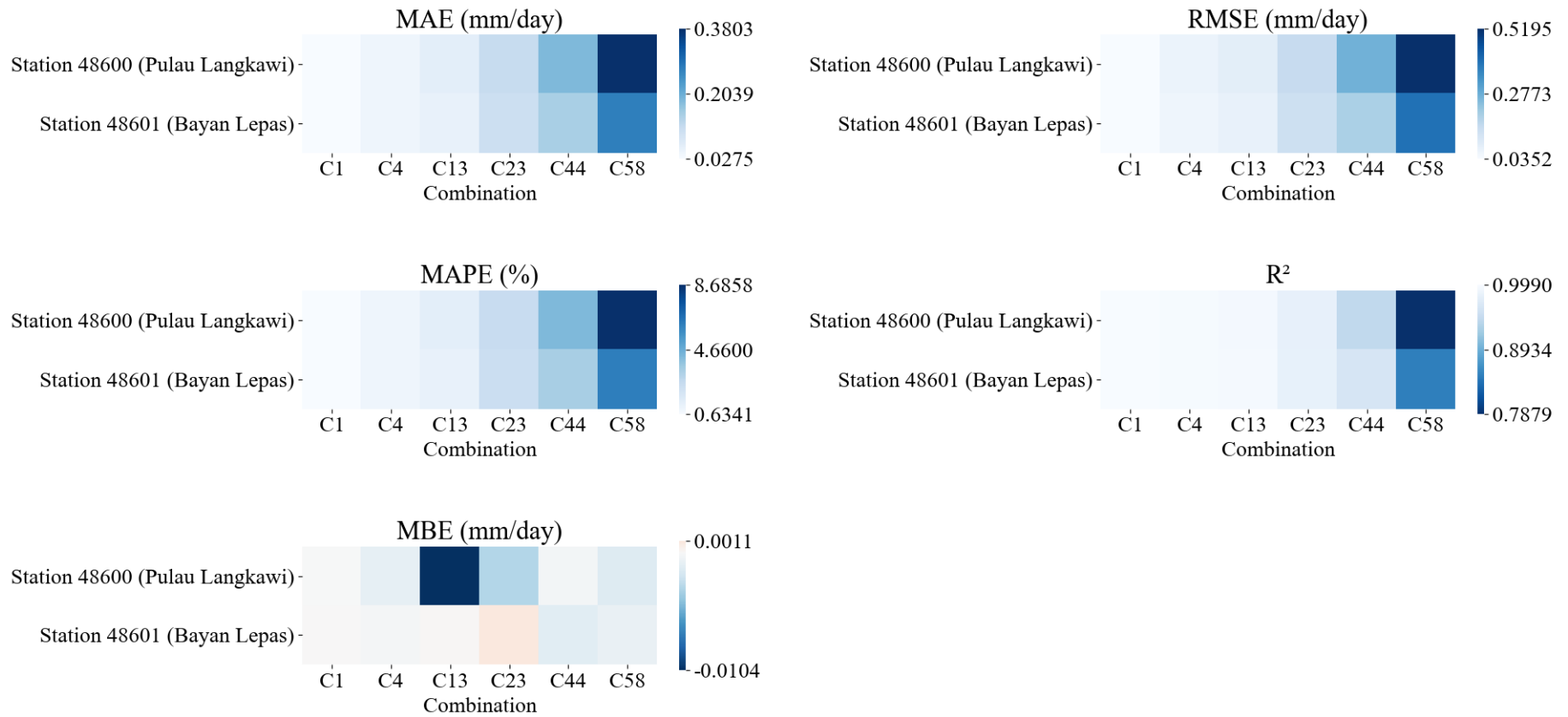
Figure 4.21: Performance of BMA-E in ET₀ Estimation using Different Input Combinations for Stations in Cluster 3
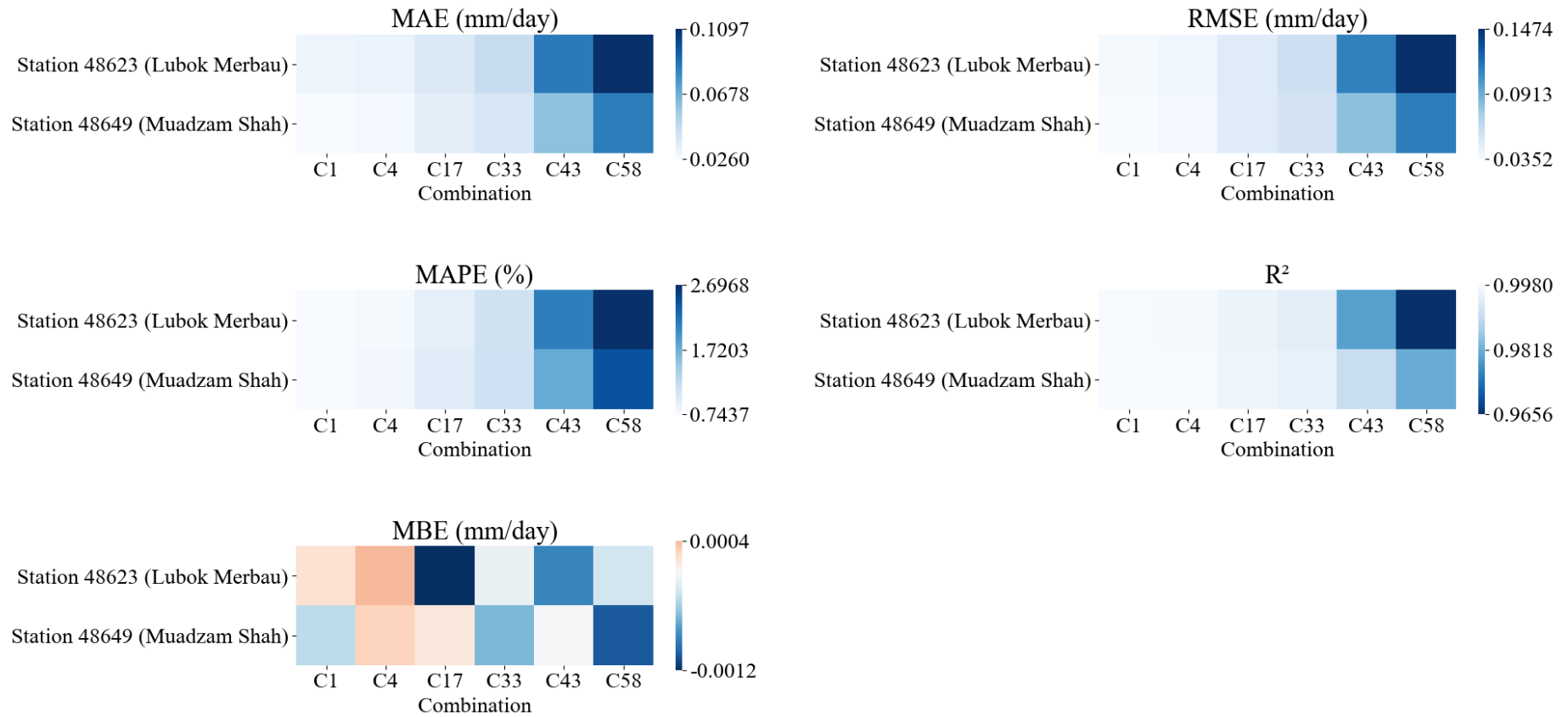
Figure 4.22: Performance of BMA-E in ET₀ Estimation using Different Input Combinations for Stations in Cluster 4

151

MAE (mm/day)

Station 48632 (Cameron Highlands)

C1 C4 C13 C33 C53 C58
Combination

0.1645
0.0270

RMSE (mm/day)

Station 48632 (Cameron Highlands)

C1 C4 C13 C33 C53 C58
Combination

0.2055
0.0323

MAPE (%)

Station 48632 (Cameron Highlands)

C1 C4 C13 C33 C53 C58
Combination

9.5461
5.2003
0.8545

R²

Station 48632 (Cameron Highlands)

C1 C4 C13 C33 C53 C58
Combination

0.9991
0.9367
0.8744

MBE (mm/day)

Station 48632 (Cameron Highlands)

C1 C4 C13 C33 C53 C58
Combination

0.0002
0.0011
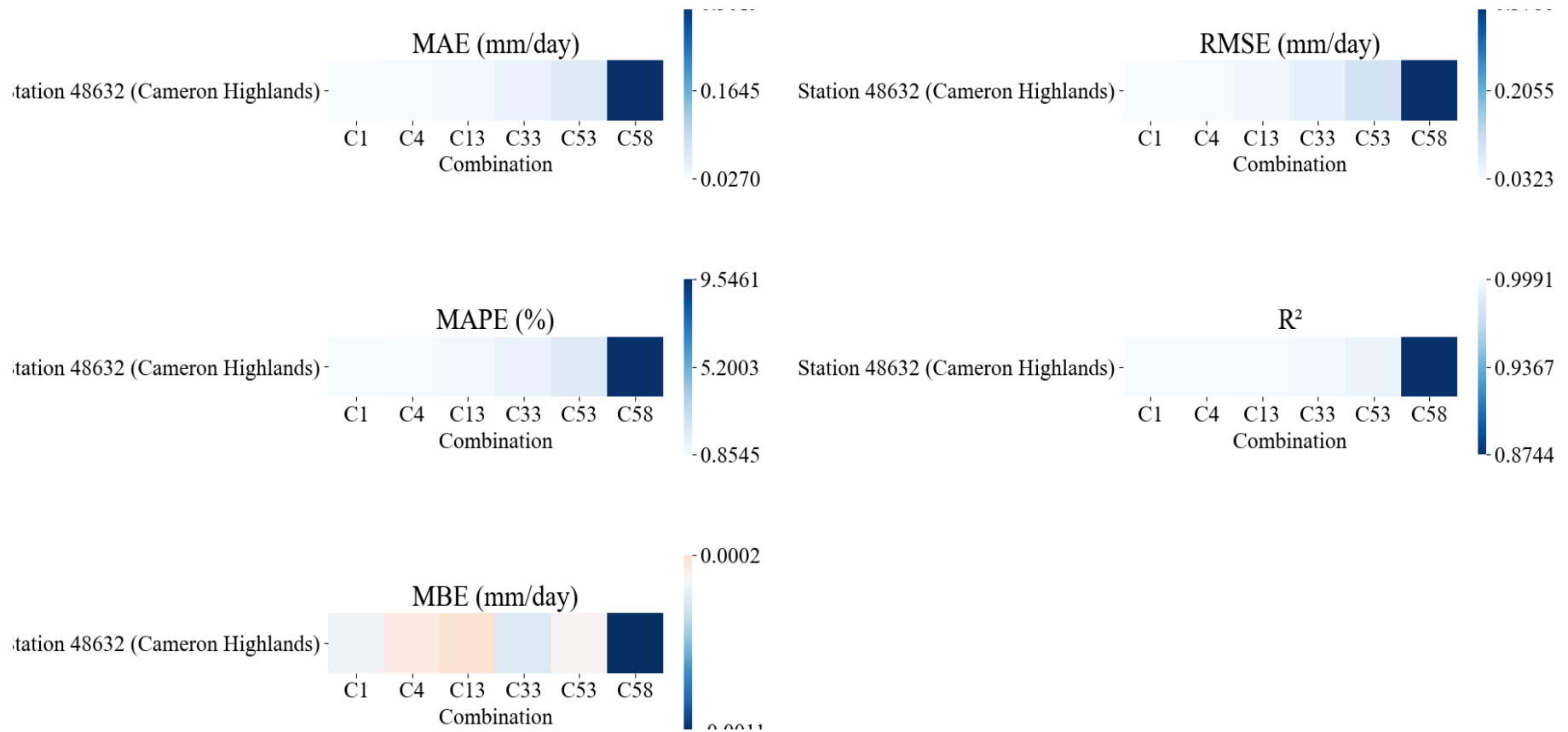
Figure 4.23: Performance of BMA-E in ET$_0$ Estimation using Different Input Combinations for Stations in Cluster 5

152

As compared to the base MLP, SVM and ANFIS, the BMA-E evidently enhanced the ET_0 estimation, as shown in Figure 4.19 to Figure 4.23. There were several cases whereby the BMA algorithm was switched to BMS as shown by the assignment of a unity weight, comparison of such BMA-E is less meaningful as the resultant BMA-E merely inherited the performance of the selected base models. This can be observed from Table 4.3 where some models were given Bayesian weights of 1. Attention shall be given to the BMA-E where several base models were involved in the decision committee.

For illustration purposes, the BMA-E developed at Station 48603 (Alor Setar) using the C44 as the input combination was chosen for the detailed explanation. In this BMA-E, the MLP, SVM and ANFIS were assigned with Bayesian weights of 0.3057, 0.3027 and 0.3916, respectively (refer to Table 4.3). In other words, the BMA algorithm opined that the ANFIS was the most truthful model with 39.16 % confidence, followed by MLP (30.57 %) and SVM (30.27 %). Again, the Bayesian weights assigned were solely based on the available model candidates, which can be different if another set is given. Although not remarkable, the BMA-E exhibited better performance at almost all aspects as compared to the base models. The comparison of the BMA-E with the base models is summarised in Table 4.4.

Table 4.4 clearly shows how the BMA algorithm managed to combine the beneficial attributes or characteristics of different base models to produce a better decision committee. In this case, the base SVM exhibited good accuracy

153

with low MAE and MAPE values. However, this model also suffered from the lack of generalisability as depicted by the obvious overfitting. On the other hand, the MLP and ANFIS excelled in maintaining high R^2 and low MBE which corresponded to good generalisability. As a result, the BMA-E which considered all the three models in its decision committee managed to achieve better accuracy than the SVM alone, at the same time had its generalisability improved by the inclusion of the MLP and ANFIS. The BMA-E in fact, outperformed all its constituent base models. Although the improvement was marginal, however, this proved that the model centric data fusion technique based on the BMA showed a positive and promising impact in improving the estimation of ET_0.

Table 4.4: Comparison of BMA-E with the Base Models at Station 48603 (Alor Setar) using C44 as Input Combination

Model	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE (mm/day)
BMA-E	0.1373	0.1826	3.2080	0.9707	-0.0026
MLP	0.1383	0.1835	3.2312	0.9704	-0.0011
SVM	0.1376	0.1835	3.2160	0.9705	-0.0072
ANFIS	0.1379	0.1834	3.2236	0.9703	-0.0000

Besides, there were also several occasions where only two of the base models were considered by the BMA algorithm. This can be seen in the BMA-E developed at Station 48603 (Alor Setar) when C58 as the input combination (Table 4.5). In this case, the BMA algorithm was confident that the MLP was 44.40 % true, while the remaining 56.60 % confidence was given to the ANFIS (Bayesian weights can be referred from Table 4.3).

Table 4.5: Comparison of BMA-E with the Base Models at Station 48603 (Alor Setar) using C58 as Input Combination

Model	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE (mm/day)
BMA-E	0.2588	0.3620	6.0474	0.8848	-0.0005
MLP	0.2589	0.3622	6.0496	0.8848	-0.0012
SVM	0.2499	0.3744	5.8405	0.8839	-0.0840
ANFIS	0.2591	0.3621	6.0538	0.8847	-0.0000

The BMA algorithm omitted the SVM in this case, probably due to the extremely high MBE which could possibly deteriorate the performance of the BMA-E as a whole. Hence, by only comparing the BMA-E with the base MLP and ANFIS, the BMA data fusion technique once again proved its ability in improving the ET_0 performance based on the desirable properties of the selected constituent base models.

4.3.3 Summary

Figure 4.19 to Figure 4.23 show the overall performance of the BMA-E at different stations using their preferred input combinations (according to clusters) was satisfying. This is the combinatory effect of the selection of suitable input combinations, coupled with the positive influence brought about by the integration of the BMA algorithm. This study has proved the feasibility of this model centric data fusion approach for better ET_0 estimation in Peninsular Malaysia. In fact, the Bayesian weights of the models in the ensemble were revealed, for the first time in the article published under this research work (Chia, Huang and Koo, 2021a). This is inspired by a previous

work that utilised the Bayesian-based algorithms for ET estimation using the product of remote sensing database (He, et al., 2020; Sun, et al., 2019). This makes the resultant machine learning model (BMA-E) more explainable and can be studied at greater resolution by the scientific community.

Nevertheless, through this research work, the shortcomings of the BMA data fusion technique were also exposed. Firstly, it is difficult for the BMA data fusion technique to include all the models as its constituents. Consequentially, some favourable properties of the neglected or omitted models are being sacrificed in the process of developing the BMA-E. Most of the time, this algorithm is reasonable in ensuring the optimum performance of the ensemble. However, this would also hamper the propagation of information extracted by the base models into the ensemble. In other words, the rigid structure of BMA data fusion technique lacks the flexibility in dealing with the noble characteristics of its possible constituent models.

Secondly, the rigidity of the BMA data fusion technique had resulted on many occasions where the BMA-E were not technically hybrid models or ensembles. This can be seen in many cases when the Bayesian weight of 1 was given to only one model (MLP in most of the cases). That is to say, the so-called BMA-E was only inheriting everything from the MLP instead of being a hybrid model. This had caused many hybridisations failing to take place and enhancement in ET_0 estimation cannot be realised. Having said that, the

statistics-based BMA data fusion technique is still preferred over the base models, given a conducive modelling environment and favourable conditions.

4.4 Data Fusion III: Non-Linear Neural Ensemble

Like the BMA, the NNE also improves the estimation of ET_0 by aggregating the outputs of the individual base models. However, it does not have any statistical theory or rule to govern the process of the aggregation. This study used a separate neural network as the meta-learner that performed the aggregation. The following subsection is dedicated to analysing and discussing the results for this part of the research work.

4.4.1 WOA-ELM as Meta-Learner

Unlike the data centric bootstrap aggregation and the model centric BMA, the NNE approaches the problem by resorting to the black-box operation (hence the name non-linear). In this study, a novel meta-learner was developed as the "black-box agent" to complete the NNE. The ELM, an efficient variant of ANN was hybridised using the metaheuristic WOA. The output of the base MLP, SVM and ANFIS was fed into the WOA-ELM meta-learner, which resulted in the final NNE-based hybrid model, addressed as the WOA-ELM-E in the text.

The ELM contained a single hidden layer with a tuneable number of hidden neurons. By using the WOA, the number of hidden neurons was fine-tuned through a feedback mechanism based on a fitness value calculated using Equation 3.16. Hence, the stochastic training attribute was inculcated into the base ELM to enable it to have the self-tune ability. The "secondary" training stage allowed the WOA-ELM-E to have a second look at the target values in the training data. This enabled it to make appropriate adjustments in the WOA-ELM meta-learner for better ET_0 estimation. This distinguishes the WOA-ELM-E from the bootstrap aggregating (mere transformation of data structure) and the BMA (mere combination of models by posterior probability).

4.4.2 Inter-Model Ensemble using NNE

Like the bootstrap aggregating and the BMA, the WOA-ELM-E were tested at different stations according to their affiliated clusters. The performance of the WOA-ELM-E, measured in terms of MAE, RMSE, MAPE, R^2 and MBE are presented as heat maps as shown in Figure 4.24 to Figure 4.28. The numerical values of the performance evaluation metrics can be consulted at Appendix D (Table D1 – Table D5).

It was well-proven in Section 4.3 that the model centric BMA-E was performing better than the individual base MLP, SVM and ANFIS. Hence, if the WOA-ELM-E managed to outperform the BMA-E, it would also mean that the WOA-ELM-E has better performance than the base models.

158
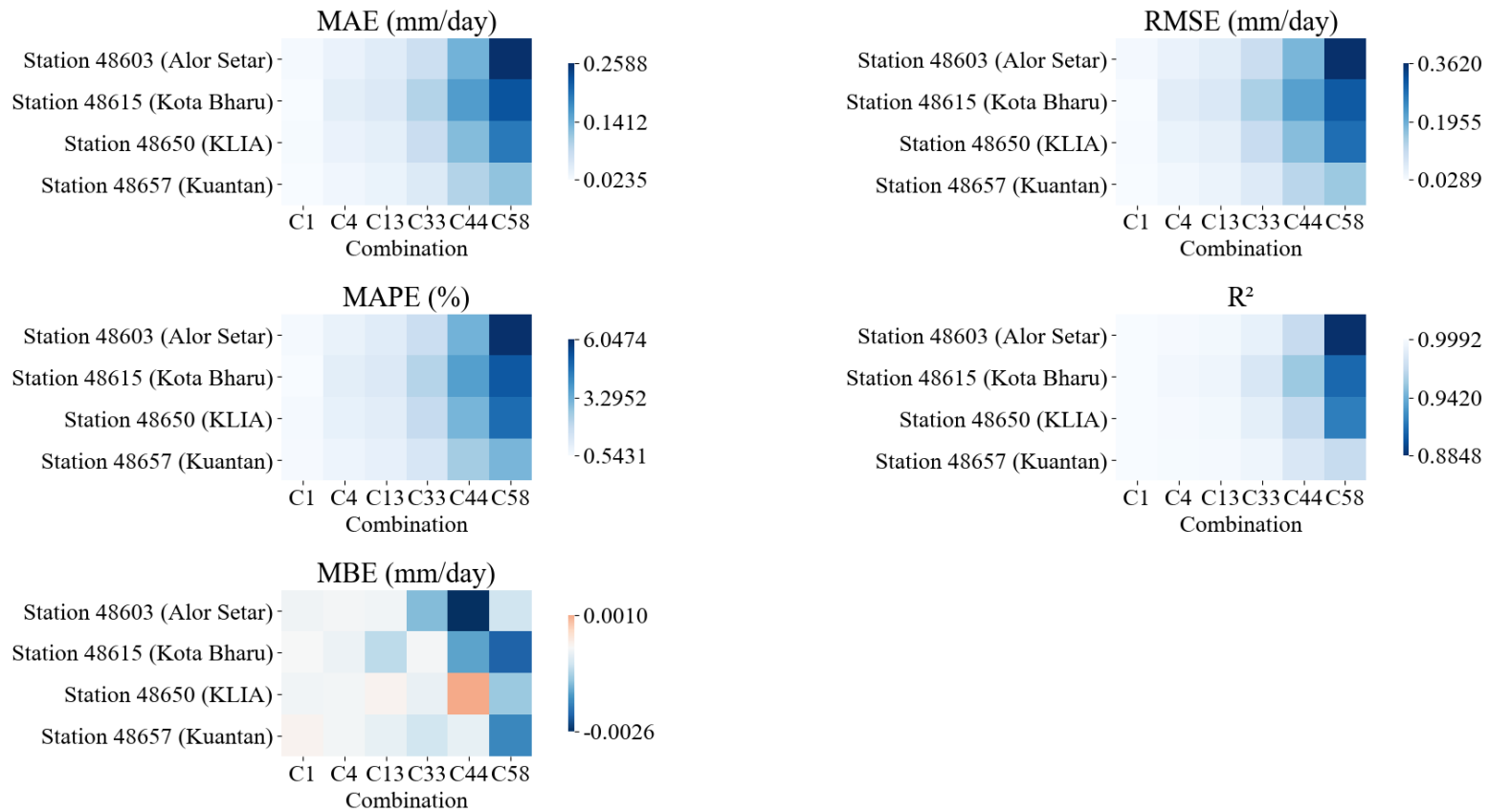
Figure 4.24: Performance of WOA-ELM-E in ET₀ Estimation using Different Input Combinations for Stations in Cluster 1
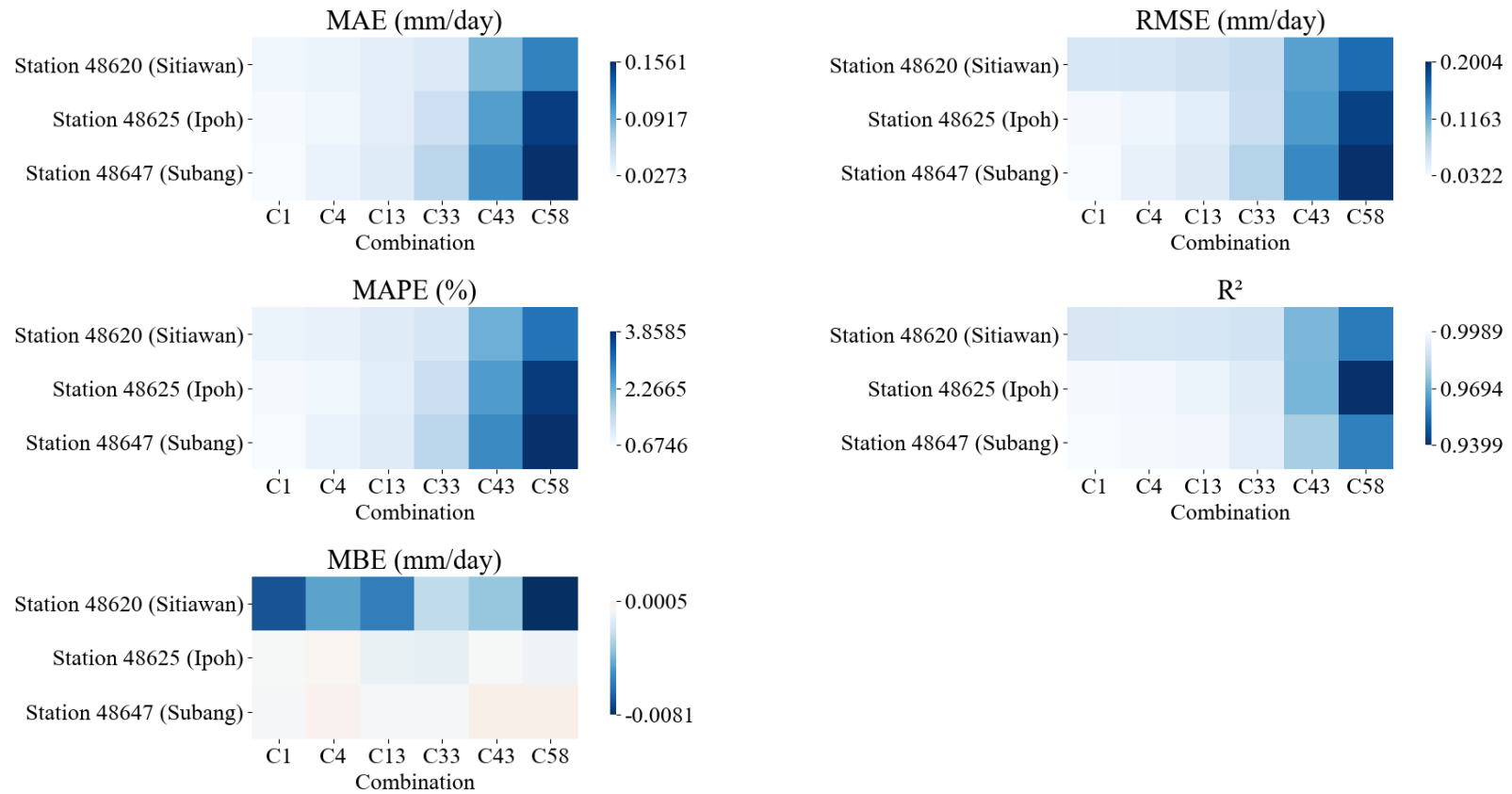
Figure 4.25: Performance of WOA-ELM-E in ET₀ Estimation using Different Input Combinations for Stations in Cluster 2
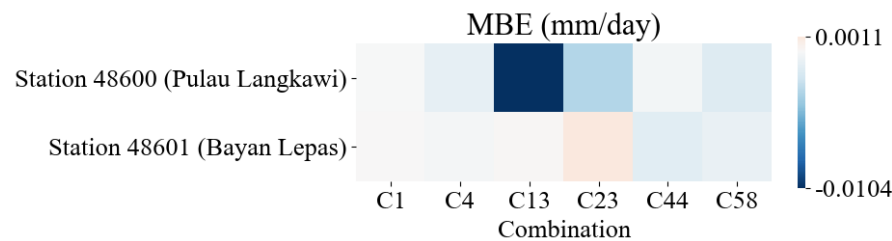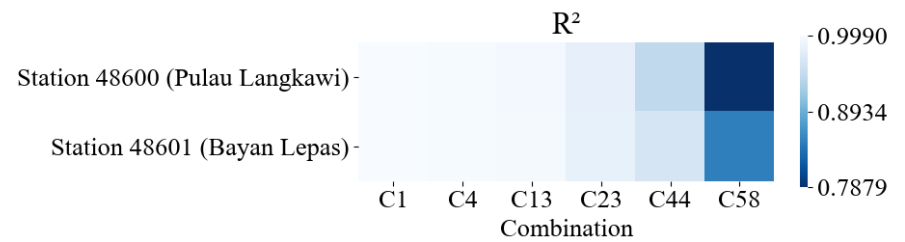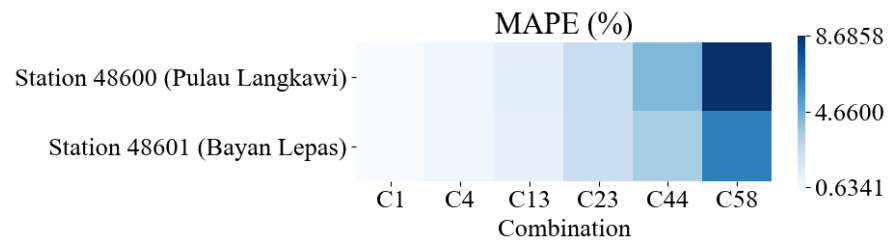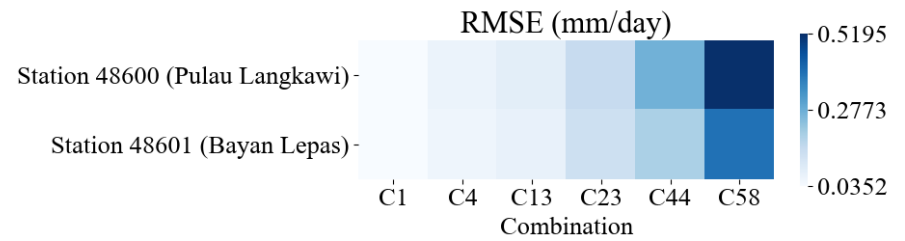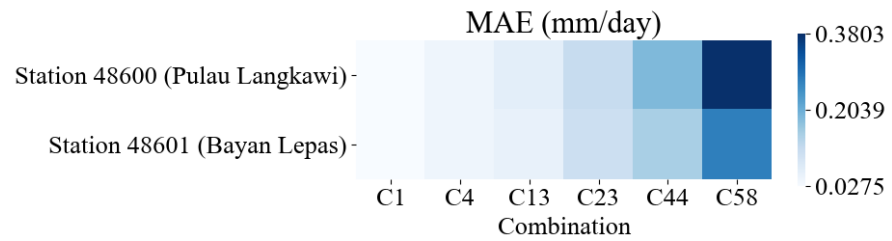
Figure 4.26: Performance of WOA-ELM-E in ET₀ Estimation using Different Input Combinations for Stations in Cluster 3
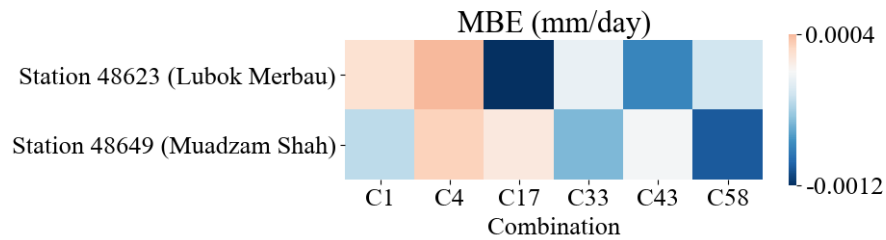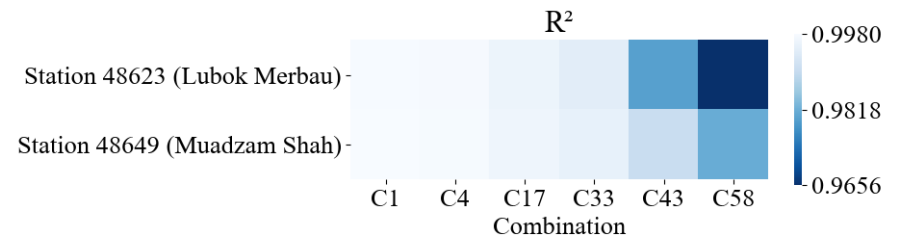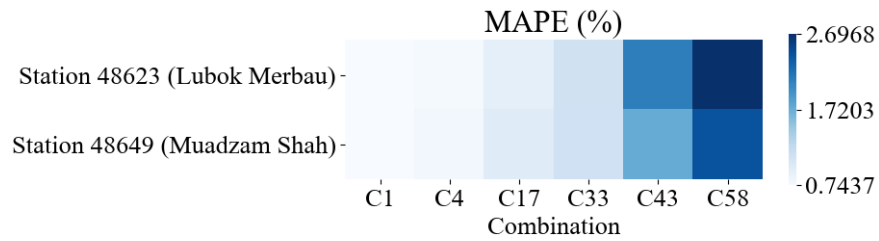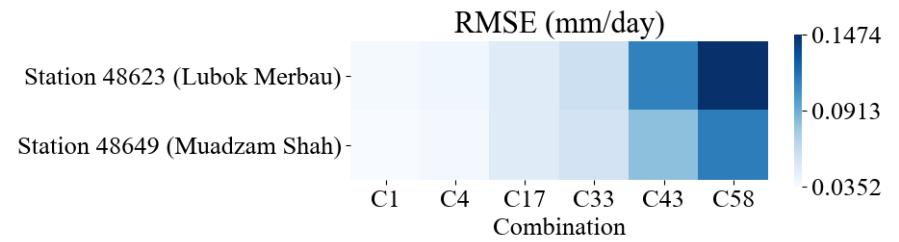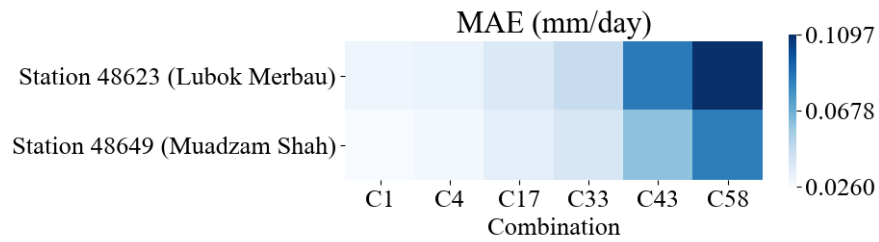
Figure 4.27: Performance of WOA-ELM-E in ET₀ Estimation using Different Input Combinations for Stations in Cluster 4
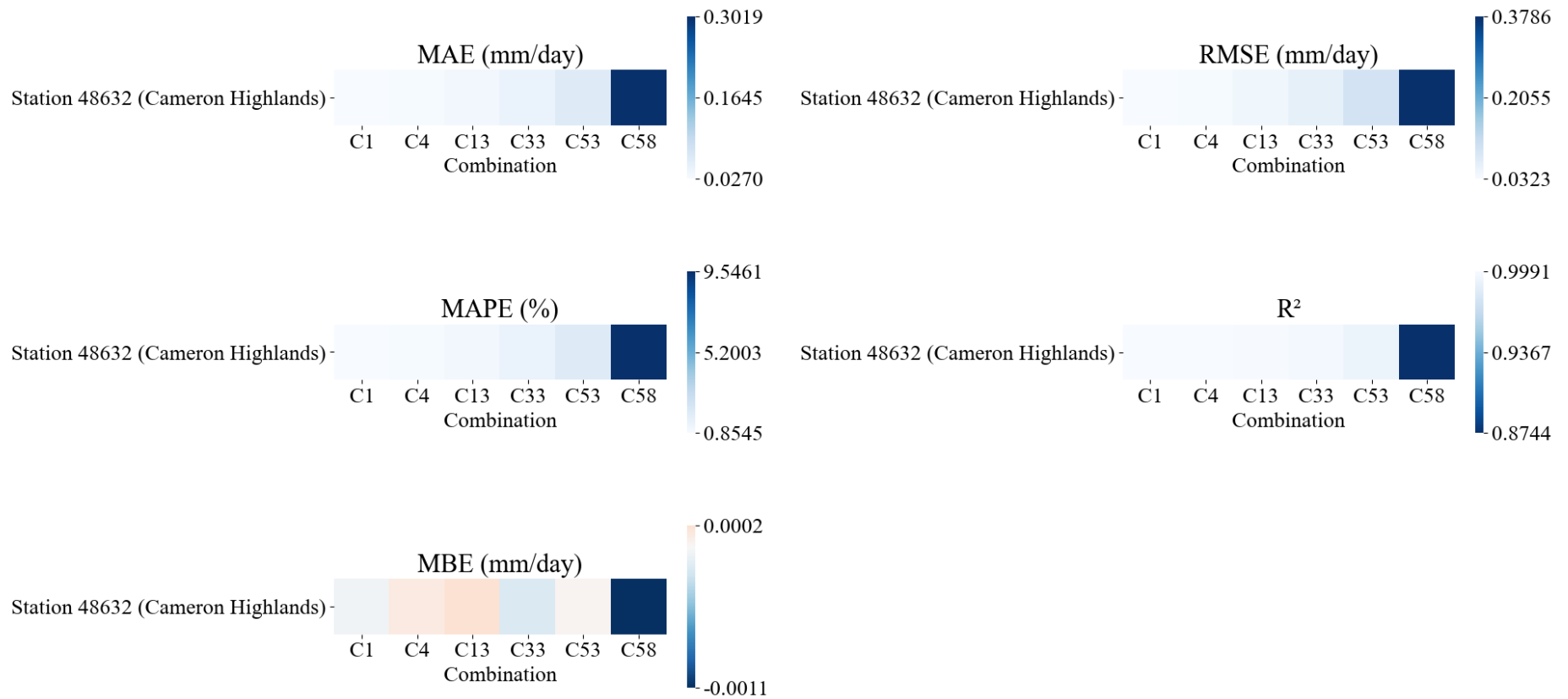
Figure 4.28: Performance of WOA-ELM-E in ET₀ Estimation using Different Input Combinations for Stations in Cluster 5

From Figure 4.24 to Figure 4.28, it was observable that the MBE of the WOA-ELM-E decreased to the 10^{-4} scale as compared to the 10^{-3} scale of the BMA-E. This means that the bias of WOA-ELM-E's estimations of ET_0 was even lower than those of BMA-E's. Comparing the WOA-ELM-E and BMA-E, the former had better generalisation ability that could cope with data from all clusters.

Meanwhile, the R^2 of the WOA-ELM-E was comparable to the BMA-E, which means that both ensembles fit well into the data of each station, whereby different input combinations have been considered. On the other hand, the error metrics, including the MAE, RMSE and MAPE showed improvement when the model centric BMA data fusion was replaced with the black-box NNE approach.

4.4.3 Summary

It should be clear that the improvement in the generalisability would mean that some degrees of accuracy must be sacrificed. Conversely, the low error reported could indicate the risk of overfitting (Belkin, et al., 2019). Thus, it is less practical if the developed machine learning models are evaluated and compared from separate aspects. For instance, model A could outperform model B in terms of accuracy, but at the same time lose out in terms of generalisability. Therefore, in this study, the author adopted an all-rounded comparison approach to assess the developed models in a fairer and more comprehensive environment.

A scoring system introduced by Despotovic, et al. (2015), known as the GPI, was calculated for every model. The positive-oriented GPI was inclusive of all the performance evaluation metrics used in this study and can provide an overall score for the models. In this study, the GPI scores of the models were calculated based on the normalised values of the performance evaluation metrics according to the different number of input meteorological variables. The comparison of GPI scores of the models developed in this research work (MLP, SVM, ANFIS, BMLP, BSVM, BANFIS, BMA-E and WOA-ELM-E) are shown in Figure 4.29 to Figure 4.33 (from Cluster 1 to Cluster 5). The actual values of the GPI can be referred to Appendix D (Table D6 to Table D10).
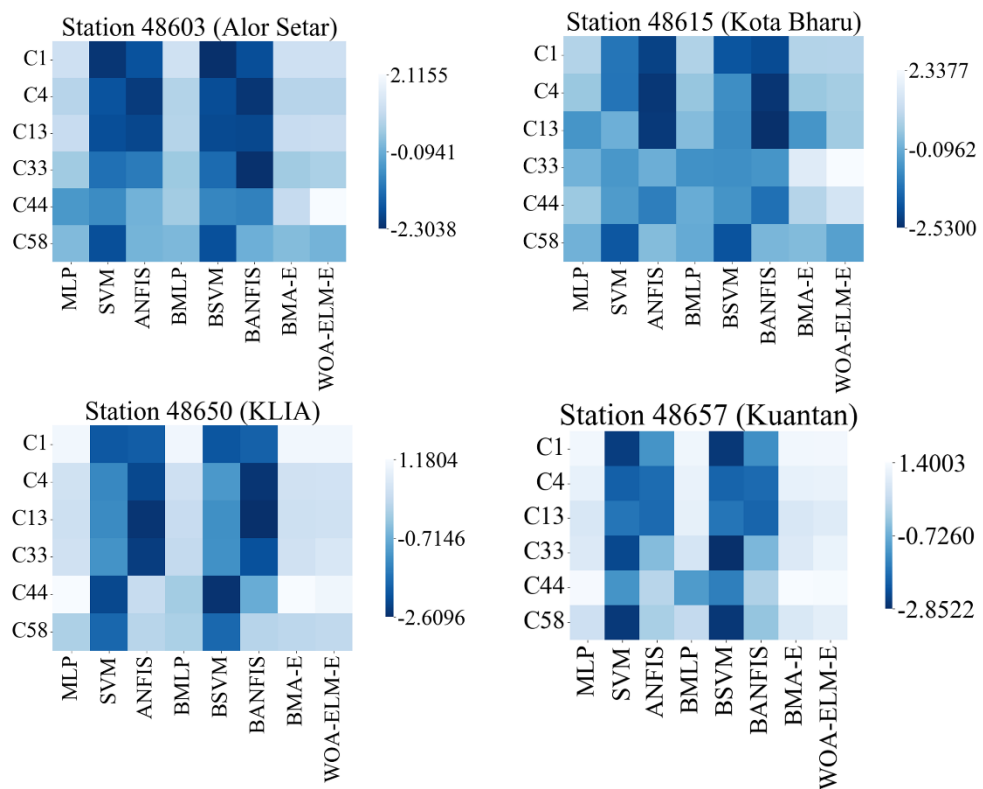
Figure 4.29: GPI Scores of Different Machine Learning Models at Stations in Cluster 1
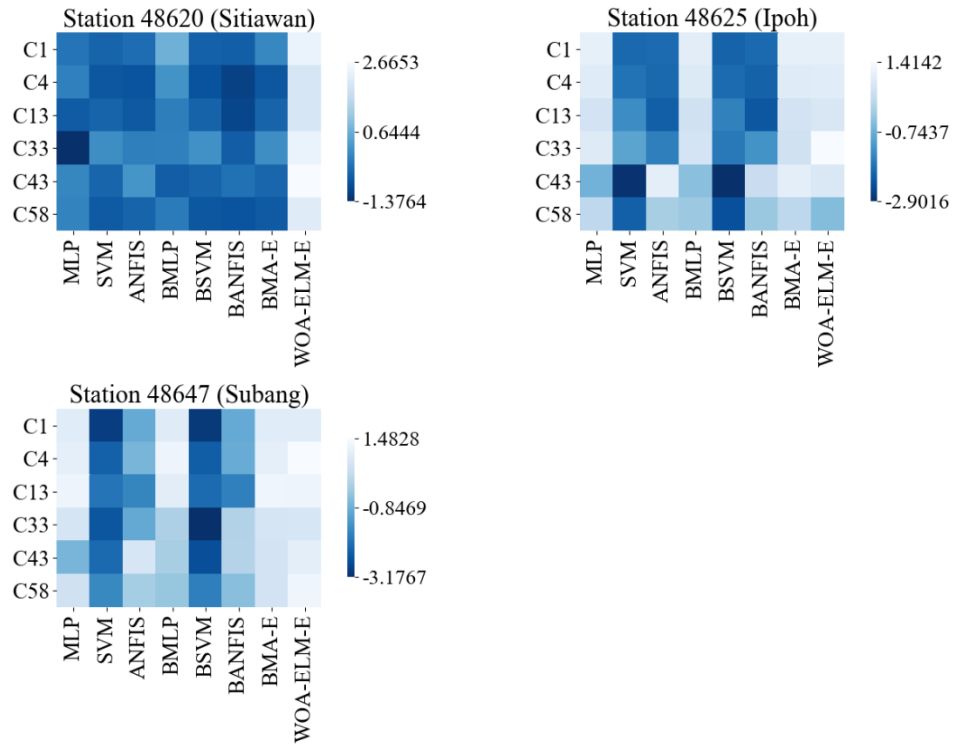
Figure 4.30: GPI Scores of Different Machine Learning Models at Stations in Cluster 2
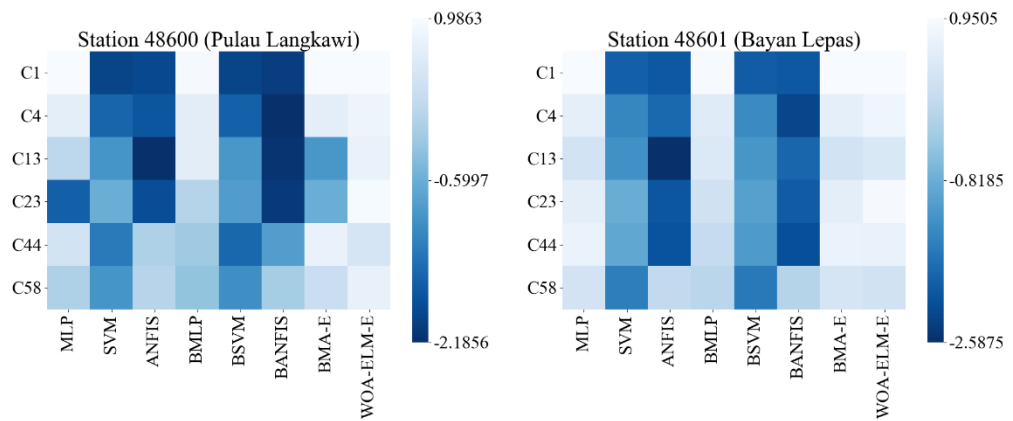
Figure 4.31: GPI Scores of Different Machine Learning Models at Stations in Cluster 3
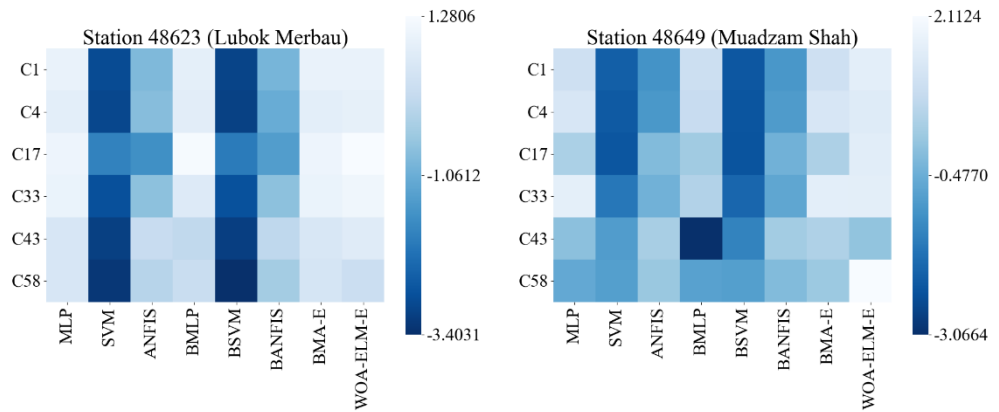
Figure 4.32: GPI Scores of Different Machine Learning Models at Stations in Cluster 4
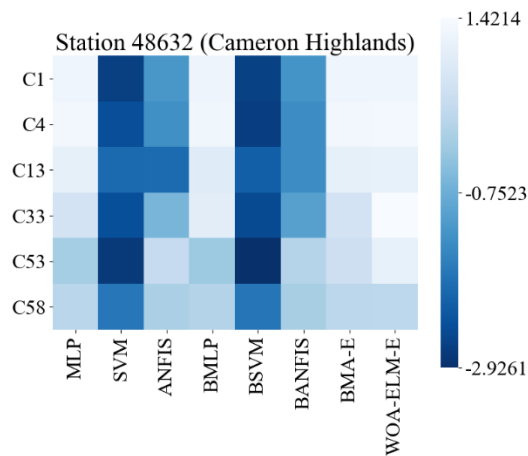
Figure 4.33: GPI Scores of Different Machine Learning Models at Stations in Cluster 5

In Figure 4.29 to Figure 4.33, WOA-ELM-E appeared to be the most stable model that maintained constantly in the upper bracket in terms of the GPI values. This can be used as strong evidence to justify the NNE developed in this study as a promising tool in improving the estimation of ET_0 in various regions in Peninsular Malaysia, even though with only one input meteorological

variable (the essential R_s). Another inter-model ensemble, the BMA-E also had satisfactory performance in most cases. However, the performance of the BMA-E was strongly correlated to the performance of the base models, in which their poor performance would result in a poor BMA-E (Chen, et al., 2015). This can be clearly seen at Station 48620 (Sitiawan) in Figure 4.30. The BMA-E had low GPI as the base model also performed poorly at that station.

The base MLP appeared to be sufficient for ET0 estimation at some of the stations, such as Station 48623 (Lubok Merbau). A similar observation is obtained for the BMLP. However, the performance of the MLP and BMLP was not as stable as the BMA-E or WOA-ELM-E. In other words, the outstanding performance of MLP and BMLP are only exceptions that happened in certain regions only.

Table 4.6 tabulates the best model for daily ET_0 estimation at all stations. The WOA-ELM-E was selected as the best ET_0 estimating model at most of the stations, regardless of the number of input meteorological variables. On the contrary, the BMA-E was only selected when the number of input meteorological variables was low. This is because the BMA-E could only manage to improve the model when different weights were assigned to different base models. Consequently, the favourable traits of different base models can be incorporated into the BMA-E to achieve estimations with higher quality. Although BMA-E was also selected at Station 48601 (Bayan Lepas) and Station 48603 (Alor Setar) for six and five input meteorological variables, respectively,

however, the BMA-E was merely the MLP as the Bayesian weight of a unit was given to the MLP in both cases. As for the other models such as the MLP, BMLP and ANFIS, they were selected in some cases where they perform better marginally than the WOA-ELM-E and the BMA-E. Therefore, a conclusion can be reached at this stage, in which the WOA-ELM-E was the best ET_0 estimating model in Peninsular Malaysia, in which it has wider spatial applicability among the eight models developed in this study.

The next section of this thesis will discuss the transferability of the locally developed WOA-ELM-E at external stations to investigate their spatial robustness. In other words, the local ET_0 values were estimated using exogenous models to eliminate the need for local data collection and model development/calibration.

Table 4.6: Best Models at Different Stations using Different Numbers of Input Meteorological Variables

Station	Number of Meteorological Variables					
	1	2	3	4	5	6
Station 48600 (Pulau Langkawi)	WOA-ELM-E	BMA-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	MLP
Station 48601 (Bayan Lepas)	BMA-E	MLP	WOA-ELM-E	BMLP	WOA-ELM-E	BMA-E
Station 48603 (Alor Setar)	BMA-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	BMA-E	WOA-ELM-E
Station 48615 (Kota Bharu)	ANFIS	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E
Station 48620 (Sitiawan)	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E
Station 48623 (Lubok Merbau)	MLP	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E
Station 48625 (Ipoh)	MLP	BMA-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E
Station 48632 (Cameron Highlands)	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E
Station 48647 (Subang)	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	BMA-E	WOA-ELM-E	MLP
Station 48649 (Muadzam Shah)	WOA-ELM-E	BMA-E	MLP	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E
Station 48650 (KLIA)	WOA-ELM-E	BMA-E	WOA-ELM-E	WOA-ELM-E	WOA-ELM-E	MLP
Station 48657 (Kuantan)	WOA-ELM-E	BMA-E	WOA-ELM-E	BMLP	BMLP	WOA-ELM-E

4.5   Models Transferability

Results and discussion from Section 4.1 to Section 4.4 focussed on
Scenario 1, in which the machine learning models (both base and hybrid) were
trained and tested locally. However, in order to curb the issue of qualitative data
hunger, the machine learning models should be transferable. In other words,
they must be spatially robust so that they can be applied elsewhere other than
the location where they were trained. In this section, results involving Scenario
2 and Scenario 3 will be discussed further to prove the spatial robustness of the
machine learning models developed in this research work.

4.5.1   Performance of Exogenous Models (Scenario 2)

This section is dedicated to discussing the transferability of locally
developed models to external stations. That is to say, the exogenous models
were used to estimate the local ET_0 in an effort to avoid local model
development in the future. To test the robustness of the exogenous models by
using a stricter standard, the input combinations fed to the exogenous models
(for ET_0 estimation) were the best input combinations obtained in their stations
of origin. For instance, the best model using two input meteorological variables
at Station 48620 (Sitiawan) was the WOA-ELM-E with C43 being used as the
input combination. When this model was tested in another station, say Station
48600 (Pulau Langkawi), C43 of Station 48600 (Pulau Langkawi) was used as
the input to the model although Station 48600 (Pulau Langkawi) was

categorised in Cluster 3 that preferred C44 as the input combination for two meteorological variables. This was to simulate a more realistic scenario in which meteorological data will be collected according to the requirement of the exogenous models in the future. In this study, this part is known as Scenario 2, whereby the ET_0 was not estimated by a locally trained model anymore (the exogenous model instead). The tested exogenous models were based on the results presented in Table 4.6.

Figure 4.34 illustrates the performance of the exogenous models at different stations in terms of the GPI using different numbers of meteorological variables in the form of heat maps. The actual GPI values can be consulted in Appendix E (Table E6, Table E12, Table E18, Table E24, Table E30 and Table E36). In Figure 4.34, lighter tones correspond to higher GPI (better performance) and vice versa. The heat maps can be read horizontally and vertically. By reading horizontally, one can be informed on the performance of the best models trained at different stations when they were tested at other stations. Conversely, a vertical view of the heat maps contains information on the dependency of the stations on the locally trained model. For instance, if there is a vertical dark band on any of the stations, it means that the exogenous model performed poorly (low GPI) at that station, and that station would need a locally trained model for its ET_0 estimation.

(a)

Six Input Meteorological Variables

(b)

Five Input Meteorological Variables

(c)

(d)

(e)

Two Input Meteorological Variables

(f)

One Input Meteorological Variable
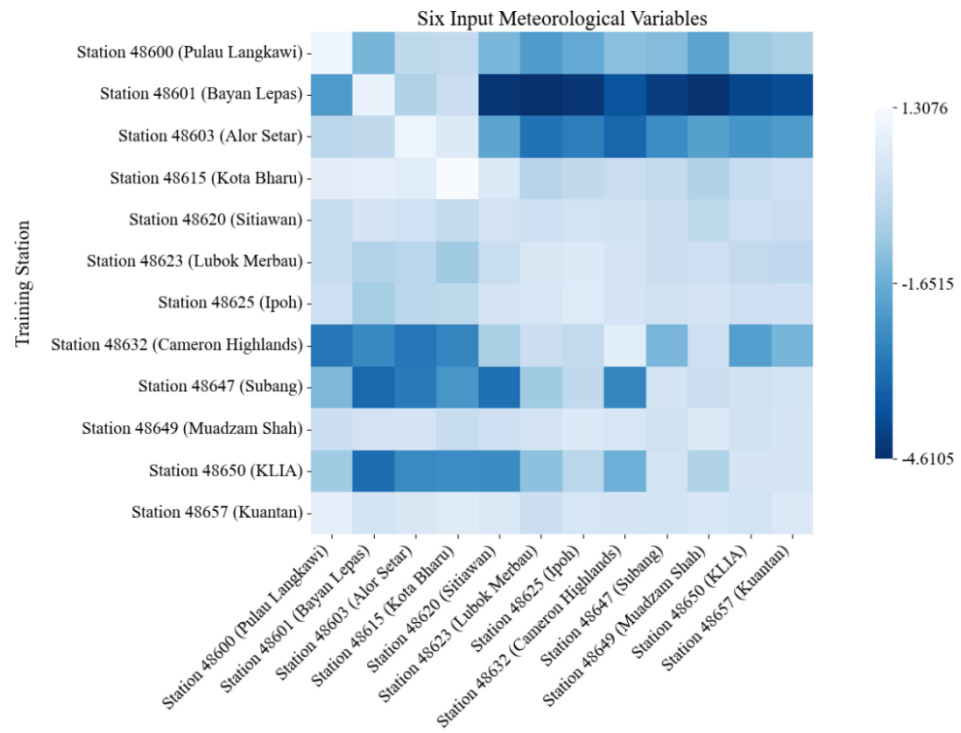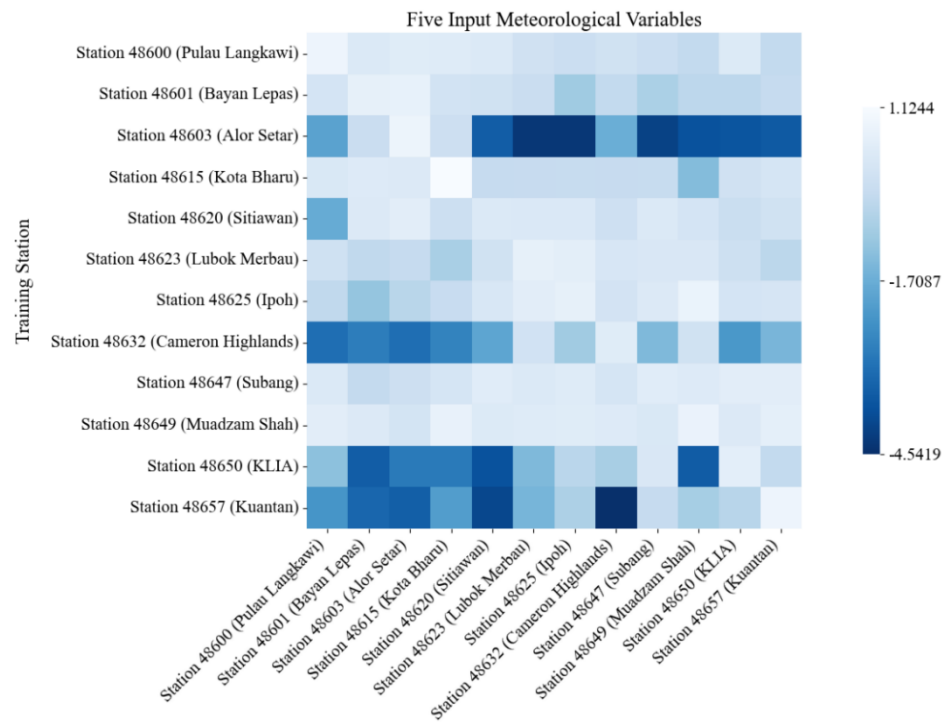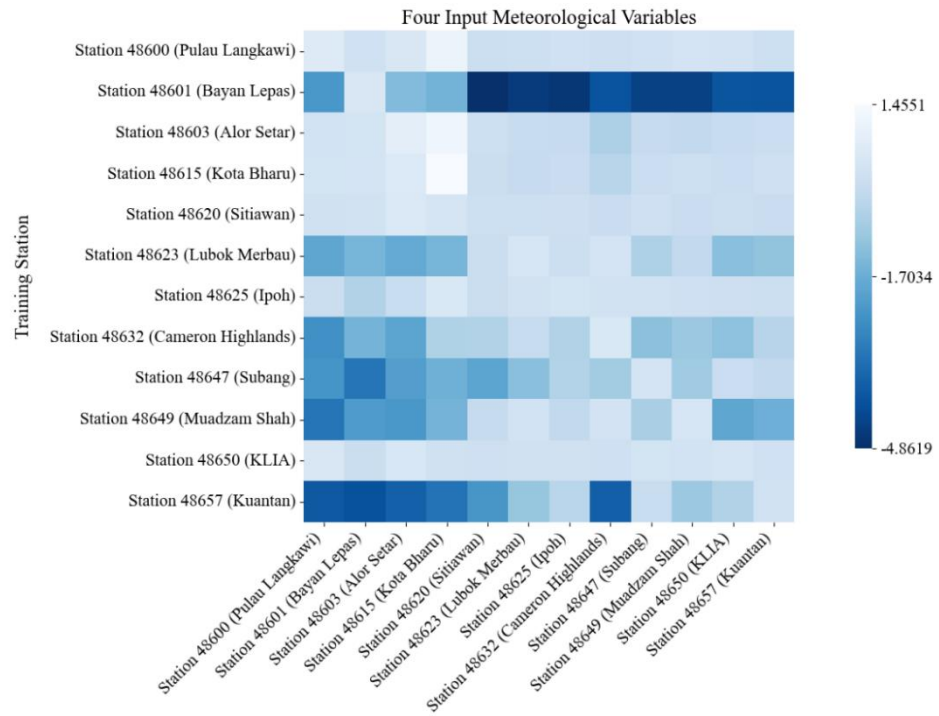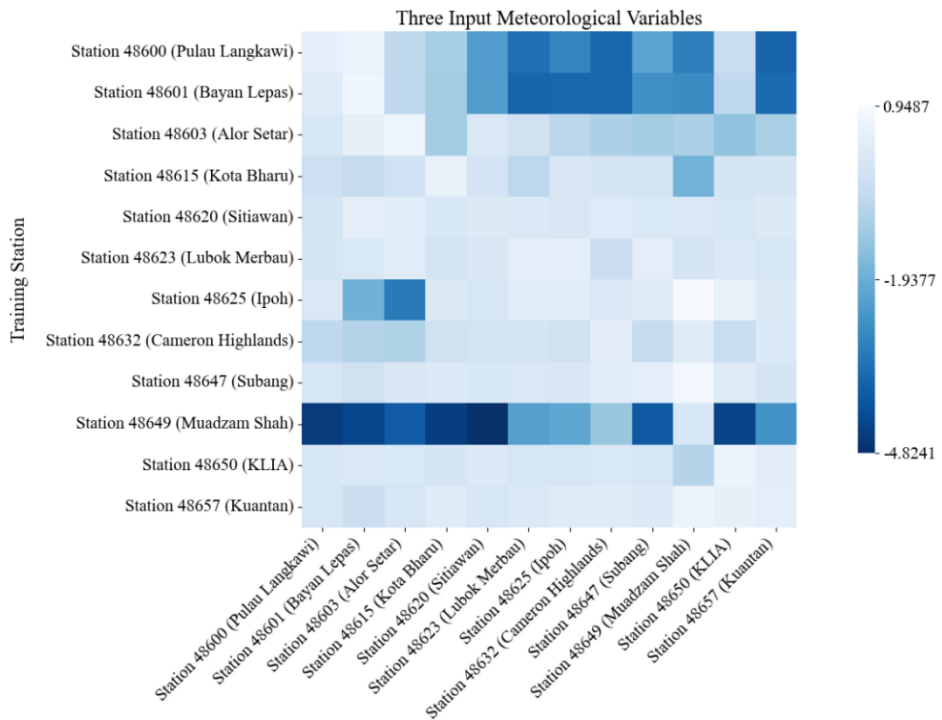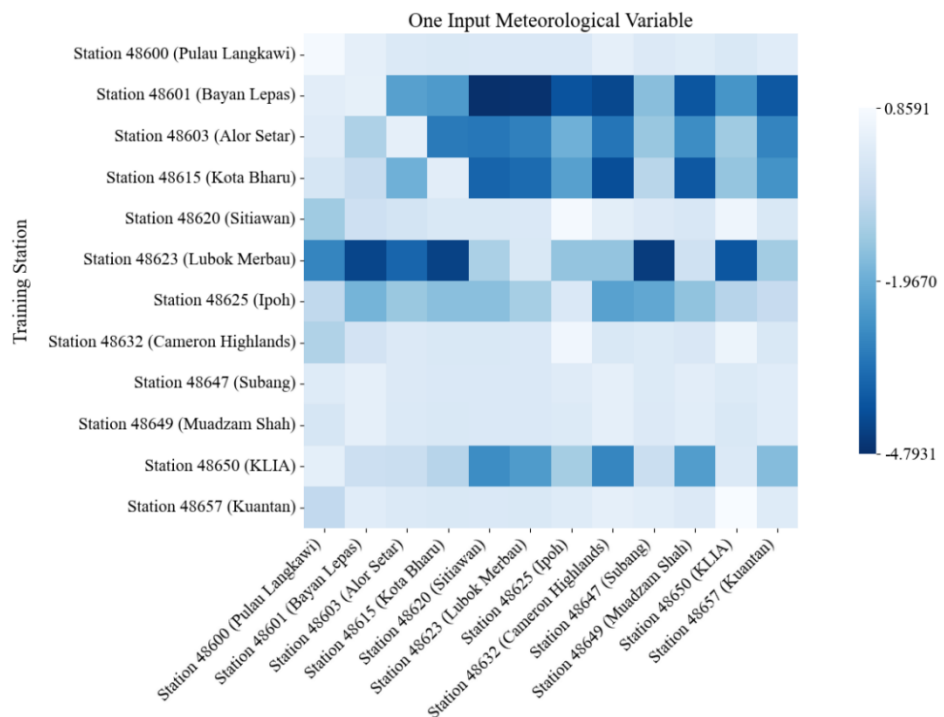
Figure 4.34: GPI Score for Cross-Station Testing using Best Models Trained at Different Stations with (a) Six, (b) Five, (c) Four, (d) Three, (e) Two and (f) One Input Meteorological Variable(s)

From Figure 4.34, no vertical dark bands were found. In other words, the hypothesis that certain stations would require local model development can be nullified. This finding can be taken positively, as it can be interpreted that the ET_0 of any region can be estimated using suitable exogenous models. Consequently, local data collection to train a machine learning model from scratch becomes unnecessary.

For the horizontal interpretation, several observations can be obtained from the heat maps in Figure 4.34. Some models were found to be incompetent for ET_0 estimation when they were deployed in other stations, in other words, not spatially robust. Some examples of non-spatially robust model for Scenario 2 are shown in Table 4.7.

Table 4.7: Examples of Non-Spatially Robust Model for Scenario 2

Training Station	Model Type	Input Combination
Station 48601 (Bayan Lepas)	BMA-E	C1
Station 48603 (Alor Setar)	BMA-E	C4
Station 48601 (Bayan Lepas)	BMLP	C13
Station 48657 (Kuantan)	BMLP	C13
Station 48649 (Muadzam Shah)	MLP	C33
Station 48649 (Muadzam Shah)	BMA-E	C43
Station 48601 (Bayan Lepas)	BMA-E	C58
Station 48603 (Alor Setar)	BMA-E	C58
Station 48615 (Kota Bharu)	ANFIS	C58
Station 48623 (Lubok Merbau)	MLP	C58

The non-spatially robust models reported in Table 4.7 consist of various model types, including the MLP, ANFIS, BMLP and BMA-E. Therefore, it can be deduced that if the best model selected at a particular station (and input combination) is not the WOA-ELM-E, the model would most likely perform poorly at external stations. However, there was one exception case that occurred

at Station 48650 (KLIA) when C44 was used as the input combination. In this case, the BMA-E was selected as the best local model and found to be one of the most stable models when it was deployed at other stations in Peninsular Malaysia using two meteorological variables.

Having said that, models other than the WOA-ELM-E could perform at external stations under special conditions. The most prominent observation is the clustering effect. For instance, models trained at Station 48601 (Bayan Lepas) performed poorly at most external stations. However, their performance at Station 48600 (Pulau Langkawi), which was also classified in Cluster 3 with Station 48601 (Bayan Lepas) was satisfactory. A similar observation was also obtained for the case of Station 48649 (Muadzam Shah) when C33 (MLP) and C43 (BMA-E) were the input combinations. These models delivered tolerable accuracy when deployed at another station in Cluster 4, which was Station 48623 (Lubok Merbau). Hence, clustering effect (location with similar geographical characteristics) increases the spatial coverage of the locally developed models. Although the models could be non-spatially robust across the whole Peninsular Malaysia, however, they can still be applied in regions with similar geographical characteristics, thereby reducing the need for model calibration/development in new areas.

Even though the WOA-ELM-E appeared to be the most spatially robust model for ET_0 estimation in Peninsular Malaysia, the exception case also happened at Station 48601 (Bayan Lepas). When C23 was fed to the WOA-

ELM-E trained at Station 48601 (Bayan Lepas), the model also suffered from low ET_0 estimation accuracy at most of the external stations. However, the poor performance of the WOA-ELM-E for this case may be attributed to the characteristics of the data collected at Station 48601 (Bayan Lepas), which render all the models trained at that station to be non-spatially robust. This argument can be supported by the performance of the WOA-ELM-E trained at Station 48600 (Pulau Langkawi) using C23, where equally poor performance was observed. The C23 (only selected by Cluster 2) did not have any temperature related meteorological variables, which render this input combination that was only suitable for stations in Cluster 2 and not spatially robust. In fact, except for MLP trained using C4 (five meteorological variables), all the models developed at Station 48601 (Bayan Lepas) were poor in external applications, suggesting that models trained with data from Cluster 2 had limited spatial applicability. In other words, besides the selection of model variants, the data collected locally is also another deterministic factor that affects the transferability of the machine learning models. Of course, the clustering effect discussed previously can help to mitigate this issue.

Across the 12 meteorological stations and six possible number of input meteorological variables, it was found that the models trained at Station 48620 (Sitiawan) were the most stable yet outstanding among the 72 models tested. It is noteworthy to mention that all the best models selected at Station 48620 (Sitiawan) were the WOA-ELM-E, which could possibly contribute to their outstanding performance due to the secondary training mechanism of the NNE

approach as discussed in the previous section. In Figure 4.34, the horizontal band of Station 48620 (Sitiawan) constantly showed tiles with light tones, regardless of the testing station as well as the number of meteorological variables. Explicitly, the local models trained at Station 48620 (Sitiawan) can be regarded as the most stable models for ET_0 estimation in Peninsular Malaysia, across regions with different geographical characteristics (different clusters).

4.5.2   The Global Model (Scenario 3)

This subsection is devoted to discussing the feasibility of developing a global model for ET_0 estimation in Peninsular Malaysia. Specifically, all the meteorological data collected from different stations were combined in a global data pool and reshuffled to remove the plausible geographical effects. These data were used to train WOA-ELM-E with different input combinations. The WOA-ELM-E was selected because it was found to be the most spatially robust model according to the findings obtained through the investigation on Scenario 2. As for the input combination, the optimum input combinations of Cluster 1 were applied since Cluster 1 was the largest cluster in Peninsular Malaysia.

Figure 5.35 shows the GPI scores of the global WOA-ELM-E at different stations with the various input combinations (refer Appendix E (Table E37) for a detailed breakdown of performance evaluation metrics). The figure suggests that the performance of the global WOA-ELM-E was good and stable generally. This can be deduced from the GPI scores of the global WOA-ELM-

E, which were higher than zero (the median value) most of the time. When comparing Figure 5.35 with Figure 5.34, it can be seen that the GPI scores of poor models could be as low as -4, which was not found in the case of the global WOA-ELM-E (minimum GPI of -0.3269). The data management scheme proposed under Scenario 3, coupled with NNE data fusion technique (WOA-ELM-E) was proven to be a successful strategy in resolving the need for *in-situ* meteorological data collection which requires burdensome fiscal input.
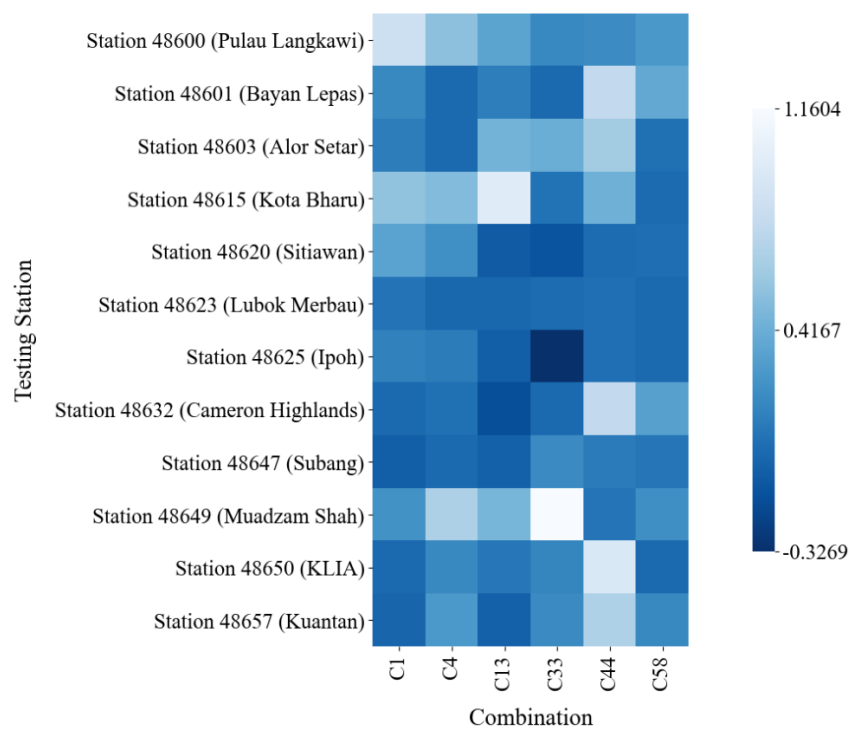
Figure 4.35: GPI Scores of WOA-ELM-E Model Trained using Global Data

It is interesting to compare the performance of global WOA-ELM-E with the local WOA-ELM-E trained at Station 48620 (Sitiawan). The former was trained using a dataset that contained data of great variabilities, whereas the latter performed satisfactorily well without the need for a large amount of data.

Table 4.8 summarises the comparison of GPI scores for WOA-ELM-E models trained using meteorological data from Station 48620 (Sitiawan) and the global data pool. The table directly contrasts the local and global WOA-ELM-E, in terms of their GPI scores, at different meteorological stations in Peninsular Malaysia.

The GPI scores tabulated in Table 4.8 generally exceed zero, which means that the performance of the WOA-ELM-E, regardless of the training data, was better than the median model. In other words, it can be said that the local and global WOA-ELM-E positioned themselves in the upper bracket among all the models developed in this study. To conduct a fair comparison between the local and global WOA-ELM-E, the GPI scores of the ET_0 estimating models were compared based on the number of input meteorological variables. The GPI scores of the better model under various circumstances are highlighted in Table 4.8.

It was found that the local WOA-ELM-E trained using the data from Station 48620 (Sitiawan) outperformed its global counterpart at Station 48625 (Ipoh) and Station 48647 (Subang). On the contrary, the global WOA-ELM-E trained using the global dataset had the upper hand at Station 48603 (Alor Setar), Station 48615 (Kota Bharu), Station 48650 (KLIA) and Station 48657 (Kuantan). Coincidently, the local WOA-ELM-E had better performance at stations in Cluster 2 (its origin cluster) whereas the global WOA-ELM-E performed better at stations in Cluster 1 (from which it inherited the training

input combinations). Therefore, the selection of input combinations and geographical characteristics clustering once again exhibited their impacts on the exogenous models of ET_0 estimation.

As for the stations in Cluster 3, Cluster 4 and Cluster 5, the local and global WOA-ELM-E had comparable results. Both models excelled at one station each for Cluster 3 and Cluster 4. At the sole station of Cluster 5, Station 48632 (Cameron Highlands), the two models performed equally well where they were reported to be suitable for three input combinations each. This finding further highlights that the local and global WOA-ELM-E were comparable to each other and could be a potential choice of ET_0 estimating model at another new region in Peninsular Malaysia.

Table 4.8: Comparison of GPI Scores for WOA-ELM-E Models Trained using Meteorological Data from Station 48620 (Sitiawan) and Global Data Pool

| Testing Station | Data Sources and Number of Input Meteorological Variables | | | | | | | | | | | |
| | Scenario 2: Station 48620 (Sitiawan) (Local WOA-ELM-E) | | | | | | Scenario 3: Global Data Pool (Global WOA-ELM-E) | | | | | |
	1	2	3	4	5	6	1	2	3	4	5	6
Station 48600 (Pulau Langkawi)	-1.218	0.096	-0.091	0.200	-1.792	-0.131	0.267	0.198	0.173	0.349	0.536	0.842
Station 48601 (Bayan Lepas)	-0.313	-0.116	0.394	0.240	0.343	0.259	0.379	0.774	0.001	0.120	0.001	0.182
Station 48603 (Alor Setar)	-0.178	-0.757	0.298	0.549	0.521	0.098	0.044	0.621	0.415	0.447	0.000	0.112
Station 48615 (Kota Bharu)	0.013	-0.028	0.028	0.389	-0.116	-0.199	0.006	0.430	0.054	0.980	0.503	0.551
Station 48620 (Sitiawan)	0.002	0.606	0.173	0.117	0.306	0.208	0.030	0.019	-0.127	-0.079	0.220	0.339
Station 48623 (Lubok Merbau)	0.020	0.287	0.117	0.107	0.297	0.019	0.001	0.039	0.018	-0.013	-0.009	0.061
Station 48625 (Ipoh)	0.768	0.639	0.052	0.077	0.295	0.188	0.001	0.035	-0.327	-0.055	0.104	0.137
Station 48632 (Cameron Highlands)	0.273	-0.305	0.189	-0.014	-0.061	0.161	0.325	0.772	0.001	-0.149	0.044	0.001
Station 48647 (Subang)	0.079	0.987	0.088	0.131	0.314	0.000	0.070	0.098	0.195	-0.048	0.000	-0.056
Station 48649 (Muadzam Shah)	-0.031	0.311	0.098	-0.003	0.111	-0.313	0.219	0.063	1.160	0.458	0.671	0.234
Station 48650 (KLIA)	0.590	0.012	-0.001	0.065	-0.160	0.032	0.003	0.936	0.157	0.081	0.181	0.001
Station 48657 (Kuantan)	-0.001	-0.854	0.182	-0.021	-0.001	-0.069	0.182	0.682	0.191	-0.042	0.275	-0.020

Nevertheless, the pros and cons of the local and global WOA-ELM-E have to be commented prior to the selection of models. Table 4.9 shows the properties of the two models, in terms of various aspects.

Table 4.9: Comparison of the Properties of the Local and Global WOA-ELM-E

	Local WOA-ELM-E	Global WOA-ELM-E
Weight	Lightweight	Heavyweight
Data Collection	Tougher to collect	Easier to collect
Future Improvement	Can be updated using new data from Station 48620 (Sitiawan)	Can be updated using data from any other regions

The local and global WOA-ELM-E are compared in terms of their weights during the deployment, the ease of data collection as well as the potential for future improvement on the model. Firstly, the local WOA-ELM-E is more advantageous from the deployment weight perspective. As the model was trained with a smaller amount of data, the resultant hyper-parameters of the local WOA-ELM-E became lower as compared to the global WOA-ELM-E. This leads to the smaller size and lighter weight of the local WOA-ELM-E which makes it easier to be deployed elsewhere, especially when cloud computing is not accessible.

In terms of the ease of data collection, both models agreed with most of the optimum input combinations apart from the input combination with only two meteorological variables. The local WOA-ELM-E originated from a Cluster 2 station favoured C43 (u and R_s) whereas the global WOA-ELM-E was

using C44 (*RH* and *R_s*). Comparing the *u* and *RH*, collecting the data for the former variable is more challenging due to the high volatility (magnitude and direction), complex stochastic changes in the environment, noises and equipment failure (Wang, et al., 2021). On the other hand, the measurement of *RH* is relatively easier as the variable is rather stable with many accurate measuring devices available in the market. The ease of data collection can determine the model selection of the end-users and render the global WOA-ELM-E more useful.

Lastly, the global WOA-ELM-E has a greater potential to be improved to become a better model. This is because new data collected at any new (or existing) stations can be included into the global data pool to be used as its training data. Contrarily, the local WOA-ELM-E, by its design can only accept the new data collected from Station 48620 (Sitiawan) for further improvement. Consequently, the variability of the data received by the global WOA-ELM-E is larger than the local WOA-ELM-E. The variability of the training data has a positive correlation with the performance of the machine learning models (Can, et al., 2021).

By considering the differences in the properties of the local and global WOA-ELM-E for ET_0 estimation, the policy makers can make an appropriate selection on the most suitable model to be deployed in a new region by addressing the needs in computational power as well as data collection infrastructure. The two data management schemes investigated in this research

work, Scenario 2 and Scenario 3, were found to be promising methods with different integrations and applications. The machine learning models (MLP, SVM and ANFIS) hybridised using the NNE data fusion method is a highly transferable model when trained using local data as well as the global dataset.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1   Conclusion

This thesis presents the research work that focusses on devising a reliable resolution to the problem of estimation of ET_0 in Peninsular Malaysia. At present, physical measurement and empirical calculation of ET_0 are facing daunting challenges from various aspects, including financial costs, technical feasibility and skilled labour forces involved. The interception of the black-box machine learning models showed positive outcomes, but they suffered from qualitative as well as quantitative data hunger.

The work reported in this thesis aimed to develop a robust machine learning model for ET_0 estimation in Peninsular Malaysia with minimum data requirements. This was done through a minimalistic approach which employed three base machine learning models, namely the MLP, the SVM and the ANFIS and hybridising them through the implementation of the various data fusion techniques. Three data fusion techniques, including the data centric bootstrap aggregating, the model centric BMA and the black-box-based NNE were investigated.

Among the three elementary machine learning models, the MLP was found to be the best model in terms of estimation accuracy, while the SVM and the ANFIS were less sensitive towards the decrement of input meteorological variables. Regardless of the machine learning models used, each meteorological station showed preference input combinations with different numbers of input meteorological variables. By assessing the optimum input combinations at the meteorological stations, it was found that the meteorological stations can be grouped into five clusters of similar geographical characteristics and the members of each cluster shared the same optimum input combinations.

Of all the meteorological variables, the R_s was the most essential feature that had to be included into the input combinations. The omission of the R_s would render the model to have increased estimation error and reduced goodness-of-fit. This was because the inbound radiation was the main driving factor and energy source for the water molecules to escape into the atmosphere, especially in proximity to the Equator. The other meteorological variables would act as complementary features that touched up the accuracy of the machine learning models (Objective 1).

Bootstrap aggregating was integrated into the individual MLP, SVM and ANFIS models to create the BMLP, BSVM and BANFIS models, respectively. Although the bootstrap aggregating improved some of the base models at certain stations, overall, bagged models performed poorer than the base models. This was because the number of samples involved in this study had already

overwhelmed the dimensionality of the ET_0 estimation problem, subsequently rendering the bootstrap aggregating to be ineffective. In fact, the integration of bootstrap aggregating could be risky when "bad" data points were given a chance to have a higher influence on the models.

On the other hand, the BMA could assign Bayesian weights to the base machine learning models by evaluating and normalising their posterior probability. For the first time, the Bayesian weights of the MLP, SVM and ANFIS for an ET_0 estimating ensemble were documented and reported in detail. When the number of meteorological variables was high, the BMA algorithm tended to assign the Bayesian weight of one to the MLP. However, when input meteorological variables were reduced, the influence of the SVM and ANFIS on the ensemble increased. The BMA-E model undoubtedly improved the ET_0 estimation as compared to the base models, but the algorithm lacked flexibility, and could possibly sacrifice the constituent models which led to the loss of information. Besides, the frequent assignment of the unit weight to the MLP caused the resultant ensemble to be technically a base model that did not have favourable traits of different models.

To integrate the NNE data fusion technique, a novel meta-learner based on the ELM optimised with the WOA was developed in this work (WOA-ELM). The outputs of the MLP, SVM and ANFIS were fed into the WOA-ELM so that a "secondary training" could be carried out. By comparing the GPI, the WOA-ELM-E was found to be the model with the highest consistency, in terms of

189

making it into the top brackets for model performance. The outstanding performance of the WOA-ELM-E was due to the flexibility of the ANN-based meta-learner (unlike the rigid BMA algorithm) and the development strategy that allowed it to have a "second look" at the target values during the training process. The best model (out of eight) was selected at each meteorological station, with the WOA-ELM-E being the dominating model at most of the stations (Objective 2).

The spatial robustness of the best models was tested under Scenario 2 to ensure that these models could be applied externally so that the data collection process for local model calibration could be eliminated. The best models acted as an exogenous model at all the meteorological stations. It was found that none of the 12 meteorological stations studied in this research work, showed a high dependency on the local data because there was at least one exogenous model that can estimate the local ET_0 at satisfactory accuracy. The best models at Station 48620 (Sitiawan) performed consistently at all the stations. All models selected at Station 48620 (Sitiawan) were of the WOA-ELM-E variant. The local model trained at Station 48620 (Sitiawan) had the best spatial robustness in Peninsular Malaysia (Objective 3).

A global WOA-ELM-E was developed by using the data from all the meteorological stations as training data. The global dataset was shuffled to remove any possible geographical influence. Like the local WOA-ELM-E from Station 48620 (Sitiawan), the global WOA-ELM-E also showed consistent

performance at all the stations. To be exact, the local and global WOA-ELM-E had comparable performance, depending on the testing meteorological stations. This meant that both models could be deployed across the whole Peninsular Malaysia. The pros and cons of the local and global WOA-ELM-E were contrasted, in which the former was easier to deploy due to its light weight, while the training data of the latter were easier to be collected and more diversified (Objective 4).

The spatial robustness of the local and global WOA-ELM-E developed in this study means that in the future, there is no need for meteorological data collection in a new region for the local model calibration or development. In other words, the qualitative as well as quantitative hunger aspect had been resolved with the introduction of the local and global WOA-ELM-E. Furthermore, this thesis reported that the R_s was the most essential meteorological variable for ET_0 estimation in Peninsular Malaysia. Optimum input combinations with low number of meteorological variables were also reported. This means that the authorities only need to collect the essential data, instead of collecting all meteorological variables which could be costly and troublesome. In view of this, the issue of qualitative hunger has also been addressed.

The outcome and output of this research work had significant contribution to the scientific community in understanding the relationship of the meteorological variable, base machine learning models and data fusion

techniques. At the national level, the development of the robust ET_0 estimating model could provide valuable information to the authorities and decision makers in drawing appropriate strategies of water resources allocation and management in Peninsular Malaysia.

5.2 Future Works and Recommendations

Even though this investigation was designed and executed carefully, however, due to many limitations, some flaws were still being spotted during the research phase. Many future works can be extended based on the findings of this study. Some important recommendations are included in this section as the afterthought of the entire research phase.

5.2.1 Deployment of the Models at Different Regions

The work reported in the thesis was constrained by time and financial resources. The authors did not have the opportunity to deploy the developed models, particularly the local and global WOA-ELM-E in other regions of the world. Future works can be focussed on the application of the local and global WOA-ELM-E in Southeast Asia, regions with a tropical climate and then other parts of the world.

5.2.2 A Continual Improvement Framework

The models developed in this study were trained using static historical data. However, the ET_0 is highly affected by the global climate change, which is a dynamic process. New data can be collected at the meteorological stations and fed into the developed models to fine-tune the models so that their relevancy can be updated. The continual improvement framework of the models has to be designed (update threshold, degree of fine-tuning) in future works.

REFERENCES

Abdullah, S. S., Malek, M. A., Abdullah, N. S., Kisi, O. and Yap, K. S., 2015. Extreme Learning Machines: A new approach for prediction of reference evapotranspiration. *Journal of Hydrology,* 527**,** pp. 184-195.

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A. and Arshad, H., 2018. State-of-the-art in artificial neural network applications: A survey. *Heliyon,* 4**,** pp. e00938.

Adamala, S., 2018. Temperature based generalized wavelet-neural network models to estimate evapotranspiration in India. *Information Processing in Agriculture,* 5**,** pp. 149-155.

Alizadeh, M. R. and Nikoo, M. R., 2018. A fusion-based methodology for meteorological drought estimation using remote sensing data. *Remote Sensing of Environment,* 211**,** pp. 229-247.

Alizamir, M., Kisi, O., Muhammad Adnan, R. and Kuriqi, A., 2020. Modelling reference evapotranspiration by combining neuro-fuzzy and evolutionary strategies. *Acta Geophysica,* 68**,** pp. 1113-1126.

Allan, R. G., Pereira, L., Raes, D. and Smith, M., 1998. *Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and Drainage Paper 56.*

Anapalli, S. S., Ahuja, L. R., Gowda, P. H., Ma, L., Marek, G., Evett, S. R. and Howell, T. A., 2016. Simulation of crop evapotranspiration and crop coefficients with data in weighing lysimeters. *Agricultural Water Management,* 177**,** pp. 274-283.

Antonopoulos, V. Z. and Antonopoulos, A. V., 2017. Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. *Computers and Electronics in Agriculture,* 132**,** pp. 86-96.

Belayneh, A., Adamowski, J., Khalil, B. and Quilty, J., 2016. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. *Atmospheric Research,* 172-173**,** pp. 37-47.

Belkin, M., Hsu, D., Ma, S. and Mandal, S., 2019. Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc Natl Acad Sci U S A,* 116**,** pp. 15849-15854.

Berti, A., Tardivo, G., Chiaudani, A., Rech, F. and Borin, M., 2014. Assessing reference evapotranspiration by the Hargreaves method in north-eastern Italy. *Agricultural Water Management,* 140**,** pp. 20-25.

Breiman, L., 1996. Bagging predictors. *Machine Learning,* 24**,** pp. 123-140.

Cammalleri, C., Anderson, M. C., Gao, F., Hain, C. R. and Kustas, W. P., 2013. A data fusion approach for mapping daily evapotranspiration at field scale. *Water Resources Research,* 49**,** pp. 4672-4686.

Cammalleri, C., Anderson, M. C., Gao, F., Hain, C. R. and Kustas, W. P., 2014. Mapping daily evapotranspiration at field scales over rainfed and

irrigated agricultural areas using remote sensing data fusion. *Agricultural and Forest Meteorology,* 186**,** pp. 1-11.

Can, G., Mantegazza, D., Abbate, G., Chappuis, S. and Giusti, A., 2021. Semantic segmentation on Swiss3DCities: A benchmark study on aerial photogrammetric 3D pointcloud dataset. *Pattern Recognition Letters,* 150**,** pp. 108-114.

Carter, C. and Liang, S., 2019. Evaluation of ten machine learning methods for estimating terrestrial evapotranspiration from remote sensing. *International Journal of Applied Earth Observation and Geoinformation,* 78**,** pp. 86-92.

Cascone, S., Coma, J., Gagliano, A. and Pérez, G., 2019. The evapotranspiration process in green roofs: A review. *Building and Environment,* 147**,** pp. 337-355.

Chacon-Hurtado, J. C., Alfonso, L. and Solomatine, D. P., 2017. Rainfall and streamflow sensor network design: a review of applications, classification, and a proposed framework. *Hydrology and Earth System Sciences,* 21**,** pp. 3071-3091.

Chen, T. and Guestrin, C., XGBoost: a scalable tree boosting system. In*: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016. San Francisco, California, USA. 785-794.

Chen, Y., Yuan, W., Xia, J., Fisher, J. B., Dong, W., Zhang, X., Liang, S., Ye, A., Cai, W. and Feng, J., 2015. Using Bayesian model averaging to estimate terrestrial evapotranspiration in China. *Journal of Hydrology,* 528**,** pp. 537-549.

Chia, M. Y., Huang, Y. F. and Koo, C. H., 2021a. Improving reference evapotranspiration estimation using novel inter-model ensemble approaches. *Computers and Electronics in Agriculture,* 187**,** pp. 106227.

Chia, M. Y., Huang, Y. F. and Koo, C. H., 2021b. Swarm-based optimization as stochastic training strategy for estimation of reference evapotranspiration using extreme learning machine. *Agricultural Water Management,* 243**,** pp. 106447.

Chia, M. Y., Huang, Y. F., Koo, C. H. and Fung, K. F., 2020. Recent advances in evapotranspiration estimation using artificial intelligence approaches with a focus on hybridization techniques—a review. *Agronomy,* 10**,** pp. 101.

Chia, S. L., Chia, M. Y., Koo, C. H. and Huang, Y. F., 2022. Integration of advanced optimization algorithms into least-square support vector machine (LSSVM) for water quality index prediction. *Water Supply,* 22**,** pp. 1951-1963.

Cobaner, M., 2011. Evapotranspiration estimation by two different neuro-fuzzy inference systems. *Journal of Hydrology,* 398**,** pp. 292-302.

Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine Learning,* 20**,** pp. 273-297.

Dantas, T. M. and Cyrino Oliveira, F. L., 2018. Improving time series forecasting: An approach combining bootstrap aggregation, clusters and exponential smoothing. *International Journal of Forecasting,* 34**,** pp. 748-761.

Despotovic, M., Nedic, V., Despotovic, D. and Cvetanovic, S., 2015. Review and statistical analysis of different global solar radiation sunshine models. *Renewable and Sustainable Energy Reviews,* 52**,** pp. 1869-1880.

Dorigo, M. and Blum, C., 2005. Ant colony optimization theory: A survey. *Theoretical Computer Science,* 344**,** pp. 243-278.

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.

Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X. and Xiang, Y., 2018. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agricultural and Forest Meteorology,* 263**,** pp. 225-241.

Fatchurrachman, Rudiyanto, Soh, N. C., Shah, R. M., Giap, S. G. E., Setiawan, B. I. and Minasny, B., 2022. High-Resolution Mapping of Paddy Rice Extent and Growth Stages across Peninsular Malaysia Using a Fusion of Sentinel-1 and 2 Time Series Data in Google Earth Engine. *Remote Sensing,* 14**,** pp. 1875.

Feng, Y., Cui, N., Gong, D., Zhang, Q. and Zhao, L., 2017a. Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. *Agricultural Water Management,* 193**,** pp. 163-173.

Feng, Y., Peng, Y., Cui, N., Gong, D. and Zhang, K., 2017b. Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. *Computers and Electronics in Agriculture,* 136**,** pp. 71-78.

Ferreira, L. B. and Da Cunha, F. F., 2020. New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning. *Agricultural Water Management,* 234**,** pp. 106113.

Ferreira, L. B., Da Cunha, F. F., De Oliveira, R. A. and Fernandes Filho, E. I., 2019. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – A new approach. *Journal of Hydrology,* 572**,** pp. 556-570.

Freund, Y. and Schapire, R. E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences,* 55**,** pp. 119-139.

Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics,* 29**,** pp. 1189-1232.

Gocić, M., Motamedi, S., Shamshirband, S., Petković, D., Ch, S., Hashim, R. and Arif, M., 2015. Soft computing approaches for forecasting reference evapotranspiration. *Computers and Electronics in Agriculture,* 113**,** pp. 164-173.

Gocic, M., Petković, D., Shamshirband, S. and Kamsin, A., 2016. Comparative analysis of reference evapotranspiration equations modelling by extreme learning machine. *Computers and Electronics in Agriculture,* 127**,** pp. 56-63.

Goh, R. Y. and Lee, L. S., 2019. Credit scoring: a review on support vector machines and metaheuristic approaches. *Advances in Operations Research,* 2019**,** pp. 1-30.

Gonzalez Del Cerro, R. T., Subathra, M. S. P., Manoj Kumar, N., Verrastro, S. and Thomas George, S., 2021. Modelling the daily reference evapotranspiration in semi-arid region of South India: A case study comparing ANFIS and empirical models. *Information Processing in Agriculture,* 8**,** pp. 173-184.

Granata, F., 2019. Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agricultural Water Management,* 217**,** pp. 303-315.

Grandvalet, Y., 2004. Bagging equalizes influence. *Machine Learning,* 55**,** pp. 251-270.

Güçlü, Y. S., Subyani, A. M. and Şen, Z., 2017. Regional fuzzy chain model for evapotranspiration estimation. *Journal of Hydrology,* 544**,** pp. 233-241.

Hargreaves, G. H. and Samani, Z. A., 1985. Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture,* 1**,** pp. 96-99.

Hassan, M. A., Khalil, A., Kaseb, S. and Kassem, M. A., 2017. Exploring the potential of tree-based ensemble methods in solar radiation modeling. *Applied Energy,* 203**,** pp. 897-916.

Hatamlou, A., 2013. Black hole: A new heuristic optimization approach for data clustering. *Information Sciences,* 222**,** pp. 175-184.

He, X., Xu, T., Xia, Y., Bateni, S. M., Guo, Z., Liu, S., Mao, K., Zhang, Y., Feng, H. and Zhao, J., 2020. A bayesian three-cornered hat (BTCH) method: improving the terrestrial evapotranspiration estimation. *Remote Sensing,* 12**,** pp. 878.

Hebbalaguppae Krishnashetty, P., Balasangameshwara, J., Sreeman, S., Desai, S. and Bengaluru Kantharaju, A., 2021. Cognitive computing models for estimation of reference evapotranspiration: A review. *Cognitive Systems Research,* 70**,** pp. 109-116.

Höge, M., Guthke, A. and Nowak, W., 2019. The hydrologist's guide to Bayesian model selection, averaging and combination. *Journal of Hydrology,* 572**,** pp. 96-107.

Holmes, J. W., 1984. Measuring evapotranspiration by hydrological methods. *Agricultural Water Management,* 8**,** pp. 29-40.

Hu, C., Wu, Q., Li, H., Jian, S., Li, N. and Lou, Z., 2018. Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water,* 10**,** pp. 1543.

Huang, G.-B., Zhu, Q.-Y. and Siew, C.-K., 2006. Extreme learning machine: Theory and applications. *Neurocomputing,* 70**,** pp. 489-501.

Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W. and Zhou, H., 2019. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology,* 574**,** pp. 1029-1041.

Huo, Z., Feng, S., Kang, S. and Dai, X., 2012. Artificial neural network models for reference evapotranspiration in an arid area of northwest China. *Journal of Arid Environments,* 82**,** pp. 81-90.

Hussain, M. I., Muscolo, A., Farooq, M. and Ahmad, W., 2019. Sustainable use and management of non-conventional water resources for rehabilitation of marginal lands in arid and semiarid environments. *Agricultural Water Management,* 221**,** pp. 462-476.

Janga Reddy, M. and Nagesh Kumar, D., 2020. Evolutionary algorithms, swarm intelligence methods, and their applications in water resources engineering: a state-of-the-art review. *H2Open Journal,* 3**,** pp. 135-188.

Jing, W., Yaseen, Z. M., Shahid, S., Saggi, M. K., Tao, H., Kisi, O., Salih, S. Q., Al-Ansari, N. and Chau, K.-W., 2019. Implementation of evolutionary computing models for reference evapotranspiration modeling: short review, assessment and possible future research directions. *Engineering Applications of Computational Fluid Mechanics,* 13**,** pp. 811-823.

Johnvictor, A. C., Durgamahanthi, V., Pariti Venkata, R. M. and Jethi, N., 2020. Critical review of bio‑inspired optimization techniques. *WIREs Computational Statistics,* 14**,** pp. e1528.

Jovic, S., Nedeljkovic, B., Golubovic, Z. and Kostic, N., 2018. Evolutionary algorithm for reference evapotranspiration analysis. *Computers and Electronics in Agriculture,* 150**,** pp. 1-4.

Karaboga, D. and Basturk, B., 2007. Artificial Bee Colony (ABC) Optimization Algorithm for Solving Constrained Optimization Problems. 4529**,** pp. 789-798.

Kass, R. E. and Raftery, A. E., 1995. Bayes Factors. *Journal of the American Statistical Association,* 90**,** pp. 773-795.

Kaya, Y. Z., Zelenakova, M., Üneş, F., Demirci, M., Hlavata, H. and Mesaros, P., 2021. Estimation of daily evapotranspiration in Košice City (Slovakia) using several soft computing techniques. *Theoretical and Applied Climatology,* 144**,** pp. 287-298.

Kennedy, J. and Eberhart, R., 1995. Particle swarm optimization. 4**,** pp. 1942-1948.

Khoshravesh, M., Sefidkouhi, M. a. G. and Valipour, M., 2015. Estimation of reference evapotranspiration using multivariate fractional polynomial, Bayesian regression, and robust regression models in three arid environments. *Applied Water Science,* 7**,** pp. 1911-1922.

Kim, D., Ha, K. J. and Yeo, J. H., 2021. New drought projections over east asia using evapotranspiration deficits from the CMIP6 warming scenarios. *Earth's Future,* 9**,** pp. e2020EF001697.

Kirkpatrick, S., Gelatt, C. D., Jr. and Vecchi, M. P., 1983. Optimization by simulated annealing. *Science,* 220**,** pp. 671-80.

Kisi, O., 2012. Least squares support vector machine for modeling daily reference evapotranspiration. *Irrigation Science,* 31**,** pp. 611-619.

Kisi, O., 2013. Applicability of Mamdani and Sugeno fuzzy genetic approaches for modeling reference evapotranspiration. *Journal of Hydrology,* 504**,** pp. 160-170.

Kisi, O. and Alizamir, M., 2018. Modelling reference evapotranspiration using a new wavelet conjunction heuristic method: Wavelet extreme learning machine vs wavelet neural networks. *Agricultural and Forest Meteorology,* 263**,** pp. 41-48.

Kisi, O. and Cimen, M., 2010. Evapotranspiration modelling using support vector machines. *Hydrological Sciences Journal,* 54**,** pp. 918-928.

Kisi, Ö. and Öztürk, Ö., 2007. Adaptive neurofuzzy computing technique for evapotranspiration estimation. *Journal of Irrigation and Drainage Engineering,* 133**,** pp. 368-379.

Kisi, O. and Zounemat-Kermani, M., 2014. Comparison of two different adaptive neuro-fuzzy inference systems in modelling daily reference evapotranspiration. *Water Resources Management,* 28**,** pp. 2655-2675.

Knipper, K. R., Kustas, W. P., Anderson, M. C., Alfieri, J. G., Prueger, J. H., Hain, C. R., Gao, F., Yang, Y., Mckee, L. G., Nieto, H., Hipps, L. E., Alsina, M. M. and Sanchez, L., 2018. Evapotranspiration estimates derived using thermal-based satellite remote sensing and data fusion for irrigation management in California vineyards. *Irrigation Science,* 37**,** pp. 431-449.

Kumar, D., Adamowski, J., Suresh, R. and Ozga-Zielinski, B., 2016. Estimating evapotranspiration using an extreme learning machine model: case study in North Bihar, India. *Journal of Irrigation and Drainage Engineering,* 142**,** pp. 04016032.

Kumar, M., Raghuwanshi, N. S., Singh, R., Wallender, W. W. and Pruitt, W. O., 2002. Estimating evapotranspiration using artificial neural network. *Journal of Irrigation and Drainage Engineering,* 128**,** pp. 224-233.

Kundu, S., Khare, D. and Mondal, A., 2017. Future changes in rainfall, temperature and reference evapotranspiration in the central India by least square support vector machine. *Geoscience Frontiers,* 8**,** pp. 583-596.

Ladlani, I., Houichi, L., Djemili, L., Heddam, S. and Belouz, K., 2012. Modeling daily reference evapotranspiration (ET0) in the north of Algeria using generalized regression neural networks (GRNN) and radial basis function neural networks (RBFNN): a comparative study. *Meteorology and Atmospheric Physics,* 118**,** pp. 163-178.

Lee, S., Ahmad, A. and Jeon, G., 2018. Combining bootstrap aggregation with support vector regression for small blood pressure measurement. *J Med Syst,* 42**,** pp. 63.

Li, Y., Huang, C., Hou, J., Gu, J., Zhu, G. and Li, X., 2017. Mapping daily evapotranspiration based on spatiotemporal fusion of ASTER and MODIS images over irrigated agricultural areas in the Heihe River Basin, Northwest China. *Agricultural and Forest Meteorology,* 244-245**,** pp. 82-97.

Liu, J., Jia, B., Xie, Z. and Shi, C., 2016. Ensemble simulation of land evapotranspiration in China based on a multi-forcing and multi-model approach. *Advances in Atmospheric Sciences,* 33**,** pp. 673-684.

Liu, X., Xu, C., Zhong, X., Li, Y., Yuan, X. and Cao, J., 2017. Comparison of 16 models for reference crop evapotranspiration against weighing lysimeter measurement. *Agricultural Water Management,* 184**,** pp. 145-155.

Logue, B. A. and Manandhar, E., 2018. Percent residual accuracy for quantifying goodness-of-fit of linear calibration curves. *Talanta,* 189**,** pp. 527-533.

Luo, C., 2018. A comparison analysis for credit scoring using bagging ensembles. *Expert Systems*, pp. e12297.

Luo, Y., Chang, X., Peng, S., Khan, S., Wang, W., Zheng, Q. and Cai, X., 2014. Short-term forecasting of daily reference evapotranspiration using the Hargreaves–Samani model and temperature forecasts. *Agricultural Water Management,* 136, pp. 42-51.

Luo, Y., Traore, S., Lyu, X., Wang, W., Wang, Y., Xie, Y., Jiao, X. and Fipps, G., 2015. Medium range daily reference evapotranspiration forecasting by using ANN and public weather forecasts. *Water Resources Management,* 29, pp. 3863-3876.

Ma, Y., Liu, S., Song, L., Xu, Z., Liu, Y., Xu, T. and Zhu, Z., 2018. Estimation of daily evapotranspiration and irrigation water efficiency at a Landsat-like scale for an arid irrigation area using multi-source remote sensing data. *Remote Sensing of Environment,* 216, pp. 715-734.

Mahidin, M. U. 2021. Gross Domestic Product (GDP) by State 2020. Department of Statistics Malaysia.

Mehdizadeh, S., Behmanesh, J. and Khalili, K., 2017. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Computers and Electronics in Agriculture,* 139, pp. 103-114.

Meng, T., Jing, X., Yan, Z. and Pedrycz, W., 2020. A survey on machine learning for data fusion. *Information Fusion,* 57, pp. 115-129.

Mirjalili, S. and Lewis, A., 2016. The whale optimization algorithm. *Advances in Engineering Software,* 95, pp. 51-67.

Nagappan, M., Gopalakrishnan, V. and Alagappan, M., 2020. Prediction of reference evapotranspiration for irrigation scheduling using machine learning. *Hydrological Sciences Journal,* 65, pp. 2669-2677.

Ndiaye, P. M., Bodian, A., Diop, L. and Djaman, K., 2017. Sensitivity analysis of the penman-monteith reference evapotranspiration to climatic variables: case of Burkina Faso. *Journal of Water Resource and Protection,* 09, pp. 1364-1376.

Nourani, V., Elkiran, G. and Abdullahi, J., 2019. Multi-station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements. *Journal of Hydrology,* 577, pp. 123958.

Odongo, V. O., Van Oel, P. R., Van Der Tol, C. and Su, Z., 2019. Impact of land use and land cover transitions and climate on evapotranspiration in the Lake Naivasha Basin, Kenya. *Sci Total Environ,* 682, pp. 19-30.

Pal, M. and Deswal, S., 2009. M5 model tree based modelling of reference evapotranspiration. *Hydrological Processes,* 23, pp. 1437-1443.

Patil, A. P. and Deka, P. C., 2016. An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs. *Computers and Electronics in Agriculture,* 121, pp. 385-392.

Pereira, L. S., Allen, R. G., Smith, M. and Raes, D., 2015. Crop evapotranspiration estimation with FAO56: Past and future. *Agricultural Water Management,* 147, pp. 4-20.

Petković, D., Gocic, M., Shamshirband, S., Qasem, S. N. and Trajkovic, S., 2015. Particle swarm optimization-based radial basis function network

for estimation of reference evapotranspiration. *Theoretical and Applied Climatology,* 125**,** pp. 555-563.

Pham, Q.-V., Nguyen, D. C., Mirjalili, S., Hoang, D. T., Nguyen, D. N., Pathirana, P. N. and Hwang, W.-J., 2021. Swarm intelligence for next-generation networks: Recent advances and applications. *Journal of Network and Computer Applications,* 191**,** pp. 103141.

Pokorny, J., 2019. Evapotranspiration. *In:* Fath, B. (ed.). *Encyclopedia of Ecology*. 2 ed. Elsevier.

Pour, S. H., Wahab, A. K. A., Shahid, S. and Ismail, Z. B., 2020. Changes in reference evapotranspiration and its driving factors in peninsular Malaysia. *Atmospheric Research,* 246**,** pp. 105096.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. and Gulin, A., CatBoost: unbiased boosting with categorical features. In*: Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018. Montréal, Canada. 3327770: Curran Associates Inc., 6639-6649.

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.

Rahimikhoob, A., 2009. Estimation of evapotranspiration based on only air temperature data using artificial neural networks for a subtropical climate in Iran. *Theoretical and Applied Climatology,* 101**,** pp. 83-91.

Rahimikhoob, A., 2014. Comparison between M5 model tree and neural networks for estimating reference evapotranspiration in an arid environment. *Water Resources Management,* 28**,** pp. 657-669.

Ritchie, J. T., 1972. Model for predicting evaporation from a row crop with incomplete cover. *Water Resources Research,* 8**,** pp. 1204-1213.

Rostami, M., Berahmand, K., Nasiri, E. and Forouzandeh, S., 2021. Review of swarm intelligence-based feature selection methods. *Engineering Applications of Artificial Intelligence,* 100**,** pp. 104210.

Roy, D. K., Barzegar, R., Quilty, J. and Adamowski, J., 2020. Using ensembles of adaptive neuro-fuzzy inference system and optimization algorithms to predict reference evapotranspiration in subtropical climatic zones. *Journal of Hydrology,* 591**,** pp. 125509.

Saggi, M. K. and Jain, S., 2019. Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning. *Computers and Electronics in Agriculture,* 156**,** pp. 387-398.

Semmens, K. A., Anderson, M. C., Kustas, W. P., Gao, F., Alfieri, J. G., Mckee, L., Prueger, J. H., Hain, C. R., Cammalleri, C., Yang, Y., Xia, T., Sanchez, L., Mar Alsina, M. and Vélez, M., 2016. Monitoring daily evapotranspiration over two California vineyards using Landsat 8 in a multi-sensor data fusion approach. *Remote Sensing of Environment,* 185**,** pp. 155-170.

Şen, D., Dönmez, C. Ç. and Yıldırım, U. M., 2020. A hybrid bi-level metaheuristic for credit scoring. *Information Systems Frontiers,* 22**,** pp. 1009-1019.

Shiri, J., Marti, P., Karimi, S. and Landeras, G., 2019. Data splitting strategies for improving data driven models for reference evapotranspiration

estimation among similar stations. *Computers and Electronics in Agriculture,* 162**,** pp. 70-81.

Shiri, J., Nazemi, A. H., Sadraddini, A. A., Landeras, G., Kisi, O., Fakheri Fard, A. and Marti, P., 2014. Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran. *Computers and Electronics in Agriculture,* 108**,** pp. 230-241.

Shrestha, N. K. and Shukla, S., 2015. Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment. *Agricultural and Forest Meteorology,* 200**,** pp. 172-184.

Skurichina, M., Kuncheva, L. I. and Duin, R. P. W., 2002. Bagging and boosting for the nearest mean classifier: effects of sample size on diversity and accuracy. 2364**,** pp. 62-71.

Sowmya, M. R., Santosh Kumar, M. B. and Ambat, S. K., 2020. Comparison of deep neural networks for reference evapotranspiration prediction using minimal meteorological data. pp. 27-33.

Stanhill, G. 2005. Evapotranspiration. *In:* Hillel, D. (ed.) *Encyclopedia of Soils in the Environment.* Elsevier.

Sun, H., Yang, Y., Wu, R., Gui, D., Xue, J., Liu, Y. and Yan, D., 2019. Improving estimation of cropland evapotranspiration by the bayesian model averaging method with surface energy balance models. *Atmosphere,* 10**,** pp. 188.

Szafranek, K., 2019. Bagged neural networks for forecasting Polish (low) inflation. *International Journal of Forecasting,* 35**,** pp. 1042-1059.

Tabari, H., Grismer, M. E. and Trajkovic, S., 2011. Comparative analysis of 31 reference evapotranspiration methods under humid conditions. *Irrigation Science,* 31**,** pp. 107-117.

Tabari, H. and Hosseinzadeh Talaee, P., 2012. Multilayer perceptron for reference evapotranspiration estimation in a semiarid region. *Neural Computing and Applications,* 23**,** pp. 341-348.

Tabari, H., Kisi, O., Ezani, A. and Hosseinzadeh Talaee, P., 2012. SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *Journal of Hydrology,* 444-445**,** pp. 78-89.

Tang, J., Liu, G. and Pan, Q., 2021. A review on representative swarm intelligence algorithms for solving optimization problems: applications and trends. *IEEE/CAA Journal of Automatica Sinica,* 8**,** pp. 1627-1643.

Tao, H., Diop, L., Bodian, A., Djaman, K., Ndiaye, P. M. and Yaseen, Z. M., 2018. Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: Regional case study in Burkina Faso. *Agricultural Water Management,* 208**,** pp. 140-151.

Theng Hue, H., Ng, J. L., Huang, Y. F. and Tan, Y. X., 2022. Evaluation of temporal variability and stationarity of potential evapotranspiration in Peninsular Malaysia. *Water Supply,* 22**,** pp. 1360-1374.

Tikhamarine, Y., Malik, A., Souag-Gamane, D. and Kisi, O., 2020. Artificial intelligence models versus empirical equations for modeling monthly reference evapotranspiration. *Environ Sci Pollut Res Int,* 27**,** pp. 30001-30019.

Trajkovic, S., 2007. Hargreaves versus Penman-Monteith under Humid Conditions. *Journal of Irrigation and Drainage Engineering,* 133**,** pp. 38-42.

Traore, S., Luo, Y. and Fipps, G., 2016. Deployment of artificial neural network for short-term forecasting of evapotranspiration using public weather forecast restricted messages. *Agricultural Water Management,* 163**,** pp. 363-379.

Traore, S., Wang, Y.-M. and Kerh, T., 2010. Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone. *Agricultural Water Management,* 97**,** pp. 707-714.

United Nations, 2019. *World Population Prospects: The 2019 Highlights.* United Nations. Department of Economic and Social Affairs/Population Division.

Valiantzas, J. D., 2013. Simplified forms for the standardized FAO-56 Penman–Monteith reference evapotranspiration using limited weather data. *Journal of Hydrology,* 505**,** pp. 13-23.

Vapnik, V., 1995. *The Nature of Statistical Learning Theory,* New York, Springer-Verlag.

Wang, K. and Dickinson, R. E., 2012. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Reviews of Geophysics,* 50**,** pp. RG2005.

Wang, Y., Zou, R., Liu, F., Zhang, L. and Liu, Q., 2021. A review of wind speed and wind power forecasting with deep neural networks. *Applied Energy,* 304**,** pp. 117766.

Wang, Z., Wu, P., Zhao, X., Cao, X. and Gao, Y., 2013. GANN models for reference evapotranspiration estimation developed with weather data from different climatic regions. *Theoretical and Applied Climatology,* 116**,** pp. 481-489.

Wen, X., Si, J., He, Z., Wu, J., Shao, H. and Yu, H., 2015. Support-Vector-Machine-Based Models for Modeling Daily Reference Evapotranspiration With Limited Climatic Data in Extreme Arid Regions. *Water Resources Management,* 29**,** pp. 3195-3209.

Xiang, K., Li, Y., Horton, R. and Feng, H., 2020. Similarity and difference of potential evapotranspiration and reference crop evapotranspiration – a review. *Agricultural Water Management,* 232**,** pp. 106043.

Xiao, J., 2019. SVM and KNN ensemble learning for traffic incident detection. *Physica A: Statistical Mechanics and its Applications,* 517**,** pp. 29-35.

Xu, C. Y. and Singh, V. P., 2000. Evaluation and generalization of radiation-based methods for calculating evaporation. *Hydrological Processes,* 14**,** pp. 339-349.

Xu, T., Guo, Z., Liu, S., He, X., Meng, Y., Xu, Z., Xia, Y., Xiao, J., Zhang, Y., Ma, Y. and Song, L., 2018. Evaluating different machine learning methods for upscaling evapotranspiration from flux towers to the regional scale. *Journal of Geophysical Research: Atmospheres,* 123**,** pp. 8674-8690.

Yadav, A., Chatterjee, S. and Equeenuddin, S. M., 2021. Suspended sediment yield modeling in Mahanadi River, India by multi-objective

optimization hybridizing artificial intelligence algorithms. *International Journal of Sediment Research,* 36**,** pp. 76-91.

Yang, X.-S., Bekdaş, G. and Nigdeli, S. M., 2016. Review and Applications of Metaheuristic Algorithms in Civil Engineering. 7**,** pp. 1-24.

Yao, Y., Liang, S., Li, X., Zhang, Y., Chen, J., Jia, K., Zhang, X., Fisher, J. B., Wang, X., Zhang, L., Xu, J., Shao, C., Posse, G., Li, Y., Magliulo, V., Varlagin, A., Moors, E. J., Boike, J., Macfarlane, C., Kato, T., Buchmann, N., Billesbach, D. P., Beringer, J., Wolf, S., Papuga, S. A., Wohlfahrt, G., Montagnani, L., Ardö, J., Paul-Limoges, E., Emmel, C., Hörtnagl, L., Sachs, T., Gruening, C., Gioli, B., López-Ballesteros, A., Steinbrecher, R. and Gielen, B., 2017. Estimation of high-resolution terrestrial evapotranspiration from Landsat data using a simple Taylor skill fusion method. *Journal of Hydrology,* 553**,** pp. 508-526.

Yin, J., Deng, Z., Ines, A. V. M., Wu, J. and Rasu, E., 2020. Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bi-directional long short-term memory model (Bi-LSTM). *Agricultural Water Management,* 242**,** pp. 106386.

Zadeh, L. A., 1965. Fuzzy sets. *Information and Control,* 8**,** pp. 38-53.

Zanetti, S. S., Dohler, R. E., Cecílio, R. A., Pezzopane, J. E. M. and Xavier, A. C., 2019. Proposal for the use of daily thermal amplitude for the calibration of the Hargreaves-Samani equation. *Journal of Hydrology,* 571**,** pp. 193-201.

Zendehboudi, A., Baseer, M. A. and Saidur, R., 2018. Application of support vector machine models for forecasting solar and wind energy resources: A review. *Journal of Cleaner Production,* 199**,** pp. 272-285.

Zhao, L., Zhao, X., Zhou, H., Wang, X. and Xing, X., 2021. Prediction model for daily reference crop evapotranspiration based on hybrid algorithm and principal components analysis in Southwest China. *Computers and Electronics in Agriculture,* 190**,** pp. 106424.

Zhao, T., Wang, Q. J. and Schepen, A., 2019. A Bayesian modelling approach to forecasting short-term reference crop evapotranspiration from GCM outputs. *Agricultural and Forest Meteorology,* 269-270**,** pp. 88-101.

Zhu, B., Feng, Y., Gong, D., Jiang, S., Zhao, L. and Cui, N., 2020. Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data. *Computers and Electronics in Agriculture,* 173**,** pp. 105430.

Zhu, G., Li, X., Zhang, K., Ding, Z., Han, T., Ma, J., Huang, C., He, J. and Ma, T., 2016. Multi-model ensemble prediction of terrestrial evapotranspiration across north China using Bayesian model averaging. *Hydrological Processes,* 30**,** pp. 2861-2879.

Zou, L., Zhan, C., Xia, J., Wang, T. and Gippel, C. J., 2017. Implementation of evapotranspiration data assimilation with catchment scale distributed hydrological model via an ensemble Kalman Filter. *Journal of Hydrology,* 549**,** pp. 685-702.

LIST OF PUBLICATIONS

Chia, M.Y., Huang, Y.F., Koo, C.H. and Fung, K.F., 2020. Recent advances in evapotranspiration estimation using artificial intelligence approaches with a focus on hybridization techniques—a review. *Agronomy*, 10(1), p.101.

Chia, M.Y., Huang, Y.F. and Koo, C.H., 2020. Support vector machine enhanced empirical reference evapotranspiration estimation with limited meteorological parameters. *Computers and Electronics in Agriculture*, 175, p.105577.

Chia, M.Y., Huang, Y.F. and Koo, C.H., 2021. Swarm-based optimization as stochastic training strategy for estimation of reference evapotranspiration using extreme learning machine. *Agricultural Water Management*, 243, p.106447.

Chia, M.Y., Huang, Y.F. and Koo, C.H., 2021. Improving reference evapotranspiration estimation using novel inter-model ensemble approaches. *Computers and Electronics in Agriculture*, 187, p.106227.

Chia, M.Y., Huang, Y.F. and Koo, C.H., 2022. Resolving data-hungry nature of machine learning reference evapotranspiration estimating models using inter-model ensembles with various data management schemes. *Agricultural Water Management*, 261, p.107343.

APPENDICES
Appendix A: Performance of Base Machine Learning Models at Different Stations
Table A1: Performance of MLP at Station 48600 (Pulau Langkawi)

Combinations	Variables	Hidden Neurons	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	7	0.0280	0.0365	0.640	0.9990	0.0000
2	5	6	0.0810	0.1147	1.850	0.9896	-0.0003
3	5	6	0.0533	0.0751	1.216	0.9956	-0.0004
4	5	6	0.0445	0.0638	1.015	0.9968	-0.0009
5	5	6	0.1018	0.1526	2.326	0.9817	0.0003
6	5	6	0.1300	0.2006	2.968	0.9685	-0.0003
7	5	6	0.3932	0.4970	8.980	0.8060	-0.0007
8	4	5	0.0861	0.1233	1.966	0.9880	-0.0003
9	4	5	0.0891	0.1249	2.034	0.9877	0.0001
10	4	5	0.1451	0.2230	3.314	0.9609	-0.0002
11	4	5	0.1680	0.2421	3.836	0.9540	-0.0012
12	4	5	0.4240	0.5376	9.683	0.7727	-0.0002
13	4	5	0.0605	0.0855	1.382	0.9942	-0.0005
14	4	5	0.1103	0.1662	2.518	0.9783	-0.0004
15	4	5	0.1458	0.2216	3.330	0.9615	0.0001
16	4	5	0.4159	0.5236	9.497	0.7845	0.0062
17	4	5	0.1048	0.1605	2.392	0.9798	-0.0002
18	4	5	0.1397	0.2123	3.190	0.9647	-0.0002
19	4	5	0.4257	0.5330	9.721	0.7768	-0.0012
20	4	5	0.2474	0.3622	5.650	0.8969	0.0017
21	4	5	0.4333	0.5515	9.896	0.7610	-0.0023
22	4	5	0.4184	0.5267	9.556	0.7820	0.0032
23	3	4	0.1147	0.1576	2.619	0.9804	-0.0007
24	3	4	0.1515	0.2301	3.460	0.9584	-0.0024
25	3	4	0.1727	0.2500	3.944	0.9510	-0.0019
26	3	4	0.4634	0.5818	10.584	0.7339	0.0009
27	3	4	0.1659	0.2457	3.788	0.9526	-0.0002
28	3	4	0.1714	0.2455	3.915	0.9527	-0.0015
29	3	4	0.5396	0.6752	12.323	0.6416	-0.0020
30	3	4	0.3210	0.4515	7.331	0.8399	-0.0015
31	3	4	0.4757	0.6160	10.863	0.7021	-0.0045
32	3	4	0.4544	0.5711	10.377	0.7438	0.0037
33	3	4	0.1145	0.1702	2.615	0.9772	0.0006
34	3	4	0.1471	0.2240	3.360	0.9607	-0.0020
35	3	4	0.4323	0.5403	9.872	0.7707	-0.0003
36	3	4	0.2527	0.3703	5.770	0.8922	-0.0029
37	3	4	0.4643	0.5824	10.603	0.7337	0.0046
38	3	4	0.4396	0.5477	10.038	0.7643	0.0014
39	3	4	0.2481	0.3648	5.665	0.8954	0.0001
40	3	4	0.4907	0.6131	11.207	0.7048	0.0025
41	3	4	0.4470	0.5542	10.207	0.7588	0.0001
42	3	4	0.4997	0.6361	11.412	0.6822	0.0025
43	2	3	0.1943	0.2760	4.437	0.9403	-0.0029
44	2	3	0.1875	0.2701	4.282	0.9427	-0.0003
45	2	3	0.5439	0.6814	12.421	0.6352	-0.0008
46	2	3	0.3290	0.4608	7.514	0.8332	-0.0034
47	2	3	0.5314	0.6766	12.136	0.6407	-0.0124
48	2	3	0.4857	0.6048	11.093	0.7126	-0.0036
49	2	3	0.3538	0.4890	8.080	0.8122	-0.0024
50	2	3	0.7460	0.9353	17.037	0.3131	0.0101
51	2	3	0.5540	0.6905	12.651	0.6256	-0.0038
52	2	3	0.5527	0.7167	12.623	0.5963	-0.0019
53	2	3	0.2546	0.3728	5.814	0.8907	0.0009
54	2	3	0.4934	0.6153	11.268	0.7026	-0.0032
55	2	3	0.4499	0.5581	10.274	0.7552	0.0007
56	2	3	0.5225	0.6597	11.933	0.6581	-0.0001
57	2	3	0.5427	0.6785	12.393	0.6383	0.0005
58	1	2	0.3799	0.5196	8.676	0.7878	-0.0027
59	1	2	0.7733	0.9781	17.660	0.2494	-0.0014
60	1	2	0.5597	0.6980	12.783	0.6174	-0.0006
61	1	2	0.6071	0.7791	13.864	0.5248	-0.0017
62	1	2	0.8457	1.0645	19.313	0.1125	0.0006
63	1	2	0.5423	0.6784	12.385	0.6391	0.0019

Table A2: Performance of MLP at Station 48601 (Bayan Lepas)

Combinations	Variables	Hidden Neurons	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	7	0.0269	0.0335	0.621	0.9989	-0.0002
2	5	6	0.0705	0.0926	1.624	0.9919	-0.0005
3	5	6	0.0494	0.0652	1.139	0.9960	0.0001
4	5	6	0.0425	0.0565	0.981	0.9970	-0.0002
5	5	6	0.0940	0.1367	2.168	0.9823	-0.0003
6	5	6	0.1083	0.1764	2.496	0.9698	0.0008
7	5	6	0.3779	0.4804	8.710	0.7815	-0.0013
8	4	5	0.0760	0.1010	1.751	0.9904	-0.0002
9	4	5	0.0743	0.0974	1.712	0.9910	0.0002
10	4	5	0.1356	0.1955	3.125	0.9638	-0.0028
11	4	5	0.1262	0.1927	2.910	0.9643	-0.0012
12	4	5	0.4032	0.5162	9.294	0.7478	0.0000
13	4	5	0.0517	0.0683	1.191	0.9956	0.0003
14	4	5	0.0990	0.1431	2.282	0.9807	-0.0005
15	4	5	0.1257	0.1944	2.898	0.9638	-0.0017
16	4	5	0.3977	0.5046	9.168	0.7592	0.0016
17	4	5	0.0993	0.1462	2.289	0.9797	-0.0005
18	4	5	0.1175	0.1867	2.709	0.9664	-0.0015
19	4	5	0.3897	0.4940	8.982	0.7691	0.0011
20	4	5	0.2073	0.3303	4.778	0.8964	0.0033
21	4	5	0.3996	0.5093	9.212	0.7544	-0.0038
22	4	5	0.4435	0.5612	10.223	0.7020	-0.0010
23	3	4	0.0979	0.1281	2.257	0.9845	-0.0003
24	3	4	0.1414	0.2034	3.259	0.9608	0.0028
25	3	4	0.1372	0.2046	3.163	0.9600	0.0014
26	3	4	0.4408	0.5591	10.161	0.7050	0.0059
27	3	4	0.1551	0.2177	3.575	0.9552	-0.0011
28	3	4	0.1286	0.1953	2.964	0.9634	0.0009
29	3	4	0.4733	0.6012	10.910	0.6582	0.0037
30	3	4	0.2384	0.3679	5.494	0.8716	-0.0023
31	3	4	0.4391	0.5624	10.122	0.7004	-0.0044
32	3	4	0.4632	0.5876	10.676	0.6731	-0.0028
33	3	4	0.1021	0.1480	2.353	0.9793	-0.0009
34	3	4	0.1281	0.1980	2.952	0.9625	-0.0006
35	3	4	0.4020	0.5085	9.265	0.7557	-0.0017
36	3	4	0.2110	0.3350	4.863	0.8936	0.0000
37	3	4	0.4350	0.5509	10.026	0.7130	-0.0081
38	3	4	0.4534	0.5751	10.450	0.6867	0.0015
39	3	4	0.2070	0.3341	4.771	0.8942	-0.0021
40	3	4	0.4307	0.5450	9.928	0.7188	-0.0023
41	3	4	0.4504	0.5698	10.382	0.6926	-0.0028
42	3	4	0.5232	0.6670	12.060	0.5787	-0.0052
43	2	3	0.1793	0.2412	4.133	0.9449	0.0005
44	2	3	0.1506	0.2183	3.472	0.9547	0.0019
45	2	3	0.4789	0.6067	11.039	0.6520	-0.0005
46	2	3	0.2387	0.3685	5.502	0.8713	-0.0021
47	2	3	0.5066	0.6388	11.677	0.6137	-0.0049
48	2	3	0.4908	0.6217	11.312	0.6348	-0.0061
49	2	3	0.2502	0.3803	5.767	0.8630	-0.0019
50	2	3	0.6490	0.8218	14.959	0.3610	0.0069
51	2	3	0.5204	0.6576	11.996	0.5907	0.0033
52	2	3	0.5557	0.7177	12.810	0.5131	-0.0066
53	2	3	0.2129	0.3376	4.908	0.8919	-0.0010
54	2	3	0.4404	0.5561	10.150	0.7074	0.0010
55	2	3	0.4571	0.5780	10.535	0.6838	0.0016
56	2	3	0.5441	0.6938	12.542	0.5442	-0.0030
57	2	3	0.5475	0.6920	12.620	0.5461	-0.0017
58	1	2	0.2767	0.4041	6.377	0.8454	-0.0009
59	1	2	0.6579	0.8381	15.164	0.3350	-0.0001
60	1	2	0.5253	0.6612	12.108	0.5862	0.0011
61	1	2	0.6077	0.7809	14.007	0.4225	-0.0019
62	1	2	0.7844	0.9935	18.081	0.0654	0.0023
63	1	2	0.5500	0.6980	12.678	0.5385	-0.0014

Table A3: Performance of MLP at Station 48603 (Alor Setar)

Combinations	Variables	Hidden Neurons	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	7	0.0277	0.0357	0.648	0.9989	-0.0001
2	5	6	0.0879	0.1149	2.055	0.9884	-0.0013
3	5	6	0.0453	0.0587	1.060	0.9970	0.0008
4	5	6	0.0400	0.0514	0.935	0.9977	-0.0001
5	5	6	0.0722	0.1034	1.688	0.9906	-0.0006
6	5	6	0.0901	0.1225	2.106	0.9868	-0.0007
7	5	6	0.3119	0.4028	7.290	0.8576	-0.0014
8	4	5	0.0911	0.1185	2.129	0.9877	0.0000
9	4	5	0.0972	0.1286	2.273	0.9854	0.0005
10	4	5	0.1214	0.1653	2.837	0.9760	0.0005
11	4	5	0.1138	0.1508	2.659	0.9801	0.0015
12	4	5	0.3676	0.4708	8.591	0.8052	-0.0008
13	4	5	0.0508	0.0654	1.186	0.9963	-0.0001
14	4	5	0.0726	0.1042	1.698	0.9904	-0.0008
15	4	5	0.1071	0.1457	2.504	0.9813	-0.0004
16	4	5	0.3357	0.4302	7.844	0.8373	0.0013
17	4	5	0.0718	0.1035	1.677	0.9906	0.0000
18	4	5	0.0977	0.1317	2.282	0.9848	0.0007
19	4	5	0.3395	0.4337	7.933	0.8346	-0.0004
20	4	5	0.1827	0.2783	4.269	0.9318	-0.0027
21	4	5	0.3187	0.4118	7.447	0.8509	0.0009
22	4	5	0.3837	0.4919	8.968	0.7873	-0.0012
23	3	4	0.1117	0.1478	2.611	0.9808	-0.0006
24	3	4	0.1404	0.1899	3.281	0.9683	0.0002
25	3	4	0.1244	0.1645	2.906	0.9762	-0.0001
26	3	4	0.4184	0.5331	9.778	0.7502	0.0012
27	3	4	0.1645	0.2293	3.843	0.9537	-0.0007
28	3	4	0.1187	0.1565	2.773	0.9785	0.0000
29	3	4	0.5052	0.6419	11.805	0.6382	0.0068
30	3	4	0.2063	0.2988	4.821	0.9215	0.0032
31	3	4	0.3736	0.4792	8.732	0.7986	0.0043
32	3	4	0.4204	0.5349	9.825	0.7486	0.0028
33	3	4	0.0758	0.1063	1.771	0.9901	-0.0011
34	3	4	0.1119	0.1535	2.614	0.9793	-0.0003
35	3	4	0.3481	0.4450	8.134	0.8257	-0.0012
36	3	4	0.1889	0.2829	4.415	0.9296	-0.0007
37	3	4	0.3511	0.4516	8.205	0.8207	0.0023
38	3	4	0.3994	0.5079	9.333	0.7733	0.0006
39	3	4	0.1867	0.2861	4.363	0.9280	-0.0025
40	3	4	0.3692	0.4708	8.627	0.8050	-0.0061
41	3	4	0.4107	0.5202	9.598	0.7622	0.0026
42	3	4	0.4239	0.5551	9.905	0.7288	-0.0013
43	2	3	0.1708	0.2371	3.991	0.9505	-0.0030
44	2	3	0.1383	0.1835	3.231	0.9704	-0.0011
45	2	3	0.5059	0.6429	11.821	0.6364	0.0029
46	2	3	0.2414	0.3336	5.641	0.9022	-0.0031
47	2	3	0.4674	0.5960	10.922	0.6877	-0.0004
48	2	3	0.4527	0.5776	10.580	0.7066	-0.0007
49	2	3	0.2512	0.3543	5.869	0.8896	-0.0012
50	2	3	0.6899	0.8742	16.123	0.3278	-0.0020
51	2	3	0.5310	0.6709	12.409	0.6041	0.0001
52	2	3	0.4597	0.5962	10.742	0.6873	-0.0039
53	2	3	0.1880	0.2873	4.393	0.9273	-0.0024
54	2	3	0.3704	0.4738	8.655	0.8024	-0.0007
55	2	3	0.4169	0.5259	9.743	0.7567	-0.0017
56	2	3	0.4645	0.5998	10.855	0.6833	-0.0038
57	2	3	0.4923	0.6264	11.504	0.6545	-0.0007
58	1	2	0.2589	0.3622	6.050	0.8848	-0.0012
59	1	2	0.7012	0.8881	16.386	0.3062	0.0004
60	1	2	0.5390	0.6785	12.597	0.5952	0.0011
61	1	2	0.5640	0.7301	13.180	0.5309	0.0040
62	1	2	0.8156	1.0428	19.060	0.0431	0.0029
63	1	2	0.4947	0.6301	11.560	0.6507	0.0003

Table A4: Performance of MLP at Station 48615 (Kota Bharu)

Combinations	Variables	Hidden Neurons	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	7	0.0236	0.0297	0.543	0.9992	0.0000
2	5	6	0.0888	0.1169	2.046	0.9869	0.0007
3	5	6	0.0527	0.0726	1.213	0.9950	-0.0006
4	5	6	0.0480	0.0651	1.106	0.9959	-0.0002
5	5	6	0.0910	0.1288	2.095	0.9841	-0.0007
6	5	6	0.0929	0.1318	2.139	0.9834	-0.0005
7	5	6	0.3583	0.4531	8.253	0.8041	0.0020
8	4	5	0.0925	0.1230	2.131	0.9855	0.0005
9	4	5	0.0934	0.1217	2.150	0.9858	-0.0001
10	4	5	0.1292	0.1716	2.975	0.9718	0.0014
11	4	5	0.1265	0.1654	2.913	0.9739	0.0003
12	4	5	0.3987	0.5118	9.183	0.7502	0.0000
13	4	5	0.0566	0.0779	1.303	0.9942	-0.0007
14	4	5	0.0961	0.1407	2.214	0.9810	-0.0013
15	4	5	0.1132	0.1633	2.607	0.9746	0.0008
16	4	5	0.3845	0.4870	8.854	0.7738	-0.0028
17	4	5	0.0925	0.1321	2.130	0.9833	0.0000
18	4	5	0.1039	0.1431	2.392	0.9805	-0.0002
19	4	5	0.3752	0.4741	8.640	0.7856	-0.0017
20	4	5	0.1564	0.2255	3.603	0.9514	0.0004
21	4	5	0.4000	0.5082	9.211	0.7538	0.0001
22	4	5	0.4232	0.5355	9.747	0.7263	0.0016
23	3	4	0.1226	0.1622	2.825	0.9749	-0.0004
24	3	4	0.1312	0.1762	3.021	0.9703	-0.0009
25	3	4	0.1432	0.1942	3.298	0.9640	0.0009
26	3	4	0.4339	0.5574	9.993	0.7035	-0.0017
27	3	4	0.1457	0.1893	3.354	0.9657	-0.0008
28	3	4	0.1367	0.1802	3.149	0.9691	0.0013
29	3	4	0.4717	0.6037	10.863	0.6522	0.0002
30	3	4	0.1848	0.2475	4.256	0.9416	0.0017
31	3	4	0.4369	0.5612	10.061	0.6994	0.0034
32	3	4	0.4571	0.5805	10.527	0.6782	-0.0009
33	3	4	0.0968	0.1407	2.230	0.9811	0.0005
34	3	4	0.1233	0.1789	2.840	0.9695	-0.0014
35	3	4	0.3923	0.4965	9.034	0.7645	-0.0021
36	3	4	0.1703	0.2437	3.921	0.9435	0.0015
37	3	4	0.4560	0.5740	10.503	0.6859	0.0003
38	3	4	0.4424	0.5590	10.189	0.7016	0.0018
39	3	4	0.1632	0.2304	3.758	0.9495	0.0011
40	3	4	0.4408	0.5539	10.151	0.7077	-0.0002
41	3	4	0.4385	0.5528	10.098	0.7081	0.0020
42	3	4	0.4994	0.6397	11.502	0.6094	0.0004
43	2	3	0.1752	0.2270	4.034	0.9508	0.0004
44	2	3	0.1602	0.2160	3.690	0.9555	0.0001
45	2	3	0.4875	0.6209	11.228	0.6317	0.0009
46	2	3	0.1955	0.2691	4.502	0.9309	-0.0012
47	2	3	0.5407	0.6901	12.451	0.5459	0.0029
48	2	3	0.4988	0.6325	11.487	0.6180	0.0016
49	2	3	0.1996	0.2646	4.597	0.9333	-0.0012
50	2	3	0.6655	0.8582	15.326	0.2976	0.0041
51	2	3	0.5371	0.6748	12.369	0.5651	0.0016
52	2	3	0.5275	0.6776	12.147	0.5625	0.0024
53	2	3	0.1822	0.2697	4.196	0.9305	0.0014
54	2	3	0.4695	0.5897	10.813	0.6689	-0.0009
55	2	3	0.4458	0.5630	10.268	0.6972	0.0025
56	2	3	0.5643	0.7015	12.996	0.5307	0.0024
57	2	3	0.5554	0.6902	12.791	0.5458	-0.0008
58	1	2	0.2257	0.3080	5.199	0.9095	-0.0056
59	1	2	0.6900	0.8902	15.891	0.2460	-0.0066
60	1	2	0.5394	0.6775	12.423	0.5618	0.0031
61	1	2	0.6327	0.7948	14.572	0.3975	0.0019
62	1	2	0.7931	1.0059	18.266	0.0351	0.0103
63	1	2	0.5733	0.7085	13.203	0.5214	-0.0010

Table A5: Performance of MLP at Station 48620 (Sitiawan)

Combinations	Variables	Hidden Neurons	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	7	0.0345	0.0635	0.907	0.9884	-0.0048
2	5	6	0.0529	0.0869	1.390	0.9835	0.0024
3	5	6	0.0407	0.0696	1.070	0.9879	-0.0050
4	5	6	0.0359	0.0638	0.942	0.9890	-0.0001
5	5	6	0.0414	0.0719	1.088	0.9867	-0.0022
6	5	6	0.0733	0.1126	1.924	0.9766	-0.0030
7	5	6	0.2954	0.3713	7.757	0.7830	-0.0009
8	4	5	0.0534	0.0891	1.403	0.9824	-0.0004
9	4	5	0.0559	0.0908	1.468	0.9820	-0.0012
10	4	5	0.0590	0.0961	1.550	0.9804	-0.0035
11	4	5	0.0777	0.1170	2.040	0.9748	-0.0021
12	4	5	0.3209	0.4028	8.426	0.7444	0.0028
13	4	5	0.0434	0.0735	1.141	0.9870	-0.0013
14	4	5	0.0432	0.0739	1.135	0.9867	-0.0015
15	4	5	0.0812	0.1186	2.133	0.9752	-0.0016
16	4	5	0.3135	0.3936	8.233	0.7559	0.0004
17	4	5	0.0417	0.0723	1.096	0.9869	-0.0030
18	4	5	0.0744	0.1129	1.953	0.9762	-0.0004
19	4	5	0.3046	0.3826	8.000	0.7693	0.0012
20	4	5	0.0849	0.1255	2.230	0.9718	-0.0054
21	4	5	0.2989	0.3750	7.850	0.7785	0.0007
22	4	5	0.3511	0.4366	9.219	0.6998	-0.0007
23	3	4	0.0697	0.1074	1.830	0.9780	-0.0031
24	3	4	0.0598	0.0950	1.571	0.9818	0.0023
25	3	4	0.0855	0.1260	2.245	0.9718	-0.0020
26	3	4	0.3637	0.4570	9.550	0.6713	-0.0010
27	3	4	0.0726	0.1120	1.906	0.9755	0.0007
28	3	4	0.0798	0.1204	2.096	0.9736	0.0002
29	3	4	0.3761	0.4745	9.875	0.6456	-0.0009
30	3	4	0.0894	0.1307	2.347	0.9692	-0.0013
31	3	4	0.3257	0.4080	8.553	0.7382	0.0011
32	3	4	0.3667	0.4568	9.631	0.6712	0.0002
33	3	4	0.0486	0.0830	1.276	0.9839	-0.0027
34	3	4	0.0887	0.1299	2.329	0.9704	-0.0010
35	3	4	0.3160	0.3976	8.298	0.7512	0.0021
36	3	4	0.0879	0.1281	2.309	0.9710	-0.0015
37	3	4	0.3420	0.4263	8.981	0.7142	0.0015
38	3	4	0.3624	0.4514	9.518	0.6791	-0.0042
39	3	4	0.0860	0.1257	2.259	0.9723	-0.0030
40	3	4	0.3295	0.4146	8.653	0.7293	-0.0025
41	3	4	0.3596	0.4479	9.442	0.6843	-0.0020
42	3	4	0.3662	0.4564	9.617	0.6718	0.0064
43	2	3	0.0863	0.1279	2.266	0.9707	0.0001
44	2	3	0.0976	0.1423	2.562	0.9649	0.0028
45	2	3	0.3801	0.4791	9.981	0.6389	-0.0003
46	2	3	0.0914	0.1328	2.399	0.9687	-0.0023
47	2	3	0.4081	0.5106	10.717	0.5896	0.0009
48	2	3	0.3935	0.4905	10.333	0.6211	-0.0004
49	2	3	0.0968	0.1390	2.541	0.9665	0.0006
50	2	3	0.5405	0.6814	14.195	0.2684	0.0014
51	2	3	0.4097	0.5135	10.758	0.5846	0.0002
52	2	3	0.3819	0.4766	10.028	0.6425	-0.0050
53	2	3	0.0980	0.1411	2.573	0.9660	-0.0012
54	2	3	0.3453	0.4328	9.067	0.7052	-0.0003
55	2	3	0.3620	0.4514	9.506	0.6789	0.0001
56	2	3	0.4048	0.5022	10.629	0.6032	-0.0031
57	2	3	0.4115	0.5137	10.807	0.5845	0.0017
58	1	2	0.1163	0.1625	3.054	0.9558	-0.0001
59	1	2	0.5484	0.6891	14.400	0.2518	-0.0052
60	1	2	0.4107	0.5150	10.786	0.5822	-0.0001
61	1	2	0.4550	0.5648	11.947	0.4989	-0.0126
62	1	2	0.6362	0.7908	16.707	0.0142	0.0068
63	1	2	0.4170	0.5197	10.952	0.5751	-0.0021

Table A6: Performance of MLP at Station 48623 (Lubok Merbau)

Combinations	Variables	Hidden Neurons	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	7	0.0302	0.0370	0.744	0.9978	0.0002
2	5	6	0.0558	0.0720	1.371	0.9918	-0.0006
3	5	6	0.0370	0.0463	0.910	0.9966	-0.0004
4	5	6	0.0317	0.0391	0.779	0.9976	0.0004
5	5	6	0.0363	0.0465	0.892	0.9966	0.0000
6	5	6	0.0651	0.0864	1.600	0.9882	0.0000
7	5	6	0.2656	0.3339	6.529	0.8236	0.0032
8	4	5	0.0598	0.0767	1.471	0.9907	-0.0011
9	4	5	0.0598	0.0767	1.471	0.9907	-0.0010
10	4	5	0.0594	0.0773	1.462	0.9906	0.0000
11	4	5	0.0756	0.0986	1.859	0.9846	-0.0004
12	4	5	0.3507	0.4402	8.624	0.6944	-0.0027
13	4	5	0.0423	0.0536	1.040	0.9955	-0.0001
14	4	5	0.0400	0.0516	0.983	0.9958	-0.0003
15	4	5	0.0705	0.0932	1.733	0.9862	-0.0013
16	4	5	0.2771	0.3476	6.813	0.8087	0.0005
17	4	5	0.0379	0.0490	0.931	0.9962	-0.0012
18	4	5	0.0663	0.0884	1.631	0.9876	-0.0005
19	4	5	0.2755	0.3457	6.772	0.8107	0.0010
20	4	5	0.0772	0.1090	1.898	0.9812	-0.0012
21	4	5	0.2684	0.3369	6.598	0.8204	-0.0015
22	4	5	0.3080	0.3863	7.572	0.7639	0.0039
23	3	4	0.0767	0.1001	1.886	0.9841	0.0000
24	3	4	0.0617	0.0799	1.517	0.9899	0.0004
25	3	4	0.0808	0.1044	1.986	0.9828	-0.0009
26	3	4	0.3765	0.4725	9.257	0.6474	0.0018
27	3	4	0.0718	0.0954	1.765	0.9856	-0.0012
28	3	4	0.0772	0.1005	1.898	0.9840	-0.0002
29	3	4	0.4246	0.5395	10.440	0.5399	0.0003
30	3	4	0.0870	0.1175	2.138	0.9781	-0.0004
31	3	4	0.3519	0.4424	8.653	0.6907	-0.0023
32	3	4	0.3883	0.4834	9.548	0.6310	0.0013
33	3	4	0.0459	0.0591	1.130	0.9945	-0.0001
34	3	4	0.0747	0.0994	1.837	0.9844	-0.0013
35	3	4	0.2822	0.3530	6.937	0.8027	0.0015
36	3	4	0.0780	0.1094	1.917	0.9810	0.0000
37	3	4	0.2847	0.3559	7.001	0.7994	0.0009
38	3	4	0.3165	0.3977	7.783	0.7499	0.0032
39	3	4	0.0794	0.1128	1.953	0.9798	-0.0007
40	3	4	0.2838	0.3569	6.977	0.7984	-0.0025
41	3	4	0.3143	0.3956	7.728	0.7524	-0.0032
42	3	4	0.3146	0.3970	7.734	0.7508	-0.0018
43	2	3	0.0858	0.1127	2.110	0.9799	-0.0008
44	2	3	0.0921	0.1202	2.265	0.9771	-0.0009
45	2	3	0.4232	0.5391	10.406	0.5405	0.0014
46	2	3	0.0887	0.1196	2.180	0.9774	0.0003
47	2	3	0.3861	0.4849	9.494	0.6282	-0.0013
48	2	3	0.4100	0.5118	10.081	0.5870	0.0027
49	2	3	0.0991	0.1351	2.438	0.9711	-0.0013
50	2	3	0.5242	0.6691	12.888	0.2914	0.0051
51	2	3	0.4535	0.5713	11.149	0.4841	0.0000
52	2	3	0.3946	0.4905	9.701	0.6202	0.0018
53	2	3	0.0831	0.1175	2.044	0.9781	-0.0001
54	2	3	0.2911	0.3651	7.157	0.7890	-0.0006
55	2	3	0.3180	0.4005	7.819	0.7463	-0.0009
56	2	3	0.3385	0.4253	8.322	0.7140	0.0023
57	2	3	0.3404	0.4303	8.370	0.7073	-0.0015
58	1	2	0.1096	0.1474	2.696	0.9656	-0.0003
59	1	2	0.5292	0.6753	13.012	0.2786	0.0021
60	1	2	0.4546	0.5730	11.177	0.4810	0.0004
61	1	2	0.4379	0.5460	10.766	0.5285	-0.0007
62	1	2	0.6184	0.7818	15.206	0.0332	0.0114
63	1	2	0.3484	0.4392	8.566	0.6947	-0.0010

Table A7: Performance of MLP at Station 48625 (Ipoh)

Combinations	Variables	Hidden Neurons	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	7	0.0287	0.0344	0.717	0.9980	-0.0001
2	5	6	0.0627	0.0809	1.570	0.9890	-0.0001
3	5	6	0.0368	0.0457	0.921	0.9965	0.0004
4	5	6	0.0324	0.0395	0.810	0.9974	0.0002
5	5	6	0.0441	0.0567	1.102	0.9946	-0.0003
6	5	6	0.0851	0.1117	2.129	0.9790	-0.0003
7	5	6	0.2779	0.3481	6.953	0.7961	0.0013
8	4	5	0.0651	0.0838	1.629	0.9882	-0.0003
9	4	5	0.0648	0.0833	1.621	0.9884	-0.0010
10	4	5	0.0692	0.0906	1.731	0.9862	-0.0003
11	4	5	0.1081	0.1375	2.706	0.9682	-0.0005
12	4	5	0.3245	0.4106	8.120	0.7164	-0.0015
13	4	5	0.0411	0.0512	1.027	0.9956	-0.0005
14	4	5	0.0454	0.0583	1.135	0.9943	-0.0005
15	4	5	0.0990	0.1290	2.476	0.9720	-0.0005
16	4	5	0.3005	0.3750	7.520	0.7631	-0.0006
17	4	5	0.0461	0.0595	1.154	0.9940	-0.0002
18	4	5	0.0988	0.1292	2.472	0.9719	-0.0001
19	4	5	0.2962	0.3697	7.411	0.7698	-0.0024
20	4	5	0.1156	0.1526	2.893	0.9608	-0.0003
21	4	5	0.2814	0.3531	7.040	0.7903	0.0015
22	4	5	0.3196	0.3978	7.996	0.7338	0.0001
23	3	4	0.0791	0.1010	1.979	0.9829	0.0000
24	3	4	0.0703	0.0913	1.758	0.9860	-0.0001
25	3	4	0.1164	0.1479	2.912	0.9632	-0.0019
26	3	4	0.3705	0.4672	9.270	0.6325	0.0050
27	3	4	0.0840	0.1101	2.102	0.9796	-0.0008
28	3	4	0.1126	0.1427	2.818	0.9657	0.0008
29	3	4	0.4009	0.5106	10.029	0.5613	0.0011
30	3	4	0.1340	0.1716	3.353	0.9505	-0.0013
31	3	4	0.3284	0.4156	8.216	0.7095	0.0019
32	3	4	0.3782	0.4703	9.463	0.6280	0.0019
33	3	4	0.0551	0.0700	1.377	0.9918	-0.0004
34	3	4	0.1021	0.1330	2.553	0.9702	0.0003
35	3	4	0.3044	0.3792	7.617	0.7577	-0.0003
36	3	4	0.1170	0.1541	2.928	0.9600	-0.0005
37	3	4	0.3170	0.3949	7.931	0.7376	-0.0026
38	3	4	0.3294	0.4096	8.242	0.7175	-0.0046
39	3	4	0.1154	0.1522	2.886	0.9610	0.0002
40	3	4	0.3297	0.4122	8.249	0.7141	0.0015
41	3	4	0.3241	0.4035	8.108	0.7259	0.0012
42	3	4	0.3379	0.4207	8.454	0.7021	-0.0007
43	2	3	0.1025	0.1326	2.566	0.9704	-0.0003
44	2	3	0.1187	0.1505	2.969	0.9619	-0.0006
45	2	3	0.4036	0.5133	10.099	0.5570	0.0037
46	2	3	0.1352	0.1728	3.382	0.9498	-0.0006
47	2	3	0.3902	0.4922	9.762	0.5923	-0.0010
48	2	3	0.4053	0.5079	10.140	0.5659	0.0003
49	2	3	0.1390	0.1772	3.478	0.9472	0.0013
50	2	3	0.5498	0.7058	13.755	0.1612	-0.0031
51	2	3	0.4235	0.5329	10.595	0.5224	0.0027
52	2	3	0.3941	0.4893	9.860	0.5972	-0.0007
53	2	3	0.1186	0.1554	2.968	0.9593	-0.0006
54	2	3	0.3350	0.4178	8.383	0.7062	-0.0009
55	2	3	0.3309	0.4117	8.280	0.7145	-0.0006
56	2	3	0.3704	0.4616	9.268	0.6412	-0.0018
57	2	3	0.3741	0.4652	9.361	0.6355	-0.0015
58	1	2	0.1486	0.1891	3.718	0.9398	-0.0008
59	1	2	0.5551	0.7133	13.890	0.1431	-0.0039
60	1	2	0.4229	0.5322	10.582	0.5236	-0.0008
61	1	2	0.4568	0.5704	11.429	0.4526	0.0030
62	1	2	0.5995	0.7625	14.999	0.0222	-0.0003
63	1	2	0.3810	0.4742	9.532	0.6213	-0.0008

Table A8: Performance of MLP at Station 48632 (Cameron Highlands)

Combinations	Variables	Hidden Neurons	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	7	0.0270	0.0323	0.854	0.9991	0.0000
2	5	6	0.0412	0.0529	1.303	0.9975	-0.0004
3	5	6	0.0304	0.0376	0.962	0.9988	-0.0002
4	5	6	0.0295	0.0359	0.932	0.9989	0.0001
5	5	6	0.0389	0.0532	1.231	0.9975	-0.0002
6	5	6	0.0430	0.0588	1.360	0.9969	-0.0004
7	5	6	0.3468	0.4605	10.964	0.8133	-0.0014
8	4	5	0.0424	0.0545	1.341	0.9974	0.0001
9	4	5	0.0481	0.0626	1.521	0.9966	0.0003
10	4	5	0.0508	0.0701	1.607	0.9957	-0.0012
11	4	5	0.0520	0.0703	1.644	0.9956	0.0004
12	4	5	0.3795	0.4949	11.999	0.7845	0.0021
13	4	5	0.0354	0.0460	1.120	0.9981	0.0002
14	4	5	0.0400	0.0553	1.265	0.9973	-0.0002
15	4	5	0.0472	0.0659	1.493	0.9962	-0.0002
16	4	5	0.3635	0.4803	11.492	0.7972	0.0011
17	4	5	0.0396	0.0537	1.252	0.9975	-0.0001
18	4	5	0.0446	0.0613	1.410	0.9967	-0.0007
19	4	5	0.3583	0.4734	11.327	0.8030	-0.0012
20	4	5	0.0586	0.0927	1.853	0.9923	-0.0008
21	4	5	0.3693	0.4913	11.675	0.7882	-0.0023
22	4	5	0.3781	0.4924	11.955	0.7866	0.0018
23	3	4	0.1543	0.2308	4.879	0.9531	0.0004
24	3	4	0.0521	0.0723	1.647	0.9954	0.0007
25	3	4	0.0531	0.0722	1.679	0.9954	-0.0012
26	3	4	0.4249	0.5491	13.434	0.7345	0.0009
27	3	4	0.0627	0.0868	1.982	0.9934	0.0003
28	3	4	0.0565	0.0757	1.786	0.9950	0.0004
29	3	4	0.4661	0.5908	14.737	0.6926	0.0019
30	3	4	0.0658	0.1009	2.081	0.9909	0.0008
31	3	4	0.4028	0.5309	12.734	0.7525	-0.0009
32	3	4	0.4072	0.5255	12.874	0.7570	0.0048
33	3	4	0.0451	0.0621	1.425	0.9966	-0.0002
34	3	4	0.0521	0.0736	1.647	0.9952	-0.0006
35	3	4	0.3742	0.4972	11.830	0.7824	0.0029
36	3	4	0.0593	0.0934	1.875	0.9922	-0.0001
37	3	4	0.4097	0.5403	12.952	0.7436	-0.0020
38	3	4	0.3873	0.5058	12.247	0.7746	-0.0006
39	3	4	0.0587	0.0927	1.855	0.9923	0.0000
40	3	4	0.3956	0.5213	12.506	0.7610	-0.0026
41	3	4	0.3847	0.5002	12.164	0.7800	0.0051
42	3	4	0.4344	0.5630	13.733	0.7209	-0.0021
43	2	3	0.2870	0.3645	9.074	0.8837	-0.0034
44	2	3	0.1644	0.2411	5.198	0.9490	-0.0015
45	2	3	0.4822	0.6111	15.247	0.6711	0.0011
46	2	3	0.0660	0.1013	2.085	0.9909	-0.0006
47	2	3	0.5113	0.6576	16.165	0.6193	0.0009
48	2	3	0.4420	0.5659	13.975	0.7179	-0.0074
49	2	3	0.0724	0.1093	2.289	0.9894	0.0000
50	2	3	0.6150	0.7793	19.443	0.4654	0.0078
51	2	3	0.4738	0.5998	14.980	0.6832	-0.0048
52	2	3	0.4627	0.6003	14.629	0.6827	-0.0013
53	2	3	0.0629	0.0965	1.990	0.9917	0.0001
54	2	3	0.4250	0.5626	13.438	0.7218	-0.0002
55	2	3	0.4006	0.5248	12.667	0.7576	0.0016
56	2	3	0.4798	0.6167	15.171	0.6646	0.0012
57	2	3	0.4589	0.5886	14.509	0.6946	0.0008
58	1	2	0.3020	0.3786	9.548	0.8744	-0.0013
59	1	2	0.7815	0.9861	24.709	0.1431	-0.0056
60	1	2	0.4858	0.6151	15.358	0.6669	-0.0036
61	1	2	0.5548	0.7090	17.541	0.5569	0.0012
62	1	2	0.6435	0.8137	20.346	0.4186	-0.0016
63	1	2	0.4922	0.6364	15.561	0.6434	-0.0089

Table A9: Performance of MLP at Station 48647 (Subang)

Combinations	Variables	Hidden Neurons	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	7	0.0273	0.0322	0.675	0.9989	-0.0001
2	5	6	0.0671	0.0852	1.659	0.9924	0.0000
3	5	6	0.0416	0.0522	1.027	0.9971	-0.0002
4	5	6	0.0361	0.0447	0.893	0.9979	0.0003
5	5	6	0.0557	0.0725	1.376	0.9945	0.0001
6	5	6	0.1011	0.1343	2.499	0.9811	0.0003
7	5	6	0.3760	0.4898	9.294	0.7491	-0.0037
8	4	5	0.0709	0.0901	1.751	0.9915	0.0000
9	4	5	0.0706	0.0895	1.744	0.9916	-0.0003
10	4	5	0.0768	0.0991	1.899	0.9897	0.0000
11	4	5	0.1100	0.1434	2.720	0.9785	-0.0004
12	4	5	0.4721	0.5993	11.668	0.6250	0.0020
13	4	5	0.0433	0.0548	1.070	0.9969	-0.0001
14	4	5	0.0587	0.0764	1.452	0.9939	-0.0005
15	4	5	0.1045	0.1392	2.583	0.9797	-0.0007
16	4	5	0.3843	0.4985	9.499	0.7402	-0.0038
17	4	5	0.0565	0.0735	1.396	0.9944	-0.0002
18	4	5	0.1022	0.1352	2.526	0.9809	-0.0003
19	4	5	0.3816	0.4963	9.432	0.7425	-0.0002
20	4	5	0.1152	0.1514	2.847	0.9760	0.0004
21	4	5	0.4094	0.5303	10.119	0.7058	-0.0069
22	4	5	0.4239	0.5468	10.477	0.6873	-0.0021
23	3	4	0.0885	0.1118	2.188	0.9870	-0.0009
24	3	4	0.0784	0.1006	1.937	0.9894	-0.0001
25	3	4	0.1125	0.1465	2.781	0.9776	-0.0006
26	3	4	0.4977	0.6265	12.303	0.5900	0.0002
27	3	4	0.0883	0.1157	2.182	0.9860	0.0001
28	3	4	0.1135	0.1467	2.805	0.9775	-0.0008
29	3	4	0.5276	0.6644	13.040	0.5386	0.0017
30	3	4	0.1192	0.1554	2.945	0.9748	0.0001
31	3	4	0.4862	0.6190	12.016	0.5994	-0.0016
32	3	4	0.4987	0.6318	12.326	0.5836	0.0023
33	3	4	0.0640	0.0831	1.582	0.9928	-0.0001
34	3	4	0.1105	0.1468	2.732	0.9775	-0.0003
35	3	4	0.3919	0.5059	9.686	0.7324	-0.0051
36	3	4	0.1152	0.1515	2.848	0.9761	0.0008
37	3	4	0.4323	0.5546	10.686	0.6783	-0.0006
38	3	4	0.4321	0.5564	10.679	0.6763	-0.0011
39	3	4	0.1222	0.1587	3.021	0.9737	-0.0004
40	3	4	0.4272	0.5514	10.560	0.6816	-0.0002
41	3	4	0.4295	0.5533	10.615	0.6800	0.0040
42	3	4	0.4615	0.5938	11.407	0.6310	-0.0024
43	2	3	0.1113	0.1436	2.750	0.9785	0.0017
44	2	3	0.1301	0.1685	3.215	0.9703	0.0000
45	2	3	0.5294	0.6668	13.086	0.5356	0.0110
46	2	3	0.1196	0.1557	2.957	0.9747	0.0000
47	2	3	0.5337	0.6728	13.192	0.5271	0.0022
48	2	3	0.5264	0.6616	13.011	0.5426	0.0006
49	2	3	0.1342	0.1721	3.318	0.9691	0.0009
50	2	3	0.6675	0.8391	16.498	0.2628	0.0021
51	2	3	0.5628	0.7080	13.910	0.4760	0.0035
52	2	3	0.5188	0.6583	12.824	0.5470	0.0002
53	2	3	0.1326	0.1729	3.277	0.9688	-0.0001
54	2	3	0.4375	0.5608	10.814	0.6710	-0.0019
55	2	3	0.4351	0.5601	10.754	0.6721	-0.0025
56	2	3	0.4907	0.6271	12.128	0.5885	-0.0001
57	2	3	0.5091	0.6525	12.584	0.5549	-0.0018
58	1	2	0.1561	0.2005	3.859	0.9580	0.0011
59	1	2	0.6779	0.8535	16.755	0.2376	-0.0004
60	1	2	0.5648	0.7116	13.961	0.4709	0.0066
61	1	2	0.5717	0.7195	14.130	0.4597	-0.0017
62	1	2	0.7626	0.9501	18.851	0.0562	0.0159
63	1	2	0.5133	0.6560	12.688	0.5501	-0.0012

Table A10: Performance of MLP at Station 48649 (Muadzam Shah)

Combinations	Variables	Hidden Neurons	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	7	0.0260	0.0352	0.746	0.9980	-0.0003
2	5	6	0.0393	0.0535	1.130	0.9959	0.0002
3	5	6	0.0298	0.0410	0.855	0.9974	-0.0001
4	5	6	0.0280	0.0382	0.804	0.9977	0.0003
5	5	6	0.0329	0.0473	0.944	0.9967	0.0003
6	5	6	0.0589	0.0809	1.693	0.9907	-0.0008
7	5	6	0.2655	0.3352	7.631	0.8421	-0.0027
8	4	5	0.0406	0.0554	1.167	0.9956	0.0002
9	4	5	0.0454	0.0615	1.304	0.9946	0.0008
10	4	5	0.0431	0.0596	1.239	0.9949	0.0000
11	4	5	0.0595	0.0812	1.709	0.9907	-0.0012
12	4	5	0.3332	0.4173	9.575	0.7553	-0.0014
13	4	5	0.0354	0.0490	1.017	0.9965	-0.0013
14	4	5	0.0347	0.0497	0.998	0.9964	-0.0005
15	4	5	0.0653	0.0892	1.877	0.9888	-0.0004
16	4	5	0.2786	0.3514	8.005	0.8265	-0.0020
17	4	5	0.0342	0.0486	0.982	0.9965	0.0001
18	4	5	0.0599	0.0823	1.722	0.9904	0.0009
19	4	5	0.2743	0.3471	7.882	0.8308	0.0028
20	4	5	0.0730	0.0998	2.096	0.9860	0.0007
21	4	5	0.2766	0.3491	7.947	0.8288	-0.0020
22	4	5	0.3207	0.4000	9.216	0.7752	-0.0004
23	3	4	0.0572	0.0775	1.644	0.9915	0.0001
24	3	4	0.0443	0.0614	1.272	0.9946	-0.0007
25	3	4	0.0662	0.0904	1.902	0.9884	-0.0005
26	3	4	0.3825	0.4793	10.990	0.6776	-0.0019
27	3	4	0.0506	0.0691	1.455	0.9932	0.0000
28	3	4	0.0610	0.0834	1.753	0.9902	-0.0001
29	3	4	0.4321	0.5451	12.417	0.5823	-0.0007
30	3	4	0.0739	0.1017	2.123	0.9854	-0.0010
31	3	4	0.3412	0.4283	9.804	0.7423	-0.0019
32	3	4	0.3655	0.4547	10.502	0.7094	0.0020
33	3	4	0.0398	0.0559	1.143	0.9955	-0.0005
34	3	4	0.0708	0.0961	2.034	0.9870	0.0003
35	3	4	0.2834	0.3569	8.145	0.8210	-0.0014
36	3	4	0.0763	0.1036	2.193	0.9849	0.0004
37	3	4	0.3090	0.3902	8.879	0.7861	-0.0020
38	3	4	0.3261	0.4059	9.370	0.7685	-0.0019
39	3	4	0.0741	0.1015	2.130	0.9855	0.0001
40	3	4	0.3034	0.3831	8.718	0.7939	-0.0052
41	3	4	0.3267	0.4069	9.388	0.7675	0.0041
42	3	4	0.3771	0.4729	10.836	0.6858	0.0005
43	2	3	0.0611	0.0826	1.756	0.9904	-0.0002
44	2	3	0.0733	0.0983	2.107	0.9864	-0.0008
45	2	3	0.4341	0.5469	12.475	0.5797	-0.0024
46	2	3	0.0773	0.1073	2.221	0.9838	-0.0009
47	2	3	0.4354	0.5473	12.512	0.5792	-0.0038
48	2	3	0.4003	0.4979	11.503	0.6523	-0.0037
49	2	3	0.0752	0.1034	2.160	0.9849	-0.0004
50	2	3	0.6180	0.7877	17.759	0.1280	0.0027
51	2	3	0.4417	0.5562	12.693	0.5654	0.0029
52	2	3	0.3991	0.5012	11.467	0.6468	-0.0012
53	2	3	0.0835	0.1129	2.398	0.9820	-0.0009
54	2	3	0.3169	0.4000	9.106	0.7751	-0.0014
55	2	3	0.3300	0.4100	9.484	0.7640	-0.0026
56	2	3	0.4164	0.5194	11.967	0.6209	-0.0023
57	2	3	0.4236	0.5281	12.173	0.6081	-0.0002
58	1	2	0.0851	0.1146	2.446	0.9815	-0.0019
59	1	2	0.6254	0.7947	17.973	0.1136	0.0053
60	1	2	0.4421	0.5563	12.704	0.5651	0.0022
61	1	2	0.4792	0.6026	13.770	0.4903	-0.0075
62	1	2	0.6623	0.8368	19.032	0.0162	0.0029
63	1	2	0.4352	0.5391	12.506	0.5917	0.0036

Table A11: Performance of MLP at Station 48650 (KLIA)

Combinations	Variables	Hidden Neurons	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	7	0.0262	0.0308	0.646	0.9990	-0.0001
2	5	6	0.0745	0.0960	1.841	0.9906	0.0004
3	5	6	0.0438	0.0572	1.083	0.9967	0.0001
4	5	6	0.0395	0.0502	0.976	0.9974	-0.0001
5	5	6	0.0709	0.0994	1.752	0.9900	-0.0015
6	5	6	0.0896	0.1289	2.213	0.9831	-0.0018
7	5	6	0.3802	0.4793	9.389	0.7666	-0.0004
8	4	5	0.0791	0.1021	1.953	0.9894	-0.0002
9	4	5	0.0770	0.0990	1.902	0.9901	0.0003
10	4	5	0.1074	0.1498	2.651	0.9773	-0.0014
11	4	5	0.1135	0.1536	2.803	0.9760	-0.0005
12	4	5	0.4246	0.5333	10.486	0.7111	-0.0018
13	4	5	0.0459	0.0597	1.134	0.9964	0.0001
14	4	5	0.0750	0.1061	1.852	0.9886	-0.0009
15	4	5	0.1027	0.1495	2.537	0.9773	-0.0003
16	4	5	0.3969	0.4990	9.801	0.7471	0.0006
17	4	5	0.0720	0.1014	1.779	0.9896	-0.0009
18	4	5	0.0976	0.1412	2.411	0.9797	-0.0003
19	4	5	0.3933	0.4949	9.713	0.7516	0.0041
20	4	5	0.1348	0.2117	3.329	0.9545	-0.0032
21	4	5	0.4015	0.5063	9.917	0.7397	-0.0012
22	4	5	0.4197	0.5306	10.365	0.7140	-0.0066
23	3	4	0.0891	0.1152	2.201	0.9865	0.0002
24	3	4	0.1078	0.1495	2.662	0.9774	-0.0010
25	3	4	0.1214	0.1689	2.999	0.9713	0.0001
26	3	4	0.4542	0.5711	11.217	0.6685	0.0014
27	3	4	0.1236	0.1704	3.052	0.9706	0.0006
28	3	4	0.1165	0.1595	2.876	0.9741	-0.0003
29	3	4	0.4904	0.6244	12.110	0.6037	0.0048
30	3	4	0.1606	0.2409	3.967	0.9411	-0.0010
31	3	4	0.4510	0.5656	11.138	0.6750	0.0025
32	3	4	0.4657	0.5831	11.501	0.6549	0.0074
33	3	4	0.0785	0.1084	1.939	0.9881	-0.0002
34	3	4	0.1034	0.1502	2.554	0.9771	0.0011
35	3	4	0.3977	0.4999	9.821	0.7460	0.0040
36	3	4	0.1392	0.2185	3.437	0.9515	-0.0008
37	3	4	0.4245	0.5334	10.484	0.7110	-0.0047
38	3	4	0.4309	0.5444	10.642	0.6988	-0.0014
39	3	4	0.1353	0.2131	3.341	0.9538	-0.0017
40	3	4	0.4369	0.5486	10.789	0.6943	0.0017
41	3	4	0.4284	0.5415	10.581	0.7023	0.0012
42	3	4	0.4668	0.5922	11.528	0.6447	0.0083
43	2	3	0.1419	0.1928	3.504	0.9623	0.0000
44	2	3	0.1271	0.1725	3.139	0.9697	0.0006
45	2	3	0.4931	0.6271	12.179	0.6006	-0.0005
46	2	3	0.1610	0.2418	3.975	0.9406	-0.0006
47	2	3	0.4959	0.6182	12.246	0.6115	-0.0020
48	2	3	0.4856	0.6105	11.993	0.6211	0.0010
49	2	3	0.1733	0.2564	4.281	0.9332	-0.0018
50	2	3	0.6653	0.8333	16.432	0.2946	0.0020
51	2	3	0.5192	0.6570	12.821	0.5611	0.0028
52	2	3	0.5110	0.6468	12.620	0.5751	0.0025
53	2	3	0.1431	0.2224	3.534	0.9497	-0.0016
54	2	3	0.4391	0.5520	10.845	0.6908	-0.0001
55	2	3	0.4334	0.5468	10.703	0.6962	-0.0006
56	2	3	0.4857	0.6157	11.994	0.6152	-0.0083
57	2	3	0.5030	0.6355	12.423	0.5899	0.0025
58	1	2	0.1928	0.2824	4.761	0.9191	-0.0028
59	1	2	0.6848	0.8573	16.913	0.2539	0.0067
60	1	2	0.5194	0.6577	12.827	0.5603	0.0034
61	1	2	0.5515	0.6970	13.620	0.5070	-0.0001
62	1	2	0.7566	0.9529	18.684	0.0790	-0.0003
63	1	2	0.5044	0.6379	12.457	0.5866	-0.0062

216

Table A12: Performance of MLP at Station 48657 (Kuantan)

Combinations	Variables	Hidden Neurons	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	7	0.0235	0.0289	0.611	0.9990	0.0001
2	5	6	0.0637	0.0835	1.655	0.9918	0.0002
3	5	6	0.0347	0.0439	0.902	0.9977	-0.0005
4	5	6	0.0321	0.0402	0.833	0.9981	-0.0001
5	5	6	0.0502	0.0665	1.303	0.9948	-0.0001
6	5	6	0.0664	0.0873	1.725	0.9910	-0.0003
7	5	6	0.2784	0.3591	7.227	0.8474	-0.0012
8	4	5	0.0660	0.0858	1.713	0.9913	0.0001
9	4	5	0.0675	0.0879	1.751	0.9909	0.0005
10	4	5	0.0787	0.1028	2.043	0.9875	0.0001
11	4	5	0.0834	0.1072	2.164	0.9864	-0.0006
12	4	5	0.3493	0.4480	9.068	0.7627	-0.0001
13	4	5	0.0388	0.0494	1.008	0.9971	-0.0002
14	4	5	0.0514	0.0679	1.335	0.9945	0.0002
15	4	5	0.0759	0.0990	1.971	0.9884	-0.0005
16	4	5	0.2968	0.3823	7.706	0.8271	-0.0021
17	4	5	0.0508	0.0672	1.318	0.9947	-0.0001
18	4	5	0.0733	0.0952	1.904	0.9893	0.0009
19	4	5	0.2905	0.3747	7.542	0.8337	-0.0008
20	4	5	0.0939	0.1236	2.438	0.9819	-0.0012
21	4	5	0.2884	0.3703	7.487	0.8378	-0.0029
22	4	5	0.3343	0.4251	8.678	0.7860	0.0004
23	3	4	0.0874	0.1150	2.269	0.9844	-0.0006
24	3	4	0.0791	0.1034	2.054	0.9874	0.0009
25	3	4	0.0878	0.1123	2.278	0.9851	-0.0004
26	3	4	0.4026	0.5099	10.452	0.6925	0.0018
27	3	4	0.0903	0.1200	2.343	0.9829	-0.0004
28	3	4	0.0838	0.1075	2.175	0.9864	-0.0006
29	3	4	0.4483	0.5675	11.638	0.6190	0.0012
30	3	4	0.1043	0.1351	2.708	0.9784	0.0011
31	3	4	0.3596	0.4611	9.336	0.7487	0.0000
32	3	4	0.3911	0.4952	10.155	0.7100	-0.0025
33	3	4	0.0553	0.0721	1.435	0.9939	-0.0005
34	3	4	0.0787	0.1020	2.043	0.9877	0.0001
35	3	4	0.3039	0.3906	7.889	0.8194	-0.0045
36	3	4	0.0965	0.1258	2.506	0.9812	0.0005
37	3	4	0.3267	0.4173	8.483	0.7938	-0.0016
38	3	4	0.3520	0.4478	9.137	0.7625	0.0012
39	3	4	0.0944	0.1241	2.451	0.9817	0.0003
40	3	4	0.3166	0.4059	8.219	0.8050	-0.0017
41	3	4	0.3468	0.4401	9.003	0.7705	0.0015
42	3	4	0.3672	0.4647	9.533	0.7444	0.0002
43	2	3	0.1103	0.1441	2.863	0.9754	-0.0005
44	2	3	0.0969	0.1251	2.516	0.9815	-0.0002
45	2	3	0.4534	0.5719	11.772	0.6132	-0.0004
46	2	3	0.1046	0.1355	2.716	0.9783	-0.0006
47	2	3	0.4729	0.5980	12.278	0.5771	0.0044
48	2	3	0.4359	0.5501	11.316	0.6419	0.0028
49	2	3	0.1069	0.1404	2.775	0.9767	0.0001
50	2	3	0.6773	0.8544	17.584	0.1366	0.0035
51	2	3	0.4715	0.5933	12.242	0.5834	0.0031
52	2	3	0.4175	0.5301	10.838	0.6675	0.0024
53	2	3	0.0981	0.1282	2.546	0.9805	-0.0010
54	2	3	0.3368	0.4296	8.745	0.7815	-0.0005
55	2	3	0.3618	0.4588	9.392	0.7508	-0.0039
56	2	3	0.4221	0.5343	10.958	0.6619	0.0002
57	2	3	0.4037	0.5098	10.480	0.6923	0.0016
58	1	2	0.1197	0.1560	3.107	0.9712	-0.0025
59	1	2	0.6881	0.8679	17.865	0.1083	-0.0012
60	1	2	0.4724	0.5939	12.265	0.5827	0.0046
61	1	2	0.5424	0.6889	14.082	0.4382	-0.0029
62	1	2	0.7280	0.9104	18.901	0.0191	0.0006
63	1	2	0.4373	0.5516	11.354	0.6398	-0.0052

Table A13: Performance of SVM at Station 48600 (Pulau Langkawi)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	0.0480	0.0581	1.097	0.9976	-0.0180
2	5	0.0824	0.1167	1.882	0.9893	0.0024
3	5	0.0589	0.0788	1.345	0.9952	-0.0110
4	5	0.0512	0.0693	1.169	0.9963	-0.0090
5	5	0.1062	0.1591	2.426	0.9802	-0.0078
6	5	0.1310	0.2034	2.991	0.9678	-0.0137
7	5	0.4062	0.5150	9.276	0.7916	-0.0002
8	4	0.0866	0.1246	1.978	0.9878	0.0010
9	4	0.0895	0.1259	2.043	0.9876	0.0004
10	4	0.1491	0.2307	3.404	0.9582	0.0030
11	4	0.1663	0.2449	3.797	0.9535	-0.0250
12	4	0.4333	0.5557	9.896	0.7580	0.0190
13	4	0.0632	0.0846	1.443	0.9945	-0.0105
14	4	0.1131	0.1704	2.584	0.9772	-0.0085
15	4	0.1433	0.2232	3.272	0.9613	-0.0205
16	4	0.4271	0.5369	9.753	0.7736	-0.0019
17	4	0.1072	0.1626	2.448	0.9793	-0.0084
18	4	0.1400	0.2154	3.198	0.9639	-0.0166
19	4	0.4362	0.5451	9.961	0.7666	-0.0068
20	4	0.2325	0.3737	5.309	0.8953	-0.0771
21	4	0.4421	0.5680	10.097	0.7469	-0.0036
22	4	0.4317	0.5413	9.859	0.7698	-0.0004
23	3	0.1131	0.1547	2.583	0.9812	-0.0031
24	3	0.1530	0.2372	3.494	0.9560	0.0050
25	3	0.1709	0.2526	3.903	0.9505	-0.0261
26	3	0.4680	0.5910	10.687	0.7263	0.0284
27	3	0.1665	0.2485	3.802	0.9516	-0.0013
28	3	0.1694	0.2490	3.868	0.9521	-0.0283
29	3	0.5453	0.6855	12.453	0.6322	0.0423
30	3	0.3017	0.4755	6.890	0.8371	-0.1274
31	3	0.4774	0.6276	10.903	0.6917	0.0208
32	3	0.4603	0.5831	10.511	0.7332	0.0096
33	3	0.1166	0.1730	2.662	0.9765	-0.0078
34	3	0.1461	0.2261	3.337	0.9603	-0.0212
35	3	0.4393	0.5487	10.032	0.7637	-0.0091
36	3	0.2375	0.3834	5.423	0.8904	-0.0843
37	3	0.4718	0.5959	10.775	0.7215	-0.0116
38	3	0.4459	0.5563	10.183	0.7568	-0.0015
39	3	0.2344	0.3778	5.354	0.8934	-0.0799
40	3	0.4965	0.6195	11.338	0.6987	-0.0069
41	3	0.4540	0.5637	10.369	0.7504	-0.0086
42	3	0.5036	0.6456	11.501	0.6729	-0.0224
43	2	0.1921	0.2778	4.388	0.9395	-0.0018
44	2	0.1852	0.2720	4.230	0.9424	-0.0252
45	2	0.5510	0.6915	12.583	0.6250	0.0332
46	2	0.3060	0.4799	6.987	0.8333	-0.1271
47	2	0.5304	0.6794	12.112	0.6385	0.0148
48	2	0.4858	0.6102	11.095	0.7083	0.0264
49	2	0.3283	0.5156	7.498	0.8115	-0.1484
50	2	0.7414	0.9414	16.931	0.3118	0.0983
51	2	0.5589	0.6990	12.763	0.6171	0.0294
52	2	0.5565	0.7277	12.708	0.5847	-0.0206
53	2	0.2387	0.3854	5.452	0.8891	-0.0850
54	2	0.4967	0.6199	11.344	0.6985	-0.0054
55	2	0.4557	0.5656	10.407	0.7487	-0.0102
56	2	0.5233	0.6634	11.951	0.6548	-0.0265
57	2	0.5438	0.6801	12.419	0.6367	-0.0122
58	1	0.3509	0.5453	8.014	0.7878	-0.1592
59	1	0.7662	0.9853	17.498	0.2492	0.1168
60	1	0.5653	0.7072	12.911	0.6079	0.0344
61	1	0.6046	0.7796	13.807	0.5226	-0.0125
62	1	0.8467	1.0663	19.337	0.1090	0.0331
63	1	0.5440	0.6805	12.424	0.6363	-0.0132

Table A14: Performance of SVM at Station 48601 (Bayan Lepas)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	0.0429	0.0542	0.989	0.9975	-0.0159
2	5	0.0829	0.1112	1.910	0.9884	-0.0017
3	5	0.0542	0.0727	1.248	0.9950	-0.0030
4	5	0.0474	0.0641	1.092	0.9962	0.0011
5	5	0.1027	0.1464	2.367	0.9800	-0.0051
6	5	0.1033	0.1480	2.381	0.9795	-0.0037
7	5	0.3822	0.4952	8.809	0.7696	-0.0123
8	4	0.0860	0.1155	1.982	0.9875	-0.0019
9	4	0.0895	0.1228	2.063	0.9859	-0.0043
10	4	0.1404	0.1975	3.236	0.9635	-0.0057
11	4	0.1275	0.1742	2.940	0.9717	-0.0098
12	4	0.4167	0.5398	9.605	0.7263	0.0062
13	4	0.0562	0.0755	1.296	0.9947	-0.0038
14	4	0.1067	0.1515	2.459	0.9785	-0.0032
15	4	0.1209	0.1718	2.787	0.9724	-0.0079
16	4	0.4054	0.5194	9.343	0.7464	-0.0132
17	4	0.1041	0.1485	2.399	0.9794	-0.0046
18	4	0.1130	0.1606	2.605	0.9759	-0.0052
19	4	0.3975	0.5066	9.163	0.7588	-0.0169
20	4	0.2017	0.3482	4.649	0.8912	-0.0728
21	4	0.4039	0.5249	9.311	0.7409	-0.0092
22	4	0.4340	0.5543	10.004	0.7111	-0.0025
23	3	0.1072	0.1405	2.471	0.9815	-0.0031
24	3	0.1433	0.2018	3.304	0.9620	-0.0084
25	3	0.1358	0.1856	3.129	0.9679	-0.0134
26	3	0.4708	0.5967	10.852	0.6654	0.0083
27	3	0.1628	0.2263	3.753	0.9522	-0.0142
28	3	0.1324	0.1814	3.052	0.9693	-0.0127
29	3	0.5206	0.6592	11.999	0.5919	0.0234
30	3	0.2244	0.3717	5.172	0.8766	-0.0815
31	3	0.4419	0.5734	10.186	0.6914	0.0119
32	3	0.4634	0.5928	10.682	0.6702	0.0140
33	3	0.1114	0.1559	2.568	0.9772	-0.0024
34	3	0.1241	0.1753	2.860	0.9713	-0.0084
35	3	0.4068	0.5194	9.377	0.7466	-0.0160
36	3	0.2070	0.3534	4.772	0.8877	-0.0735
37	3	0.4368	0.5606	10.068	0.7046	-0.0095
38	3	0.4489	0.5722	10.348	0.6919	-0.0045
39	3	0.2033	0.3503	4.687	0.8898	-0.0731
40	3	0.4383	0.5593	10.102	0.7060	-0.0131
41	3	0.4452	0.5645	10.261	0.7002	-0.0081
42	3	0.5196	0.6792	11.976	0.5673	-0.0319
43	2	0.1821	0.2438	4.198	0.9443	-0.0078
44	2	0.1485	0.1995	3.424	0.9629	-0.0113
45	2	0.5216	0.6603	12.024	0.5905	0.0218
46	2	0.2261	0.3735	5.211	0.8757	-0.0840
47	2	0.5181	0.6557	11.943	0.5961	0.0224
48	2	0.5045	0.6373	11.630	0.6184	0.0174
49	2	0.2395	0.3874	5.520	0.8667	-0.0878
50	2	0.6673	0.8552	15.382	0.3194	0.0858
51	2	0.5556	0.6962	12.807	0.5454	0.0278
52	2	0.5481	0.7176	12.633	0.5159	-0.0155
53	2	0.2135	0.3572	4.922	0.8848	-0.0711
54	2	0.4464	0.5686	10.289	0.6962	-0.0112
55	2	0.4503	0.5718	10.380	0.6925	-0.0077
56	2	0.5440	0.7043	12.538	0.5339	-0.0277
57	2	0.5626	0.7178	12.968	0.5164	-0.0362
58	1	0.2643	0.4054	6.092	0.8513	-0.0793
59	1	0.6698	0.8625	15.438	0.3086	0.0928
60	1	0.5556	0.6960	12.806	0.5455	0.0305
61	1	0.6116	0.7854	14.096	0.4197	-0.0039
62	1	0.7975	1.0143	18.381	0.0379	0.0568
63	1	0.5633	0.7194	12.983	0.5139	-0.0340

Table A15: Performance of SVM at Station 48603 (Alor Setar)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	0.0437	0.0538	1.021	0.9976	-0.0134
2	5	0.0883	0.1156	2.064	0.9883	0.0001
3	5	0.0521	0.0677	1.217	0.9961	-0.0095
4	5	0.0460	0.0599	1.075	0.9969	-0.0045
5	5	0.0765	0.1080	1.788	0.9898	-0.0062
6	5	0.0906	0.1229	2.117	0.9868	-0.0022
7	5	0.3188	0.4122	7.449	0.8507	0.0111
8	4	0.0915	0.1192	2.139	0.9875	-0.0011
9	4	0.0978	0.1298	2.286	0.9852	-0.0017
10	4	0.1236	0.1668	2.888	0.9756	-0.0023
11	4	0.1142	0.1517	2.669	0.9798	-0.0052
12	4	0.3699	0.4750	8.643	0.8021	0.0212
13	4	0.0563	0.0720	1.316	0.9956	-0.0111
14	4	0.0779	0.1095	1.821	0.9895	-0.0059
15	4	0.1082	0.1482	2.528	0.9807	-0.0017
16	4	0.3449	0.4434	8.060	0.8271	0.0075
17	4	0.0768	0.1076	1.796	0.9899	-0.0056
18	4	0.0981	0.1322	2.292	0.9847	-0.0015
19	4	0.3552	0.4519	8.300	0.8203	0.0026
20	4	0.1728	0.2840	4.038	0.9312	-0.0464
21	4	0.3280	0.4219	7.665	0.8436	0.0091
22	4	0.3874	0.4989	9.054	0.7815	0.0231
23	3	0.1118	0.1487	2.613	0.9806	-0.0043
24	3	0.1426	0.1918	3.333	0.9677	-0.0031
25	3	0.1238	0.1642	2.893	0.9763	-0.0068
26	3	0.4231	0.5434	9.887	0.7414	0.0342
27	3	0.1678	0.2334	3.921	0.9522	-0.0103
28	3	0.1186	0.1570	2.773	0.9784	-0.0071
29	3	0.5142	0.6548	12.017	0.6259	0.0604
30	3	0.2027	0.3071	4.737	0.9196	-0.0518
31	3	0.3779	0.4851	8.832	0.7935	0.0207
32	3	0.4235	0.5418	9.896	0.7427	0.0298
33	3	0.0797	0.1102	1.863	0.9894	-0.0054
34	3	0.1129	0.1557	2.639	0.9787	-0.0019
35	3	0.3582	0.4545	8.370	0.8182	0.0049
36	3	0.1792	0.2904	4.189	0.9285	-0.0522
37	3	0.3571	0.4590	8.345	0.8147	0.0052
38	3	0.4035	0.5154	9.430	0.7668	0.0194
39	3	0.1779	0.2918	4.158	0.9278	-0.0519
40	3	0.3755	0.4779	8.775	0.7991	0.0016
41	3	0.4180	0.5279	9.768	0.7551	0.0187
42	3	0.4261	0.5585	9.958	0.7257	0.0092
43	2	0.1729	0.2379	4.040	0.9503	-0.0110
44	2	0.1376	0.1835	3.216	0.9705	-0.0075
45	2	0.5181	0.6577	12.109	0.6230	0.0617
46	2	0.2310	0.3455	5.398	0.9017	-0.0793
47	2	0.4678	0.5999	10.932	0.6841	0.0298
48	2	0.4531	0.5818	10.590	0.7031	0.0328
49	2	0.2424	0.3675	5.664	0.8882	-0.0831
50	2	0.6859	0.8860	16.028	0.3198	0.1059
51	2	0.5416	0.6870	12.657	0.5881	0.0602
52	2	0.4623	0.6009	10.804	0.6825	0.0193
53	2	0.1808	0.2939	4.225	0.9266	-0.0517
54	2	0.3765	0.4794	8.799	0.7979	0.0027
55	2	0.4214	0.5319	9.849	0.7514	0.0169
56	2	0.4653	0.6030	10.873	0.6799	-0.0027
57	2	0.4928	0.6283	11.517	0.6527	0.0169
58	1	0.2499	0.3744	5.840	0.8839	-0.0840
59	1	0.6925	0.8965	16.184	0.3069	0.1242
60	1	0.5492	0.6941	12.835	0.5806	0.0677
61	1	0.5643	0.7318	13.186	0.5287	0.0017
62	1	0.8159	1.0498	19.068	0.0373	0.0901
63	1	0.4952	0.6310	11.573	0.6496	0.0143

Table A16: Performance of SVM at Station 48615 (Kota Bharu)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	0.0369	0.0490	0.851	0.9977	-0.0004
2	5	0.0911	0.1199	2.097	0.9863	-0.0027
3	5	0.0555	0.0760	1.279	0.9945	-0.0035
4	5	0.0519	0.0701	1.196	0.9953	0.0010
5	5	0.0915	0.1304	2.106	0.9838	-0.0038
6	5	0.0967	0.1376	2.226	0.9820	-0.0056
7	5	0.3728	0.4740	8.587	0.7865	0.0308
8	4	0.0933	0.1246	2.149	0.9852	-0.0032
9	4	0.0959	0.1248	2.209	0.9851	-0.0031
10	4	0.1301	0.1729	2.997	0.9715	-0.0060
11	4	0.1292	0.1702	2.976	0.9725	-0.0077
12	4	0.4076	0.5274	9.388	0.7361	0.0413
13	4	0.0578	0.0791	1.331	0.9940	-0.0001
14	4	0.0962	0.1400	2.216	0.9813	-0.0025
15	4	0.1168	0.1740	2.690	0.9712	-0.0120
16	4	0.3964	0.5026	9.129	0.7597	0.0274
17	4	0.0934	0.1338	2.151	0.9829	-0.0041
18	4	0.1092	0.1513	2.514	0.9782	-0.0076
19	4	0.3834	0.4857	8.830	0.7758	0.0305
20	4	0.1597	0.2378	3.679	0.9474	-0.0358
21	4	0.4113	0.5265	9.473	0.7369	0.0331
22	4	0.4299	0.5522	9.901	0.7123	0.0577
23	3	0.1240	0.1642	2.855	0.9743	0.0023
24	3	0.1312	0.1758	3.022	0.9705	-0.0062
25	3	0.1416	0.1956	3.261	0.9638	-0.0164
26	3	0.4499	0.5827	10.361	0.6785	0.0493
27	3	0.1462	0.1904	3.366	0.9654	-0.0047
28	3	0.1361	0.1810	3.135	0.9689	-0.0103
29	3	0.4922	0.6334	11.336	0.6202	0.0544
30	3	0.1875	0.2630	4.318	0.9357	-0.0389
31	3	0.4474	0.5810	10.303	0.6802	0.0469
32	3	0.4588	0.5913	10.566	0.6701	0.0616
33	3	0.0967	0.1407	2.227	0.9811	-0.0012
34	3	0.1237	0.1863	2.850	0.9671	-0.0176
35	3	0.4014	0.5088	9.243	0.7539	0.0321
36	3	0.1679	0.2622	3.868	0.9367	-0.0460
37	3	0.4681	0.5903	10.780	0.6693	0.0235
38	3	0.4407	0.5667	10.150	0.6976	0.0625
39	3	0.1633	0.2440	3.762	0.9449	-0.0382
40	3	0.4443	0.5585	10.233	0.7034	0.0290
41	3	0.4374	0.5604	10.073	0.7038	0.0563
42	3	0.5114	0.6598	11.778	0.5880	0.0364
43	2	0.1761	0.2286	4.055	0.9501	-0.0063
44	2	0.1587	0.2167	3.655	0.9553	-0.0101
45	2	0.5043	0.6473	11.613	0.6035	0.0562
46	2	0.1906	0.2742	4.390	0.9304	-0.0434
47	2	0.5564	0.7160	12.814	0.5147	0.0540
48	2	0.4944	0.6402	11.386	0.6158	0.0783
49	2	0.1964	0.2696	4.523	0.9322	-0.0361
50	2	0.6935	0.9152	15.971	0.2330	0.1629
51	2	0.5334	0.6807	12.284	0.5634	0.0691
52	2	0.5311	0.6858	12.232	0.5548	0.0447
53	2	0.1721	0.2798	3.963	0.9290	-0.0549
54	2	0.4740	0.5956	10.917	0.6629	0.0189
55	2	0.4452	0.5702	10.252	0.6934	0.0574
56	2	0.5650	0.7111	13.012	0.5229	0.0561
57	2	0.5586	0.6994	12.866	0.5370	0.0548
58	1	0.2226	0.3128	5.127	0.9088	-0.0430
59	1	0.7103	0.9395	16.357	0.1838	0.1603
60	1	0.5362	0.6825	12.349	0.5604	0.0665
61	1	0.6299	0.8024	14.507	0.3925	0.0651
62	1	0.7798	1.0232	17.960	0.0374	0.1871
63	1	0.5742	0.7162	13.223	0.5150	0.0604

Table A17: Performance of SVM at Station 48620 (Sitiawan)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	0.0347	0.0600	0.912	0.9901	-0.0094
2	5	0.0514	0.0820	1.351	0.9855	0.0023
3	5	0.0376	0.0631	0.987	0.9898	-0.0032
4	5	0.0349	0.0602	0.916	0.9900	-0.0043
5	5	0.0409	0.0678	1.073	0.9887	-0.0016
6	5	0.0717	0.1080	1.882	0.9789	-0.0018
7	5	0.2959	0.3716	7.771	0.7826	0.0013
8	4	0.0519	0.0823	1.364	0.9856	0.0024
9	4	0.0550	0.0861	1.443	0.9843	0.0005
10	4	0.0580	0.0909	1.524	0.9831	0.0002
11	4	0.0768	0.1141	2.017	0.9766	-0.0016
12	4	0.3222	0.4044	8.462	0.7426	0.0069
13	4	0.0399	0.0659	1.048	0.9893	-0.0056
14	4	0.0417	0.0686	1.094	0.9886	-0.0004
15	4	0.0803	0.1164	2.110	0.9763	-0.0036
16	4	0.3161	0.3974	8.301	0.7514	0.0053
17	4	0.0410	0.0683	1.077	0.9885	-0.0027
18	4	0.0732	0.1092	1.923	0.9785	-0.0018
19	4	0.3053	0.3836	8.017	0.7683	0.0028
20	4	0.0830	0.1204	2.178	0.9745	-0.0044
21	4	0.2999	0.3771	7.877	0.7763	0.0014
22	4	0.3510	0.4380	9.217	0.6982	0.0036
23	3	0.0667	0.1010	1.752	0.9806	-0.0003
24	3	0.0579	0.0906	1.522	0.9833	-0.0004
25	3	0.0823	0.1193	2.162	0.9750	-0.0033
26	3	0.3631	0.4576	9.534	0.6711	0.0183
27	3	0.0722	0.1092	1.897	0.9773	-0.0024
28	3	0.0777	0.1153	2.041	0.9760	-0.0025
29	3	0.3756	0.4760	9.864	0.6444	0.0208
30	3	0.0879	0.1266	2.308	0.9719	-0.0062
31	3	0.3267	0.4095	8.580	0.7361	0.0066
32	3	0.3672	0.4591	9.642	0.6683	0.0114
33	3	0.0452	0.0730	1.187	0.9877	-0.0020
34	3	0.0865	0.1243	2.271	0.9736	-0.0017
35	3	0.3174	0.3995	8.335	0.7490	0.0067
36	3	0.0861	0.1231	2.262	0.9738	-0.0054
37	3	0.3423	0.4276	8.990	0.7123	0.0088
38	3	0.3624	0.4530	9.518	0.6772	0.0042
39	3	0.0832	0.1213	2.185	0.9740	-0.0041
40	3	0.3288	0.4144	8.633	0.7297	-0.0008
41	3	0.3598	0.4490	9.447	0.6830	0.0067
42	3	0.3665	0.4580	9.625	0.6703	-0.0020
43	2	0.0854	0.1256	2.243	0.9717	-0.0032
44	2	0.0960	0.1371	2.521	0.9681	-0.0015
45	2	0.3798	0.4805	9.974	0.6377	0.0204
46	2	0.0891	0.1274	2.340	0.9718	-0.0068
47	2	0.4087	0.5124	10.732	0.5873	0.0143
48	2	0.3931	0.4922	10.323	0.6194	0.0165
49	2	0.0954	0.1375	2.504	0.9673	-0.0083
50	2	0.5377	0.6834	14.120	0.2697	0.0607
51	2	0.4096	0.5155	10.757	0.5828	0.0230
52	2	0.3822	0.4780	10.036	0.6402	0.0081
53	2	0.0949	0.1350	2.492	0.9691	-0.0033
54	2	0.3462	0.4344	9.092	0.7031	0.0075
55	2	0.3634	0.4539	9.544	0.6761	0.0038
56	2	0.4050	0.5031	10.636	0.6021	0.0010
57	2	0.4128	0.5148	10.839	0.5827	-0.0011
58	1	0.1156	0.1613	3.035	0.9568	-0.0081
59	1	0.5454	0.6923	14.323	0.2510	0.0630
60	1	0.4104	0.5167	10.778	0.5808	0.0225
61	1	0.4535	0.5655	11.909	0.4975	0.0158
62	1	0.6323	0.7959	16.605	0.0144	0.0907
63	1	0.4178	0.5202	10.971	0.5740	0.0029

Table A18: Performance of SVM at Station 48623 (Lubok Merbau)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	0.0375	0.0462	0.921	0.9968	-0.0113
2	5	0.0576	0.0740	1.417	0.9914	0.0002
3	5	0.0403	0.0509	0.991	0.9959	-0.0043
4	5	0.0374	0.0460	0.919	0.9969	-0.0100
5	5	0.0404	0.0517	0.992	0.9958	-0.0049
6	5	0.0661	0.0877	1.624	0.9878	-0.0006
7	5	0.2674	0.3365	6.574	0.8208	-0.0074
8	4	0.0608	0.0781	1.496	0.9904	-0.0002
9	4	0.0615	0.0789	1.513	0.9902	-0.0014
10	4	0.0614	0.0800	1.510	0.9899	-0.0003
11	4	0.0764	0.0995	1.879	0.9844	-0.0023
12	4	0.3555	0.4459	8.741	0.6865	0.0118
13	4	0.0450	0.0571	1.106	0.9949	-0.0055
14	4	0.0425	0.0549	1.046	0.9953	-0.0031
15	4	0.0713	0.0945	1.753	0.9859	-0.0025
16	4	0.2803	0.3516	6.892	0.8042	-0.0054
17	4	0.0414	0.0530	1.018	0.9957	-0.0055
18	4	0.0670	0.0888	1.646	0.9875	-0.0006
19	4	0.2775	0.3487	6.823	0.8074	-0.0072
20	4	0.0769	0.1097	1.891	0.9810	-0.0053
21	4	0.2697	0.3390	6.631	0.8182	-0.0095
22	4	0.3083	0.3885	7.579	0.7617	-0.0100
23	3	0.0777	0.1014	1.909	0.9837	-0.0023
24	3	0.0634	0.0820	1.558	0.9894	-0.0002
25	3	0.0813	0.1053	1.999	0.9825	-0.0031
26	3	0.3777	0.4757	9.287	0.6429	0.0207
27	3	0.0746	0.0986	1.835	0.9847	-0.0023
28	3	0.0782	0.1019	1.922	0.9836	-0.0033
29	3	0.4278	0.5458	10.518	0.5307	0.0342
30	3	0.0878	0.1200	2.158	0.9774	-0.0086
31	3	0.3585	0.4492	8.815	0.6817	0.0097
32	3	0.3925	0.4886	9.651	0.6238	0.0217
33	3	0.0481	0.0624	1.183	0.9939	-0.0044
34	3	0.0745	0.0992	1.832	0.9844	-0.0016
35	3	0.2832	0.3548	6.963	0.8007	-0.0069
36	3	0.0776	0.1106	1.908	0.9807	-0.0065
37	3	0.2860	0.3580	7.032	0.7971	-0.0077
38	3	0.3178	0.4009	7.813	0.7461	-0.0049
39	3	0.0795	0.1138	1.954	0.9795	-0.0059
40	3	0.2835	0.3569	6.970	0.7982	-0.0036
41	3	0.3155	0.3986	7.756	0.7492	-0.0094
42	3	0.3152	0.3986	7.750	0.7492	-0.0113
43	2	0.0882	0.1164	2.169	0.9786	-0.0041
44	2	0.0926	0.1209	2.276	0.9769	-0.0050
45	2	0.4289	0.5468	10.545	0.5285	0.0307
46	2	0.0887	0.1213	2.180	0.9769	-0.0105
47	2	0.3884	0.4882	9.550	0.6235	0.0180
48	2	0.4113	0.5146	10.113	0.5827	0.0241
49	2	0.0994	0.1383	2.444	0.9703	-0.0149
50	2	0.5260	0.6734	12.934	0.2856	0.0464
51	2	0.4588	0.5796	11.281	0.4707	0.0329
52	2	0.3994	0.4964	9.819	0.6114	0.0158
53	2	0.0834	0.1186	2.050	0.9778	-0.0069
54	2	0.2908	0.3650	7.150	0.7891	-0.0037
55	2	0.3191	0.4029	7.845	0.7437	-0.0079
56	2	0.3380	0.4260	8.311	0.7131	-0.0047
57	2	0.3395	0.4309	8.348	0.7067	-0.0103
58	1	0.1103	0.1512	2.711	0.9644	-0.0151
59	1	0.5301	0.6786	13.032	0.2747	0.0471
60	1	0.4599	0.5810	11.306	0.4680	0.0310
61	1	0.4404	0.5488	10.827	0.5243	0.0116
62	1	0.6189	0.7868	15.216	0.0294	0.0555
63	1	0.3480	0.4400	8.557	0.6941	-0.0070

Table A19: Performance of SVM at Station 48625 (Ipoh)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	0.0345	0.0419	0.864	0.9972	-0.0096
2	5	0.0642	0.0822	1.607	0.9887	-0.0009
3	5	0.0395	0.0489	0.987	0.9960	-0.0029
4	5	0.0356	0.0437	0.890	0.9968	-0.0047
5	5	0.0455	0.0588	1.137	0.9942	-0.0014
6	5	0.0856	0.1126	2.142	0.9787	0.0011
7	5	0.2798	0.3511	7.001	0.7925	-0.0023
8	4	0.0661	0.0845	1.655	0.9880	-0.0001
9	4	0.0664	0.0850	1.662	0.9879	-0.0009
10	4	0.0709	0.0923	1.774	0.9857	-0.0010
11	4	0.1085	0.1384	2.715	0.9679	-0.0068
12	4	0.3284	0.4171	8.216	0.7084	0.0241
13	4	0.0427	0.0532	1.068	0.9953	-0.0040
14	4	0.0467	0.0601	1.169	0.9939	-0.0018
15	4	0.0991	0.1294	2.479	0.9718	0.0000
16	4	0.3027	0.3778	7.574	0.7599	-0.0109
17	4	0.0472	0.0610	1.181	0.9938	-0.0020
18	4	0.0995	0.1302	2.490	0.9715	-0.0013
19	4	0.2985	0.3730	7.469	0.7660	-0.0123
20	4	0.1155	0.1534	2.890	0.9605	-0.0065
21	4	0.2832	0.3551	7.085	0.7877	-0.0032
22	4	0.3214	0.4009	8.042	0.7296	-0.0024
23	3	0.0798	0.1018	1.996	0.9826	-0.0022
24	3	0.0715	0.0929	1.790	0.9855	-0.0015
25	3	0.1171	0.1488	2.929	0.9628	-0.0074
26	3	0.3731	0.4725	9.336	0.6259	0.0299
27	3	0.0851	0.1121	2.128	0.9789	-0.0030
28	3	0.1131	0.1438	2.830	0.9653	-0.0082
29	3	0.4098	0.5238	10.254	0.5413	0.0411
30	3	0.1345	0.1737	3.364	0.9496	-0.0147
31	3	0.3324	0.4216	8.316	0.7020	0.0255
32	3	0.3820	0.4766	9.557	0.6194	0.0308
33	3	0.0555	0.0707	1.388	0.9916	-0.0008
34	3	0.1022	0.1334	2.556	0.9700	-0.0003
35	3	0.3047	0.3800	7.625	0.7570	-0.0105
36	3	0.1174	0.1552	2.937	0.9595	-0.0063
37	3	0.3180	0.3969	7.957	0.7352	-0.0079
38	3	0.3324	0.4137	8.316	0.7119	-0.0063
39	3	0.1155	0.1536	2.891	0.9604	-0.0064
40	3	0.3291	0.4123	8.234	0.7140	-0.0048
41	3	0.3256	0.4058	8.147	0.7230	-0.0090
42	3	0.3380	0.4223	8.457	0.6999	0.0048
43	2	0.1031	0.1339	2.581	0.9700	-0.0059
44	2	0.1189	0.1511	2.975	0.9617	-0.0078
45	2	0.4097	0.5235	10.251	0.5421	0.0427
46	2	0.1347	0.1740	3.371	0.9494	-0.0146
47	2	0.3926	0.4961	9.823	0.5869	0.0253
48	2	0.4076	0.5125	10.198	0.5597	0.0324
49	2	0.1390	0.1800	3.477	0.9461	-0.0176
50	2	0.5478	0.7100	13.707	0.1579	0.0631
51	2	0.4277	0.5412	10.702	0.5105	0.0408
52	2	0.3973	0.4945	9.940	0.5898	0.0266
53	2	0.1185	0.1560	2.964	0.9591	-0.0046
54	2	0.3347	0.4181	8.374	0.7058	-0.0054
55	2	0.3338	0.4152	8.353	0.7098	-0.0063
56	2	0.3715	0.4628	9.295	0.6394	-0.0043
57	2	0.3741	0.4652	9.360	0.6358	-0.0051
58	1	0.1484	0.1917	3.714	0.9390	-0.0213
59	1	0.5546	0.7183	13.876	0.1370	0.0575
60	1	0.4281	0.5415	10.711	0.5098	0.0423
61	1	0.4590	0.5731	11.485	0.4481	0.0214
62	1	0.5972	0.7641	14.941	0.0227	0.0525
63	1	0.3812	0.4745	9.537	0.6212	-0.0105

Table A20: Performance of SVM at Station 48632 (Cameron Highlands)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	0.0416	0.0517	1.315	0.9979	-0.0117
2	5	0.0467	0.0599	1.478	0.9969	0.0067
3	5	0.0412	0.0518	1.304	0.9978	-0.0128
4	5	0.0391	0.0495	1.238	0.9979	-0.0068
5	5	0.0455	0.0619	1.440	0.9966	0.0016
6	5	0.0486	0.0664	1.535	0.9961	-0.0049
7	5	0.3515	0.4683	11.114	0.8081	-0.0252
8	4	0.0466	0.0598	1.473	0.9969	0.0051
9	4	0.0515	0.0666	1.628	0.9961	0.0022
10	4	0.0561	0.0758	1.775	0.9950	0.0017
11	4	0.0545	0.0734	1.722	0.9953	0.0043
12	4	0.3839	0.5042	12.138	0.7768	-0.0186
13	4	0.0437	0.0559	1.382	0.9975	-0.0151
14	4	0.0458	0.0622	1.448	0.9966	-0.0015
15	4	0.0507	0.0708	1.602	0.9956	0.0047
16	4	0.3702	0.4891	11.703	0.7901	-0.0223
17	4	0.0455	0.0621	1.439	0.9966	-0.0014
18	4	0.0482	0.0668	1.524	0.9962	0.0074
19	4	0.3610	0.4777	11.413	0.8000	-0.0250
20	4	0.0634	0.0962	2.004	0.9919	0.0103
21	4	0.3743	0.5024	11.833	0.7797	-0.0357
22	4	0.3807	0.4983	12.035	0.7824	-0.0306
23	3	0.1649	0.2404	5.215	0.9495	0.0106
24	3	0.0565	0.0765	1.787	0.9949	0.0008
25	3	0.0555	0.0753	1.756	0.9951	0.0055
26	3	0.4336	0.5599	13.708	0.7242	-0.0119
27	3	0.0650	0.0891	2.055	0.9930	0.0009
28	3	0.0573	0.0772	1.813	0.9948	0.0001
29	3	0.4698	0.5958	14.852	0.6876	-0.0129
30	3	0.0690	0.1039	2.182	0.9905	0.0067
31	3	0.4116	0.5475	13.012	0.7378	-0.0270
32	3	0.4074	0.5288	12.879	0.7545	-0.0234
33	3	0.0480	0.0651	1.518	0.9963	-0.0059
34	3	0.0542	0.0762	1.713	0.9949	-0.0024
35	3	0.3959	0.5321	12.518	0.7509	-0.0003
36	3	0.0637	0.0970	2.015	0.9918	0.0089
37	3	0.4197	0.5543	13.270	0.7312	-0.0384
38	3	0.3919	0.5124	12.389	0.7697	-0.0283
39	3	0.0634	0.0960	2.003	0.9919	0.0104
40	3	0.3973	0.5261	12.562	0.7586	-0.0430
41	3	0.3851	0.5028	12.175	0.7783	-0.0287
42	3	0.4404	0.5762	13.925	0.7112	-0.0619
43	2	0.2712	0.3728	8.573	0.8806	-0.0486
44	2	0.1682	0.2458	5.316	0.9473	0.0110
45	2	0.4837	0.6126	15.291	0.6695	-0.0144
46	2	0.0688	0.1036	2.174	0.9905	0.0059
47	2	0.5344	0.6951	16.897	0.5758	-0.0212
48	2	0.4433	0.5710	14.017	0.7131	-0.0142
49	2	0.0733	0.1103	2.316	0.9893	0.0019
50	2	0.6180	0.7829	19.538	0.4607	-0.0092
51	2	0.4763	0.6025	15.059	0.6804	-0.0111
52	2	0.4663	0.6111	14.744	0.6736	-0.0495
53	2	0.0647	0.0982	2.044	0.9915	0.0044
54	2	0.4531	0.6127	14.327	0.6702	-0.0093
55	2	0.4133	0.5478	13.066	0.7360	-0.0056
56	2	0.4826	0.6254	15.259	0.6590	-0.0630
57	2	0.4595	0.5940	14.528	0.6921	-0.0552
58	1	0.2807	0.3902	8.875	0.8706	-0.0651
59	1	0.7936	1.0119	25.092	0.1077	-0.1042
60	1	0.4872	0.6163	15.403	0.6655	-0.0163
61	1	0.5772	0.7447	18.248	0.5141	-0.0415
62	1	0.6449	0.8169	20.388	0.4137	-0.0247
63	1	0.5134	0.6759	16.233	0.5998	-0.0351

Table A21: Performance of SVM at Station 48647 (Subang)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	0.0401	0.0503	0.991	0.9977	-0.0131
2	5	0.0693	0.0876	1.712	0.9920	0.0022
3	5	0.0455	0.0583	1.124	0.9965	-0.0036
4	5	0.0410	0.0526	1.014	0.9971	-0.0006
5	5	0.0582	0.0762	1.438	0.9940	0.0012
6	5	0.1018	0.1355	2.516	0.9809	-0.0046
7	5	0.3760	0.4915	9.293	0.7481	-0.0225
8	4	0.0723	0.0913	1.786	0.9913	0.0013
9	4	0.0720	0.0910	1.780	0.9913	0.0016
10	4	0.0783	0.1010	1.936	0.9893	0.0004
11	4	0.1105	0.1443	2.731	0.9783	-0.0070
12	4	0.4716	0.5990	11.657	0.6260	-0.0229
13	4	0.0463	0.0594	1.145	0.9963	-0.0035
14	4	0.0603	0.0788	1.492	0.9935	0.0019
15	4	0.1050	0.1400	2.597	0.9796	-0.0058
16	4	0.3837	0.4993	9.484	0.7399	-0.0276
17	4	0.0587	0.0767	1.451	0.9939	0.0022
18	4	0.1026	0.1359	2.535	0.9807	-0.0048
19	4	0.3801	0.4955	9.396	0.7439	-0.0233
20	4	0.1151	0.1520	2.846	0.9760	-0.0096
21	4	0.4089	0.5305	10.108	0.7062	-0.0222
22	4	0.4246	0.5496	10.494	0.6853	-0.0315
23	3	0.0891	0.1123	2.202	0.9869	-0.0012
24	3	0.0792	0.1019	1.958	0.9892	0.0009
25	3	0.1125	0.1473	2.781	0.9774	-0.0072
26	3	0.4984	0.6277	12.319	0.5889	-0.0171
27	3	0.0896	0.1172	2.215	0.9857	-0.0006
28	3	0.1140	0.1476	2.817	0.9773	-0.0061
29	3	0.5299	0.6667	13.099	0.5357	-0.0105
30	3	0.1196	0.1563	2.955	0.9746	-0.0105
31	3	0.4872	0.6204	12.042	0.5982	-0.0208
32	3	0.5010	0.6338	12.384	0.5811	-0.0233
33	3	0.0656	0.0855	1.622	0.9924	0.0008
34	3	0.1109	0.1474	2.740	0.9773	-0.0036
35	3	0.3917	0.5069	9.682	0.7320	-0.0270
36	3	0.1152	0.1523	2.847	0.9759	-0.0094
37	3	0.4333	0.5566	10.710	0.6770	-0.0319
38	3	0.4324	0.5592	10.689	0.6742	-0.0358
39	3	0.1217	0.1597	3.008	0.9735	-0.0109
40	3	0.4266	0.5521	10.544	0.6816	-0.0220
41	3	0.4288	0.5545	10.599	0.6800	-0.0370
42	3	0.4605	0.5960	11.382	0.6302	-0.0436
43	2	0.1119	0.1444	2.766	0.9782	-0.0011
44	2	0.1305	0.1689	3.224	0.9702	-0.0047
45	2	0.5311	0.6685	13.126	0.5331	-0.0135
46	2	0.1197	0.1566	2.959	0.9745	-0.0096
47	2	0.5332	0.6724	13.180	0.5270	-0.0132
48	2	0.5267	0.6621	13.019	0.5424	-0.0143
49	2	0.1340	0.1733	3.313	0.9688	-0.0131
50	2	0.6676	0.8392	16.501	0.2628	0.0062
51	2	0.5634	0.7090	13.926	0.4748	-0.0112
52	2	0.5212	0.6606	12.883	0.5447	-0.0279
53	2	0.1319	0.1740	3.261	0.9686	-0.0141
54	2	0.4366	0.5609	10.791	0.6715	-0.0264
55	2	0.4346	0.5610	10.741	0.6717	-0.0302
56	2	0.4911	0.6294	12.139	0.5868	-0.0385
57	2	0.5080	0.6530	12.556	0.5553	-0.0400
58	1	0.1559	0.2016	3.853	0.9578	-0.0171
59	1	0.6774	0.8532	16.745	0.2383	0.0146
60	1	0.5654	0.7122	13.976	0.4700	-0.0098
61	1	0.5709	0.7183	14.110	0.4603	-0.0060
62	1	0.7620	0.9495	18.836	0.0563	0.0152
63	1	0.5118	0.6563	12.651	0.5505	-0.0389

Table A22: Performance of SVM at Station 48649 (Muadzam Shah)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	0.0330	0.0441	0.948	0.9972	-0.0051
2	5	0.0419	0.0561	1.203	0.9955	0.0032
3	5	0.0333	0.0445	0.957	0.9971	-0.0057
4	5	0.0334	0.0446	0.958	0.9971	-0.0058
5	5	0.0365	0.0512	1.050	0.9962	-0.0018
6	5	0.0595	0.0810	1.710	0.9907	0.0009
7	5	0.2661	0.3374	7.646	0.8401	0.0025
8	4	0.0424	0.0565	1.217	0.9955	0.0038
9	4	0.0468	0.0630	1.345	0.9943	-0.0006
10	4	0.0450	0.0619	1.293	0.9945	-0.0001
11	4	0.0604	0.0818	1.736	0.9906	0.0011
12	4	0.3365	0.4207	9.669	0.7514	-0.0012
13	4	0.0374	0.0504	1.075	0.9963	-0.0055
14	4	0.0372	0.0518	1.070	0.9961	-0.0043
15	4	0.0658	0.0890	1.892	0.9888	-0.0012
16	4	0.2796	0.3541	8.034	0.8240	0.0026
17	4	0.0376	0.0524	1.080	0.9960	-0.0038
18	4	0.0608	0.0826	1.747	0.9904	0.0007
19	4	0.2739	0.3472	7.872	0.8306	0.0035
20	4	0.0730	0.1000	2.098	0.9859	-0.0027
21	4	0.2774	0.3510	7.973	0.8269	0.0002
22	4	0.3220	0.4022	9.253	0.7730	0.0062
23	3	0.0577	0.0778	1.659	0.9914	0.0005
24	3	0.0454	0.0621	1.304	0.9945	0.0000
25	3	0.0664	0.0901	1.908	0.9885	-0.0011
26	3	0.3851	0.4825	11.065	0.6733	0.0079
27	3	0.0522	0.0710	1.501	0.9928	0.0000
28	3	0.0617	0.0838	1.773	0.9901	0.0010
29	3	0.4367	0.5524	12.550	0.5722	0.0301
30	3	0.0737	0.1021	2.119	0.9853	-0.0027
31	3	0.3433	0.4309	9.865	0.7391	0.0000
32	3	0.3659	0.4558	10.516	0.7082	0.0046
33	3	0.0417	0.0577	1.199	0.9952	-0.0022
34	3	0.0706	0.0954	2.029	0.9871	-0.0006
35	3	0.2831	0.3575	8.137	0.8206	0.0003
36	3	0.0761	0.1041	2.187	0.9848	-0.0045
37	3	0.3099	0.3919	8.905	0.7842	-0.0002
38	3	0.3270	0.4078	9.397	0.7666	0.0075
39	3	0.0738	0.1014	2.121	0.9856	-0.0038
40	3	0.3039	0.3839	8.732	0.7929	0.0058
41	3	0.3261	0.4071	9.370	0.7674	0.0039
42	3	0.3764	0.4752	10.817	0.6833	-0.0061
43	2	0.0620	0.0835	1.781	0.9901	-0.0001
44	2	0.0732	0.0978	2.105	0.9865	0.0001
45	2	0.4370	0.5527	12.559	0.5718	0.0286
46	2	0.0767	0.1073	2.204	0.9838	-0.0054
47	2	0.4380	0.5511	12.586	0.5737	0.0068
48	2	0.4003	0.4985	11.503	0.6512	0.0100
49	2	0.0753	0.1036	2.163	0.9849	-0.0024
50	2	0.6158	0.8039	17.697	0.1117	0.1166
51	2	0.4420	0.5582	12.701	0.5634	0.0287
52	2	0.3991	0.5025	11.470	0.6453	0.0019
53	2	0.0827	0.1128	2.375	0.9821	-0.0054
54	2	0.3172	0.4005	9.116	0.7746	0.0002
55	2	0.3295	0.4103	9.469	0.7637	0.0049
56	2	0.4151	0.5211	11.929	0.6194	-0.0117
57	2	0.4230	0.5288	12.155	0.6080	-0.0051
58	1	0.0848	0.1149	2.438	0.9814	-0.0065
59	1	0.6214	0.8095	17.856	0.1006	0.1219
60	1	0.4419	0.5585	12.699	0.5631	0.0302
61	1	0.4782	0.6025	13.741	0.4902	0.0041
62	1	0.6562	0.8436	18.857	0.0160	0.1076
63	1	0.4343	0.5402	12.481	0.5907	-0.0117

Table A23: Performance of SVM at Station 48650 (KLIA)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	0.0383	0.0492	0.947	0.9978	-0.0130
2	5	0.0758	0.0978	1.872	0.9903	0.0014
3	5	0.0472	0.0621	1.165	0.9961	-0.0042
4	5	0.0430	0.0559	1.062	0.9968	-0.0009
5	5	0.0729	0.1028	1.800	0.9893	-0.0027
6	5	0.0927	0.1350	2.290	0.9815	-0.0017
7	5	0.3849	0.4885	9.506	0.7586	-0.0310
8	4	0.0796	0.1029	1.966	0.9893	0.0010
9	4	0.0779	0.1004	1.923	0.9898	0.0004
10	4	0.1079	0.1505	2.664	0.9771	-0.0058
11	4	0.1158	0.1589	2.861	0.9744	-0.0062
12	4	0.4307	0.5416	10.637	0.7023	-0.0166
13	4	0.0483	0.0634	1.192	0.9959	-0.0041
14	4	0.0766	0.1081	1.891	0.9881	-0.0029
15	4	0.1040	0.1533	2.568	0.9763	-0.0061
16	4	0.3986	0.5039	9.844	0.7428	-0.0262
17	4	0.0733	0.1036	1.811	0.9891	-0.0025
18	4	0.1001	0.1461	2.472	0.9784	-0.0038
19	4	0.3942	0.4990	9.734	0.7482	-0.0330
20	4	0.1318	0.2169	3.254	0.9533	-0.0290
21	4	0.4060	0.5130	10.028	0.7336	-0.0298
22	4	0.4240	0.5369	10.470	0.7079	-0.0267
23	3	0.0897	0.1167	2.216	0.9862	-0.0020
24	3	0.1088	0.1513	2.688	0.9768	-0.0046
25	3	0.1219	0.1688	3.010	0.9712	-0.0098
26	3	0.4573	0.5759	11.293	0.6631	-0.0050
27	3	0.1238	0.1712	3.057	0.9704	-0.0076
28	3	0.1170	0.1614	2.891	0.9736	-0.0076
29	3	0.4913	0.6255	12.133	0.6024	0.0053
30	3	0.1551	0.2468	3.830	0.9403	-0.0395
31	3	0.4527	0.5679	11.181	0.6728	-0.0118
32	3	0.4670	0.5870	11.534	0.6502	-0.0162
33	3	0.0800	0.1104	1.977	0.9877	-0.0022
34	3	0.1048	0.1542	2.589	0.9760	-0.0057
35	3	0.3998	0.5051	9.875	0.7418	-0.0286
36	3	0.1353	0.2231	3.341	0.9507	-0.0308
37	3	0.4280	0.5387	10.571	0.7059	-0.0253
38	3	0.4346	0.5494	10.732	0.6939	-0.0265
39	3	0.1320	0.2178	3.259	0.9530	-0.0296
40	3	0.4376	0.5524	10.807	0.6916	-0.0357
41	3	0.4307	0.5452	10.636	0.6989	-0.0288
42	3	0.4659	0.5970	11.505	0.6415	-0.0515
43	2	0.1425	0.1939	3.519	0.9619	-0.0061
44	2	0.1275	0.1749	3.150	0.9691	-0.0105
45	2	0.4935	0.6288	12.188	0.5982	0.0111
46	2	0.1558	0.2480	3.848	0.9397	-0.0405
47	2	0.4971	0.6198	12.277	0.6099	-0.0050
48	2	0.4885	0.6141	12.064	0.6165	-0.0051
49	2	0.1679	0.2629	4.146	0.9325	-0.0446
50	2	0.6651	0.8329	16.426	0.2956	0.0228
51	2	0.5212	0.6593	12.873	0.5580	0.0097
52	2	0.5106	0.6500	12.611	0.5735	-0.0383
53	2	0.1389	0.2270	3.430	0.9491	-0.0333
54	2	0.4395	0.5545	10.855	0.6891	-0.0337
55	2	0.4355	0.5505	10.755	0.6929	-0.0292
56	2	0.4851	0.6184	11.979	0.6142	-0.0466
57	2	0.5014	0.6386	12.383	0.5886	-0.0521
58	1	0.1885	0.2885	4.656	0.9183	-0.0467
59	1	0.6834	0.8560	16.878	0.2559	0.0236
60	1	0.5227	0.6611	12.908	0.5556	0.0057
61	1	0.5503	0.6965	13.591	0.5084	-0.0209
62	1	0.7544	0.9502	18.631	0.0827	0.0083
63	1	0.5036	0.6407	12.436	0.5855	-0.0495

Table A24: Performance of SVM at Station 48657 (Kuantan)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	0.0347	0.0440	0.900	0.9978	-0.0088
2	5	0.0655	0.0851	1.701	0.9915	0.0022
3	5	0.0399	0.0508	1.035	0.9970	-0.0062
4	5	0.0367	0.0471	0.953	0.9974	-0.0046
5	5	0.0528	0.0699	1.372	0.9943	-0.0026
6	5	0.0684	0.0893	1.776	0.9906	-0.0014
7	5	0.2826	0.3660	7.337	0.8414	0.0071
8	4	0.0676	0.0875	1.755	0.9910	0.0020
9	4	0.0684	0.0887	1.775	0.9907	0.0017
10	4	0.0801	0.1039	2.079	0.9872	-0.0008
11	4	0.0839	0.1079	2.179	0.9863	-0.0020
12	4	0.3513	0.4513	9.120	0.7594	0.0127
13	4	0.0412	0.0526	1.069	0.9968	-0.0064
14	4	0.0539	0.0710	1.399	0.9941	-0.0040
15	4	0.0771	0.1003	2.002	0.9881	-0.0008
16	4	0.2994	0.3855	7.773	0.8242	0.0048
17	4	0.0529	0.0701	1.374	0.9942	-0.0032
18	4	0.0744	0.0967	1.931	0.9890	0.0009
19	4	0.2918	0.3763	7.577	0.8324	0.0045
20	4	0.0943	0.1246	2.449	0.9817	-0.0050
21	4	0.2947	0.3801	7.651	0.8291	0.0096
22	4	0.3414	0.4332	8.864	0.7778	0.0053
23	3	0.0875	0.1153	2.273	0.9843	0.0020
24	3	0.0803	0.1045	2.086	0.9871	-0.0007
25	3	0.0885	0.1134	2.297	0.9848	-0.0035
26	3	0.4026	0.5129	10.452	0.6901	0.0229
27	3	0.0913	0.1215	2.370	0.9825	-0.0018
28	3	0.0852	0.1092	2.211	0.9859	-0.0030
29	3	0.4497	0.5719	11.674	0.6145	0.0284
30	3	0.1039	0.1357	2.697	0.9783	-0.0080
31	3	0.3610	0.4643	9.371	0.7453	0.0159
32	3	0.3914	0.4971	10.161	0.7083	0.0231
33	3	0.0566	0.0738	1.471	0.9936	-0.0023
34	3	0.0789	0.1023	2.049	0.9876	-0.0001
35	3	0.3046	0.3914	7.908	0.8188	0.0063
36	3	0.0959	0.1259	2.489	0.9813	-0.0052
37	3	0.3333	0.4250	8.652	0.7862	0.0013
38	3	0.3538	0.4496	9.186	0.7607	0.0091
39	3	0.0945	0.1247	2.453	0.9816	-0.0052
40	3	0.3195	0.4093	8.294	0.8018	0.0061
41	3	0.3473	0.4400	9.017	0.7708	0.0042
42	3	0.3756	0.4764	9.751	0.7313	-0.0013
43	2	0.1105	0.1452	2.870	0.9750	-0.0025
44	2	0.0975	0.1257	2.532	0.9813	-0.0014
45	2	0.4549	0.5766	11.810	0.6083	0.0304
46	2	0.1042	0.1356	2.706	0.9783	-0.0077
47	2	0.4709	0.5988	12.225	0.5770	0.0152
48	2	0.4351	0.5519	11.295	0.6407	0.0279
49	2	0.1071	0.1414	2.779	0.9764	-0.0089
50	2	0.6730	0.8695	17.472	0.1259	0.1311
51	2	0.4717	0.5962	12.247	0.5807	0.0313
52	2	0.4193	0.5324	10.886	0.6650	0.0216
53	2	0.0975	0.1281	2.530	0.9806	-0.0066
54	2	0.3395	0.4322	8.814	0.7789	0.0024
55	2	0.3617	0.4590	9.389	0.7507	0.0110
56	2	0.4295	0.5431	11.150	0.6507	-0.0023
57	2	0.4053	0.5109	10.521	0.6911	0.0065
58	1	0.1201	0.1567	3.117	0.9710	-0.0077
59	1	0.6806	0.8806	17.669	0.1038	0.1342
60	1	0.4729	0.5963	12.276	0.5805	0.0284
61	1	0.5410	0.6902	14.045	0.4375	0.0199
62	1	0.7206	0.9180	18.708	0.0198	0.1192
63	1	0.4368	0.5519	11.340	0.6398	-0.0031

Table A25: Performance of ANFIS at Station 48600 (Pulau Langkawi)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	0.0491	0.0714	1.122	0.9960	0.0000
2	5	0.0871	0.1238	1.989	0.9879	0.0000
3	5	0.0644	0.0900	1.472	0.9936	0.0000
4	5	0.0533	0.0762	1.218	0.9954	0.0000
5	5	0.1148	0.1702	2.623	0.9772	0.0003
6	5	0.1363	0.2060	3.112	0.9667	-0.0001
7	5	0.4070	0.5136	9.296	0.7926	0.0005
8	4	0.0902	0.1300	2.060	0.9867	-0.0001
9	4	0.0925	0.1307	2.113	0.9866	0.0000
10	4	0.1522	0.2264	3.477	0.9597	0.0000
11	4	0.1702	0.2442	3.886	0.9532	-0.0001
12	4	0.4403	0.5604	10.055	0.7532	0.0006
13	4	0.0679	0.0953	1.550	0.9929	0.0000
14	4	0.1237	0.1833	2.826	0.9736	-0.0001
15	4	0.1492	0.2254	3.408	0.9602	0.0000
16	4	0.4256	0.5341	9.719	0.7757	0.0001
17	4	0.1210	0.1752	2.763	0.9759	-0.0001
18	4	0.1437	0.2167	3.281	0.9632	-0.0001
19	4	0.4368	0.5458	9.976	0.7658	0.0000
20	4	0.2469	0.3625	5.639	0.8967	-0.0002
21	4	0.4518	0.5781	10.318	0.7376	-0.0004
22	4	0.4263	0.5350	9.736	0.7750	0.0004
23	3	0.1154	0.1583	2.635	0.9803	-0.0001
24	3	0.1545	0.2305	3.528	0.9583	0.0000
25	3	0.1737	0.2510	3.967	0.9506	-0.0001
26	3	0.4713	0.5928	10.763	0.7240	0.0001
27	3	0.1660	0.2438	3.791	0.9534	0.0001
28	3	0.1722	0.2463	3.933	0.9524	-0.0001
29	3	0.5429	0.6815	12.399	0.6351	-0.0001
30	3	0.3239	0.4536	7.396	0.8384	-0.0001
31	3	0.4701	0.6113	10.736	0.7066	0.0008
32	3	0.4588	0.5774	10.477	0.7380	0.0001
33	3	0.1277	0.1837	2.917	0.9735	-0.0003
34	3	0.1498	0.2263	3.422	0.9599	-0.0001
35	3	0.4384	0.5473	10.011	0.7645	0.0001
36	3	0.2520	0.3698	5.755	0.8925	-0.0003
37	3	0.4644	0.5874	10.605	0.7297	0.0009
38	3	0.4429	0.5514	10.114	0.7610	0.0000
39	3	0.2515	0.3670	5.744	0.8941	-0.0003
40	3	0.4904	0.6123	11.199	0.7055	0.0004
41	3	0.4488	0.5577	10.250	0.7556	0.0003
42	3	0.4987	0.6460	11.390	0.6741	0.0026
43	2	0.1903	0.2708	4.346	0.9425	0.0000
44	2	0.1875	0.2704	4.283	0.9426	0.0000
45	2	0.5484	0.6881	12.523	0.6278	0.0006
46	2	0.3288	0.4598	7.509	0.8339	-0.0001
47	2	0.5293	0.6711	12.087	0.6460	0.0000
48	2	0.4849	0.6051	11.073	0.7123	-0.0001
49	2	0.3546	0.4890	8.098	0.8121	-0.0002
50	2	0.7436	0.9305	16.982	0.3197	-0.0002
51	2	0.5540	0.6910	12.653	0.6249	0.0001
52	2	0.5538	0.7219	12.648	0.5910	0.0008
53	2	0.2564	0.3740	5.856	0.8900	0.0000
54	2	0.4935	0.6150	11.270	0.7029	0.0000
55	2	0.4537	0.5619	10.360	0.7518	0.0002
56	2	0.5191	0.6558	11.855	0.6621	0.0001
57	2	0.5442	0.6797	12.427	0.6370	-0.0003
58	1	0.3808	0.5196	8.697	0.7878	-0.0001
59	1	0.7732	0.9768	17.657	0.2508	-0.0002
60	1	0.5602	0.6989	12.792	0.6162	0.0001
61	1	0.6038	0.7758	13.789	0.5269	-0.0001
62	1	0.8443	1.0622	19.281	0.1149	-0.0003
63	1	0.5441	0.6795	12.425	0.6373	0.0000

Table A26: Performance of ANFIS at Station 48601 (Bayan Lepas)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	0.0466	0.0659	1.075	0.9959	0.0000
2	5	0.0895	0.1200	2.063	0.9865	0.0000
3	5	0.0587	0.0785	1.353	0.9942	0.0000
4	5	0.0527	0.0702	1.214	0.9954	0.0000
5	5	0.1152	0.1610	2.656	0.9756	0.0000
6	5	0.1089	0.1540	2.510	0.9778	-0.0001
7	5	0.3844	0.4957	8.860	0.7689	-0.0001
8	4	0.0936	0.1257	2.158	0.9852	-0.0001
9	4	0.0933	0.1282	2.151	0.9845	-0.0001
10	4	0.1449	0.2029	3.339	0.9614	0.0000
11	4	0.1344	0.1816	3.099	0.9691	-0.0003
12	4	0.4230	0.5442	9.751	0.7220	0.0010
13	4	0.0644	0.0867	1.485	0.9930	-0.0002
14	4	0.1182	0.1655	2.724	0.9743	0.0000
15	4	0.1251	0.1756	2.883	0.9711	-0.0001
16	4	0.4044	0.5166	9.322	0.7490	0.0003
17	4	0.1109	0.1572	2.557	0.9768	-0.0002
18	4	0.1266	0.1773	2.919	0.9706	-0.0003
19	4	0.3983	0.5060	9.182	0.7592	-0.0001
20	4	0.2162	0.3382	4.982	0.8922	0.0000
21	4	0.4073	0.5281	9.389	0.7377	0.0002
22	4	0.4363	0.5560	10.057	0.7093	-0.0004
23	3	0.1096	0.1450	2.525	0.9803	-0.0001
24	3	0.1473	0.2059	3.395	0.9602	0.0000
25	3	0.1400	0.1895	3.227	0.9664	-0.0004
26	3	0.4769	0.6022	10.991	0.6593	0.0007
27	3	0.1658	0.2284	3.821	0.9511	0.0002
28	3	0.1384	0.1878	3.189	0.9670	0.0000
29	3	0.5107	0.6473	11.772	0.6061	0.0004
30	3	0.2394	0.3619	5.518	0.8767	0.0000
31	3	0.4368	0.5668	10.067	0.6981	-0.0002
32	3	0.4647	0.5911	10.712	0.6716	0.0002
33	3	0.1183	0.1636	2.726	0.9748	-0.0001
34	3	0.1343	0.1875	3.095	0.9671	-0.0001
35	3	0.4004	0.5116	9.230	0.7539	0.0000
36	3	0.2206	0.3441	5.085	0.8884	-0.0001
37	3	0.4402	0.5638	10.146	0.7008	0.0004
38	3	0.4487	0.5696	10.343	0.6946	0.0000
39	3	0.2165	0.3403	4.991	0.8909	0.0000
40	3	0.4345	0.5544	10.015	0.7108	-0.0003
41	3	0.4396	0.5577	10.132	0.7074	-0.0001
42	3	0.5217	0.6798	12.026	0.5653	0.0002
43	2	0.1820	0.2439	4.195	0.9442	0.0001
44	2	0.1515	0.2034	3.492	0.9613	-0.0001
45	2	0.5144	0.6502	11.858	0.6024	0.0001
46	2	0.2404	0.3628	5.542	0.8760	0.0000
47	2	0.5153	0.6509	11.877	0.6015	0.0001
48	2	0.5049	0.6365	11.637	0.6193	0.0002
49	2	0.2541	0.3759	5.857	0.8669	0.0000
50	2	0.6692	0.8465	15.425	0.3262	0.0002
51	2	0.5510	0.6910	12.700	0.5514	0.0002
52	2	0.5459	0.7147	12.583	0.5196	-0.0003
53	2	0.2255	0.3495	5.198	0.8849	-0.0001
54	2	0.4436	0.5648	10.224	0.6998	0.0000
55	2	0.4473	0.5686	10.311	0.6958	-0.0001
56	2	0.5459	0.7039	12.583	0.5337	-0.0001
57	2	0.5602	0.7138	12.912	0.5204	-0.0002
58	1	0.2734	0.3967	6.301	0.8517	0.0000
59	1	0.6720	0.8558	15.490	0.3112	0.0001
60	1	0.5530	0.6924	12.747	0.5493	0.0000
61	1	0.6117	0.7855	14.099	0.4196	-0.0001
62	1	0.7973	1.0096	18.378	0.0435	0.0001
63	1	0.5617	0.7181	12.947	0.5147	0.0000

Table A27: Performance of ANFIS at Station 48603 (Alor Setar)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	0.0438	0.0586	1.024	0.9970	0.0000
2	5	0.0916	0.1202	2.140	0.9873	0.0001
3	5	0.0563	0.0738	1.315	0.9952	0.0001
4	5	0.0507	0.0672	1.184	0.9960	0.0000
5	5	0.0880	0.1226	2.057	0.9868	-0.0004
6	5	0.0947	0.1287	2.214	0.9854	-0.0001
7	5	0.3187	0.4109	7.447	0.8516	0.0007
8	4	0.0942	0.1229	2.201	0.9867	0.0002
9	4	0.0990	0.1319	2.314	0.9847	0.0000
10	4	0.1301	0.1749	3.041	0.9731	0.0000
11	4	0.1167	0.1547	2.727	0.9790	-0.0001
12	4	0.3743	0.4776	8.747	0.7994	-0.0003
13	4	0.0580	0.0764	1.355	0.9949	0.0000
14	4	0.0882	0.1221	2.061	0.9869	-0.0004
15	4	0.1129	0.1542	2.638	0.9791	-0.0002
16	4	0.3434	0.4407	8.024	0.8290	0.0000
17	4	0.0879	0.1206	2.054	0.9872	-0.0001
18	4	0.1035	0.1399	2.420	0.9828	0.0000
19	4	0.3521	0.4477	8.229	0.8237	0.0001
20	4	0.1861	0.2823	4.349	0.9299	0.0000
21	4	0.3240	0.4180	7.571	0.8464	-0.0002
22	4	0.3882	0.4969	9.072	0.7828	0.0008
23	3	0.1140	0.1516	2.664	0.9798	0.0001
24	3	0.1464	0.1952	3.421	0.9665	-0.0001
25	3	0.1244	0.1648	2.907	0.9761	-0.0001
26	3	0.4229	0.5398	9.883	0.7437	0.0000
27	3	0.1684	0.2337	3.935	0.9519	-0.0001
28	3	0.1192	0.1578	2.786	0.9781	0.0000
29	3	0.4990	0.6358	11.661	0.6444	-0.0004
30	3	0.2142	0.3055	5.005	0.9180	-0.0001
31	3	0.3710	0.4767	8.669	0.8004	-0.0010
32	3	0.4262	0.5405	9.959	0.7431	0.0010
33	3	0.0817	0.1139	1.908	0.9886	0.0000
34	3	0.1167	0.1601	2.726	0.9775	0.0000
35	3	0.3582	0.4535	8.371	0.8190	0.0000
36	3	0.1894	0.2851	4.425	0.9284	0.0000
37	3	0.3560	0.4579	8.320	0.8154	-0.0006
38	3	0.4015	0.5113	9.382	0.7700	0.0001
39	3	0.1879	0.2859	4.390	0.9281	0.0001
40	3	0.3663	0.4689	8.559	0.8066	0.0001
41	3	0.4055	0.5135	9.477	0.7681	-0.0001
42	3	0.4234	0.5549	9.894	0.7291	0.0002
43	2	0.1717	0.2367	4.013	0.9507	0.0000
44	2	0.1379	0.1834	3.224	0.9704	0.0000
45	2	0.5096	0.6446	11.909	0.6343	-0.0002
46	2	0.2421	0.3338	5.657	0.9020	0.0000
47	2	0.4670	0.5953	10.914	0.6883	0.0001
48	2	0.4537	0.5790	10.602	0.7051	0.0002
49	2	0.2528	0.3555	5.907	0.8888	0.0000
50	2	0.6903	0.8750	16.131	0.3267	0.0000
51	2	0.5328	0.6723	12.452	0.6024	0.0001
52	2	0.4568	0.5931	10.675	0.6905	0.0000
53	2	0.1891	0.2879	4.419	0.9270	0.0000
54	2	0.3694	0.4729	8.632	0.8031	0.0000
55	2	0.4157	0.5239	9.715	0.7585	0.0001
56	2	0.4654	0.6022	10.875	0.6807	0.0000
57	2	0.4900	0.6244	11.451	0.6568	-0.0001
58	1	0.2590	0.3621	6.054	0.8847	0.0000
59	1	0.7012	0.8878	16.387	0.3066	0.0001
60	1	0.5382	0.6774	12.578	0.5964	0.0001
61	1	0.5639	0.7299	13.179	0.5310	0.0000
62	1	0.8172	1.0434	19.098	0.0418	-0.0001
63	1	0.4941	0.6297	11.546	0.6510	0.0000

Table A28: Performance of ANFIS at Station 48615 (Kota Bharu)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	0.0457	0.0649	1.052	0.9960	0.0001
2	5	0.0967	0.1266	2.226	0.9847	0.0000
3	5	0.0634	0.0882	1.460	0.9926	0.0000
4	5	0.0596	0.0830	1.372	0.9934	0.0000
5	5	0.0939	0.1338	2.162	0.9828	0.0001
6	5	0.1130	0.1572	2.601	0.9764	-0.0001
7	5	0.3895	0.4908	8.971	0.7701	0.0004
8	4	0.0973	0.1289	2.241	0.9841	0.0001
9	4	0.0999	0.1291	2.300	0.9841	0.0000
10	4	0.1305	0.1739	3.005	0.9711	0.0000
11	4	0.1367	0.1788	3.148	0.9695	-0.0001
12	4	0.4137	0.5313	9.528	0.7310	0.0001
13	4	0.0645	0.0894	1.486	0.9924	0.0000
14	4	0.0972	0.1407	2.238	0.9810	0.0000
15	4	0.1346	0.1917	3.099	0.9649	0.0000
16	4	0.4027	0.5079	9.274	0.7537	0.0002
17	4	0.0952	0.1364	2.191	0.9822	0.0000
18	4	0.1240	0.1698	2.855	0.9725	0.0001
19	4	0.3766	0.4771	8.672	0.7827	-0.0004
20	4	0.1683	0.2361	3.876	0.9468	0.0002
21	4	0.4299	0.5432	9.901	0.7190	0.0002
22	4	0.4346	0.5507	10.009	0.7104	0.0005
23	3	0.1257	0.1652	2.895	0.9739	0.0001
24	3	0.1309	0.1756	3.015	0.9705	0.0000
25	3	0.1463	0.1983	3.370	0.9625	0.0000
26	3	0.4375	0.5610	10.075	0.6994	0.0003
27	3	0.1463	0.1906	3.369	0.9653	0.0000
28	3	0.1379	0.1817	3.177	0.9685	0.0000
29	3	0.4874	0.6238	11.226	0.6280	0.0012
30	3	0.1942	0.2580	4.473	0.9365	-0.0001
31	3	0.4367	0.5642	10.057	0.6963	-0.0004
32	3	0.4619	0.5884	10.638	0.6697	-0.0015
33	3	0.0975	0.1407	2.244	0.9810	0.0000
34	3	0.1311	0.1896	3.018	0.9657	0.0001
35	3	0.3958	0.5003	9.115	0.7610	-0.0010
36	3	0.1782	0.2541	4.103	0.9384	0.0000
37	3	0.4699	0.5871	10.822	0.6715	-0.0001
38	3	0.4443	0.5604	10.232	0.7001	-0.0004
39	3	0.1699	0.2373	3.913	0.9463	0.0000
40	3	0.4577	0.5706	10.541	0.6894	0.0001
41	3	0.4411	0.5563	10.160	0.7044	-0.0002
42	3	0.5087	0.6488	11.715	0.5985	-0.0001
43	2	0.1742	0.2263	4.011	0.9511	0.0000
44	2	0.1612	0.2171	3.712	0.9550	0.0000
45	2	0.5047	0.6385	11.624	0.6104	-0.0025
46	2	0.1964	0.2697	4.523	0.9306	0.0000
47	2	0.5485	0.6991	12.631	0.5343	0.0011
48	2	0.4978	0.6306	11.466	0.6203	-0.0001
49	2	0.2009	0.2646	4.627	0.9332	0.0000
50	2	0.6747	0.8672	15.539	0.2825	0.0001
51	2	0.5356	0.6730	12.334	0.5675	-0.0002
52	2	0.5170	0.6651	11.907	0.5783	-0.0005
53	2	0.1811	0.2696	4.171	0.9306	0.0000
54	2	0.4700	0.5902	10.824	0.6680	-0.0001
55	2	0.4461	0.5628	10.273	0.6974	-0.0001
56	2	0.5664	0.7029	13.045	0.5288	0.0000
57	2	0.5575	0.6917	12.840	0.5434	-0.0009
58	1	0.2260	0.3077	5.205	0.9096	0.0000
59	1	0.6996	0.9001	16.111	0.2268	-0.0002
60	1	0.5405	0.6780	12.447	0.5610	0.0001
61	1	0.6343	0.7962	14.608	0.3954	-0.0002
62	1	0.7937	1.0039	18.280	0.0388	-0.0001
63	1	0.5750	0.7092	13.243	0.5204	-0.0001

Table A29: Performance of ANFIS at Station 48620 (Sitiawan)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	0.0376	0.0676	0.986	0.9874	-0.0001
2	5	0.0539	0.0876	1.414	0.9829	-0.0002
3	5	0.0422	0.0716	1.109	0.9869	-0.0001
4	5	0.0388	0.0685	1.020	0.9872	0.0000
5	5	0.0450	0.0768	1.183	0.9857	-0.0001
6	5	0.0736	0.1129	1.932	0.9760	-0.0001
7	5	0.2986	0.3742	7.841	0.7797	0.0004
8	4	0.0534	0.0868	1.401	0.9831	-0.0001
9	4	0.0570	0.0909	1.498	0.9820	-0.0003
10	4	0.0606	0.0967	1.593	0.9805	-0.0004
11	4	0.0786	0.1184	2.065	0.9739	-0.0002
12	4	0.3230	0.4052	8.483	0.7416	-0.0001
13	4	0.0445	0.0748	1.170	0.9862	0.0000
14	4	0.0450	0.0753	1.183	0.9860	0.0000
15	4	0.0819	0.1208	2.150	0.9737	-0.0001
16	4	0.3198	0.4007	8.399	0.7472	0.0004
17	4	0.0454	0.0774	1.192	0.9854	0.0000
18	4	0.0753	0.1141	1.977	0.9757	0.0000
19	4	0.3024	0.3804	7.942	0.7720	-0.0006
20	4	0.0852	0.1253	2.236	0.9716	-0.0002
21	4	0.3031	0.3795	7.959	0.7735	0.0001
22	4	0.3504	0.4367	9.201	0.6997	0.0002
23	3	0.0690	0.1059	1.813	0.9781	-0.0001
24	3	0.0582	0.0934	1.529	0.9815	-0.0002
25	3	0.0834	0.1228	2.189	0.9726	-0.0001
26	3	0.3634	0.4566	9.542	0.6718	0.0004
27	3	0.0724	0.1108	1.901	0.9762	-0.0001
28	3	0.0794	0.1190	2.084	0.9738	-0.0002
29	3	0.3741	0.4725	9.824	0.6485	-0.0003
30	3	0.0897	0.1308	2.356	0.9692	-0.0002
31	3	0.3239	0.4066	8.506	0.7396	0.0000
32	3	0.3682	0.4595	9.669	0.6672	0.0004
33	3	0.0485	0.0798	1.274	0.9850	0.0000
34	3	0.0885	0.1295	2.325	0.9705	0.0000
35	3	0.3173	0.3987	8.333	0.7497	0.0003
36	3	0.0875	0.1271	2.298	0.9712	-0.0001
37	3	0.3435	0.4280	9.021	0.7116	0.0002
38	3	0.3629	0.4525	9.530	0.6775	0.0001
39	3	0.0863	0.1274	2.266	0.9708	0.0000
40	3	0.3269	0.4128	8.585	0.7315	-0.0004
41	3	0.3581	0.4454	9.403	0.6875	-0.0002
42	3	0.3655	0.4562	9.598	0.6724	0.0001
43	2	0.0856	0.1274	2.248	0.9705	0.0000
44	2	0.0977	0.1409	2.566	0.9658	-0.0001
45	2	0.3788	0.4776	9.948	0.6408	-0.0002
46	2	0.0899	0.1300	2.360	0.9698	-0.0001
47	2	0.4070	0.5091	10.688	0.5917	0.0000
48	2	0.3938	0.4908	10.342	0.6204	0.0001
49	2	0.0965	0.1402	2.535	0.9653	-0.0001
50	2	0.5390	0.6789	14.154	0.2739	0.0000
51	2	0.4091	0.5122	10.744	0.5867	0.0000
52	2	0.3809	0.4755	10.003	0.6438	-0.0002
53	2	0.0974	0.1407	2.557	0.9659	0.0000
54	2	0.3445	0.4323	9.046	0.7057	-0.0001
55	2	0.3625	0.4520	9.519	0.6782	0.0000
56	2	0.4050	0.5023	10.637	0.6028	-0.0002
57	2	0.4110	0.5133	10.792	0.5851	-0.0001
58	1	0.1173	0.1646	3.080	0.9545	-0.0001
59	1	0.5472	0.6882	14.371	0.2537	0.0001
60	1	0.4110	0.5151	10.792	0.5821	0.0000
61	1	0.4536	0.5636	11.911	0.4995	-0.0001
62	1	0.6365	0.7901	16.714	0.0155	0.0000
63	1	0.4166	0.5193	10.941	0.5754	-0.0001

Table A30: Performance of ANFIS at Station 48623 (Lubok Merbau)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	0.0343	0.0425	0.842	0.9971	0.0000
2	5	0.0562	0.0721	1.382	0.9918	0.0001
3	5	0.0404	0.0512	0.993	0.9958	0.0001
4	5	0.0345	0.0428	0.849	0.9971	0.0000
5	5	0.0401	0.0517	0.985	0.9958	0.0000
6	5	0.0671	0.0893	1.651	0.9874	-0.0001
7	5	0.2717	0.3407	6.679	0.8162	0.0002
8	4	0.0604	0.0776	1.484	0.9905	-0.0002
9	4	0.0602	0.0770	1.480	0.9906	0.0000
10	4	0.0603	0.0785	1.482	0.9903	-0.0001
11	4	0.0770	0.1000	1.893	0.9842	-0.0001
12	4	0.3550	0.4444	8.729	0.6882	-0.0001
13	4	0.0449	0.0577	1.103	0.9947	-0.0001
14	4	0.0425	0.0547	1.045	0.9953	0.0000
15	4	0.0724	0.0961	1.781	0.9854	0.0000
16	4	0.2833	0.3547	6.966	0.8007	-0.0003
17	4	0.0417	0.0545	1.025	0.9953	0.0000
18	4	0.0682	0.0905	1.676	0.9870	0.0000
19	4	0.2751	0.3451	6.764	0.8113	-0.0002
20	4	0.0778	0.1097	1.913	0.9809	0.0000
21	4	0.2695	0.3384	6.627	0.8189	-0.0002
22	4	0.3099	0.3900	7.619	0.7596	-0.0001
23	3	0.0767	0.1001	1.886	0.9841	0.0000
24	3	0.0621	0.0804	1.526	0.9898	0.0000
25	3	0.0814	0.1056	2.002	0.9824	0.0000
26	3	0.3765	0.4736	9.258	0.6454	0.0013
27	3	0.0718	0.0955	1.766	0.9856	0.0002
28	3	0.0775	0.1011	1.906	0.9838	-0.0002
29	3	0.4242	0.5380	10.431	0.5420	0.0004
30	3	0.0879	0.1187	2.161	0.9777	0.0000
31	3	0.3510	0.4405	8.631	0.6939	0.0003
32	3	0.3884	0.4836	9.550	0.6310	-0.0001
33	3	0.0471	0.0610	1.159	0.9941	0.0000
34	3	0.0750	0.0989	1.844	0.9845	0.0000
35	3	0.2795	0.3506	6.872	0.8053	-0.0003
36	3	0.0778	0.1097	1.912	0.9809	0.0000
37	3	0.2845	0.3559	6.995	0.7994	0.0001
38	3	0.3177	0.4001	7.812	0.7467	-0.0004
39	3	0.0800	0.1134	1.966	0.9796	0.0000
40	3	0.2832	0.3564	6.963	0.7989	0.0002
41	3	0.3132	0.3939	7.700	0.7546	-0.0002
42	3	0.3143	0.3964	7.728	0.7515	0.0002
43	2	0.0863	0.1134	2.122	0.9796	0.0000
44	2	0.0920	0.1202	2.262	0.9771	0.0000
45	2	0.4252	0.5397	10.454	0.5391	0.0000
46	2	0.0889	0.1199	2.185	0.9773	0.0000
47	2	0.3852	0.4841	9.470	0.6296	0.0001
48	2	0.4098	0.5112	10.075	0.5870	0.0000
49	2	0.0990	0.1341	2.434	0.9716	0.0000
50	2	0.5244	0.6685	12.892	0.2924	-0.0002
51	2	0.4553	0.5733	11.194	0.4802	0.0001
52	2	0.3935	0.4892	9.676	0.6224	0.0000
53	2	0.0836	0.1179	2.056	0.9779	0.0000
54	2	0.2900	0.3641	7.131	0.7900	0.0000
55	2	0.3167	0.3986	7.787	0.7485	0.0000
56	2	0.3369	0.4236	8.283	0.7162	-0.0002
57	2	0.3396	0.4298	8.350	0.7076	-0.0001
58	1	0.1100	0.1477	2.705	0.9655	0.0000
59	1	0.5297	0.6755	13.025	0.2778	0.0000
60	1	0.4572	0.5759	11.241	0.4755	0.0001
61	1	0.4379	0.5462	10.767	0.5284	0.0000
62	1	0.6181	0.7810	15.196	0.0348	0.0001
63	1	0.3484	0.4392	8.567	0.6947	0.0000

Table A31: Performance of ANFIS at Station 48625 (Ipoh)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	0.0359	0.0454	0.899	0.9965	-0.0001
2	5	0.0648	0.0828	1.622	0.9885	0.0000
3	5	0.0430	0.0541	1.075	0.9951	-0.0001
4	5	0.0373	0.0464	0.933	0.9964	0.0000
5	5	0.0462	0.0594	1.157	0.9941	0.0000
6	5	0.0862	0.1127	2.156	0.9786	0.0000
7	5	0.2839	0.3552	7.104	0.7876	0.0002
8	4	0.0662	0.0848	1.657	0.9879	0.0000
9	4	0.0655	0.0837	1.638	0.9882	-0.0001
10	4	0.0698	0.0910	1.747	0.9861	0.0000
11	4	0.1081	0.1376	2.704	0.9682	-0.0001
12	4	0.3334	0.4214	8.342	0.7010	-0.0002
13	4	0.0445	0.0561	1.113	0.9947	0.0000
14	4	0.0471	0.0603	1.179	0.9939	0.0002
15	4	0.0994	0.1293	2.488	0.9718	0.0000
16	4	0.3066	0.3817	7.671	0.7545	0.0001
17	4	0.0468	0.0602	1.170	0.9939	-0.0001
18	4	0.0996	0.1300	2.493	0.9715	0.0000
19	4	0.2971	0.3719	7.435	0.7671	-0.0009
20	4	0.1156	0.1523	2.892	0.9609	0.0000
21	4	0.2818	0.3533	7.050	0.7899	0.0000
22	4	0.3248	0.4043	8.127	0.7247	0.0000
23	3	0.0791	0.1009	1.980	0.9829	-0.0001
24	3	0.0708	0.0918	1.770	0.9858	0.0000
25	3	0.1163	0.1478	2.911	0.9632	-0.0001
26	3	0.3769	0.4745	9.430	0.6208	0.0001
27	3	0.0828	0.1090	2.072	0.9800	-0.0001
28	3	0.1123	0.1425	2.811	0.9659	0.0000
29	3	0.4023	0.5122	10.066	0.5585	-0.0002
30	3	0.1346	0.1722	3.368	0.9501	0.0001
31	3	0.3284	0.4160	8.217	0.7088	-0.0004
32	3	0.3831	0.4757	9.585	0.6189	-0.0001
33	3	0.0559	0.0711	1.399	0.9915	0.0000
34	3	0.1022	0.1330	2.557	0.9702	0.0000
35	3	0.3012	0.3761	7.536	0.7617	0.0000
36	3	0.1175	0.1544	2.940	0.9598	0.0000
37	3	0.3162	0.3945	7.910	0.7380	-0.0002
38	3	0.3343	0.4151	8.364	0.7096	0.0001
39	3	0.1158	0.1529	2.896	0.9606	0.0000
40	3	0.3290	0.4114	8.232	0.7151	0.0000
41	3	0.3228	0.4023	8.076	0.7274	-0.0004
42	3	0.3360	0.4191	8.408	0.7041	0.0002
43	2	0.1014	0.1313	2.536	0.9710	0.0000
44	2	0.1184	0.1501	2.962	0.9621	0.0000
45	2	0.4024	0.5116	10.069	0.5592	0.0000
46	2	0.1350	0.1726	3.378	0.9498	0.0000
47	2	0.3884	0.4903	9.718	0.5954	-0.0002
48	2	0.4070	0.5096	10.182	0.5627	-0.0001
49	2	0.1387	0.1770	3.470	0.9473	0.0000
50	2	0.5477	0.7036	13.704	0.1662	0.0000
51	2	0.4229	0.5323	10.581	0.5233	-0.0001
52	2	0.3934	0.4886	9.843	0.5982	-0.0001
53	2	0.1186	0.1554	2.969	0.9593	0.0000
54	2	0.3343	0.4171	8.365	0.7071	0.0001
55	2	0.3293	0.4098	8.239	0.7171	0.0001
56	2	0.3701	0.4613	9.261	0.6417	-0.0002
57	2	0.3737	0.4647	9.351	0.6363	0.0000
58	1	0.1487	0.1891	3.720	0.9398	0.0000
59	1	0.5546	0.7131	13.877	0.1435	0.0000
60	1	0.4234	0.5328	10.594	0.5222	0.0000
61	1	0.4569	0.5704	11.432	0.4526	-0.0001
62	1	0.5986	0.7618	14.978	0.0244	0.0000
63	1	0.3809	0.4739	9.531	0.6218	0.0000

Table A32: Performance of ANFIS at Station 48632 (Cameron Highlands)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	0.0366	0.0490	1.157	0.9979	0.0000
2	5	0.0474	0.0626	1.499	0.9965	0.0000
3	5	0.0404	0.0545	1.279	0.9974	0.0000
4	5	0.0372	0.0498	1.175	0.9978	0.0000
5	5	0.0460	0.0640	1.453	0.9964	0.0000
6	5	0.0488	0.0687	1.542	0.9958	0.0000
7	5	0.3586	0.4769	11.338	0.8002	0.0001
8	4	0.0479	0.0633	1.514	0.9965	0.0000
9	4	0.0510	0.0676	1.613	0.9960	0.0001
10	4	0.0534	0.0736	1.688	0.9952	0.0001
11	4	0.0556	0.0767	1.757	0.9948	0.0000
12	4	0.3908	0.5111	12.355	0.7704	0.0000
13	4	0.0446	0.0608	1.411	0.9967	-0.0002
14	4	0.0460	0.0642	1.455	0.9964	0.0000
15	4	0.0516	0.0735	1.633	0.9952	-0.0001
16	4	0.3723	0.4925	11.770	0.7867	0.0000
17	4	0.0462	0.0640	1.461	0.9964	-0.0001
18	4	0.0492	0.0696	1.555	0.9957	0.0000
19	4	0.3681	0.4865	11.639	0.7920	-0.0001
20	4	0.0611	0.0950	1.931	0.9919	0.0000
21	4	0.3786	0.5038	11.969	0.7769	0.0002
22	4	0.3846	0.5027	12.159	0.7777	-0.0001
23	3	0.1631	0.2388	5.157	0.9499	-0.0001
24	3	0.0541	0.0745	1.712	0.9951	0.0001
25	3	0.0567	0.0782	1.791	0.9946	0.0000
26	3	0.4312	0.5557	13.632	0.7283	0.0001
27	3	0.0627	0.0856	1.982	0.9935	-0.0001
28	3	0.0576	0.0784	1.822	0.9946	-0.0001
29	3	0.4845	0.6138	15.317	0.6685	0.0004
30	3	0.0667	0.1021	2.108	0.9907	0.0001
31	3	0.4091	0.5425	12.935	0.7414	0.0002
32	3	0.4094	0.5301	12.944	0.7528	-0.0002
33	3	0.0464	0.0641	1.468	0.9964	0.0000
34	3	0.0534	0.0765	1.688	0.9948	-0.0001
35	3	0.3792	0.5036	11.988	0.7769	0.0003
36	3	0.0606	0.0946	1.917	0.9920	0.0000
37	3	0.4172	0.5477	13.192	0.7362	0.0001
38	3	0.3913	0.5123	12.372	0.7691	-0.0002
39	3	0.0607	0.0946	1.920	0.9920	0.0001
40	3	0.4027	0.5277	12.732	0.7553	0.0001
41	3	0.3888	0.5072	12.294	0.7737	0.0003
42	3	0.4415	0.5708	13.957	0.7132	0.0002
43	2	0.2787	0.3617	8.810	0.8851	0.0000
44	2	0.1648	0.2415	5.209	0.9488	0.0000
45	2	0.4830	0.6126	15.270	0.6691	-0.0001
46	2	0.0667	0.1018	2.108	0.9908	0.0001
47	2	0.5115	0.6571	16.171	0.6200	0.0001
48	2	0.4371	0.5626	13.820	0.7214	0.0000
49	2	0.0720	0.1089	2.276	0.9895	0.0000
50	2	0.6115	0.7758	19.333	0.4703	-0.0004
51	2	0.4860	0.6150	15.365	0.6670	0.0004
52	2	0.4627	0.6026	14.630	0.6804	0.0002
53	2	0.0624	0.0962	1.974	0.9918	0.0000
54	2	0.4235	0.5575	13.391	0.7265	0.0002
55	2	0.3964	0.5201	12.534	0.7620	0.0001
56	2	0.4753	0.6122	15.027	0.6698	0.0001
57	2	0.4594	0.5891	14.526	0.6943	0.0002
58	1	0.3020	0.3796	9.548	0.8737	0.0001
59	1	0.7897	0.9925	24.968	0.1319	0.0002
60	1	0.4860	0.6152	15.364	0.6665	0.0001
61	1	0.5533	0.7067	17.492	0.5600	0.0001
62	1	0.6401	0.8106	20.238	0.4219	0.0001
63	1	0.4804	0.6204	15.187	0.6607	0.0002

Table A33: Performance of ANFIS at Station 48647 (Subang)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	0.0348	0.0440	0.860	0.9980	0.0000
2	5	0.0691	0.0878	1.709	0.9920	0.0000
3	5	0.0468	0.0593	1.156	0.9963	0.0000
4	5	0.0402	0.0507	0.995	0.9973	0.0000
5	5	0.0564	0.0733	1.395	0.9944	-0.0001
6	5	0.1015	0.1346	2.510	0.9810	0.0002
7	5	0.3818	0.4963	9.438	0.7424	-0.0005
8	4	0.0722	0.0918	1.785	0.9912	0.0000
9	4	0.0724	0.0921	1.790	0.9911	0.0000
10	4	0.0772	0.0997	1.908	0.9896	0.0000
11	4	0.1108	0.1440	2.740	0.9783	0.0000
12	4	0.4752	0.6024	11.746	0.6209	-0.0003
13	4	0.0474	0.0606	1.172	0.9962	0.0000
14	4	0.0588	0.0764	1.452	0.9939	0.0000
15	4	0.1048	0.1391	2.589	0.9798	0.0000
16	4	0.3857	0.4998	9.532	0.7387	-0.0002
17	4	0.0572	0.0744	1.415	0.9942	0.0000
18	4	0.1021	0.1350	2.523	0.9809	0.0000
19	4	0.3828	0.4961	9.463	0.7425	-0.0002
20	4	0.1151	0.1511	2.845	0.9761	0.0000
21	4	0.4131	0.5329	10.211	0.7029	0.0001
22	4	0.4288	0.5530	10.598	0.6803	0.0000
23	3	0.0903	0.1140	2.233	0.9864	0.0000
24	3	0.0782	0.1007	1.933	0.9894	0.0000
25	3	0.1128	0.1469	2.787	0.9774	0.0000
26	3	0.4973	0.6266	12.293	0.5898	0.0000
27	3	0.0883	0.1157	2.183	0.9860	0.0000
28	3	0.1132	0.1464	2.798	0.9776	0.0000
29	3	0.5303	0.6667	13.106	0.5353	0.0003
30	3	0.1193	0.1554	2.949	0.9748	0.0000
31	3	0.4906	0.6229	12.127	0.5944	0.0001
32	3	0.5040	0.6366	12.457	0.5766	-0.0002
33	3	0.0650	0.0846	1.607	0.9925	0.0001
34	3	0.1102	0.1463	2.725	0.9776	0.0000
35	3	0.3922	0.5061	9.695	0.7321	-0.0001
36	3	0.1152	0.1516	2.848	0.9760	0.0000
37	3	0.4332	0.5554	10.707	0.6772	-0.0008
38	3	0.4322	0.5577	10.684	0.6747	-0.0001
39	3	0.1224	0.1594	3.024	0.9735	0.0000
40	3	0.4257	0.5503	10.522	0.6831	-0.0001
41	3	0.4294	0.5533	10.613	0.6798	0.0001
42	3	0.4641	0.5964	11.472	0.6278	-0.0002
43	2	0.1110	0.1434	2.744	0.9785	0.0000
44	2	0.1298	0.1682	3.209	0.9704	0.0000
45	2	0.5296	0.6670	13.091	0.5349	-0.0002
46	2	0.1194	0.1556	2.952	0.9747	0.0000
47	2	0.5321	0.6720	13.153	0.5275	-0.0002
48	2	0.5261	0.6611	13.005	0.5435	-0.0003
49	2	0.1341	0.1720	3.314	0.9691	0.0000
50	2	0.6677	0.8391	16.503	0.2629	0.0001
51	2	0.5630	0.7083	13.915	0.4754	-0.0002
52	2	0.5191	0.6591	12.830	0.5459	-0.0002
53	2	0.1326	0.1731	3.277	0.9687	0.0000
54	2	0.4375	0.5605	10.813	0.6712	0.0000
55	2	0.4349	0.5599	10.749	0.6721	-0.0003
56	2	0.4890	0.6254	12.086	0.5906	0.0001
57	2	0.5078	0.6511	12.551	0.5563	0.0000
58	1	0.1563	0.2006	3.862	0.9579	0.0000
59	1	0.6775	0.8533	16.747	0.2380	0.0000
60	1	0.5648	0.7117	13.960	0.4704	0.0000
61	1	0.5704	0.7181	14.099	0.4606	-0.0002
62	1	0.7620	0.9492	18.836	0.0565	-0.0001
63	1	0.5127	0.6554	12.673	0.5504	0.0000

Table A34: Performance of ANFIS at Station 48649 (Muadzam Shah)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	0.0317	0.0441	0.912	0.9970	-0.0001
2	5	0.0404	0.0551	1.160	0.9956	0.0001
3	5	0.0339	0.0467	0.974	0.9967	0.0000
4	5	0.0317	0.0443	0.911	0.9970	0.0001
5	5	0.0343	0.0493	0.986	0.9964	0.0001
6	5	0.0604	0.0826	1.735	0.9903	0.0000
7	5	0.2711	0.3422	7.792	0.8354	-0.0006
8	4	0.0415	0.0566	1.192	0.9953	0.0002
9	4	0.0459	0.0625	1.320	0.9944	0.0001
10	4	0.0430	0.0600	1.237	0.9948	0.0000
11	4	0.0604	0.0825	1.736	0.9904	0.0000
12	4	0.3409	0.4257	9.795	0.7453	0.0001
13	4	0.0371	0.0513	1.067	0.9961	0.0000
14	4	0.0358	0.0511	1.029	0.9961	0.0000
15	4	0.0669	0.0908	1.924	0.9883	0.0000
16	4	0.2846	0.3585	8.177	0.8194	0.0005
17	4	0.0350	0.0501	1.005	0.9962	0.0000
18	4	0.0614	0.0839	1.763	0.9900	0.0001
19	4	0.2796	0.3526	8.033	0.8253	-0.0004
20	4	0.0735	0.1006	2.113	0.9857	0.0000
21	4	0.2813	0.3543	8.083	0.8235	0.0000
22	4	0.3222	0.4016	9.259	0.7735	-0.0001
23	3	0.0574	0.0779	1.650	0.9914	0.0000
24	3	0.0440	0.0609	1.264	0.9946	0.0000
25	3	0.0669	0.0913	1.924	0.9882	0.0000
26	3	0.3853	0.4823	11.072	0.6732	-0.0001
27	3	0.0501	0.0689	1.440	0.9932	0.0000
28	3	0.0624	0.0849	1.794	0.9898	0.0001
29	3	0.4328	0.5454	12.438	0.5818	0.0004
30	3	0.0739	0.1021	2.124	0.9853	0.0000
31	3	0.3446	0.4321	9.902	0.7376	-0.0004
32	3	0.3657	0.4547	10.507	0.7093	0.0000
33	3	0.0410	0.0576	1.177	0.9951	0.0000
34	3	0.0710	0.0959	2.040	0.9870	0.0000
35	3	0.2867	0.3606	8.238	0.8172	-0.0003
36	3	0.0772	0.1053	2.217	0.9844	-0.0001
37	3	0.3124	0.3945	8.978	0.7813	0.0003
38	3	0.3287	0.4093	9.447	0.7648	-0.0001
39	3	0.0743	0.1016	2.134	0.9855	0.0000
40	3	0.3046	0.3844	8.754	0.7923	0.0001
41	3	0.3265	0.4070	9.382	0.7673	0.0001
42	3	0.3777	0.4752	10.854	0.6827	0.0001
43	2	0.0609	0.0825	1.750	0.9903	0.0001
44	2	0.0744	0.0996	2.137	0.9860	0.0000
45	2	0.4338	0.5460	12.466	0.5808	0.0001
46	2	0.0768	0.1069	2.208	0.9839	0.0000
47	2	0.4364	0.5490	12.539	0.5766	0.0000
48	2	0.3997	0.4974	11.485	0.6523	-0.0001
49	2	0.0750	0.1032	2.154	0.9850	0.0000
50	2	0.6178	0.7892	17.752	0.1253	0.0000
51	2	0.4420	0.5563	12.701	0.5649	0.0001
52	2	0.3990	0.5015	11.467	0.6465	0.0001
53	2	0.0831	0.1125	2.387	0.9821	0.0000
54	2	0.3180	0.4006	9.138	0.7745	-0.0005
55	2	0.3296	0.4101	9.470	0.7637	-0.0001
56	2	0.4158	0.5188	11.948	0.6219	0.0000
57	2	0.4238	0.5282	12.180	0.6080	0.0000
58	1	0.0849	0.1145	2.441	0.9815	0.0000
59	1	0.6256	0.7963	17.979	0.1097	0.0000
60	1	0.4420	0.5562	12.703	0.5651	0.0001
61	1	0.4778	0.6017	13.730	0.4910	0.0000
62	1	0.6621	0.8364	19.025	0.0172	0.0000
63	1	0.4345	0.5390	12.487	0.5917	-0.0001

Table A35: Performance of ANFIS at Station 48650 (KLIA)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
1	6	0.0405	0.0572	1.000	0.9967	-0.0001
2	5	0.0785	0.1013	1.939	0.9896	0.0001
3	5	0.0553	0.0764	1.365	0.9941	0.0001
4	5	0.0503	0.0686	1.243	0.9952	0.0000
5	5	0.0748	0.1051	1.847	0.9888	0.0000
6	5	0.0945	0.1367	2.335	0.9810	-0.0002
7	5	0.3892	0.4921	9.612	0.7540	0.0003
8	4	0.0827	0.1070	2.042	0.9884	-0.0001
9	4	0.0812	0.1043	2.004	0.9890	-0.0001
10	4	0.1104	0.1523	2.725	0.9765	0.0000
11	4	0.1166	0.1589	2.881	0.9743	-0.0001
12	4	0.4394	0.5522	10.851	0.6902	0.0010
13	4	0.0551	0.0755	1.361	0.9942	0.0000
14	4	0.0823	0.1161	2.032	0.9863	-0.0001
15	4	0.1057	0.1544	2.611	0.9758	-0.0001
16	4	0.4002	0.5052	9.885	0.7406	0.0002
17	4	0.0780	0.1101	1.927	0.9877	-0.0001
18	4	0.1012	0.1465	2.500	0.9782	0.0000
19	4	0.3973	0.5007	9.812	0.7454	-0.0003
20	4	0.1355	0.2117	3.347	0.9544	-0.0001
21	4	0.4033	0.5087	9.960	0.7370	0.0012
22	4	0.4233	0.5354	10.454	0.7086	-0.0002
23	3	0.0906	0.1177	2.238	0.9859	-0.0001
24	3	0.1098	0.1518	2.712	0.9766	0.0002
25	3	0.1228	0.1690	3.032	0.9710	0.0000
26	3	0.4600	0.5779	11.359	0.6605	-0.0003
27	3	0.1234	0.1713	3.048	0.9703	0.0001
28	3	0.1171	0.1606	2.892	0.9738	0.0000
29	3	0.4880	0.6215	12.053	0.6072	0.0003
30	3	0.1601	0.2405	3.955	0.9412	0.0002
31	3	0.4511	0.5658	11.140	0.6749	0.0000
32	3	0.4650	0.5844	11.485	0.6528	-0.0007
33	3	0.0848	0.1172	2.093	0.9861	-0.0001
34	3	0.1061	0.1543	2.619	0.9758	-0.0001
35	3	0.3961	0.4977	9.782	0.7482	0.0002
36	3	0.1392	0.2181	3.437	0.9516	0.0000
37	3	0.4228	0.5313	10.441	0.7132	-0.0001
38	3	0.4326	0.5464	10.685	0.6965	0.0003
39	3	0.1370	0.2137	3.382	0.9536	0.0000
40	3	0.4362	0.5485	10.773	0.6946	0.0001
41	3	0.4253	0.5381	10.502	0.7057	0.0004
42	3	0.4637	0.5900	11.453	0.6465	-0.0006
43	2	0.1415	0.1923	3.494	0.9625	0.0000
44	2	0.1274	0.1732	3.146	0.9695	0.0000
45	2	0.4918	0.6254	12.145	0.6023	0.0001
46	2	0.1612	0.2416	3.982	0.9407	0.0000
47	2	0.4965	0.6191	12.262	0.6107	0.0000
48	2	0.4834	0.6073	11.938	0.6250	-0.0001
49	2	0.1731	0.2551	4.274	0.9339	0.0000
50	2	0.6657	0.8329	16.442	0.2950	0.0001
51	2	0.5162	0.6543	12.748	0.5646	0.0001
52	2	0.5104	0.6464	12.606	0.5757	0.0000
53	2	0.1433	0.2219	3.539	0.9499	0.0000
54	2	0.4383	0.5511	10.824	0.6916	0.0000
55	2	0.4315	0.5441	10.656	0.6991	0.0000
56	2	0.4817	0.6106	11.897	0.6211	-0.0001
57	2	0.5018	0.6348	12.392	0.5905	0.0001
58	1	0.1930	0.2821	4.766	0.9192	0.0000
59	1	0.6839	0.8559	16.891	0.2555	0.0000
60	1	0.5189	0.6570	12.816	0.5611	0.0001
61	1	0.5512	0.6954	13.613	0.5087	-0.0001
62	1	0.7546	0.9505	18.637	0.0821	0.0003
63	1	0.5044	0.6379	12.458	0.5865	-0.0001

Table A36: Performance of ANFIS at Station 48657 (Kuantan)

Combinations	Variables	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
1	6	0.0326	0.0419	0.847	0.9979	0.0000
2	5	0.0666	0.0868	1.730	0.9911	0.0000
3	5	0.0415	0.0532	1.077	0.9967	-0.0001
4	5	0.0386	0.0493	1.003	0.9971	0.0000
5	5	0.0520	0.0690	1.351	0.9944	0.0000
6	5	0.0686	0.0899	1.782	0.9904	-0.0002
7	5	0.2876	0.3712	7.467	0.8369	-0.0001
8	4	0.0690	0.0896	1.792	0.9905	0.0001
9	4	0.0699	0.0906	1.814	0.9903	0.0000
10	4	0.0792	0.1034	2.056	0.9873	0.0000
11	4	0.0840	0.1085	2.181	0.9861	0.0000
12	4	0.3543	0.4536	9.199	0.7566	0.0001
13	4	0.0430	0.0550	1.117	0.9964	0.0000
14	4	0.0533	0.0702	1.383	0.9942	0.0000
15	4	0.0782	0.1017	2.029	0.9878	-0.0001
16	4	0.3047	0.3908	7.909	0.8192	-0.0003
17	4	0.0521	0.0691	1.352	0.9944	0.0000
18	4	0.0762	0.0988	1.978	0.9885	0.0000
19	4	0.2932	0.3774	7.612	0.8313	-0.0004
20	4	0.0946	0.1241	2.456	0.9817	0.0000
21	4	0.2982	0.3837	7.743	0.8257	0.0000
22	4	0.3470	0.4388	9.009	0.7719	0.0003
23	3	0.0892	0.1170	2.316	0.9838	0.0000
24	3	0.0795	0.1039	2.063	0.9872	0.0001
25	3	0.0885	0.1137	2.297	0.9847	0.0000
26	3	0.4049	0.5132	10.513	0.6885	-0.0001
27	3	0.0903	0.1201	2.344	0.9829	0.0000
28	3	0.0847	0.1090	2.198	0.9860	0.0000
29	3	0.4532	0.5724	11.766	0.6123	0.0003
30	3	0.1041	0.1349	2.703	0.9785	-0.0001
31	3	0.3609	0.4629	9.370	0.7465	0.0001
32	3	0.3927	0.4972	10.196	0.7074	-0.0001
33	3	0.0562	0.0730	1.458	0.9937	0.0000
34	3	0.0798	0.1036	2.071	0.9873	0.0000
35	3	0.3046	0.3911	7.907	0.8189	0.0002
36	3	0.0962	0.1255	2.497	0.9814	0.0000
37	3	0.3390	0.4318	8.801	0.7793	0.0000
38	3	0.3585	0.4547	9.306	0.7551	0.0000
39	3	0.0946	0.1243	2.457	0.9817	0.0000
40	3	0.3179	0.4072	8.252	0.8037	0.0001
41	3	0.3465	0.4393	8.997	0.7714	0.0000
42	3	0.3780	0.4786	9.813	0.7286	0.0001
43	2	0.1104	0.1443	2.866	0.9753	0.0000
44	2	0.0973	0.1256	2.525	0.9814	0.0000
45	2	0.4534	0.5721	11.770	0.6128	0.0001
46	2	0.1046	0.1351	2.716	0.9784	0.0000
47	2	0.4711	0.5965	12.232	0.5791	-0.0001
48	2	0.4354	0.5498	11.303	0.6423	0.0001
49	2	0.1064	0.1401	2.763	0.9767	0.0000
50	2	0.6773	0.8546	17.585	0.1362	-0.0002
51	2	0.4718	0.5934	12.248	0.5832	0.0000
52	2	0.4189	0.5315	10.876	0.6656	0.0002
53	2	0.0978	0.1277	2.538	0.9807	0.0000
54	2	0.3366	0.4296	8.737	0.7816	0.0002
55	2	0.3616	0.4584	9.387	0.7512	-0.0001
56	2	0.4298	0.5422	11.159	0.6517	0.0000
57	2	0.4040	0.5098	10.488	0.6922	0.0000
58	1	0.1199	0.1561	3.112	0.9711	0.0000
59	1	0.6873	0.8675	17.844	0.1092	-0.0001
60	1	0.4724	0.5937	12.264	0.5827	0.0000
61	1	0.5419	0.6886	14.069	0.4386	0.0000
62	1	0.7277	0.9100	18.893	0.0201	0.0000
63	1	0.4372	0.5514	11.350	0.6398	-0.0001

Appendix B: Performance of Bootstrap Aggregating Integrated Machine Learning Models at Different Stations

Table B1: Performance of BMLP at Cluster 1 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48603 (Alor Setar)	1	0.0277 (-0.23)	0.0356 (-0.27)	0.646 (-0.23)	0.9989 (0.00)	0.0000
	4	0.0402 (0.41)	0.0517 (0.67)	0.939 (0.41)	0.9977 (-0.00)	-0.0001
	13	0.0514 (1.27)	0.0663 (1.38)	1.201 (1.27)	0.9961 (-0.01)	-0.0005
	33	0.0764 (0.82)	0.1073 (0.92)	1.786 (0.82)	0.9899 (-0.02)	-0.0003
	44	0.1379 (-0.27)	0.1829 (-0.30)	3.223 (-0.27)	0.9706 (0.02)	0.0001
	58	0.2583 (-0.22)	0.3622 (-0.01)	6.036 (-0.22)	0.8847 (-0.01)	-0.0035
Station 48615 (Kota Bharu)	1	0.0237 (0.35)	0.0297 (0.25)	0.545 (0.35)	0.9992 (-0.00)	0.0000
	4	0.0484 (0.84)	0.0663 (1.81)	1.116 (0.84)	0.9958 (-0.01)	0.0000
	13	0.0560 (-1.04)	0.0773 (-0.80)	1.290 (-1.04)	0.9943 (0.01)	-0.0002
	33	0.0972 (0.37)	0.1411 (0.28)	2.238 (0.37)	0.9810 (-0.01)	-0.0004
	44	0.1604 (0.13)	0.2165 (0.25)	3.695 (0.13)	0.9553 (-0.02)	-0.0004
	58	0.2263 (0.24)	0.3081 (0.03)	5.212 (0.24)	0.9094 (-0.01)	-0.0018
Station 48650 (KLIA)	1	0.0263 (0.41)	0.0310 (0.73)	0.649 (0.41)	0.9990 (-0.00)	0.0000
	4	0.0396 (0.15)	0.0504 (0.35)	0.978 (0.15)	0.9974 (-0.00)	-0.0001
	13	0.0461 (0.30)	0.0601 (0.70)	1.138 (0.30)	0.9963 (-0.01)	-0.0002
	33	0.0789 (0.47)	0.1090 (0.56)	1.949 (0.47)	0.9879 (-0.01)	-0.0002
	44	0.1275 (0.31)	0.1734 (0.54)	3.148 (0.31)	0.9695 (-0.03)	-0.0007
	58	0.1925 (-0.14)	0.2825 (0.02)	4.755 (-0.14)	0.9190 (-0.01)	-0.0032
Station 48657 (Kuantan)	1	0.0237 (0.72)	0.0291 (0.73)	0.616 (0.72)	0.9990 (-0.00)	0.0000
	4	0.0320 (-0.25)	0.0401 (-0.15)	0.831 (-0.25)	0.9981 (0.00)	0.0001
	13	0.0384 (-1.00)	0.0487 (-1.26)	0.998 (-1.00)	0.9972 (0.01)	-0.0001
	33	0.0555 (0.35)	0.0723 (0.29)	1..440 (0.35)	0.9938 (-0.00)	-0.0001
	44	0.0973 (0.37)	0.1262 (0.84)	2.525 (0.37)	0.9812 (-0.03)	-0.0003
	58	0.1198 (0.15)	0.1560 (0.02)	3.111 (0.15)	0.9712 (-0.00)	-0.0005

* Change (in %) in bracket as compared to the MLP.

Table B2: Performance of BMLP at Cluster 2 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48620 (Sitiawan)	1	0.0305 (-11.58)	0.0576 (-9.21)	0.802 (-11.58)	0.9889 (0.05)	-0.0022
	4	0.0333 (-7.17)	0.0603 (-5.45)	0.875 (-7.17)	0.9890 (0.00)	-0.0008
	13	0.0405 (-6.69)	0.0694 (-5.45)	1.065 (-6.69)	0.9878 (0.07)	-0.0015
	33	0.0461 (-5.10)	0.0773 (-6.96)	1.211 (-5.10)	0.9857 (0.18)	-0.0013
	43	0.0862 (-0.09)	0.1287 (0.59)	2.264 (-0.09)	0.9701 (-0.06)	-0.0019
	58	0.1161 (-0.19)	0.1624 (-0.05)	3.048 (-0.19)	0.9559 (0.01)	-0.0018
Station 48625 (Ipoh)	1	0.0288 (0.61)	0.0347 (0.79)	0.722 (0.61)	0.9980 (-0.00)	-0.0002
	4	0.0325 (0.43)	0.0397 (0.47)	0.814 (0.43)	0.9974 (-0.00)	-0.0001
	13	0.0412 (0.30)	0.0515 (0.53)	1.030 (0.30)	0.9956 (-0.00)	-0.0002
	33	0.0551 (0.05)	0.0701 (0.19)	1.378 (0.05)	0.9917 (-0.00)	-0.0003
	43	0.1024 (-0.14)	0.1324 (-0.16)	2.562 (-0.14)	0.9706 (0.01)	-0.0002
	58	0.1161 (0.07)	0.1893 (0.07)	3.720 (0.07)	0.9398 (-0.01)	-0.0013
Station 48647 (Subang)	1	0.0274 (0.26)	0.0324 (0.57)	0.676 (0.26)	0.9989 (-0.00)	-0.0001
	4	0.0364 (0.83)	0.0452 (1.03)	0.901 (0.83)	0.9979 (-0.00)	-0.0001
	13	0.0437 (0.88)	0.0553 (0.89)	1.080 (0.88)	0.9968 (-0.01)	-0.0002
	33	0.0644 (0.56)	0.0836 (0.65)	1.591 (0.56)	0.9927 (-0.01)	-0.0003
	43	0.1114 (0.11)	0.1437 (0.09)	2.753 (0.11)	0.9785 (-0.00)	-0.0003
	58	0.1562 (0.04)	0.2009 (0.20)	3.860 (0.04)	0.9579 (-0.01)	-0.0009

* Change (in %) in bracket as compared to the MLP.

Table B3: Performance of BMLP at Cluster 3 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48600 (Pulau Langkawi)	1	0.0281 (0.50)	0.0371 (-9.21)	0.643 (0.50)	0.9989 (-0.00)	-0.0001
	4	0.0449 (0.96)	0.0642 (-5.45)	1.025 (0.96)	0.9968 (-0.01)	-0.0003
	13	0.0593 (-1.98)	0.0832 (-5.45)	1.355 (-1.98)	0.9946 (0.04)	-0.0002
	23	0.1133 (-1.19)	0.1550 (-6.96)	2.588 (-1.19)	0.9811 (0.07)	-0.0002
	44	0.1879 (0.20)	0.2705 (0.59)	4.290 (0.20)	0.9426(-0.01)	0.0001
	58	0.3809 (0.26)	0.5198 (-0.05)	8.698 (0.26)	0.7877 (-0.01)	0.0000
Station 48601 (Bayan Lepas)	1	0.0277 (2.95)	0.0353 (1.91)	0.639 (2.95)	0.9988 (-0.01)	0.0000
	4	0.0437 (2.77)	0.0580 (0.75)	1.008 (2.77)	0.9968 (-0.01)	-0.0003
	13	0.0538 (4.18)	0.0715 (-2.76)	1.241 (4.18)	0.9952 (-0.04)	-0.0001
	23	0.1067 (8.94)	0.1402 (-1.64)	2.458 (8.94)	0.9815 (-0.30)	0.0003
	44	0.1488 (-1.22)	0.1990 (0.15)	3.430 (-1.22)	0.9629 (0.86)	0.0001
	58	0.2730 (-1.31)	0.3969 (0.03)	6.294 (-1.31)	0.8517 (0.74)	-0.0028

* Change (in %) in bracket as compared to the MLP.

Table B4: Performance of BMLP at Cluster 4 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48623 (Lubok Merbau)	1	0.0304 (0.44)	0.0374 (0.99)	0.747 (0.44)	0.9978 (-0.00)	0.0002
	4	0.0318 (0.29)	0.0393 (0.49)	0.781 (0.29)	0.9976 (-0.00)	0.0001
	17	0.0377 (-0.52)	0.0491 (0.25)	0.927 (-0.52)	0.9962 (-0.00)	0.0000
	33	0.0461 (0.24)	0.0596 (0.76)	1.132 (0.24)	0.9944 (-0.01)	0.0000
	43	0.0863 (0.50)	0.1135 (0.71)	2.121 (0.50)	0.9796 (-0.03)	-0.0005
	58	0.1098 (0.16)	0.1476 (0.14)	2.700 (0.16)	0.9655 (-0.01)	-0.0005
Station 48649 (Muadzam Shah)	1	0.0260 (0.02)	0.0357 (1.55)	0.746 (0.02)	0.9980 (-0.01)	-0.0002
	4	0.0286 (2.04)	0.0394 (3.19)	0.821 (2.04)	0.9976 (-0.01)	-0.0001
	17	0.0343 (0.46)	0.0492 (1.16)	0.987 (0.46)	0.9964 (-0.01)	-0.0001
	33	0.0404 (1.50)	0.0569 (1.82)	1.160 (1.50)	0.9953 (-0.02)	-0.0003
	43	0.0615 (0.73)	0.0877 (6.28)	1.769 (0.73)	0.9874 (0.30)	-0.0006
	58	0.0852 (0.09)	0.1148 (0.14)	2.448 (0.09)	0.9814 (-0.01)	-0.0004

* Change (in %) in bracket as compared to the MLP.

Table B5: Performance of BMLP at Cluster 5 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48632 (Cameron Highlands)	1	0.0272 (0.64)	0.0326 (0.76)	0.860 (0.64)	0.9991 (-0.00)	0.0000
	4	0.0297 (0.87)	0.0362 (0.92)	0.940 (0.87)	0.9988 (-0.00)	0.0000
	13	0.0358 (1.10)	0.0469 (1.96)	1.132 (1.10)	0.9981 (-0.01)	0.0001
	33	0.0447 (-0.74)	0.0616 (-0.83)	1.414 (-0.74)	0.9967 (0.01)	-0.0001
	53	0.0630 (0.05)	0.0967 (0.14)	1.990 (0.05)	0.9917 (-0.00)	0.0000
	58	0.3020 (-0.01)	0.3790 (0.11)	9.547 (-0.01)	0.8742 (-0.01)	-0.0026

* Change (in %) in bracket as compared to the MLP.

Table B6: Performance of BSVM at Cluster 1 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48603 (Alor Setar)	1	0.0443 (1.43)	0.0548 (1.81)	1.035 (1.43)	0.9975 (-0.01)	-0.0133
	4	0.0463 (0.61)	0.0603 (0.64)	1.082 (0.61)	0.9968 (-0.00)	-0.0047
	13	0.0565 (0.35)	0.0723 (0.40)	1.321 (0.35)	0.9955 (-0.00)	-0.0111
	33	0.0799 (0.18)	0.1104 (0.21)	1.867 (0.18)	0.9893 (-0.00)	-0.0056
	44	0.1378 (0.11)	0.1836 (0.05)	3.220 (0.11)	0.9705 (0.00)	-0.0068
	58	0.2499 (-0.02)	0.3741 (-0.07)	5.840 (-0.02)	0.8838 (-0.00)	-0.0824
Station 48615 (Kota Bharu)	1	0.0376 (1.71)	0.0500 (2.17)	0.865 (1.71)	0.9976 (-0.01)	-0.0009
	4	0.0522 (0.59)	0.0705 (0.60)	1.203 (0.59)	0.9953 (-0.01)	0.0005
	13	0.0581 (0.59)	0.0795 (0.52)	1.339 (0.59)	0.9940 (-0.01)	-0.0005
	33	0.0968 (0.17)	0.1409 (0.12)	2.230 (0.17)	0.9810 (-0.00)	-0.0011
	44	0.1587 (0.02)	0.2168 (0.06)	3.656 (0.02)	0.9553 (-0.00)	-0.0106
	58	0.2228 (0.08)	0.3128 (-0.00)	5.131 (0.08)	0.9088 (-0.00)	-0.0428
Station 48650 (KLIA)	1	0.0386 (0.67)	0.0498 (1.37)	0.953 (0.67)	0.9977 (-0.01)	-0.0124
	4	0.0434 (0.87)	0.0565 (1.28)	1.072 (0.87)	0.9968 (-0.01)	-0.0006
	13	0.0485 (0.43)	0.0638 (0.49)	1.197 (0.43)	0.9959 (-0.00)	-0.0035
	33	0.0803 (0.29)	0.1108 (0.44)	1.982 (0.29)	0.9876 (-0.01)	-0.0020
	44	0.1276 (0.07)	0.1751 (0.12)	3.152 (0.07)	0.9690 (-0.01)	-0.0104
	58	0.1886 (0.06)	0.2882 (-0.08)	4.659 (0.06)	0.9183 (-0.00)	-0.0456
Station 48657 (Kuantan)	1	0.0349 (0.76)	0.0291 (1.27)	0.907 (0.76)	0.9978 (-0.01)	-0.0086
	4	0.0369 (0.60)	0.0401 (0.85)	0.958 (0.60)	0.9973 (0.01)	-0.0038
	13	0.0413 (0.36)	0.0487 (0.48)	1.073 (0.36)	0.9967 (-0.00)	-0.0058
	33	0.0569 (0.38)	0.0723 (0.32)	1.476 (0.38)	0.9936 (-0.00)	-0.0024
	44	0.0976 (0.04)	0.1262 (0.06)	2.533 (0.04)	0.9813 (-0.00)	-0.0017
	58	0.1201 (0.03)	0.1560 (0.00)	3.118 (0.03)	0.9710 (0.00)	-0.0076

* Change (in %) in bracket as compared to the SVM.

Table B7: Performance of BSVM at Cluster 2 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48620 (Sitiawan)	1	0.0351 (1.28)	0.0606 (1.03)	0.924 (1.28)	0.9901 (-0.01)	-0.0091
	4	0.0353 (1.13)	0.0607 (0.81)	0.927 (1.13)	0.9900 (-0.00)	-0.0044
	13	0.0401 (0.41)	0.0662 (0.43)	1.052 (0.41)	0.9893 (-0.00)	-0.0055
	33	0.0453 (0.22)	0.0732 (0.18)	1.189 (0.22)	0.9877 (-0.00)	-0.0018
	43	0.0855 (0.09)	0.1258 (0.10)	2.245 (0.09)	0.9717 (-0.00)	-0.0031
	58	0.1156 (0.04)	0.1614 (0.02)	3.036 (0.04)	0.9568 (-0.00)	-0.0083
Station 48625 (Ipoh)	1	0.0348 (0.70)	0.0423 (0.89)	0.871 (0.70)	0.9972 (-0.01)	-0.0094
	4	0.0358 (0.61)	0.0440 (0.75)	0.896 (0.61)	0.9968 (-0.00)	-0.0046
	13	0.0429 (0.37)	0.0534 (0.44)	1.072 (0.37)	0.9953 (-0.00)	-0.0043
	33	0.0556 (0.17)	0.0708 (0.17)	1.391 (0.17)	0.9916 (-0.00)	-0.0014
	43	0.1033 (0.11)	0.1341 (0.14)	2.584 (0.11)	0.9699 (-0.01)	-0.0051
	58	0.1485 (0.04)	0.1917 (0.02)	3.715 (0.04)	0.9390 (-0.00)	-0.0210
Station 48647 (Subang)	1	0.0405 (0.94)	0.0509 (1.29)	1.000 (0.94)	0.9976 (-0.01)	-0.0125
	4	0.0413 (0.71)	0.0531 (0.97)	1.021 (0.71)	0.9971 (-0.01)	-0.0006
	13	0.0465 (0.39)	0.0597 (0.51)	1.149 (0.39)	0.9963 (-0.00)	-0.0036
	33	0.0658 (0.31)	0.0858 (0.35)	1.627 (0.31)	0.9923 (-0.00)	0.0013
	43	0.1120 (0.05)	0.1445 (0.09)	2.767 (0.05)	0.9782 (-0.00)	-0.0015
	58	0.1559 (0.03)	0.2017 (0.02)	3.855 (0.03)	0.9578 (-0.00)	-0.0165

* Change (in %) in bracket as compared to the SVM.

Table B8: Performance of BSVM at Cluster 3 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
Station 48600 (Pulau Langkawi)	1	0.0481 (0.21)	0.0587 (0.94)	1.099 (0.21)	0.9976 (-0.01)	-0.0172
	4	0.0515 (0.57)	0.0699 (0.92)	1.176 (0.57)	0.9962 (-0.01)	-0.0084
	13	0.0632 (0.03)	0.0849 (0.33)	1.443 (0.03)	0.9944 (-0.00)	-0.0098
	23	0.1133 (0.13)	0.1549 (0.16)	2.587 (0.13)	0.9812 (-0.01)	-0.0034
	44	0.1854 (0.10)	0.2722 (0.07)	4.234 (0.10)	0.9424(-0.00)	-0.0254
	58	0.3510 (0.02)	0.5452 (-0.01)	8.016 (0.02)	0.7878 (-0.00)	-0.1587
Station 48601 (Bayan Lepas)	1	0.0432 (0.78)	0.0551 (1.56)	0.996 (0.78)	0.9988 (-0.01)	-0.0155
	4	0.0477 (0.64)	0.0646 (0.85)	1.099 (0.64)	0.9968 (-0.01)	0.0009
	13	0.0564 (0.34)	0.0759 (0.55)	1.301 (0.34)	0.9952 (-0.01)	-0.0033
	23	0.1074 (0.13)	0.1408 (0.19)	2.474 (0.13)	0.9815 (-0.01)	-0.0034
	44	0.1488 (0.15)	0.1998 (0.19)	3.429 (0.15)	0.9629 (-0.01)	-0.0116
	58	0.2643 (0.00)	0.4056 (0.06)	6.092 (0.00)	0.8517 (-0.00)	-0.0803

* Change (in %) in bracket as compared to the SVM.

Table B9: Performance of BSVM at Cluster 4 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
Station 48623 (Lubok Merbau)	1	0.0377 (0.71)	0.0465 (0.75)	0.928 (0.71)	0.9968 (-0.00)	-0.0115
	4	0.0376 (0.47)	0.0462 (0.61)	0.923 (0.47)	0.9968 (-0.00)	-0.0099
	17	0.0415 (0.18)	0.0532 (0.47)	1.019 (0.18)	0.9956 (-0.00)	-0.0057
	33	0.0481 (0.00)	0.0625 (0.18)	1.183 (0.00)	0.9939 (-0.00)	-0.0039
	43	0.0884 (0.14)	0.1166 (0.16)	2.173 (0.14)	0.9786 (-0.01)	-0.0036
	58	0.1104 (0.09)	0.1513 (0.02)	2.714 (0.09)	0.9643 (-0.00)	-0.0146
Station 48649 (Muadzam Shah)	1	0.0334 (1.24)	0.0447 (1.34)	0.960 (1.24)	0.9971 (-0.01)	-0.0052
	4	0.0336 (0.69)	0.0450 (0.95)	0.65 (0.69)	0.9971 (-0.01)	-0.0055
	17	0.0378 (0.48)	0.0528 (0.78)	1.085 (0.48)	0.9959 (-0.01)	-0.0034
	33	0.0419 (0.38)	0.0580 (0.64)	1.204 (0.38)	0.9951 (-0.01)	-0.0020
	43	0.0620 (0.09)	0.0837 (0.21)	1.783 (0.09)	0.9901 (-0.00)	-0.0003
	58	0.0849 (0.04)	0.1149 (-0.02)	2.439 (0.04)	0.9814 (0.00)	-0.0060

* Change (in %) in bracket as compared to the SVM.

Table B10: Performance of BSVM at Cluster 5 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
Station 48632 (Cameron Highlands)	1	0.0416 (-0.03)	0.0522 (0.96)	1.314 (-0.03)	0.9978 (-0.01)	-0.0099
	4	0.0400 (2.06)	0.0508 (2.58)	1.263 (2.06)	0.9978 (-0.01)	-0.0069
	13	0.0443 (1.25)	0.0567 (1.37)	1.399 (1.25)	0.9974 (-0.01)	-0.0154
	33	0.0482 (0.38)	0.0654 (0.37)	1.524 (0.38)	0.9963 (0.00)	-0.0056
	53	0.0648 (0.24)	0.0984 (0.13)	2.049 (0.24)	0.9915 (-0.00)	0.0046
	58	0.2808 (0.05)	0.3905 (0.07)	8.879 (0.05)	0.8705 (-0.00)	-0.0644

* Change (in %) in bracket as compared to the SVM.

Table B11: Performance of BANFIS at Cluster 1 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48603 (Alor Setar)	1	0.0443 (0.99)	0.0589 (0.58)	1.034 (0.99)	0.9989 (0.00)	0.0000
	4	0.0510 (0.64)	0.0676 (0.62)	1.192 (0.64)	0.9977 (-0.00)	-0.0001
	13	0.0580 (-0.03)	0.0763 (-0.13)	1.354 (-0.03)	0.9961 (-0.01)	-0.0001
	33	0.0863 (5.65)	0.1186 (4.44)	2.016 (5.65)	0.9899 (-0.02)	0.0000
	44	0.1383 (0.29)	0.1839 (0.28)	3.233 (0.29)	0.9706 (0.02)	-0.0004
	58	0.2590 (-0.00)	0.3623 (0.04)	6.054 (-0.00)	0.8847 (-0.01)	-0.0007
Station 48615 (Kota Bharu)	1	0.0442 (-3.29)	0.0626 (-3.54)	1.017 (-3.29)	0.9992 (-0.00)	0.0001
	4	0.0597 (0.22)	0.0829 (-0.12)	1.375 (0.22)	0.9958 (-0.01)	0.0001
	13	0.0648 (0.35)	0.0897 (0.36)	1.491 (0.35)	0.9943 (0.01)	0.0001
	33	0.0976 (0.14)	0.1409 (0.16)	2.247 (0.14)	0.9810 (-0.01)	0.0002
	44	0.1614 (0.14)	0.2173 (0.07)	3.718 (0.14)	0.9553 (-0.02)	-0.0002
	58	0.2262 (0.07)	0.3078 (0.03)	5.209 (0.07)	0.9094 (-0.01)	0.0001
Station 48650 (KLIA)	1	0.0402 (-0.72)	0.0567 (-0.89)	0.993 (-0.72)	0.9990 (-0.00)	-0.0002
	4	0.0512 (1.62)	0.0700 (2.02)	1.263 (1.62)	0.9974 (-0.00)	0.0000
	13	0.0553 (0.36)	0.0759 (0.52)	1.366 (0.36)	0.9963 (-0.01)	-0.0001
	33	0.0841 (-0.78)	0.1163 (-0.82)	2.077 (-0.78)	0.9879 (-0.01)	-0.0001
	44	0.1277 (0.27)	0.1737 (0.31)	3.154 (0.27)	0.9695 (-0.03)	0.0001
	58	0.1930 (0.01)	0.2822 (0.04)	4.767 (0.01)	0.9190 (-0.01)	-0.0003
Station 48657 (Kuantan)	1	0.0328 (0.56)	0.0422 (0.91)	0.851 (0.56)	0.9990 (-0.00)	-0.0001
	4	0.0387 (0.04)	0.0494 (0.21)	1.003 (0.04)	0.9981 (0.00)	0.0000
	13	0.0431 (0.24)	0.0551 (0.30)	1.120 (0.24)	0.9972 (0.01)	-0.0001
	33	0.0562 (0.06)	0.0731 (0.08)	1..459 (0.06)	0.9938 (-0.00)	0.0000
	44	0.0973 (0.03)	0.1257 (0.03)	2.526 (0.03)	0.9812 (-0.03)	0.0000
	58	0.1199 (0.05)	0.1561 (0.03)	3.114 (0.05)	0.9712 (-0.00)	0.0001

* Change (in %) in bracket as compared to the ANFIS.

Table B12: Performance of BANFIS at Cluster 2 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48620 (Sitiawan)	1	0.0383 (2.00)	0.0693 (2.49)	1.006 (2.00)	0.9870 (-0.04)	-0.0005
	4	0.0395 (1.67)	0.0700 (2.23)	1.037 (1.67)	0.9869 (-0.03)	-0.0005
	13	0.0457 (2.49)	0.0769 (2.74)	1.199 (2.49)	0.9858 (-0.04)	-0.0003
	33	0.0500 (3.11)	0.0825 (3.35)	1.313 (3.11)	0.9844 (-0.06)	-0.0004
	43	0.0869 (1.47)	0.1293 (1.51)	2.281 (1.47)	0.9699 (-0.07)	0.0001
	58	0.1178 (0.45)	0.1660 (0.85)	3.094 (0.45)	0.9538 (-0.07)	-0.0002
Station 48625 (Ipoh)	1	0.0360 (0.32)	0.0455 (0.22)	0.901 (0.32)	0.9965 (-0.00)	0.0000
	4	0.0374 (0.35)	0.0466 (0.56)	0.936 (0.35)	0.9963 (-0.00)	0.0001
	13	0.0447 (0.42)	0.0563 (0.29)	1.118 (0.42)	0.9947 (-0.00)	0.0000
	33	0.0558 (-0.25)	0.0709 (-0.20)	1.395 (-0.25)	0.9915 (0.00)	0.0001
	43	0.1018 (0.42)	0.1318 (0.33)	2.547 (0.42)	0.9708 (-0.02)	-0.0001
	58	0.1487 (0.03)	0.1891 (0.02)	3.722 (0.03)	0.9398 (-0.00)	0.0001
Station 48647 (Subang)	1	0.0348 (0.10)	0.0440 (0.10)	0.860 (0.10)	0.9980 (-0.00)	0.0001
	4	0.0402 (-0.00)	0.0508 (0.01)	0.994 (-0.00)	0.9973 (0.00)	0.0001
	13	0.0475 (0.22)	0.0606(0.13)	1.175 (0.22)	0.9962 (-0.00)	0.0001
	33	0.0645 (-0.85)	0.0838(-0.96)	1.593 (-0.85)	0.9927 (0.01)	0.0000
	43	0.1112 (0.19)	0.1437 (0.21)	2.749 (0.19)	0.9785 (-0.01)	-0.0001
	58	0.1563 (0.05)	0.2007 (0.05)	3.864 (0.05)	0.9579 (-0.00)	0.0000

* Change (in %) in bracket as compared to the ANFIS.

246

Table B13: Performance of BANFIS at Cluster 3 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
Station 48600 (Pulau Langkawi)	1	0.0503 (2.41)	0.0724 (1.39)	1.149 (2.41)	0.9959 (-0.01)	0.0000
	4	0.0549 (2.85)	0.0785 (3.04)	1.253 (2.85)	0.9951 (-0.03)	0.0000
	13	0.0678 (-0.09)	0.0949 (-0.35)	1.549 (-0.09)	0.9929 (0.01)	-0.0001
	23	0.1156 (0.18)	0.1586 (0.21)	2.639 (0.18)	0.9802 (-0.01)	-0.0002
	44	0.1879 (0.19)	0.2709 (0.18)	4.291 (0.19)	0.9424(-0.02)	0.0000
	58	0.3807 (-0.04)	0.5197 (0.02)	8.693 (-0.04)	0.7878 (-0.00)	-0.0008
Station 48601 (Bayan Lepas)	1	0.0467 (0.24)	0.0659 (0.04)	1.077 (0.24)	0.9959 (-0.00)	0.0000
	4	0.0534 (2.32)	0.0721 (2.63)	1.242 (2.32)	0.9951 (-0.03)	0.0001
	13	0.0620 (-3.82)	0.0833 (-3.89)	1.428 (-3.82)	0.9935 (0.05)	0.0000
	23	0.1096 (0.03)	0.1448 (-0.13)	2.526 (0.03)	0.9803 (-0.00)	0.0001
	44	0.1516 (0.05)	0.2035 (0.05)	3.494 (0.05)	0.9613 (-0.00)	0.0000
	58	0.2736 (0.07)	0.3969 (0.04)	6.306 (0.07)	0.8516 (-0.01)	0.0001

* Change (in %) in bracket as compared to the ANFIS.

Table B14: Performance of BANFIS at Cluster 4 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
Station 48623 (Lubok Merbau)	1	0.0344 (0.46)	0.0428 (0.70)	0.846 (0.46)	0.9971 (-0.00)	0.0001
	4	0.0351 (1.50)	0.0436 (2.00)	0.862 (1.50)	0.9970 (-0.01)	0.0000
	17	0.0414 (-0.71)	0.0539 (-1.01)	1.017 (-0.71)	0.9954 (0.01)	0.0001
	33	0.0472 (0.08)	0.0610 (-0.05)	1.160 (0.08)	0.9941 (0.00)	0.0001
	43	0.0864 (0.11)	0.1135 (0.14)	2.124 (0.11)	0.9796 (-0.01)	0.0001
	58	0.1102 (0.13)	0.1478 (0.08)	2.708 (0.13)	0.9654 (-0.00)	0.0003
Station 48649 (Muadzam Shah)	1	0.0316 (-0.62)	0.0436 (-1.02)	0.907 (-0.62)	0.9971 (0.00)	0.0001
	4	0.0316 (-0.21)	0.0440 (-0.61)	0.909 (-0.21)	0.9970 (0.00)	0.0001
	17	0.0353 (0.99)	0.0507 (1.04)	1.015 (0.99)	0.9962 (-0.01)	0.0001
	33	0.0411 (0.45)	0.0578 (0.34)	1.182 (0.45)	0.9951 (-0.00)	0.0001
	43	0.0610 (0.19)	0.0826 (0.11)	1.754 (0.19)	0.9903 (-0.00)	0.0000
	58	0.0850 (0.11)	0.1146 (0.07)	2.444 (0.11)	0.9815 (-0.00)	0.0000

* Change (in %) in bracket as compared to the ANFIS.

Table B15: Performance of BANFIS at Cluster 5 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
Station 48632 (Cameron Highlands)	1	0.0369 (0.89)	0.0494 (0.70)	1.167 (0.89)	0.9979 (-0.00)	0.0000
	4	0.0373 (0.36)	0.0500 (0.37)	1.179 (0.36)	0.9978 (-0.00)	0.0000
	13	0.0428 (-4.16)	0.0583 (-4.18)	1.352 (-4.16)	0.9970 (0.03)	-0.0001
	33	0.0468 (0.90)	0.0647 (1.04)	1.481 (0.90)	0.9963 (-0.01)	0.0001
	53	0.0626 (0.33)	0.0964 (0.20)	1.980 (0.33)	0.9917 (-0.00)	0.0001
	58	0.3021 (0.03)	0.3798 (0.07)	9.551 (0.03)	0.8736 (-0.01)	0.0000

* Change (in %) in bracket as compared to the ANFIS.

Appendix C: Performance of BMA-E at Different Stations

Table C1: Performance of BMA-E at Cluster 1 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48603 (Alor Setar)	1	0.0277	0.0357	0.648	0.9989	-0.0001
	4	0.0400	0.0514	0.935	0.9977	-0.0001
	13	0.0508	0.0654	1.186	0.9963	-0.0001
	33	0.0758	0.1063	1.771	0.9901	-0.0011
	44	0.1373	0.1826	3.208	0.9707	-0.0026
	58	0.2588	0.3620	6.047	0.8848	-0.0005
Station 48615 (Kota Bharu)	1	0.0236	0.0297	0.543	0.9992	0.0000
	4	0.0480	0.0651	1.106	0.9959	-0.0002
	13	0.0566	0.0779	1.303	0.9942	-0.0007
	33	0.0963	0.1399	2.217	0.9813	-0.0001
	44	0.1597	0.2157	3.678	0.9556	-0.0014
	58	0.2258	0.3078	5.201	0.9095	-0.0021
Station 48650 (KLIA)	1	0.0262	0.0308	0.646	0.9990	-0.0001
	4	0.0395	0.0502	0.976	0.9974	-0.0001
	13	0.0459	0.0597	1.134	0.9964	0.0001
	33	0.0785	0.1084	1.939	0.9881	-0.0002
	44	0.1271	0.1725	3.139	0.9698	0.0010
	58	0.1928	0.2821	4.762	0.9192	-0.0010
Station 48657 (Kuantan)	1	0.0235	0.0289	0.611	0.9990	0.0001
	4	0.0321	0.0402	0.833	0.9981	-0.0001
	13	0.0388	0.0494	1.008	0.9971	-0.0002
	33	0.0553	0.0721	1.435	0.9939	-0.0005
	44	0.0969	0.1251	2.515	0.9815	-0.0002
	58	0.1196	0.1559	3.106	0.9712	-0.0017

Table C2: Performance of BMA-E at Cluster 2 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48620 (Sitiawan)	1	0.0325	0.0585	0.855	0.9897	-0.0070
	4	0.0349	0.0602	0.916	0.9900	-0.0043
	13	0.0399	0.0659	1.048	0.9893	-0.0056
	33	0.0452	0.0730	1.187	0.9877	-0.0020
	43	0.0854	0.1256	2.242	0.9717	-0.0031
	58	0.1156	0.1613	3.035	0.9568	-0.0081
Station 48625 (Ipoh)	1	0.0287	0.0344	0.717	0.9980	-0.0001
	4	0.0324	0.0395	0.810	0.9974	0.0002
	13	0.0411	0.0512	1.027	0.9956	-0.0005
	33	0.0551	0.0700	1.378	0.9918	-0.0008
	43	0.1014	0.1313	2.536	0.9710	0.0000
	58	0.1486	0.1891	3.718	0.9399	-0.0004
Station 48647 (Subang)	1	0.0273	0.0322	0.675	0.9989	-0.0001
	4	0.0361	0.0447	0.893	0.9979	0.0003
	13	0.0433	0.0548	1.070	0.9969	-0.0001
	33	0.0640	0.0831	1.582	0.9928	-0.0001
	43	0.1110	0.1433	2.744	0.9786	0.0005
	58	0.1561	0.2004	3.858	0.9580	0.0005

Table C3: Performance of BMA-E at Cluster 3 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48600 (Pulau Langkawi)	1	0.0280	0.0365	0.640	0.9990	0.0000
	4	0.0445	0.0638	1.015	0.9968	-0.0009
	13	0.0631	0.0845	1.442	0.9945	-0.0104
	23	0.1131	0.1547	2.583	0.9812	-0.0031
	44	0.1872	0.2699	4.275	0.9428	-0.0003
	58	0.3803	0.5195	8.686	0.7879	-0.0013
Station 48601 (Bayan Lepas)	1	0.0275	0.0352	0.634	0.9988	0.0000
	4	0.0437	0.0578	1.007	0.9969	-0.0002
	13	0.0544	0.0724	1.253	0.9951	0.0001
	23	0.1060	0.1390	2.443	0.9819	0.0011
	44	0.1479	0.1976	3.409	0.9635	-0.0012
	58	0.2731	0.3966	6.294	0.8519	-0.0008

Table C4: Performance of BMA-E at Cluster 4 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48623 (Lubok Merbau)	1	0.0302	0.0370	0.744	0.9978	0.0002
	4	0.0317	0.0391	0.779	0.9976	0.0004
	17	0.0379	0.0490	0.931	0.9962	-0.0012
	33	0.0459	0.0591	1.130	0.9945	-0.0001
	43	0.0858	0.1127	2.110	0.9799	-0.0008
	58	0.1097	0.1474	2.697	0.9656	-0.0002
Station 48649 (Muadzam Shah)	1	0.0260	0.0352	0.746	0.9980	-0.0003
	4	0.0280	0.0382	0.804	0.9977	0.0003
	17	0.0342	0.0486	0.982	0.9965	0.0001
	33	0.0398	0.0559	1.143	0.9955	-0.0005
	43	0.0609	0.0824	1.750	0.9904	0.0000
	58	0.0849	0.1145	2.441	0.9815	-0.0010

Table C5: Performance of BMA-E at Cluster 5 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48632 (Cameron Highlands)	1	0.0270	0.0323	0.854	0.9991	0.0000
	4	0.0295	0.0359	0.932	0.9989	0.0001
	13	0.0354	0.0460	1.120	0.9981	0.0002
	33	0.0451	0.0621	1.425	0.9966	-0.0002
	53	0.0623	0.0960	1.969	0.9918	0.0000
	58	0.3019	0.3786	9.546	0.8744	-0.0011

Appendix D: Performance of WOA-ELM-E at Different Stations and GPI Scores of Different Machine Learning Models

Table D1: Performance of WOA-ELM-E at Cluster 1 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
Station 48603 (Alor Setar)	1	0.0277	0.0357	0.648	0.9989	0.0000
	4	0.0401	0.0514	0.936	0.9977	0.0000
	13	0.0507	0.0652	1.184	0.9962	0.0000
	33	0.0759	0.1061	1.774	0.9900	-0.0002
	44	0.1370	0.1820	3.201	0.9708	0.0000
	58	0.2586	0.3613	6.044	0.8845	0.0000
Station 48615 (Kota Bharu)	1	0.0235	0.0294	0.542	0.9992	0.0000
	4	0.0479	0.0650	1.104	0.9960	0.0000
	13	0.0560	0.0769	1.290	0.9943	0.0000
	33	0.0956	0.1393	2.203	0.9812	0.0000
	44	0.1593	0.2149	3.668	0.9556	0.0003
	58	0.2260	0.3075	5.204	0.9090	0.0000
Station 48650 (KLIA)	1	0.0262	0.0308	0.647	0.9990	0.0000
	4	0.0395	0.0501	0.975	0.9975	-0.0001
	13	0.0458	0.0596	1.132	0.9964	0.0000
	33	0.0784	0.1079	1.936	0.9881	0.0001
	44	0.1271	0.1728	3.139	0.9697	-0.0001
	58	0.1929	0.2816	4.764	0.9192	0.0001
Station 48657 (Kuantan)	1	0.0235	0.0289	0.611	0.9990	0.0000
	4	0.0320	0.0401	0.830	0.9981	0.0000
	13	0.0387	0.0490	1.005	0.9971	0.0000
	33	0.0552	0.0720	1.433	0.9939	-0.0001
	44	0.0968	0.1251	2.514	0.9814	-0.0001
	58	0.1197	0.1557	3.108	0.9712	0.0001

Table D2: Performance of WOA-ELM-E at Cluster 2 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R²	MBE
Station 48620 (Sitiawan)	1	0.0299	0.0392	0.786	0.9975	0.0001
	4	0.0319	0.0393	0.838	0.9976	0.0000
	13	0.0376	0.0472	0.988	0.9965	0.0000
	33	0.0435	0.0547	1.141	0.9953	0.0000
	43	0.0836	0.1082	2.194	0.9815	0.0000
	58	0.1142	0.1454	2.998	0.9666	0.0000
Station 48625 (Ipoh)	1	0.0287	0.0344	0.717	0.9980	0.0000
	4	0.0324	0.0395	0.810	0.9974	0.0000
	13	0.0410	0.0511	1.026	0.9956	0.0000
	33	0.0550	0.0698	1.375	0.9918	0.0000
	43	0.1014	0.1313	2.538	0.9708	0.0000
	58	0.1487	0.1890	3.720	0.9394	0.0001
Station 48647 (Subang)	1	0.0273	0.0322	0.675	0.9989	0.0000
	4	0.0362	0.0448	0.894	0.9979	0.0000
	13	0.0433	0.0548	1.071	0.9968	0.0000
	33	0.0640	0.0830	1.582	0.9928	0.0000
	43	0.1108	0.1430	2.739	0.9786	-0.0003
	58	0.1558	0.2000	3.851	0.9580	0.0000

Table D3: Performance of WOA-ELM-E at Cluster 3 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48600 (Pulau Langkawi)	1	0.0280	0.0365	0.640	0.9990	0.0000
	4	0.0444	0.0634	1.014	0.9968	0.0000
	13	0.0598	0.0821	1.365	0.9947	0.0000
	23	0.1123	0.1534	2.565	0.9815	0.0001
	44	0.1871	0.2698	4.273	0.9427	0.0001
	58	0.3804	0.5191	8.687	0.7880	0.0000
Station 48601 (Bayan Lepas)	1	0.0275	0.0351	0.635	0.9988	0.0000
	4	0.0437	0.0577	1.006	0.9969	0.0000
	13	0.0540	0.0716	1.245	0.9952	0.0001
	23	0.1058	0.1389	2.440	0.9818	0.0000
	44	0.1479	0.1974	3.410	0.9632	-0.0002
	58	0.2733	0.3955	6.299	0.8518	-0.0001

Table D4: Performance of WOA-ELM-E at Cluster 4 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48623 (Lubok Merbau)	1	0.0303	0.0370	0.744	0.9978	0.0000
	4	0.0317	0.0390	0.780	0.9976	0.0000
	17	0.0378	0.0488	0.930	0.9962	0.0000
	33	0.0459	0.0589	1.128	0.9945	0.0000
	43	0.0858	0.1127	2.109	0.9799	-0.0001
	58	0.1098	0.1473	2.701	0.9655	0.0002
Station 48649 (Muadzam Shah)	1	0.0257	0.0326	0.740	0.9985	-0.0002
	4	0.0286	0.0379	0.823	0.9979	0.0001
	17	0.0337	0.0450	0.968	0.9972	0.0000
	33	0.0401	0.0562	1.151	0.9955	-0.0001
	43	0.0614	0.0822	1.764	0.9905	0.0000
	58	0.0848	0.1131	2.438	0.9820	0.0000

Table D5: Performance of WOA-ELM-E at Cluster 5 Stations

Stations	Combinations	MAE (mm/day)	RMSE (mm/day)	MAPE (%)	R^2	MBE
Station 48632 (Cameron Highlands)	1	0.0270	0.0323	0.853	0.9991	0.0000
	4	0.0294	0.0358	0.931	0.9989	0.0000
	13	0.0353	0.0459	1.115	0.9981	0.0000
	33	0.0445	0.0607	1.407	0.9967	0.0001
	53	0.0620	0.0956	1.960	0.9919	0.0000
	58	0.3007	0.3781	9.506	0.8740	0.0000

Table D6: GPI Score of Different Models at Cluster 1 Stations

Stations	Combinations	MLP	SVM	ANFIS	BMLP	BSVM	BANFIS	BMA-E	WOA-ELM-E
Station 48603 (Alor Setar)	1	1.192	-2.183	-1.730	1.212	-2.304	-1.786	1.192	1.197
	4	0.801	-1.703	-2.109	0.744	-1.807	-2.198	0.802	0.791
	13	1.050	-1.799	-1.914	0.776	-1.870	-1.903	1.050	1.082
	33	0.505	-1.210	-1.031	0.446	-1.296	-2.279	0.507	0.643
	44	-0.539	-0.746	-0.015	0.524	-0.851	-0.940	0.994	2.116
	58	0.134	-1.771	0.013	0.100	-1.759	-0.069	0.172	-0.007
Station 48615 (Kota Bharu)	1	0.828	-1.231	-2.199	0.796	-1.841	-2.017	0.829	0.848
	4	0.468	-1.255	-2.395	0.423	-0.779	-2.423	0.467	0.623
	13	-0.640	-0.067	-2.359	0.225	-0.831	-2.530	-0.646	0.560
	33	-0.010	-0.598	-0.092	-0.714	-0.760	-0.639	1.735	2.338
	44	0.490	-0.498	-1.066	-0.145	-0.655	-1.324	0.856	1.450
	58	-0.020	-1.804	0.205	-0.188	-1.870	0.088	0.185	-0.416
Station 48650 (KLIA)	1	1.057	-2.025	-1.944	1.038	-2.059	-1.886	1.056	1.054
	4	0.429	-1.328	-2.263	0.396	-1.076	-2.529	0.425	0.449
	13	0.370	-1.287	-2.522	0.274	-1.207	-2.610	0.372	0.398
	33	0.419	-1.162	-2.408	0.214	-1.198	-2.128	0.420	0.583
	44	1.170	-2.269	0.252	-0.195	-2.551	-0.756	1.180	1.019
	58	-0.072	-1.800	0.068	-0.087	-1.790	0.041	0.124	0.164
Station 48657 (Kuantan)	1	1.262	-2.620	-1.223	1.233	-2.710	-1.291	1.265	1.270
	4	1.065	-2.062	-1.858	1.100	-1.997	-1.892	1.066	1.096
	13	0.720	-1.707	-1.900	1.019	-1.711	-1.979	0.723	0.865
	33	0.803	-2.457	-0.454	0.669	-2.852	-0.544	0.814	1.118
	44	1.347	-1.210	0.145	-1.087	-1.572	0.029	1.400	1.363
	58	0.508	-2.693	-0.053	0.315	-2.731	-0.284	0.767	0.960

Table D7: GPI Score of Different Models at Cluster 2 Stations

Stations	Combinations	MLP	SVM	ANFIS	BMLP	BSVM	BANFIS	BMA-E	WOA-ELM-E
Station 48620 (Sitiawan)	1	-0.314	-0.563	-0.414	0.680	-0.611	-0.640	-0.023	2.402
	4	-0.130	-0.767	-0.827	0.157	-0.845	-1.097	-0.766	2.000
	13	-0.710	-0.564	-0.736	-0.172	-0.582	-1.039	-0.563	2.011
	33	-1.376	0.076	-0.158	-0.131	0.131	-0.667	0.077	2.404
	43	-0.042	-0.548	0.198	-0.679	-0.563	-0.351	-0.535	2.665
	58	-0.091	-0.725	-0.562	-0.218	-0.762	-0.839	-0.726	2.180
Station 48625 (Ipoh)	1	1.039	-1.964	-1.918	0.955	-2.058	-1.952	1.040	1.041
	4	0.898	-1.796	-1.966	0.832	-1.915	-2.073	0.894	0.915
	13	0.614	-1.365	-2.124	0.549	-1.564	-2.249	0.610	0.741
	33	0.866	-0.935	-1.593	0.654	-1.702	-1.255	0.562	1.414
	43	-0.668	-2.834	0.959	-0.410	-2.902	0.410	0.961	0.774
	58	0.243	-2.106	-0.081	-0.224	-2.375	-0.249	0.214	-0.490
Station 48647 (Subang)	1	0.953	-2.904	-0.925	0.929	-2.994	-0.934	0.953	0.950
	4	1.047	-2.294	-0.699	1.231	-2.341	-0.889	1.051	1.483
	13	1.242	-1.948	-1.659	0.983	-2.112	-1.747	1.251	1.243
	33	0.689	-2.488	-0.949	-0.064	-3.177	0.012	0.689	0.736
	43	-0.701	-2.147	0.733	-0.127	-2.634	0.045	0.623	1.023
	58	0.546	-1.580	-0.156	-0.352	-1.763	-0.511	0.624	1.249

Table D8: GPI Score of Different Models at Cluster 3 Stations

Stations	Combinations	MLP	SVM	ANFIS	BMLP	BSVM	BANFIS	BMA-E	WOA-ELM-E
Station 48600 (Pulau Langkawi)	1	0.986	-1.943	-1.900	0.940	-1.948	-2.015	0.985	0.982
	4	0.671	-1.538	-1.718	0.639	-1.584	-2.183	0.670	0.814
	13	0.096	-0.949	-2.186	0.646	-0.923	-2.124	-0.932	0.762
	23	-1.593	-0.607	-1.838	0.032	-0.848	-2.064	-0.606	0.954
	44	0.388	-1.287	-0.052	-0.200	-1.507	-0.839	0.752	0.453
	58	-0.035	-0.951	0.035	-0.297	-1.028	-0.129	0.277	0.746
Station 48601 (Bayan Lepas)	1	0.950	-1.929	-2.041	0.929	-1.987	-2.048	0.951	0.949
	4	0.616	-1.425	-1.819	0.510	-1.345	-2.306	0.614	0.786
	13	0.283	-1.265	-2.587	0.452	-1.198	-1.847	0.284	0.413
	23	0.587	-0.838	-2.074	0.238	-1.041	-2.005	0.589	0.899
	44	0.703	-0.939	-2.124	0.063	-1.136	-2.165	0.702	0.693
	58	0.300	-1.532	0.033	-0.072	-1.602	-0.120	0.336	0.262

Table D9: GPI Score of Different Models at Cluster 4 Stations

Stations	Combinations	MLP	SVM	ANFIS	BMLP	BSVM	BANFIS	BMA-E	WOA-ELM-E
Station 48623 (Lubok Merbau)	1	0.936	-2.924	-0.828	0.841	-3.046	-0.924	0.940	0.949
	4	0.822	-2.991	-0.736	0.771	-3.098	-1.098	0.816	0.878
	17	1.029	-1.929	-1.676	1.228	-2.065	-1.412	1.025	1.281
	33	0.977	-2.819	-0.662	0.670	-2.802	-0.672	0.971	1.082
	43	0.528	-3.117	0.161	0.020	-3.142	0.013	0.544	0.710
	58	0.517	-3.274	-0.121	0.175	-3.403	-0.392	0.459	0.223
Station 48649 (Muadzam Shah)	1	1.030	-2.135	-1.098	0.976	-2.292	-1.015	1.033	1.605
	4	1.277	-2.217	-0.986	0.873	-2.332	-0.906	1.274	1.456
	17	0.392	-2.309	-0.190	0.211	-2.339	-0.413	0.394	1.538
	33	1.607	-1.647	-0.397	0.486	-1.993	-0.671	1.595	1.570
	43	-0.054	-0.900	0.331	-3.066	-1.416	0.246	0.444	0.025
	58	-0.614	-0.831	0.129	-0.770	-0.828	-0.191	0.132	2.112

Table D10: GPI Score of Different Models at Cluster 5 Stations

Stations	Combinations	MLP	SVM	ANFIS	BMLP	BSVM	BANFIS	BMA-E	WOA-ELM-E
Station 48632 (Cameron Highlands)	1	1.218	-2.663	-1.202	1.187	-2.621	-1.256	1.219	1.227
	4	1.301	-2.420	-1.327	1.247	-2.676	-1.373	1.298	1.324
	13	1.040	-1.962	-1.932	0.893	-2.132	-1.366	1.040	1.080
	33	0.618	-2.402	-0.603	0.963	-2.501	-1.016	0.618	1.421
	53	-0.123	-2.760	0.329	-0.223	-2.926	0.103	0.487	1.067
	58	0.180	-1.728	-0.039	0.096	-1.754	-0.084	0.194	0.203

Appendix E: Results for Scenario 2 and Scenario 3

Table E1: MAE (in mm/day) of Local Best Models when Estimating ET$_0$ at External Stations using Six Meteorological Variables

| | | Testing Station | | | | | | | | | | | |
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.0280	0.0964	0.0500	0.0508	0.0304	0.1121	0.1060	0.1558	0.0689	0.0535	0.0573	0.0400
	Station 48601 (Bayan Lepas)	0.1115	0.0275	0.0526	0.0449	0.0564	0.1984	0.2006	0.3147	0.1711	0.1004	0.1544	0.1036
	Station 48603 (Alor Setar)	0.0581	0.0595	0.0277	0.0367	0.0326	0.1452	0.1444	0.2918	0.1104	0.0562	0.0989	0.0678
	Station 48615 (Kota Bharu)	0.0404	0.0308	0.0357	0.0235	0.0281	0.0497	0.0486	0.0788	0.0374	0.0365	0.0387	0.0298
	Station 48620 (Sitiawan)	0.0628	0.0448	0.0457	0.0503	0.0299	0.0399	0.0378	0.0614	0.0330	0.0331	0.0329	0.0310
	Station 48623 (Lubok Merbau)	0.0670	0.0689	0.0568	0.0665	0.0335	0.0303	0.0297	0.0539	0.0325	0.0301	0.0380	0.0335
	Station 48625 (Ipoh)	0.0597	0.0797	0.0570	0.0569	0.0302	0.0313	0.0287	0.0563	0.0301	0.0291	0.0323	0.0292
	Station 48632 (Cameron Highlands)	0.1508	0.1217	0.1004	0.0991	0.0383	0.0400	0.0476	0.0270	0.0573	0.0316	0.0758	0.0465
	Station 48647 (Subang)	0.0876	0.1536	0.0995	0.0910	0.0432	0.0690	0.0536	0.2419	0.0273	0.0297	0.0282	0.0270
	Station 48649 (Muadzam Shah)	0.0575	0.0496	0.0440	0.0500	0.0321	0.0334	0.0301	0.0508	0.0301	0.0257	0.0289	0.0264
	Station 48650 (KLIA)	0.0781	0.1515	0.0930	0.0950	0.0381	0.0815	0.0602	0.1618	0.0281	0.0374	0.0262	0.0259
	Station 48657 (Kuantan)	0.0365	0.0497	0.0408	0.0385	0.0283	0.0400	0.0346	0.0555	0.0286	0.0273	0.0282	0.0235

Table E2: RMSE (in mm/day) of Local Best Models when Estimating ET$_0$ at External Stations using Six Meteorological Variables

| | | Testing Station | | | | | | | | | | | |
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.0365	0.1206	0.0638	0.0638	0.0831	0.1356	0.1247	0.1913	0.0845	0.0715	0.0724	0.0496
	Station 48601 (Bayan Lepas)	0.1538	0.0352	0.0723	0.0583	0.1009	0.2266	0.2255	0.3621	0.1882	0.1179	0.1812	0.1145
	Station 48603 (Alor Setar)	0.0953	0.0752	0.0356	0.0556	0.0865	0.1730	0.1663	0.3754	0.1275	0.0744	0.1204	0.0792
	Station 48615 (Kota Bharu)	0.0518	0.0407	0.0457	0.0294	0.0351	0.0647	0.0614	0.1067	0.0473	0.0493	0.0489	0.0376
	Station 48620 (Sitiawan)	0.0993	0.0696	0.0654	0.0804	0.0392	0.0498	0.0470	0.0883	0.0408	0.0460	0.0410	0.0391
	Station 48623 (Lubok Merbau)	0.0986	0.1070	0.0794	0.0928	0.0418	0.0370	0.0360	0.0810	0.0405	0.0405	0.0525	0.0432
	Station 48625 (Ipoh)	0.0856	0.1126	0.0763	0.0779	0.0375	0.0385	0.0344	0.0754	0.0371	0.0364	0.0422	0.0369
	Station 48632 (Cameron Highlands)	0.2127	0.1916	0.1566	0.1384	0.0495	0.0525	0.0613	0.0323	0.0766	0.0406	0.1132	0.0616
	Station 48647 (Subang)	0.1284	0.1874	0.1295	0.1083	0.0937	0.0872	0.0671	0.2862	0.0322	0.0425	0.0356	0.0326
	Station 48649 (Muadzam Shah)	0.0944	0.0669	0.0571	0.0758	0.0394	0.0418	0.0365	0.0657	0.0371	0.0326	0.0361	0.0329
	Station 48650 (KLIA)	0.0973	0.1823	0.1194	0.1127	0.0896	0.0994	0.0737	0.2144	0.0343	0.0525	0.0308	0.0316
	Station 48657 (Kuantan)	0.0496	0.0659	0.0534	0.0513	0.0342	0.0500	0.0425	0.0731	0.0340	0.0351	0.0339	0.0289

Table E3: MAPE (in %) of Local Best Models when Estimating ET$_0$ at External Stations using Six Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.640	2.502	1.297	1.318	0.788	2.910	2.752	4.046	1.789	1.389	1.487	1.038
	Station 48601 (Bayan Lepas)	2.895	0.634	1.367	1.166	1.464	5.151	5.208	8.170	4.441	2.607	4.008	2.689
	Station 48603 (Alor Setar)	1.508	1.545	0.646	0.952	0.847	3.771	3.749	7.576	2.866	1.459	2.568	1.761
	Station 48615 (Kota Bharu)	1.049	0.799	0.926	0.542	0.731	1.289	1.261	2.046	0.970	0.947	1.005	0.775
	Station 48620 (Sitiawan)	1.629	1.162	1.186	1.306	0.786	1.035	0.982	1.594	0.856	0.860	0.854	0.805
	Station 48623 (Lubok Merbau)	1.740	1.788	1.475	1.726	0.869	0.744	0.772	1.400	0.844	0.781	0.988	0.870
	Station 48625 (Ipoh)	1.551	2.068	1.481	1.478	0.784	0.812	0.717	1.462	0.781	0.756	0.839	0.757
	Station 48632 (Cameron Highlands)	3.914	3.159	2.606	2.574	0.994	1.039	1.235	0.853	1.488	0.820	1.969	1.208
	Station 48647 (Subang)	2.273	3.988	2.584	2.363	1.120	1.792	1.392	6.279	0.675	0.771	0.731	0.702
	Station 48649 (Muadzam Shah)	1.492	1.288	1.141	1.297	0.834	0.867	0.781	1.319	0.780	0.740	0.750	0.684
	Station 48650 (KLIA)	2.028	3.934	2.415	2.466	0.989	2.117	1.562	4.201	0.730	0.971	0.646	0.673
	Station 48657 (Kuantan)	0.947	1.291	1.060	0.999	0.735	1.039	0.897	1.441	0.744	0.708	0.733	0.611

256

Table E4: R^2 of Local Best Models when Estimating ET_0 at External Stations using Six Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.9990	0.9957	0.9983	0.9977	0.9892	0.9931	0.9937	0.9834	0.9977	0.9959	0.9981	0.9985
	Station 48601 (Bayan Lepas)	0.9941	0.9988	0.9979	0.9983	0.9885	0.9881	0.9881	0.9703	0.9959	0.9951	0.9943	0.9975
	Station 48603 (Alor Setar)	0.9952	0.9978	0.9989	0.9968	0.9888	0.9912	0.9917	0.9383	0.9970	0.9960	0.9966	0.9980
	Station 48615 (Kota Bharu)	0.9979	0.9984	0.9982	0.9992	0.9980	0.9934	0.9936	0.9900	0.9975	0.9964	0.9976	0.9983
	Station 48620 (Sitiawan)	0.9923	0.9954	0.9962	0.9938	0.9975	0.9961	0.9962	0.9931	0.9982	0.9969	0.9983	0.9982
	Station 48623 (Lubok Merbau)	0.9923	0.9893	0.9945	0.9918	0.9973	0.9978	0.9978	0.9942	0.9982	0.9976	0.9972	0.9978
	Station 48625 (Ipoh)	0.9943	0.9881	0.9949	0.9942	0.9978	0.9977	0.9980	0.9950	0.9986	0.9981	0.9982	0.9984
	Station 48632 (Cameron Highlands)	0.9645	0.9660	0.9784	0.9817	0.9961	0.9956	0.9937	0.9991	0.9936	0.9977	0.9871	0.9955
	Station 48647 (Subang)	0.9943	0.9915	0.9952	0.9961	0.9882	0.9960	0.9970	0.9705	0.9989	0.9978	0.9988	0.9988
	Station 48649 (Muadzam Shah)	0.9930	0.9958	0.9971	0.9945	0.9976	0.9972	0.9977	0.9962	0.9986	0.9985	0.9987	0.9987
	Station 48650 (KLIA)	0.9974	0.9924	0.9959	0.9965	0.9884	0.9961	0.9971	0.9777	0.9989	0.9977	0.9990	0.9988
	Station 48657 (Kuantan)	0.9981	0.9959	0.9975	0.9975	0.9982	0.9961	0.9970	0.9953	0.9988	0.9982	0.9988	0.9990

Table E5: MBE of Local Best Models when Estimating ET₀ at External Stations using Six Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.0000	-0.0940	-0.0401	-0.0391	-0.0071	0.1116	0.1048	-0.0766	0.0673	0.0459	0.0547	0.0344
	Station 48601 (Bayan Lepas)	0.1098	0.0000	0.0490	0.0398	0.0524	0.1984	0.2003	0.3117	0.1710	0.0987	0.1543	0.1033
	Station 48603 (Alor Setar)	0.0496	-0.0538	0.0000	-0.0055	0.0177	0.1448	0.1438	0.0542	0.1099	0.0491	0.0985	0.0669
	Station 48615 (Kota Bharu)	0.0003	0.0000	0.0000	0.0000	-0.0001	0.0000	0.0001	0.0004	0.0000	-0.0001	0.0001	0.0000
	Station 48620 (Sitiawan)	-0.0004	-0.0004	0.0001	-0.0001	0.0001	0.0001	0.0000	-0.0001	-0.0001	-0.0002	-0.0001	-0.0001
	Station 48623 (Lubok Merbau)	-0.0001	0.0001	-0.0002	0.0001	0.0000	0.0000	0.0000	0.0004	0.0000	-0.0002	0.0001	0.0001
	Station 48625 (Ipoh)	0.0000	-0.0001	0.0000	0.0001	0.0000	0.0000	0.0000	0.0002	0.0000	-0.0001	-0.0001	0.0000
	Station 48632 (Cameron Highlands)	0.0001	0.0002	0.0002	-0.0002	-0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0002	0.0001
	Station 48647 (Subang)	-0.0819	-0.1530	-0.0969	-0.0836	-0.0354	0.0674	0.0504	-0.1950	-0.0001	0.0140	-0.0042	-0.0040
	Station 48649 (Muadzam Shah)	0.0000	0.0001	-0.0001	-0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0002	0.0001	0.0000
	Station 48650 (KLIA)	-0.0718	-0.1508	-0.0891	-0.0922	-0.0260	0.0811	0.0588	0.1410	0.0088	0.0311	-0.0001	0.0021
	Station 48657 (Kuantan)	0.0001	-0.0001	0.0001	0.0000	0.0001	-0.0001	0.0000	-0.0002	0.0000	-0.0002	0.0000	0.0000

Table E6: GPI Score of Local Best Models when Estimating ET$_0$ at External Stations using Six Meteorological Variables

	Testing Station											
	Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Station 48600 (Pulau Langkawi)	1.008	-1.480	-0.317	-0.256	-1.435	-2.159	-1.786	-1.209	-1.264	-1.924	-0.926	-0.665
Station 48601 (Bayan Lepas)	-2.170	0.880	-0.569	-0.030	-4.479	-4.611	-4.456	-3.792	-4.279	-4.506	-4.124	-3.954
Station 48603 (Alor Setar)	-0.399	-0.285	0.998	0.500	-1.892	-3.107	-2.851	-3.370	-2.495	-2.022	-2.306	-2.157
Station 48615 (Kota Bharu)	0.661	0.757	0.627	1.308	0.460	-0.451	-0.282	-0.067	-0.220	-0.577	-0.171	0.019
Station 48620 (Sitiawan)	-0.131	0.259	0.098	-0.199	0.208	0.019	0.188	0.161	0.000	-0.313	0.032	-0.069
Station 48623 (Lubok Merbau)	-0.193	-0.543	-0.403	-0.849	-0.096	0.389	0.496	0.252	0.010	0.030	-0.221	-0.292
Station 48625 (Ipoh)	0.056	-0.785	-0.364	-0.325	0.253	0.343	0.544	0.264	0.138	0.263	0.024	0.060
Station 48632 (Cameron Highlands)	-2.993	-2.621	-3.003	-2.694	-0.666	-0.041	-0.257	0.643	-1.426	0.027	-2.053	-1.487
Station 48647 (Subang)	-1.380	-3.318	-2.934	-2.288	-3.143	-0.869	-0.278	-2.681	0.283	-0.055	0.153	0.238
Station 48649 (Muadzam Shah)	0.007	0.214	0.257	-0.110	0.068	0.257	0.484	0.352	0.138	0.475	0.159	0.271
Station 48650 (KLIA)	-0.868	-3.209	-2.561	-2.499	-2.520	-1.148	-0.417	-1.617	0.187	-0.587	0.276	0.308
Station 48657 (Kuantan)	0.744	0.222	0.391	0.588	0.475	0.013	0.322	0.282	0.222	0.367	0.198	0.470

Training Station

259

Table E7: MAE (in mm/day) of Local Best Models when Estimating ET$_0$ at External Stations using Five Meteorological Variables

| | | Testing Station | | | | | | | | | | | |
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.0444	0.0510	0.0455	0.0585	0.0321	0.0416	0.0423	0.0633	0.0439	0.0362	0.0427	0.0397
	Station 48601 (Bayan Lepas)	0.0622	0.0437	0.0416	0.0600	0.0332	0.0427	0.0516	0.0785	0.0513	0.0371	0.0554	0.0391
	Station 48603 (Alor Setar)	0.0862	0.0602	0.0400	0.0600	0.0372	0.1338	0.1252	0.2009	0.1147	0.0625	0.1082	0.0636
	Station 48615 (Kota Bharu)	0.0587	0.0493	0.0472	0.0479	0.0341	0.0445	0.0436	0.0820	0.0453	0.0431	0.0491	0.0363
	Station 48620 (Sitiawan)	0.1006	0.0503	0.0435	0.0640	0.0319	0.0368	0.0366	0.0627	0.0376	0.0325	0.0491	0.0369
	Station 48623 (Lubok Merbau)	0.0658	0.0672	0.0585	0.0749	0.0332	0.0317	0.0330	0.0486	0.0384	0.0316	0.0483	0.0399
	Station 48625 (Ipoh)	0.0736	0.0824	0.0616	0.0662	0.0323	0.0330	0.0324	0.0564	0.0376	0.0291	0.0453	0.0363
	Station 48632 (Cameron Highlands)	0.1412	0.1210	0.0996	0.1014	0.0404	0.0394	0.0503	0.0294	0.0575	0.0341	0.0767	0.0454
	Station 48647 (Subang)	0.0561	0.0667	0.0555	0.0609	0.0312	0.0366	0.0352	0.0543	0.0362	0.0315	0.0400	0.0341
	Station 48649 (Muadzam Shah)	0.0510	0.0506	0.0519	0.0534	0.0319	0.0349	0.0347	0.0367	0.0386	0.0286	0.0426	0.0337
	Station 48650 (KLIA)	0.0811	0.1471	0.0966	0.1053	0.0383	0.0753	0.0499	0.1126	0.0384	0.0575	0.0395	0.0388
	Station 48657 (Kuantan)	0.0928	0.1309	0.1008	0.0834	0.0385	0.0721	0.0537	0.3378	0.0451	0.0377	0.0516	0.0320

Table E8: RMSE (in mm/day) of Local Best Models when Estimating ET$_0$ at External Stations using Five Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.0634	0.0688	0.0589	0.0775	0.0389	0.0517	0.0527	0.0912	0.0563	0.0479	0.0540	0.0500
	Station 48601 (Bayan Lepas)	0.0885	0.0577	0.0530	0.0906	0.0410	0.0553	0.0653	0.1286	0.0672	0.0491	0.0727	0.0498
	Station 48603 (Alor Setar)	0.1505	0.0774	0.0514	0.0940	0.0889	0.1627	0.1482	0.2344	0.1382	0.0804	0.1347	0.0786
	Station 48615 (Kota Bharu)	0.0817	0.0661	0.0623	0.0650	0.0441	0.0568	0.0542	0.1154	0.0574	0.0557	0.0629	0.0462
	Station 48620 (Sitiawan)	0.1693	0.0675	0.0583	0.0917	0.0393	0.0454	0.0451	0.1007	0.0472	0.0445	0.0683	0.0476
	Station 48623 (Lubok Merbau)	0.0951	0.0998	0.0792	0.1046	0.0413	0.0390	0.0406	0.0719	0.0481	0.0435	0.0661	0.0514
	Station 48625 (Ipoh)	0.1117	0.1280	0.0861	0.0934	0.0400	0.0409	0.0395	0.0837	0.0470	0.0384	0.0618	0.0457
	Station 48632 (Cameron Highlands)	0.2033	0.1861	0.1542	0.1412	0.0522	0.0518	0.0652	0.0358	0.0769	0.0452	0.1127	0.0599
	Station 48647 (Subang)	0.0826	0.0949	0.0736	0.0857	0.0398	0.0450	0.0430	0.0762	0.0448	0.0423	0.0510	0.0428
	Station 48649 (Muadzam Shah)	0.0736	0.0678	0.0678	0.0725	0.0391	0.0432	0.0423	0.0484	0.0483	0.0379	0.0552	0.0424
	Station 48650 (KLIA)	0.1132	0.1819	0.1270	0.1301	0.0889	0.0923	0.0626	0.1479	0.0486	0.0765	0.0502	0.0495
	Station 48657 (Kuantan)	0.1728	0.1846	0.1468	0.1254	0.0901	0.0923	0.0683	0.4161	0.0599	0.0534	0.0757	0.0401

Table E9: MAPE (in %) of Local Best Models when Estimating ET$_0$ at External Stations using Five Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	1.014	1.325	1.181	1.518	0.834	1.080	1.097	1.644	1.140	0.939	1.109	1.032
	Station 48601 (Bayan Lepas)	1.614	1.006	1.080	1.557	0.863	1.109	1.341	2.039	1.333	0.962	1.437	1.014
	Station 48603 (Alor Setar)	2.238	1.563	0.935	1.558	0.965	3.474	3.250	5.217	2.977	1.622	2.810	1.652
	Station 48615 (Kota Bharu)	1.524	1.280	1.225	1.104	0.885	1.155	1.132	2.129	1.177	1.118	1.274	0.942
	Station 48620 (Sitiawan)	2.611	1.305	1.129	1.663	0.838	0.955	0.951	1.628	0.977	0.843	1.275	0.957
	Station 48623 (Lubok Merbau)	1.707	1.744	1.519	1.943	0.862	0.780	0.857	1.261	0.996	0.821	1.253	1.035
	Station 48625 (Ipoh)	1.910	2.138	1.600	1.718	0.839	0.857	0.810	1.464	0.977	0.756	1.176	0.944
	Station 48632 (Cameron Highlands)	3.666	3.140	2.587	2.632	1.048	1.023	1.307	0.931	1.493	0.884	1.990	1.177
	Station 48647 (Subang)	1.457	1.731	1.440	1.581	0.809	0.949	0.913	1.409	0.894	0.817	1.038	0.885
	Station 48649 (Muadzam Shah)	1.324	1.314	1.348	1.386	0.828	0.906	0.901	0.953	1.001	0.823	1.106	0.874
	Station 48650 (KLIA)	2.106	3.818	2.508	2.734	0.995	1.956	1.295	2.923	0.996	1.493	0.976	1.006
	Station 48657 (Kuantan)	2.409	3.399	2.617	2.164	0.998	1.873	1.394	8.770	1.171	0.979	1.339	0.831

Table E10: R^2 of Local Best Models when Estimating ET_0 at External Stations using Five Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.9968	0.9956	0.9970	0.9943	0.9976	0.9958	0.9953	0.9927	0.9967	0.9967	0.9970	0.9971
	Station 48601 (Bayan Lepas)	0.9938	0.9969	0.9975	0.9922	0.9974	0.9952	0.9928	0.9853	0.9953	0.9966	0.9947	0.9971
	Station 48603 (Alor Setar)	0.9879	0.9964	0.9977	0.9918	0.9884	0.9920	0.9925	0.9783	0.9951	0.9964	0.9945	0.9970
	Station 48615 (Kota Bharu)	0.9947	0.9959	0.9966	0.9960	0.9969	0.9949	0.9950	0.9884	0.9964	0.9956	0.9960	0.9975
	Station 48620 (Sitiawan)	0.9775	0.9957	0.9970	0.9920	0.9976	0.9967	0.9965	0.9909	0.9976	0.9971	0.9953	0.9973
	Station 48623 (Lubok Merbau)	0.9928	0.9907	0.9946	0.9895	0.9973	0.9976	0.9972	0.9954	0.9974	0.9972	0.9955	0.9969
	Station 48625 (Ipoh)	0.9902	0.9846	0.9935	0.9917	0.9975	0.9974	0.9974	0.9938	0.9977	0.9979	0.9961	0.9975
	Station 48632 (Cameron Highlands)	0.9676	0.9677	0.9791	0.9809	0.9957	0.9958	0.9929	0.9989	0.9936	0.9971	0.9872	0.9958
	Station 48647 (Subang)	0.9946	0.9915	0.9952	0.9930	0.9975	0.9968	0.9969	0.9949	0.9979	0.9975	0.9974	0.9978
	Station 48649 (Muadzam Shah)	0.9958	0.9957	0.9960	0.9950	0.9976	0.9971	0.9970	0.9979	0.9976	0.9979	0.9969	0.9979
	Station 48650 (KLIA)	0.9939	0.9907	0.9941	0.9930	0.9882	0.9964	0.9963	0.9857	0.9976	0.9954	0.9974	0.9971
	Station 48657 (Kuantan)	0.9804	0.9841	0.9905	0.9878	0.9883	0.9953	0.9960	0.9445	0.9969	0.9969	0.9945	0.9981

Table E11: MBE of Local Best Models when Estimating ET$_0$ at External Stations using Five Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.0000	0.0001	0.0000	0.0002	0.0001	0.0001	0.0000	-0.0002	-0.0001	0.0001	0.0000	0.0001
	Station 48601 (Bayan Lepas)	-0.0002	0.0000	0.0000	-0.0005	0.0002	0.0000	0.0000	0.0005	-0.0001	0.0001	-0.0001	0.0000
	Station 48603 (Alor Setar)	0.0773	-0.0435	-0.0001	0.0079	0.0209	0.1335	0.1242	0.0451	0.1129	0.0576	0.1067	0.0600
	Station 48615 (Kota Bharu)	-0.0001	0.0000	-0.0002	0.0000	0.0003	0.0001	0.0000	0.0007	-0.0001	0.0001	0.0000	0.0000
	Station 48620 (Sitiawan)	0.0006	-0.0001	-0.0001	0.0000	0.0000	0.0000	0.0000	-0.0010	0.0001	-0.0001	-0.0002	0.0001
	Station 48623 (Lubok Merbau)	-0.0004	0.0004	-0.0003	-0.0001	0.0000	0.0000	0.0000	-0.0003	0.0000	-0.0003	0.0000	0.0001
	Station 48625 (Ipoh)	-0.0002	0.0001	0.0000	-0.0001	0.0001	0.0000	0.0000	0.0000	0.0001	-0.0001	0.0000	0.0000
	Station 48632 (Cameron Highlands)	-0.0002	0.0002	0.0005	0.0000	-0.0002	-0.0001	0.0001	0.0000	0.0000	0.0001	-0.0002	0.0000
	Station 48647 (Subang)	0.0002	0.0002	-0.0001	-0.0003	0.0004	0.0000	0.0000	-0.0003	0.0000	-0.0003	-0.0001	0.0000
	Station 48649 (Muadzam Shah)	0.0000	0.0001	0.0000	0.0002	0.0000	-0.0001	0.0000	-0.0001	0.0001	0.0001	0.0000	0.0000
	Station 48650 (KLIA)	-0.0624	-0.1446	-0.0887	-0.0942	-0.0200	0.0723	0.0378	-0.0591	0.0050	0.0460	-0.0001	-0.0056
	Station 48657 (Kuantan)	-0.0598	-0.1263	-0.0942	-0.0615	-0.0257	0.0681	0.0434	-0.2504	0.0165	0.0227	0.0163	0.0001

Table E12: GPI Score of Local Best Models when Estimating ET₀ at External Stations using Five Meteorological Variables

| | | Testing Station | | | | | | | | | | | |
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.815	0.313	0.446	0.409	0.312	-0.020	-0.142	-0.003	-0.143	-0.340	0.364	-0.358
	Station 48601 (Bayan Lepas)	0.120	0.628	0.661	0.041	-0.004	-0.182	-0.975	-0.339	-0.764	-0.467	-0.452	-0.299
	Station 48603 (Alor Setar)	-2.005	-0.199	0.796	-0.109	-3.549	-4.384	-4.384	-1.724	-4.160	-3.805	-3.725	-3.615
	Station 48615 (Kota Bharu)	0.270	0.379	0.339	1.124	-0.303	-0.273	-0.256	-0.270	-0.264	-1.368	-0.023	0.145
	Station 48620 (Sitiawan)	-1.792	0.343	0.521	-0.116	0.306	0.297	0.295	-0.061	0.314	0.111	-0.160	-0.001
	Station 48623 (Lubok Merbau)	-0.035	-0.403	-0.296	-0.811	-0.003	0.616	0.547	0.195	0.259	0.240	-0.090	-0.472
	Station 48625 (Ipoh)	-0.398	-1.117	-0.519	-0.230	0.230	0.519	0.616	0.084	0.349	0.766	0.109	0.173
	Station 48632 (Cameron Highlands)	-3.188	-2.879	-3.172	-2.745	-1.946	0.018	-0.939	0.458	-1.428	0.015	-2.267	-1.524
	Station 48647 (Subang)	0.311	-0.327	-0.106	0.133	0.477	0.316	0.416	0.135	0.475	0.377	0.508	0.521
	Station 48649 (Muadzam Shah)	0.520	0.332	0.102	0.690	0.359	0.409	0.447	0.381	0.283	0.744	0.341	0.578
	Station 48650 (KLIA)	-1.240	-3.550	-2.939	-2.929	-3.791	-1.441	-0.500	-0.838	0.250	-3.588	0.567	-0.343
	Station 48657 (Kuantan)	-2.330	-3.364	-3.516	-2.135	-4.051	-1.539	-0.756	-4.542	-0.305	-0.867	-0.554	0.833

Table E13: MAE (in mm/day) of Local Best Models when Estimating ET_0 at External Stations using Four Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.0598	0.0636	0.0556	0.0614	0.0384	0.0457	0.0438	0.0559	0.0463	0.0353	0.0472	0.0403
	Station 48601 (Bayan Lepas)	0.1170	0.0538	0.0818	0.0864	0.0871	0.2290	0.1997	0.2479	0.1749	0.1383	0.1646	0.1232
	Station 48603 (Alor Setar)	0.0690	0.0584	0.0507	0.0590	0.0376	0.0498	0.0493	0.1200	0.0498	0.0407	0.0551	0.0414
	Station 48615 (Kota Bharu)	0.0673	0.0569	0.0539	0.0560	0.0388	0.0513	0.0469	0.0818	0.0474	0.0387	0.0532	0.0399
	Station 48620 (Sitiawan)	0.0673	0.0605	0.0541	0.0720	0.0376	0.0463	0.0452	0.0641	0.0461	0.0388	0.0521	0.0421
	Station 48623 (Lubok Merbau)	0.1209	0.1094	0.0928	0.1018	0.0389	0.0378	0.0445	0.0463	0.0568	0.0382	0.0785	0.0540
	Station 48625 (Ipoh)	0.0748	0.0838	0.0641	0.0704	0.0391	0.0431	0.0410	0.0505	0.0451	0.0369	0.0509	0.0410
	Station 48632 (Cameron Highlands)	0.1391	0.1063	0.0903	0.0863	0.0457	0.0493	0.0543	0.0353	0.0637	0.0470	0.0718	0.0466
	Station 48647 (Subang)	0.1282	0.1664	0.0977	0.1033	0.0445	0.0989	0.0622	0.0948	0.0433	0.0525	0.0500	0.0428
	Station 48649 (Muadzam Shah)	0.1499	0.1242	0.0968	0.0992	0.0405	0.0419	0.0493	0.0494	0.0589	0.0337	0.0885	0.0620
	Station 48650 (KLIA)	0.0622	0.0684	0.0557	0.0895	0.0376	0.0444	0.0428	0.0562	0.0438	0.0354	0.0458	0.0394
	Station 48657 (Kuantan)	0.1567	0.1911	0.1211	0.1268	0.0491	0.0866	0.0561	0.2059	0.0485	0.0531	0.0579	0.0384

Table E14: RMSE (in mm/day) of Local Best Models when Estimating ET$_0$ at External Stations using Four Meteorological Variables

| | | Testing Station | | | | | | | | | | | |
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.0821	0.0842	0.0720	0.0842	0.0483	0.0586	0.0550	0.0870	0.0589	0.0464	0.0615	0.0514
	Station 48601 (Bayan Lepas)	0.1895	0.0715	0.1143	0.1470	0.1271	0.2662	0.2285	0.3294	0.2003	0.1581	0.1955	0.1395
	Station 48603 (Alor Setar)	0.0953	0.0774	0.0652	0.0829	0.0476	0.0642	0.0618	0.0733	0.0639	0.0540	0.0743	0.0527
	Station 48615 (Kota Bharu)	0.0928	0.0762	0.0696	0.0769	0.0498	0.0674	0.0591	0.1183	0.0609	0.0495	0.0700	0.0510
	Station 48620 (Sitiawan)	0.1028	0.0803	0.0704	0.1018	0.0472	0.0589	0.0569	0.0909	0.0587	0.0514	0.0694	0.0535
	Station 48623 (Lubok Merbau)	0.1860	0.1595	0.1315	0.1434	0.0494	0.0488	0.0578	0.0667	0.0742	0.0550	0.1103	0.0731
	Station 48625 (Ipoh)	0.1078	0.1162	0.0871	0.0985	0.0491	0.0547	0.0511	0.0728	0.0570	0.0498	0.0696	0.0519
	Station 48632 (Cameron Highlands)	0.1993	0.1629	0.1395	0.1240	0.0587	0.0647	0.0713	0.0459	0.0859	0.0644	0.1137	0.0617
	Station 48647 (Subang)	0.1808	0.2129	0.1294	0.1376	0.0937	0.1230	0.0790	0.1312	0.0548	0.0706	0.0730	0.0554
	Station 48649 (Muadzam Shah)	0.2278	0.1883	0.1470	0.1465	0.0512	0.0537	0.0637	0.0671	0.0769	0.0450	0.1332	0.0811
	Station 48650 (KLIA)	0.0865	0.0916	0.0736	0.0895	0.0472	0.0563	0.0536	0.0801	0.0554	0.0477	0.0596	0.0502
	Station 48657 (Kuantan)	0.2360	0.2484	0.1665	0.1705	0.0986	0.1109	0.0724	0.2651	0.0633	0.0714	0.0918	0.0487

Table E15: MAPE (in %) of Local Best Models when Estimating ET₀ at External Stations using Four Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	1.365	1.651	1.442	1.595	0.996	1.187	1.138	1.452	1.201	0.917	1.225	1.048
	Station 48601 (Bayan Lepas)	3.038	1.241	2.124	2.242	2.261	5.945	5.185	6.436	4.541	3.591	4.274	3.198
	Station 48603 (Alor Setar)	1.791	1.517	1.184	1.532	0.977	1.293	1.279	3.115	1.292	1.055	1.432	1.076
	Station 48615 (Kota Bharu)	1.747	1.477	1.398	1.290	1.006	1.332	1.218	2.123	1.231	1.004	1.381	1.037
	Station 48620 (Sitiawan)	1.748	1.572	1.405	1.869	0.988	1.202	1.173	1.665	1.196	1.008	1.354	1.094
	Station 48623 (Lubok Merbau)	3.139	2.840	2.409	2.642	1.009	0.930	1.156	1.201	1.475	0.992	2.038	1.403
	Station 48625 (Ipoh)	1.941	2.177	1.665	1.828	1.016	1.118	1.026	1.312	1.170	0.959	1.321	1.064
	Station 48632 (Cameron Highlands)	3.612	2.761	2.344	2.239	1.186	1.281	1.410	1.115	1.654	1.220	1.864	1.211
	Station 48647 (Subang)	3.328	4.321	2.537	2.681	1.155	2.567	1.614	2.461	1.070	1.363	1.297	1.110
	Station 48649 (Muadzam Shah)	3.893	3.225	2.513	2.575	1.052	1.088	1.280	1.282	1.528	0.968	2.297	1.611
	Station 48650 (KLIA)	1.615	1.776	1.447	2.323	0.976	1.152	1.110	1.460	1.138	0.920	1.132	1.023
	Station 48657 (Kuantan)	4.069	4.962	3.145	3.291	1.276	2.249	1.455	5.346	1.258	1.379	1.503	0.998

Table E16: R^2 of Local Best Models when Estimating ET_0 at External Stations using Four Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.9947	0.9933	0.9955	0.9932	0.9963	0.9946	0.9949	0.9933	0.9964	0.9970	0.9962	0.9969
	Station 48601 (Bayan Lepas)	0.9821	0.9952	0.9940	0.9830	0.9865	0.9855	0.9883	0.9563	0.9930	0.9944	0.9918	0.9960
	Station 48603 (Alor Setar)	0.9929	0.9944	0.9962	0.9934	0.9964	0.9935	0.9936	0.9873	0.9955	0.9958	0.9944	0.9967
	Station 48615 (Kota Bharu)	0.9932	0.9946	0.9958	0.9943	0.9961	0.9928	0.9941	0.9877	0.9959	0.9965	0.9951	0.9969
	Station 48620 (Sitiawan)	0.9917	0.9939	0.9957	0.9901	0.9965	0.9945	0.9945	0.9928	0.9962	0.9963	0.9951	0.9966
	Station 48623 (Lubok Merbau)	0.9727	0.9761	0.9850	0.9803	0.9962	0.9962	0.9943	0.9961	0.9939	0.9957	0.9874	0.9937
	Station 48625 (Ipoh)	0.9909	0.9873	0.9933	0.9907	0.9962	0.9953	0.9956	0.9953	0.9966	0.9965	0.9951	0.9968
	Station 48632 (Cameron Highlands)	0.9688	0.9752	0.9829	0.9852	0.9946	0.9934	0.9915	0.9981	0.9920	0.9941	0.9870	0.9955
	Station 48647 (Subang)	0.9858	0.9854	0.9928	0.9874	0.9873	0.9938	0.9946	0.9882	0.9969	0.9958	0.9948	0.9968
	Station 48649 (Muadzam Shah)	0.9593	0.9667	0.9810	0.9795	0.9959	0.9954	0.9931	0.9960	0.9938	0.9972	0.9820	0.9922
	Station 48650 (KLIA)	0.9941	0.9921	0.9952	0.9924	0.9965	0.9950	0.9951	0.9943	0.9968	0.9968	0.9964	0.9970
	Station 48657 (Kuantan)	0.9768	0.9800	0.9892	0.9845	0.9864	0.9931	0.9944	0.9637	0.9960	0.9957	0.9922	0.9972

Table E17: MBE of Local Best Models when Estimating ET$_0$ at External Stations using Four Meteorological Variables

| | | Testing Station | | | | | | | | | | | |
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.0000	-0.0002	0.0001	0.0000	-0.0001	-0.0001	0.0001	0.0000	-0.0001	0.0000	0.0000	0.0000
	Station 48601 (Bayan Lepas)	0.1009	-0.0001	0.0671	0.0688	0.0827	0.2286	0.1990	-0.0570	0.1743	0.1343	0.1640	0.1216
	Station 48603 (Alor Setar)	-0.0003	0.0000	0.0000	0.0003	-0.0001	-0.0002	-0.0001	0.0002	0.0002	0.0004	-0.0001	0.0000
	Station 48615 (Kota Bharu)	0.0000	0.0000	-0.0001	0.0000	-0.0003	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	Station 48620 (Sitiawan)	0.0007	0.0001	0.0001	0.0000	0.0000	0.0000	0.0001	-0.0002	-0.0001	0.0002	0.0001	-0.0001
	Station 48623 (Lubok Merbau)	-0.0004	-0.0002	0.0003	0.0003	0.0000	0.0000	0.0000	-0.0002	0.0000	0.0002	0.0001	0.0000
	Station 48625 (Ipoh)	-0.0003	0.0001	0.0000	-0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
	Station 48632 (Cameron Highlands)	0.0001	0.0005	0.0004	0.0001	-0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0005	-0.0001
	Station 48647 (Subang)	-0.1042	-0.1622	-0.0858	-0.0704	-0.0275	0.0968	0.0500	0.0293	-0.0001	0.0380	-0.0031	0.0189
	Station 48649 (Muadzam Shah)	0.0005	-0.0008	0.0003	-0.0006	0.0000	0.0004	0.0002	0.0000	-0.0001	0.0000	0.0002	0.0000
	Station 48650 (KLIA)	0.0002	0.0001	0.0002	0.0004	0.0001	0.0000	0.0000	-0.0001	0.0000	-0.0001	0.0000	0.0000
	Station 48657 (Kuantan)	-0.1418	-0.1887	-0.1134	-0.1126	-0.0330	0.0813	0.0401	-0.1097	-0.0109	0.0404	-0.0208	-0.0001

Table E18: GPI Score of Local Best Models when Estimating ET₀ at External Stations using Four Meteorological Variables

| | | Testing Station | | | | | | | | | | | |
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.644	0.152	0.481	1.073	0.065	0.119	0.161	0.093	0.156	0.344	0.278	0.102
	Station 48601 (Bayan Lepas)	-2.331	0.472	-1.343	-1.572	-4.862	-4.588	-4.683	-4.006	-4.451	-4.460	-3.970	-3.994
	Station 48603 (Alor Setar)	0.252	0.303	0.803	1.164	0.116	-0.052	-0.130	-0.618	-0.101	-0.196	-0.073	0.026
	Station 48615 (Kota Bharu)	0.313	0.337	0.571	1.455	0.003	-0.144	-0.020	-0.398	0.032	0.110	0.038	0.128
	Station 48620 (Sitiawan)	0.200	0.240	0.549	0.389	0.117	0.107	0.077	-0.014	0.131	-0.003	0.065	-0.021
	Station 48623 (Lubok Merbau)	-1.943	-1.530	-1.813	-1.527	0.016	0.412	0.050	0.323	-0.612	-0.214	-1.215	-1.101
	Station 48625 (Ipoh)	0.000	-0.527	-0.041	0.507	0.015	0.231	0.317	0.243	0.235	0.133	0.082	0.069
	Station 48632 (Cameron Highlands)	-2.499	-1.538	-1.966	-0.564	-0.534	-0.058	-0.537	0.514	-1.193	-0.958	-1.158	-0.442
	Station 48647 (Subang)	-2.416	-3.177	-2.170	-1.651	-1.975	-1.226	-0.502	-0.824	0.343	-0.871	0.035	-0.174
	Station 48649 (Muadzam Shah)	-3.171	-2.238	-2.340	-1.546	-0.101	0.261	-0.206	0.291	-0.679	0.411	-1.926	-1.676
	Station 48650 (KLIA)	0.481	0.000	0.446	0.196	0.127	0.183	0.213	0.139	0.303	0.266	0.347	0.165
	Station 48657 (Kuantan)	-3.863	-4.060	-3.659	-3.209	-2.380	-1.036	-0.361	-3.679	-0.049	-0.939	-0.524	0.238

Table E19: MAE (in mm/day) of Local Best Models when Estimating ET$_0$ at External Stations using Three Meteorological Variables

| | | Testing Station | | | | | | | | | | | |
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.1123	0.1066	0.1120	0.1263	0.0646	0.0841	0.0826	0.1795	0.0913	0.0605	0.0909	0.0910
	Station 48601 (Bayan Lepas)	0.1179	0.1058	0.1122	0.1268	0.0646	0.0868	0.0878	0.1860	0.0963	0.0611	0.0931	0.0894
	Station 48603 (Alor Setar)	0.1200	0.1083	0.0759	0.1203	0.0439	0.0509	0.0619	0.0902	0.0754	0.0464	0.0971	0.0653
	Station 48615 (Kota Bharu)	0.1318	0.1306	0.0963	0.0956	0.0456	0.0563	0.0562	0.0611	0.0679	0.0541	0.0840	0.0592
	Station 48620 (Sitiawan)	0.1251	0.1092	0.0838	0.1017	0.0435	0.0477	0.0565	0.0525	0.0661	0.0459	0.0833	0.0569
	Station 48623 (Lubok Merbau)	0.1252	0.1209	0.0858	0.1043	0.0446	0.0459	0.0533	0.0699	0.0628	0.0435	0.0826	0.0584
	Station 48625 (Ipoh)	0.1194	0.1839	0.1741	0.0997	0.0450	0.0455	0.0550	0.0547	0.0647	0.0417	0.0791	0.0571
	Station 48632 (Cameron Highlands)	0.1451	0.1436	0.1138	0.1061	0.0462	0.0493	0.0584	0.0445	0.0708	0.0458	0.0870	0.0572
	Station 48647 (Subang)	0.1224	0.1295	0.0904	0.0997	0.0444	0.0480	0.0567	0.0498	0.0640	0.0419	0.0817	0.0595
	Station 48649 (Muadzam Shah)	0.2451	0.2343	0.1674	0.1556	0.0695	0.0667	0.0670	0.0727	0.1098	0.0398	0.1138	0.0742
	Station 48650 (KLIA)	0.1200	0.1181	0.0886	0.1021	0.0439	0.0491	0.0569	0.0546	0.0675	0.0506	0.0784	0.0556
	Station 48657 (Kuantan)	0.1228	0.1337	0.0921	0.0978	0.0450	0.0472	0.0547	0.0520	0.0656	0.0422	0.0793	0.0552

Table E20: RMSE (in mm/day) of Local Best Models when Estimating ET$_0$ at External Stations using Three Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.1534	0.1406	0.1481	0.1668	0.0839	0.1116	0.1053	0.2589	0.1151	0.0838	0.1174	0.1199
	Station 48601 (Bayan Lepas)	0.1610	0.1389	0.1490	0.1690	0.0834	0.1139	0.1119	0.2555	0.1209	0.0813	0.1200	0.1186
	Station 48603 (Alor Setar)	0.1791	0.1527	0.1061	0.1788	0.0564	0.0666	0.0816	0.1371	0.1041	0.0681	0.1389	0.0925
	Station 48615 (Kota Bharu)	0.1986	0.2055	0.1434	0.1393	0.0581	0.0729	0.0716	0.0851	0.0883	0.0732	0.1188	0.0767
	Station 48620 (Sitiawan)	0.1869	0.1552	0.1206	0.1467	0.0547	0.0618	0.0714	0.0736	0.0854	0.0644	0.1175	0.0737
	Station 48623 (Lubok Merbau)	0.1822	0.1679	0.1194	0.1474	0.0564	0.0589	0.0676	0.1024	0.0820	0.0622	0.1135	0.0766
	Station 48625 (Ipoh)	0.1747	0.2500	0.2405	0.1422	0.0568	0.0587	0.0698	0.0786	0.0838	0.0579	0.1084	0.0741
	Station 48632 (Cameron Highlands)	0.2137	0.2183	0.1680	0.1513	0.0592	0.0661	0.0753	0.0607	0.0933	0.0630	0.1247	0.0750
	Station 48647 (Subang)	0.1802	0.1841	0.1310	0.1433	0.0572	0.0616	0.0715	0.0692	0.0830	0.0591	0.1125	0.0765
	Station 48649 (Muadzam Shah)	0.4063	0.3552	0.2603	0.2558	0.1194	0.0865	0.0870	0.1010	0.1401	0.0559	0.1831	0.0981
	Station 48650 (KLIA)	0.1820	0.1692	0.1313	0.1511	0.0553	0.0634	0.0724	0.0781	0.0871	0.0702	0.1079	0.0720
	Station 48657 (Kuantan)	0.1806	0.1911	0.1320	0.1414	0.0568	0.0622	0.0698	0.0718	0.0857	0.0585	0.1097	0.0720

Table E21: MAPE (in %) of Local Best Models when Estimating ET$_0$ at External Stations using Three Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	2.565	2.769	2.907	3.280	1.677	2.183	2.145	4.661	2.371	1.570	2.359	2.363
	Station 48601 (Bayan Lepas)	3.060	2.440	2.913	3.292	1.676	2.253	2.279	4.829	2.500	1.587	2.418	2.321
	Station 48603 (Alor Setar)	3.115	2.811	1.774	3.122	1.140	1.321	1.607	2.342	1.957	1.204	2.520	1.695
	Station 48615 (Kota Bharu)	3.423	3.391	2.500	2.203	1.183	1.462	1.460	1.587	1.763	1.405	2.182	1.536
	Station 48620 (Sitiawan)	3.247	2.835	2.176	2.640	1.141	1.238	1.467	1.363	1.716	1.192	2.164	1.477
	Station 48623 (Lubok Merbau)	3.250	3.138	2.229	2.708	1.158	1.128	1.384	1.815	1.630	1.129	2.144	1.517
	Station 48625 (Ipoh)	3.100	4.775	4.519	2.589	1.168	1.181	1.375	1.420	1.680	1.082	2.053	1.484
	Station 48632 (Cameron Highlands)	3.767	3.727	2.956	2.755	1.198	1.280	1.516	1.407	1.837	1.189	2.258	1.485
	Station 48647 (Subang)	3.177	3.361	2.346	2.588	1.153	1.246	1.473	1.294	1.582	1.087	2.121	1.546
	Station 48649 (Muadzam Shah)	6.363	6.084	4.345	4.041	1.805	1.733	1.739	1.888	2.851	1.143	2.954	1.927
	Station 48650 (KLIA)	3.114	3.065	2.299	2.651	1.140	1.275	1.478	1.417	1.752	1.313	1.936	1.443
	Station 48657 (Kuantan)	3.188	3.471	2.390	2.538	1.168	1.224	1.421	1.351	1.704	1.096	2.059	1.433

Table E22: R^2 of Local Best Models when Estimating ET_0 at External Stations using Three Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.9815	0.9814	0.9807	0.9735	0.9889	0.9803	0.9814	0.9413	0.9862	0.9901	0.9860	0.9830
	Station 48601 (Bayan Lepas)	0.9796	0.9818	0.9805	0.9727	0.9891	0.9795	0.9789	0.9428	0.9848	0.9907	0.9856	0.9834
	Station 48603 (Alor Setar)	0.9748	0.9781	0.9900	0.9694	0.9950	0.9930	0.9888	0.9835	0.9881	0.9934	0.9805	0.9899
	Station 48615 (Kota Bharu)	0.9688	0.9604	0.9820	0.9812	0.9947	0.9916	0.9913	0.9936	0.9914	0.9924	0.9858	0.9931
	Station 48620 (Sitiawan)	0.9727	0.9775	0.9873	0.9795	0.9953	0.9940	0.9914	0.9953	0.9920	0.9942	0.9861	0.9936
	Station 48623 (Lubok Merbau)	0.9738	0.9735	0.9876	0.9792	0.9950	0.9945	0.9922	0.9908	0.9926	0.9945	0.9867	0.9931
	Station 48625 (Ipoh)	0.9761	0.9412	0.9492	0.9807	0.9949	0.9945	0.9918	0.9946	0.9927	0.9953	0.9881	0.9935
	Station 48632 (Cameron Highlands)	0.9643	0.9555	0.9753	0.9781	0.9945	0.9931	0.9904	0.9967	0.9905	0.9944	0.9844	0.9933
	Station 48647 (Subang)	0.9743	0.9681	0.9850	0.9804	0.9949	0.9940	0.9914	0.9958	0.9928	0.9950	0.9871	0.9931
	Station 48649 (Muadzam Shah)	0.9175	0.9292	0.9632	0.9471	0.9828	0.9925	0.9879	0.9911	0.9904	0.9955	0.9731	0.9896
	Station 48650 (KLIA)	0.9741	0.9730	0.9848	0.9782	0.9952	0.9936	0.9911	0.9946	0.9921	0.9931	0.9881	0.9939
	Station 48657 (Kuantan)	0.9744	0.9657	0.9847	0.9809	0.9949	0.9939	0.9918	0.9955	0.9924	0.9951	0.9878	0.9939

Table E23: MBE of Local Best Models when Estimating ET$_0$ at External Stations using Three Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.0001	-0.0002	0.0001	-0.0001	0.0002	0.0002	0.0000	0.0004	0.0002	-0.0002	0.0000	-0.0003
	Station 48601 (Bayan Lepas)	0.0000	0.0000	0.0003	0.0001	-0.0002	0.0000	0.0000	-0.0001	-0.0002	0.0001	0.0001	0.0000
	Station 48603 (Alor Setar)	0.0002	0.0001	-0.0002	0.0003	-0.0002	-0.0001	0.0002	-0.0005	0.0004	-0.0004	-0.0004	-0.0001
	Station 48615 (Kota Bharu)	-0.0008	0.0006	0.0002	0.0000	-0.0001	0.0002	-0.0002	-0.0002	-0.0001	0.0003	0.0001	0.0001
	Station 48620 (Sitiawan)	0.0004	0.0002	0.0003	0.0000	0.0000	0.0001	0.0002	-0.0001	0.0000	0.0000	0.0001	0.0002
	Station 48623 (Lubok Merbau)	-0.0006	0.0006	0.0004	-0.0003	0.0000	0.0000	0.0000	0.0006	0.0001	-0.0003	0.0001	0.0000
	Station 48625 (Ipoh)	-0.0001	0.0008	-0.0001	0.0001	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	0.0001	0.0000
	Station 48632 (Cameron Highlands)	0.0000	-0.0003	0.0003	0.0000	0.0002	0.0000	-0.0001	0.0001	-0.0001	0.0000	0.0005	0.0000
	Station 48647 (Subang)	0.0001	0.0000	0.0000	0.0002	-0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000
	Station 48649 (Muadzam Shah)	-0.2229	-0.2183	-0.1425	-0.0955	-0.0562	0.0514	-0.0108	0.0064	-0.0992	-0.0005	-0.0668	0.0276
	Station 48650 (KLIA)	0.0000	0.0000	-0.0001	-0.0004	-0.0001	0.0000	0.0000	0.0003	0.0000	0.0000	0.0001	0.0000
	Station 48657 (Kuantan)	-0.0001	0.0009	0.0001	0.0003	-0.0002	0.0002	0.0001	0.0000	0.0000	0.0001	-0.0001	-0.0001

Table E24: GPI Score of Local Best Models when Estimating ET$_0$ at External Stations using Three Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.455	0.575	-0.592	-1.080	-2.405	-3.414	-2.956	-3.591	-2.216	-3.086	-0.409	-3.663
	Station 48601 (Bayan Lepas)	0.225	0.687	-0.610	-1.135	-2.385	-3.638	-3.588	-3.593	-2.691	-2.791	-0.596	-3.500
	Station 48603 (Alor Setar)	0.046	0.433	0.686	-1.116	0.107	-0.185	-0.685	-0.950	-1.118	-0.967	-1.399	-1.015
	Station 48615 (Kota Bharu)	-0.297	-0.484	-0.225	0.482	-0.072	-0.654	0.055	-0.037	-0.115	-1.836	-0.080	-0.047
	Station 48620 (Sitiawan)	-0.091	0.394	0.298	0.028	0.173	0.117	0.052	0.189	0.088	0.098	-0.001	0.182
	Station 48623 (Lubok Merbau)	-0.058	0.083	0.274	-0.071	0.054	0.345	0.400	-0.370	0.351	-0.079	0.129	-0.004
	Station 48625 (Ipoh)	0.093	-1.850	-3.185	0.161	0.012	0.315	0.280	0.111	0.255	0.949	0.480	0.162
	Station 48632 (Cameron Highlands)	-0.614	-0.828	-0.893	-0.188	-0.151	-0.096	-0.202	0.330	-0.433	0.229	-0.415	0.123
	Station 48647 (Subang)	0.002	-0.219	0.048	0.145	0.046	0.110	0.052	0.274	0.373	0.774	0.221	-0.060
	Station 48649 (Muadzam Shah)	-4.544	-4.313	-3.836	-4.518	-4.824	-2.326	-2.148	-1.295	-3.883	0.040	-4.374	-2.650
	Station 48650 (KLIA)	0.026	0.114	0.077	-0.065	0.141	0.000	0.004	0.091	0.007	-0.802	0.625	0.324
	Station 48657 (Kuantan)	-0.005	-0.364	0.001	0.232	0.011	0.127	0.236	0.228	0.149	0.613	0.432	0.343

Table E25: MAE (in mm/day) of Local Best Models when Estimating ET₀ at External Stations using Two Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.1872	0.1553	0.1560	0.1626	0.1358	0.1974	0.1942	0.3925	0.2170	0.2114	0.1631	0.1069
	Station 48601 (Bayan Lepas)	0.1987	0.1506	0.1435	0.1636	0.1763	0.2402	0.2133	0.4619	0.2223	0.2645	0.1785	0.1167
	Station 48603 (Alor Setar)	0.1993	0.1500	0.1370	0.1614	0.0947	0.1000	0.1228	0.1605	0.1375	0.0800	0.1283	0.0968
	Station 48615 (Kota Bharu)	0.1973	0.1489	0.1374	0.1593	0.0955	0.0956	0.1196	0.1592	0.1315	0.0755	0.1294	0.0984
	Station 48620 (Sitiawan)	0.1945	0.1800	0.1764	0.1743	0.0836	0.0860	0.1022	0.2866	0.1088	0.0649	0.1400	0.1128
	Station 48623 (Lubok Merbau)	0.2017	0.1866	0.1741	0.1731	0.0809	0.0858	0.0983	0.2819	0.1076	0.0623	0.1398	0.1106
	Station 48625 (Ipoh)	0.2173	0.2116	0.1942	0.2390	0.0885	0.1061	0.1014	0.5491	0.1423	0.0689	0.1494	0.1388
	Station 48632 (Cameron Highlands)	0.2599	0.2228	0.1882	0.1799	0.0931	0.0830	0.1174	0.0620	0.1321	0.0849	0.1419	0.0977
	Station 48647 (Subang)	0.2086	0.1851	0.1783	0.1747	0.0836	0.0870	0.1019	0.2861	0.1108	0.0639	0.1428	0.1207
	Station 48649 (Muadzam Shah)	0.3105	0.2850	0.2506	0.1959	0.0954	0.0922	0.1233	0.4602	0.1969	0.0609	0.1736	0.1157
	Station 48650 (KLIA)	0.1997	0.1536	0.1377	0.1665	0.0948	0.0983	0.1218	0.1665	0.1386	0.0742	0.1271	0.0968
	Station 48657 (Kuantan)	0.2146	0.1795	0.1552	0.1751	0.1222	0.1645	0.1402	0.3709	0.1480	0.1887	0.1337	0.0969

Table E26: RMSE (in mm/day) of Local Best Models when Estimating ET$_0$ at External Stations using Two Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.2699	0.2078	0.2054	0.2234	0.1860	0.2681	0.2672	0.4744	0.2988	0.2498	0.2222	0.1351
	Station 48601 (Bayan Lepas)	0.2764	0.2183	0.1919	0.2220	0.2219	0.2970	0.2783	0.5435	0.2999	0.2942	0.2330	0.1448
	Station 48603 (Alor Setar)	0.2824	0.2040	0.1820	0.2185	0.1225	0.1334	0.1559	0.2417	0.1805	0.1073	0.1742	0.1248
	Station 48615 (Kota Bharu)	0.2816	0.2025	0.1828	0.2149	0.1234	0.1246	0.1515	0.2408	0.1693	0.1004	0.1746	0.1268
	Station 48620 (Sitiawan)	0.2821	0.2441	0.2407	0.2259	0.1082	0.1128	0.1320	0.3652	0.1394	0.0876	0.1900	0.1473
	Station 48623 (Lubok Merbau)	0.2882	0.2528	0.2399	0.2236	0.1044	0.1127	0.1268	0.3630	0.1375	0.0846	0.1910	0.1442
	Station 48625 (Ipoh)	0.3207	0.3006	0.2833	0.3102	0.1415	0.1341	0.1313	0.7060	0.1802	0.0957	0.2031	0.1905
	Station 48632 (Cameron Highlands)	0.3805	0.3471	0.2873	0.2675	0.1202	0.1172	0.1523	0.0956	0.1723	0.1132	0.2205	0.1276
	Station 48647 (Subang)	0.2962	0.2466	0.2469	0.2259	0.1084	0.1140	0.1318	0.3683	0.1430	0.0882	0.1942	0.1567
	Station 48649 (Muadzam Shah)	0.5024	0.4181	0.3817	0.2695	0.1493	0.1240	0.1662	0.5579	0.2441	0.0824	0.2634	0.1487
	Station 48650 (KLIA)	0.2844	0.2069	0.1829	0.2273	0.1227	0.1297	0.1556	0.2511	0.1801	0.0989	0.1728	0.1248
	Station 48657 (Kuantan)	0.3385	0.2589	0.2142	0.2502	0.1706	0.2052	0.1789	0.4579	0.1962	0.2185	0.1855	0.1251

Table E27: MAPE (in %) of Local Best Models when Estimating ET$_0$ at External Stations using Two Meteorological Variables

| | | Testing Station | | | | | | | | | | | |
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	4.275	4.031	4.049	4.222	3.525	5.126	5.041	10.189	5.635	5.488	4.234	2.774
	Station 48601 (Bayan Lepas)	5.158	3.472	3.727	4.248	4.576	6.235	5.537	11.993	5.770	6.868	4.634	3.031
	Station 48603 (Alor Setar)	5.175	3.893	3.201	4.191	2.459	2.596	3.189	4.168	3.571	2.077	3.332	2.512
	Station 48615 (Kota Bharu)	5.121	3.867	3.568	3.668	2.480	2.483	3.104	4.133	3.413	1.961	3.359	2.555
	Station 48620 (Sitiawan)	5.051	4.672	4.580	4.525	2.194	2.232	2.653	7.440	2.825	1.686	3.634	2.929
	Station 48623 (Lubok Merbau)	5.238	4.844	4.519	4.493	2.100	2.109	2.551	7.320	2.794	1.617	3.630	2.871
	Station 48625 (Ipoh)	5.643	5.494	5.042	6.206	2.296	2.755	2.536	14.255	3.694	1.789	3.879	3.604
	Station 48632 (Cameron Highlands)	6.748	5.785	4.887	4.671	2.418	2.154	3.048	1.960	3.429	2.204	3.684	2.535
	Station 48647 (Subang)	5.416	4.806	4.628	4.535	2.170	2.259	2.646	7.429	2.739	1.658	3.707	3.133
	Station 48649 (Muadzam Shah)	8.062	7.398	6.506	5.085	2.476	2.393	3.202	11.949	5.112	1.750	4.508	3.003
	Station 48650 (KLIA)	5.186	3.987	3.575	4.323	2.461	2.553	3.162	4.322	3.597	1.927	3.139	2.512
	Station 48657 (Kuantan)	5.572	4.660	4.030	4.545	3.173	4.271	3.640	9.630	3.842	4.900	3.472	2.515

Table E28: R^2 of Local Best Models when Estimating ET_0 at External Stations using Two Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.9428	0.9611	0.9638	0.9540	0.9629	0.9398	0.9266	0.9347	0.9350	0.9728	0.9596	0.9791
	Station 48601 (Bayan Lepas)	0.9410	0.9547	0.9679	0.9531	0.9650	0.9502	0.9385	0.9299	0.9447	0.9779	0.9644	0.9809
	Station 48603 (Alor Setar)	0.9374	0.9612	0.9708	0.9544	0.9764	0.9718	0.9592	0.9484	0.9644	0.9838	0.9694	0.9816
	Station 48615 (Kota Bharu)	0.9381	0.9617	0.9707	0.9556	0.9760	0.9755	0.9610	0.9494	0.9685	0.9858	0.9692	0.9811
	Station 48620 (Sitiawan)	0.9375	0.9441	0.9494	0.9514	0.9815	0.9798	0.9704	0.8831	0.9787	0.9892	0.9636	0.9744
	Station 48623 (Lubok Merbau)	0.9344	0.9398	0.9500	0.9520	0.9828	0.9799	0.9725	0.8857	0.9790	0.9899	0.9621	0.9756
	Station 48625 (Ipoh)	0.9247	0.9270	0.9414	0.9446	0.9694	0.9787	0.9710	0.8375	0.9765	0.9890	0.9597	0.9651
	Station 48632 (Cameron Highlands)	0.8865	0.8874	0.9274	0.9317	0.9773	0.9783	0.9610	0.9919	0.9677	0.9820	0.9509	0.9807
	Station 48647 (Subang)	0.9308	0.9428	0.9464	0.9513	0.9815	0.9794	0.9707	0.8808	0.9786	0.9890	0.9618	0.9710
	Station 48649 (Muadzam Shah)	0.8619	0.8962	0.9200	0.9374	0.9695	0.9769	0.9659	0.8693	0.9749	0.9904	0.9466	0.9744
	Station 48650 (KLIA)	0.9369	0.9597	0.9706	0.9506	0.9763	0.9734	0.9591	0.9447	0.9660	0.9862	0.9697	0.9816
	Station 48657 (Kuantan)	0.9214	0.9521	0.9680	0.9462	0.9667	0.9681	0.9577	0.9290	0.9620	0.9833	0.9651	0.9815

Table E29: MBE of Local Best Models when Estimating ET$_0$ at External Stations using Two Meteorological Variables

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	-0.0003	-0.0429	-0.0329	-0.0401	0.1053	0.1754	0.1356	0.3806	0.1488	0.2059	0.0915	0.0033
	Station 48601 (Bayan Lepas)	0.0371	0.0019	0.0173	0.0124	0.1641	0.2336	0.1854	0.4575	0.1800	0.2629	0.1340	0.0687
	Station 48603 (Alor Setar)	-0.0003	0.0000	0.0000	-0.0003	-0.0001	0.0003	0.0000	0.0002	-0.0003	-0.0002	-0.0002	-0.0001
	Station 48615 (Kota Bharu)	0.0002	-0.0002	-0.0001	0.0003	0.0000	0.0000	0.0002	-0.0004	0.0000	0.0001	-0.0001	0.0003
	Station 48620 (Sitiawan)	-0.0001	-0.0010	-0.0003	-0.0002	0.0000	0.0001	0.0000	-0.0010	0.0000	-0.0003	0.0003	0.0003
	Station 48623 (Lubok Merbau)	-0.0006	-0.0002	0.0000	-0.0001	0.0000	-0.0001	-0.0001	0.0002	0.0002	-0.0002	-0.0004	0.0000
	Station 48625 (Ipoh)	-0.0837	-0.1103	-0.1116	0.1421	-0.0158	0.0672	0.0000	0.5356	-0.0948	0.0272	0.0134	0.0747
	Station 48632 (Cameron Highlands)	0.0001	-0.0001	-0.0002	-0.0001	0.0000	-0.0001	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000
	Station 48647 (Subang)	0.0006	-0.0002	-0.0003	0.0001	0.0001	0.0000	0.0001	-0.0007	-0.0003	0.0001	0.0005	-0.0001
	Station 48649 (Muadzam Shah)	-0.2633	-0.2508	-0.2163	-0.0786	-0.0542	0.0070	-0.0797	0.4027	-0.1862	0.0000	-0.1218	-0.0129
	Station 48650 (KLIA)	0.0000	-0.0004	0.0001	-0.0002	-0.0002	0.0004	0.0000	0.0002	0.0004	0.0001	-0.0001	0.0000
	Station 48657 (Kuantan)	-0.1130	-0.1237	-0.0875	-0.0761	0.0878	0.1461	0.0715	0.3541	0.0392	0.1838	0.0115	-0.0002

Table E30: GPI Score of Local Best Models when Estimating ET$_0$ at External Stations using Two Meteorological Variables

| | | Testing Station | | | | | | | | | | | |
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.477	0.454	-0.107	0.094	-2.716	-3.715	-3.533	-1.299	-3.656	-3.611	-2.045	-0.137
	Station 48601 (Bayan Lepas)	-0.038	0.661	0.320	0.244	-4.124	-4.401	-3.949	-1.877	-3.700	-4.269	-2.844	-1.519
	Station 48603 (Alor Setar)	0.021	0.718	0.723	0.470	0.005	-0.202	-0.119	0.847	-0.087	-0.237	0.867	0.693
	Station 48615 (Kota Bharu)	0.064	0.746	0.601	0.790	-0.040	-0.008	0.004	0.860	0.179	-0.045	0.816	0.549
	Station 48620 (Sitiawan)	0.096	-0.116	-0.757	-0.028	0.606	0.287	0.639	-0.305	0.987	0.311	0.012	-0.854
	Station 48623 (Lubok Merbau)	-0.078	-0.302	-0.700	0.051	0.771	0.320	0.788	-0.264	1.027	0.393	-0.058	-0.623
	Station 48625 (Ipoh)	-0.888	-1.478	-1.937	-3.668	-0.473	-0.399	0.704	-3.250	-0.403	0.120	-0.747	-4.303
	Station 48632 (Cameron Highlands)	-1.936	-1.935	-1.617	-1.414	0.104	0.263	0.038	1.750	0.132	-0.414	-0.940	0.555
	Station 48647 (Subang)	-0.260	-0.213	-0.876	-0.037	0.617	0.257	0.653	-0.322	0.971	0.312	-0.217	-1.574
	Station 48649 (Muadzam Shah)	-4.521	-4.010	-4.272	-2.112	-0.916	0.045	-0.480	-2.184	-2.265	0.414	-3.680	-1.175
	Station 48650 (KLIA)	0.001	0.634	0.595	0.104	-0.002	-0.123	-0.101	0.782	-0.067	0.000	1.048	0.696
	Station 48657 (Kuantan)	-1.076	-0.560	-0.309	-1.051	-2.005	-2.126	-0.993	-1.170	-0.629	-2.556	0.274	0.676

Table E31: MAE (in mm/day) of Local Best Models when Estimating ET$_0$ at External Stations using One Meteorological Variable

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.3804	0.2722	0.2583	0.2259	0.1140	0.1096	0.1485	0.2996	0.1551	0.0848	0.1924	0.1195
	Station 48601 (Bayan Lepas)	0.3723	0.2731	0.2950	0.2687	0.3319	0.3663	0.2615	0.6454	0.1904	0.4260	0.2439	0.2827
	Station 48603 (Alor Setar)	0.3615	0.2724	0.2588	0.2698	0.2350	0.2844	0.1930	0.5637	0.1716	0.3115	0.2090	0.2009
	Station 48615 (Kota Bharu)	0.3585	0.2690	0.2750	0.2260	0.2534	0.2699	0.2057	0.5983	0.1737	0.3531	0.2149	0.2111
	Station 48620 (Sitiawan)	0.3941	0.2748	0.2597	0.2259	0.1142	0.1096	0.1448	0.3020	0.1530	0.0871	0.1887	0.1205
	Station 48623 (Lubok Merbau)	0.4211	0.3764	0.3053	0.2833	0.1168	0.1096	0.1744	0.4104	0.2585	0.1063	0.2523	0.1376
	Station 48625 (Ipoh)	0.3656	0.2972	0.2536	0.2383	0.1428	0.1659	0.1486	0.4699	0.1818	0.2007	0.1941	0.1283
	Station 48632 (Cameron Highlands)	0.3913	0.2766	0.2586	0.2259	0.1142	0.1096	0.1463	0.3007	0.1532	0.0868	0.1890	0.1205
	Station 48647 (Subang)	0.3805	0.2721	0.2581	0.2259	0.1140	0.1097	0.1474	0.2997	0.1558	0.0848	0.1920	0.1195
	Station 48649 (Muadzam Shah)	0.3845	0.2722	0.2583	0.2259	0.1140	0.1096	0.1478	0.2997	0.1551	0.0848	0.1924	0.1195
	Station 48650 (KLIA)	0.3521	0.2668	0.2533	0.2351	0.2109	0.2452	0.1766	0.5341	0.1581	0.2924	0.1928	0.1758
	Station 48657 (Kuantan)	0.3884	0.2729	0.2583	0.2258	0.1140	0.1096	0.1479	0.2998	0.1544	0.0855	0.1871	0.1197

Table E32: RMSE (in mm/day) of Local Best Models when Estimating ET$_0$ at External Stations using One Meteorological Variable

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.5191	0.3950	0.3613	0.3074	0.1453	0.1473	0.1890	0.3771	0.1993	0.1138	0.2810	0.1557
	Station 48601 (Bayan Lepas)	0.5218	0.3966	0.3786	0.3404	0.3673	0.3886	0.2987	0.7416	0.2338	0.4424	0.3103	0.3139
	Station 48603 (Alor Setar)	0.5346	0.4149	0.3620	0.3579	0.2810	0.3145	0.2322	0.6703	0.2302	0.3345	0.2942	0.2421
	Station 48615 (Kota Bharu)	0.5501	0.4193	0.3918	0.3077	0.2960	0.2980	0.2447	0.6889	0.2183	0.3718	0.2996	0.2463
	Station 48620 (Sitiawan)	0.5344	0.4008	0.3629	0.3074	0.1454	0.1474	0.1839	0.3782	0.1961	0.1154	0.2768	0.1571
	Station 48623 (Lubok Merbau)	0.6454	0.5318	0.4574	0.3908	0.1718	0.1474	0.2315	0.4840	0.3195	0.1335	0.3595	0.1827
	Station 48625 (Ipoh)	0.5809	0.4581	0.3951	0.3383	0.1918	0.1955	0.1891	0.5651	0.2466	0.2246	0.2983	0.1651
	Station 48632 (Cameron Highlands)	0.5317	0.4005	0.3613	0.3075	0.1457	0.1473	0.1854	0.3781	0.1966	0.1156	0.2774	0.1569
	Station 48647 (Subang)	0.5200	0.3951	0.3612	0.3074	0.1452	0.1473	0.1879	0.3770	0.2000	0.1131	0.2803	0.1556
	Station 48649 (Muadzam Shah)	0.5233	0.3951	0.3613	0.3074	0.1453	0.1473	0.1882	0.3771	0.1994	0.1131	0.2810	0.1557
	Station 48650 (KLIA)	0.5438	0.4172	0.3706	0.3228	0.2554	0.2708	0.2130	0.6349	0.2079	0.3126	0.2821	0.2104
	Station 48657 (Kuantan)	0.5287	0.3967	0.3613	0.3072	0.1452	0.1473	0.1886	0.3771	0.1978	0.1143	0.2738	0.1557

Table E33: MAPE (in %) of Local Best Models when Estimating ET$_0$ at External Stations using One Meteorological Variable

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	8.687	7.068	6.706	5.864	2.960	2.845	3.856	7.777	4.028	2.202	4.995	3.103
	Station 48601 (Bayan Lepas)	9.665	6.294	7.659	6.977	8.618	9.511	6.788	16.755	4.942	11.059	6.332	7.339
	Station 48603 (Alor Setar)	9.386	7.072	6.047	7.005	6.101	7.382	5.010	14.634	4.454	8.087	5.426	5.216
	Station 48615 (Kota Bharu)	9.308	6.984	7.140	5.205	6.578	7.006	5.341	15.532	4.510	9.166	5.580	5.482
	Station 48620 (Sitiawan)	10.232	7.135	6.742	5.865	2.998	2.846	3.759	7.840	3.972	2.262	4.900	3.130
	Station 48623 (Lubok Merbau)	10.933	9.771	7.927	7.356	3.032	2.696	4.528	10.655	6.712	2.760	6.550	3.572
	Station 48625 (Ipoh)	9.491	7.716	6.583	6.186	3.707	4.308	3.718	12.199	4.719	5.211	5.038	3.330
	Station 48632 (Cameron Highlands)	10.158	7.180	6.714	5.866	2.965	2.845	3.797	9.506	3.978	2.254	4.907	3.129
	Station 48647 (Subang)	9.879	7.064	6.701	5.866	2.960	2.847	3.826	7.781	3.851	2.201	4.985	3.103
	Station 48649 (Muadzam Shah)	9.983	7.067	6.706	5.866	2.961	2.846	3.837	7.779	4.028	2.438	4.996	3.103
	Station 48650 (KLIA)	9.142	6.927	6.576	6.103	5.476	6.366	4.585	13.867	4.103	7.590	4.762	4.563
	Station 48657 (Kuantan)	10.083	7.086	6.706	5.862	2.959	2.846	3.841	7.783	4.008	2.219	4.858	3.108

Table E34: R² of Local Best Models when Estimating ET₀ at External Stations using One Meteorological Variable

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.7880	0.8528	0.8850	0.9099	0.9668	0.9657	0.9400	0.8751	0.9585	0.9818	0.9198	0.9714
	Station 48601 (Bayan Lepas)	0.7870	0.8519	0.8815	0.9086	0.9556	0.9636	0.9393	0.8697	0.9561	0.9813	0.9185	0.9710
	Station 48603 (Alor Setar)	0.7859	0.8477	0.8848	0.9048	0.9550	0.9649	0.9387	0.8714	0.9512	0.9810	0.9169	0.9683
	Station 48615 (Kota Bharu)	0.7848	0.8505	0.8802	0.9096	0.9532	0.9630	0.9385	0.8664	0.9557	0.9787	0.9174	0.9699
	Station 48620 (Sitiawan)	0.7760	0.8498	0.8842	0.9099	0.9666	0.9657	0.9425	0.8747	0.9577	0.9811	0.9228	0.9708
	Station 48623 (Lubok Merbau)	0.7863	0.8477	0.8844	0.9061	0.9560	0.9656	0.9395	0.8721	0.9516	0.9813	0.9178	0.9693
	Station 48625 (Ipoh)	0.7866	0.8493	0.8838	0.9080	0.9557	0.9653	0.9398	0.8710	0.9498	0.9812	0.9186	0.9700
	Station 48632 (Cameron Highlands)	0.7783	0.8505	0.8852	0.9099	0.9666	0.9657	0.9422	0.8740	0.9580	0.9812	0.9226	0.9708
	Station 48647 (Subang)	0.7868	0.8528	0.8851	0.9100	0.9668	0.9657	0.9404	0.8752	0.9580	0.9820	0.9202	0.9714
	Station 48649 (Muadzam Shah)	0.7857	0.8528	0.8850	0.9099	0.9668	0.9657	0.9402	0.8751	0.9584	0.9820	0.9197	0.9714
	Station 48650 (KLIA)	0.7874	0.8512	0.8832	0.9080	0.9556	0.9650	0.9396	0.8715	0.9554	0.9815	0.9192	0.9706
	Station 48657 (Kuantan)	0.7809	0.8522	0.8850	0.9099	0.9668	0.9657	0.9404	0.8752	0.9595	0.9815	0.9241	0.9712

Table E35: MBE of Local Best Models when Estimating ET$_0$ at External Stations using One Meteorological Variable

		Testing Station											
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.0000	0.0000	0.0000	-0.0001	0.0000	0.0000	0.0001	0.0001	-0.0001	-0.0002	0.0001	0.0001
	Station 48601 (Bayan Lepas)	-0.0334	-0.0008	0.0746	0.1215	0.3261	0.3577	0.2615	0.6323	0.1128	0.4246	0.1268	0.2720
	Station 48603 (Alor Setar)	-0.1051	-0.0777	-0.0005	0.0696	0.2195	0.2736	0.1286	0.5395	0.0279	0.3070	0.0431	0.1717
	Station 48615 (Kota Bharu)	-0.1449	-0.1082	-0.0366	0.0000	0.2364	0.2434	0.1365	0.5682	0.0234	0.3507	0.0368	0.1792
	Station 48620 (Sitiawan)	-0.0007	-0.0003	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000	0.0001
	Station 48623 (Lubok Merbau)	-0.3805	-0.3471	-0.2741	-0.2155	-0.0383	-0.0003	-0.1330	0.2903	-0.2347	0.0622	-0.2201	-0.0864
	Station 48625 (Ipoh)	-0.2570	-0.2218	0.2536	-0.0933	0.0914	0.1277	-0.0008	0.4099	-0.1058	0.1878	-0.0940	0.0406
	Station 48632 (Cameron Highlands)	0.0000	0.0000	0.0000	0.0000	-0.0002	-0.0001	0.0001	0.0000	-0.0001	-0.0001	0.0000	-0.0001
	Station 48647 (Subang)	-0.0001	0.0001	-0.0001	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	-0.0001	0.0001	0.0000
	Station 48649 (Muadzam Shah)	0.0002	-0.0001	-0.0001	-0.0001	0.0000	0.0000	0.0001	0.0000	-0.0001	0.0000	0.0000	0.0000
	Station 48650 (KLIA)	-0.1591	-0.1264	-0.0503	0.0023	0.1912	0.2263	0.0974	0.5039	-0.0158	0.2877	-0.0010	0.1388
	Station 48657 (Kuantan)	0.0001	0.0001	0.0000	0.0000	0.0001	0.0000	0.0002	0.0000	-0.0001	-0.0003	-0.0002	0.0001

Table E36: GPI Score of Local Best Models when Estimating ET$_0$ at External Stations using One Meteorological Variable

| | | Testing Station | | | | | | | | | | | |
		Station 48600 (Pulau Langkawi)	Station 48601 (Bayan Lepas)	Station 48603 (Alor Setar)	Station 48615 (Kota Bharu)	Station 48620 (Sitiawan)	Station 48623 (Lubok Merbau)	Station 48625 (Ipoh)	Station 48632 (Cameron Highlands)	Station 48647 (Subang)	Station 48649 (Muadzam Shah)	Station 48650 (KLIA)	Station 48657 (Kuantan)
Training Station	Station 48600 (Pulau Langkawi)	0.765	0.346	0.054	0.018	0.022	0.025	0.034	0.335	0.086	0.182	0.015	0.200
	Station 48601 (Bayan Lepas)	0.261	0.364	-2.303	-2.463	-4.793	-4.727	-4.031	-4.280	-1.568	-3.962	-2.606	-3.910
	Station 48603 (Alor Setar)	0.157	-0.989	0.342	-3.198	-3.234	-3.069	-1.907	-3.271	-1.337	-2.783	-1.226	-2.968
	Station 48615 (Kota Bharu)	-0.078	-0.526	-1.906	0.245	-3.653	-3.508	-2.320	-4.142	-0.793	-3.930	-1.373	-2.638
	Station 48620 (Sitiawan)	-1.218	-0.313	-0.178	0.013	0.002	0.020	0.768	0.273	0.079	-0.031	0.590	-0.001
	Station 48623 (Lubok Merbau)	-2.960	-4.351	-3.627	-4.415	-1.030	0.003	-1.400	-1.407	-4.513	-0.293	-3.955	-1.176
	Station 48625 (Ipoh)	-0.654	-1.826	-1.315	-1.528	-1.539	-1.114	0.030	-2.282	-2.135	-1.447	-0.827	-0.551
	Station 48632 (Cameron Highlands)	-0.932	-0.198	0.086	0.008	0.007	0.031	0.646	0.004	0.099	-0.016	0.549	0.002
	Station 48647 (Subang)	0.131	0.344	0.071	0.022	0.028	0.024	0.163	0.339	0.087	0.246	0.091	0.207
	Station 48649 (Muadzam Shah)	-0.090	0.338	0.053	0.016	0.026	0.024	0.098	0.331	0.083	0.228	0.008	0.198
	Station 48650 (KLIA)	0.308	-0.394	-0.422	-0.831	-2.766	-2.435	-1.138	-2.936	-0.422	-2.404	0.042	-1.594
	Station 48657 (Kuantan)	-0.626	0.203	0.052	0.013	0.028	0.016	0.137	0.336	0.219	0.076	0.859	0.139

Table E37: Performance of WOA-ELM-E Trained with Global Pooled Data when Estimating ET$_0$ at All Stations using Different Input Combinations

		MAE (mm/day)					
		C1	C4	C13	C33	C44	C58
Testing Station	Station 48600 (Pulau Langkawi)	0.0330	0.0503	0.0667	0.1155	0.1921	0.3803
	Station 48601 (Bayan Lepas)	0.0506	0.0603	0.0643	0.1225	0.1488	0.2720
	Station 48603 (Alor Setar)	0.0468	0.0540	0.0560	0.0823	0.1369	0.2584
	Station 48615 (Kota Bharu)	0.0396	0.0565	0.0633	0.1015	0.1618	0.2259
	Station 48620 (Sitiawan)	0.0295	0.0324	0.0382	0.0449	0.0945	0.1140
	Station 48623 (Lubok Merbau)	0.0380	0.0405	0.0481	0.0490	0.0931	0.1096
	Station 48625 (Ipoh)	0.0387	0.0392	0.0474	0.0601	0.1189	0.1485
	Station 48632 (Cameron Highlands)	0.0742	0.0605	0.0707	0.0601	0.1695	0.2998
	Station 48647 (Subang)	0.0345	0.0423	0.0495	0.0654	0.1340	0.1550
	Station 48649 (Muadzam Shah)	0.0295	0.0307	0.0344	0.0410	0.0731	0.0850
	Station 48650 (KLIA)	0.0339	0.0456	0.0519	0.0826	0.1272	0.1924
	Station 48657 (Kuantan)	0.0302	0.0356	0.0422	0.0567	0.0970	0.1195

		RMSE (mm/day)					
		C1	C4	C13	C33	C44	C58
Testing Station	Station 48600 (Pulau Langkawi)	0.0442	0.0735	0.0906	0.1717	0.2752	0.5190
	Station 48601 (Bayan Lepas)	0.0683	0.0797	0.0862	0.1736	0.1996	0.3951
	Station 48603 (Alor Setar)	0.0622	0.0699	0.0731	0.1131	0.1817	0.3613
	Station 48615 (Kota Bharu)	0.0518	0.0762	0.0860	0.1454	0.2196	0.3072
	Station 48620 (Sitiawan)	0.0361	0.0398	0.0533	0.0612	0.1224	0.1453
	Station 48623 (Lubok Merbau)	0.0488	0.0520	0.0634	0.0630	0.1235	0.1472
	Station 48625 (Ipoh)	0.0490	0.0486	0.0603	0.0759	0.1509	0.1890
	Station 48632 (Cameron Highlands)	0.1007	0.0860	0.1002	0.0829	0.2500	0.3769
	Station 48647 (Subang)	0.0431	0.0536	0.0636	0.0846	0.1744	0.1997
	Station 48649 (Muadzam Shah)	0.0372	0.0388	0.0445	0.0555	0.0964	0.1136
	Station 48650 (KLIA)	0.0421	0.0584	0.0685	0.1132	0.1728	0.2808
	Station 48657 (Kuantan)	0.0383	0.0449	0.0541	0.0738	0.1248	0.1556

		MAPE (%)					
		C1	C4	C13	C33	C44	C58
Testing Station	Station 48600 (Pulau Langkawi)	0.858	1.305	1.732	2.999	4.988	9.873
	Station 48601 (Bayan Lepas)	1.315	1.565	1.670	3.181	3.863	7.063
	Station 48603 (Alor Setar)	1.215	1.402	1.453	2.136	3.555	6.709
	Station 48615 (Kota Bharu)	1.029	1.467	1.644	2.635	4.201	5.864
	Station 48620 (Sitiawan)	0.766	0.842	0.991	1.167	2.453	2.960
	Station 48623 (Lubok Merbau)	0.986	1.052	1.248	1.273	2.417	2.846
	Station 48625 (Ipoh)	1.005	1.017	1.231	1.560	3.087	3.855
	Station 48632 (Cameron Highlands)	1.925	1.571	1.836	1.561	4.400	7.784
	Station 48647 (Subang)	0.895	1.099	1.286	1.699	3.478	4.024
	Station 48649 (Muadzam Shah)	0.765	0.798	0.894	1.066	1.898	2.206
	Station 48650 (KLIA)	0.880	1.183	1.348	2.143	3.303	4.996
	Station 48657 (Kuantan)	0.785	0.925	1.096	1.472	2.519	3.104

		R^2					
		C1	C4	C13	C33	C44	C58
Tes	Station 48600 (Pulau Langkawi)	0.9985	0.9958	0.9936	0.9769	0.9405	0.7883

Station 48601 (Bayan Lepas)	0.9956	0.9940	0.9930	0.9717	0.9627	0.8530
Station 48603 (Alor Setar)	0.9966	0.9957	0.9953	0.9888	0.9710	0.8850
Station 48615 (Kota Bharu)	0.9974	0.9945	0.9929	0.9798	0.9539	0.9099
Station 48620 (Sitiawan)	0.9979	0.9975	0.9954	0.9940	0.9765	0.9669
Station 48623 (Lubok Merbau)	0.9962	0.9957	0.9937	0.9937	0.9758	0.9656
Station 48625 (Ipoh)	0.9959	0.9960	0.9939	0.9903	0.9617	0.9399
Station 48632 (Cameron Highlands)	0.9911	0.9935	0.9911	0.9939	0.9451	0.8751
Station 48647 (Subang)	0.9981	0.9970	0.9958	0.9925	0.9681	0.9583
Station 48649 (Muadzam Shah)	0.9981	0.9979	0.9972	0.9957	0.9870	0.9819
Station 48650 (KLIA)	0.9982	0.9965	0.9952	0.9870	0.9697	0.9197
Station 48657 (Kuantan)	0.9983	0.9976	0.9965	0.9936	0.9816	0.9713

		MBE					
		C1	C4	C13	C33	C44	C58
Testing Station	Station 48600 (Pulau Langkawi)	0.0000	0.0000	-0.0001	0.0001	-0.0001	0.0000
	Station 48601 (Bayan Lepas)	0.0004	0.0000	0.0001	-0.0001	0.0000	-0.0002
	Station 48603 (Alor Setar)	0.0001	-0.0001	-0.0004	0.0001	0.0000	0.0000
	Station 48615 (Kota Bharu)	0.0000	0.0001	-0.0002	0.0000	0.0002	0.0001
	Station 48620 (Sitiawan)	-0.0002	0.0000	0.0003	-0.0004	0.0000	0.0000
	Station 48623 (Lubok Merbau)	0.0001	0.0000	-0.0001	0.0000	0.0002	0.0000
	Station 48625 (Ipoh)	0.0000	0.0001	0.0001	0.0001	0.0000	0.0000
	Station 48632 (Cameron Highlands)	0.0002	0.0002	0.0002	0.0002	0.0015	-0.0001
	Station 48647 (Subang)	0.0000	-0.0001	0.0000	0.0001	0.0001	-0.0001
	Station 48649 (Muadzam Shah)	0.0002	0.0001	0.0001	0.0000	0.0001	-0.0001
	Station 48650 (KLIA)	0.0000	0.0000	-0.0001	0.0001	0.0002	-0.0001
	Station 48657 (Kuantan)	0.0000	0.0000	-0.0001	-0.0002	-0.0001	0.0000

		GPI					
		C1	C4	C13	C33	C44	C58
Testing Station	Station 48600 (Pulau Langkawi)	0.842	0.536	0.349	0.173	0.198	0.267
	Station 48601 (Bayan Lepas)	0.182	0.001	0.120	0.001	0.774	0.379
	Station 48603 (Alor Setar)	0.112	0.000	0.447	0.415	0.621	0.044
	Station 48615 (Kota Bharu)	0.551	0.503	0.980	0.054	0.430	0.006
	Station 48620 (Sitiawan)	0.339	0.220	-0.079	-0.127	0.019	0.030
	Station 48623 (Lubok Merbau)	0.061	-0.009	-0.013	0.018	0.039	0.001
	Station 48625 (Ipoh)	0.137	0.104	-0.055	-0.327	0.035	0.001
	Station 48632 (Cameron Highlands)	0.001	0.044	-0.149	0.001	0.772	0.325
	Station 48647 (Subang)	-0.056	0.000	-0.048	0.195	0.098	0.070
	Station 48649 (Muadzam Shah)	0.234	0.671	0.458	1.160	0.063	0.219
	Station 48650 (KLIA)	0.001	0.181	0.081	0.157	0.936	0.003
	Station 48657 (Kuantan)	-0.020	0.275	-0.042	0.191	0.682	0.182