Crop Evapotranspiration Prediction using Machine Learning

Project-I (AG47005) report submitted to

Indian Institute of Technology Kharagpur

in partial fulfilment for the award of the degree of

Integrated Dual Degree

in

Agricultural and Food Engineering

by

Abhirama Gorti

(20AG38002)

Under the supervision of

Dr. Damodhara Rao Mailapalli

Department of Agricultural and Food Engineering

Indian Institute of Technology Kharagpur

Autumn Semester, 2023-24

November 28, 2023

DECLARATION

I certify that

(a) The work contained in this report has been done by me under the guidance of my supervisor.

(b) The work has not been submitted to any other Institute for any degree or diploma.

(c) I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

(d) Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Date: November 28, 2023 (Abhirama Gorti)
Place: Kharagpur (20AG38002)

DEPARTMENT OF AGRICULTURAL AND FOOD ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

KHARAGPUR - 721302, INDIA

CERTIFICATE

This is to certify that the project report entitled "**Crop Evapotranspiration Prediction using Machine Learning**" submitted by **Abhirama Gorti** (Roll No. 20AG38002) to Indian Institute of Technology Kharagpur towards partial fulfilment of requirements for the award of degree of Integrated Dual Degree in Agricultural and Food Engineering is a record of bona fide work carried out by him under my supervision and guidance during Autumn Semester, 2023-24.

Date: November 28, 2023

Place: Kharagpur

Dr. Damodhara Rao Mailapalli
Department of Agricultural and Food Engineering
Indian Institute of Technology Kharagpur
Kharagpur - 721302, India

Abstract

Name of the student: **Abhirama Gorti** Roll No: **20AG38002**

Degree for which submitted: **Integrated Dual Degree**

Department: **Department of Agricultural and Food Engineering**

Thesis title: **Crop Evapotranspiration Prediction using Machine Learning**

Thesis supervisor: **Dr. Damodhara Rao Mailapalli**

Month and year of thesis submission: **November 28, 2023**

Crop evapotranspiration (ETc) is the combined process of evaporation from the soil surface and transpiration from plants. It is a key component of the agricultural water balance, as it represents the amount of water that is lost from the system to the atmosphere. Accurate prediction of ETc is essential for irrigation scheduling, which is the process of applying water to crops in the right amount at the right time to ensure optimal growth and yield.

Traditional methods of ETc prediction rely on empirical equations and crop coefficients. The problem can be broken down to predicting Reference Evapotranspiration(ETo). Machine learning (ML) models have the potential to improve the accuracy of ETo prediction by forming a black-box model that helps connect different features for predicting the target variable.

In this bachelor thesis, I developed a ML model to predict ETo using meteorological variables. We trained and evaluated several ML models such as SVMs and Random Forests. They are also refered to as universal predictors for their robust performance in predictions without any implicit changes in the model structure. Followed by this,

the recent trend of Deep Learning models have also inspired me to include a few of the architectures such as such as feedforward neural networks (FFNs), convolutional neural networks (CNNs), and long short-term memory (LSTM) networks. %

I plan to integrate our ML model into an irrigation scheduling application. This application will use ETc predictions to recommend irrigation schedules to farmers, based on their crop type, soil type, and climatic conditions. The application will also take into account the farmer's water availability and budget.

The integration of the ML model into an irrigation scheduling application has the potential to revolutionize the way that farmers manage their water resources. The application will be able to provide farmers with accurate and personalized irrigation recommendations, based on their specific crop type, soil type, and climatic conditions. This will help farmers to improve water use efficiency, increase crop yields, and reduce environmental impact.

Acknowledgements

I express my profound gratitude to all those who have contributed to the completion of this thesis. First and foremost, I would like to extend my heartfelt thanks to my advisor, Dr. Damodhara Rao Mailapalli, for his unwavering support, insightful guidance, and invaluable expertise throughout this research journey. His dedication and encouragement played a pivotal role in shaping the direction of this study.

I am deeply thankful to the faculty members of the Indian Institute of Technology, Kharagpur, whose knowledge and feedback enriched my understanding and inspired me to delve deeper into the intricate domain of evapotranspiration prediction.

My sincere appreciation goes to my colleagues and friends who provided endless encouragement, engaging discussions, and moral support. Your camaraderie made this academic endeavor a truly enriching experience.

I am indebted to my family for their unconditional love, understanding, and encouragement. Their unwavering belief in my abilities kept me motivated during challenging times. Ramesh lanja

Lastly, I extend my gratitude to the academic community, researchers, and authors whose work laid the foundation for this study. Your contributions have been invaluable in shaping my perspective and enhancing the quality of this research.

This thesis stands as a testament to the collective support and inspiration I have received, and I am profoundly grateful for each individual who has been part of this academic endeavor.

Abhirama Gorti
November, 2023
Indian Institute of Technology, Kharagpur

Contents

List of Figures

List of Tables

Abbreviations

FEA	**F**inite **E**lement **A**nalysis
FEM	**F**inite **E**lement **M**ethod
LVDT	**L**inear **V**ariable **D**ifferential **T**ransformer
RC	**R**einforced **C**oncrete

Symbols

D^{el}	elasticity tensor
σ	stress tensor
ε	strain tensor

Chapter 1

Introduction

1.1 Background and Motivation

1.1.1 Introduction to Evapotranspiration

Evapotranspiration (ET) is one of the major components of the hydrologic cycle and affects crop water demand. Therefore, its quantification is necessary for proper irrigation planning. The term evapotranspiration refers to the combination of two processes, namely, evaporation and transpiration. Evaporation is a process by which water is lost in the form of vapor from natural surfaces, such as freewater surface, bare soil, from live or dead vegetation. Transpiration is a process by which water is lost in the form of vapour through plant leaves. Therefore, evapotranspiration is a combined loss of water from the soil (evaporation) and plant (transpiration) surfaces to the atmosphere through vaporization of liquid water, and is expressed in depth per unit time (for example mm/day).

1.1.2 Origins of the term

From early times, many scientists have been trying to define Evapotranspiration. Earliest definitions of the term were coined by Howard Penman in the Regarding the Potential Evapotranspiration. He defined it as "the amount of water transpired in a given time by a short green crop, completely shading the ground, of uniform height and with adequate water status in the soil profile".

Here, it does not refer to any specific crop which creates a confusion on Short Green Crop. So, scientists may be confused as to which crop should be selected to be used as a short green crop because the evapotranspiration rates from well-watered crops may be as much as 10 to 30% greater than that occurring from a short green grass.

Allan et al. (1998) modified the definition and replaced it with the term "Reference Evapotranspiration" and is defined as: The rate of evapotranspiration from a hypothetical reference crop with an assumed crop height of 0.12 m (4.72 in), a fixed surface resistance of 70 secm^{-1} and an albedo of 0.23, closely resembling the evapotranspiration from an extensive surface of green grass of uniform height, actively growing, well-watered, and completely shading the ground. The reference crop is short green grass with active water supply and is denoted by the term ETo.

1.2 Research Objectives

In the field of physical sciences, predicting physical variables involves complex relationships with various factors, making it a challenging task. Traditional methods for predicting these variables often rely on empirical equations and coefficients, which may not capture the full complexity of the system

This research aims to investigate the potential of machine learning (ML) and deep learning (DL) algorithms in predicting evapotranspiration (ET) across India. The specific research objectives are:

- To collect meteorological data of various locations across India and identify the most influential features for ET prediction.
- To train and test different ML models, including Support Vector Regression (SVR), Random Forest (RF), and Artificial Neural Networks (ANNs), to accurately predict ET using the most influential meteorological features.
- To compare the performance of the developed ML-based ET prediction models with the FAO-PM 56 ET model and validate them using measured ET data.

1.3 Scope and Limitations

As we are using Machine Learning models for prediction, we are confined with the physical context of the prediction, i.e. Evapotranspiration might be a factor of many physical processes in a catchment and we choose to focus on meteorological which are the most contributing thus bringing in Structural Uncertainty. Hence as far as model is considered we are assuming that other physical characteristics measured at the sites from which data is extracted are more or less similar.

This also brings in the limitation to our training, i.e. the data on which our model is trained is obtained from standard fields across different locations. Hence, a certain extent of ideal conditions is observed at the test sites, but differences may be observed if the test site conditions are drastically changes. This topic would be further discussed in future sections of Data collection and processing.

Since, Machine Learning models are nothing more than rough mathematical approximators the errors obtained in them for training and validation data are explained in great detail. Sometimes, the model's predictions may not follow the same trajectory due to systematic and non-systematic errors and might be prone to hallucinations, which is out of the scope of this project.

Chapter 2

Literature Review

2.1 ETo empirical models

We start by picking one of the most important papers of all times in the history of Evapotranspiration modelling which is Allan et al. (1998). It provides detailed guidelines for computing crop water requirements, focusing on the concepts of evapotranspiration. The paper introduces the need for a standardized method to compute reference evapotranspiration (ETo) from meteorological data. It recommends the FAO Penman-Monteith method as the sole ETo method for determining reference evapotranspiration. Equation 3.1 gives the structure of the equation.

$$ET_o = \frac{0.408\Delta(R_n - G) + \gamma U_2\left(\frac{900}{T+273}\right)}{\Delta + \gamma(1 + 0.34U_2)} \tag{2.1}$$

ET_o : Reference evapotranspiration (mm/day)

Δ : Slope of the saturation vapor pressure function (kPa (°C)-1)

R_n : Net solar radiations (MJ m-2 day-1)

G : Earth heat flux thickness (MJ m-2 day-1)

T : Average atmospheric temperature

γ : Psychrometric constant (kPa °C-1)

U_2 : Wind speed at 2m height (m s-1)

Similarly, there are other empirical models such as Hargreaves Equation 2.2 by Hargreaves and Samani (1985), Thortwaite Equation 2.3 by Thornthwaite (1948) and FAO Blaney-Criddle Equation 2.4 by Blaney et al. (1952)

$$ET_o = 0.0023 \times (T_{avg} + 17.8) \times (T_{max} - T_{min})^{0.5} \times Ra \tag{2.2}$$

ET_o : Reference evapotranspiration (mm/day)

T_{avg} : Average daily temperature (°C)

T_{max} : Maximum daily temperature (°C)

T_{min} : Minimum daily temperature (°C)

Ra : Extraterrestrial Radiation (mm/day)

$$PET = 16 \times \left(\frac{10 \times T_{avg}}{I}\right)^a \times \left(\frac{N}{12}\right) \times \left(\frac{d}{30}\right) \tag{2.3}$$

PET : Potential evapotranspiration (mm/month)

T_{avg} : Average daily temperature (°C)

I : Heat index which depends on the 12 monthly mean temperatures

a : Empirical constant

N : Average day length (hours) of the month being calculated

d : Number of days in a month

$$ET_o = p \times (0.457 \times T_{mean} + 8.128) \tag{2.4}$$

ET_o : Reference evapotranspiration (mm/day)

T_{mean} : Mean daily temperature (°C)

p : Mean daily percentage of annual daytime hours

2.2 ML and DL models for ETo

The usage of artificial intelligence models picked up in the 2000s. One of the earliest papers include Kumar et al. (2002) which uses the vanilla neural network with 1 hidden layer of 7 neurons, 6 input features and a regression output layer. The model has a WSEE (Weighted Standard Error of estimate) 0.3mm/day.

Later as the computational power rose, some of the more common Machine Learning models known today such as Support Vector Machine, Random Forest picked up as explained in one of the recent review papers. After KIŞI and ÇIMEN (2009) there were many models developed using traditional ML models and tested for different regions across the world.

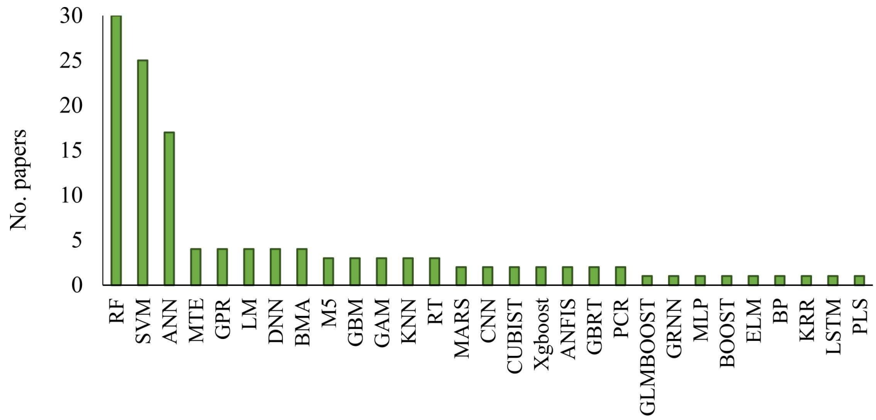

FIGURE 2.1: Machine learning models for ET estimation using remote sensing data from Amani and Shafizadeh-Moghadam (2023)

Similarly, there are multiple papers that attempt to use different machine learning papers like Pagano et al. (2023) using Multi Layered Perceptron (MLP), Adamala et al. (2014) using Second order neural network, Ravindran et al. (2021) using Random Forests and XGBoost to find the most important feature and pass it through a neural network for ETo prediction, etc.

2.3 Research Gaps

While there have been many models developed, all of them face some similar issues or challenges. Some of them include but not restricted to:

- **Data aquisition and scarcity**: Existing methods for predicting evapotranspiration require extensive parametrization and may lack relevant data in some regions. This can limit the accuracy and applicability of the models.
- **Reliable Results**: The results obtained from these models are hyper-local to the training data partly acounting for the reason that the relation between physical variables is dependent upon the global position.

Chapter 3

Materials and Methods

3.1 Data Collection and Preprocessing

3.1.1 Study Area and Climate Data

For the purpose of this study, 25 different meteorological locations in India were chosen. The data for this study was collected from the All India Coordinated Research Project on Agrometeorology (AICRPAM), CRIDA, Hyderabad, Telangana, India. These locations are having daily meteorological data of five years (2001–05) of variables such as minimum temperature (T_{min}), maximum temperature (T_{max}), minimum relative humidity (RH_{min}),maximum relative humidity (RH_{max}), mean wind speed (W_s), Incident solar radiation (R_s) and Sunshine hours (n).

3.1 shows the locations of selected sites,whereas presents information related to altitude, latitude,longitude, and mean climatic characteristics of the chosen sites. The altitude of selected sites varied from 10 m at Mohanpur to 1600 m at Ranichauri above mean sea level. The mean T_{min} and T_{max} range from 9.66 to 23.38 °C and 20.08 to 35.11 °C,respectively. The mean RH_{min} and RH_{max} range from 33.91 to 75.27% and 69.70 to 96.18%, respectively. The mean wind speed and incident

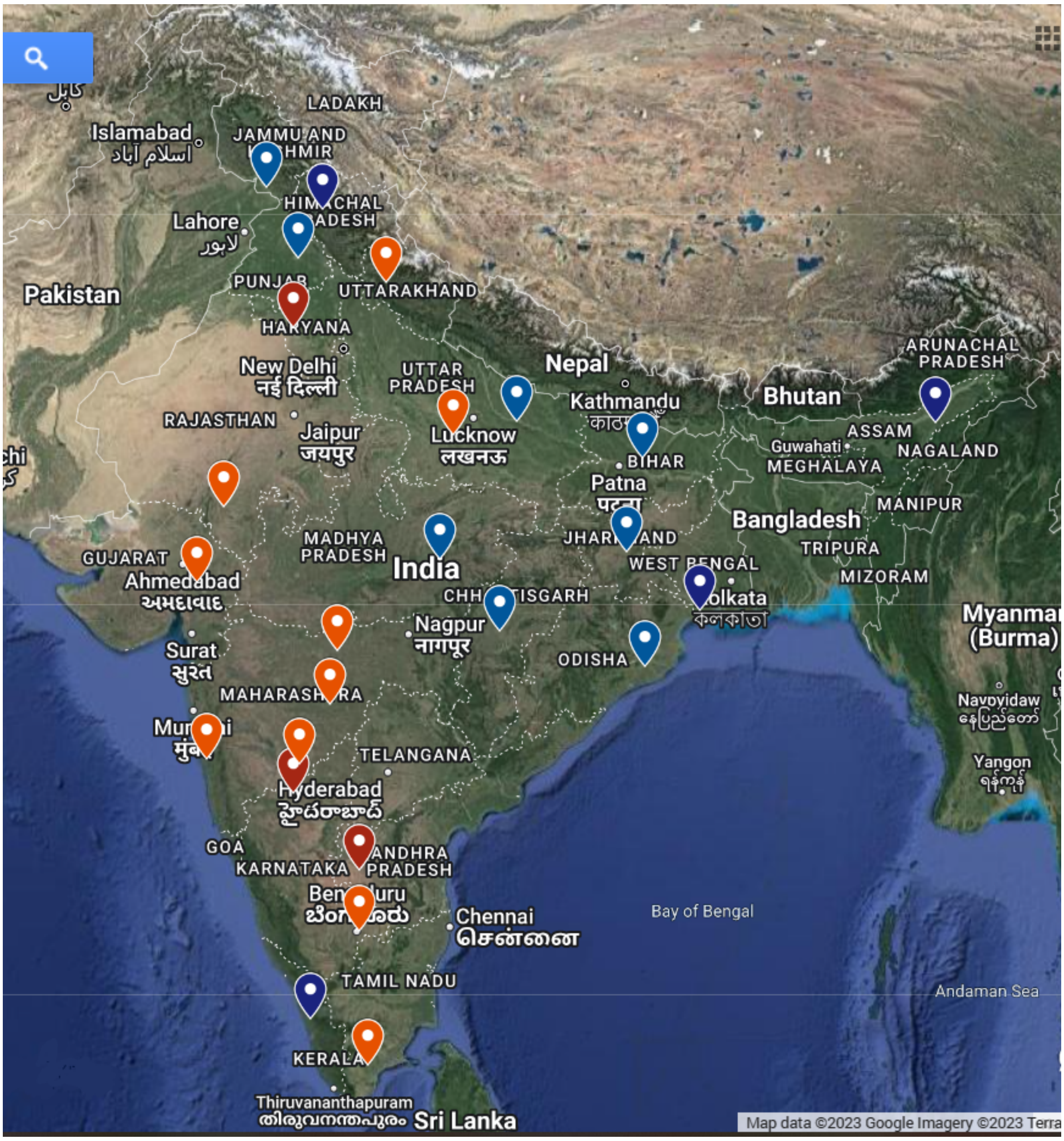

FIGURE 3.1: 25 stations across India with stations marked according to the type of climate prevalent

solar radiation ranged from 1.27 to 9.64 kmh^{-1} and 14.68 to 20.87 $MJm^{-2}day^{-1}$, respectively. The climate in the selected locations of the study area is classified as Semiarid, Arid, Subhumid, and Humid as marked on the map with different colored markers. explained in 3.1.

The data used for validating the results are obtained from two different sources. One being IIT Kharagpur Agricultural Field's Metereological Station from which data for the last 5 years, i.e. 2018 January to 2023 April for 1499 days has been collected. The data has been cleaned thoroughly and useful variables needed for ETo regression have been separated. Apart from the basic weather features, there are other features such as amount of evaporation on a daily basis.

The second source is the UC Davis Standard Calibrated Lysimeter Data from the Campbell Tract. Davis Lysimeters are large pans of soil (6 meters across and 1 meter deep) that are set on a weighing device. The basic utility is to be able to very accurately measure mass flux into and out of the lysimeter, thus directly measure precipitation, irrigation and evapotranspiration at the field scale. The lysimeters were built in the late 50's. The data available to use ranges from 1959 July to June 1963, i.e. 1438 days in total. The different types of data available from the sources were Evapotranspiration in inches, Evaporation, Radiation and Climatic Data.

Station	**Climate**	**Lat.**	**Long.**	**Alt.**	**Period**	**Tmax**	**Tmin**	**RHmax**	**RHmin**	**U2**	**Rs**	**n**
Anantapur	Arid	14°41’	77°37’	350	2001-2005	34.68	21.91	72.18	33.07	9.81	1.39	8.08
Bijapur	Arid	16°49’	75°43’	594	2001-2005	32.76	19.30	77.60	43.91	5.86	1.48	7.01
Hissar	Arid	29°10’	75°44’	215	2001-2005	31.38	16.23	80.47	43.16	5.34	1.31	7.33
Jorhat	Humid	26°47’	94°12’	86	2001-2005	28.58	18.68	92.20	71.51	2.30	4.68	5.06
Mohanpur	Humid	21°52’	87°26’	10	2001-2005	31.94	20.88	96.49	62.64	1.53	4.12	7.06
Palampur	Humid	32°06’	76°03’	1291	2001-2005	24.53	13.31	65.43	53.20	5.46	5.34	7.00
Thrissur	Humid	10°31’	76°13’	26	2001-2005	31.97	23.39	85.11	59.83	4.98	6.87	6.22
Akola	Semi-arid	20°42’	77°02’	282	2001-2003	33.93	19.88	64.27	37.15	8.01	1.59	6.87
Anand	Semi-arid	22°33’	72°58’	45	2001-2004	33.66	19.73	79.74	42.23	3.62	2.73	8.36
Bangalore	Semi-arid	12°58’	77°35’	930	2001-2005	29.13	18.03	87.60	47.31	8.48	2.53	6.92
Dapoli	Semi-arid	17°46’	73°12’	250	2001-2005	31.09	19.17	93.56	69.99	4.99	8.45	6.53
Kanpur	Semi-arid	26°26’	80°22’	126	2004-2005	31.50	19.04	79.59	54.08	5.60	2.01	6.36
Kovilpatti	Semi-arid	9°10’	77°52’	90	2001-2005	35.16	22.21	78.52	45.90	6.73	1.91	7.09
Parbhani	Semi-arid	19°08’	76°50’	423	2001-2005	33.70	18.40	70.59	40.02	5.09	2.53	8.76
Ranichauri	Semi-arid	30°52’	78°02’	1600	2001-2005	19.88	9.49	79.53	60.64	4.98	3.05	6.57
Solapur	Semi-arid	17°41’	75°56’	25	2001-2005	34.53	20.45	72.55	43.95	6.00	1.66	7.35
Udaipur	Semi-arid	25°21’	74°38’	433	2001-2005	31.71	16.51	70.24	36.54	4.02	1.57	8.28
Bhubaneswar	Sub-humid	20°15’	85°50’	25	2002-2005	32.83	22.23	91.41	58.39	6.53	4.25	6.98
Faizabad	Sub-humid	26°47’	82°08’	133	2001-2005	31.43	18.29	85.03	53.47	3.46	2.48	6.97
Jabalpur	Sub-humid	23°09’	79°58’	393	2002-2005	31.45	18.57	77.71	42.77	3.78	4.42	6.81
Ludhiana	Sub-humid	30°56’	75°52’	247	2001-2005	30.06	17.42	83.97	49.14	4.26	1.84	7.82
Raipur	Sub-humid	21°14’	81°39’	298	2001-2005	32.85	20.08	79.39	43.91	5.45	3.24	6.82
Rakh Dhiansar	Sub-humid	32°39’	74°58’	332	2005-2005	29.10	15.40	83.46	50.02	1.76	2.25	6.28
Ranchi	Sub-humid	23°17’	85°19’	625	2005-2005	29.73	17.08	88.09	53.92	1.86	2.98	6.97
Samastipur	Sub-humid	25°53’	85°48’	52	2004-2005	30.53	19.99	84.67	56.12	4.66	2.33	6.92

TABLE 3.1: Climate Data of 25 stations across India

3.2 FAO-56 Penman Monteith Equation

FAO-56 PM Method is recommended as the standard method for estimating ETo when other standard measuring equipments like Lysimeter are not available. 3.1 gives the equation for calculation of daily ETo.

$$ET_o = \frac{0.408\Delta(R_n - G) + \gamma U_2 \left(\frac{900}{T+273}\right)(e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \tag{3.1}$$

where ETo is the Reference evapotranspiration (mm/day) Δ is the Slope of the saturation vapor pressure function (kPa (°C)-1), R_n is Net solar radiations (MJ m-2 day-1), G is the Earth heat flux thickness (MJ m-2 day-1), T is the Average atmospheric temperature, γ is the Psychrometric constant (kPa °C-1), U_2 is the Wind speed at 2m height (m s-1), e_s is the saturation vapour pressure (kPa) and e_a is the actual vapour pressure (kPa).

I have used this equation as my reference to count the different features for passing it through the Machine and Deep Learning models with ETo as the regressor variable. 3.2 depicts the different terms being used in different parts of the equation as described from Allan et al. (1998).

The colored and highlighted boxes indicate the key variable needed to formulate the structure of the equation. There are 7 key variables in total, they are J - Julian Day, Φ - Latitude of the test site (rad), T - Daily Mean Air Temperature (°C), n - number of sunshine hours. z - Elevation of the test site (m), U_2 - Wind Speed at 2m above the test site (m/s) and T_D - Daily Mean Air Dew Point Temperature (°C).

The constants inlcude λ - Latent Heat of vaporization = 2.45 MJ kg^{-1}, α - Albedo or Crop Canopy Coefficient = 0.23, σ - Stefan / Boltmann Constant = 4.903×10^{-9} MJK^{-4}m$-$2day$-$1 and G - Solar Constant = 0.082 MJm^{-2}min^{-1}.

Functional variables like Solar Declination, Sunset hour angle, Clear Sky Radiation, etc. are calculated from these values and variables.

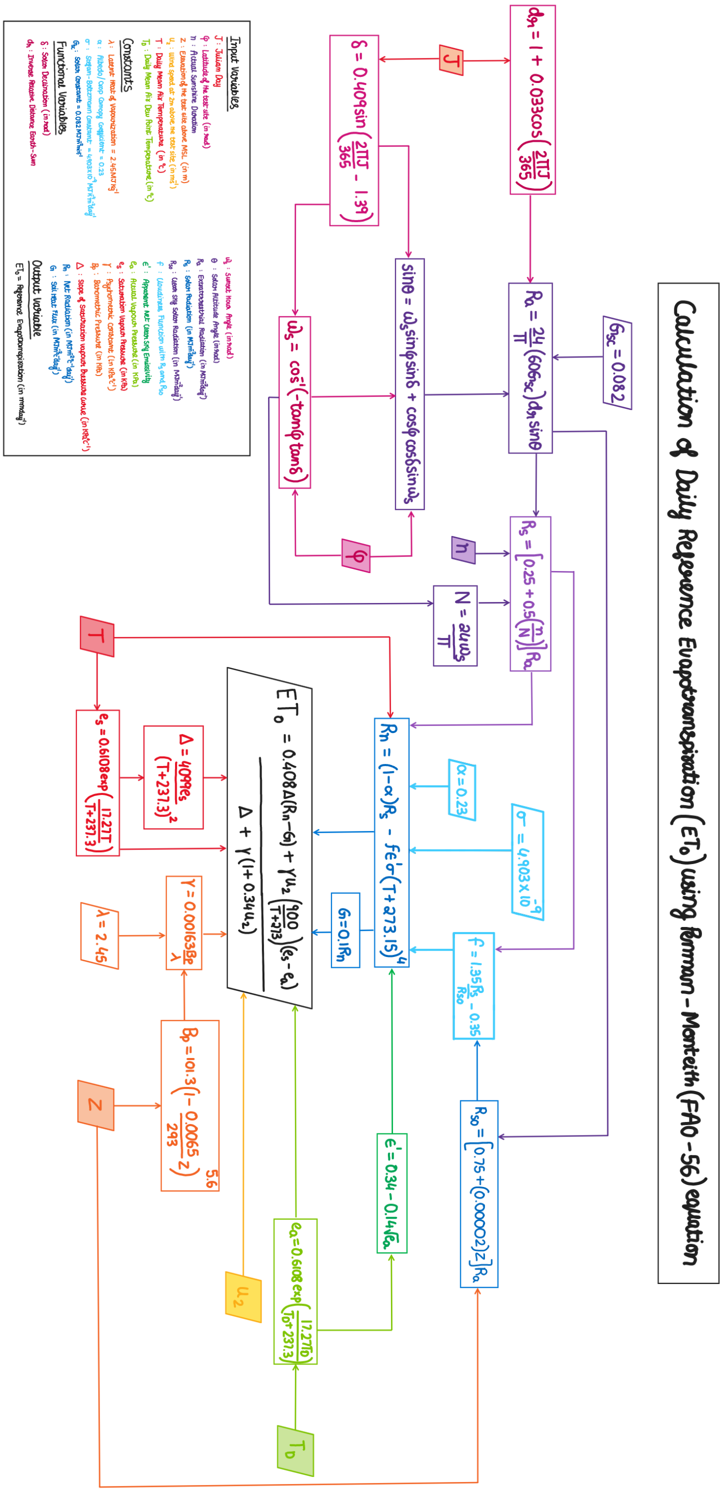

FIGURE 3.2: Flowchart of decomposition of the Penman Monteith Equation

Appendix A

Appendix A

Write your Appendix content here.

Bibliography

Adamala, S., Raghuwanshi, N. S., Mishra, A., and Tiwari, M. K. (2014). Evapotranspiration modeling using second-order neural networks. *Journal of Hydrologic Engineering*, 19(6):1131–1140.

Allan, R., Pereira, L., and Smith, M. (1998). *Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56*, volume 56.

Amani, S. and Shafizadeh-Moghadam, H. (2023). A review of machine learning models and influential factors for estimating evapotranspiration using remote sensing and ground-based data. *Agricultural Water Management*, 284:108324.

Blaney, H. F. et al. (1952). Determining water requirements in irrigated areas from climatological and irrigation data.

Hargreaves, G. H. and Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Applied engineering in agriculture*, 1(2):96–99.

KIŞI, O. and ÇIMEN, M. (2009). Evapotranspiration modelling using support vector machines / modélisation de l'évapotranspiration à l'aide de 'support vector machines'. *Hydrological Sciences Journal*, 54(5):918–928.

Kumar, M., Raghuwanshi, N. S., Singh, R., Wallender, W. W., and Pruitt, W. O. (2002). Estimating evapotranspiration using artificial neural network. *Journal of Irrigation and Drainage Engineering*, 128(4):224–233.

Pagano, A., Amato, F., Ippolito, M., De Caro, D., Croce, D., Motisi, A., Provenzano, G., and Tinnirello, I. (2023). Machine learning models to predict daily actual evapotranspiration of citrus orchards under regulated deficit irrigation. *Ecological Informatics*, 76:102133.

Ravindran, S. M., Bhaskaran, S. K. M., and Ambat, S. K. N. (2021). A deep neural network architecture to model reference evapotranspiration using a single input meteorological parameter. *Environmental Processes*, 8(4):1567–1599.

Thornthwaite, C. W. (1948). An approach toward a rational classification of climate. *Geographical review*, 38(1):55–94.