

Estimation of Potential and Actual Evapotranspiration under Limited Climate Data

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science in Civil Engineering by Research

by

Adeeba Ayaz

201971003

adeeba.ayaz@research.iiit.ac.in



International Institute of Information Technology, Hyderabad
(Deemed to be University)

Hyderabad -500 032, INDIA

April-2022

Copyright © Adeeba Ayaz, 2021
All Rights Reserved

CERTIFICATE

It is certified that the work contained in this thesis, titled “**Estimation of Potential and Actual Evapotranspiration under Limited Climate Data**” by ADEEBA AYZAZ , has been carried out under our supervision and is not submitted elsewhere for a degree.

Dr. Shaik Rehana

Assistant Professor

International Institute of Information Technology

Hyderabad, India

Dr. Shailesh Kumar Singh

Senior Scientist, Hydrologist

National Institute of Water and Atmospheric

Research, Christchurch, New Zealand

To the pursuit of knowledge

Acknowledgements

I would like to take this opportunity to thank my research advisor, Dr. Shaik Rehana, without whose vision this work would not have taken a form. I am grateful to Dr. Rehana for never doubting my ability to do things in different research areas and helping me in pursuing the right problem statements in areas, I was passionate about. I thank her for being constantly supportive in the roughest times over the past few years and motivating me to work hard through paper submissions and to meet conference deadlines.

I would also like to thank my co-advisor Dr. Shailesh Kumar Singh, for helping me in my research and for constantly supporting me and motivating me for paper submission. I would also like to thank Dr K.S. Rajan for introducing and arousing interest in the domain of Spatial Informatics.

I would like to thank my friend and colleague Rajesh for being a constant companion in this journey and helping me in solving the challenges in this work, and Sharas Chandra, for providing valuable inputs and the domain knowledge. My appreciation goes to the institute IIIT-Hyderabad for providing exposure to different research domains and the Lab for Spatial Informatics for focusing on real world problems like global environmental change, drought analysis and remote sensing impacting society.

I specially thank my friends Priyanka Sah and Avantika Latwal for patiently bearing me and getting me out of the darkest times in my life. I also thank them for being curious about my work, providing insights. Above everyone else, I thank my parents for inspiring and guiding me and for believing in me and never giving up on me, and my sister Aiman Ayaz for caring and providing me with all the love and comfort. I thank my family members for the faith in me and celebrating my every achievement as their own.

The research work presented in this thesis is funded by Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India through Start-up Grant for Young Scientists (YSS) Project no. YSS/2015/002111 received by Dr. Shaik Rehana. The authors sincerely thank Professor Jayashankar Telangana State Agricultural University (PJTSAU), Rajendranagar, Hyderabad, India, for providing meteorological data.

Abstract

Evapotranspiration (ET) is one of the prominent hydrologic variables affecting water and energy balances and critical factors for crop water requirements and irrigation scheduling. ET is a complex hydrological variable defined by various hydro climatological variables. The major forms of ET which are widely applicable in hydrological, water balance, drought assessments and other ecological assessments are Potential (PET) and Actual Evapotranspiration (AET). PET represents the atmospheric water demand with a focus on climatological variables. AET is influenced by climate, vegetation, soil moisture, and the amount of available water and presents the annual water balance between precipitation and latent heat exchange. Various empirical formulations have been developed to estimate PET and AET depending upon the availability of hydrometeorological variables. These empirical formulations are region-specific and developed for particular climatic conditions. Furthermore, the empirical family of models has a major limitation as they require a large number of hydrometeorological inputs, limiting their utility in data-scarce areas of ungauged basins.

In this context, empirical and mathematical models have emerged as simple and readily implementable for estimating PET and AET with measured hydrometeorological parameters as independent variables. Such mathematical models can be valuable to predict PET and AET when climate data is insufficient. The present study compared various empirical models and data-driven algorithms to predict PET and AET using various hydroclimatic variables. Four empirical methods, such as FAO-based Penman-Monteith method, temperature-based Hargreaves method, and radiation-based Turc, and Priestley-Taylor method, were used to estimate PET at a daily time scale. Five data-driven algorithms, such as Long short-term memory neural networks (LSTM), Artificial Neural Network (ANN), Gradient Boosting Regressor (GBR), Random Forest (RF) and Support Vector Regression (SVR), were implemented. Two empirical AET models, such as Budyko and Turc methods, were utilized in estimating AET. These models were evaluated over two different climatic regions, Hyderabad, the largest city of the Indian state, Telangana and Waipara in New Zealand, both with semi-arid climates. Dataset consists of daily meteorological data of maximum and minimum air temperature, relative humidity, solar radiation, and wind speed over a period of 51 years (1965 - 2015) for Hyderabad and for a period of 6 years (2010- 2016) for Waipara station.

The Penman-Monteith method was considered as the standard method to compare the different models and various empirical models of PET. The models were trained and tested with climate variables as input variables and various empirical models as reference models. The most influencing climate variables on PET were found in the order of temperature, solar radiation, wind speed, and relative humidity, which formed as the basis to choose different datasets to train over models and compare the results to validate. Temperature and radiation-based models of Turc and Priestley-Taylor methods can be used to estimate PET when all other climate variables are not available as they are also promising with the Penman-Monteith method. The results indicated that 99% accuracy could be achieved with all climatic input, whereas accuracy drops to 86% with limited data. Both LSTM and ANN models have been noted as the most robust models for estimating PET with minimal climate data. Even though the excellent performance can be achievable when all input variables are used, the study, however, found that even a three-parameter combination (temperature, wind speed and relative humidity values) or two-parameter combination (temperature and relative humidity, temperature and wind speed) can also be promising in PET estimation for a semi-arid climate.

The AET over semi-arid climatic conditions of Hyderabad, Telangana, India and Waipara (New Zealand) was estimated using modelled and different empirical methods-based PET using Budyko and Turc models. The proposed empirical-based AET models, Budyko and Turc, showed that the AET process has the potential to be estimated by structurally simple methods. Equation-based AET methods made it possible to extract useful information about the hydrological process. It was observed that the meteorological variables of temperature and solar radiation have more significant contributions than other variables in the estimation of AET. In addition, the effects of the meteorological variables were found to be essential and effective in the estimation of AET. The research findings of the study reveal that under limited data availability, the best input combinations were identified as temperature and wind speed for estimating PET; temperature, wind speed and precipitation for estimating AET for semi-arid climatology. Overall, the research findings of the study stress on the use of limited data in understanding the complex hydrological processes such as PET and AET using data-driven and empirical-based approaches for diverse climatological conditions.

Contents

1	Introduction.....	1
1.1	Background.....	1
1.2	Motivation.....	1
1.3	Introduction.....	2
1.4	Problem Definition.....	4
1.5	Objectives of the thesis.....	5
1.6	Scope of the study.....	7
1.7	Thesis Organization.....	8
2	Literature Review.....	10
2.1	Introduction.....	10
2.2	Developments in estimation of potential evapotranspiration.....	14
2.3	Estimation of PET using machine learning models.....	17
2.4	Estimation of AET from ML models.....	17
2.5	Conclusion.....	19
3	Study area and data.....	20
3.1	Study area.....	20
3.2	Data sets.....	23
3.2.1	Hyderabad data set.....	23
3.2.2	Waipara data set.....	24
3.2.3	Climate classification.....	26
3.3	Conclusion.....	27
4	Machine Learning Models.....	28
4.1	Pre-processing methods.....	28
4.1.1	Pearson Correlation.....	28
4.1.2	Principal Component Analysis.....	28
4.1.3	LASSO Model.....	29
4.2	Artificial Neural Network.....	30
4.3	Gradient Boosting Regressor.....	32
4.4	Random Forest.....	33
4.5	Support Vector Regression.....	33
4.6	Long Short Term Memory.....	35

4.7 Model evaluation.....	36
4.8 Conclusion.....	37
5 Estimation of Potential Evapotranspiration using machine learning models with limited data.....	38
5.1 Introduction.....	38
5.2 Estimation of potential evapotranspiration using empirical methods.....	38
5.2.1 Penman-Monteith method.....	38
5.2.2 Priestley-Taylor method.....	39
5.2.3 Turc Method.....	40
5.2.4 Hargreaves method.....	40
5.3 Modelling of potential evapotranspiration using machine learning models.....	40
5.4 Results and discussions.....	41
5.4.1 ML models performance with various input combinations.....	48
5.4.2 Discussions.....	61
5.5 Conclusions.....	62
6 Estimation of Actual Evapotranspiration with Limited Data.....	63
6.1 Introduction.....	64
6. Methods to estimate actual evapotranspiration.....	65
6.2.1 Budyko method.....	66
6.2.2 Turc method.....	66
6.3 Results.....	67
6.3.1 ML models performance with various AET methods under different Input combinations.....	73
6.3.2 Discussions.....	84
6.4 Conclusion.....	85
7 Conclusions.....	87
8 Publications.....	90
9 Bibliography.....	91

List of Figures

Figures

3.1 Case study: Hyderabad, Telangana, India.....	22
3.2 Case study: Waipara, New Zealand.....	23
4.1 Structure of ANN used for training a model with hidden layer and weights and the output layer showing a feed-forward pass. $x_i w_i$	31
4.2 Structure of SVR.....	34
4.3 Overview diagram of Long short-term memory (LSTM).....	36
5.1 Flowchart for the estimation of PET using ML models.....	38
5.2 Comparison of observed and estimated PET by different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period at Hyderabad (Top) and Waipara Stations (Bottom) for the Penman-Monteith method.....	49
5.3 Comparison of observed and estimated PET by different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period at Hyderabad (Top) and Waipara Stations (Bottom) for Priestley Taylor method.	50
5.4 Comparison of observed and estimated PET by different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period at Hyderabad (Top) and Waipara Stations (Bottom) for Hargreaves method.....	51
5.5 Comparison of observed and estimated PET by different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period at Hyderabad (Top) and Waipara Stations (Bottom) for Turc method.....	52
5.6 Comparison of RMSE values for Hyderabad (Top) and Waipara station (Bottom) for different input combinations for Penman-Monteith Method.....	56
5.7 Comparison of RMSE values for Hyderabad (Top) and Waipara station (Bottom) for different input combinations for Priestley-Taylor Method	57
5.8 Comparison of RMSE values for Hyderabad (Top) and Waipara station (Bottom) for different input combinations for Hargreaves Method.....	58
5.9 Comparison of RMSE values for Hyderabad (Top) and Waipara station (Bottom) for different input combinations for Turc Method.....	59
5.10 Correlation of main meteorological parameters such as temperature, relative humidity, solar radiation, wind speed to PET.....	64
6.1 Flowchart for estimation of AET using modelled PET.....	65
6.2 Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF and SVR) with varying parameters of input for the validation period for Budyko (Top) and Turc methods (Bottom) at Hyderabad station for Penman-Monteith method.....	74

6.3 Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF and SVR) with varying parameters of input for the validation period for Budyko (Top) and Turc methods (Bottom) at Waipara station for Penman-Monteith method.....	75
6.4 Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF and SVR) with varying parameters of input for the validation period for Budyko (Top) and Turc methods (Bottom) at Hyderabad station for Priestley Taylor method.....	76
6.5 Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF and SVR) with varying parameters of input for the validation period for Budyko (Top) and Turc methods (Bottom) at Waipara station for Priestley Taylor method.....	77
6.6 Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF and SVR) with varying parameters of input for the validation period for Budyko (Top) and Turc methods (Bottom) at Hyderabad station for Hargreaves method.....	78
6.7 Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF and SVR) with varying parameters of input for the validation period for Budyko (Top) and Turc methods (Bottom) at Waipara station for Hargreaves method.....	79
6.8 Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF and SVR) with varying parameters of input for the validation period for Budyko (Top) and Turc methods (Bottom) at Hyderabad station for Turc method.....	80
6.9 Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF and SVR) with varying parameters of input for the validation period for Budyko (Top) and Turc methods (Bottom) at Waipara station for Turc method.....	81

List of Tables

Table	Page
3.1 Statistical values of available meteorological variables and PET at Hyderabad station	25
3.2 Statistical values of available meteorological variables and PET at Waipara station	25
3.3 Climate classification according to aridity index	26
3.4 Climate classification of Hyderabad and Waipara stations	26
5.1 Performance of Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long Short-Term Memory (LSTM) and Artificial Neural Network(ANN) for Penman-Monteith Method.	43
5.2 Performance of Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long Short-Term Memory (LSTM) and Artificial Neural Network(ANN) for Priestley-Taylor Method.	44
5.3 Performance of Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long Short-Term Memory (LSTM) and Artificial Neural Network(ANN) for Hargreaves Method.	45
5.4 Performance of Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long Short-Term Memory (LSTM) and Artificial Neural Network(ANN) for Turc Method.	46
6.1 Performance of Budyko and Turc methods in estimating AET using PET modelled from Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long Short-Term Memory (LSTM) and Artificial Neural Network(ANN) for Penman-Monteith Method under different input combinations.	68
6.2 Performance of Budyko and Turc methods in estimating AET using PET modelled from Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long Short-Term Memory (LSTM) and Artificial Neural Network(ANN) for Priestley-Taylor Method under different input combinations.	69
6.3 Performance of Budyko and Turc methods in estimating AET using PET modelled from Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long Short-Term Memory (LSTM) and Artificial Neural Network(ANN) for Hargreaves Method under different input combinations.	70
6.4 Performance of Budyko and Turc methods in estimating AET using PET modelled from Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long Short-Term Memory (LSTM) and Artificial Neural Network(ANN) for Turc Method	

under different input combinations.

Chapter1

Introduction

1.1 Background

Among the many issues associated with prolonged drought is plants inability to extract water at a rate fast enough to keep up with the rates of Evapotranspiration (the combined loss of water from plant transpiration and soil evapotranspiration) that atmospheric conditions allow. Evapotranspiration (ET) is when water starting from an expansive scope of sources is moved from the soil and vegetation layer to the atmosphere. Water loss from a vegetative surface through the consolidated cycles of plant transpiration, soil and atmospheric evaporation. ET from the land surface is critical for maintaining the balance of land surface water-lakes-reservoirs as well as the energy balance of the earth's surface. The best possible assessment of ET is an essential issue in food security research, land management frameworks, contamination recognition, irrigation planning and scheduling, hydrological balance studies, and watershed hydrology. Knowledge of ET is essential while managing water resources and management problems, such as the stipulation of the water for irrigation, agriculture, drinking and industrial use, or water reserve management. ET also offers potential advantages for irrigation management. Many studies have also shown that at least 70% of surface precipitation returns to the atmosphere via ET, with the figure rising to over 90% in drought areas. Accurate ET estimations have critical applications in water resource assessment, vegetation drought monitoring, and ecological water use.

1.2 Motivation

Many researchers are interested in monitoring and simulating various hydrological processes in different regions [1]. The various hydrological processes that drive the hydrology of different regions can be simulated as a single system, which is complicated by the interdependence of the various processes. By monitoring and simulating these processes, one can gain a better understanding of hydrology and allow the implementation of more efficient water resource management and future reclamation designs. ET is a necessary hydrological process that must be monitored and modelled. Almost 62 percent of the precipitation that falls on continents is returned to the atmosphere via ET [2]. ET is the most significant annual loss of water in the sub-humid climate of northern Alberta [3], demonstrating its critical role in the hydrological system. ET can be expressed conceptually as either potential or actual Evapotranspiration. The maximum loss of water from a short green crop under specific climatic conditions when unlimited water is available is referred to as potential

evapotranspiration (PET). The rate of Evapotranspiration from a well-defined reference environment is defined as reference evapotranspiration (ET_0), a commonly used concept in engineering and scientific practises (e.g., well-watered short grass). The actual evapotranspiration (AET) rate is the rate at which water is removed from a surface to the atmosphere as a result of the evapotranspiration process. In hydrological analysis, AET is the preferred form of ET because, in most cases, limited water is available for ET, and the actual rate of water loss is of interest.

AET and PET are critical from a variety of perspectives, including reliable quantification of hydrological water balance, hydrological design, water resource planning and management, irrigation system design, and crop yield simulation. The understanding of the temporal variations of PET and AET time series, as well as the meteorological variables influencing them, can be considered a step forward in the overall goal of better understanding and management of hydrological systems. Hence, the exact calculation of PET and AET is fundamental in improving irrigation efficiency, water reuse and seepage control [4], [5]. Motivated by the increased number of ML models and neural networks involved in environmental and hydrological monitoring, the study proposed a method to estimate PET and AET from limited climate data using ML algorithms. There are several modifications to the standard FAO Penman-Monteith equation that enables us to use limited climatic data for estimating PET. However, these equations have to be adjusted locally depending on the different climatic conditions. This study has used five different ML models in order to determine the uncertainty of different models that explain PET and AET processes. AET and PET were modelled, estimated, and analysed for two different climatic stations to study the suitability of the models and input variables for diverse climatic conditions. More specifically, the uncertainties of hydrological data limitations and implementation of state-of-the-art ML models to estimate ET with a focus on region-specific parameters has formed the basis for the present study. The primary motivation of the study is to address the uncertainty due to limited hydrological data in estimating ET, the most prominent hydrological variable using the state-of-the-art ML models. The main hypothesis of the study is that compared to standard empirical-based models, which are region-specific and require various hydro climatological data, ML models can be promising in estimating ET, which can serve as a basis for many hydrological applications. The proposed hypothesis was studied by comparing the PET and AET estimates resulting from various empirical models and ML models for two different climatic regions of Hyderabad (India) and Waipara (New Zealand) under limited data scenarios.

1.3 Introduction

Evapotranspiration (ET) is one of the most critical components of the hydrological water cycle affecting terrestrial water-energy balances. AET is a significant component of the water balance and is generally utilised in agronomy, hydrology, climatology, meteorology, ecology and environmental sciences [6]–

[9]. Two more closely related types of ET are PET and reference evapotranspiration (ET_0). Although both PET and ET_0 provide estimates of atmospheric evaporative demand, they are based on different ideas, concepts, application fields and have different equations that can help to differentiate the terms. However, many researchers have treated PET and ET_0 as identical concepts and used similar equations for their estimation [8], [10]–[12]. Designers and specialists of hydrology presented the PET idea in the last part of the 1970s and mid-80s to stay away from the ambiguity that existed in the definition of PET [13]. The first idea of PET was proposed by Thornthwaite [14], and that core idea with improvements are being used now. The PET was characterized by Penman [15] as the maximal water amount moved to the climate, from vegetation spread in a condition of full physiological activity and unlimited water and supplement accessibility [14]. PET has been applied primarily in hydrology, meteorology, climatology, agronomy, agriculture, irrigation, and ecology. The accurate estimation of PET is essential in irrigation planning, scheduling, hydrological balance studies, and watershed hydrology [16], [17]. It has a broader significance in numerous fields of research, including crop yield simulation, optimization of water loss, management and irrigation system design, water usage improvement in agriculture, and hydrologic water balance.

The laws of mass or energy conservation or both were always associated with the evaluation of PET. In recent years, various PET estimating procedures and modelling methods are available in the literature. The PET equations were classified as temperature, radiation, and pan-evaporation based on [6]. PET assessments can be performed utilizing distinctive exploratory methods, for example, the leaf (porometer), an individual plant (for example, lysimeter), at the field scale (for example, field water balance, Bowen proportion, scintillometer) and landscape scale (for instance eddy correlation and catchment water balance) [18]. However, some of these techniques are not practical over a vast region because of regular maintenance and significant expense [19]. Allen [10] referenced that the main factor influencing PET is climatic variables so that PET can be surveyed by experimental and semi-experimental equations from meteorological data. Numerous strategies dependent on climatic data have already been proposed. However, the FAO suggested that the Penman-Monteith model can be utilized as the standard method to assess PET [10]. This equation has been utilized worldwide for benchmark ET assessments [10]. Nonetheless, when computing PET by the Penman-Monteith model, loads of climatic factors, including temperature, wind speed, solar radiation, and relative humidity, are required, which can be considered as significant limitations. In some cases, these factors are incomplete or not accessible in a given meteorological station, particularly in developing nations [20]. Consequently, it is fundamental to build up a more precise methodology that could compute PET with high accuracy, particularly in data-scarce regions.

Explicit standardized equations and procedures are being suggested for PET estimates and are typically modelled utilizing climate data and algorithms that depict surface vitality and aerodynamic qualities of the vegetation. Several empirical models for assessing PET with limited data can also be categorized

as mass exchange-based, temperature-based, radiation-based, pan-evaporation based, and combination type [20]. Penman[15]inferred a model for measuring evaporation from open surfaces by the mix of vitality offset with mass exchange techniques. Priestley & Taylor [21] proposed the radiation-based Priestley-Taylor model, a rearrangement of the Penman model. George H. Hargreaves & Zohrab A. Samani [22]proposed the temperature-based Hargreaves model, which was perhaps the least complicated strategy. Eleven temperature-based PET techniques like Thornthwaite [14]Mc cloud [23], Hamon [24], Hamon [25], Oudin [26], Baier and Robertson[27], Papadakis [28], Malmström [29], Hargreaves and Samani, Camargo [30]were evaluated for assessing PET and were discovered that the Hargreaves method gave an excellent performance in arid, semi-arid, temperate, cold, and polar atmospheres [13]. Because of the reliance on different climatic components, several researchers considered the calculation of PET as a complicated non-direct regression process and have progressed the assessment of PET models using computing procedures, for example, prepared Artificial Intelligence (AI), Machine learning models, and statistical regression approaches [31].

1.4 Problem definition

Evapotranspiration is one of the least measured components of the water cycle due to the expensive and time-consuming requirements for direct measurement methods [32]. As a result, for estimating PET, indirect methods ranging from empirical relationships to complex combined equations like Soil water balance, Weighing lysimeters, Micrometeorological approaches, Energy balance and Bowen ratio, Aerodynamic method, Eddy covariance plant physiology approaches, Sap flow method, Chambers System, Penman-Monteith, crop coefficient methods are used.

PET measurement methods are based on concepts that can be critical in semi-arid environments for a variety of reasons, including representativeness (for example, the weighing lysimeter), instrumentation (for example, air humidity sensors), microclimate (advection regime), and the hypothesis of applicability for the simplified aerodynamic method. Thus, in order to determine the degree of accuracy of the obtained PET measurement and the validity of a method, all of these parameters must be considered. Estimation methods based on an analytical approach may be very accurate, but they are usually insufficiently practical. AET is dependent on available soil moisture and, as a result, is region-specific, making its modelling or estimation more difficult than PET. AET is currently estimated indirectly and in relation to PET estimation models, using approaches that require soil moisture information in order to account for the water supply deficit in the estimation of the AET. However, in many cases, soil moisture is not readily available, limiting the applicability of this method. Given the drawbacks and limitations of current PET and AET modelling methods, there is an urgent need to develop techniques that can accurately estimate PET and AET values based on conventionally and readily available meteorological variables while also being simple to implement. Furthermore, it is

challenging to develop mechanistic models for some complex hydrological processes, such as PET and AET, because the underlying physics of the PET and AET processes can sometimes be too complicated to be accurately represented in a physically based manner. As a result, the data-driven modelling approach that can provide a model to predict and investigate the process without requiring a complete understanding can be helpful. The knowledge discovery property of the modelling approach is also appealing. Using machine learning models, one can extract useful implicit information from large amounts of data and improve one's understanding of the underlying process.

Machine learning (ML) techniques are modern data-driven modelling methods that arose as a result of advancements in computer technology and mathematical algorithms. These techniques are typically used to characterise complex systems that are difficult to understand, analyse, and model. Artificial neural networks (ANNs), Long Short term memory (LSTM), Random Forest (RF), Support vector regression (SVR), Gradient boosting regression (GBR) are few ML techniques that use artificial intelligence to model complex systems. These are the computational models that simulate the functional aspects and can be used to model complex relationships. These have been frequently employed in the characterization of the PET/AET process, and in the selection of limited inputs and in the selection of the best method to be employed in estimating PET/AET. The current study gives a comparison of five ML-based models to discover the best model for assessing daily PET and generating AET using the modelled PET under the state of minimal input variables in the semi-arid atmospheres in two unique areas, for example, Hyderabad, India and Waipara, New Zealand.

1.5 Objectives of the Thesis

The overarching goal of this research is to model and analyse the PET and AET hydrological process using data-driven techniques under limited input combinations. The purpose of this research is to develop different ML models, SVR, LSTM, RF and GBR, for modelling PET in Hyderabad and Waipara Stations and to assess the performance and stability of these models with different input combinations over the two stations. The purpose is also to find an appropriate approach to boost the modelling performance under the limited input factors condition. It would also be interesting to see if data-driven models can reveal anything about the PET and AET function and its most influential variables. The contribution of meteorological variables to PET and AET is also of interest and will be investigated using four different machine learning models as a modelling input determination approach. The study's specific goals are as follows:

- 1) To assess the applicability and validity of different empirical based-PET methods such as Priestley-Taylor, Hargreaves, and Turc methods in comparison with the standard Penman-Monteith model estimates for two different climatological regions.

- 2) To assess the applicability of different ML models, ANN, SVR, LSTM, RF and GBR in comparison with the standard Penman-Monteith model estimates for two different climatological regions.
- 3) To utilise and assess the best performance of the modelled PET in estimating AET for two different climatological stations.
- 4) To identify the most prominent meteorological variables influencing the PET and AET process under diverse climatic conditions
- 5) To find an appropriate empirical and ML approach to boost the modelling performance under the limited input factors condition.

1.6 Scope of the Study

The current study aims to create a framework for better understanding the dynamics of various hydrological functions that drive hydrology in different climatic regions. The overall findings of this study will aid the scientists in developing a better understanding of the hydrological processes and their need in estimating the terms of hydrological processes under limited data.

1.7 Thesis Organisation:

This thesis is organised into five chapters as follows:

- Chapter 1 provides a background, introduction to PET and AET and its motivation and objectives.
- Chapter 2 contains the detailed literature on PET and AET and the methods to estimate them. It also contained the detailed literature on the machine learning models employed in estimating PET and AET.
- Chapter 3 contains an introduction to the study area and the datasets.
- Chapter 4 contains different ML models used in the study.
- Chapter 5 contains estimation of PET using ML models and empirical methods in the context of limited data for different stations and their comparison to find the best method to be utilised to estimate PET under limited data.
- Chapter 6 contains an estimation of AET utilising the modelled PET from chapter 5 and their comparison between the two AET methods.

- Chapter 7 summarizes the study's research findings and discusses the applications, conclusions and potential extensions to the study.

Chapter 2

Literature Review

2.1 Introduction

Water availability is an essential factor in almost every other economic activity, including industry, the energy sector, and public use. Water availability has become an issue in recent years in most parts of the world, as periods of prolonged drought have stressed both agricultural and non-agricultural sectors. Since the 1980s, global water consumption has increased by about 1% per year, owing to a combination of population growth, socio-economic development, and changing consumption patterns. Stress levels will continue to rise as water demand rises and the effects of climate change worsen [33]. As a result of this increase, groundwater supplies have been depleted, which worsens during droughts as industries that usually use surface water switch to subsurface aquifers for water. In agricultural water management also the water availability and demand of crops play a major role. Water is naturally provided to crops through precipitation and subsurface moisture, but when these supplies are insufficient for crop use, agricultural water managers must resort to irrigation. Agricultural water managers must understand the environmental demand for surface water in order to schedule irrigation properly. This surface water loss is primarily caused by evapotranspiration (ET). Thornthwaite [14] defined evapotranspiration, both actual and potential, in 1944, and the term became widely known and used following his 1948 publication, in which PET was calculated as a complex empirical function of air temperature and day length. However, the word first appeared in print in 1937, albeit without explanation or definition and in a hyphenated form.

ET is the process of returning water to the atmosphere via evaporation from open water, soil, and plant surfaces, as well as transpiration from plants [34]. In theory, evaporation is a diffusive process that obeys Fick's first law and can be expressed as a function of vapour pressure deficit (at the evaporating surface and overlying air) and wind speed. Evaporation causes heat loss from the evaporating surface in the form of latent heat, which can be compensated for by radiative or sensible-heat transfer or by heat transfer from within the evaporating body to the surface [35]. Energy availability, water availability, vapour pressure gradient, wind, and atmospheric conductance are the four basic factors involving the evaporation mechanism. Any other parameters such as atmospheric pressure, water quality, water depth and soil type also affect the evaporation process [36].

Transpiration is the evaporation of water from a plant's vascular system into the atmosphere as a result of photosynthesis. Water is absorbed from the soil through the roots and transported to the leaves via the vascular system of the roots, stem, and branches. The water is then transferred from the leaf's vascular system to the stomatal walls, where it evaporates. Water vapour is then released into the

atmosphere via the leaf's openings known as stomata. Energy availability, water availability, humidity, temperature, ambient CO₂ concentration, and wind speed all limit transpiration. Plant species play a role by influencing the leaf conductance and the plant adaptation to water availability [35].

Evaporation component, including the return of water to the atmosphere via direct evaporative loss from the soil surface, standing water (depression storage), and water on surfaces (intercepted water) such as leaves or roofs, can be combinedly termed as Evapotranspiration (ET) [37]. Evaporation and transpiration happen simultaneously, and there is no easy way to tell them apart. Aside from water availability in the topsoil, evaporation from a cropped soil is primarily determined by the fraction of solar radiation that reaches the soil surface. As the crop develops and the crop canopy shades more and more of the ground area, this fraction decreases over the growing season. When the crop is small, water is lost primarily through soil evaporation; however, once the crop is mature and completely covers the soil, transpiration becomes the primary process. Because of the importance of ET in the water cycle and hydrological management, as well as the high cost and sensitivity of measuring equipment, extensive efforts have been made to model the ET mechanism. Many methods for estimating ET in various climatic conditions using various predictor variables have been developed, revised, and proposed. There are numerous methods for estimating evapotranspiration, ranging from those that account for evaporation from the water surface to a variety of PET estimates, but the majority of them only consider evapotranspiration from a single underlying surface, such as water, bare soil, or vegetation, while ignoring water balance. These methods treat evapotranspiration as a static quantity to be estimated rather than as an important process of the hydrological cycle. Recent advances in remote sensing methods can estimate basin-scale evapotranspiration, but due to technological limitations, this estimation is difficult to meet the diverse spatiotemporal scale requirement.

A related concept is PET, which is simply the amount of water lost from the surface to the atmosphere if the soil/vegetation mass had an infinite supply of water [34], [35], [37]. PET is the sum of soil evaporation and plant transpiration. It only happens at the potential rate when the amount of water available for this process is unlimited. The rate of evaporation is affected by climatic conditions, specifically the sun's radiative energy, wind, the air's vapour deficit, and temperature. The Penman-Monteith equation is frequently used to calculate PET from these measurements. It can also be estimated from readily available rainfall and temperature data using simple equations like Thornthwaite's, which has been done for over sixty years. Open pan evaporation measurements are another source of information [14]. Because PET assumes that water availability is not an issue, vegetation would never wilt (the point where there is not enough water left in the soil for a plant to transpire). As a result, the only limitation to the plant's transpiration rate is due to the plant's physiology, not to any atmospheric or soil moisture restrictions [34]. As a result, PET is defined as the maximum ET rate achievable with a given set of meteorological and physical parameters [35].

Another major form of ET which has wide application in water balance and crop water assessment studies is Actual Evapotranspiration (AET). AET is the rate at which water is removed from a surface to the atmosphere as a result of the evapotranspiration process. In hydrological analysis, AET is the preferred form of ET because, in most cases, limited water is available for evapotranspiration and the actual rate of water loss is of interest. Estimates of AET based on hydrological models take into account the influence of water and energy and can be calculated at various spatial and temporal scales, allowing the results of such estimations to meet the demand for water resource assessment and management [38]. There are numerous ET estimation methods based on existing hydrological models, each with its own set of data input requirements. Simultaneously, the accuracy of their output results is rarely compared. Several variables influence the AET process, such as (solar radiation, temperature, humidity, and wind speed [10], [35], [39]. Furthermore, the physical characteristics of the vegetation and soil play an important role in the ET process. Leaf shape, growth stage, crop height, and leaf albedo, for example, are all important factors in controlling transpiration functions [10]. Furthermore, stomatal resistance is an important factor. Stomatal resistance is defined as the restriction of water vapour diffusion back to the atmosphere by the guard cells surrounding the stomatal opening [39]. Finally, soil properties such as heat capacity, albedo, and soil chemistry can all impact ET [10]. These factors, together with stomatal resistance, are referred to as bulk surface resistance [10].

2.2 Developments in Estimation of Potential Evapotranspiration

As previously stated, PET is the ET from a vegetated surface with an infinite supply of water. However, because PET is still dependent on vegetation-specific characteristics (as previously mentioned) rather than solely meteorological variables, there was a determined need for a reference surface that was independent of vegetation and soil characteristics [10], [40]. This reference surface would allow for the analysis of the "atmospheric evaporative demand," leaving only meteorological factors to be considered [10], [40].

This simplifies ET calculation by generating a single surface against which other surfaces (e.g., different vegetation types) can be compared. Furthermore, using such an ET term would eliminate the need to vary the ET equation at various stages of vegetative growth [10]. This new type of ET referred to as PET, simply "expresses the evaporating power of the atmosphere at a specific location and time of year" [10]. PET is a term in which the transpiring vegetation has explicitly been defined.

Short, clipped grass and alfalfa have been used as common reference surfaces [10], [15], [40]–[46]. The reference surface (grass) has typically been chosen by researchers based on the availability of relevant data. Grass has bulk stomatal resistance and exchange values similar to many agricultural crops, but short, clipped grass has more experimental data. As a result, the FAO chose grass as the primary reference surface for international use [47]. It was also debated which model should be used as the

standard model for computing PET. Doorenbos [43] proposed four methods (FAO-24 Blaney-Criddle, FAO-24 Penman, FAO 24 Radiation, and FAO-24 Pan Evaporation) for estimation of PET. Smith [48] was the first to advocate using the Penman-Monteith model as the primary model for computing PET. This recommendation was made based on the model's previous performance and the model's incorporation of plant physiological and aerodynamic micrometeorological factors [10], [40], [44]. With the publication of Penman-Monteith in 1998, the Penman-Monteith equation was officially adopted as the FAO-recommended model [10].

With the Penman-Monteith equation chosen as the most dependent PET equation, the physical, physiological, and aerodynamic parameters for the reference grass had to be determined. The FAO selected parameters for a hypothetical grass with a crop height of 0.12 m, an albedo of 0.23, and a fixed surface resistance value of 70 s m^{-1} [10]. These parameters are very similar to the parameters of clipped Alfa fescue grass found in the weighing lysimeters in Davis, California- a site used extensively in PET research [10], [42]. The Penman-Monteith model was chosen as the PET standard, and the fixed hypothetical parameters standardizing the calculation of PET. As a result, the plant physiological and soil factors are ignored in the PET calculation. Furthermore, a baseline value of PET is calculated, allowing for objective comparison of PET across climates. It also allows for the simplified calculation of crop coefficients for various crop varieties, which are used to adjust ET to a value specific to a specific crop at a specific time in the crop's growth [10].

The modified Penman-Monteith equation being recommended for the calculation of PET and calibration of other equations by various international organizations such as the United Nations Food and Agriculture Organization (FAO) and World Meteorological Organization [10], [49]. The Penman-Monteith equation has two critical advantages. First, it can be used in a wide variety of environments and climate scenarios without any local calibrations because of its physical basis. Second, it is a well-documented method that has been validated using lysimeters under a wide range of climate conditions [50]. The main drawback of this equation is that it requires data on a large number of climate variables that are unavailable in many regions. The major limitation to the Penman family of models is that they require many meteorological inputs, thereby limiting their utility in data-sparse areas [35], [37]. The equation employs aerodynamic and surface resistance terms, with the aerodynamic resistance being relatively simple to calculate. Surface resistance, on the other hand, is difficult to calculate. According to the researchers [51], using the Penman-Monteith equation in advective conditions would underestimate PET because the equation is not completely capable of incorporating the horizontal flow of sensible heat flux.

Apart from the Penman-Monteith model, various PET methods have been developed and are being utilized, depending upon the availability of meteorological variables. Empirical models for PET estimation, i.e., statistical functions of approximation between meteorological variables and values, can overcome the difficulties associated with data availability for ET estimation [52], [53]. Among these,

Priestley-Taylor [21], Thorn Waite[14], Hargreaves [42], Turc (1954) are well-established models. These empirical models vary in terms of solar radiation, temperature considering the physical processes of radiation and transport characteristics of natural surfaces. An empirical model such as The Priestley-Taylor equation [21] can estimate regional monthly ET provided that the adjustment factor is adapted to different site conditions [54]. The Hargreaves and Samani equation is an empirical approximation of the PET calculation based on maximum and minimum temperatures and extra-terrestrial radiation data [42]. The Hargreaves model has conceptually similar versions, both of which are intended to be computationally simple and applicable to a wide range of climates using only commonly available meteorological data. This model was designed to simplify previous versions even further by limiting the amount of measured meteorological data to air temperature and by substituting extra-terrestrial radiation (R_a) for measured sunshine or radiation data [10], [42]. The FAO later adopted the Hargreaves and Samani model for use in areas where air temperature alone is insufficient.

Nevertheless, the superiority of the Penman-Monteith method over the Priestley Taylor equation has recently been demonstrated carried out surface polynomial regression analysis using hourly solar radiation, air temperature, and relative humidity (RH) to estimate PET [55], [56]. A much simpler alternative is the Thornthwaite scheme [14], as it requires only temperature as input data. However, this approach has been found to underestimate PET under arid conditions and overestimate in a humid climate [57].

To this end, over the last 50 years, numerous scientists and specialists worldwide have developed a large number of more or less empirical methods for estimating PET from various climatic variables. Relationships were frequently subjected to stringent local calibrations and were found to have limited global validity. Testing the accuracy of the methods under new conditions is time-consuming, labour-intensive, and expensive, yet evapotranspiration data are frequently required on short notice for project planning or irrigation scheduling design. Guidelines were developed and published in the FAO Irrigation and Drainage Paper No. 24 'Crop water requirements' to meet this need. To accommodate users with varying data availability, four methods for calculating PET were presented: the Blaney-Criddle, radiation, modified penman, and pan evaporation methods. In comparison to a living grass reference crop, the modified Penman method was thought to provide the best results with the least amount of error. The pan method was expected to provide reasonable estimates depending on the location of the pan. The radiation method was proposed for areas where available climatic data included measured air temperature, sunshine, cloudiness, or radiation but not measured wind speed or air humidity.

Several researchers examined the performance of the four methods (Blaney-Criddle, radiation, modified penman, and pan evaporation) in various locations. Although the results of such analyses could have been influenced by site or measurement conditions or by bias in weather data collection, it became clear that the proposed methods do not behave consistently across the globe [15], [41], [58].

Deviations from computed to observed values were frequently found to exceed FAO ranges. For low evaporative conditions, the modified Penman-Monteith was frequently found to overestimate PET by up to 20%. The other FAO-recommended equations demonstrated varying adherence to the standard PET. The comparative studies also summarised that to achieve satisfactory results; the penman methods may necessitate local calibration of the wind function. The radiation methods perform well in humid climates where the aerodynamic term is small, but in arid conditions, the performance is erratic and tends to underestimate PET. Temperature methods are still empirical and require local calibration to produce acceptable results. A possible exception is the Hargreaves method from 1985, which has produced reasonable PET results with global validity. The shortcomings of predicting PET from open water evapotranspiration are clearly reflected in pan evapotranspiration methods.

As the Penman-Monteith model has come to be accepted as the standard PET equation, many studies are looking into how other PET models, particularly those with fewer data requirements, compare to the PM model. Amatya [59] used data collected intermittently from 1982 to 1994 to compare Hargreaves-Samani, Priestley-Taylor, Makkink, and Turc to penman-Monteith at three sites in North Carolina. Amatya [59] discovered that Turc was the best model for simulating PET estimates at annual and monthly time scales comparable to the Penman-Monteith model. On a daily basis, Turc performed best at one site, while Priestley-Taylor and Makkink performed best at the other two sites. It was also discovered that the Makkink underpredicted PET during peak months, whereas temperature-based methods (including Hargreaves-Samani) overpredicted PET [59]. PET models typically outperform in the climates for which they were designed. According to Amatya [59], while the Makkink model performed well overall in North Carolina, it underestimated PET during the peak months of summer. Nonetheless, the Makkink model performs admirably in Western Europe, where it was developed, both in comparison to Penman-Monteith and measured PET data [32], [38], [60]. This observation is supported by research by [11], [61]. The implication is that some models, such as Makkink, may not perform well in humid climates. This may not be true in all circumstances. Several authors [11], [59], [62] demonstrated that the Turc model, designed in Western Europe [40], performs well in warm, humid climates such as those found in North Carolina [59], India [62], and Florida [11]. The best model to simulate the Penman-Monteith method is frequently determined by data availability. George [62] investigated a decision support system that chose the best PET model based on data availability and the climate of the location under consideration. They discovered that specific models, such as Hargreaves-Samani, perform best when only maximum and minimum air temperature data are available. George [62] also looked at PET estimation at three different locations, with daily and monthly comparisons in India and Davis, California, and a monthly-only site in India. Both sites in India have humid climates, whereas Davis has an arid climate [62].

When compared to the Penman-Monteith model, the FAO 24 Radiation model overestimated PET for Davis, while Priestley-Taylor and Turc both underestimated PET [62]. The results of [62] show that the

FAO 24 Blaney-Criddle model overestimated the Penman-Monteith method, whereas the Hargreaves-Samani model was within 1% of the penman model. The results of [62] show that the FAO 24 Blaney-Criddle model overestimated the penman method, whereas the Hargreaves-Samani model was within 1% of the Penman model. This is not surprising given that Hargreaves-Samani was created using Davis data [11], [62]. Priestley-Taylor and Turc tended to underestimate PET at the Indian sites, with the behaviour of the remaining models being site-dependent [62]. The study by [62] is similar to that of [11], which looks at the performance of PET models in Gainesville, Florida. The conclusions were similar, with the recommendation that most temperature models need to be calibrated locally if they were not designed for the climate in which they were used. Irmak [11] also pointed out that model selection is heavily influenced by the availability and quality of meteorological data. It should also be noted that while Irmak provides an informative study of PET models in a humid climate, the study only uses data from one location. Ravazzani [63] also compared the Hargreaves and Samani (1985) equation to the Penman-Monteith equation for daily time steps in alpine river basins and found the Penman-Monteith method performed well in India to calculate PET. They found that Penman's method seemed more realistic in estimating the mean annual PET distribution over India. The advantage of the Blaney Criddle method is the simplicity, and the disadvantage is that it underestimates PET grossly compared to the measured PET values [64]. The empirical models have the advantage of being simple and employing a limited number of meteorological variables; however, reasonable estimation of model parameters is required for local applications. This is regarded as a limitation of empirical PET prediction models.

2.3 Estimation of PET using Machine Learning Models

Most hydrologists and modellers are concerned with improving the predictive ability to understand the important terms of hydrological processes [65]. These data-driven modelling techniques have been developed as a new approach for simulation and prediction resulting from recent developments in hydrology using computer technologies and new mathematical algorithms [66]. In the past decades, there has been a widespread interest in the application of data-driven modelling and ML techniques in the field of water resources and hydrology [67], [68].

Although several methods for predicting PET have been developed around the world, there are only a few models for estimating PET when meteorological data is limited or insufficient. Empirical models can be the best choice for estimating the PET given the availability of meteorological variables. As a result, to overcome this limitation, several researchers considered the calculation of PET as a complicated non-direct regression process and have progressed the assessment of PET models using computing procedures, for example, Artificial Intelligence (AI), Machine learning models, and statistical regression approaches. Extensive research has been conducted in order to develop more

reliable and efficient models to estimate PET. The major types of ML models used for PET estimation were based on Artificial Neural Network (ANN), Support vector regression (SVR), Gradient boosting methods (GBR), Random forest methods (RF), Long short term memory (LSTM), deep learning models, gene expression programming, etc. Granata (Granata, 2019) applied Support Vector Regression (SVR) and Regression Trees to predict wastewater quality indicators in urban catchments. Tree models were used to evaluate sediment transport [69], to predict flood [38], to make predictions of the mean annual flood [70], to estimate scour depth due to waves [71], to forecast sediment yield in rivers [72]. Ensemble methods were employed to tackle wastewater hydraulics problems [73]. Further examples of relevant applications of machine learning techniques to hydrological prediction problems are provided by the works of [38], [73]–[78]. RF models are also one of the supervised ML approaches, which are popular in ML [69], [79]–[82] and are frequently used in hydrology [7], [83], [84]. They tend to have a low bias; also, you can increase the depth to decrease bias error. In several research studies, SVR, which is focused on systemic risk minimization to prevent overfitting [84]. SVR was adopted over ANN due to the solution's uniqueness and globalization [85]. The SVR has been commonly used in engineering [79], [86], [87]. Its evapotranspiration applications are also quite impressive [74], [88]. Dibike [2] firstly applied the SVM approach for rainfall-runoff modelling in hydrology. As a data-driven technique, ANNs can determine the critical model inputs [89]. The Artificial Neural Network (ANN) has also been a widely applied machine learning algorithm in water resources engineering, including PET [67].

In recent years, ANN algorithms have largely been applied in the field of PET estimation. Kumar [67] developed ANN models for the estimation of PET and found that the ANNs could predict PET better than the conventional empirical methods. More recently, Ehteram [90] investigated the modelling of PET using ANNs with the Levenberg Marquardt training algorithm and inferred that ANNs could be employed successfully in modelling PET from available climate data. Jain [91] interpreted the physical meanings of ANNs for PET estimation. Some of them utilized the comparable climatic data required for the application of the FAO Penman-Monteith method [47], [92], [93]. These researchers reported that the ANN can anticipate PET ever better than the FAO Penman-Monteith conventional method. [94], [95] simplified the input variables, and PET was evaluated as a function of air temperature, extra-terrestrial solar radiation, and daylight hours. The performance of four climate-based methods and Artificial Neural Networks (ANNs) was compared for PET estimation when input climatic parameters are insufficient to apply the FAO Penman-Monteith method [96]. They concluded that ANN models performed better than the PET climatic methods. Suryavanshi [97] examined the trend in temperature and PET over the Betwa basin, India. In another study, Rahimikhoob [98] applied the different ML techniques to estimate PET based on air temperature data under humid subtropical conditions on the southern coast of the Caspian Sea situated in the north of Iran. The study showed that ANN successfully estimated the daily PET better than the Hargreaves

classical equation. Adamala [99]made a comparison of developed models with the artificial neural network models and also with the linear and wavelet regression and conventional methods to estimate PET using temperature-based generalized wavelet-neural network models.

Moreover, Traore[31]detailed gene expression programming (GEP), with its capacity to represent algebraic equations, to be a reliable procedure for modelling the decadal PET of six districts in Burkina Faso. Citakoglu [100]found a versatile neuro-fuzzy inference system (ANFIS) that outperformed Artificial neural networks (ANN) in modelling the monthly mean PET. Additionally, Kisi [101]evaluated the ANN, GEP, and ANFIS-grid partitioning (ANFIS-GP) models to display monthly PET, discovered that the ANFIS-GP model performed well compared to other models. Yassin [78]discovered ANN models to give more superior accuracy than GEP models in comparing GEP and ANN for modelling daily ET. Moreover, Patil [102],assessing week by week PET at Pali and Jodhpur, India, related the performance of ANN, least-square SVM, and extreme learning machine (ELM) approaches. The ELM gave preferable PET estimates over the other two models. Antonopoulos [103]found that the ANN model gives preferred outcomes over the experimental conditions in assessing ANN's performance and four empirical methods for demonstrating everyday PET at the Amineto climate station in Greece. Granata[73]assessed the PET in Florida, USA, utilizing M5P regression tree, SVR, and RF procedures and announced that these techniques were all prepared to model PET in the study area. Nourani [104]used ANFIS, feed-forward neural networks (FFNN), and SVR ensemble-based models to model PET at 14 stations in Iran, Iraq, Libya, Turkey, and Cyprus. They found the ensemble techniques improved the performance of traditional ML models. PET demonstrating methods that likewise incorporate SVR and RF were applied by [36], [88], [101], [105], [106]. Assessment of the precision of regression and machine learning approaches recommends that such methodologies estimate better PET over empirical models, such as the Thornthwaite and Hargreaves, which utilized limited meteorological factors. Chauhan and Cobaner [96], [107]compared the performance of four atmosphere-based methods and ANNs to assess PET when input climatic parameters were insufficient to apply the FAO Penman-Monteith method. They inferred that ANN models performed superior to climatic techniques. Suryavanshi [97]inspected the temperature and PET pattern over the Betwa bowl, India. Sonali and Nagesh [93]examined the pattern of minimum and maximum temperature of yearly, monthly, winter, pre-monsoon, monsoon, and post-monsoon. Assessment of the PET of Punjab was done by Saggi [108]dependent on different machine learning models, including Deep Learning-Multilayer Perceptron's (DL), Generalized Linear Model (GLM), Random Forest (RF), and Gradient-Boosting Machine (GBM) models. It was compared in predicting daily ET with the deep learning model's performance and was compared with the Penman-Monteith model and found that deep learning models performed superior to the considered models for training, validation and testing sets. Pal [109]examined the unique data-driven based regression approaches to deal with daily PET modelling utilizing four data inputs, including average air temperature, average wind speed, average relative

humidity, and solar radiation. Results from their examination proposed that the unique data-driven and AI models could effectively be utilized in modelling the PET. [106] investigated the viability of the deep learning neural network (DLNN) for estimating the PET. Most of the existing studies [78], [86], [90] have used machine learning models, such as Artificial Neural Network (ANN), Fuzzy Logic, Gene Expression Programming (GEP), Multivariate Adaptive Regression Splines (MARS), Decision Tree (DT), Random Forests (RFs), Support Vector Machine (SVM), Extreme Learning Machine (ELM) and Adaptive Neuro-fuzzy Inference System (ANFIS) to estimate PET. Different studies used ML models to estimate PET, but the present study emphasised on to estimate PET using ML models under limited climate data to achieve best input combination for diverse climatic conditions. To this end, machine learning methods have been promising in data-scarce regions under limitations of the empirical methods and as the Penman-Monteith equation requires a large amount of data input, which is not available in many ungauged basins.

2.4 Estimation of AET from ML models

AET is the quantity of water that is actually removed from a surface due to the evaporation and transpiration processes. It explains how water and energy are exchanged between the soil, land surface, and atmosphere. Improved characterization of AET is significant for modelling and management of water resources and related ecosystem services, such as provisioning, supporting, and regulating services and addressing global climate change. Because climate change affects AET rates, soil moisture, vegetation productivity, the carbon cycle, and water budgets may all be impacted [110]. The rising water demand across the regions also highlights the importance of finding alternative methods to quantify AET.

Earlier studies have estimated AET using time-consuming, labour-intensive methods such as water-balance, energy-balance, Bowen-ratio (EBBR), eddy-covariance (EC), and a few modelled based methods. The modelled based methods include Thornthwaite and Mather equation, Coutagne, Penman, Serra, Cappus, Kessler, Jensen and Haise, Blaney and Criddle, and Papadakis [111]. AET has been recently measured using the eddy-covariance (EC) method [112]. Although this method has proven to be a reliable technique, it is costly and still rarely available worldwide. Weighing lysimeters have also been considered as a viable tool for measuring AET [113]. However, most of the studies focus on the use of weighing lysimeters [113], whose construction and operation is still costly. For that matter, some authors have constructed volumetric lysimeters as a lower-cost alternative [114], [115]. When due to their high cost, complex installation, and/or intensive maintenance, no measurement techniques are available, estimating AET can be a valuable alternative for agricultural or hydrological studies. AET estimation is also difficult because it is influenced by climate, vegetation, soil moisture, and the amount

of available water, among other factors. Thus, there are various methods, including the water balance, the energy balance, and hydrological models. The water balance has been used as a reference method to estimate AET globally. However, closing the water balance requires not only the measurement of precipitation and discharge but also, in some cases, the measurement of other variables that are difficult to quantify. Also, such water balance estimates work efficiently at higher temporal time scales such as annual. Such traditional methods for estimating AET at catchment scales uses the water balance equation, which is commonly referred to as observed AET. However, the water balance AET estimates are inapplicable for analysing the spatial variability across large study areas. Various modelling approaches based on readily available hydro-meteorological data have been developed to estimate AET. Such approaches are based on first estimating PET and then applying a limiting factor to account for water availability [116], [117]. However, AET modelling approaches are region-specific for a given set of climatological conditions, and implementation of such models necessitates parameter calibration [58]. Catchment parameter estimation dynamically with precipitation, evapotranspiration and runoff using machine learning algorithms was done for semi-arid river basin of India by Sireesha [118]. Ochoa [119] estimated AET using the following methods: eddy-covariance (EC), volumetric lysimeters (Lys), water balance (WB), energy balance (EB), the calibrated Penman-Monteith equation (PMCal), and two hydrological models [the Probability Distribution Model (PDM) and the Hydrologiska Byran's Vattenbalans Avdelning model (HBV-light)].

As the vast majority of studies in the literature have focused on modelling of the PET process, in which evaporation occurs from soil and plant surfaces under no water stress. However, AET occurs under actual water supply conditions. AET can now be quantified using time-consuming and labour-intensive methods such as water-balance, energy-balance-Bowen-ratio (EBBR), and eddy-covariance (EC), Micrometeorological estimation (observation) methods such as EBBR and EC [120]. Because theoretical modelling of the AET mechanism is difficult, its values are currently estimated using available PET models, crop coefficient (K_c) (as an indicator of actual vegetation), and soil moisture information. This method essentially adjusts the estimated PET values for the actual plant and soil water conditions investigated.

In comparison to PET, there were few studies in the literature that investigated the modelling of the AET mechanism. Slabbers [121] developed a simplified method for predicting AET based on PET, crop-dependent critical leaf water potential, and the fraction of available soil moisture. Poulouvassilis [122] used meteorological, crop, and soil data to develop a simple semi-empirical approach for estimating AET. AET was also estimated using the relationships developed between AET and pan evaporation [45], [123], [124] and between AET and Penman's potential evaporation. According to Slabbers [121], the concept of AET is frequently restricted to semi-empirical models [125], [126], which have several limitations. Some equation-based PET models, such as the Penman-Monteith equation and Shuttleworth and Wallace's work, have also been adapted for AET estimation (1985). In cases where

the theoretical assumptions of the Penman-Monteith method are not valid, the model parameters (e.g. aerodynamic resistance of leaf surface) should be specified in the Penman-Monteith method for the estimation of AET (e.g., low soil moisture conditions). The Priestley-Taylor (PT) method was also modified to estimate AET using an empirical parameter [127]. Gavin [128] discovered a strong relationship between the empirical parameter of the Priestley Taylor method and the soil moisture condition. The proposed AET models necessitate many predictor variables, including meteorological parameters, soil moisture information, leaf area, and canopy aerodynamic characteristics. The most common issue encountered in the application of currently available models is a lack of required information. According to [122], determining critical parameters (e.g., threshold soil moisture and threshold leaf water potential) is also a significant challenge in estimating AET using available models. In this context, some efforts were made on the use of ML models to estimate AET. In this context, [1] modelled the AET process using GP and ANN as a function of net radiation, ground temperature, air temperature, wind speed, and relative humidity. Their analysis revealed that the effects of net radiation and ground temperature on AET outweighed the effects of other variables. The majority of studies have focused on modelling potential and reference crop evapotranspiration rather than actual evapotranspiration (AET). Most of the existing studies by [1], [73], [94] have also used ML models like support vector regression, random forest and regression trees to estimate AET. Different studies used different ML models to estimate PET, but the present study has highlighted the use of ML modelled PET for estimating AET under limited climate data for distinct climatic conditions. Currently, very little is known about which methods are suitable for accurately measuring or estimating AET at many regions, and studies have not compared AET by implementing several methods simultaneously at the same regions under limited data.

2.5 Conclusions

Despite its importance in water balance, evapotranspiration is one of the least measured components of the water cycle, most likely due to the expensive and time-consuming requirements for direct measurement methods. As a result, indirect methods ranging from empirical relationships to complex combined equations have been used for estimating ET. The major forms of ET in understanding the hydrological process are PET and AET. PET represents the atmospheric water demand with a focus on climatological variables. Whereas AET is dependent on available soil moisture and, as a result, is region-specific, making its modelling or estimation more difficult than PET. PET is currently estimated by considering Penman-Monteith developed by FAO as the standard model, which requires various climatological information. Whereas AET is currently estimated indirectly using approaches that require soil moisture information in order to account for the water supply deficits. To use these methods, users must make reasonable estimates for some of the parameters which involve some uncertainties and

may not yield reliable PET and AET estimates. Furthermore, in many cases, soil moisture and other vast climatological variables are not readily available, limiting the applicability of conventional PET and AET methods. Given the drawbacks and limitations of current modelling methods, there is an urgent need to develop techniques that can accurately estimate PET and AET values based on readily available meteorological variables while also being simple to implement. It is challenging to develop mechanistic models for complex hydrological processes, such as AET/PET, because the underlying physics of the AET/PET process are too complicated to be accurately represented in a physically based manner. As a result, an inductive (data-driven) modelling approach that can provide a model to predict and investigate the process without requiring a complete understanding of it can be useful. In literature, there has been wide usage of PET to estimate AET, but the usage of modelled PET from machine learning models has been limited. Furthermore, how the limited data availability can impact the PET and AET estimates for diverse climatological conditions has not been explored. Therefore, the present study made use of different ML models to estimate PET and AET under limited data for two different climatic regions.

Penman-Monteith method is a well-known empirical method and has been widely used in the past for data mining and function estimation problems. Despite significant advancements in data-driven modelling, penman-monteith remains popular and is used for a variety of modelling and model comparison problems. In this study, this technique was used as a benchmark modelling method. Artificial neural networks (ANNs), Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting Regression (GBR), Long short term Memory (LSTM) are a few such robust machine learning (ML) techniques that were used to model PET. These are the computational models that simulate the functional aspects of biological neural networks and can be used to model complex relationships between various meteorological variables and PET/AET. These techniques have been widely used for modelling the PET process but not significantly in the generation of AET. The current study gives a comparison of five ML-based models to discover the best model for assessing daily PET under the state of minimal input variables in the semi-arid atmospheres in two unique climatic conditions, for example, Hyderabad, India and Waipara, New Zealand. By utilising available time-series data and data-driven algorithms, we can also improve our understanding of PET and AET as well as its correlation with the interacting meteorological variables. More importantly, the study made efforts to use data-driven algorithms in the prediction of PET and AET with limited data, which are more promising with the standard modelling approaches for various climatic conditions.

Chapter 3

Study Area and Data

3.1 Study Area

The study has selected two weather stations with diverse climatological conditions. The first one is Hyderabad which is the largest city of the Indian state of Telangana which lies between latitude 17.3850°N , 78.4867°E located on the Deccan Plateau in the northern part of South India and covers an area of 650 square kilometres (250 sq. mi) at an elevation of 542m (Figure 3.1). Based on the Koppen climate classification, the climate is tropical wet and dry bordering on a hot semi-arid, with an average annual precipitation of about 171 mm(<https://en.climate-data.org/asia/india/hyderabad>). Much of Hyderabad is built on hilly terrain around artificial lakes, including the Hussain Sagar lake, which predates the city's founding and is located north of the city centre. According to the 2011 Census of India, Hyderabad is the fourth-most populous city in India, with a population of 6.9 million within city limits and a population of 9.7 million in the metropolitan region, making it the sixth-most populous metropolitan area in India. Hyderabad has the fifth-largest urban economy in India, with a GDP of US \$74 billion.

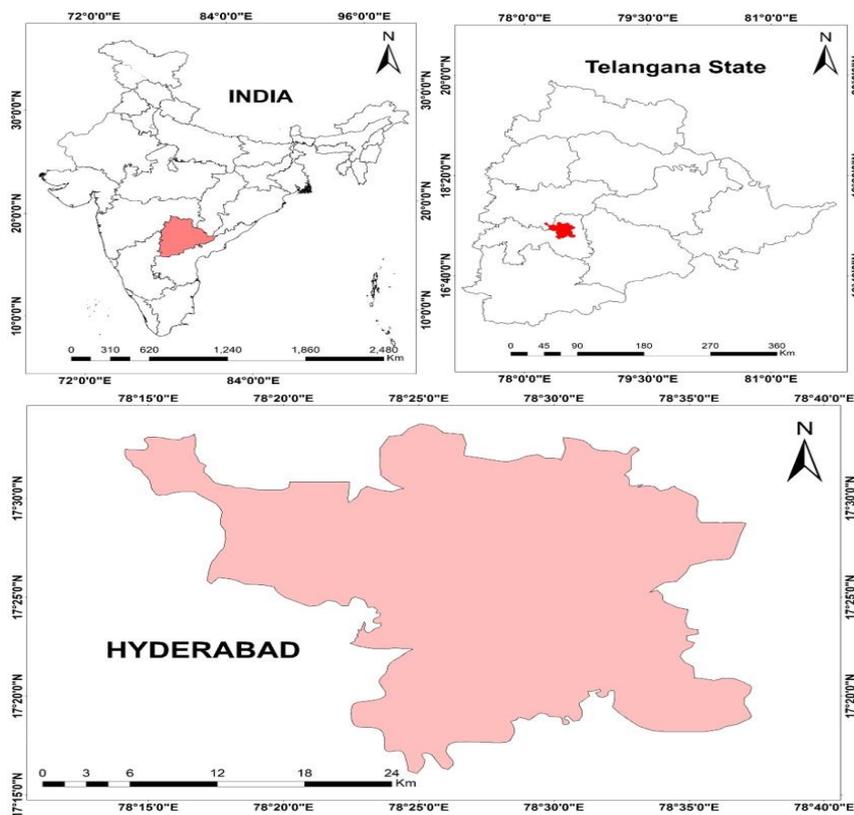


Figure 3.1: Case study: Hyderabad, Telangana, India

The next station was at the Waipara experimental catchment (WARVEX), situated in the South Island of New Zealand, in the Waipara River (Figure 3.2). The annual average weather data of the meteorological station is introduced in Table 2 (McMillan et al., 2011; Singh et al., 2017) provides a detailed overview of the basin and its monitoring network. Only a description of those elements of the basin important to the current study is given here. WARVEX is set up in Langs gully.

The catchment area of the Langs gully is 0.7 km². The elevation varies between 500 m and 723 m above sea level. The annual rainfall varies from 500 to 1100 mm/yr. It contains a surface slope of 0.22-34 degrees with a mean slope of 17 degrees. Soils are gravelly sandy loam, depth ranges from 0.25 to 1.5 m and averages 0.5 m. Grass and exotic forests are the primary vegetation. An ephemeral stream flows approximately from late March through early November. The catchment has fairly regular frosts and occasional snow in winter.

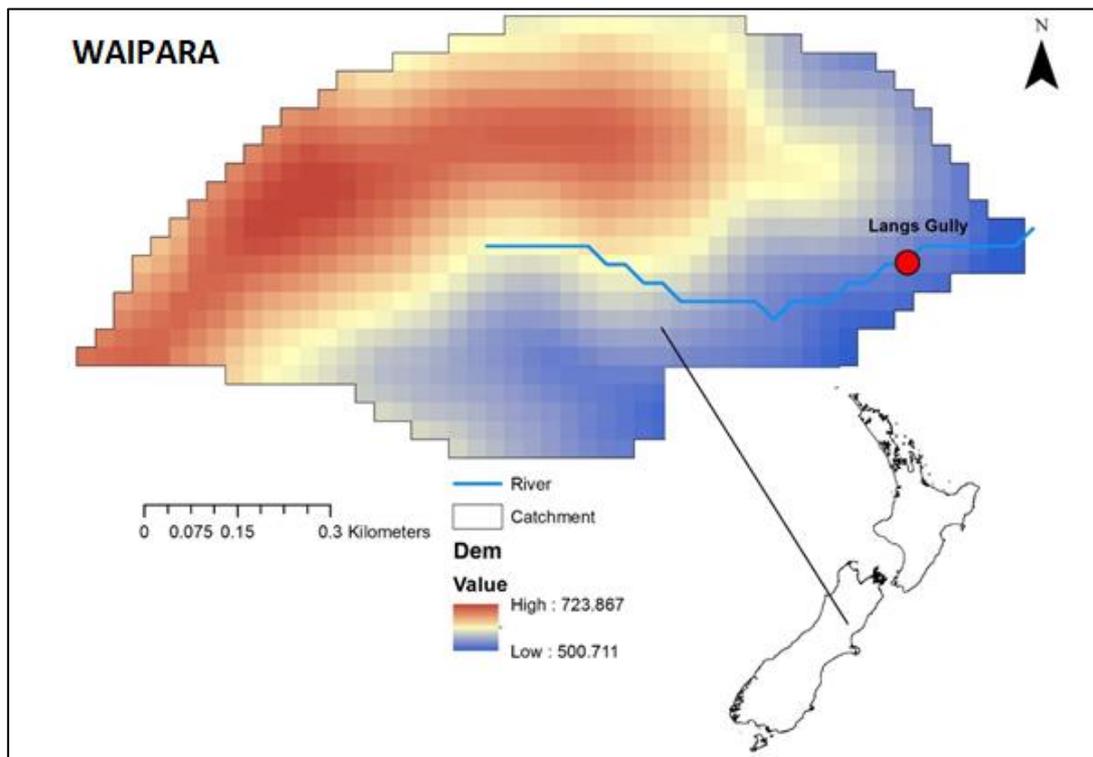


Figure 3.2: Case study: Waipara, New Zealand.

3.2 Data sets

3.2.1 Hyderabad Data Set

Daily meteorological data were obtained from January 1965 through December 2015 (51 years) (612 months) from weather station situated in Professor Jayashankar Telangana State Agricultural

University, Rajendran agar Mandal Hyderabad, Telangana. The annual average weather data of the meteorological station is presented in Table 3. 1. Six meteorological variables were recorded at a daily time scale, including (1) maximum air temperature (T_x °C) (2) minimum air temperature (T_n °C) (3) minimum relative humidity (RH, %) (4) wind speed (U_2 , $m\ s^{-1}$) and (5) solar radiation (R_s , $MJ\ m^{-2}\ d^{-1}$). Measurements were carried out at 2 m (air temperature and relative humidity) and 10 m (wind speed) above the soil's surface. Data on wind speeds at 2 m (U_2) were obtained from those taken at 10 m using the log-wind profile equation.

On an average, throughout the year, rainfall duration exists for 64 days in Hyderabad, collecting up to 828.5mm (32.62") of precipitation. February has the most sunshine of the year in Hyderabad, with an average of 9.6 hours of sunshine per day. March through September have the highest UV index, with an average maximum UV index of 12. April is the least humid month in Hyderabad, India, with an average relative humidity of 37%. May has the highest average high temperature of 39°C (102.2°F) and the lowest average temperature of 26.2°C (79.2°F). June has the longest days of the year in Hyderabad, India, with an average of 13.2 hours of daylight. With an average relative humidity of 74%, August is the wettest month in Hyderabad, India. The month with the most rainfall is August, when it rains for 14.1 days and averages 178.7 mm (7.04 inches) in the form of precipitation. August has the least amount of sunshine in Hyderabad, with an average of 4.4 hours. December is the coldest month in Hyderabad, with an average high temperature of 28°C (82.4°F) and a low temperature of 14.5°C (58.1°F). December has the least amount of rain for an average of about 0.7 days with a total of 5.9 mm (0.23 inches). December has the shortest days of the year, with an average of 11.1 hours of daylight.

3.2.2 Waipara Data Set

Field data from Lang gully was collected from 2010 to 2016. All data were stored in data loggers and had temporal resolutions of 10 minutes and have been aggregated to the hourly time series for this study to match the model time step. The annual average weather data of the meteorological station is presented in Table 3.2. For more details please see [70], [129]. Seasons in the Waipara, as in the rest of Canterbury, vary dramatically due to the influence of the Southern Alps. Three mesoscale wind systems characterize the region: westerly fronts blowing from the Tasman Sea that bring rainfall across the Southern Alps to the western parts of the catchment, moist southerly cold wind blowing across the Tasman Sea and the Southern Ocean across the South Island, and easterly fronts typically blowing from the north over the Pacific Ocean. Long dry spells are common in summer when hot, dry north-westerly winds blow, leaving temperatures ranging from 21°C to 32°C, creating dry microclimatic conditions that are frequently cooled by a north-easterly sea breeze. Snow is also common in the mountains during the winter, lowering daytime winter temperatures to around 7°C to 14°C(<http://www.northcanterbury.co.nz/NorthCanterbury/location-climate/>)

Table 3.1: Statistical values of available meteorological variables and PET at Hyderabad station.

Parameters	T_x (°C)	T_n	RH	Rs	U₂	PET
Maximum	45.5	33.0	139	14.45	36.00	13.17
Minimum	17.6	5.00	10	4.0	0.00	0.48
Mean	32.37	19.88	60.70	9.32	4.69	3.76
Standard deviation	4.1	4.79	14.93	2.44	4.62	1.72

Table 3.2: Statistical values of available meteorological variables and PET at Waipara station.

Parameters	T_x	T_n	RH	Rs	U₂	PET
Maximum	24.32	24.15	93.54	47.00	6.00	8.27
Minimum	-3.00	-3.00	0.00	0.00	0.00	0.00
Mean	10.17	9.95	65.61	15.29	1.75	1.50
Standard deviation	4.92	4.88	15.27	10.11	0.82	1.18

Table 3.3: Climate classification according to Aridity Index.

Climate Type	Aridity Index (AI)
Hyper arid	$AI < 0.05$
Arid	$0.05 \leq AI < 0.2$
Semi-arid	$0.2 \leq AI < 0.5$
Dry Subhumid	$0.5 \leq AI < 0.65$
Humid	$AI > 0.65$

Table 3.4: Climatic classification of Hyderabad and Waipara Stations

Station	Precipitation	PET	P/PET (calculated)	P/PET (original)
Hyderabad	384.24	434.42	0.33	0.2-0.5 (Semi-arid)
Waipara	146.98	1596.38	0.24	0.2-0.5 (Semi-arid)

3.2.3 Climate classification

The Aridity Index (AI) is a simple but useful numerical indicator of aridity that is calculated as the P/PET ratio and is based on long-term climatic water deficits. The AI is a widely used measure of the dryness of a given location's climate. The AI classifies arid lands or drylands into six subtypes: cold, hyper-arid, arid, semi-arid, dry sub-humid, and humid.

$$AI = \frac{P}{PET} \quad (3.1)$$

Where P-precipitation, PET-potential evapotranspiration.

In reference to Table 3.3, it was observed that both the stations fall under semi-arid regions (Table 4). The Food and Agriculture Organization, Agro-Ecological Zone concept, classifies parts of the world as

semi-arid if the annual precipitation totals between one fifth and one half of the PET (i.e., $0.2 < P/PET \leq 0.5$) from Table 3.4 . This equates to 60 to 180 days of plant growth opportunity on average. It is the climate of a region with precipitation that is less than PET but not as low as a desert climate.

3.3 Conclusions

The study has selected two different climatic regions of Hyderabad and Waipara. The study has concluded that the climatic zones fall under the semi-arid category for both the stations. However, the conditions vary from extremely wet on the West Coast of the South Island to almost semi-arid in Central Otago and subtropical in the Northland of New Zealand. It was observed that minimum temperatures have varied predominantly. It was also concluded that they both are different climatological conditions as they may not follow the same months of higher temperatures and rainfall. Due to the diverse climatological conditions of these two stations, the present study considered Hyderabad and Waipara as case studies to understand the complex phenomenon of PET and various climatological variables.

Chapter 4

Machine Learning Models

Data-driven modelling techniques have been developed as a new approach for simulation and prediction of various natural and artificial phenomena due to recent developments in computer technologies and new mathematical algorithms. These new techniques are especially relevant in hydrological modelling, which is widely used to simulate complex and poorly understood natural processes. ANNs, SVR, GBR, RF, LSTM are few data-driven techniques that do not require a thorough understanding of the physics of the processes under investigation. These machine learning techniques used in this study for modelling the PET process are listed. Using these techniques for modelling and predicting various complex processes has been studied in the literature and is promising. Furthermore, in the literature review, a brief history of each technique's development is provided to highlight the long road acquisition (measuring) methods; these modelling techniques are becoming even more popular.

4.1 PRE-PROCESSING METHODS

To determine the ranking of the feature set and to normalize the data set, variable variance and linear methods such as Principal Component Analysis (PCA), Pearson Correlation (COR) were applied to abstract the most correlated variables. The data were first normalized to within a range of [0-1] using :

$$X_n = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (4.1)$$

4.1.1 Pearson Correlation (COR)

The Pearson's Correlation Coefficient is also known as the Pearson Product-Moment Correlation Coefficient. It is a measure of the linear relationship between two random variables, x and y. It is the covariance of two variables divided by the product of their standard deviations.

$$\text{COR} = \frac{\sum(X_i - X_{\text{mean}})(Y_i - Y_{\text{mean}})}{\sqrt{\sum(X_i - X_{\text{mean}})^2 \sum(Y_i - Y_{\text{mean}})^2}} \quad (4.2)$$

X_i -values of the x-variable in a sample, X_{mean} -mean of the values of the x-variable, Y_i -Values of y-variable in a sample, Y_{mean} - mean of the values of the y-variable

4.1.2 Principal Component Analysis

PCA is a statistical procedure employed in extracting variables in the form of components from a broad set of variables within a dataset, and the resultant component is the linear composition of the main variables. PCA is commonly used to reduce the data dimension (i.e., the number of variables measured

per view) using the communication description between variables, thereby drawing complete information from the original data set while maintaining a minimal error between the original data set and the new dimensions obtained.

The PCA extracts a set with a low dimension of features from one of high dimension. In doing so, variables in a multimode correlated area are a collection of non-correlated components, each of which is a linear combination of the main variables. In general, the PCA method is mainly applied to reduce the number of variables and find the relationship structure between similarly classified variables. In doing so, the primary input is converted to new variables that are non-correlated, i.e., the resultant component is a linear composition of the primary variable [130]. The PCA will have changed the input variables to main components that are independent and linear combinations of the input variables, thereby suffering minimal losses in the main components [131]. The critical advantage of this method is to eliminate the linear model contained within the models, given the many influential variables in the model. In the present study, the main components of the first component, which bore at least eighty percent of the total variance in the data, served as the input to the estimation models.

The k th principal component of a data vector $x_{(i)}$ can be given as a score $t_{k(i)} = x_{(i)} \cdot w_{(k)}$ in the transformed coordinates, or as the corresponding vector in the space of the original variables, $\{x_{(i)} \cdot w_{(k)}\} w_{(k)}$, where $w_{(k)}$ is the k th eigenvector of $X^T X$.

4.1.3 LASSOMODEL

In statistics and machine learning, lasso (least absolute shrinkage and selection operator; also LASSO) is a regression analysis method that performs both variable selection and regularization to enhance the prediction accuracy and interpretability of the statistical model it produces. Though originally defined for least squares, lasso regularization is easily extended to a wide variety of statistical models, including generalized linear models, generalized estimating equations, proportional hazards models, and M-estimators, in a straightforward fashion. Lasso's ability to perform subset selection relies on the form of the constraint and has a variety of interpretations, including in terms of geometry, Bayesian statistics, and convex analysis.

The goal of Lasso Regression is to obtain the subset of predictors that minimizes the prediction error for a quantitative response variable. The lasso does this by imposing a constraint on the model parameters that causes regression coefficients for some variables to shrink toward zero. Variables with a regression coefficient equal to zero after the shrinkage process are excluded from the model. Variables with non-zero regression coefficient variables are most strongly associated with the response variable. Therefore, when you conduct a regression model, it can be helpful to do a lasso regression to predict how many variables your model should contain.

4.2 Artificial Neural Network (ANN)

ANNs are massive networks of parallel information processing systems that resemble (simulate) the analytical function of the human brain, and they have an inherent ability to learn and recognize highly nonlinear and complex relationships through experience. It is a computational model inspired by networks of biological neurons, wherein the neurons compute output values from inputs. It learns from its past experience and errors in a nonlinear parallel processing manner. ANNs are fully connected neural nets that consist of an input layer, hidden layers (multiple or single), output layer. Each node can be considered as a neuron. It is also composed of a large number of highly interconnected processing elements (neurons) working in unison to solve a specific problem. The neuron is the primary calculating entity that computes from several inputs and delivers one output compared with a threshold value and turned on (fired).

Neurons:

Biological Neurons (also called nerve cells) or simply neurons are the fundamental units of the brain and nervous system, the cells responsible for receiving sensory input from the external world via dendrites, process it and give the output through Axons.

Activation Function:

The activation function decides whether a neuron should be activated or not by calculating the weighted sum and further adding bias to it. The motive is to introduce non-linearity into the output of a neuron.

The computational processing is done by internal structural arrangement consisting of hidden layers that utilize the backpropagation and feed-forward mechanism to deliver output close to accuracy. Fully connected neural nets are those where each node in a layer is connected to every other node in the next layer (right). Each node takes the weighted sum of its inputs which then passes through a nonlinear activation function (like RELU, sigmoid, tanh, etc.), which then becomes the input of other nodes in the next layer. In Eq.4, the function, f , represents the activation function and w is the weight matrix, X is the set of input vectors. (<https://medium.com/@ariesiitr/an-artificial-neural-network-ann-is-a-computational-model-that-is-inspired-by-the-way-biological-c17b07166d4c>)

$$Z = f(x \cdot w) = f\left(\sum_{i=1}^n x_i w_i\right) \quad x \in d_{1 \times n}, w \in d_{n \times 1}, z \in d_{1 \times 1}, \quad (4.3)$$

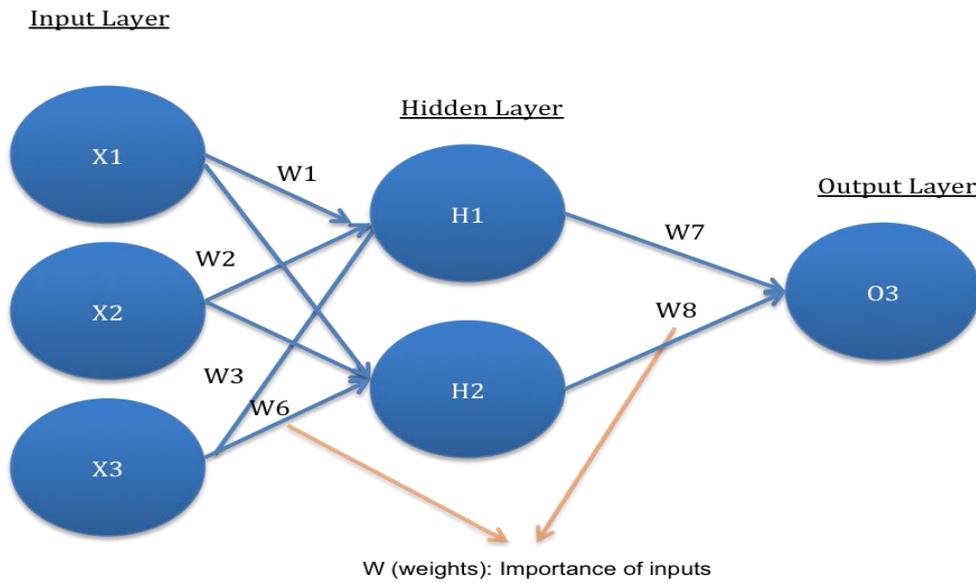


Figure 4.1: Structure of ANN used for training a model with hidden layer and weights and the output layer showing a feed-forward pass. $x_i w_i$

This study used a feed-forward backpropagation neural network. The weights are initially randomly assigned. The train: test split on the dataset is 7:3. A forward pass is performed for every training data using the current weights, and the output is calculated for each node. At the last node, the final output is acquired, and the error is calculated with a loss function. Now, a backward pass is performed to calculate the contribution of each node in error calculated. The error is propagated to every single node using backpropagation. Once the contribution of each node has been calculated, the weights are adjusted accordingly using gradient descent. The present study used gradient descent with momentum and adaptive linear regression. The procedure is repeated until the loss function gives an error that is less than the threshold value, and the weights and bias of the required network are thus obtained. Thus, the model converges, and a definite result can be obtained for any type of testing dataset.

The data from 1951-2015 were randomly divided into two subsets of training and testing using the earlier approach. The employed approach for splitting the data ensured that the sub-datasets were fairly representing the population to be modelled. The training subset was used for optimizing the connection weight matrices and bias vectors of the network. Once the network was trained, the generalization and predictive ability of the network was evaluated using a completely unseen subset called the testing subset.

No prior knowledge of the physics of the PET mechanism or the relationships between variables was assumed in this approach. As ANN model input sets, all possible combinations of input variables, a total of 26 combinations, were considered. Using the model development approach described earlier,

separate optimal ANN models were developed and trained for each input combination set. The prediction accuracy of the developed ANN models was compared to identify the most appropriate and efficient combinations of inputs for PET estimation. This method is known as a trial-and-error procedure, and it falls under the category of heuristic approaches.

4.3 Gradient Boosting Regressor (GBR)

Gradient Boosting for regression, GB builds an additive model in a forward stage-wise fashion. It allows for the optimization of arbitrary differentiable loss functions. In each stage, a regression tree is fit on the negative gradient of the given loss function. Gradient boosting involves three elements: A loss function to be optimized, A weak learner to make predictions and an additive model to add weak learners to minimize the loss function.

Gradient Boosting is a greedy algorithm and can overfit a training dataset quickly. It can benefit from regularization a method that penalize various algorithm parts and improves the algorithm's performance by reducing overfitting. There are four improvements primary gradient boosting:

1. Tree Constraints
2. Shrinkage
3. Random Sampling
4. Penalized learning

Gradient Boosting trains many models in a gradual, additive and sequential manner. The significant difference between AdaBoost and Gradient Boosting Algorithms is how the two algorithms identify the shortcomings of weak learners (e.g., decision trees). While the AdaBoost model identifies the shortcomings by using high weight data points, gradient boosting performs the same by using gradients in the loss function ($y=ax+b+e$, e needs a special mention as it is the error term).

The loss function is a measure indicating how good a model's coefficients are at fitting the underlying data. A logical understanding of loss function would depend on what we are trying to optimize.

For example, if we try to predict the sales prices using a regression, then the loss function would be based on the error between actual and predicted house prices. Similarly, if our goal is to classify credit defaults, then the loss function would be a measure of how good our predictive model is at classifying bad loans.

One of the biggest motivations of using gradient boosting is that it allows one to optimize a user-specified cost function instead of a loss function that usually offers less control and does not essentially correspond with real-world applications.

4.4 Random Forest (RF)

A Random Forest is an ensemble technique capable of performing regression and classification tasks using multiple decision trees and a technique called Bootstrap Aggregation, commonly known as bagging. The Sum of Squared Error (SSE) has been calculated between the observed values and the predicted values. This procedure will recursively continue until the entire data is being covered. The model can be written as:

$$f(x) = f_0(x) + f_1(x) + f_2(x) + \dots \quad (4.4)$$

Where the ultimate model f is the sum of simple base models f_i . Where each base regressor portion is the simple decision tree.

The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.

Approach:

1. Pick at random K data points from the training set.
2. Build the decision tree associated with those K data points.
3. Choose the number N_{tree} of trees you want to build and repeat steps 1 & 2.
4. For a new data point, make each one of your N_{tree} trees predict the value of Y for the data point, and assign the new data point the average across all of the predicted Y values.

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set.

Each Decision Tree in the Extra Trees Forest is constructed from the original training sample. Then, at each test node, each tree is provided with a random sample of k features from the feature-set from which each decision tree must select the best feature to split the data based on some mathematical criteria (typically the Gini Index). This random sample of features leads to the creation of multiple de-correlated decision trees.

4.5 SUPPORT VECTOR REGRESSION (SVR)

Support Vector Machines (SVMs) are well known in classification problems. When SVM is used for regression, these types of models are known as Support Vector Regression (SVR). SVR gives us the flexibility to define how much error is acceptable in our model and will find an appropriate line to fit

the data. SVR can learn and model nonlinear relationships between the input and output data in a higher dimension, thereby minimizing the observed training error and distribution error sufficiently. SVR implements the principle of inductive minimization of the structural error to attain a general optimal solution.

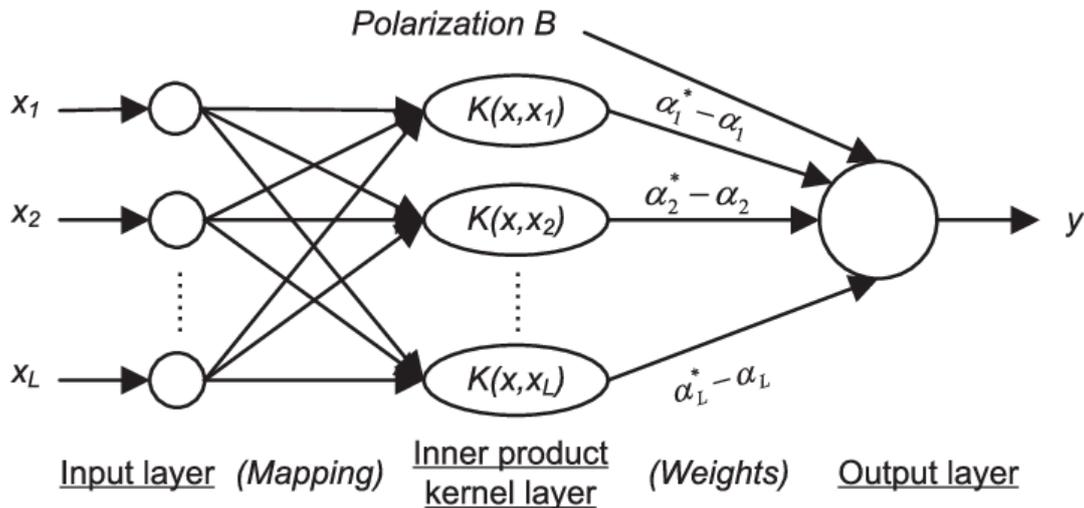


Figure 4.2: Structure of SVR

In simple regression, we try to minimize the error rate. While in SVR, we try to fit the error within a certain threshold.

Our objective when we are moving on with SVR is to consider the points that are within the boundary line, basically. Our best fit line is the line hyperplane that has a maximum number of points.

In the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM, which would have already been requested from the problem. But besides this fact, there is also a more complicated reason; the algorithm is more complicated, therefore to be taken into consideration. However, the main idea is always the same: to minimize error, to individualize the hyperplane, which maximizes the margin, keeping in mind that part of the error is tolerated. The goal of linear regression is to minimize the error between the prediction and data. In SVR, the goal is to make sure that the errors do not exceed the threshold.

SVRs are today known as efficient and robust ML algorithms for predictions. When the training data of $\{(x_1, y_1), \dots, (x_n, y_n)\}$ with n patterns, a function $f(x)$ will be identified with the consideration of the deviation from the actually observed target variables y_i for all the training data (Lima et al., 2012). Using a nonlinear mapping function ϕ , X will be mapping the input variables to a higher dimensional feature space.

$$f(x; w) = \langle W, \phi(x) \rangle + b \quad (4.5)$$

where W and b are the regression coefficients and \langle, \rangle denotes the inner product. SVR uses the ϵ -insensitive error to measure the error between $f(x)$ and the observed values of y .

$$|f(x; w) - y|_{\epsilon} = \begin{cases} 0, & \text{if } |f(x; w) - y| < \epsilon \\ |f(x; w) - y| - \epsilon, & \text{otherwise,} \end{cases} \quad (4.6)$$

Using the training data of (x_i, y_i) the values of w and b are calculated by minimizing the objective function:

$$F = \frac{C}{N} \sum_{i=1}^n |f(x_i, w) - y_i|_{\epsilon} + \frac{1}{2} \|w\|^2 \quad (4.7)$$

Where ϵ and C are the hyper-parameters. The minimization of the objective function, F , uses the Lagrange multiplier method. The ultimate regression equation with kernel function $K(X, X')$ can be in the form:

$$f(X) = \sum_i K(X, X_i) + b \quad (4.8)$$

Based on earlier studies (Y. Dibike et al., 2001), the kernel function RBF was chosen to measure the performance of the model for the PET. A complete overview of the SVR method may be found in (Y. Dibike et al., 2001).

4.6 Long short-term memory (LSTM)

Long short-term memory neural networks are similar to Recurrent neural networks (RNN), which have the capability to learn more significant data compared to normal RNNs. This is done by controlling the hidden state in LSTM and solving the vanishing gradient problem. LSTM has feedback connections. An LSTM unit has an input gate, an output gate, and a forget gate (Figure 5). LSTM calculates a gate's values using the previous cell value C_{t-1} , previously hidden values h_{t-1} , and input x_t .

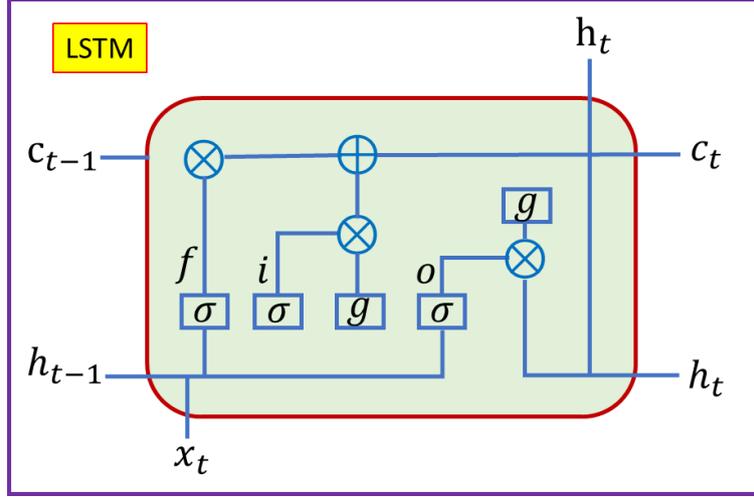


Figure 4.3: Overview diagram of Long short-term memory (LSTM).

Where f , i , and o denotes the forget gate, input gate, an output gate, h_t denotes hidden state, c_t denotes cell state, σ is the sigmoid function, g is the activation function.

$$i_t = F(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + bias_i) \quad (4.9)$$

$$o_t = F(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + bias_o) \quad (4.10)$$

$$f_t = F(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + bias_f) \quad (4.11)$$

And the cell value is calculated using

$$C_t = f_t c_{t-1} + i_t F(W_{xc}x_t + W_{hc}h_{t-1} + bias_c) \quad (4.12)$$

$$h_t = o_t \tanh(C_t) \quad (4.13)$$

LSTM is like RNN, but by using the three gates, it can process longer lengths of data, and it is also able to solve the vanishing gradient problem.

4.7 Model evaluation

The accuracy of the ML models was calculated using the coefficient of determination (R^2) (4.14), the root mean squared error (RMSE) (4.15), and the mean absolute error (MAE) (4.16). The equations are as follows:

$$R^2 = 1 - \frac{\sum(PET_{sim} - PET_{obs})^2}{\sum(ET_{obs} - ET_{mean})^2} \quad (4.14)$$

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{\sum_{i=1}^n (\text{PET}_{\text{sim}} - \text{PET}_{\text{obs}})^2}{n}} \quad (4.15)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^n (\text{PET}_{\text{sim}} - \text{PET}_{\text{obs}}) \quad (4.16)$$

where PET_{sim} is the simulated PET at time step i in mm/day; PET_{obs} is the observed PET at time step i in; PET_{mean} is the average PET at time step i in mm/day; n is the number of data pairs, respectively.

4.8 Conclusions

As a new approach for the simulation and prediction of PET, the use of ML models has been increased gradually. Five different ML regression models were implemented for modelling the PET of Hyderabad station and Waipara station. The study compared various PET simulations based on Long Short Term memory (LSTM), Artificial Neural Networks (ANN) regressor, Gradient boosting regression (GBR), Support Vector regressor (SVR) and Random Forest (RF) regressor as discussed in the present chapter for two stations. As the performance of any modelling technique is dependent upon the selection of predictors, the machine learning models are applied by using data pre-processing techniques, which are: Random Forests, Principal Component Analysis, Pearson Correlation to find optimal parameters for the model and also to give the ranking on features. The methodology, use and limitations of the pre-processing methods were discussed. The model's performance criteria was validated using different standard statistical methods, such as coefficient of Determination (R^2), Root Mean Square Error (RMSE), and the Mean Absolute Error (MAE) as validation criteria for selecting the best model as discussed in the present Chapter. Various empirical models used in the estimation of PET and AET were discussed in the Chapter 5&6.

Chapter -5

Estimation of potential evapotranspiration using machine learning models with limited data.

5.1 Introduction

PET is the enormous water flux representing the climatic demand of the water and the second most significant component of the terrestrial hydrological cycle next to rainfall. It is the process of returning water to the atmosphere via evaporation from open water, soil, and plant surfaces, as well as transpiration from plants. Water is absorbed from the soil through the roots and transported to the leaves via the vascular system of the roots, stem, and branches. The water is then transferred from the leaf's vascular system to the stomatal walls. PET is affected by climatic conditions, specifically the sun's radiative energy, wind, the air's vapour deficit, and temperature. PET is used with precipitation data to calculate water balances. Changes in values enable the impact of land-use change to be seen. PET calculations help to calculate plant water requirements and also the losses of water to evaporation from open areas such as irrigated fields and reservoirs. As it relates to climate change and food security, organizations such as the World Meteorological Organization and the Food and Agriculture Organization have already used PET in various ways.

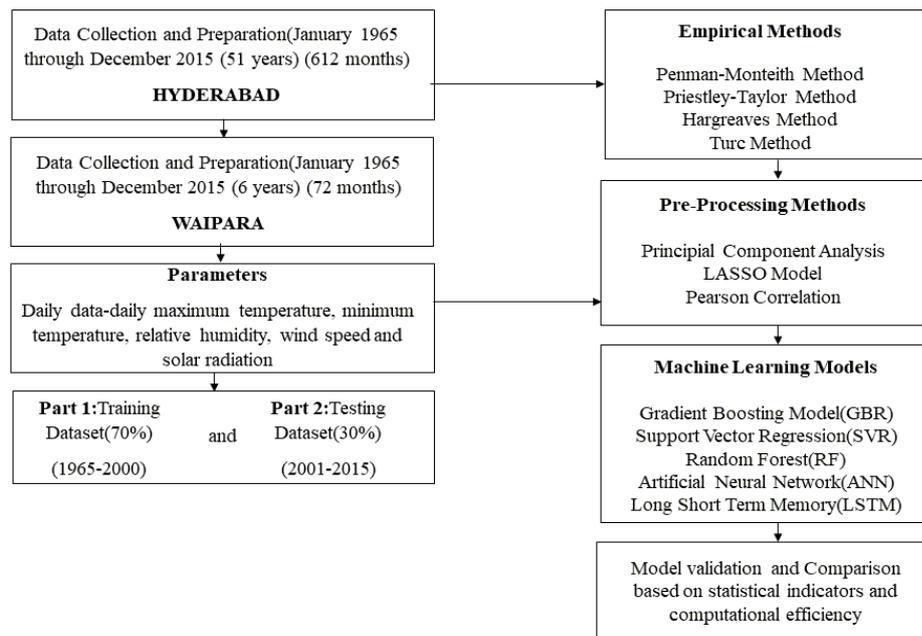


Figure 5.1: Flowchart for the estimation of PET using ML models

5.2 Estimation of Potential Evapotranspiration using empirical methods

PET is a measure of the ability of the atmosphere to remove water from the surface through the processes of evaporation and transpiration, assuming no control over the water supply. PET can be estimated based on energy balance and water vapour mass flux transfer methodologies. These empirical models vary in terms of solar radiation, temperature considering the physical processes of radiation and transport characteristics of natural surfaces.

5.2.1 FAO Penman-Monteith method

The Food and Agriculture Organization of the United Nations (FAO) suggested that the Penman-Monteith be utilized as the standard method to assess PET. This equation has picked up the parcel of acknowledgement and utilized worldwide for benchmark evapotranspiration assessments [10]. Nonetheless, when computing PET by Penman-Monteith model, loads of climatic factors, including temperature, wind speed, solar radiation, and relative humidity, are required.

$$PET = \frac{0.408D(Rn-G) + g\left(\frac{900}{T+273}\right)U_2(es-ea)}{D + g(1+0.34U_2)} \quad (5.1)$$

Where, R_n = net radiation ($MJ\ m^{-2}\ d^{-1}$), G = soil heat flux ($MJ\ m^{-2}\ d^{-1}$), T = average temperature at 2 m height ($^{\circ}C$), U_2 = wind speed measured at 2 m height [$m\ s^{-1}$], $(es - ea)$ = pressure deficit for measurement at 2 m height [$k\ Pa$], D = slope vapor pressure curve [$k\ pa^{\circ}C^{-1}$], g = psychrometric constant [$k\ pa^{\circ}C^{-1}$], 900 = coefficient for the reference crop [$l\ J^{-1}\ Kg\ K\ d^{-1}$], 0.34 = wind coefficient for the reference crop [$s\ m^{-1}$](<http://www.fao.org/3/x0490e/x0490e06.htm#TopOfPage>).

5.2.2 Turc Method

Turc developed an equation to simplify an older equation [40] for calculating daily PET as a function of air temperature, relative humidity and solar radiation. The Turc method estimates PET based on mean temperature and solar radiation on the daily time scale. The equation is given by:

$$PET = 0.013 \frac{T_m}{T_m + 15} (23.88R_s + 50) \quad (5.2)$$

where T_m is mean temperature ($^{\circ}C$), solar radiation (R_s) is $[0.25 + 0.5 (n/N)] R_a$, R_a is extra-terrestrial radiation (mm/d), n is actual hours of bright sunshine (hrs), N is maximum possible hours of sunshine (hrs).

5.2.3 Priestly and Taylor method

The Priestley-Taylor model is essentially a condensed version of the original Penman combination equation, which was developed in [21]. The model's original purpose was to be used in large-scale numerical modelling where it is assumed that because advection is negligible, the aerodynamic component of the original Penman equation can be ignored. This method is calculated using net radiation and latent heat of vaporization on a daily time scale. The equation is given by:

$$PET = A \left(\frac{D}{D+g} \right) \left(\frac{R_n - G}{L} \right) \quad (5.3)$$

$$D = \frac{4098 \left[0.6108 \exp \left(\frac{17.27 + T_m}{T_m + 237.3} \right) \right]}{(T_m + 237.3)^2} \quad (5.4)$$

Where D is slope vapour pressure curve [k pa°C⁻¹], g is psychrometric constant [k pa°C⁻¹], R_n is the net radiation at crop surface (MJ m⁻² d⁻¹), A is a calibration constant 1.26, L is the latent heat of vaporization and can be considered as 2.45 (MJ/kg) which is constant.

5.2.4 Hargreaves Method

The Hargreaves-Samani 1985 model is one of the more well-known versions of an older evapotranspiration model [42]. The Hargreaves method estimates PET based on maximum and minimum air temperature on a daily time scale.

$$PET = 0.0023 R_a \left(\frac{T_m}{2} + 17.8 \right) (T_d)^{0.5} \quad (5.5)$$

Where, T_d = difference between maximum temperature and min temperature (°C), T_m = mean temperature (°C), R_a = extra-terrestrial radiation (mm/d)

5.3 Modelling of Potential Evapotranspiration using machine learning models

Data-driven modelling techniques have been developed as a new approach for simulation and prediction of various natural and artificial phenomena due to recent developments in computer technologies and new mathematical algorithms. These new techniques are especially relevant in hydrological modelling, which is widely used to simulate complex and poorly understood natural processes. ANNs, SVR, GBR, RF, LSTM are few data-driven techniques that do not require a thorough understanding of the physics

of the processes under investigation. These machine learning techniques used in this study for modelling the PET process are listed. Using these techniques for modelling and predicting various complex processes has been studied in the literature and is promising. Furthermore, in the literature review, a brief history of each technique's development is provided to highlight the long road acquisition (measuring) methods; these modelling techniques are becoming even more popular.

5.4 RESULTS AND DISCUSSIONS

The current study analysed the performance of LSTM, ANN, SVR, GBR, and RF models in estimating daily PET. The PET was calculated using different empirical methods for Hyderabad and Waipara stations. To study data emerging from nonlinear phenomena, the determination of ideal model inputs is a critical problem. The four pre-processing data methods, i.e. Pearson correlation, Principal component analysis (PCA), Lasso model, and Random forest pre-processing methods, were evaluated to select the best set of input variables. In the present study, the main components of the first component, which bore at least eighty percent of the total variance in the data, served as the input to the estimation models. The climate variables considered for estimating daily PET using five different models and Penman-Monteith methods were the daily maximum temperature, minimum temperature, relative humidity, and solar radiation. The four various optimal input combinations for modelling this daily PET were all available meteorological parameters; temperature, wind speed, and relative humidity; temperature and wind speed; and temperature and relative humidity. Similarly, for the Turc and corresponding ML models, the input variable considered is the mean temperature. Whereas, for the Hargreaves and corresponding ML models, the input variables considered are maximum and minimum temperatures. The two optimal input combinations for modelling daily PET for the Hargreaves and Turc methods were all available meteorological parameters; minimum temperature and solar radiation. Furthermore, for the Priestley Taylor method, the input variables used in ML models are temperature, solar radiation, and relative humidity. The three various optimal input combinations for modelling daily PET for the Priestley Taylor method were all available meteorological parameters; temperature, wind speed, and relative humidity; temperature, and solar radiation and temperature and relative humidity. The models were run using the above input combinations as parameters for 500 iterations. Convergence was obtained for the datasets of all four empirical methods. The prediction values have been calculated by fitting the test data on the trained model. As the number of meteorological variables for each empirical method is different, therefore, for each empirical model, the five different ML models were trained, and results were tested. Input vector has the features considered in each method (Penman, Hargreaves, Turc, and Priestley-Taylor) have been used for each method, and the output vector is the expected PET value calculated from each method. The performance criteria considered were MAE, RMSE, and R^2 values from a test sample. In this study, ML models were trained, and statistical parameters were calculated

using only the whole test data set after each training run. The training period considered is from 1965 to 2000, and the testing period is considered from 2001 to 2015 for Hyderabad station. The training and testing data was divided into 70% training data and 30% testing data for the two stations. The validity and efficiency of the model can be seen when the training dataset is fit on the trained model, and high accuracy and minimal values of RMSE were obtained. The performance of each empirical model corresponding with the different ML models in terms of R^2 , RMSE, and MAE was listed in Tables 4-7. Figures 7-10 shows the comparison between daily PET values from empirical models of Penman-Monteith, Priestley-Taylor, Hargreaves, Turc, and ML methodologies for the testing datasets.

Table 5.1: Performance of Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long Short-Term Memory (LSTM) and Artificial Neural Network (ANN) for Penman-Monteith Method.

Parameters	Model	Hyderabad						Waipara					
		Calibration			Validation			Calibration			Validation		
		R ²	RMSE (mm/d)	MAE (mm/d)									
All parameters	RF	0.99	0.06	0.03	0.99	0.12	0.08	0.99	0.07	0.04	0.96	0.24	0.16
	SVR	0.99	0.16	0.11	0.99	0.15	0.11	0.99	0.08	0.06	0.99	0.11	0.08
	GBR	0.99	0.16	0.12	0.98	0.18	0.13	0.98	0.12	0.09	0.96	0.22	0.16
	LSTM	0.99	0.04	0.07	0.99	0.11	0.07	0.99	0.06	0.03	0.99	0.07	0.05
	ANN	0.99	0.06	0.05	0.99	0.06	0.05	0.98	0.15	0.11	0.97	0.17	0.12
Temperature, Wind Speed, Relative Humidity	RF	0.98	0.18	0.10	0.91	0.43	0.26	0.95	0.26	0.18	0.54	0.79	0.57
	SVR	0.92	0.46	0.27	0.92	0.42	0.26	0.74	0.59	0.41	0.59	0.74	0.54
	GBR	0.93	0.45	0.28	0.90	0.45	0.29	0.78	0.54	0.41	0.61	0.73	0.56
	LSTM	0.93	0.47	0.29	0.93	0.44	0.28	0.73	0.42	0.31	0.64	0.49	0.37
	ANN	0.93	0.46	0.28	0.92	0.40	0.22	0.76	0.57	0.42	0.65	0.69	0.51
Temperature and Wind Speed	RF	0.88	0.60	0.39	0.50	1.06	0.68	0.77	0.56	0.38	0.31	0.97	0.77
	SVR	0.71	0.93	0.62	0.55	1.00	0.64	0.45	0.86	0.60	0.35	0.94	0.74
	GBR	0.73	0.90	0.61	0.57	0.98	0.63	0.57	0.76	0.57	0.41	0.90	0.73

	LSTM	0.72	0.91	0.60	0.60	0.85	0.60	0.48	0.59	0.46	0.37	0.66	0.54
	ANN	0.72	0.93	0.62	0.58	0.96	0.62	0.51	0.80	0.60	0.37	0.93	0.75
Temperature and Relative Humidity	RF	0.73	0.92	0.67	0.19	1.36	1.03	0.86	0.42	0.29	0.49	0.83	0.59
	SVR	0.53	1.21	0.90	0.39	1.18	0.91	0.71	0.63	0.45	0.62	0.73	0.54
	GBR	0.55	1.19	0.89	0.37	1.20	0.94	0.77	0.56	0.42	0.58	0.76	0.57
	LSTM	0.50	1.45	0.98	0.46	1.09	0.85	0.47	0.58	0.45	0.36	0.65	0.54
	ANN	0.53	1.20	0.90	0.359	1.22	0.94	0.731	0.607	0.46	0.621	0.72	0.55

Table 5.2: Performance of Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), and Long Short-Term Memory (LSTM), Artificial Neural Network, and Artificial Neural Network (ANN) for Priestley Taylor Method.

Parameters	Model	Hyderabad						Waipara					
		Calibration			Validation			Calibration			Validation		
		R ²	RMSE (mm/d)	MAE (mm/d)									
All parameters	RF	0.99	0.04	0.02	0.97	0.13	0.08	0.99	0.08	0.04	0.96	0.27	0.15
	SVR	0.96	0.15	0.07	0.98	0.11	0.08	0.98	0.18	0.12	0.96	0.26	0.14
	GBR	0.97	0.12	0.08	0.97	0.14	0.09	0.98	0.15	0.11	0.96	0.28	0.18
	LSTM	0.99	0.03	0.02	0.99	0.03	0.02	0.99	0.06	0.03	0.99	0.07	0.05

	ANN	0.98	0.09	0.06	0.98	0.10	0.07	0.99	0.13	0.08	0.98	0.19	0.10
Temperature and Rs	RF	0.60	0.51	0.40	0.32	0.68	0.53	0.62	0.42	0.24	0.39	0.86	0.51
	SVR	0.46	0.60	0.46	0.43	0.62	0.48	0.35	0.62	0.37	0.37	0.76	0.46
	GBR	0.50	0.58	0.46	0.44	0.61	0.49	0.47	0.76	0.32	0.36	0.90	0.46
	LSTM	0.54	0.44	0.33	0.56	0.44	0.34	0.68	0.39	0.34	0.50	0.52	0.34
	ANN	0.49	0.58	0.47	0.43	0.62	0.49	0.59	0.49	0.35	0.45	0.55	0.44
Temperature and Relative Humidity	RF	0.46	0.6	0.42	0.13	0.7	0.46	0.81	1.15	0.56	0.64	1.31	0.94
	SVR	0.28	0.69	0.50	0.23	0.72	0.54	0.78	1.07	0.76	0.72	1.15	0.83
	GBR	0.32	0.67	0.49	0.23	0.72	0.54	0.84	0.96	0.71	0.73	1.16	0.88
	LSTM	0.55	0.64	0.47	0.52	0.63	0.46	0.81	1.02	0.65	0.78	1.08	0.78
	ANN	0.30	0.69	0.50	0.23	0.72	0.54	0.79	1.05	0.77	0.72	1.11	0.85

Table 5.3: Performance of Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), and Long Short-Term Memory (LSTM), Artificial Neural Network, and Artificial Neural Network (ANN) for Hargreaves Method.

Parameters	Model	Hyderabad						Waipara					
		Calibration			Validation			Calibration			Validation		
		R ²	RMSE (mm/d)	MAE (mm/d)									
All parameters	RF	0.99	0.01	0.03	0.99	0.03	0.12	0.99	0.02	0.008	0.80	0.15	0.05

	SVR	0.98	0.10	0.08	0.98	0.11	0.09	0.54	0.20	0.20	0.20	0.11	0.20
	GBR	0.99	0.05	0.03	0.99	0.07	0.04	0.78	0.13	0.10	0.66	0.19	0.14
	LSTM	0.99	0.04	0.02	0.99	0.05	0.03	0.98	0.04	0.03	0.97	0.04	0.03
	ANN	0.99	0.04	0.03	0.99	0.04	0.031	0.97	0.05	0.03	0.96	0.06	0.04
Minimum Temperature , Rs	RF	0.76	0.45	0.33	0.39	0.76	0.57	0.42	0.22	0.15	0.37	0.40	0.30
	SVR	0.53	0.63	0.48	0.46	0.71	0.53	0.45	0.30	0.26	0.42	0.34	0.30
	GBR	0.56	0.62	0.47	0.50	0.69	0.51	0.49	0.27	0.21	0.35	0.37	0.29
	LSTM	0.33	0.62	0.46	0.29	0.65	0.48	0.89	0.08	0.19	0.86	0.34	0.21
	ANN	0.73	0.63	0.49	0.49	0.69	0.52	0.68	0.29	0.22	0.49	0.34	0.28

Table 5.4: Performance of Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), and Long Short-Term Memory (LSTM), Artificial Neural Network, and Artificial Neural Network (ANN) for Turc Method.

Parameters	Model	Hyderabad						Waipara					
		Calibration			Validation			Calibration			Validation		
		R ²	RMSE (mm/d)	MAE (mm/d)									
All parameters	RF	0.99	0.02	0.01	0.99	0.03	0.01	0.99	0.02	0.07	0.99	0.05	0.02
	SVR	0.97	0.07	0.06	0.98	0.06	0.05	0.99	0.13	0.11	0.99	0.13	0.11

	GBR	0.99	0.01	0.09	0.99	0.01	0.09	0.99	0.05	0.03	0.99	0.08	0.05
	LSTM	0.99	0.03	0.02	0.99	0.02	0.01	0.99	0.04	0.02	0.99	0.03	0.02
	ANN	0.99	0.01	0.01	0.99	0.01	0.11	0.99	0.08	0.06	0.99	0.10	0.06
Minimum Temperature, Rs	RF	0.99	0.02	0.01	0.99	0.03	0.02	0.99	0.03	0.01	0.99	0.06	0.04
	SVR	0.98	0.06	0.04	0.98	0.06	0.05	0.99	0.11	0.08	0.99	0.10	0.09
	GBR	0.99	0.03	0.03	0.99	0.03	0.03	0.99	0.05	0.04	0.99	0.09	0.05
	LSTM	0.60	0.41	0.33	0.99	0.02	0.03	0.97	0.04	0.02	0.98	0.05	0.02
	ANN	0.99	0.04	0.03	0.99	0.04	0.03	0.99	0.10	0.06	0.99	0.11	0.07

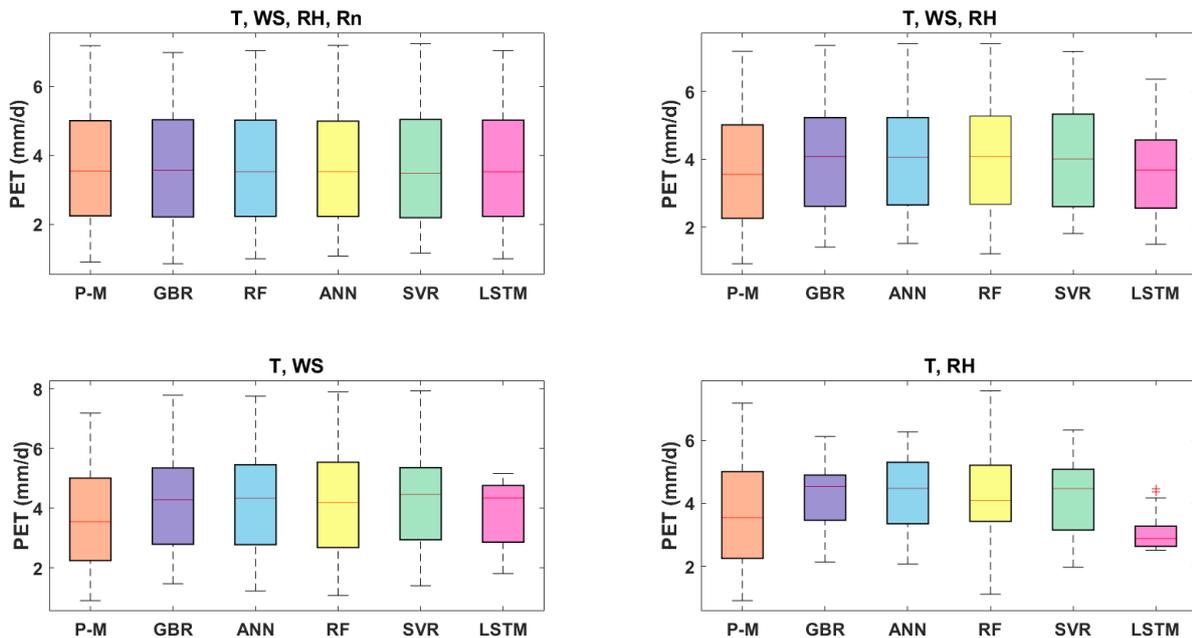
5.4.1 ML models performance with various input combinations

The performance of the LSTM, ANN, SVR, GBR, and RF models for the two stations for the Penman-Monteith method was provided in Table 5.1. Table 5.1 demonstrated that the tested models generally had comparable performance over the two stations. Figure 5.1 shows the comparisons between observed PET and model estimated values in the box plot form during the testing periods. The models LSTM, ANN and SVR estimated values showed closer agreement with those of observed PET (based on Penman-Monteith), and the LSTM and ANN models performed marginally better than the GBR, with estimated R^2 values of the two stations being 0.990 (LSTM), 0.998 (ANN), 0.990 (SVR) and 0.990 (GBR), 0.990 (RF). From Table 5, the LSTM and ANN models, for the most part, accomplished excellent performances trailed by SVR for all parameter input combinations. The performances being LSTM (RMSE: 0.02 mm/d, MAE: 0.01 mm/d, and R^2 : 0.990), ANN (RMSE: 0.06 mm/d, MAE: 0.05 mm/d, and R^2 : 0.998), SVR (RMSE: 0.155 mm/d, MAE: 0.11 mm/d, and R^2 : 0.990). The GBR model could likewise accomplish good results with (RMSE: 0.183 mm/d, MAE: 0.13 mm/d, and R^2 : 0.987), while the RF model also showed excellent performance with (RMSE: 0.123 mm/d, MAE: 0.08 mm/d, and R^2 : 0.990).

The performance of the LSTM, ANN, SVR, GBR and RF models at two stations for the Priestley Taylor method was provided in Table 5.2. Table 5.2 demonstrated that the tested models generally had comparable performance over the two stations. Figure 5.2 shows the comparisons between observed PET and model estimated values in the box plot form during the testing periods. The model's LSTM, ANN and SVR estimated values showed closer agreement with those of observed PET, and the LSTM and ANN models performed marginally better than the GBR, with estimated R^2 values of the two stations being 0.990 (LSTM), 0.985 (ANN), 0.98 (SVR) and 0.97 (GBR), 0.97 (RF). From Table 6, the LSTM and ANN models, for the most part, accomplished excellent performances trailed by the SVR model for all parameter input combinations. The performances being LSTM (RMSE: 0.03 mm/d, MAE: 0.024 mm/d, and R^2 : 0.990), ANN (RMSE: 0.10 mm/d, MAE: 0.07 mm/d, R^2 : 0.985), SVR (RMSE: 0.115 mm/d, MAE: 0.08 mm/d, and R^2 : 0.98). The GBR model could likewise accomplish good results with (RMSE: 0.14 mm/d, MAE: 0.09 mm/d and R^2 : 0.97), while the RF model also showed good performance with (RMSE: 0.13 mm/d, MAE: 0.08 mm/d, and R^2 : 0.97).

The performance of the LSTM, ANN, SVR, GBR, and RF models at the two stations for the Hargreaves and Turc methods were provided in Tables 5.1-5.2. Tables 5.1-5.2 demonstrated that the tested models generally had comparable performance over the two stations. Figures 5.1-5.2 showed the comparisons between observed PET and model estimated values in the box plot from all input combinations during the testing periods. The model's LSTM, ANN and SVR estimated values showed closer agreement with those of observed PET, and the LSTM and ANN models performed marginally better than the GBR, with estimated R^2 values of the two stations being 0.990 (LSTM), 0.99 (ANN), 0.99 (SVR) and 0.986

(GBR), 0.99 (RF) for Hargreaves and 0.990 (LSTM), 0.99 (ANN), 0.99 (SVR), 0.986 (GBR), 0.99 (RF) for Turc method. For the Hargreaves method, from Table 5.3, the LSTM and ANN models, for the most part, accomplished excellent performances trailed by SVR for all parameter input combinations. The performances being LSTM (RMSE: 0.05 mm/d, MAE: 0.03 mm/d, and R^2 : 0.990), ANN (RMSE:0.04 mm/d, MAE:0.03 mm/d, and R^2 :0.99), SVR (RMSE: 0.11 mm/d, MAE: 0.09 mm/d, and R^2 : 0.986). The GBR model could likewise accomplish good results with (RMSE: 0.012 mm/d, MAE: 0.09 mm/d, and R^2 : 0.99), while the RF model also showed acceptable performance with (RMSE: 0.03 mm/d, MAE: 0.017 mm/d, and R^2 : 0.99).For Turc method from Table 5.4, the LSTM and ANN models, for the most part, accomplished excellent performances trailed by SVR for all parameter input combinations. The performances being LSTM (RMSE: 0.02 mm/d, MAE: 0.019 mm/d, and R^2 : 0.990), ANN (RMSE:0.018 mm/d, MAE:0.011 mm/d, and R^2 :0.99), SVR (RMSE: 0.06 mm/d, MAE: 0.05 mm/d, and R^2 : 0.986).The GBR model could likewise accomplish sufficient results with (RMSE: 0.012 mm/d, MAE: 0.09 mm/d, and R^2 :0.99), while the RF model also showed acceptable performance with (RMSE: 0.03 mm/d, MAE:0.017 mm/d, and R^2 : 0.99).



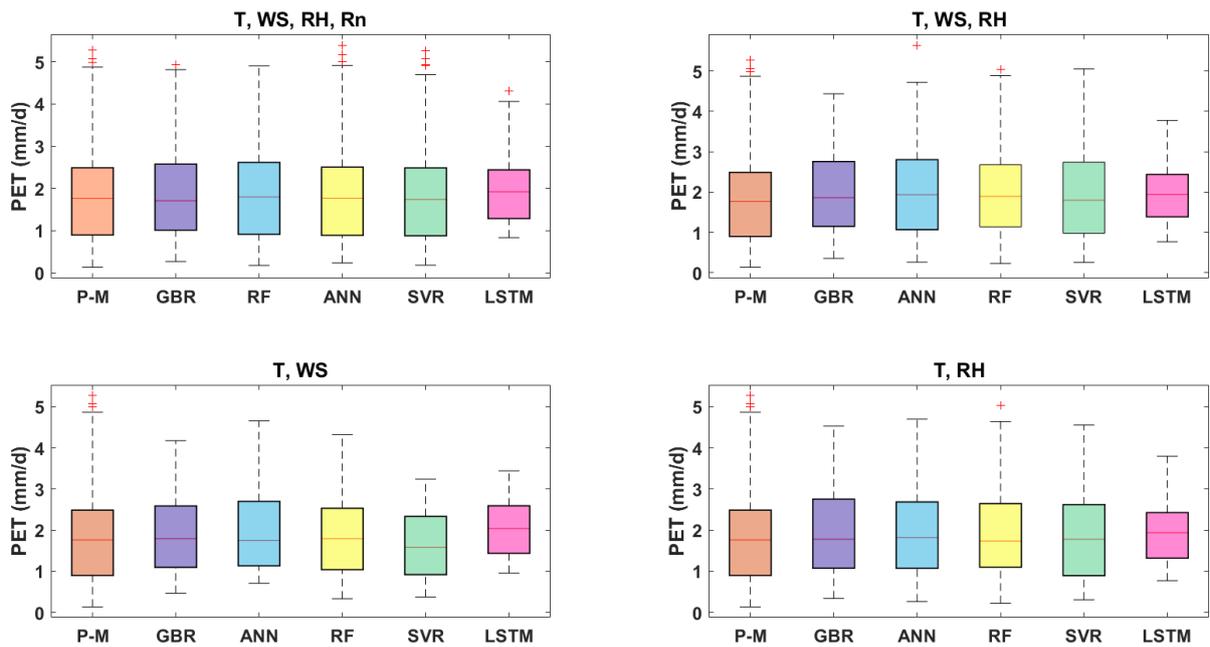
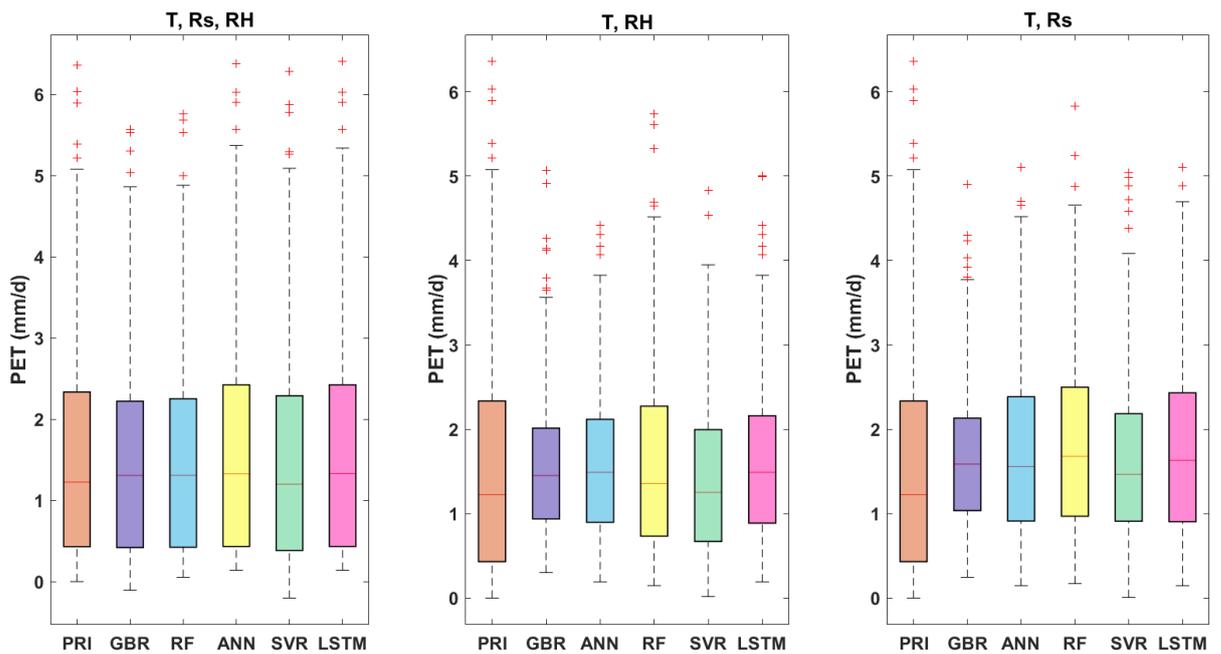


Figure 5.2: Comparison of observed and estimated PET by different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period at Hyderabad (Top) and Waipara Stations (Bottom) for Penman-Monteith method.



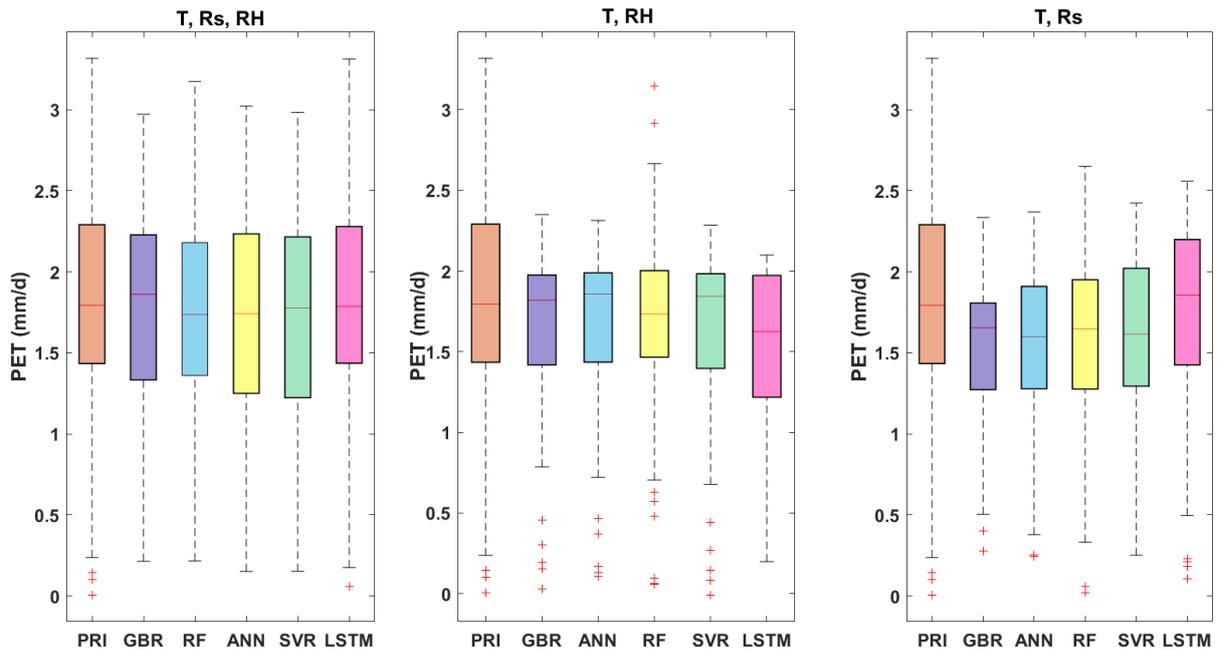
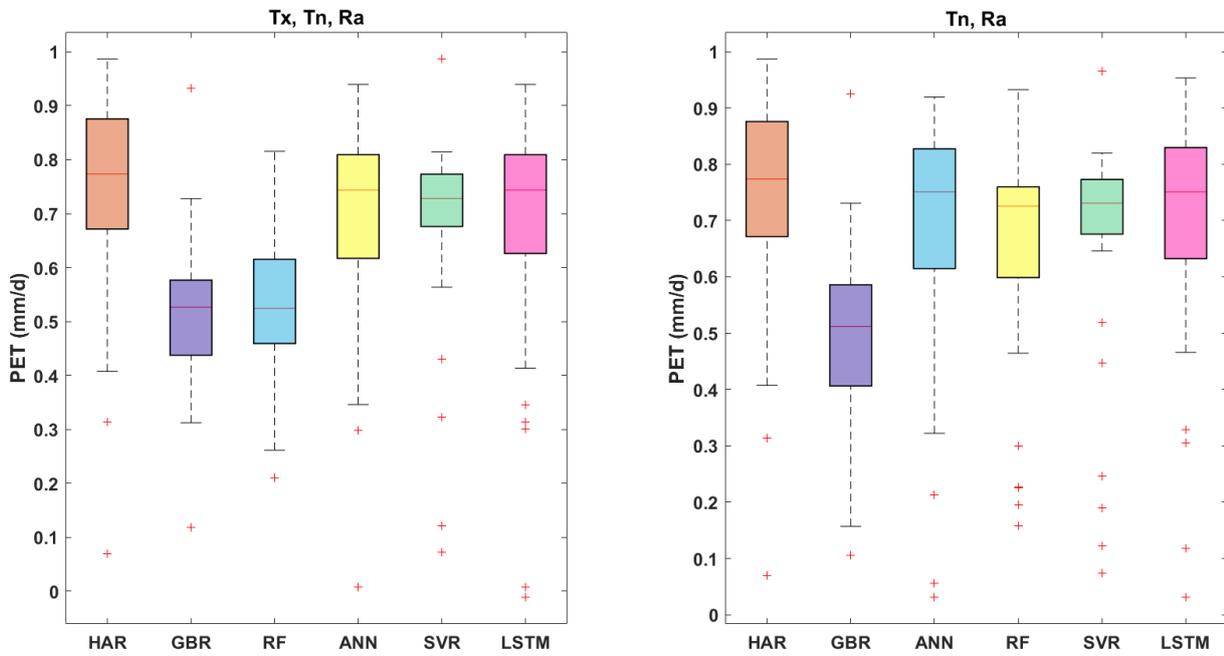


Figure 5.3: Comparison of observed and estimated PET by different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period at Hyderabad (Top) and Waipara Stations (Bottom) for Priestley Taylor method.



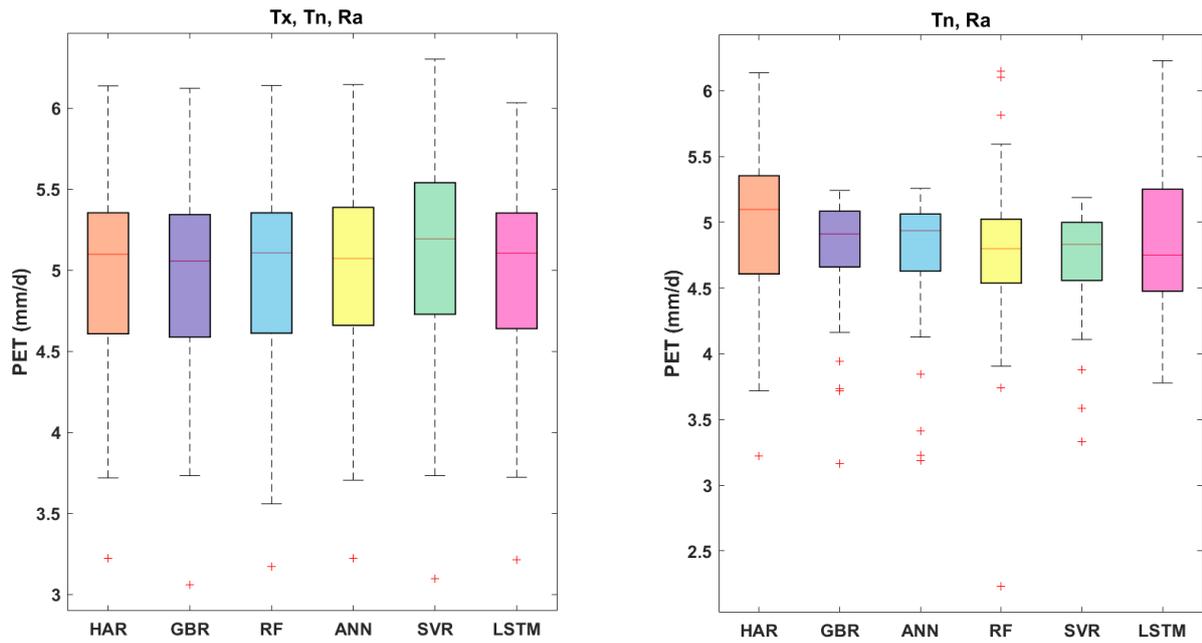
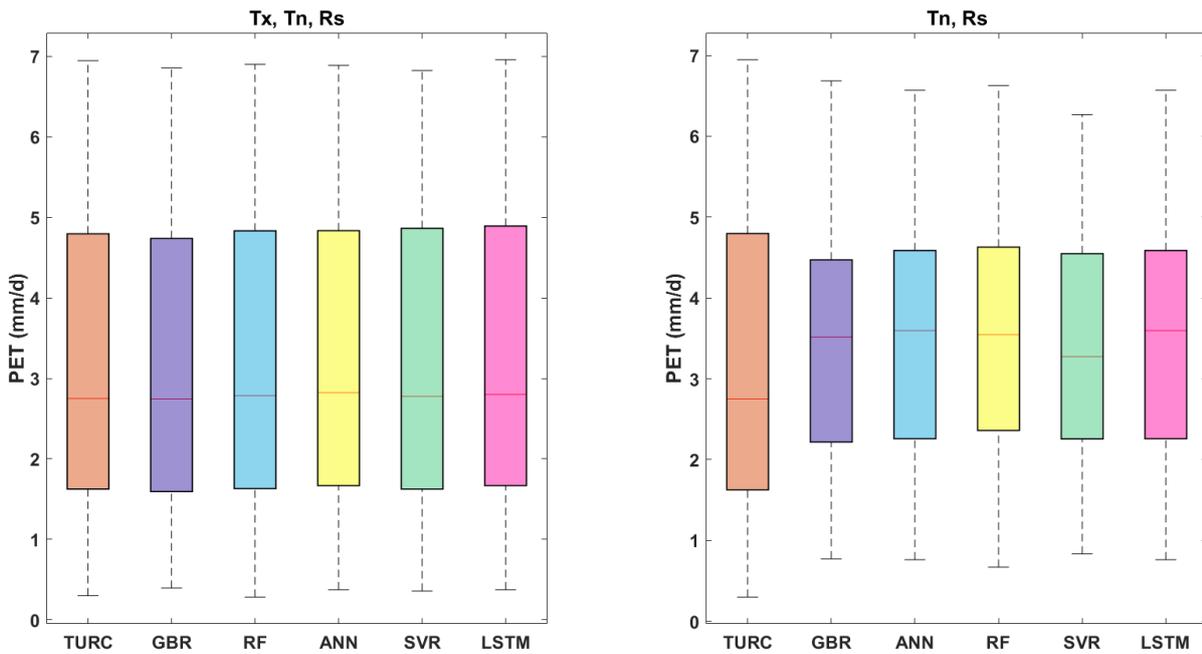


Figure 5.4: Comparison of observed and estimated PET by different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period at Hyderabad (Top) and Waipara Stations (Bottom) for Hargreaves method.



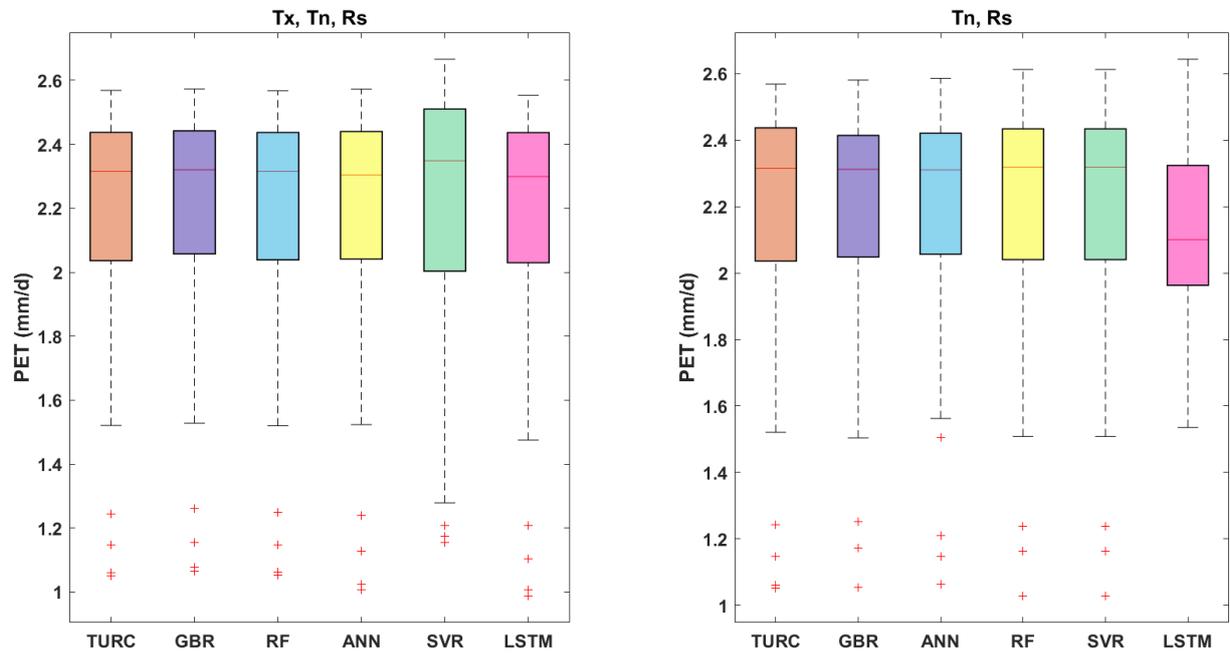


Figure 5.5: Comparison of observed and estimated PET by different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period at Hyderabad (Top) and Waipara Stations (Bottom) for Turc method.

During the validation stage at Hyderabad station, the input combinations like temperature, relative humidity, and wind speed ranging it results from (MAE: 0.07–0.13 mm/d, RMSE: 0.11–0.18 mm/d, R^2 : 0.98–0.99) played out the best for all the methods. However, all parameters input combinations outperformed in all the PET methods. Models accomplished the generally low performance when temperature and wind speed data was taken as input with (RMSE: 0.85–1.06 mm/d, MAE: 0.28–0.60 mm/d, and R^2 : 0.55–0.60), trailed by models with temperature and relative humidity taken as input (RMSE: 1.09–1.36 mm/d, MAE: 0.85–1.035 mm/d, R^2 : 0.19–0.46) for Penman-Monteith method. For the Priestley Taylor method, the ML models accomplished generally low performance when temperature and solar radiation was taken as input (RMSE: 0.44–0.68 mm/d, MAE: 0.34–0.53 mm/d, and R^2 : 0.32–0.56), followed by models dependent on temperature and relative humidity as input (RMSE: 0.63–0.725 mm/d, MAE: 0.438–0.560 mm/d, and R^2 : 0.23–0.50). For Hargreaves and Turc methods, the minimum temperature and solar radiation input combination performed equally as all

parameters input combination. It is worth seeing that models that are blends of temperature data with relative humidity and wind speed, individually, could accomplish preferred performance over models dependent on temperature and relative humidity input. The input combination of temperature and solar radiation could also accomplish outstanding performance compared to temperature and relative humidity for the Priestley Taylor method. It was also observed that the models dependent on temperature, relative humidity, wind speed and solar radiation could acquire slightly preferred performance over models dependent on three input parameter combinations. The outcomes demonstrated that LSTM and ANN performed superior to RF, SVR, and GBR when temperature and wind speed was taken as input under limited data. Furthermore, the LSTM model showed the most remarkable performance when temperature, wind speed, and relative humidity data were accessible for Penman-Monteith and when temperature and solar radiation were used as input for Priestley, Hargreaves and Turc methods. Among the five ML models, ANN came second, trailed by SVR. RF and GBR models were not as proficient as LSTM and ANN models. Hence, LSTM and ANN models came the best among any input combination, whereas other models performed low when the input parameters were reduced.

It was also observed that in all the PET methods, the results when Penman-Monteith was used as a reference method were remarkable, followed by the Priestley method. In Hargreaves and Turc methods, the inputs being the same, Turc has shown much better performance under limited combinations when compared to Hargreaves. This can be confirmed from the box plots in figures 5.4-5.6.

Tables 5.1-5.4 showed the summary of the LSTM, ANN, SVR, GBR, and RF model performances at Waipara station for different PET methods. The performance ranking of different ML models and PET methods was equivalent to the Hyderabad station. The ranking was given as LSTM>ANN>SVR>GBR>RF. Models obtained the best performance when all parameters were used as input combination with its results ranging from (RMSE: 0.11–0.242 mm/d, MAE: 0.07–0.163 mm/d, and R^2 : 0.98–0.990) compared to other input combinations. During the validation stage at Waipara station, the input combinations like temperature, relative humidity, and wind speed with (RMSE: 0.49–0.79 mm/d, MAE: 0.37–0.574 mm/d, R^2 : 0.548–0.64) played out the best for the Penman-Monteith method. The input combinations of temperature, relative humidity and solar radiation (RMSE: 1.11–1.31 mm/d, MAE: 0.28–0.6 mm/d, and R^2 : 0.64–0.72) played out the best for the Priestley method. In Hargreaves and Turc, the performances of all input combinations were almost the same, but in Hargreaves, ANN and LSTM performed much better than GBR and RF models. It is recommended to use LSTM or ANN to model PET while using the Hargreaves method.

Taking the observations from Tables 5.4-5.4 into account, the LSTM and ANN models are the most robust among the five ML models regardless of under which station or input combination, trailed by SVR and GBR models, which could generally accomplish agreeable accuracy. LSTM, ANN and

SVR models depend on just temperature, relative humidity, and solar radiation data to accomplish acceptable accuracy with the least meteorological variables. This can be considered more financially useful and helpful for advancement and application. The LSTM and ANN models built up in this study can accomplish higher performance than the other two models, and LSTM played out a little better than the ANN model under all inputs and much better under fewer input combinations. They both are successfully able to simulate PET where meteorological information is inadequate.

In this study, the stability comparison employed primarily focuses on LSTM, ANN, SVR, RF, GBR models. Among the models studied, the LSTM model achieved the lowest and the most concentrated distribution of RMSE values independent of the input combinations. It showed that the LSTM model had the best precision stability with an accuracy of 99.10 %, trailed by ANN with 92.70 % accuracy. The stability of the other two models is almost the same; therefore, when selecting one model for estimating PET between these two models, modelling accuracy should be the primary consideration. However, the accuracy of the models varies according to the number of inputs and the prediction of the time step. In terms of the input combinations effect, taking LSTM models as an example, the RMSE values of the LSTM model based on all parameters input combination gained the lowest fluctuation of (RMSE: 0.02–1.36 mm/d) across two stations. This was followed by models based on temperature, wind speed, relative humidity; temperature, relative humidity; temperature, wind speed for the Penman-Monteith method. The input combination of temperature, solar radiation, relative humidity, and temperature, solar radiation was observed to have the lowest RMSE values for the Priestley method. It was also worth noticing that although the accuracy of models with all parameters input combination was the highest in each station. Even if all parameter information is not available in a particular station, the discussed three parameters input combination or the two-parameter input combinations such as temperature and relative humidity or temperature and wind speed values could be used.

As shown in Figures 5.6-5.9, taking LSTM models as examples for Hyderabad station, the average RMSE values for input combinations of temperature, wind speed, relative humidity, and temperature and wind speed ranged from 0.11–1.36 mm/d and 0.07–1.065 mm/d, respectively. Using RH could make the average RMSE value decrease by less percentage, while using U2 could decrease it by more percentage. It is evident that (T, U2) or (T, Rs) performed even better than (T, RH). So, it can be reasoned that wind speed and solar radiation can improve the temperature-based accuracy for different models. These results generally can be relatively stable from the station's meteorological conditions in the study area and PET results from the coupling effect of other meteorological variables. Hence, temperature includes more comprehensive data and a PET pattern of the variance than the single determined RH.

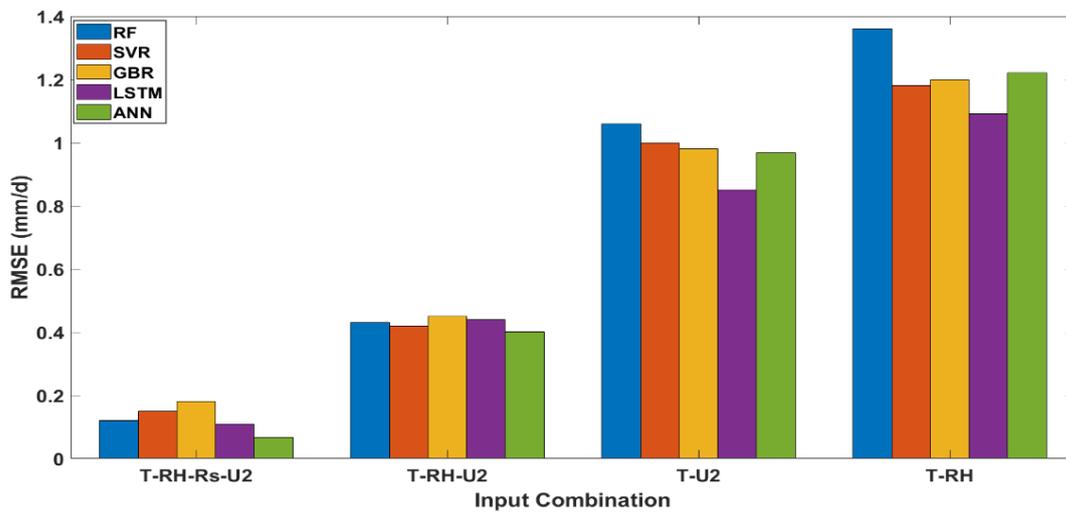
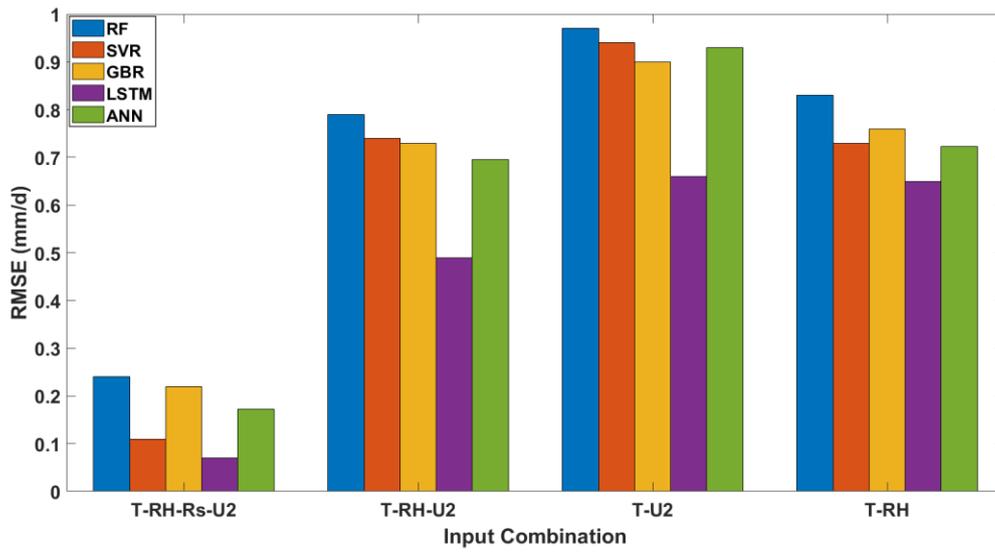


Figure 5.6: Comparison of RMSE values for Hyderabad (Top) and Waipara station (Bottom) for different input combinations for Penman-Monteith Method.

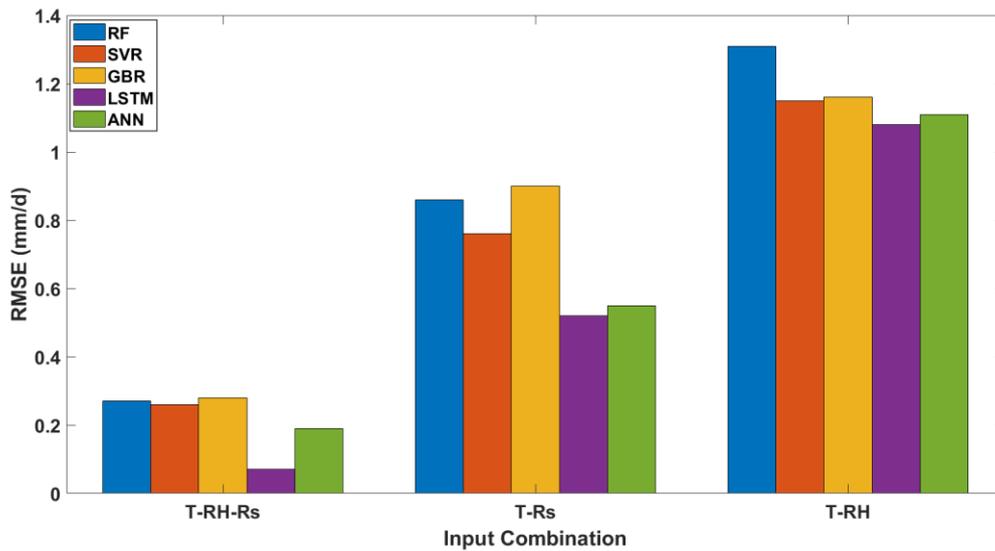
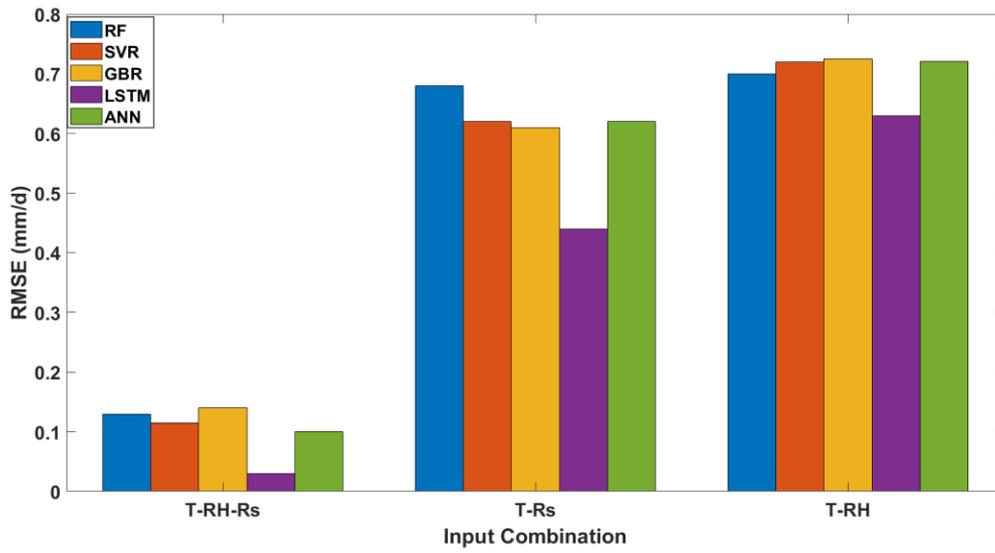


Figure 5.7: Comparison of RMSE values for Hyderabad (Top) and Waipara station (Bottom) for different input combinations for Priestley Taylor Method.

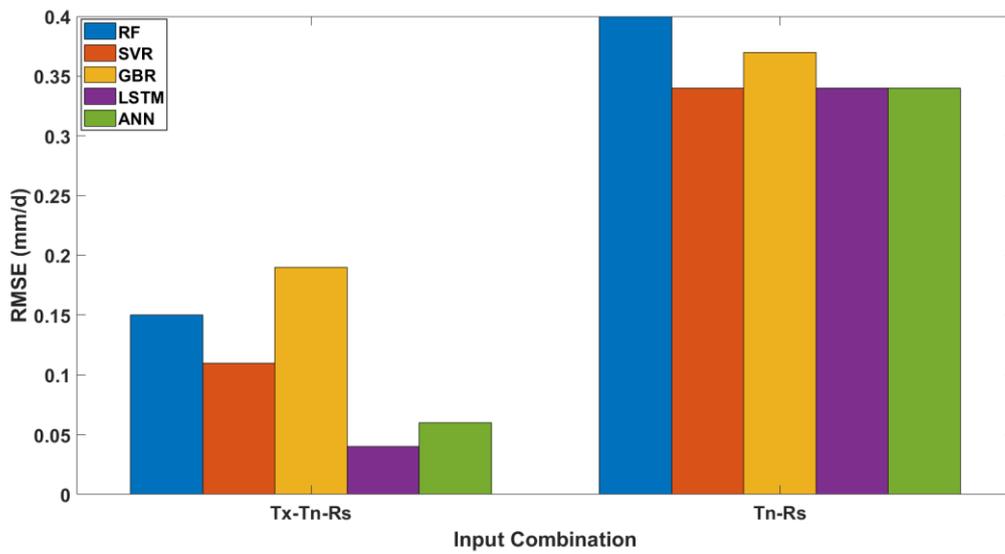
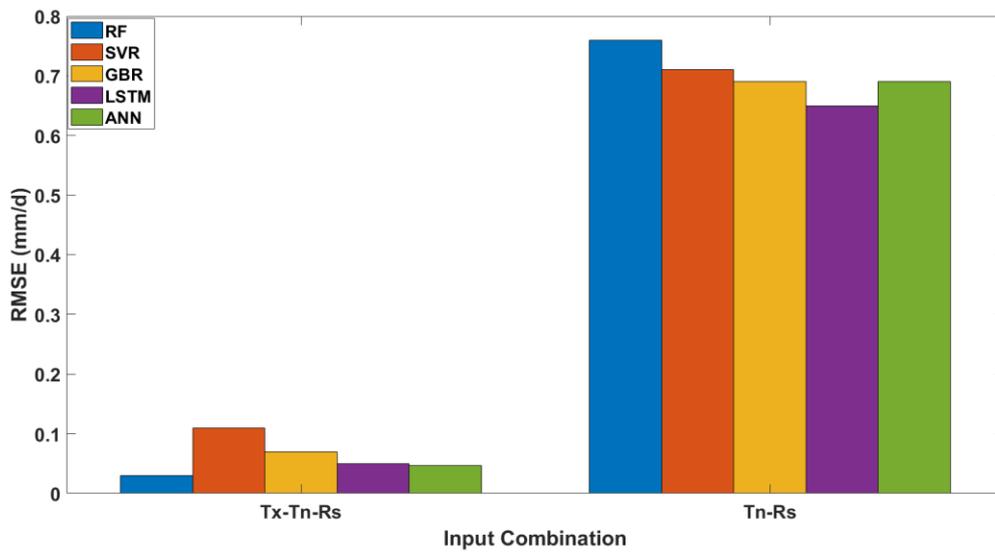


Figure 5.8: Comparison of RMSE values for Hyderabad (Top) and Waipara station (Bottom) for different input combinations for Hargreaves Method.

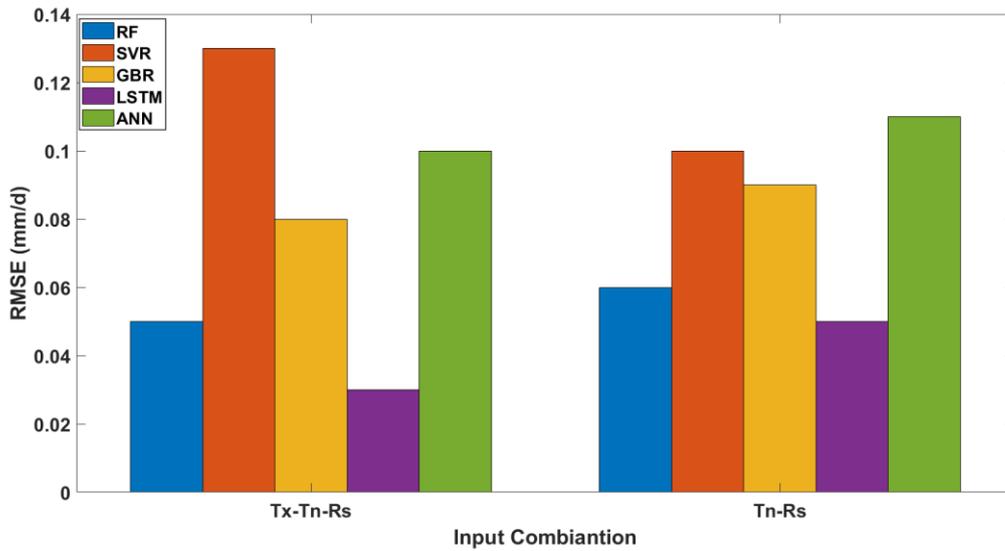
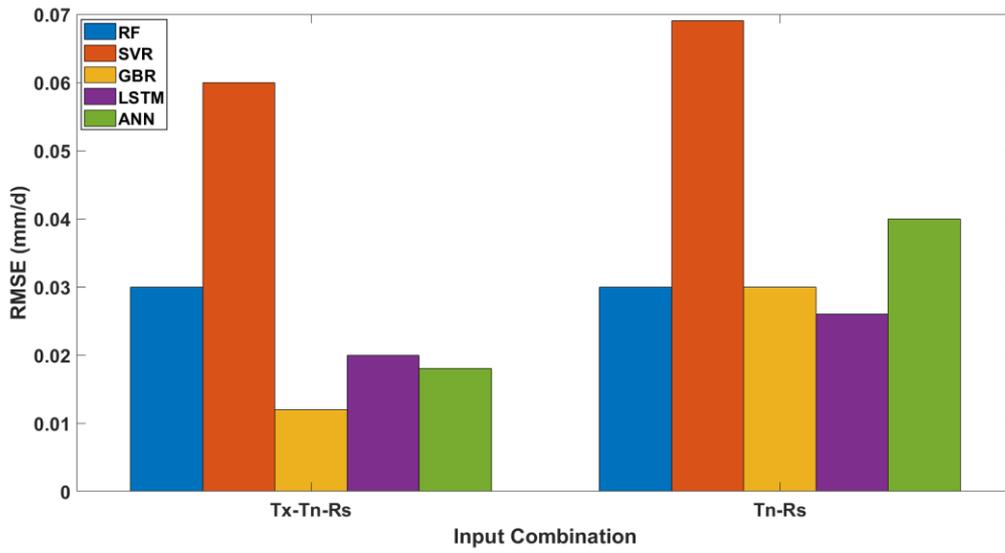


Figure 5.9: Comparison of RMSE values for Hyderabad (Top) and Waipara station (Bottom) for different input combinations for Turc Method.

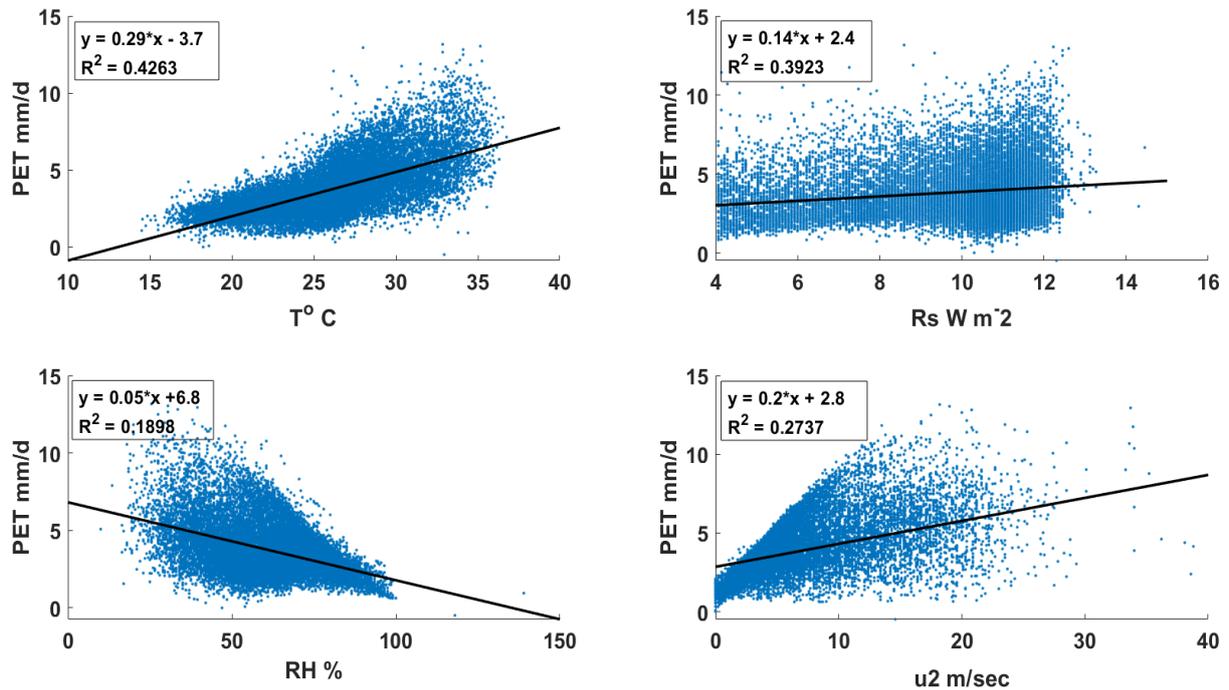


Figure 5.10: Correlation of main meteorological parameters such as temperature, relative humidity, solar radiation, wind speed to PET.

As shown in Figure 5.10, the temperature and wind speed followed by solar radiation have the most substantial influence on PET estimations based on the Penman-Monteith model. It was also observed from the study that the temperature, wind speed and solar radiation have greater than or equal to zero values. This means that these features are more practical while predicting PET value than relative humidity, which was a negative value. It was also observed that maximum temperature, minimum temperature, wind speed and relative humidity contribute more to the output variable PET. Therefore, the ML models derived based on temperature, solar radiation, wind speed as input and the PET from Penman-Monteith, Priestley Taylor, Hargreaves, and Turc methods as output variables gave a remarkable performance. The ML model results, when using only three (T, WS, Rs) from the four essential meteorological variables as input, seldom show the same values of coefficient of determination (R^2). The observed results prove that relative humidity has a meagre contribution to PET when using ML models. The overall accuracies of most models were found to be similar to each other.

5.4.2 Discussions

This study investigated the five ML models in estimating daily PET, including LSTM, ANN, GBR, RF, and SVR, for four different empirical methods under various input combinations for semi-arid climate conditions of Hyderabad and Waipara Stations. The study has reportedly found that LSTM and ANN performed the best at two stations. Three other models like SVR, GBR and RF, were also used to estimate PET. It was seen that they have also produced lower RMSE values and accurate R^2 values for these stations. Various studies in literature by (Carter & Liang, 2019; Chen et al., 2020; Raza et al., 2020; Saggi & Jain, 2019; T. Wu et al., 2020), which used ANN, LSTM, RF, SVR for PET estimation, reportedly found ANN, LSTM, and SVR performed best. The present study also acts as a proof for the performance of these models. The present study also observed that LSTM outperformed, followed by ANN, SVR and GBR. The use of different input combinations was observed in the study, and it was concluded that even when two parameters input combinations like temperature and wind speed are used, LSTM works well.

It was observed that while some of the ML models perform well in terms of both accuracy and computational demand, which can be seen clearly in the case of LSTM and ANN. The LSTM model is appealing due to its efficiency and low test RMSE, irrespective of input combination. Notably, the multi-layer neural networks tested in this study performed better than the other machine learning models. The multi neural networks and deep learning models showed more evidence of overfitting than other mentioned ML models, and it is worth noticing the performance metrics for LSTM and ANN ranging R^2 values to 0.99. Comparison of different machine learning and deep learning models and their performance when inputs were reduced showed the lowest performance in the case of RF and GBR models. The two input combinations, temperature and relative humidity, did not work well with other models, but they worked comparatively well with LSTM. So, LSTM can be used even if there are two input combinations followed by the ANN model. It has also shown that this study resembles the assessment of PET found by (Raza et al., 2020) where SVR, was found to be used as an alternative PET estimation model to the subsistence of conventional methods. Whereas the other two model's RF and GBR, performed best when more inputs were used, and their performance gradually reduced when inputs combinations were brought down. The strategy shown in the current investigation can be embraced in different domains and regions as done in this study, namely Hyderabad(India) and Waipara(New Zealand). It was also seen that among all empirical models, the Penman-Monteith has been predicted well with the ML models followed by the Priestley Taylor method. In the other two PET methods, whose inputs are the same, Turc performed better than the Hargreaves method. It can be said that when we have more available data, Penman or Priestley can be used to estimate PET, and when there is less data, Turc can be used to estimate PET. The results obtained in this study can be compared with other empirical methods and computational models in future studies. The analysis of this presented

study contributes essential guidance to use these models in PET estimations, where partial meteorological variables and topographical data are absent. This results in providing ease for agriculturists, water resources management and hydrological engineers. In the future, proposed models can be applied for irrigation scheduling, evaluating the crop coefficients for crop water modelling and in estimating AET.

To summarize, in the present study, the result of using RH with only temperature data for assessing PET was the same as previous studies. As a parameter for calculating RH, with limited inputs, U2 and Rs could increase modelling accuracy and be even better than RH. Hence, models based on U2 and Rs as input and LSTM model can be suggested for calculating PET in the light of data unavailability and models based on any combination, whether it be based on RH, ANN model can be suggested to calculate PET.

The results of the study reveal that temperature and solar radiation as the most influencing variables compared to relative humidity and wind speed for semi-arid climate conditions, as demonstrated in the present study. Given the intense data requirements for applying the Penman-Monteith model, Priestley model, Hargreaves and Turc models, the study employed ML models with minimum input variables such as temperature, solar radiation, and relative humidity. The trained and tested ML models developed based on empirical models can be valuable tools to predict PET for limited data case studies. Analysing the sensitivity of each climate variable on PET and testing their statistical dependencies, data pre-processing to acquire relevant information before developing such ML models is of most importance in the implementation. Analysis of compensating accuracies with limited climate input variables in the PET estimates compared to standard empirical models can be a potential area of research.

5.5 Conclusions

The daily PET over semi-arid climatic conditions over Hyderabad, Telangana, India and Waipara (New Zealand) were modelled using empirical and ML models. The Penman-Monteith model estimates of PET were considered as standard reference models for various temperature and radiation-based empirical methods as well as for data-driven models. The daily PET rates were estimated with five different modelling techniques, namely LSTM, ANN, SVR, GBR, RF using four input combinations such as maximum and minimum air temperatures, relative humidity and solar radiation, wind speed; three variable input combinations such as average air temperature, relative humidity, and solar radiation; two input variables combinations such as temperature and solar radiation; temperature and wind speed. The study investigated that the best performance was when all input variables were used. The study, however, finds that even three input variable combinations (temperature, wind speed and relative humidity values) or two combination input variables combination (temperature and relative

humidity, temperature and wind Speed) also can provide practically identical results. The results were discussed with the results of alternative methods of PET calculation, such as the combination-based method of Penman-Monteith, the radiation-based methods of Priestley-Taylor, the temperature-based methods of Hargreaves and Turc. The correlation coefficient values suggest that temperature is the most important factor, followed by solar radiation, wind speed, and relative humidity, respectively. LSTM and ANN with all-climate variables as input were able to simulate PET values estimated using the Penman-Monteith method. Temperature and solar radiation have a maximum correlation with PET estimates of Penman-Monteith models compared to relative humidity and wind speed. The Turc model uses temperature and solar radiation as input variables and has high accuracy with all ML models. In contrast, relative humidity has the least correlation with the PET estimates. The Priestly-Taylor model considers relative humidity, temperature, and solar radiation as input variables. Due to the lower dependency of relative humidity on the PET estimates, the Priestly-Taylor model has lower accuracy with ML models than the Turc model. The results also showed that the LSTM and ANN models could offer the most remarkable performance among five tested models regardless of station or input combination, trailed by SVR and GBR models, which could likewise accomplish moderately good performance. LSTM, ANN, and SVR models depend on just temperature, relative humidity, and wind speed data to achieve good performance with the fewest meteorological variables, which can be viewed as more practical and helpful for advancement and application. Then comes the combination of temperature and wind speed, followed by temperature and relative humidity. The study concludes that the empirical models work well with data-driven algorithms that consider the climate variables having high dependency with the standard PET estimates. Such. Further, it can be concluded that when a parameter or an input variable with a lower correlation is added to the set of features for training over ML models, the accuracy of prediction will be decreased. The results showed that LSTM and ANN models provide pretty good agreement with the PET obtained by the Penman-Monteith method. The study demonstrated that the modelling of PET through the LSTM and ANN techniques gave better estimates that proved their performance criterion as R^2 : 0.99. The study concludes that the performance of the models varies according to the number of inputs and the predicted time step. Overall, results are of significant practical use when limited climate data is available to estimate the PET. So, it can be concluded that even if not all parameter information is available in a particular station, this study proved that this three-parameter input combination or the two combinations, which are temperature and wind speed or temperature and relative humidity values, can be used to estimate PET. At spatiotemporal scales, LSTM and ANN models demonstrated extraordinary pertinence in displaying PET and can be strongly suggested for assessing PET when meteorological information is fragmented or restricted.

Chapter-6

Estimation of Actual Evapotranspiration with Limited Data

6.1 Introduction

A key aspect in the water budget is estimates of the spatial and temporal values PET. This term has been studied and modelled in Chapter 5 for different empirical methods and ML models. The uncertainties in PET estimates limits the reliability of hydrologic and restoration analyses, water resources management and planning. Along with PET estimates, hydrological implications necessitate AET estimates as well to get an actual water budget. Most of the AET empirical models depend on precipitation and PET as a limiting factor. Furthermore, due to the ambiguities in the definitions of PET and their dependency on large amounts of data, PET is not sufficient to understand the hydrology. AET expresses the annual water balance between precipitation and latent heat exchange, which is the variable most frequently correlated with biodiversity at the continental scales. Because of the close relationship between production and climatic factors, AET is also regarded as a surrogate for net primary production (Leith, 1975). The vast majority of studies in the literature have focused on modelling the PET process, in which evaporation occurs from soil and plant surfaces under no water stress. However, AET occurs under actual water supply conditions. Earlier studies have estimated AET using time-consuming and labour-intensive methods such as water-balance, energy-balance, Bowen-ratio (EBBR), and eddy-covariance (EC), and few modelled based methods. The modelled based methods include, Thornthwaite and Mather equation, Coutagne, Penman, Serra, Cappus, Kessler, Jensen and Haise, Blaney and Criddle, and Papadakis (Gudulas et al., 2013). These methods are mainly empirical and the common issue in these methods is the calculation of certain components of the water balance based on various factors such as temperature, precipitation, humidity, etc. So, the need to estimate AET using limited data has become important. The Budyko and Turc methods appear to be the most appropriate; however, their application is heavily influenced by the regional conditions prevalent in a given area. Furthermore, the more sensitive an area or station and the corresponding environmental pressures, the more important it is to apply the appropriate methods for estimating the important terms of hydrological cycle and developing an efficient water resources management plan. This study essentially adjusts the estimated PET values for the generation of AET. There are several studies which attempted to estimate AET using various empirical methods (as explained in Chapter 5), however, estimation of AET using the PET estimates in the context of limited data availability has not been addressed. The present study made an attempt to estimate AET for two different climatological conditions of Hyderabad and Waipara in the

context of limited data which can serve as input to hydrological models of ungauged river basins. In this context, estimating AET given the PET estimates under limited data availability, as demonstrated in Chapter 5 will be valuable in the water balance assessment studies for ungauged basins.

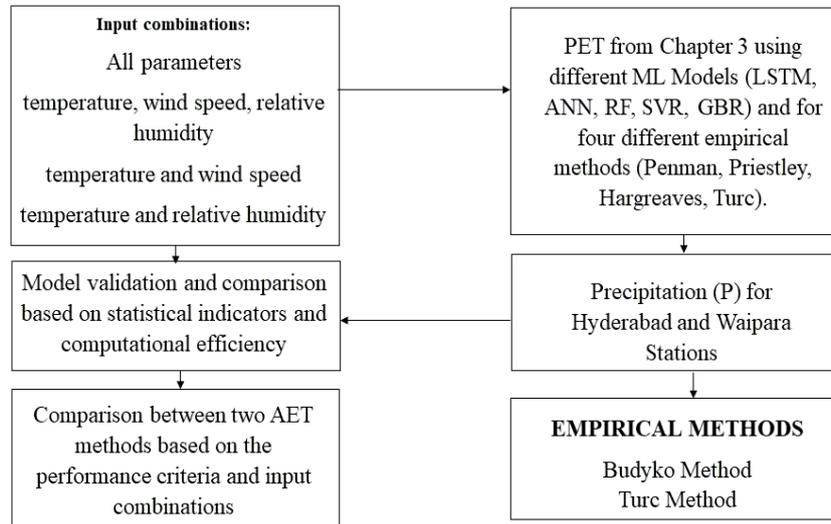


Figure 6.1: Flowchart for estimation of AET using modelled PET

6.2 Methods to Estimate Actual Evapotranspiration

AET is constrained by the availability of energy, water, and the resistance provided by the atmosphere and vegetation. Because of these constraints, AET measurement methodologies such as micrometeorological methods (energy-budget Bowen ratio or eddy correlation), and lysimeter-based techniques are costly and labour-intensive. To approximate AET over the two different stations, low-cost and simple alternatives are required. This study has provided alternatives based on PET equations to estimate AET under limited data conditions for two different climatic regions. The study has used PET from Chapter 5, the PET which was modelled using different machine learning models taking the Penman Monteith as reference method. The PET estimates from various ML models such as LSTM, ANN, RF, SVR, and GBR were utilized in place of PET in the Budyko and Turc methods. The Budyko and the Turc methods are the models used in this study to estimate AET. The meteorological variables used in estimating PET were given in Chapter 5, and the variables affecting AET are discussed in this chapter. This methodology gives us the clarity on estimating AET with readily available PET and precipitation under limited data conditions.

6.2.1 Budyko method

To estimate AET with readily available and modelled operational hydro-meteorological variables of P and PET, the study adopted the Budyko (1958) equation. Budyko Equation is a classic model for estimating AET by relating long-term-average water and energy balances at catchment scales using precipitation and PET. Budyko established a relationship between three hydro-climatic variables for a basin: precipitation (P), PET and AET. The Budyko formulation depends on the relationship between three hydro-meteorological variables: P, PET and AET, which states that the ratio of the AET over precipitation (AET/P) is fundamentally related to the ratio of the PET over precipitation (PET/P) (Wang et al., 2017), (Y. Zhang et al., 2015) as follows:

$$\frac{ET}{P} = 1 + \frac{PET}{P} - \left(1 + \left(\frac{PET}{P}\right)^\omega\right)^{(1/\omega)} \quad (6.1)$$

The parameter ‘ ω ’ accounts for the basin characteristics such as soil, vegetation, terrain, etc. ((McVicar et al., 2012). The original Budyko equation (Eq.23) was developed for a long-time scale such as annual with the parameter ‘’, representing the combined effect of climate and land surface. For a reasonable application of Budyko equation for short-term scales, the original Budyko formulation (Eq. 23) has been modified by several researchers (e.g., (Schreiber, 1904)) and one of the widely used formulation is as implemented by Zhang et al., (2012) for estimating the AET, as follows:

$$AET_{Budyko} = \left[P \left(1 - \exp\left(\frac{-PET}{P}\right) \right) PET \tanh\left(\frac{P}{PET}\right) \right]^{0.5} \quad (6.2)$$

6.2.2 Turc Method

Another well accepted and widely used AET model which considers precipitation and PET along with soil and vegetative characteristics implicitly is Turc model (Turc, 1954). It is also one of the widely used AET models in hydrological applications [58], [132], [133].

$$AET_{Turc} = \frac{P}{\sqrt{0.9 + \frac{P^2}{PET^2}}} \quad (6.3)$$

6.3 RESULTS

The modelled PET from Chapter 5 using five different ML models like LSTM, GBR, RF, ANN, SVR were considered in this analysis to estimate AET using Budyko and Turc methods. The PET modelling

was done on empirical methods such as Penman, Hargreaves, Priestley and Turc. These empirical approaches use precipitation, temperature, solar radiation, wind speed, relative humidity as parameters to determine PET. The two AET methods of Budyko and Turc use precipitation, and the PET estimates do determine AET. The input combinations were replicated from the estimation of PET to estimate different modelled combinations of AET as given in Tables 9-12. The best empirical AET methods were recorded corresponding to the four ML models and their input combinations in terms of R^2 , RMSE, and MAE for four empirical methods. These were listed in Tables 9, 10, 11, 12 for the Hyderabad and Waipara Stations.

Table 6.1: Performance of Budyko and Turc methods in estimating AET using PET modelled from Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long Short-Term Memory (LSTM) and Artificial Neural Network(ANN) for Penman-Monteith Method under different input combinations.

Parameters	Model	Hyderabad						Waipara					
		Budyko			Turc			Budyko			Turc		
		R ²	RMSE (mm/d)	MAE (mm/d)									
All parameters	RF	0.99	0.05	0.01	0.99	0.06	0.01	0.99	0.03	0.04	0.99	0.02	0.01
	SVR	0.99	0.04	0.02	0.99	0.04	0.02	0.98	0.02	0.01	0.98	0.03	0.04
	GBR	0.99	0.09	0.02	0.99	0.09	0.02	0.98	0.03	0.04	0.99	0.03	0.04
	LSTM	0.99	0.05	0.01	0.99	0.06	0.01	0.95	0.22	0.01	0.97	0.23	0.05
	ANN	0.99	0.04	0.01	0.99	0.04	0.01	0.98	0.02	0.01	0.98	0.03	0.01
Temperature, Wind Speed, Relative Humidity	RF	0.91	0.51	0.17	0.91	0.18	0.26	0.95	0.06	0.01	0.95	0.06	0.01
	SVR	0.92	0.47	0.14	0.91	0.50	0.16	0.95	0.05	0.03	0.95	0.05	0.06
	GBR	0.91	0.47	0.14	0.91	0.45	0.15	0.95	0.05	0.01	0.96	0.06	0.01
	LSTM	0.92	0.38	0.09	0.92	0.49	0.15	0.96	0.19	0.05	0.96	0.20	0.05
	ANN	0.92	0.48	0.15	0.91	0.40	0.09	0.96	0.06	0.02	0.96	0.06	0.01
Temperature and Wind Speed	RF	0.90	0.69	0.26	0.89	0.72	0.27	0.79	0.19	0.05	0.79	0.20	0.05
	SVR	0.91	0.57	0.22	0.91	0.58	0.22	0.89	0.14	0.03	0.89	0.14	0.04

	GBR	0.92	0.56	0.21	0.91	0.58	0.22	0.85	0.17	0.05	0.85	0.18	0.05
	LSTM	0.92	0.52	0.18	0.93	0.59	0.23	0.77	0.21	0.06	0.77	0.22	0.06
	ANN	0.920	0.56	0.21	0.92	0.54	0.19	0.86	0.18	0.05	0.85	0.19	0.05
Temperature and Relative Humidity	RF	0.84	0.64	0.19	0.84	0.66	0.20	0.95	0.07	0.05	0.92	0.08	0.06
	SVR	0.88	0.54	0.18	0.88	0.60	0.21	0.92	0.05	0.07	0.95	0.06	0.07
	GBR	0.881	0.58	0.21	0.88	0.60	0.21	0.92	0.06	0.05	0.99	0.06	0.01
	LSTM	0.85	0.48	0.15	0.85	0.56	0.18	0.95	0.19	0.05	0.77	0.20	0.05
	ANN	0.88	0.58	0.20	0.87	0.50	0.15	0.95	0.06	0.01	0.95	0.06	0.01

Table 6.2: Performance of Budyko and Turc methods in estimating AET using PET modelled from Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long Short-Term Memory (LSTM) and Artificial Neural Network(ANN) for Priestley Taylor Method under different input combinations.

Parameters	Model	Hyderabad						Waipara					
		Budyko			Turc			Budyko			Turc		
		R ²	RMSE (mm/d)	MAE (mm/d)									
All parameters	RF	0.94	0.09	0.02	0.94	0.20	0.03	0.93	0.04	0.04	0.93	0.04	0.05
	SVR	0.96	0.08	0.04	0.96	0.18	0.04	0.88	0.06	0.02	0.88	0.06	0.09
	GBR	0.96	0.06	0.02	0.96	0.01	0.01	0.88	0.06	0.03	0.88	0.06	0.03

	LSTM	0.99	0.01	0.022	0.99	0.03	0.024	0.92	0.05	0.03	0.92	0.05	0.08
	ANN	0.95	0.07	0.03	0.95	0.17	0.03	0.92	0.05	0.08	0.92	0.05	0.08
Temperature and Rs	RF	0.92	0.23	0.03	0.92	0.24	0.03	0.49	0.18	0.03	0.49	0.19	0.03
	SVR	0.93	0.22	0.06	0.93	0.23	0.06	0.53	0.16	0.03	0.53	0.17	0.03
	GBR	0.93	0.22	0.05	0.93	0.23	0.05	0.54	0.17	0.04	0.54	0.18	0.04
	LSTM	0.86	0.30	0.04	0.86	0.31	0.05	0.54	0.17	0.03	0.54	0.17	0.03
	ANN	0.93	0.21	0.05	0.93	0.22	0.05	0.53	0.17	0.03	0.53	0.18	0.03
Temperature and Relative Humidity	RF	0.89	0.26	0.04	0.89	0.28	0.04	0.58	0.16	0.03	0.57	0.17	0.03
	SVR	0.93	0.21	0.04	0.93	0.22	0.04	0.60	0.13	0.03	0.60	0.14	0.01
	GBR	0.93	0.22	0.03	0.92	0.23	0.04	0.60	0.15	0.03	0.60	0.16	0.03
	LSTM	0.97	0.11	0.02	0.97	0.63	0.12	0.03	0.15	0.03	0.61	0.16	0.03
	ANN	0.93	0.22	0.03	0.93	0.28	0.05	0.59	0.15	0.03	0.59	0.16	0.03

Table 6.3: Performance of Budyko and Turc methods in estimating AET using PET modelled from Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long Short-Term Memory (LSTM) and Artificial Neural Network(ANN) for Hargreaves Method under different input combinations.

Parameters	Model	Hyderabad						Waipara					
		Budyko			Turc			Budyko			Turc		
		R ²	RMSE (mm/d)	MAE (mm/d)									
All parameters	RF	0.99	0.01	0.01	0.99	0.01	0.02	0.97	0.07	0.02	0.97	0.07	0.02
	SVR	0.98	0.04	0.02	0.98	0.04	0.02	0.99	0.02	0.06	0.99	0.02	0.06
	GBR	0.99	0.01	0.03	0.99	0.01	0.04	0.97	0.07	0.02	0.98	0.07	0.02
	LSTM	0.99	0.01	0.02	0.99	0.01	0.01	0.97	0.02	0.03	0.97	0.02	0.08
	ANN	0.99	0.01	0.03	0.99	0.01	0.01	0.99	0.02	0.07	0.99	0.02	0.08
Minimum Temperature, Rs	RF	0.99	0.25	0.05	0.99	0.29	0.05	0.97	0.04	0.08	0.97	0.07	0.05
	SVR	0.98	0.13	0.03	0.98	0.13	0.04	0.99	0.02	0.06	0.98	0.02	0.06
	GBR	0.99	0.13	0.02	0.98	0.13	0.02	0.99	0.06	0.02	0.99	0.07	0.02
	LSTM	0.99	0.20	0.02	0.99	0.21	0.02	0.97	0.02	0.04	0.97	0.02	0.04
	ANN	0.99	0.13	0.02	0.98	0.14	0.03	0.99	0.02	0.05	0.99	0.02	0.06

Table 6.4: Performance of Budyko and Turc methods in estimating AET using PET modelled from Random Forest (RF), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long Short-Term Memory (LSTM) and Artificial Neural Network(ANN) for Turc Method under different input combinations.

Parameters	Model	Hyderabad						Waipara					
		Budyko			Turc			Budyko			Turc		
		R ²	RMSE (mm/d)	MAE (mm/d)									
All parameters	RF	0.99	0.02	0.01	0.99	0.04	0.05	0.99	0.06	0.04	0.99	0.06	0.04
	SVR	0.99	0.02	0.01	0.98	0.03	0.01	0.99	0.01	0.04	0.99	0.01	0.04
	GBR	0.99	0.04	0.04	0.99	0.05	0.04	0.99	0.01	0.04	0.99	0.01	0.04
	LSTM	0.99	0.01	0.02	0.99	0.01	0.03	0.99	0.02	0.01	0.99	0.02	0.01
	ANN	0.99	0.07	0.01	0.99	0.07	0.01	0.99	0.02	0.01	0.99	0.02	0.01
Minimum Temperature, Rs	RF	0.99	0.01	0.01	0.99	0.01	0.04	0.93	0.10	0.01	0.93	0.10	0.01
	SVR	0.98	0.02	0.04	0.98	0.02	0.04	0.99	0.08	0.01	0.99	0.09	0.01
	GBR	0.99	0.08	0.04	0.99	0.08	0.04	0.95	0.08	0.01	0.95	0.09	0.01
	LSTM	0.95	0.04	0.03	0.95	0.21	0.03	0.99	0.09	0.01	0.98	0.10	0.01
	ANN	0.99	0.08	0.04	0.99	0.08	0.01	0.99	0.09	0.01	0.99	0.10	0.01

6.3.1 ML models performance with various AET methods under different Input combinations

The performance of the two AET methods, Budyko and Turc, using PET obtained from ML models such as LSTM, ANN, SVR, GBR, and RF for the two stations of Hyderabad and Waipara using the Penman-Monteith method was provided in Table 6.1. The results demonstrated that the tested models generally had comparable performance over the two stations. Figure 6.1 shows the comparisons between observed AET using PET from Penman-Monteith and the AET estimated using modelled PET values in the form of a box plot for both the AET methods with all parameters as input combinations. The model estimated values using LSTM, ANN and SVR models showed closer agreement with observed AET estimates. Also during the testing period the LSTM and ANN models performed marginally better than the GBR, with estimated R^2 values of the two stations being Hyderabad 0.990 (LSTM), 0.998 (ANN), 0.990 (SVR) and 0.990 (GBR), 0.990 (RF) and for Waipara 0.990 (LSTM), 0.998 (ANN), 0.990 (SVR) and 0.990 (GBR), 0.990 (RF). For Hyderabad station, the results were the same for both the AET methods; it was seen that both Budyko and Turc performed well. It was also observed from the study that, among the evaluated models, LSTM and ANN models with all input combinations accomplished excellent performances, trailed by SVR. The performances for the models being LSTM (RMSE: 0.06 mm/d, MAE: 0.01 mm/d, and R^2 : 0.990) for Budyko method and (RMSE: 0.04 mm/d, MAE: 0.01 mm/d, and R^2 : 0.998) for Turc method, ANN (RMSE: 0.04 mm/d, MAE: 0.01 mm/d, and R^2 : 0.998), SVR (RMSE: 0.04 mm/d, MAE: 0.02 mm/d, and R^2 : 0.990), for Budyko as well as Turc methods. The GBR model could likewise accomplish good results with (RMSE: 0.09 mm/d, MAE: 0.02 mm/d, and R^2 : 0.987), while the RF model had also shown promising performance with (RMSE: 0.05 mm/d, MAE: 0.01 mm/d, and R^2 : 0.980).

The performances of the five ML models and the two AET methods for the Priestley Taylor method was provided in Table 6.2. Table 6.2 demonstrated that the tested models generally had comparable performance over the two stations. Figures 6.3-6.4 shows the comparisons between observed AET and model estimated values in the form of a box plot for two AET based methods under different input combinations for both the stations. From Table 10 for the Priestley method also results were same with the models LSTM, ANN and SVR estimated values showing closer agreement with those of observed AET followed by GBR and RF.

The performance of these ML models and the two AET methods for the Hargreaves and Turc methods was provided in Tables 6.3-6.4 for both the stations. Tables 6.3-6.4 demonstrated that the tested models generally had comparable performance over the two stations. Figures 6.4-6.9 shows the comparisons between observed AET and model estimated values in the form of a box plot for two AET based methods under different input combinations for Hargreaves and Turc methods. From Tables 6.3-6.4 the ranking of the models remained the same as Penman-Monteith and Priestley methods. We have

observed that when limited data combination PET from Chapter 5 was utilised to estimate AET, the results have been phenomenal, this can be seen in the Tables 6.3-6.9 for all the methods with their R^2 values ranging from 0.98-0.99.

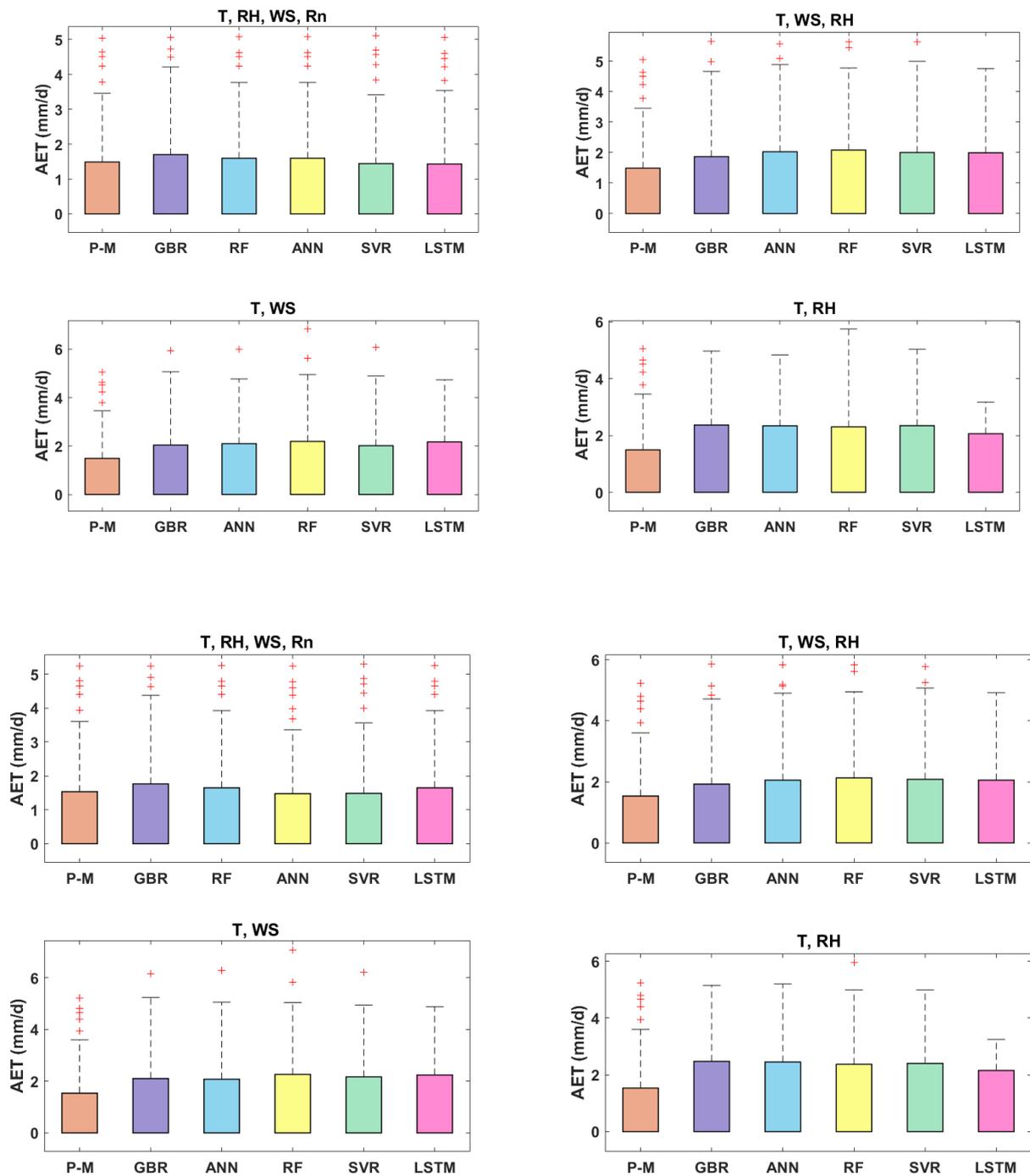


Figure 6.2: Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period for Budyko(Top) and Turc methods (Bottom) at Hyderabad station for Penman-Monteith method

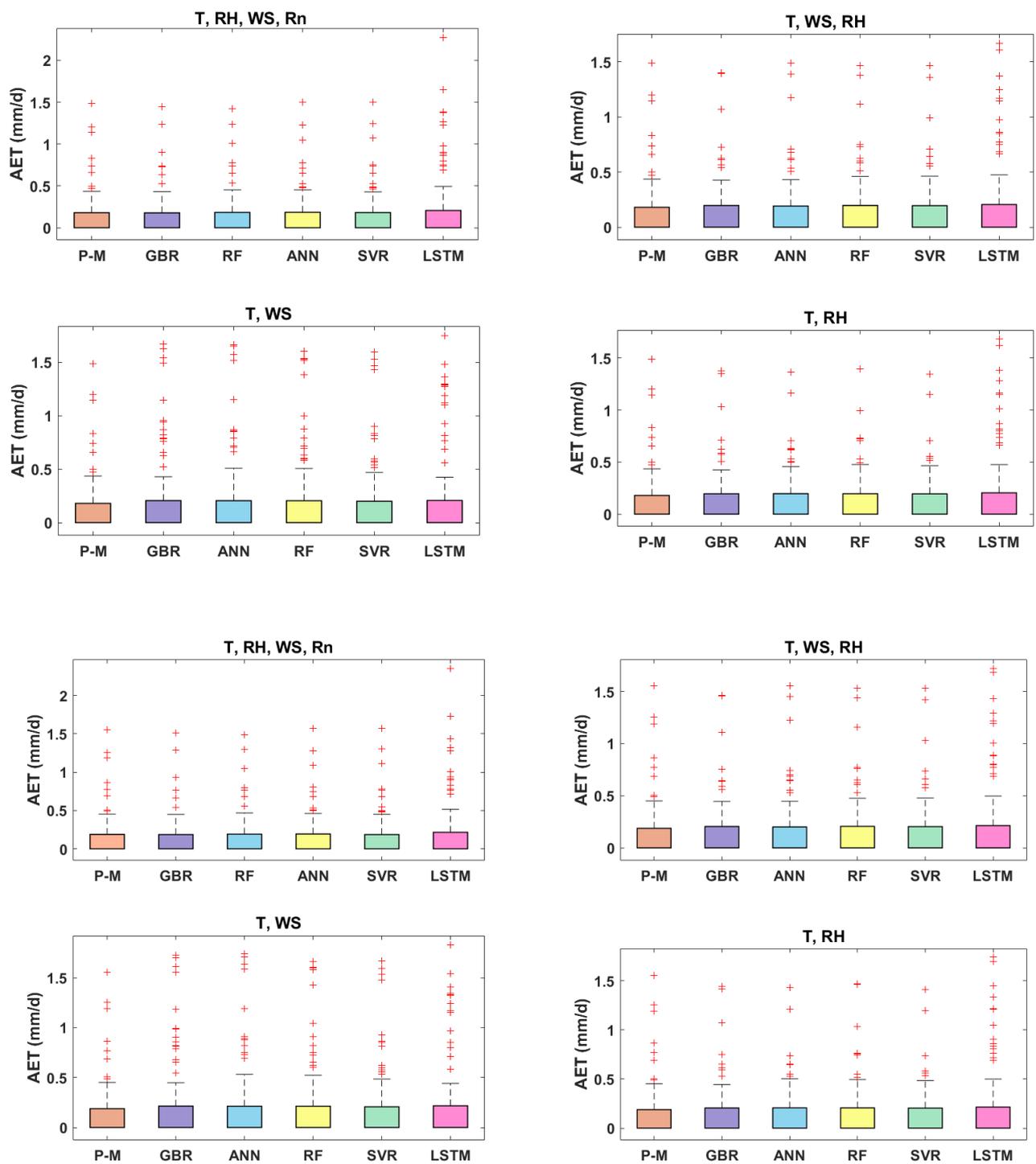


Figure 6.3: Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period for Budyko (Top) and Turc methods (Bottom) at Waipara station for penman monteith method.

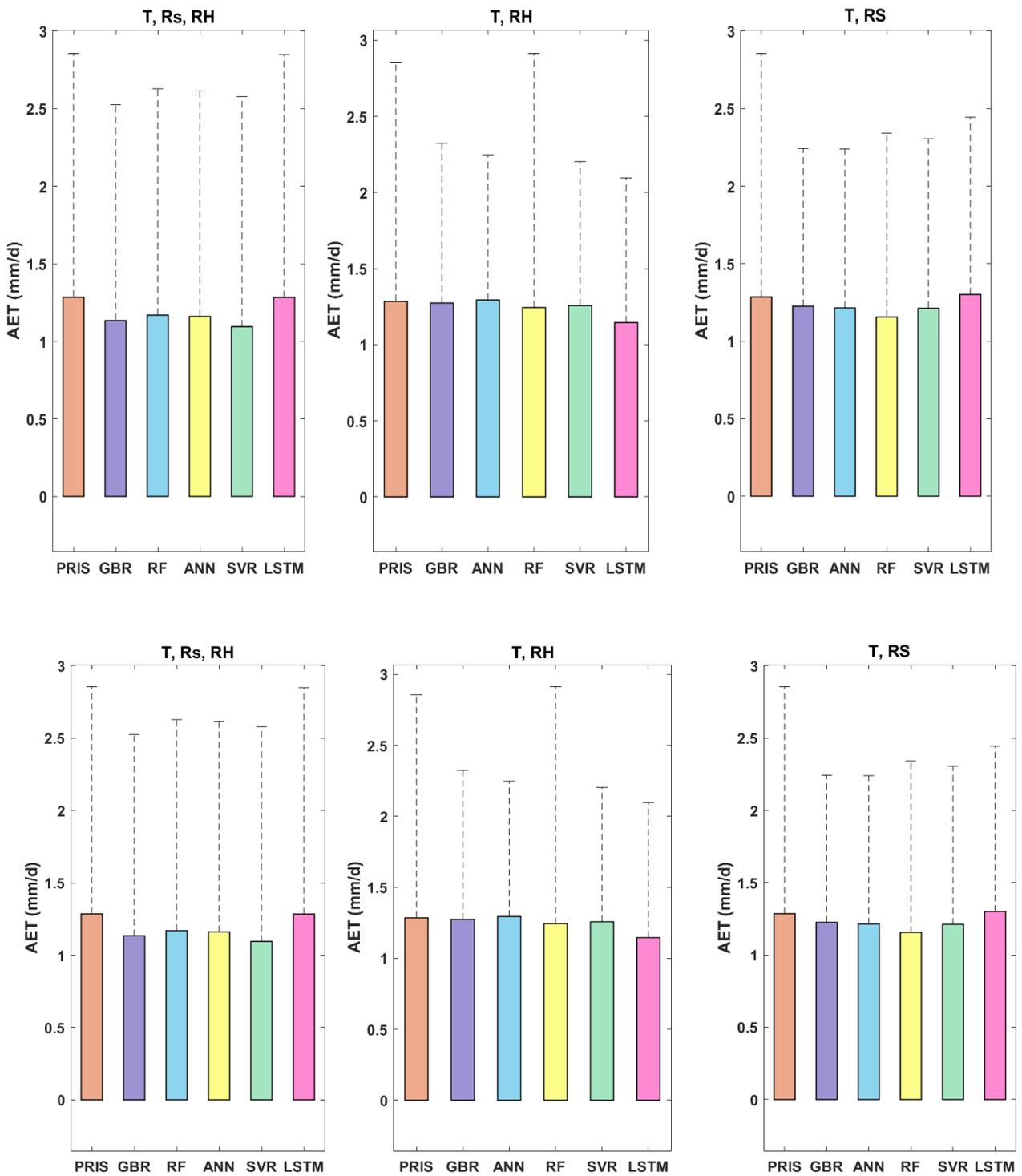


Figure 6.4: Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period for Budyko(Top) and Turc methods (Bottom) at Hyderabad station for Priestley method.

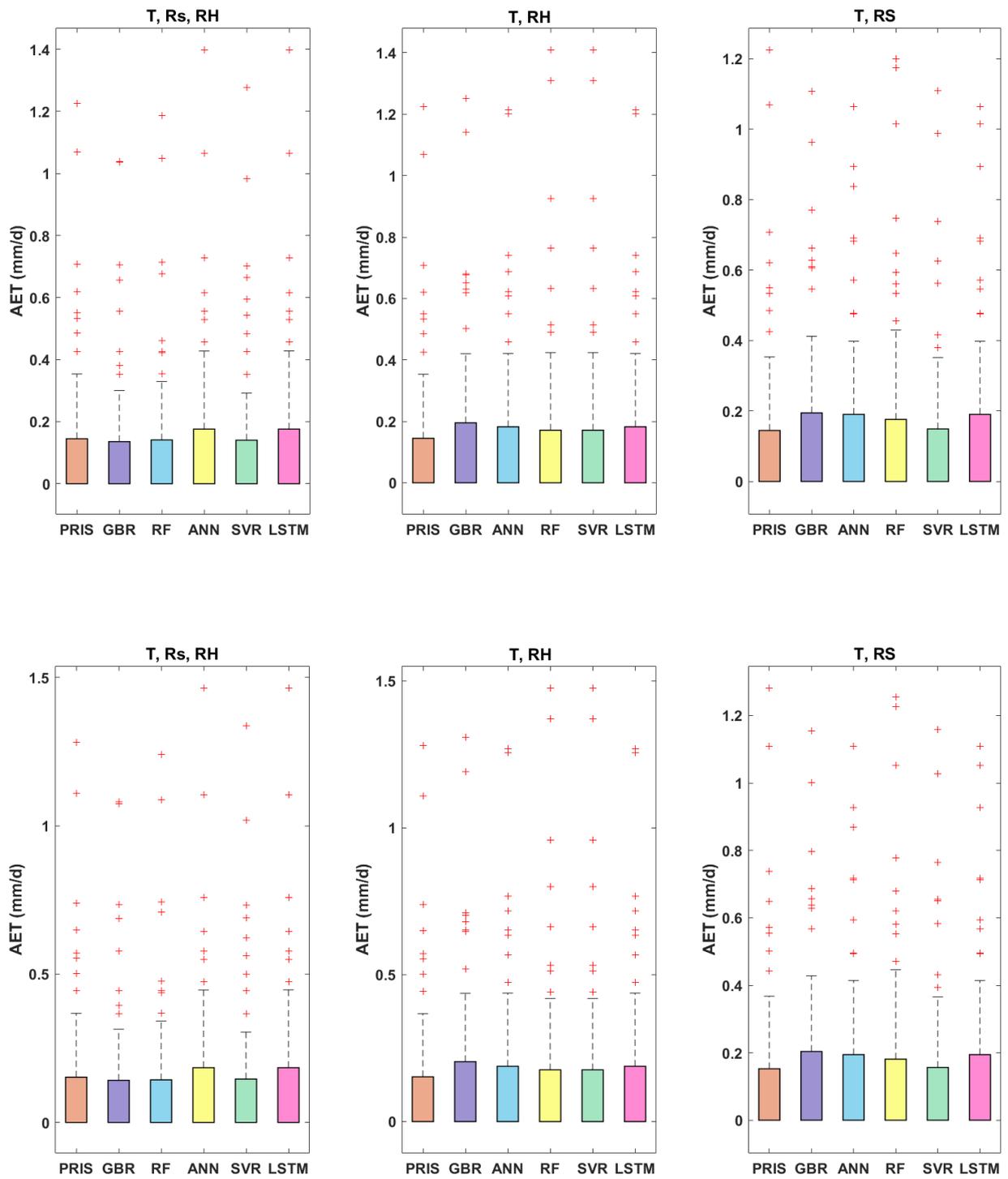


Figure 6.5: Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period for Budyko(Top) and Turc methods (Bottom) at Waipara station for Priestley method.

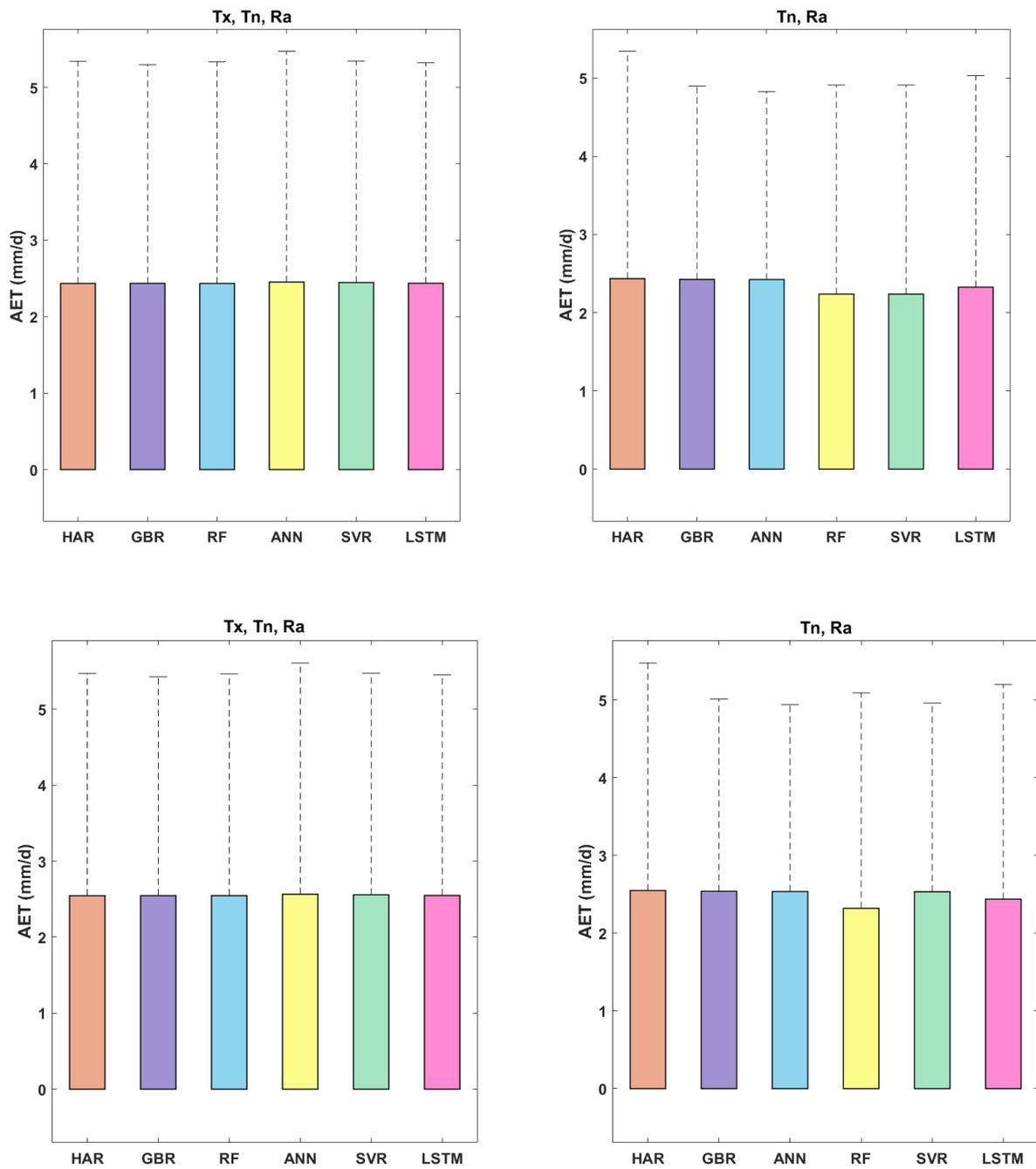


Figure 6.6: Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period for Budyko(Top) and Turc methods (Bottom) at Hyderabad station for Hargreaves method.

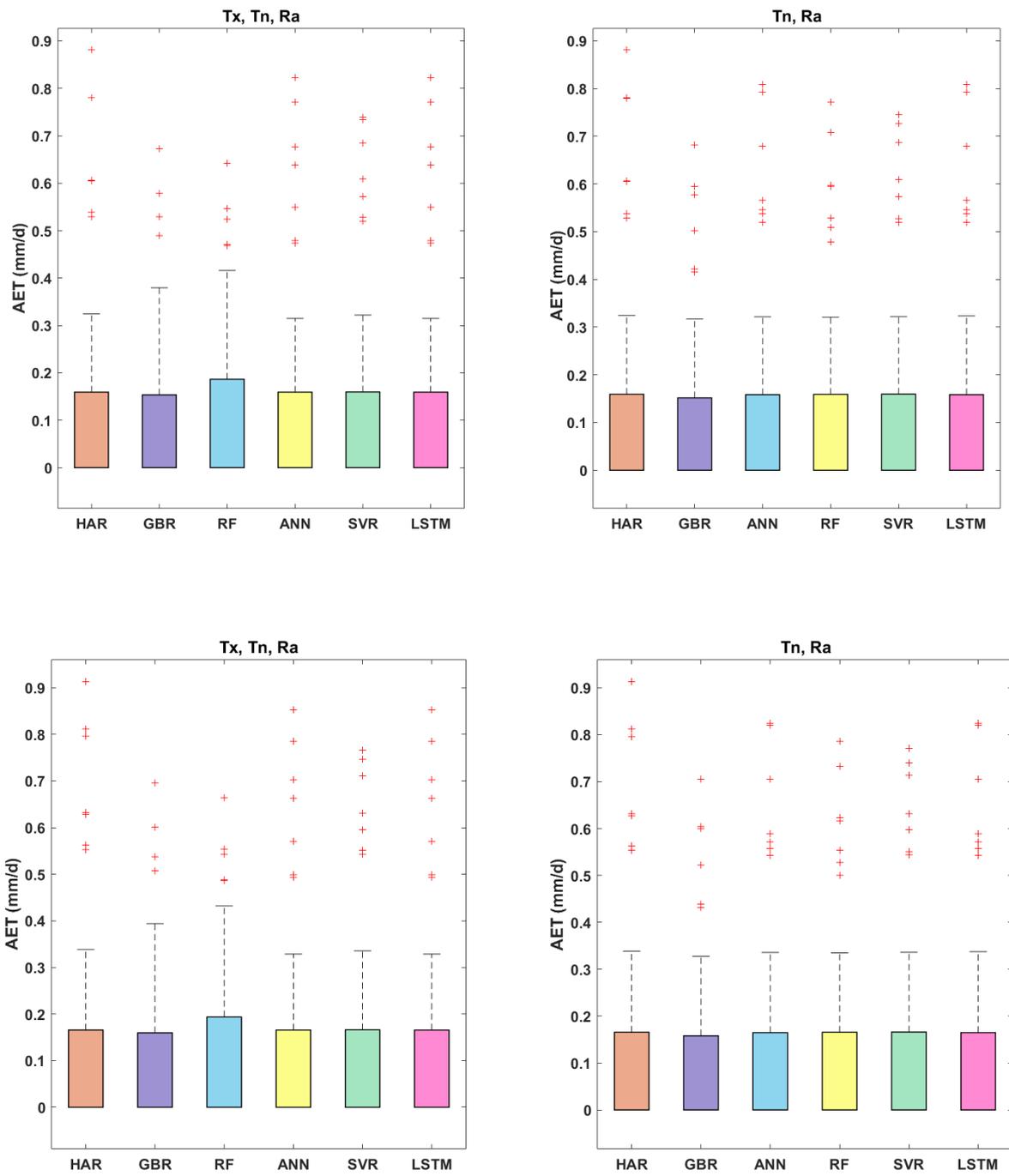


Figure 6.7: Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period for Budyko(left) and Turc methods (right) at Waipara station for Hargreaves method.

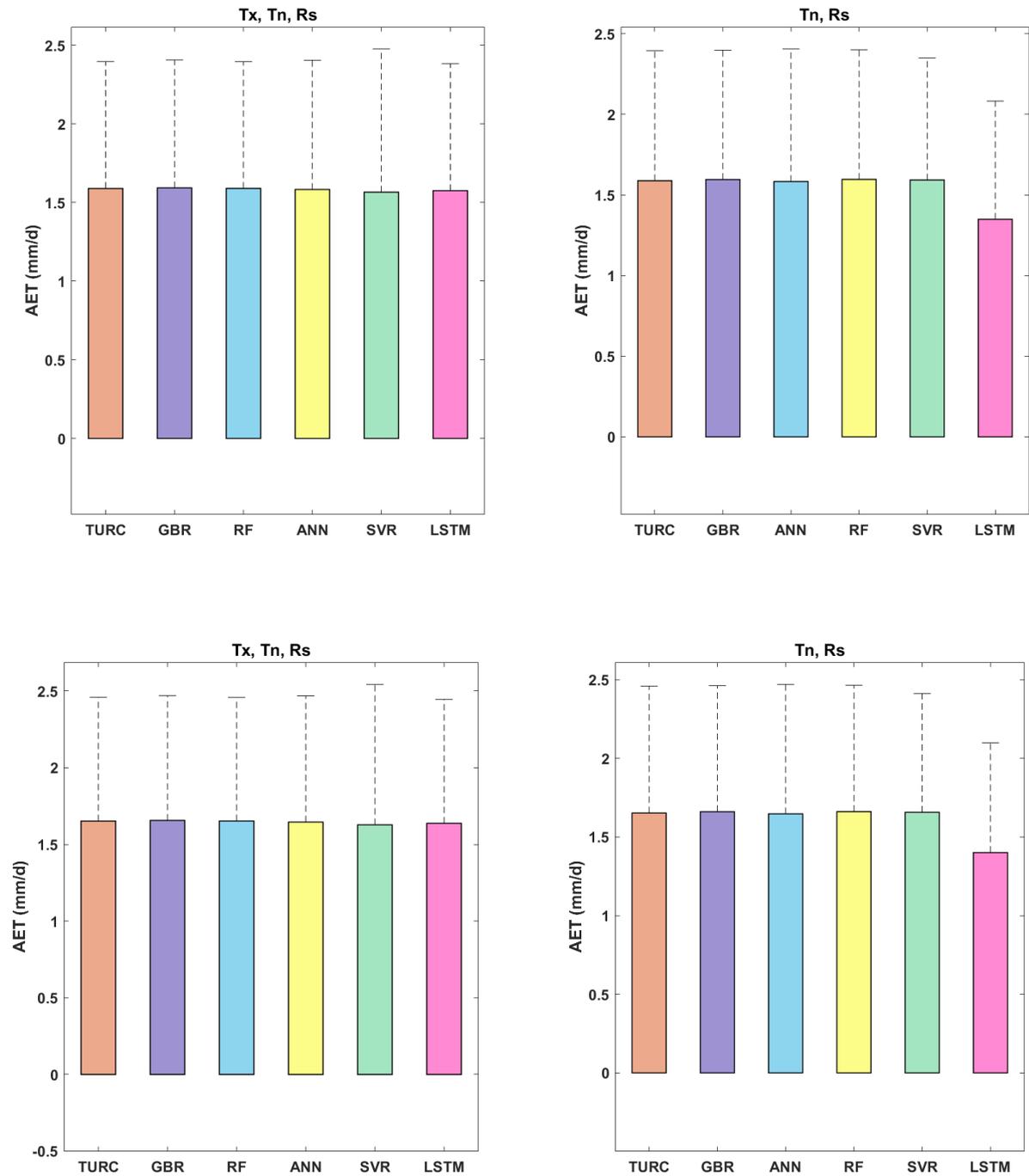


Figure 6.8: Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period for Budyko(Top) and Turc methods (bottom) at Hyderabad station for Turc method.

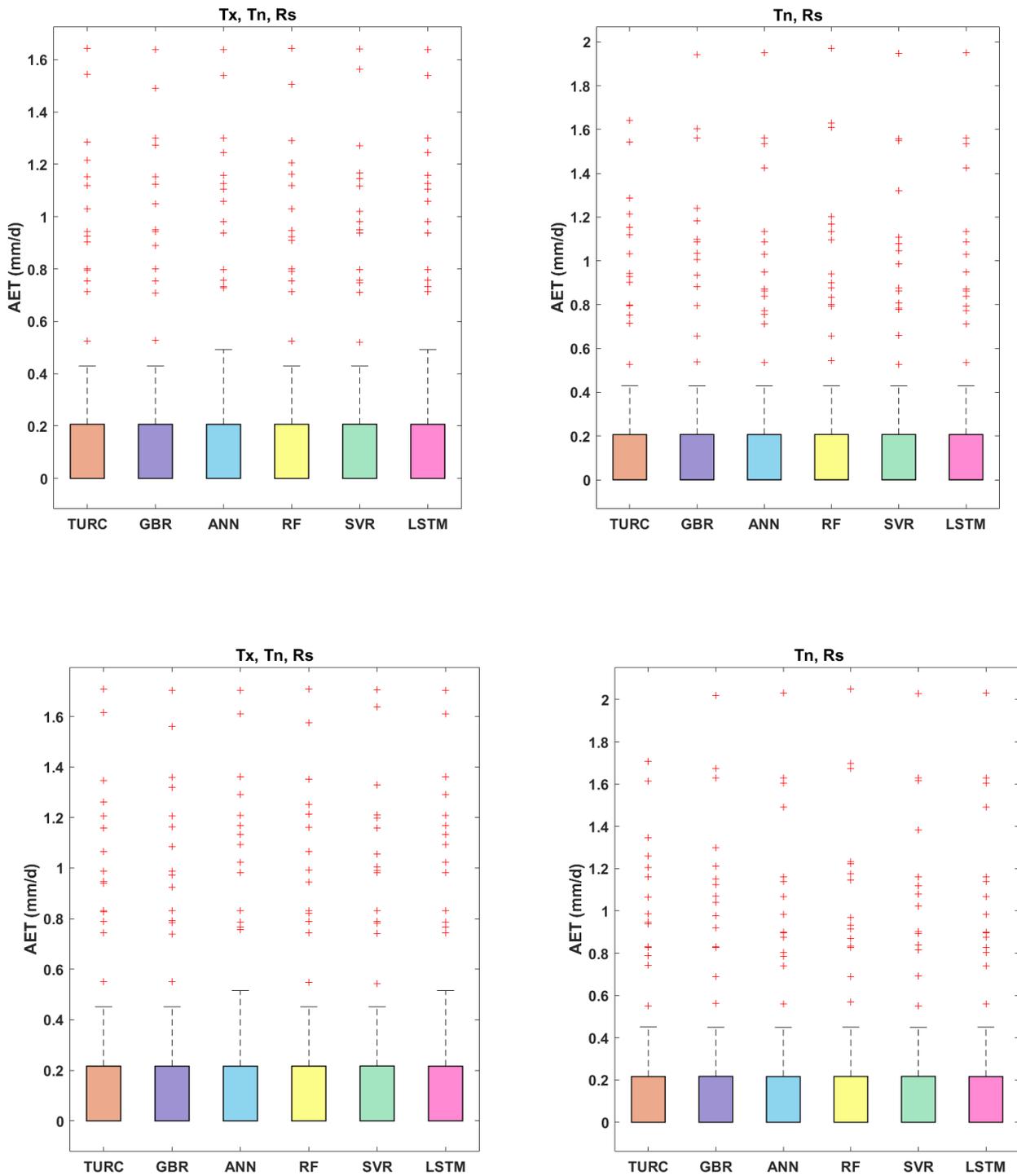


Figure 6.9: Comparison of observed and estimated AET using different models (GBR, ANN, LSTM, RF, and SVR) with varying parameters of input for the validation period for Budyko(Top) and Turc methods (bottom) at Waipara station for Turc method.

During the validation stage at Hyderabad station, the PET using the input combinations like temperature, relative humidity, and wind speed ranging results from (MAE: 0.01–0.09mm/d, RMSE :0.01-0.05 mm/d, and R^2 : 0.98–0.99) played out the best for both the AET methods of Budyko and Turc. However, all parameter input combinations outperformed among all input combinations. Methods using models with temperature and wind speed data as input combinations generally accomplished low performance of (RMSE: 0.56–0.58 mm/d, MAE: 0.17–0.21 mm/d, and R^2 : 0.88–0.90), trailed by methods dependent on temperature and relative humidity input (RMSE: 0.516-0.56 mm/d, MAE: 0.15-0.21 mm/d, and R^2 : 0.87- 0.88) for Penman-Monteith method. For the Priestley Taylor method, the AET methods generally accomplished low performance with temperature and solar radiation as input (RMSE: 0.21–0.31 mm/d, MAE: 0.04–0.12 mm/d, and R^2 : 0.865–0.939), followed by dependent on temperature and relative humidity input (RMSE: 0.23–0.31 mm/d, MAE: 0.04–0.21mm/d, and R^2 : 0.93–0.97). For Hargreaves and Turc methods, the minimum temperature and solar radiation combination performed the same as all parameters input combination for both the AET methods. It is worth to note that even for the AET estimates, the models which are blends of temperature data with relative humidity and wind speed, individually, could accomplish preferred performance over models dependent on temperature and relative humidity input. And the combination of temperature and solar radiation could also accomplish excellent performance compared to temperature and relative humidity while for the Priestley method. The outcomes demonstrated that the AET methods using PET modelled from LSTM and ANN performed superior to RF, SVR, and GBR, with temperature and wind speed as input combination. Furthermore, the AET with LSTM model showed the most remarkable performance when temperature, wind speed, and relative humidity data were accessible for the Penman-Monteith method and temperature and solar radiation when used for Priestley, Hargreaves and Turc methods. Among the five ML models performing in estimating AET, the second best model was noted as ANN, followed by SVR, RF, and GBR. Hence, LSTM and ANN can be concluded as the best ML models among any input combination which can be employed in calculating AET, whereas other models performed low when the input combinations were reduced. It was also noted that the two AET methods showed similar performance in all the used cases.

It was also observed that in all the AET methods, the results for the Penman-Monteith method were reasonable followed by the Priestley method. In Hargreaves and Turc methods, the inputs being the same, AET estimated using Turc method PET has shown much better performance under limited combinations when compared to Hargreaves. It was also confirmed that the Turc and Hargreaves methods showed better performance in estimating AET when compared to the Priestley method.

Tables 9-12 showed the summary of the Budyko and Turc methods AET estimates using modelled PET at Waipara station. The performance ranking of different ML models and AET methods was similar to the Hyderabad station, whereas the LSTM performed the best and RF as the worst. AET estimated using PET from all parameters as input combinations performed better compared to other input combinations

with its results ranging from (RMSE: 0.01–0.242mm/d, MAE: 0.02–0.07mm/d, and R^2 : 0.98–0.990). During the validation stage at Waipara station, the different input combinations like temperature, relative humidity, and wind speed (RMSE: 0.01–0.59mm/d, MAE: 0.02–0.08mm/d, and R^2 : 0.77–0.85) played out the best in estimating AET for Penman-Monteith method and temperature, relative humidity and solar radiation (RMSE: 0.01–0.561mm/d, MAE: 0.05–0.09mm/d, and R^2 : 0.76–0.84) played out the best for the Priestley method. In Hargreaves and Turc methods, the performances of all input combinations were almost the same, though the AET resulting from ANN, and LSTM performed much better than GBR and RF models. Hence, it is recommended to use either of these two PET models in estimating AET.

Taking into account the above-observed results from Tables 9-12 and Figures 17-24, the AET estimated using LSTM and ANN models are the most robust among the five ML modelled PET regardless of under which station or input combination, trailed by SVR and GBR modelled PET, which could generally accomplish agreeable accuracy in estimating AET. LSTM and ANN are both able to simulate AET where meteorological information is inadequate. Both the AET methods of Budyko and Turc gave outstanding performance. It can be concluded that the different ML modelled PET estimates employed in estimating AET at two different stations have performed promising when compared to the standard ET estimates.

With respect to the ML models, it can be concluded that ANN and LSTM can be preferred to model PET and AET with all parameters input combination as the first preference followed by three parameters input combination and then two parameters input combination. It can also be concluded that AET is highly sensitive to the number of input parameters used in the PET estimation. It's also worth noticing that the accuracy of the AET (Budyko and Turc) methods with all parameters input combination was the highest in each station. However, one can use the three input combination parameters or the two input combination parameters (e.g. temperature and relative humidity or temperature and wind speed) in PET and AET estimations under limited data.

This study has attempted to estimate AET from ML modelled PET and concluded that this could be used in future for different case studies. Using the five different ML models: ANN, LSTM, GBR, SVR, RF and different empirical methods, different AET estimates were developed. The models were analysed and compared in terms of prediction accuracy, generalization ability, complexity/simplicity of the modelling approach, and model usage. The comparison was conducted to identify the most efficient technique out of the studied five ML techniques to predict the PET in developing AET estimates. The results of the ML models were compared with those of the five empirical methods and two AET modelled methods to identify the possible advantages of the proposed models over one of the available methods for the estimation of AET under limited data. Although the generated models, based on error measures, are performing well and relatively similar, they use different combinations of inputs with different mathematical structures. This demonstrates that precise

identification of the meteorological variables driving the AET process is not a straightforward task, where different combinations of inputs may result in relatively good AET estimates.

6.3.2 Discussions

Among the proposed models in this study, Budyko and Turc are both equation-based methods. Such explicit equation-based methods are more appealing to hydrological practitioners because of the transparency and the simplicity of their application. The parameters that were observed to have the most significant contribution in modelling PET for generating AET were almost the same for both AET methods. This was observed using five different ML models, which could adequately capture the most relevant meteorological variables. According to the observed two AET methods, the meteorological variables that were found to be most important in predicting AET variations (mainly large-scale) were net Precipitation (P), Radiation (Rn) and Temperature (T). Rn is a well-known variable that serves as an energy source and is one of the critical components of the ET mechanism. The surface soil moisture, surface soil temperature, and turbulent sensible heat flux are few other essential elements that shape the physics of AET (Z. Wang et al., 2017).

The physical description of the AET mechanism, which can be found in many papers and literature ((Dingman, 2015; Z. Wang et al., 2017)), adequately explains the importance of soil moisture and its complex interaction with other land-atmosphere variables in the AET process. At daily time scales, it was interesting to investigate the level of cause and relationship between meteorological variables such as temperature, relative humidity, solar radiation and wind speed involved in AET variations. To make this comparison, the time series of PET and AET using temperature, relative humidity, solar radiation and wind speed were visually compared over a typical time window. The comparison was carried out on daily time series data for both stations. It can be concluded that temperature varies slightly over time compared to other variables. It can also be noted that the temperature and solar radiation gradually decreases over time, resulting in a lower value. As a result, it is clear that precipitation and temperature influence the AET estimation over a daily time scale. Because the process of AET is not fully understood, it is challenging to mechanistically capture the interactions that exist among the state variables to present a mathematical relationship between AET and highly correlated meteorological variables. Explicit ML models demonstrated their ability to efficiently capture PET variations in estimating AET and to induce symbolic estimation models, which are primarily dominated by net radiation, temperature and wind speed.

6.4 Conclusions

The AET over semi-arid climatic conditions of Hyderabad, Telangana, (India) and Waipara (New Zealand) was estimated using ML modelled PET with different empirical methods. This PET was utilised in estimating AET from Budyko and Turc methods. The Penman-Monteith model-based AET was considered as the standard reference method. The daily AET rates were estimated using PET obtained from five different ML techniques, namely LSTM, ANN, SVR, GBR, RF, using different input combinations. The four input variable combinations such as maximum and minimum air temperature, relative humidity, solar radiation, and wind speed; three input variable combinations such as average air temperature, relative humidity, and solar radiation; two input variable combinations such as temperature and solar radiation; and temperature and wind speed. The study investigated that the best performance was when all input variable combinations were used. However, the study also found that even three input variable combinations or two combination input variable combinations can provide practically identical results. The study also investigated that the two methods Turc and Budyko performed equally well with good R^2 values. The results were discussed and compared with the results of alternative methods of PET calculation, such as the radiation-based methods of Priestly-Taylor, the temperature-based methods of Hargreaves, and Turc. The correlation coefficient values suggest that precipitation and temperature are the most important factors, followed by solar radiation, wind speed, and relative humidity, respectively for both the regions. For both the regions, LSTM and ANN were found to be more effective than the other techniques in identifying the most relevant meteorological predictors. It can be concluded that for the two regions, temperature and solar radiation have a maximum correlation with AET estimates of Penman-Monteith models compared to relative humidity and wind speed. The Turc model used temperature and solar radiation as input variables and showed high accuracy with all ML models in estimating AET. In contrast, relative humidity has the least correlation with the AET estimates. Due to the lower dependency of relative humidity on the AET estimates, the Priestly-Taylor model has lower accuracy with ML models compared to the Turc model. The results also showed that the AET obtained from PET using LSTM and ANN models could offer the most remarkable performance among four tested models regardless of station or input combination, trailed by SVR and GBR models, which could likewise accomplish moderately good performance. The study concludes that the empirical models work well with data-driven algorithms that consider the climate variables having high dependency with the standard PET estimates in calculating AET. Such studies can be implemented for the development of ML models statistically dependent on PET in AET estimates. The study demonstrated that the modelling of PET through the LSTM and ANN techniques gave better AET estimates that proved with their performance criterion, i.e., R^2 as 0.99. The study concluded that the performance of the AET methods varies according to the number of inputs and the predicted time step. Overall, results are of significant practical use when limited climate data is available

to estimate the AET. So, it can be concluded that even if not all parameter information is available in a particular station, the three-parameter combination can be used, or the two combinations, which are temperature and wind speed or temperature and relative humidity values, to calculate AET using modelled PET. It was also concluded that the equation-based methods Budyko and Turc, which was presented in this study, are not the only capacity of the methods as an evolutionary methodology that uses ML techniques. However, they can also be implemented to evolve program-based models and estimate AET in different regions. These ML models effectively determine the best input variables for modelling small and large scale PET variations in generating AET. Further, the modelled PET was efficient in estimating AET for two different stations. However, they have the potential to provide deep insight into the larger-scale variations of AET, such as diurnal variations. It should also be noted that no single ML technique can capture all complex processes at all times. As a result, various techniques may be capable of accurately predicting various challenging components of hydrology.

Chapter 7

Conclusions

Understanding the various hydrological processes is critical for the improvement and development of water resources management systems. In this study, one of the major components of the hydrological cycle, PET and AET were modelled using data-driven modelling techniques in two different climatic regions. The selection of the most appropriate and efficient ML model is critical for the accurate assessment of PET and AET. The proposed data-driven models can be used to estimate the PET and AET in different climatic regions using standard meteorological variables. Using the developed modelled-based PET, the values of AET can be estimated using a limited number of meteorological variables that are typically measured or can be measured in most weather stations. The PET and AET methods used in this study were developed based on region specific climatological conditions and can be implementable on similar climatological conditions. Evaluation of the performance of the locally proposed models on other regions and climate zones can be potential areas of research. Data-driven techniques can be easily implementable to develop local PET models which can serve as input to AET models. This type of modelling may be more efficient and straightforward than physically-based and empirical modelling approaches. The research findings of the present study demonstrated the utility of ML models as a tool for this purpose. This study also demonstrated the significance of taking into account the time-scale of variations in the analysis and modelling of hydrological time series. The research effort of the present thesis has the potential to improve the current understanding of PET and AET variations, as well as their relation with various climatological variables, which can be very useful in the interpretation of prediction models. This type of information can be used to analyse and model any hydrological time series, which improves the success of the modelling procedures and, as a result, the understanding and monitoring of the hydrology and its processes. The proposed methodology was demonstrated on two different stations, namely Hyderabad (India) and Waipara (New Zealand). The research findings of the study are summarised as follows.

- The study estimated the PET using different empirical methods such as Penman-Monteith method, Priestley Taylor method, Hargreaves method and Turc at two different stations, namely Hyderabad and Waipara Stations. The Penman-Monteith model estimates of PET were considered as standard reference models for various temperature and radiation-based empirical models and also for data-driven models.
- The daily PET estimates were modelled using five different modelling techniques, namely LSTM, ANN, SVR, GBR, RF using four input variables as maximum and minimum air temperatures, relative humidity and solar radiation, wind speed; three input variables as average

air temperature, relative humidity, and solar radiation; two input variables as temperature and solar radiation; two input variables as temperature and wind speed.

- The study concluded that the best performance can be achieved when all meteorological variables were used; the study, however, finds that even three input variable combinations (temperature, wind speed and relative humidity values) or two combination input variables (temperature and relative humidity, temperature and wind speed) also can provide practically identical results as using all data.
- The study concluded that temperature is the most important factor followed by solar radiation, wind speed, and relative humidity, respectively. Temperature and solar radiation have a maximum correlation with PET estimates of Penman-Monteith models as compared to relative humidity and wind speed. Relative humidity has a lower dependency on the PET estimates.
- The results also showed that the LSTM and ANN models showing (LSTM (RMSE: 0.02 mm/d, MAE: 0.01 mm/d, and R^2 : 0.990), ANN (RMSE: 0.06 mm/d, MAE: 0.05 mm/d, and R^2 : 0.998), could offer the most remarkable performance among four tested models regardless of station or input combination, trailed by SVR and GBR models, which could likewise accomplish moderately good performance.
- Among the tested models, LSTM and ANN model not only achieved the smallest average RMSE value, but also the most concentrated distribution of RMSE values independent of the input combinations, which indicated that these two models had the best precision stability with accuracy of 99.10 %. Followed by SVR with 92.70 % accuracy.
- The study concludes that the performance of the models varies according to the number of inputs as well as the predicted time step. Overall, results are of significant practical use when limited climate data is available to estimate PET. So, it can be concluded that even if not all parameter information is available over a particular region, this study proved that the three-parameter or two-parameter combinations, which are temperature and wind speed or temperature and relative humidity values, can be used to estimate PET.
- The study used the most promising data-driven modelling techniques achieved in the estimation of PET in estimating the AET mechanism using two empirical based models of Budyko and Turc.
- The proposed empirical-based AET models, Budyko and Turc, showed that the AET process has the potential to be estimated by structurally simple methods. Equation-based AET methods made it possible to extract useful information about the hydrological process. It was observed that the meteorological variables of temperature and solar radiation have more significant contributions than other variables in the estimation of AET. In addition, the effects

of the meteorological variables were found to be essential and effective in the estimation of AET.

- It was also concluded that under limited data availability, the best method to be employed to estimate PET among empirical is Turc method and among machine learning models is LSTM and ANN.
- It was also concluded that under limited data availability, input combinations were identified as temperature and wind speed for estimating PET; temperature, wind speed and precipitation for estimating AET for semi-arid climatology.
- Overall, the research findings of the study stress on the use of limited data in understanding the complex hydrological processes such as PET and AET using data-driven and empirical based approaches and suitability of best method for diverse climatological conditions.

Publications:

Journal Publications:

1. Adeeba Ayaz, Maddu Rajesh, Shailesh Kumar Singh, Shaik Rehana. Estimation of reference evapotranspiration using machine learning models with limited data. AIMS Geosciences, 2021, 7(3): 268-290. Doi: 10.3934/geosci.2021016.

<http://www.aimspress.com/article/doi/10.3934/geosci.2021016>

Conference Publications:

1. Adeeba Ayaz and Shaik Rehana (2021), Modelling of reference evapotranspiration for semi-arid climate using machine learning models, International Virtual conference on Innovative Trends in Hydrological and Environmental Systems (ITHES-2021), 28th-30th April, 2021, Water & Environment Division, Department of Civil Engineering, National Institute of Technology Warangal, India.

2. Rehana, S., Sireesha Naidu, G., Apaar, A., Rajan, K.S., Ayaz, A. (2019), Spatio-temporal evaluation of evapotranspiration over Krishna River basin, India, Hydro 2019, International Conference, Hydraulics, Water Resources & Coastal Engineering, 18-20, December, 2019.

Book Chapter:

1. Ayaz A., Chandra S., Mandlecha P., Shaik R. (2021), Modelling of Reference Evapotranspiration for Semi-arid Climates Using Artificial Neural Network. In: Majumder M., Kale G.D. (eds) Water and Energy Management in India. Springer, Cham. https://doi.org/10.1007/978-3-030-66683-5_7

2. Adeeba Ayaz and Shaik Rehana (2021), Modelling of reference evapotranspiration for semi-arid climate using machine learning models (2021), In: Lecture Notes in Civil Engineering, Innovative Trends in Hydrological and Environmental Systems - Select Proceedings of ITHES 2021, Edited by: Dr. Anil Kumar Dikshit, Dr. Balaji Narasimhan, Dr. Bimlesh Kumar and Dr. Ajey Kumar Patel (under Preparation).

Bibliography

- [1] K. Parasuraman, A. Elshorbagy, and S. K. Carey, 'Modelling the dynamics of the evapotranspiration process using genetic programming', *Hydrol. Sci. J.*, vol. 52, no. 3, pp. 563–578, Jun. 2007, doi: 10.1623/hysj.52.3.563.
- [2] Y. B. Dibike, S. Velickov, D. Solomatine, and M. B. Abbott, 'Model Induction with Support Vector Machines: Introduction and Applications', *J. Comput. Civ. Eng.*, vol. 15, no. 3, pp. 208–216, Jul. 2001, doi: 10.1061/(ASCE)0887-3801(2001)15:3(208).
- [3] V. Diaz, G. A. Corzo Perez, H. A. J. Van Lanen, D. Solomatine, and E. A. Varouchakis, 'Characterisation of the dynamics of past droughts', *Sci. Total Environ.*, vol. 718, p. 134588, May 2020, doi: 10.1016/j.scitotenv.2019.134588.
- [4] J. Shiri, 'Evaluation of a neuro-fuzzy technique in estimating pan evaporation values in low-altitude locations', *Meteorol. Appl.*, vol. 26, no. 2, pp. 204–212, Apr. 2019, doi: 10.1002/met.1753.
- [5] S. Kumar Singh, N. Marcy, 1 National Institute of Water and Atmospheric Research, Christchurch, New Zealand, and 2 Ecole Polytechnique Universitaire de Montpellier, Montpellier, France, 'Comparison of Simple and Complex Hydrological Models for Predicting Catchment Discharge Under Climate Change', *AIMS Geosci.*, vol. 3, no. 3, pp. 467–497, 2017, doi: 10.3934/geosci.2017.3.467.
- [6] F. H. S. Chiew and T. A. McMahon, 'Modelling the impacts of climate change on Australian streamflow', *Hydrol. Process.*, vol. 16, no. 6, pp. 1235–1245, Apr. 2002, doi: 10.1002/hyp.1059.
- [7] Y. Liu, M. Hejazi, H. Li, X. Zhang, and G. Leng, 'A hydrological emulator for global applications – HE v1.0.0', *Geosci. Model Dev.*, vol. 11, no. 3, pp. 1077–1092, Mar. 2018, doi: 10.5194/gmd-11-1077-2018.
- [8] L. Peng, Z. Zeng, Z. Wei, A. Chen, E. F. Wood, and J. Sheffield, 'Determinants of the ratio of actual to potential evapotranspiration', *Glob. Change Biol.*, vol. 25, no. 4, pp. 1326–1343, Apr. 2019, doi: 10.1111/gcb.14577.
- [9] M. Tasumi, 'Estimating evapotranspiration using METRIC model and Landsat data for better understandings of regional hydrology in the western Urmia Lake Basin', *Agric. Water Manag.*, vol. 226, p. 105805, Dec. 2019, doi: 10.1016/j.agwat.2019.105805.
- [10] R. G. Allen and Food and Agriculture Organization of the United Nations, Eds., *Crop evapotranspiration: guidelines for computing crop water requirements*. Rome: Food and Agriculture Organization of the United Nations, 1998.
- [11] S. Irmak and D. Z. Haman, 'Evaluation of Five Methods for Estimating Class A Pan Evaporation in a Humid Climate', *HortTechnology*, vol. 13, no. 3, pp. 500–508, Jan. 2003, doi: 10.21273/HORTTECH.13.3.0500.
- [12] D. N. Yates, 'Approaches to continental scale runoff for integrated assessment models', *J. Hydrol.*, vol. 201, no. 1–4, pp. 289–310, Dec. 1997, doi: 10.1016/S0022-1694(97)00044-9.
- [13] J. Almorox, V. H. Quej, and P. Martí, 'Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Köppen climate classes', *J. Hydrol.*, vol. 528, pp. 514–522, Sep. 2015, doi: 10.1016/j.jhydrol.2015.06.057.
- [14] C. W. Thornthwaite, 'An Approach toward a Rational Classification of Climate', *Geogr. Rev.*, vol. 38, no. 1, p. 55, Jan. 1948, doi: 10.2307/210739.
- [15] H. L. Penman, 'Natural Evaporation from Open Water, Bare Soil and Grass', *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.*, vol. 193, no. 1032, pp. 120–145, 1948, [Online]. Available: <http://www.jstor.org/stable/98151>

- [16] Y. Feng, N. Cui, L. Zhao, X. Hu, and D. Gong, 'Comparison of ELM, GANN, WNN and empirical models for estimating reference evapotranspiration in humid region of Southwest China', *J. Hydrol.*, vol. 536, pp. 376–383, May 2016, doi: 10.1016/j.jhydrol.2016.02.053.
- [17] J. Yao, Y. Zhao, Y. Chen, X. Yu, and R. Zhang, 'Multi-scale assessments of droughts: A case study in Xinjiang, China', *Sci. Total Environ.*, vol. 630, pp. 444–452, Jul. 2018, doi: 10.1016/j.scitotenv.2018.02.200.
- [18] W. Verhoef and H. Bach, 'Remote sensing data assimilation using coupled radiative transfer models', *Phys. Chem. Earth Parts ABC*, vol. 28, no. 1–3, pp. 3–13, Jan. 2003, doi: 10.1016/S1474-7065(03)00003-2.
- [19] S. M. Biazar, Y. Dinpashoh, and V. P. Singh, 'Sensitivity analysis of the reference crop evapotranspiration in a humid region', *Environ. Sci. Pollut. Res.*, vol. 26, no. 31, pp. 32517–32544, Nov. 2019, doi: 10.1007/s11356-019-06419-w.
- [20] H. Tabari, M. E. Grismer, and S. Trajkovic, 'Comparative analysis of 31 reference evapotranspiration methods under humid conditions', *Irrig. Sci.*, vol. 31, no. 2, pp. 107–117, Mar. 2013, doi: 10.1007/s00271-011-0295-z.
- [21] C. H. B. Priestley and R. J. Taylor, 'On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters', *Mon. Weather Rev.*, vol. 100, no. 2, pp. 81–92, Feb. 1972, doi: 10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2.
- [22] G. H. Hargreaves, 'Closure to " Estimating Potential Evapotranspiration " by George H. Hargreaves and Zohrab A. Samani (September, 1982)', *J. Irrig. Drain. Eng.*, vol. 109, no. 3, pp. 343–344, Sep. 1983, doi: 10.1061/(ASCE)0733-9437(1983)109:3(343).
- [23] A. R. Pereira, 'The Priestley–Taylor parameter and the decoupling factor for estimating reference evapotranspiration', *Agric. For. Meteorol.*, vol. 125, no. 3–4, pp. 305–313, Oct. 2004, doi: 10.1016/j.agrformet.2004.04.002.
- [24] W. R. Hamon, 'Estimating Potential Evapotranspiration', *J. Hydraul. Div.*, vol. 87, no. 3, pp. 107–120, May 1961, doi: 10.1061/JYCEAJ.0000599.
- [25] W. R. Hamon, 'Estimating Potential Evapotranspiration', *Trans. Am. Soc. Civ. Eng.*, vol. 128, no. 1, pp. 324–338, Jan. 1963, doi: 10.1061/TACEAT.0008673.
- [26] L. Oudin *et al.*, 'Which potential evapotranspiration input for a lumped rainfall–runoff model?', *J. Hydrol.*, vol. 303, no. 1–4, pp. 290–306, Mar. 2005, doi: 10.1016/j.jhydrol.2004.08.026.
- [27] W. Baier and Geo. W. Robertson, 'A new versatile soil moisture budget', *Can. J. Plant Sci.*, vol. 46, no. 3, pp. 299–315, May 1966, doi: 10.4141/cjps66-049.
- [28] J. Papadakis, 'Climates of the world and their agricultural potentialities.', *Clim. World Their Agric. Potentialities*, 1966, Accessed: Jul. 27, 2021. [Online]. Available: <https://www.cabdirect.org/cabdirect/abstract/19671701437>
- [29] V. H. Malmström, 'A New Approach To The Classification Of Climate', *J. Geogr.*, vol. 68, no. 6, pp. 351–357, Sep. 1969, doi: 10.1080/00221346908981131.
- [30] A. R. Pereira and Â. Paes De Camargo, 'An analysis of the criticism of thornthwaite's equation for estimating potential evapotranspiration', *Agric. For. Meteorol.*, vol. 46, no. 1–2, pp. 149–157, Apr. 1989, doi: 10.1016/0168-1923(89)90118-4.
- [31] S. Traore and A. Guven, 'Regional-Specific Numerical Models of Evapotranspiration Using Gene-Expression Programming Interface in Sahel', *Water Resour. Manag.*, vol. 26, no. 15, pp. 4367–4380, Dec. 2012, doi: 10.1007/s11269-012-0149-3.
- [32] H. A. R. de Bruin and W. N. Lablans, 'Reference crop evapotranspiration determined with a modified Makkink equation', *Hydrol. Process.*, vol. 12, no. 7, pp. 1053–1062, 1998, doi: 10.1002/(SICI)1099-1085(19980615)12:7<1053::AID-HYP639>3.0.CO;2-E.
- [33] S. Mohammad Fakhru Islam and Z. Karim, 'World's Demand for Food and Water: The Consequences of Climate Change', in *Desalination - Challenges and Opportunities*, M. Hossein Davood Abadi Farahani, V. Vatanpour, and A. Hooshang Taheri, Eds. IntechOpen, 2020. doi: 10.5772/intechopen.85919.

- [34] I. Watson and A. Burnett, *Hydrology: An Environmental Approach*. 2017, p. 704. doi: 10.1201/9780203751442.
- [35] S. L. Dingman, *Physical hydrology*, Third edition. Long Grove, Illinois: Waveland Press, Inc, 2015.
- [36] Z. Wei, K. Yoshimura, L. Wang, D. G. Miralles, S. Jasechko, and X. Lee, 'Revisiting the contribution of transpiration to global terrestrial evapotranspiration: Revisiting Global ET Partitioning', *Geophys. Res. Lett.*, vol. 44, no. 6, pp. 2792–2801, Mar. 2017, doi: 10.1002/2016GL072235.
- [37] V. E. Hansen, O. W. Israelsen, and G. E. Stringham, *Irrigation principles and practices*, 4.ed. New York: Wiley, 1979.
- [38] B.-L. Xue, L. Wang, X. Li, K. Yang, D. Chen, and L. Sun, 'Evaluation of evapotranspiration estimates for two river basins on the Tibetan Plateau by a water balance method', *J. Hydrol.*, vol. 492, pp. 290–297, Jun. 2013, doi: 10.1016/j.jhydrol.2013.04.005.
- [39] R. Geigy, L. Jenni, M. Kauffmann, R. J. Onyango, and N. Weiss, 'Identification of T. brucei-subgroup strains isolated from game', *Acta Trop.*, vol. 32, no. 3, pp. 190–205, 1975.
- [40] M. E. Jensen, R. D. Burman, R. G. Allen, and American Society of Civil Engineers, Eds., *Evapotranspiration and irrigation water requirements: a manual*. New York, N.Y: The Society, 1990.
- [41] H. F. Blaney and 1892-, 'Determining water requirements in irrigated areas from climatological and irrigation data', 1952, Accessed: Jun. 24, 2021. [Online]. Available: <https://agris.fao.org/agris-search/search.do?recordID=US201300593774>
- [42] George H. Hargreaves and Zohrab A. Samani, 'Reference Crop Evapotranspiration from Temperature', *Appl. Eng. Agric.*, vol. 1, no. 2, pp. 96–99, 1985, doi: 10.13031/2013.26773.
- [43] J. DOORENBOS, 'Guidelines for predicting crop water requirements', *Food Agric. Organ. Rome Irrig Drain. Pap*, vol. 24, 1975, Accessed: Jun. 24, 2021. [Online]. Available: <https://ci.nii.ac.jp/naid/10012763449/>
- [44] M. E. Jensen and H. R. Haise, 'Estimating Evapotranspiration from Solar Radiation', *Proc. Am. Soc. Civ. Eng. J. Irrig. Drain. Div.*, vol. 89, pp. 15–41, 1963, Accessed: Jun. 24, 2021. [Online]. Available: <https://eprints.nwisrl.ars.usda.gov/id/eprint/1248/>
- [45] E. T. Linacre, 'A simple formula for estimating evaporation rates in various climates, using temperature data alone', *Agric. Meteorol.*, vol. 18, no. 6, pp. 409–424, Dec. 1977, doi: 10.1016/0002-1571(77)90007-3.
- [46] P. Mourão, 'Searching for Alternatives to Heparin Sulfated Fucans from Marine Invertebrates', *Trends Cardiovasc. Med.*, vol. 9, no. 8, pp. 225–232, Nov. 1999, doi: 10.1016/S1050-1738(00)00032-3.
- [47] S. Trajkovic, 'Temperature-Based Approaches for Estimating Reference Evapotranspiration', *J. Irrig. Drain. Eng.*, vol. 131, no. 4, pp. 316–323, Aug. 2005, doi: 10.1061/(ASCE)0733-9437(2005)131:4(316).
- [48] E. V. Smith, 'Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals', *J. Appl. Meas.*, vol. 3, no. 2, pp. 205–231, Jan. 2002.
- [49] I. A. Walter *et al.*, 'ASCE's Standardized Reference Evapotranspiration Equation', in *Watershed Management and Operations Management 2000*, Fort Collins, Colorado, United States, May 2001, pp. 1–11. doi: 10.1061/40499(2000)126.
- [50] G. Landeras, A. Ortiz-Barredo, and J. J. López, 'Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain)', *Agric. Water Manag.*, vol. 95, no. 5, pp. 553–565, May 2008, doi: 10.1016/j.agwat.2007.12.011.
- [51] R. C. Deo, P. Samui, and D. Kim, 'Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline

- models', *Stoch. Environ. Res. Risk Assess.*, vol. 30, no. 6, pp. 1769–1784, Aug. 2016, doi: 10.1007/s00477-015-1153-y.
- [52] V. Magliulo, R. d'Andria, and G. Rana, 'Use of the modified atmometer to estimate reference evapotranspiration in Mediterranean environments', *Agric. Water Manag.*, vol. 63, no. 1, pp. 1–14, Nov. 2003, doi: 10.1016/S0378-3774(03)00098-2.
- [53] S. Naoum and I. Tsanis, 'Hydroinformatics in evapotranspiration estimation', *Environ. Model. Softw.*, vol. 18, no. 3, pp. 261–271, Apr. 2003, doi: 10.1016/S1364-8152(02)00076-2.
- [54] F. Castellvi, C. O. Stockle, P. J. Perez, and M. Ibañez, 'Comparison of methods for applying the Priestley-Taylor equation at a regional scale: PRIESTLEY-TAYLOR EQUATION', *Hydrol. Process.*, vol. 15, no. 9, pp. 1609–1620, Jun. 2001, doi: 10.1002/hyp.227.
- [55] S. Alexandris and P. Kerkides, 'New empirical formula for hourly estimations of reference evapotranspiration', *Agric. Water Manag.*, vol. 60, no. 3, pp. 157–180, May 2003, doi: 10.1016/S0378-3774(02)00172-5.
- [56] A. Utset, I. Farré, A. Martínez-Cob, and J. Cavero, 'Comparing Penman–Monteith and Priestley–Taylor approaches as reference-evapotranspiration inputs for modeling maize water-use under Mediterranean conditions', *Agric. Water Manag.*, vol. 66, no. 3, pp. 205–219, May 2004, doi: 10.1016/j.agwat.2003.12.003.
- [57] A. R. Pereira and W. O. Pruitt, 'Adaptation of the Thornthwaite scheme for estimating daily reference evapotranspiration', *Agric. Water Manag.*, vol. 66, no. 3, pp. 251–257, May 2004, doi: 10.1016/j.agwat.2003.11.003.
- [58] S. M. Asokan, J. Jarsjö, and G. Destouni, 'Vapor flux by evapotranspiration: Effects of changes in climate, land use, and water use: VAPOR FLUX BY EVAPOTRANSPIRATION', *J. Geophys. Res. Atmospheres*, vol. 115, no. D24, Dec. 2010, doi: 10.1029/2010JD014417.
- [59] D. M. Amatya, R. W. Skaggs, and J. D. Gregory, 'Comparison of Methods for Estimating REF-ET', *J. Irrig. Drain. Eng.*, vol. 121, no. 6, pp. 427–435, Nov. 1995, doi: 10.1061/(ASCE)0733-9437(1995)121:6(427).
- [60] C. D. O. F. Silva and R. L. Manzione, 'MODELAGEM ESPACIAL DA EVAPOTRANSPIRAÇÃO EM AGROECOSSISTEMA DO SISTEMA AQUÍFERO BAURU', *Águas Subterrâneas*, 2018, Accessed: Jun. 24, 2021. [Online]. Available: <https://aguassubterraneas.abas.org/asubterraneas/article/view/29356>
- [61] N. Barnett, C. A. Madramootoo, and J. Perrone, 'Performance of some evapotranspiration', vol. 40, p. 7, 1998.
- [62] B. A. George, B. R. S. Reddy, N. S. Raghuvanshi, and W. W. Wallender, 'Decision Support System for Estimating Reference Evapotranspiration', *J. Irrig. Drain. Eng.*, vol. 128, no. 1, pp. 1–10, Feb. 2002, doi: 10.1061/(ASCE)0733-9437(2002)128:1(1).
- [63] G. Ravazzani, C. Corbari, S. Morella, P. Gianoli, and M. Mancini, 'Modified Hargreaves-Samani Equation for the Assessment of Reference Evapotranspiration in Alpine River Basins', *J. Irrig. Drain. Eng.*, vol. 138, no. 7, pp. 592–599, Jul. 2012, doi: 10.1061/(ASCE)IR.1943-4774.0000453.
- [64] U. Sharma and V. Dutta, 'Establishing environmental flows for intermittent tropical rivers: Why hydrological methods are not adequate?', *Int. J. Environ. Sci. Technol.*, vol. 17, no. 5, pp. 2949–2966, May 2020, doi: 10.1007/s13762-020-02680-6.
- [65] A. L. Lange, *Arthur L. Lange- Geophysical Studies at Kartchner Caverns State Park, Arizona. Journal of Cave and Karst Studies 61(2): 68-72. GEOPHYSICAL STUDIES AT KARTCHNER CAVERNS.*
- [66] M. Jing *et al.*, 'Improved regional-scale groundwater representation by the coupling of the mesoscale Hydrologic Model (mHM v5.7) to the groundwater model OpenGeoSys (OGS)', *Geosci. Model Dev.*, vol. 11, no. 5, pp. 1989–2007, Jun. 2018, doi: 10.5194/gmd-11-1989-2018.

- [67] M. Kumar, N. S. Raghuwanshi, R. Singh, W. W. Wallender, and W. O. Pruitt, 'Estimating Evapotranspiration using Artificial Neural Network', *J. Irrig. Drain. Eng.*, vol. 128, no. 4, pp. 224–233, Aug. 2002, doi: 10.1061/(ASCE)0733-9437(2002)128:4(224).
- [68] S. Rehana, 'River Water Temperature Modelling Under Climate Change Using Support Vector Regression', in *Hydrology in a Changing World*, S. K. Singh and C. T. Dhanya, Eds. Cham: Springer International Publishing, 2019, pp. 171–183. doi: 10.1007/978-3-030-02197-9_8.
- [69] M. Najafzadeh and G. Oliveto, 'Riprap incipient motion for overtopping flows with machine learning models', *J. Hydroinformatics*, vol. 22, no. 4, pp. 749–767, Jul. 2020, doi: 10.2166/hydro.2020.129.
- [70] S. K. Singh, R. Ibbitt, M. S. Srinivasan, and U. Shankar, 'Inter-comparison of experimental catchment data and hydrological modelling', *J. Hydrol.*, vol. 550, pp. 1–11, Jul. 2017, doi: 10.1016/j.jhydrol.2017.04.049.
- [71] A. Etemad-Shahidi, L. Bonakdar, and D.-S. Jeng, 'Estimation of scour depth around circular piers: applications of model tree', *J. Hydroinformatics*, vol. 17, no. 2, pp. 226–238, Mar. 2015, doi: 10.2166/hydro.2014.151.
- [72] M. K. Goyal, B. Bharti, J. Quilty, J. Adamowski, and A. Pandey, 'Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS', *Expert Syst. Appl.*, vol. 41, no. 11, pp. 5267–5276, Sep. 2014, doi: 10.1016/j.eswa.2014.02.047.
- [73] F. Granata, 'Evapotranspiration evaluation models based on machine learning algorithms—A comparative study', *Agric. Water Manag.*, vol. 217, pp. 303–315, May 2019, doi: 10.1016/j.agwat.2019.03.015.
- [74] F. Fotovatikhah, M. Herrera, S. Shamshirband, K. Chau, S. Faizollahzadeh Ardabili, and Md. J. Piran, 'Survey of computational intelligence as basis to big flood management: challenges, research directions and future work', *Eng. Appl. Comput. Fluid Mech.*, vol. 12, no. 1, pp. 411–437, Jan. 2018, doi: 10.1080/19942060.2018.1448896.
- [75] R. Taormina and K.-W. Chau, 'Data-driven input variable selection for rainfall–runoff modeling using binary-coded particle swarm optimization and Extreme Learning Machines', *J. Hydrol.*, vol. 529, pp. 1617–1632, Oct. 2015, doi: 10.1016/j.jhydrol.2015.08.022.
- [76] C. L. Wu and K. W. Chau, 'Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis', *J. Hydrol.*, vol. 399, no. 3–4, pp. 394–409, Mar. 2011, doi: 10.1016/j.jhydrol.2011.01.017.
- [77] T. Wu, W. Zhang, X. Jiao, W. Guo, and Y. A. Hamoud, 'Comparison of five Boosting-based models for estimating daily reference evapotranspiration with limited meteorological variables', *PLOS ONE*, vol. 15, no. 6, p. e0235324, Jun. 2020, doi: 10.1371/journal.pone.0235324.
- [78] M. A. Yassin, 'Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid climate', *Agric. Water Manag.*, vol. v. 163, pp. 110–124, Jan. 2016, doi: 10.1016/j.agwat.2015.09.009.
- [79] A. Antunes, A. Andrade-Campos, A. Sardinha-Lourenço, and M. S. Oliveira, 'Short-term water demand forecasting using machine learning techniques', *J. Hydroinformatics*, vol. 20, no. 6, pp. 1343–1366, Nov. 2018, doi: 10.2166/hydro.2018.163.
- [80] B. Pradhan, 'A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS', *Comput. Geosci.*, vol. 51, pp. 350–365, Feb. 2013, doi: 10.1016/j.cageo.2012.08.023.
- [81] L. Rutkowski, M. Jaworski, L. Pietruczuk, and P. Duda, 'The CART decision tree for mining data streams', *Inf. Sci.*, vol. 266, pp. 1–15, May 2014, doi: 10.1016/j.ins.2013.12.060.
- [82] P. Tsangaratos and I. Iliá, 'Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection, Greece', *Landslides*, vol. 13, no. 2, pp. 305–320, Apr. 2016, doi: 10.1007/s10346-015-0565-6.

- [83] B. Balk and K. Elder, 'Combining binary decision tree and geostatistical methods to estimate snow distribution in a mountain watershed', *Water Resour. Res.*, vol. 36, no. 1, pp. 13–26, Jan. 2000, doi: 10.1029/1999WR900251.
- [84] M. S. Tehrany, B. Pradhan, and M. N. Jebur, 'Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS', *J. Hydrol.*, vol. 504, pp. 69–79, Nov. 2013, doi: 10.1016/j.jhydrol.2013.09.034.
- [85] Z. Wang *et al.*, 'Spatiotemporal variability of reference evapotranspiration and contributing climatic factors in China during 1961–2013', *J. Hydrol.*, vol. 544, pp. 97–108, Jan. 2017, doi: 10.1016/j.jhydrol.2016.11.021.
- [86] E. Sadrfaridpour, T. Razzaghi, and I. Safro, 'Engineering fast multilevel support vector machines', *Mach. Learn.*, vol. 108, no. 11, pp. 1879–1917, Nov. 2019, doi: 10.1007/s10994-019-05800-7.
- [87] S. Salcedo-Sanz, J. L. Rojo-Álvarez, M. Martínez-Ramón, and G. Camps-Valls, 'Support vector machines in engineering: an overview: Support vector machines in engineering', *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 4, no. 3, pp. 234–267, May 2014, doi: 10.1002/widm.1125.
- [88] O. Kisi, 'Evapotranspiration modelling from climatic data using a neural computing technique', *Hydrol. Process.*, vol. 21, no. 14, pp. 1925–1934, Jul. 2007, doi: 10.1002/hyp.6403.
- [89] K. M. Andreadis, E. A. Clark, A. W. Wood, A. F. Hamlet, and D. P. Lettenmaier, 'Twentieth-Century Drought in the Conterminous United States', *J. Hydrometeorol.*, vol. 6, no. 6, pp. 985–1001, Dec. 2005, doi: 10.1175/JHM450.1.
- [90] M. Ehteram *et al.*, 'An improved model based on the support vector machine and cuckoo algorithm for simulating reference evapotranspiration', *PLOS ONE*, vol. 14, no. 5, p. e0217499, May 2019, doi: 10.1371/journal.pone.0217499.
- [91] V. K. Jain, R. P. Pandey, M. K. Jain, and H.-R. Byun, 'Comparison of drought indices for appraisal of drought characteristics in the Ken River Basin', *Weather Clim. Extrem.*, vol. 8, pp. 1–11, Jun. 2015, doi: 10.1016/j.wace.2015.05.002.
- [92] H. O. Odhiambo, C. K. Ong, J. D. Deans, J. Wilson, A. A. H. Khan, and J. I. Sprent, '[No title found]', *Plant Soil*, vol. 235, no. 2, pp. 221–233, 2001, doi: 10.1023/A:1011959805622.
- [93] P. Sonali and D. Nagesh Kumar, 'Spatio-temporal variability of temperature and potential evapotranspiration over India', *J. Water Clim. Change*, vol. 7, no. 4, pp. 810–822, Dec. 2016, doi: 10.2166/wcc.2016.230.
- [94] K. P. Sudheer, A. K. Gosain, and K. S. Ramasastri, 'Estimating Actual Evapotranspiration from Limited Climatic Data Using Neural Computing Technique', *J. Irrig. Drain. Eng.*, vol. 129, no. 3, pp. 214–218, Jun. 2003, doi: 10.1061/(ASCE)0733-9437(2003)129:3(214).
- [95] S. S. Zanetti, E. F. Sousa, V. P. Oliveira, F. T. Almeida, and S. Bernardo, 'Estimating Evapotranspiration Using Artificial Neural Network and Minimum Climatological Data', *J. Irrig. Drain. Eng.*, vol. 133, no. 2, pp. 83–89, Apr. 2007, doi: 10.1061/(ASCE)0733-9437(2007)133:2(83).
- [96] S. Chauhan and R. K. Shrivastava, 'Performance Evaluation of Reference Evapotranspiration Estimation Using Climate Based Methods and Artificial Neural Networks', *Water Resour. Manag.*, vol. 23, no. 5, pp. 825–837, Mar. 2009, doi: 10.1007/s11269-008-9301-5.
- [97] S. Suryavanshi, A. Pandey, U. C. Chaube, and N. Joshi, 'Long-term historic changes in climatic variables of Betwa Basin, India', *Theor. Appl. Climatol.*, vol. 117, no. 3–4, pp. 403–418, Aug. 2014, doi: 10.1007/s00704-013-1013-y.
- [98] A. Rahimikhoob, 'Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment', *Renew. Energy*, vol. 35, no. 9, pp. 2131–2135, Sep. 2010, doi: 10.1016/j.renene.2010.01.029.
- [99] S. Adamala, 'Temperature based generalized wavelet-neural network models to estimate evapotranspiration in India', *Inf. Process. Agric.*, vol. 5, no. 1, pp. 149–155, Mar. 2018, doi: 10.1016/j.inpa.2017.09.004.

- [100] H. Citakoglu, M. Cobaner, T. Haktanir, and O. Kisi, 'Estimation of Monthly Mean Reference Evapotranspiration in Turkey', *Water Resour. Manag.*, vol. 28, no. 1, pp. 99–113, Jan. 2014, doi: 10.1007/s11269-013-0474-1.
- [101] O. Kisi and H. Sanikhani, 'Prediction of long-term monthly precipitation using several soft computing methods without climatic data: PREDICTION OF LONG-TERM MONTHLY PRECIPITATION', *Int. J. Climatol.*, vol. 35, no. 14, pp. 4139–4150, Nov. 2015, doi: 10.1002/joc.4273.
- [102] A. P. Patil and P. C. Deka, 'An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs', *Comput. Electron. Agric.*, vol. 121, pp. 385–392, Feb. 2016, doi: 10.1016/j.compag.2016.01.016.
- [103] V. Z. Antonopoulos, 'Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables', *Comput. Electron. Agric.*, vol. v. 132, pp. 86–96, Jan. 2017, doi: 10.1016/j.compag.2016.11.011.
- [104] V. Nourani, G. Elkiran, and J. Abdullahi, 'Multi-station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements', *J. Hydrol.*, vol. 577, p. 123958, Oct. 2019, doi: 10.1016/j.jhydrol.2019.123958.
- [105] X. Dou and Y. Yang, 'Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems', *Comput. Electron. Agric.*, vol. 148, pp. 95–106, May 2018, doi: 10.1016/j.compag.2018.03.010.
- [106] L. B. Ferreira, F. F. da Cunha, R. A. de Oliveira, and E. I. Fernandes Filho, 'Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – A new approach', *J. Hydrol.*, vol. 572, pp. 556–570, May 2019, doi: 10.1016/j.jhydrol.2019.03.028.
- [107] M. Cobaner, 'Evapotranspiration estimation by two different neuro-fuzzy inference systems', *J. Hydrol.*, vol. 398, no. 3–4, pp. 292–302, Feb. 2011, doi: 10.1016/j.jhydrol.2010.12.030.
- [108] M. K. Saggi and S. Jain, 'Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning', *Comput. Electron. Agric.*, vol. 156, pp. 387–398, Jan. 2019, doi: 10.1016/j.compag.2018.11.031.
- [109] M. Pal and S. Deswal, 'M5 model tree based modelling of reference evapotranspiration', *Hydrol. Process.*, vol. 23, no. 10, pp. 1437–1443, May 2009, doi: 10.1002/hyp.7266.
- [110] S. Gu *et al.*, 'Characterizing evapotranspiration over a meadow ecosystem on the Qinghai-Tibetan Plateau', *J. Geophys. Res.*, vol. 113, no. D8, p. D08118, Apr. 2008, doi: 10.1029/2007JD009173.
- [111] K. Gudulas, K. Voudouris, G. Soulios, and G. Dimopoulos, 'Comparison of different methods to estimate actual evapotranspiration and hydrologic balance', *Desalination Water Treat.*, vol. 51, no. 13–15, pp. 2945–2954, Mar. 2013, doi: 10.1080/19443994.2012.748443.
- [112] G. Carrillo-Rojas, B. Silva, R. Rollenbeck, R. Céleri, and J. Bendix, 'The breathing of the Andean highlands: Net ecosystem exchange and evapotranspiration over the páramo of southern Ecuador', *Agric. For. Meteorol.*, vol. 265, pp. 30–47, Feb. 2019, doi: 10.1016/j.agrformet.2018.11.006.
- [113] H. Coners *et al.*, 'Evapotranspiration and water balance of high-elevation grassland on the Tibetan Plateau', *J. Hydrol.*, vol. 533, pp. 557–566, Feb. 2016, doi: 10.1016/j.jhydrol.2015.12.021.
- [114] B. R. Khan, M. Mainuddin, and M. N. Molla, 'Design, construction and testing of a lysimeter for a study of evapotranspiration of different crops', *Agric. Water Manag.*, vol. 23, no. 3, pp. 183–197, Jun. 1993, doi: 10.1016/0378-3774(93)90027-8.
- [115] J. A. Poss *et al.*, 'A volumetric lysimeter system (VLS): an alternative to weighing lysimeters for plant–water relations studies', *Comput. Electron. Agric.*, vol. 43, no. 1, pp. 55–68, Apr. 2004, doi: 10.1016/j.compag.2003.10.001.
- [116] M. I. Budyko, 'The Heat Balance of the Earth's Surface', *Sov. Geogr.*, vol. 2, no. 4, pp. 3–13, Jan. 1961, doi: 10.1080/00385417.1961.10770761.

- [117] P. C. D. Milly, 'Climate, soil water storage, and the average annual water balance', *Water Resour. Res.*, vol. 30, no. 7, pp. 2143–2156, Jul. 1994, doi: 10.1029/94WR00586.
- [118] G. Sireesha Naidu, M. Pratik, and S. Rehana, 'Modelling hydrological responses under climate change using machine learning algorithms – semi-arid river basin of peninsular India', *H2Open J.*, vol. 3, no. 1, pp. 481–498, Jan. 2020, doi: 10.2166/h2oj.2020.034.
- [119] A. Ochoa-Sánchez, P. Crespo, G. Carrillo-Rojas, A. Sucozhañay, and R. Célleri, 'Actual Evapotranspiration in the High Andean Grasslands: A Comparison of Measurement and Estimation Methods', *Front. Earth Sci.*, vol. 7, p. 55, Mar. 2019, doi: 10.3389/feart.2019.00055.
- [120] P. Droogers and R. G. Allen, 'Estimating Reference Evapotranspiration Under Inaccurate Data Conditions', *Irrig. Drain. Syst.*, vol. 16, no. 1, pp. 33–45, Feb. 2002, doi: 10.1023/A:1015508322413.
- [121] P. J. Slabbers, 'Practical prediction of actual evapotranspiration', *Irrig. Sci.*, vol. 1, no. 3, pp. 185–196, Feb. 1980, doi: 10.1007/BF00270883.
- [122] A. Poulouvassilis, M. Anadranistakis, A. Liakatas, S. Alexandris, and P. Kerkides, 'Semi-empirical approach for estimating actual evapotranspiration in Greece', *Agric. Water Manag.*, vol. 51, no. 2, pp. 143–152, Oct. 2001, doi: 10.1016/S0378-3774(01)00121-4.
- [123] T. J. Dolan, A. J. Hermann, S. E. Bayley, and J. Zoltek, 'Evapotranspiration of a Florida, U.S.A., freshwater wetland', *J. Hydrol.*, vol. 74, no. 3–4, pp. 355–371, Nov. 1984, doi: 10.1016/0022-1694(84)90024-6.
- [124] W. Koerselman and B. Beltman, 'Evapotranspiration from fens in relation to Penman's potential free water evaporation (EO) and pan evaporation', *Aquat. Bot.*, vol. 31, no. 3–4, pp. 307–320, Aug. 1988, doi: 10.1016/0304-3770(88)90019-8.
- [125] O. T. Denmead and R. H. Shaw, 'Availability of Soil Water to Plants as Affected by Soil Moisture Content and Meteorological Conditions¹', *Agron. J.*, vol. 54, no. 5, pp. 385–390, Sep. 1962, doi: 10.2134/agronj1962.00021962005400050005x.
- [126] A. M. Foyster, 'Application of the grid square technique to mapping of evapotranspiration', *J. Hydrol.*, vol. 19, no. 3, pp. 205–226, Jul. 1973, doi: 10.1016/0022-1694(73)90081-4.
- [127] V. R. N. Pauwels and R. Samson, 'Comparison of different methods to measure and model actual evapotranspiration rates for a wet sloping grassland', *Agric. Water Manag.*, vol. 82, no. 1, pp. 1–24, Apr. 2006, doi: 10.1016/j.agwat.2005.06.001.
- [128] H. Gavin and C. A. Agnew, 'Modelling actual, reference and equilibrium evaporation from a temperate wet grassland', *Hydrol. Process.*, vol. 18, no. 2, pp. 229–246, Feb. 2004, doi: 10.1002/hyp.1372.
- [129] H. K. McMillan, M. P. Clark, W. B. Bowden, M. Duncan, and R. A. Woods, 'Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure', *Hydrol. Process.*, vol. 25, no. 4, pp. 511–522, Feb. 2011, doi: 10.1002/hyp.7841.
- [130] H. Abdi and L. J. Williams, 'Principal component analysis: Principal component analysis', *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 4, pp. 433–459, Jul. 2010, doi: 10.1002/wics.101.
- [131] G. E. Hinton, 'Reducing the Dimensionality of Data with Neural Networks', *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006, doi: 10.1126/science.1127647.
- [132] J. Jarsjoe, Y. Shibuo, and G. Destouni, 'Using the PCRaster-POLFLOW approach to GIS-based modelling of coupled groundwater-surface water hydrology in the Forsmark Area', Sep. 2004, Accessed: Jul. 16, 2021. [Online]. Available: <https://www.osti.gov/etdeweb/biblio/20564022>
- [133] Y. Shibuo, J. Jarsjö, and G. Destouni, 'Hydrological responses to climate change and irrigation in the Aral Sea drainage basin', *Geophys. Res. Lett.*, vol. 34, no. 21, p. L21406, Nov. 2007, doi: 10.1029/2007GL031465.