www.cafetinnova.org

Indexed in
Scopus Compendex and Geobase Elsevier, Chemical Abstract Services-USA, Geo-Ref Information Services-USA, List B of Scientific Journals, Poland, Directory of Research Journals

ISSN 0974-5904, Volume 08, No. 02

International Journal of Earth Sciences and Engineering

April 2015, P.P.491-496

Evapotranspiration Modeling Using M5 Model Tree

NITIN SAWALKAR AND PRADNYA DIXIT
VIIT, Pune, Maharashtra, INDIA
Email: *sawalkant@gmail.com, prdxt11@gmail.com*

Abstract: Water is vital to life and development in all parts of the world. Where the agricultural sector plays an important role in economic growth of the country, the management of water resources is an item of high priorities in their developmental activities. Evaporation and evapotranspiration are two major processes involved in crop water determination to know the exact demand and supply of the water in the irrigation projects. The combination of two separate processes whereby water is lost through evaporation from the soil surface and in the next from crop-grass by transpiration is referred to as Evapotranspiration (ET). The term consumptive use is also to denote this loss by evapotranspiration. Traditionally Evapotranspiration is determined using varying methods including theoretical and empirical techniques as well as by lysimeter observations. Practically, it is not possible to install the lysimeter at each required location and the approach to empirical methods is data intensive. Therefore an alternative approach which can make use of available data and estimate evapotranspiration with reasonable accuracy is in need. The present study focuses on the calculation of evapotranspiration by a data driven technique that is Model Tree by making use of the measured metrological parameters (from 2001 to 2005) for three USGS Stations at Florida, USA and results are compared with traditional empirical method (Penman Monteith Method). The performance of estimated values is judged by the correlation coefficient between estimated (Penman Monteith and Model Tree) and actual measured evapotranspiration along with other error measures like Root Mean Square Error and Mean Absolute Error functions. The dataset includes the daily measured climatological parameters such as Solar Radiation, Air Humidity (Maximum and Minimum), Air Temperature (Maximum, Minimum and Mean) and Mean Wind Speed for five years as mentioned earlier.

Keywords: *Evapotranspiration, Evaporation, Transpiration, Model Tree (MT)*

1. Introduction

Estimation of evapotranspiration plays a key role in managing water resources and in Hydrologic Models. The combination of two separate processes whereby water is lost from the soil surface by evaporation and from grass by transpiration is referred to as evapotranspiration (ET) [1]. Thus it can be said that Evapotranspiration is nothing but the combination of Evaporation and Transpiration. For a given set of atmospheric conditions, evapotranspiration obviously depends on the availability of water. If sufficient moisture is available to meet the needs of vegetation fully covering the area, the resulting evapotranspiration is called as “Potential Evapotranspiration” (PET). Potential Evapotranspiration is no longer critically depending on the soil and plant factors but depends essentially on the climatic factors [8]. The real evapotranspiration occurring in a specific situation is called “Actual Evapotranspiration” (AET) and from reference surface it is referred as “Reference Evapotranspiration” (RET).

Traditionally evapotranspiration is measured using (i) Lysimeters and (ii) Field Plots [8]. It can also be determined by using (i) Empirical Relationship- Penman Monteith (ii) Radiation Based Method- Makkink and Priestley-Taylor, (iii) Temperature Based Method- Blaney-Criddle Method [2]. It is observed that empirical equations are data intensive, they require measurement of many meteorological variables like heat radiations, actual duration of sunshine in hours, maximum possible sunshine in hours, mean air temperature, actual vapor pressure, mean wind speed, saturation vapor pressure at mean air temperature and thus can be used only when required parameters are available. Lysimeter and Field Plots are the widely used instruments today and their applications in hydrological design and operation are long standing.

If the data which is compulsorily required in the empirical equation is not available for particular site, it is not possible to estimate evapotranspiration by using empirical methods. Although it is possible to measure evapotranspiration using Lysimeter, it is hard to place the setup at every place where there is planned or an existing hydrological models and irrigation projects. Also it is difficult to place and maintain accurate measuring devices at remote locations. It is therefore necessary to try alterative techniques to estimate evapotranspiration with reasonable accuracy and avoiding excessive data measurement.

Since last two decades many researchers have been using different soft computing techniques for various purposes, few of them are enlisted here ; Artificial Neural Network (ANN), Genetic Programming (GP) [3], Model Tree (MT) [4]. For the present study, a data driven technique; Model Tree is employed to estimate ET at one of the USGS station [5] in Florida USA by using measured data at the same location which facilitates the comparison of models with different input parameters by Model Tree and Penman Monteith method.

The aim of this paper is to estimate the evapotranspiration with number of parameters (which are needed in Penman Monteith) by using Model Tree. Model Tree is a Data Driven technique which is totally based on the mathematical approach and uses the linear equation and calculates the evapotranspiration. To achieve this, Evapotranspiration has been calculated by using daily measured meteorological data from 2000 to 2004 (Five Years) at Florida (USGS Station). The FAO Penman-Monteith method is recommended as the sole method for determining Reference Evapotranspiration (RET) [1]. Method used for calculation of ET by Empirical Method, its formulation, process followed along with the development of model to calculate ET using Model Tree and the comparison between these is organized as follows.

2. Study Area and Data Used

The present study is done at three stations of United States of Geological Survey (USGS), Florida USA. The detailed information of USGS Station 1, 2 and 3 is as given in Table 1.

Table1: *Study area location and other details*

Station	USGS Station No.	Location	Latitude	Longitude
USGS Station 1	274143080424100	St John’s Rv Marsh at Blue Cypress Nr Fellsmere, Florida, USA	27°41'43.0" N	80°42'41.0" E
USGS Station 2	281626080463400	Duda Farms Nr Lake Winder Nr Rockledge, Florida, USA	28°16'26.0” N	80°46'34.0" E
USGS Station 3	284541082163400	Ferris Farms Nr Floral City, Florida, USA	28°45'40.7”N	82°16'34.4" E

2.1. Data Used

The data has been collected from the above mentioned three USGS stations. The daily data is available for five years (from year 2000 to 2004) for each station which includes Meteorological and Atmospheric Parameters.

2.1.1. Meteorological Data

The methods for calculating evapotranspiration from meteorological data require various climatological and physical parameters. Some of the data are measured directly in weather stations. Other parameters are related to commonly measured data and can be derived with the help of a direct empirical relationship.

Meteorological factors to determine ET;

1. Solar Radiation
2. Air Temperature
3. Air Humidity
4. Wind Speed

Atmospheric Parameters;

1. Atmospheric Pressure (P)
2. Latent Heat of Vaporization
3. Psychrometric Constant
4. Air Temperature
5. Air Humidity
6. Vapor pressure
7. Dew point Temperature

3. Methodology and Model Formulation

Methodology of this study includes estimation of Evapotranspiration by Penman Monteith Method (Empirical Method) and development of models for Model Tree (a data driven technique) for available parameters.

Among many data-driven techniques, the artificial neural network (ANN) is the most widely used. The concept of ANNs is inspired by the biological neural networks of the human brain. One of the disadvantages of ANNs is that for a decision maker it is very difficult to analyze the structure of the resulting ANN and to relate it to the outputs. And therefore for the present study, M5 Model Tree is used as a data driven technique to estimate the evapotranspiration for number of input parameters.

Owing to the difficulty of obtaining accurate field measurements, ET is commonly computed from weather data [1]. A large number of empirical or semi-empirical equations have been developed for assessing crop or reference crop evapotranspiration from meteorological data [2] [6]. Some of the methods are only valid under specific climatic conditions and cannot be applied under conditions different from those under which they were originally developed [2]. Considering all, the FAO Penman-Monteith method is recommended as the sole method for determining Reference

Evapotranspiration (RET) [1] as it approximates grass ETo at the location evaluated and used worldwide [1].

3.1. Penman Monteith Equation (Empirical Method)

Numerous researchers have analyzed the performance of the various calculation methods for different locations. As a result of an Expert Consultation held in May 1990, the FAO Penman-Monteith method is now recommended as the standard method for the definition and computation of the reference evapotranspiration, ETo [1]. This empirical equation is used to calculate the evapotranspiration for daily measured data. FAO Penman Monteith Equation to calculate Reference Evapotranspiration can be expressed as- [1]

$$ETo = \frac{0.408\Delta\,(Rn - G) + \gamma \frac{900}{T + 273} u2\,(es - ea)}{\Delta + \gamma(1 + 0.34u2)} \quad ...1$$

Where

ETo = Reference Evapotranspiration in millimeter per day
Δ = slope of the saturation vapor pressure-temperature curve at the air temperature in kilopascal per ^{0}C
γ = psychrometer constant in kilopascal per ^{0}C
Rn = net radiation in million joules per meter square per day
G = soil heat flux density in million joules per meter square per day
u2 = wind speed at 2 meters height in meter per second
es = the saturation vapor pressure obtained as the average of the saturation vapor pressures at the mean maximum and mean minimum in kilo Pascal
ea = mean actual vapor pressure obtained as the saturation vapor pressure at the daily average dew point, in kilo Pascal
T= Mean daily air temperature at 2m height in ^{0}C
es - ea= saturation vapor pressure deficit

Thus it can be concluded that Reference Evapotranspiration (ETo) provides a standard to which evapotranspiration at different periods of the year or in the region can be compared and ET from other crops can be related [1]. The stepwise procedure to estimate the evapotranspiration is given in [1] and the comparison between various methods along with Penman Monteith method is given in [2].

3.2. Introduction of M5 Model Tree (A Data Driven Technique)

This is a machine-learning technique which is based on the idea of splitting of the parameter space into areas (subspaces) and building in each of them a linear regression model. In fact the resulting model can be seen as a modular model, or a committee machine, with the linear models being specialized on the particular subsets of the input space. This idea is not new. Combination of specialized models ("local" models) is used in modeling quite often. However, the M5 model tree approach, based on the principle of information theory, makes it possible to split the multi-dimensional parameter space and to generate the models automatically according to the overall quality criterion; it also allows for varying the number of models.

The splitting in MT follows the idea of a decision tree but instead of the class labels it has linear regression functions at the leaves, which can predict continuous numerical attributes. The splitting of whole dataset into number of subsets and these subsets into simple linear models is as shown in Figure 1.

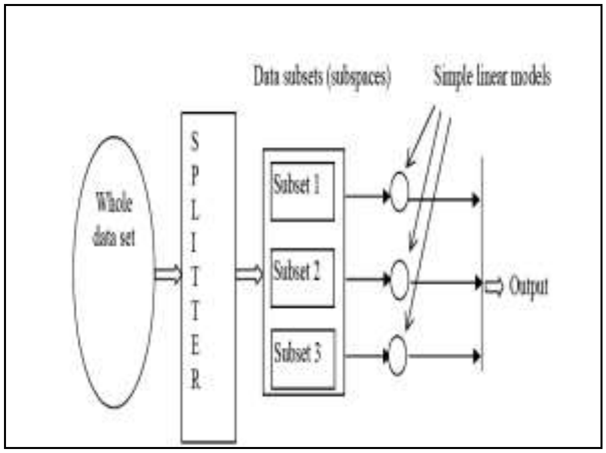

Figure 1.*The induction of a model tree as a modular model [7]*

3.3. Estimation of ET by Penman Monteith Method (Daily and Monthly)

The meteorological data for a given site (as mentioned above) is available in the prescribed format day wise. This data has been converted into per months by taking averages of each and every parameter and used in the Method of calculations for monthly estimations. The model formulation for Penman Monteith and Model Tree for Daily and Monthly variation is as shown in Table 2 for USGS Station1, 2 and 3.

Table 2: *Number of models by two methods for three stations*

Station Name	Method	No of Models	Duration	Parameters Used
USGS Station 1, 2 and 3	Penman Monteith	1	Daily	Rn, RH max, RH min, delta, G, r, T, u2, es, ea
		1	Monthly	Rn, RH max, RH min, delta, G, r, T, u2, es, ea
	Model Tree	1	Daily	Rn, RH max, RH min, delta, G, r, T, u2, es, ea
		1	Monthly	Rn, RH max, RH min, delta, G, r, T, u2, es, ea

3.3.1. Daily and Monthly Estimation

In the present study, ET has been calculated for year 2000 to 2004 (5 Year) for measured weather data from three USGS stations. The stepwise procedure which is given in [1] is followed for the calculation. The estimation of the ET using Penman Monteith method daily and monthly has been done for USGS Station1. The calculations are done for year 2000 to 2004 for three USGS Stations but for sake of understanding the results for year 2000 (For USGS Station 1) as mentioned earlier are shown in Table 3 and 4.

The direct comparison between the ET calculated and ET observed has been done along with other error measures such as CC, RMSE, MAE etc. which are discussed in the Results.

3.4. Development of model using Model Tree

To develop the model in model tree overall dataset has been divided into training and testing dataset for USGS Station 1, 2 and 3. For USGS Station 1, total 1543 daily values are available, out of which 70% values are taken for training dataset and 30% values are considered for testing dataset. The data files so prepared (training and testing) are used in the Model Tree as input. The use of Model Tree for the estimation of evapotranspiration using these input files is as done in further sections for the measured meteorological parameters. An effort has been done towards the calculation of evapotranspiration using maximum parameters (All parameters which are required in Empirical Equations).

3.4.1. Estimation of ET by Model Tree (Daily)

For station1, data is available for total 1543 days. For the model preparation whole dataset is divided into training and testing dataset as said earlier. The tree for ET for maximum parameters is as shown in figure 2 below.

Calculation of ET has been done by using Model Tree with compulsory inputs which are needed for the calculation of Evapotranspiration. The compulsory parameters which are taken into account are maximum humidity, minimum humidity, Radiation, maximum temperature, minimum temperature and wind speed, actual and saturation vapor pressure. All these parameters are used as input to get ET as output. As described earlier, overall dataset has been divided into training and testing dataset. Thus, for training dataset 70% of the values are considered and 30% values are considered for testing dataset. The Tree view of said is given in Figure 2 with Linear Equations from LM1 and LM10 (For Net Radiation Rn and Delta). From the tree it is observed that when Net Radiation (MTRn) <8.51 that is 5.4 then it choose Delta; if the delta value is > 0.15 then it uses LM4 as a linear equation. But when the Delta value is ≤0.15 it will move towards the Rn input; and if the value of Rn is ≤ 2.203 then it will use LM1 as a linear equation for the estimation purpose against the targeted input (Observed ET).

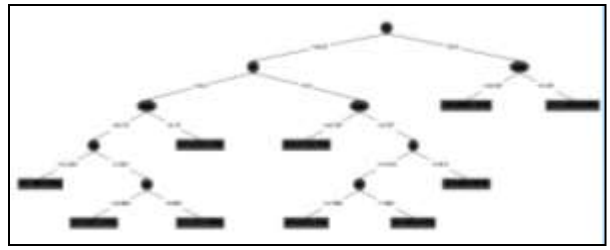

Figure 2.*Tree view for MT model having all maximum inputs parameters*

The evapotranspiration calculated by using M5 Model Tree is as shown in Table 4 along with all error measures which has discussed under the heading result and discussion.

3.4.2. Estimation of ET by using MT (Monthly)

The data used for the calculation is for total 51 months. Thus to use this data for Model Tree, the whole dataset is divided into training and testing dataset. For training dataset (70%) and for testing 30% data is considered.
Out of 51 values 37 (70%) values are taken for training and remaining 15 (30%) values are taken for testing dataset. The .csv files of training and testing dataset are used in model tree and ET is calculated. While calculation MT takes the targeted ET and generates its own linear equation and estimate the ET. Along with the estimation it also does the various error measures between estimated and targeted ET. The calculation of evapotranspiration by using MT (monthly) is given in Table 4. The equations generated are as given below.

LM num: 1
Obs ET = 0.1359 * Rn + 0.0049 * RH Max + 11.0661 * delta- 0.6554
LM num: 2
Obs ET = 0.1423 * Rn + 11.1245 * delta - 0.2005
LM num: 3
Obs ET = 0.1481 * Rn - 0.0653 * U2 + 11.7135 * delta - 0.0535
LM num: 4
Obs ET = 0.2905 * Rn + 0.0057 * RH Min + 0.0154 * T + 5.5382 * delta - 1.1782

The tree view for the monthly calculation by model tree is as shown in Figure 3 below. According to this tree, if the Rn is greater than 11.105 then it uses LM4 equation for the estimation. But when it is less than 11.105 then it uses LM1, LM2 and LM3 equations. There are again subdivisions of values of Rn below 11.105. if the value of Rn is greater than 9.085 then it uses LM3 equation and when it is less than 9.085 then it uses either LM1 or

LM2, it again depends on the value of Rn whether it is less or greater than 7.28. When the value of Rn is greater than 7.28 then it uses LM2 and when it is less then uses LM1 equation for the estimation of ET.

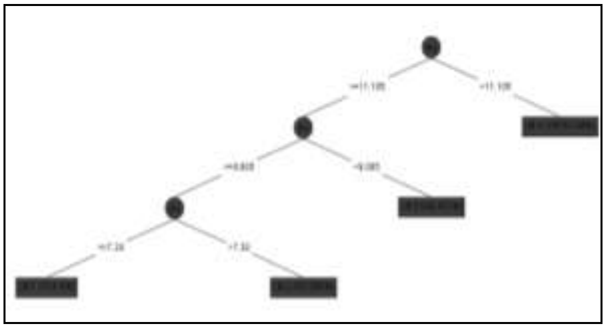

Figure3.*Tree view for MT model having maximum inputs parameters (monthly estimation)*

4. Results and Discussion

As discussed earlier, estimation of daily and monthly evapotranspiration using Penman Monteith (Empirical Method) and by using a data driven technique of Model Tree is done at three stations of Florida region (The St John's Rv Marsh at Blue Cypress Nr Fellsmere (USGS Station 1), Duda Farms Nr Lake Winder Nr Rockledge (USGS Station2) and Ferris Farms Nr Floral City (USGS Station3)). To judge these estimations, the estimated values of evapotranspiration were compared with the observed ET values at the same stations and results are discussed in this section.

4.1. Daily Calculation of ET

The comparison between three station (USGS 1, 2, 3) by using PM and MT is given in following table. Their performance is judged by means of correlation coefficient (r). The comparison between Penman Monteith Method (PM) and Model Tree (MT) has been shown in below Table 3.

Table 3: *Calculation of daily ET*

Station	ET Daily by PM (r)	ET Daily MT (r)
Station1 USGS 274143080424100	0.92	0.95
Station2 USGS 281626080463400	0.82	0.95
Station3 USGS 284541082163400	0.84	0.96

Where- 'r' is the correlation coefficient

It can be observed from the table that correlation coefficient for station 1 is 0.92 (PM) and 0.95 (MT) which is quite well, for station2 correlation coefficient is 0.82 (PM) and 0.95 (MT) which is also good. Now station 3 coefficient of correlation is 0.84 (PM) and 0.96 (MT). Thus in general for three stations correlation coefficient value varies in between 0.82 to 0.92 for PM and 0.95 to 0.96 for (MT). Thus we can say that MT performs well as compared to PM for Daily Variation.

4.2. Monthly Calculation of ET

The comparison between three station (USGS 1, 2, 3) by using PM and MT (monthly calculation) is given in following Table 4. The performance is judged by means of Correlation coefficient, Mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error.

Table 4: *Results of monthly calculation of ET for three stations using PM and MT*

Station	ET Monthly by PM (mm/day)	MT	Error Measures
Station1	0.93	Correlation Coefficient (CC)	0.98
		Mean Absolute Error (MAE)	0.32
		Root Mean Squared Error (RMSE)	0.42
		Relative Absolute Error (RAE)	37.88 %
		Root Relative Squared Error (RRSE)	41.87 %
		Total Number of Instances	16
Station2	0.82	Correlation Coefficient (CC)	0.97
		Mean Absolute Error (MAE)	0.27
		Root Mean Squared Error (RMSE)	0.33
		Relative Absolute Error (RAE)	28.85%
		Root Relative Squared Error (RRSE)	30.45%
		Total Number of Instances	14
Station3	0.86	Correlation Coefficient (CC)	0.96
		Mean Absolute Error (MAE)	0.26
		Root Mean Squared Error (RMSE)	0.34
		Relative Absolute Error (RAE)	33.18%
		Root Relative Squared Error (RRSE)	36.03%
		Total Number of Instances	15

From table it can be said that Correlation Coefficient for PM varies between 0.82 and 0.93 for station 1, 2 and 3. And in case of MT, it varies from 0.96 to 0.98 for station 1, 2, and 3. It can also be observed from the table that for station 1, 2 and 3 MT performs well as compared to PM method since it has more Correlation Coefficient values as that of the PM Method.

4.3. Discussion on Comparison of Daily and Monthly ET for Three Stations by using PM and MT

The calculation of ET by Penman Monteith (PM) and Model Tree (MT) on Daily, Monthly basis for USGS Stations 1, 2 and 3 has been done. The details for the same are given in Table 5. It is found from the table that correlation coefficient for PM (Daily) for USGS Station 1, 2 and 3 is 0.92, 0.82 and 0.84 whereas in case of MT it is 0.95, 0.95 and 0.96. That means it has more correlation coefficient for MT as compared to PM method. In case of Monthly calculation correlation coefficient for station 1 is 0.93 whereas 0.98 correlation coefficient is obtained in case of MT. The same trend is observed for station 2 and station 3. Therefore it can be stated that MT has better correlation coefficient than PM method for daily and monthly calculation. Thus it can be said that MT performs well than PM.

Table5. *Comparison of daily and monthly ET for three stations by using PM and MT*

Station	Data Used	ET (mm/day)			
		Daily (r)		Monthly (r)	
		PM	MT	PM	MT
Station1	2000 to 2004 (1543 days)	0.92	0.95	0.93	0.98
Station2	2000 to 2004 (1491 days)	0.82	0.95	0.82	0.97
Station3	2000 to 2004 (1359 days)	0.84	0.96	0.86	0.96

5. Conclusion

Estimation of daily and monthly evapotranspiration using Penman Monteith (Empirical Method) and Model Tree (Data Driven Technique) has been done at three stations of Florida region (USGS Station 1), (USGS Station2) and (USGS Station3). The performance of these methods have been judged by the coefficient of correlation between estimated and observed evapotranspiration along with other error measures such as RMSE, MAE and CE.

Following conclusions can be obtained from the results mentioned.

- It is evident from the results that the daily and monthly models of M5 Model Tree outperform than Penman Monteith models. All the results of MT models are of superior quality than empirical method for all the three stations of Florida.
- For Penman Monteith models many parameters are required whereas, MT models can perform with the available parameters and models are developed with different combinations, this is the beauty of data driven technique which has been utilized in the present study. This is the credential of the work that declares the competency of one data driven technique to estimate the evapotranspiration when traditional empirical methods are out of scope due to huge data requisite and large time consumption.

Thus it can humbly be stated that M5 Model tree is a useful data driven technique for estimation of evapotranspiration and can be used in explored further for hydrological modeling.

References

[1] Allen R.G, Pereira L.S., Raes D., Smith M. "Crop evapotranspiration - Guidelines for computing crop water requirements -FAO Irrigation and drainage paper 56", 1998.

[2] Xu and Singh V.P. "Cross Comparison of Empirical Equations for Calculating Potential Evapotranspiration with Data from Switzerland", water Resources Management 16:197-219, 2002.

[3] Kim S. and Kim H., "Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling" Journal of Hydrology 351, 299– 317, 2008.

[4] Chaudhari N., Londhe S., "Estimation of pan evaporation using soft computing tools", Int. J. Hydrology Science and Technology, Vol. 2, No. 4, 2012.

[5] http://fl.water.usgs.gov/et/etdata.html

[6] Xu and Singh 'Evaluation and Generalization of Radiation-based Methods for Calculating Evaporation', Hydrolog. Processes 15, 305–319 2001.

[7] Solomatine D.P., Bhattacharya B. "Neural networks and M5 model trees in modeling water level–discharge relationship" Neurocomputing 63 382 (2005) 381–396, 2005.

[8] Subramanya K. "Engineering Hydrology", Tata Mcgraw Hill Education Private Limited New Delhi, 2011.