Mathematical modelling of evapotranspiration by using remote sensing and data mining

Lamya Neissi (neissi.lamya@gmial.com)
Shahid Chamran University of Ahvaz
Mona Golabi
Shahid Chamran University of Ahvaz
Mohammad Albaji
Shahid Chamran University of Ahvaz
Abd Ali Naseri
Shahid Chamran University of Ahvaz

Research Article

Keywords: Remote sensing, data mining, GIS, machine learning

Posted Date: April 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-174771/v1

Mathematical modelling of evapotranspiration by using remote sensing and data mining

Lamya Neissi[1], Mona Golabi[2], Mohammad Albaji[3], Abd Ali Naseri[4]

[1]Phd Student of Irrigation and Drainage, Department of Irrigation and Drainage, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Iran. E-mail: neissi.lamya@gmail.com

[2]Assistance Professor, Department of Irrigation and Drainage, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Iran
Email: mona_golabi@yahoo.com,m.golabi@scu.ac.ir

[*3] Assistance Professor, Department of Irrigation and Drainage, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Iran.
E-mail: m_albaji2000@yahoo.co.uk; m.albaji@scu.ac.ir

[4] Professor, Department of Irrigation and Drainage, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Iran.
E-mail: abdalinaseri@yahoo.com

Abstract

Precise evaluation of evapotranspiration in an extended area is crucial for water requirement. By using remote sensing evapotranspiration algorithms, many climatological variables are needed. In case of using climatological variable measurements, many climatic stations must be established in that specific area. By using data mining method integrated with remote sensing, evapotranspiration can be calculated with high accuracy. A physical-based SEBAL evapotranspiration algorithm was modeled by GIS model builder for ET calculations. Albedo, emissivity, and Normalized Difference Water Index (NDWI) were considered as M5 decision tree model inputs. Evapotranspiration was evaluated for 3 April 2020 to 17 September 2020 and the equations were extracted in the M5 decision tree model and these equations were modeled in GIS by using python scripts for 3 April 2020 to 17 September 2020. The results make clear that the mathematical decision tree model can estimate the evapotranspiration gained by physical-based SEBAL algorithm in high accurately.

Keyword: Remote sensing; data mining; GIS; machine learning.

1. Introduction

Irrigation scheduling of crops can be done by using meteorological data for evapotranspiration calculations. By using satellite images and different algorithms, evapotranspiration can be estimated in an extended area and reach an accurate irrigation scheduling (Jaferian et al., 2019; Song et al., 2018; Diarraa et al., 2017; Colaizzi et al., 2017; Anderson et al., 2012).

Evapotranspiration estimation is a complicated process. For estimating evapotranspiration, different equations were obtained which can be used in different equations such as FAO-Penman–Monteith, Blaney-Criddle, etc. Ground observations represent the results for one

specific point in which high accuracy is needed to generalize them for extended region Hence evaporation is different from station to station. By using remote sensing technologies, one can reach acceptable and high accuracy for a specific extended region. By using satellite images as a remote sensing technic, ground observations transformed to soft data. Among different methods of data mining, the M5 decision tree was used to estimate the evapotranspiration in an extended area (Gibert et al., 2018).

This research is intended to establish an applicable different linear relation by using the M5 decision tree between independent remote sensing parameters (albedo, emissivity, and Normalized difference water index) with the dependent parameter (evapotranspiration) by using data mining which is the most important innovation of this research.

Landsat8 satellite images and SEBAL algorithm were used for evapotranspiration estimation. which was used in many evapotranspiration estimations and acceptable results obtained by these researches (Mhawej et al., 2020; Elnmer et al., 2019; Kong et al., 2019; Ochege et al., 2019; Gobbo et al., 2019; Kamali and Nazari, 2018).

Land surface temperature (LST) is one of input parameters for evapotranspiration estimation but the spatial resolution of this band is 100m and the estimated evapotranspiration image by using SEBAL algorithm by using this band has 100m spatial resolution. Which the other aim of this research is to enhance spatial resolution by using the M5 decision tree. Input parameters have 30m spatial resolution and by applying the gained equations by the M5 decision tree, an evapotranspiration map with higher spatial resolution can be obtained.

According to sugarcane plantation in an extended area in the southwest of Iran (more than 94000 ha), an extremely high volume of water is consumed in this section, so spatially enhancing

evapotranspiration estimation image, irrigation water scheduling can be calculated more precisely.

2. Materials and methods

2.1. Study area

This study was conducted in the Amir-Kabir Agro-Industry Sugarcane fields. The Sugarcane fields are located in the southwest of Iran (Figure1). The soil texture is clay-loam and annual average evapotranspiration for 20 years was 3331.812 mm. The total area of cultivated sugarcane in Khuzestan is over 84000 ha. Each farm has a low-pressure hydro flume irrigation system and a subsurface drainage system with a 40m distance and 1.8 m depth for each drain tile. The total irrigation water consumption is 3000 mm and the peak of irrigation water was applied in July.

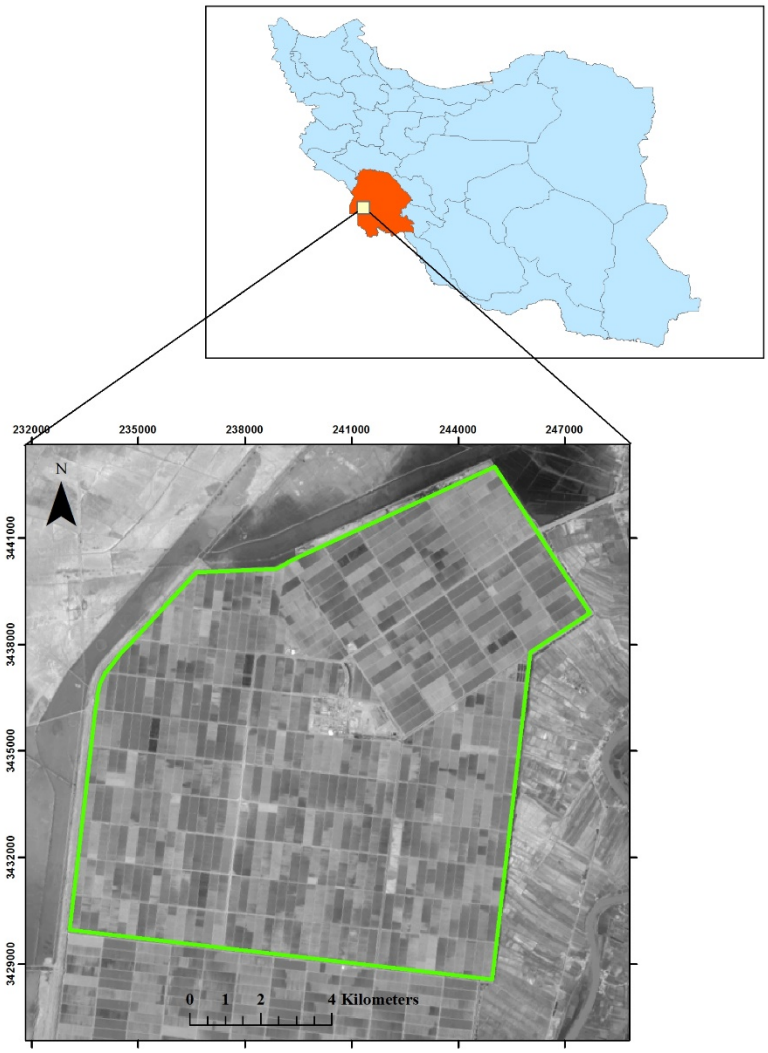


Fig. 1. Amir-Kabir sugarcane Agro-Industry location area

2.2. Remote sensing

Landsat 8 OLI satellite images were the main data for remote sensing processes (http://glovis.usgs.gov). Thermal bands have lower resolution compared to other optic bands. As for Landsat8, band10 image represents a thermal band that provides less spatial resolution (100m) but Thermal bands are critical for evapotranspiration estimation and Landsat8 has the most appropriate thermal band for agricultural evapotranspiration estimation in a great variety

93 extended regions. Amir-Kabir agro-industry plantation with area of 14000 ha is extended enough
94 for evapotranspiration calculations by using remote sensing images.

95

96 **2.3. Ground measurements**

97

98 ET estimation requires meteorological data. Meteorological data were obtained from Amir-Kabir
99 agro-industry plantation local weather station. Weather data including Max and Min temperature,
100 the relative percentage of humidity, wind speed, and sunshine hours were used for
101 evapotranspiration calculations and Ref-et software was used for reference evapotranspiration
102 calculations.

103

104 **2.4. SEBAL algorithm**

105

106 The SEBAL algorithm was used to calculate the evapotranspiration (ET) of Sugarcane for the
107 Amir-Kabir agro-industry plantation. SEBAL algorithm uses thermal and multispectral digital
108 images of Landsat or other sensors to estimate the evapotranspiration (Bastiaanssen et al., 1998).
109 The ET calculation process is obtained by the amount of energy remained from the classical
110 equation of energy balance presented in equation (1):

$$\lambda \mathrm{ET} = Rn - G - H \qquad (1)$$

111

Where λET is the latent heat flux in the atmospheric boundary layer (W/m^2), Rn is the net radiation (W/m^2), H is the sensible heat flux (W/m^2) and G is the soil heat flux (W/m^2) (Allen et al., 2002).

The net radiation (Rn) is computed by subtracting all outgoing radiant fluxes from all incoming radiant fluxes. The soil heat flux (G) and sensible heat flux (H) are subtracted from the net radiation flux at the surface Rn to compute the residual energy available for evapotranspiration (λET) (Allen et al., 2002). The latent heat flux at the moment is converted into daily λET24 assuming a constant evaporative fraction (Λ) for 24 h calculated from the instantaneous energy fluxes as observed in the satellite data as Eq. (2):

$$\Lambda = \frac{\lambda ET}{R_n - G} \quad (2)$$

The daily actual evapotranspiration (ET_{24}) can then be determined as Eq. (3):

$$ET_{24} = \frac{86400 \times \Lambda \times (Rn_{24} - G_{24})}{\rho_w \times \lambda} \quad (3)$$

Where Rn_{24} and G_{24} are the average net radiation for the day and daily soil heat flux (W/m^2) respectively which computed from raw products of instantaneous satellite spectral radiance, vegetation indices, and satellite surface temperatures which are then expressed as average day estimates (Singh et al., 2008). For this study, four cloud-free satellite images were obtained for

April 2019 and 2020. Actual evapotranspiration (ETa) maps in mm/day are generated by the SEBAL algorithm for each day.

2.5. Data mining

Data science analysis Data Mining (DM) algorithms are the most fundamental components. Certain DM techniques such as artificial neural networks, clustering, and case-based reasoning or Bayesian networks have been applied in environmental modeling (Gibert et al., 2018).

Decision Tree methods uses the explanatory variables with higher discriminant power by considering the response variable, then iteratively subdivide the training sample by building a tree where the internal nodes are associated with the variables and its corresponding branches are the possible values of the variable (Gibert et al., 2018). M5 Model Tree (introduced by Quinlan in 1992), has linear regression functions at the leaf nodes, which develops a relationship between input and output variables. Data are split into subsets and a decision tree is created. The data in child nodes of splitting criterion depends on treating the standard deviation of the class values and calculating the expected reduction in this error in consequence of testing each attribute at that node. The standard deviation reduction (SDR) is calculated as Eq. (4) (Quinlan 1992):

$$SDR = sd(T) - \sum_{i} \frac{|T_i|}{|T|} \times sd(T_i) \quad (4)$$

Where T is a set of data that reaches the node, T_i is the subset of data that have the ith outcome of the potential set and sd is the standard deviation (Rahimikhoob et al., 2013; Wang and Witten,

1997). the data in child nodes are purer due to a less standard deviation in comparison to parent nodes. The M5 tree selects the one that maximizes the expected error reduction after scanning all the possible splits.

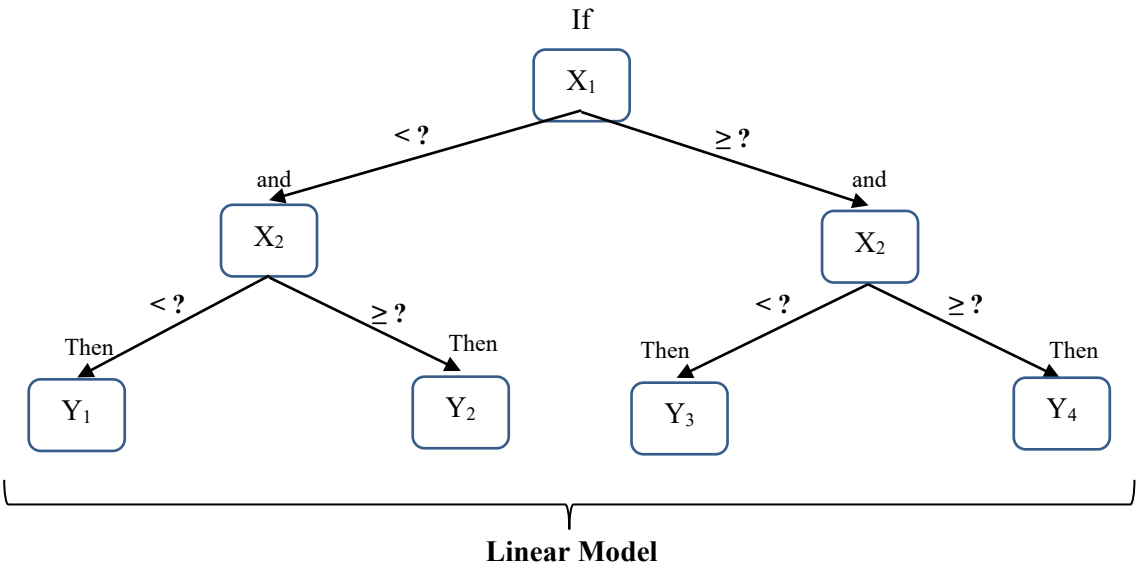


Fig. 2. Structure of M5 decision tree (Models Y1–Y4 are linear regression models)

2.6. Model Inputs and Output

For estimating evapotranspiration by SEBAL algorithm, meteorological data are needed including temperature, humidity, wind speed, etc. some inputs of SEBAL algorithm such as albedo and emissivity are affected by the land surface temperature, so for using fewer variables by data mining, albedo and emissivity were considered as inputs of M5 decision tree since these two parameters are easy to get and better show the temperature variances. Also, transpiration depends on the moisture of the plant. in data mining calculations, one vegetation index must represent the plant moisture such as Normalized Difference Water Index (NDWI). So, albedo and emissivity are represented as absorbed and transformed light to the atmosphere and NDWI is

represented as plant moisture. The main idea was to use the basic SEBAL equation (ET = Rn- G-H) as a simple equation: ET = a(Albedo) – b(emissivity) – c(NDWI) and calculate the constant values with M5 decision tree model. The three inputs of the M5 decision tree are explained hereafter in more detail.

2.6.1. Evapotranspiration

One of the crucial parameters for evapotranspiration estimation is Land surface temperature (LST) which Radiation and the exchange of energy flux between the earth's surface and atmosphere depend on it (Weng et al., 2019). A physical model such as the SEBAL algorithm has made it possible to estimate evapotranspiration for large areas. Surface biophysical characteristics such as albedo, greenness, and wetness are among the most important parameters affecting LST (Weng et al., 2019). The energy distribution is determined by albedo and emissivity of the surface and atmosphere. previous studies show that surface emissivity strongly correlates to vegetation cover (Griend and Owe, 1993; Rechid et al., 2009), vegetation also strongly affects atmospheric properties through evapotranspiration (Gordon et al., 2005). atmosphere emissivity is determined by atmospheric water vapor pressure (Staley and Jurica, 1972; Brutsaert, 1975).

2.6.2. Albedo

Albedo is a dimensionless diffuse reflectivity or reflecting power of a surface (Zhang et al., 2017) and is an important effective parameter on digital climate models and surface energy

balance equations (Zhang et al., 2017). Surface albedo is computed by correcting the α_{toa} for atmospheric transmissivity as Eq. (5):

$$\alpha = \frac{\alpha_{toa} - \alpha_{path\_radiance}}{\tau_{sw}^2} \quad (5)$$

Where; $\alpha_{path\_radiance}$ is the average portion of the incoming solar radiation by considering all bands that is back-scattered to the satellite before it reaches the earth's surface, and τ_{sw} is the atmospheric transmissivity (Allen et al., 2002).

2.6.3. Emissivity

The surface emissivity is the ratio of the actual radiation emitted by a surface to that emitted by a black body at the same surface temperature (Allen et al., 2002). Surface emissivity is an important variable for estimating land surface temperature and determining long-wave surface energy balance (Mira et al., 2010). Sobrino et al. (2004) proposed emissivity based NDVI in three different cases as Eq. (6):

$$\epsilon = \begin{cases} 0.973 & NDVI < 0.2 \\ \varepsilon_v P_v + \varepsilon_s (1 - P_v) + d\varepsilon & 0.2 \leq NDVI \leq 0.5 \\ 0.986 & NDVI > 0.5 \end{cases} \quad (6)$$

Where ε_v is the vegetation canopy emissivity and ε_s is the bare soil emissivity; in this paper ε_v = 0.986 and ε_s = 0.973. the effect of the geometrical distribution of the natural surfaces is measured as dε in Eq.6 . P_v is the vegetation proportion obtained according to (Carlson and Ripley, 1997) as Eq. (7):

$$P_v = [\frac{NDVI - NDVI_s}{NDVI_v - NDVI_s}]^2 \quad (7)$$

The minimum value of the NDVI for bare soil over the study region is presented as $NDVI_S$ and $NDVI_V$ is the highest NDVI for a fully vegetated pixel. The emissivity of land surfaces can differ significantly by vegetation, surface moisture, and roughness (Nerry *et al*. 1988, Salisbury and D'Aria 1992).

2.6.4. Vegetation Index

The NDWI spectral index which represents the crop moisture is the normalized difference water index. NDWI has been used to estimate the equivalent water thickness of vegetation canopy (Yilmaz et al., 2008). The NDWI considers two infrared bands with a central wavelength near about 0.86 μm (NIR), and a central wavelength of about 1.24 μm (SWIR). The equation is (Eq.8):

$$NDWI = \frac{\{\rho(0.86\mu m) - \rho(1.24\mu m)\}}{\{\rho(0.86\mu m) + \rho(1.24\mu m)\}} \quad (8)$$

The M5 decision tree model takes Albedo, emissivity, and a vegetation index as input and after the data mining process on these data, linear equations will be extracted. By inserting linear equations, the evapotranspiration map was obtained as an output with higher spatial resolution. Figure3 shows the flowchart of the M5 decision tree and SEBAL algorithm.

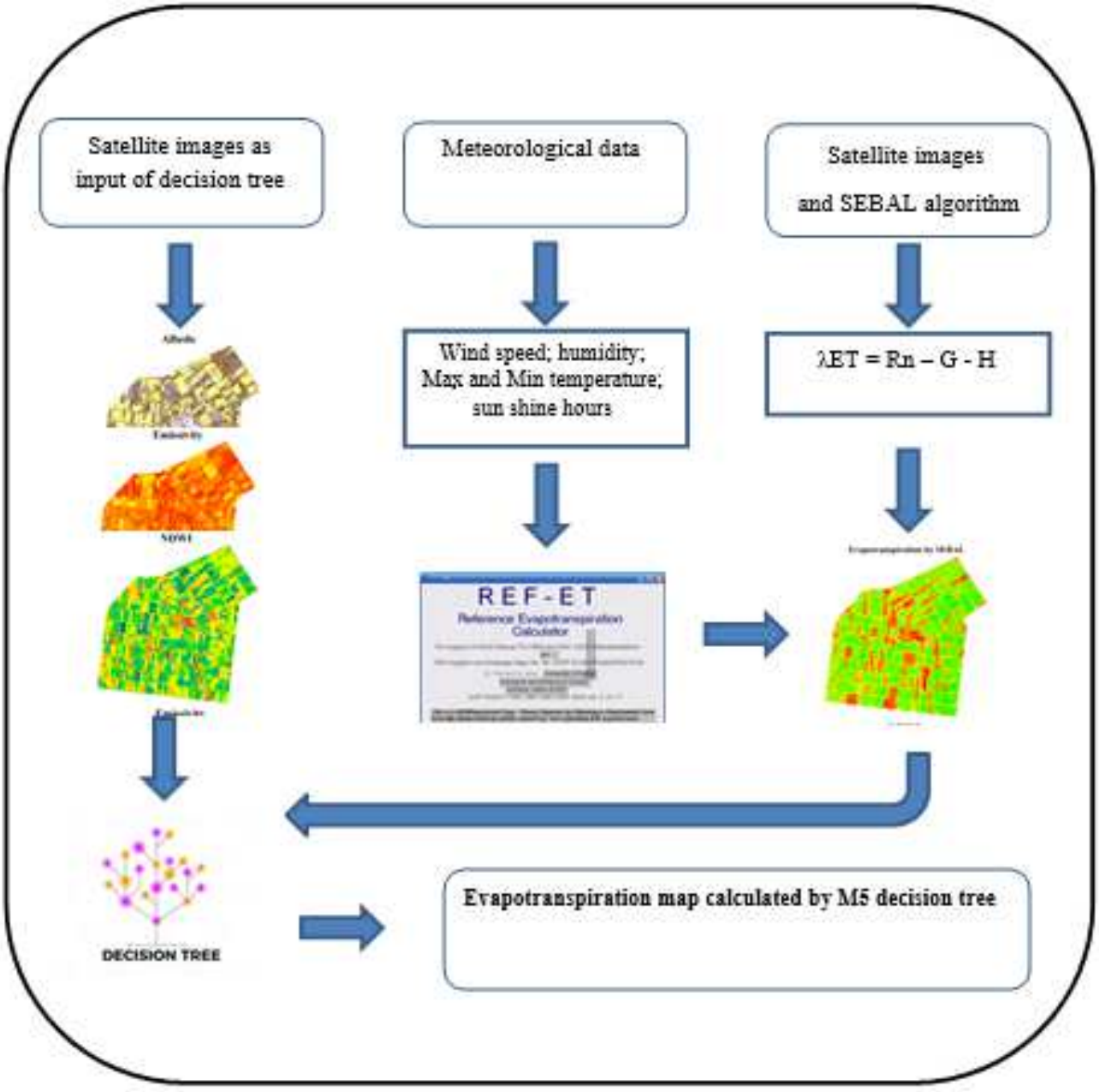


Fig. 3. Flowchart of the M5 decision tree and SEBAL algorithm

2.7. Model Validation

By using three inputs (albedo, emissivity, and NDWI) the accuracy of M5 decision tree was evaluated. The accuracy of the M5 decision tree model and the final evapotranspiration map

257 which was combined with M5 decision tree was evaluated by correlation coefficient (R^2), root
258 mean square errors (RMSE) and mean absolute errors (MAE) statistics (Eq. 9 to 11):

259

$$R^2 = 1 - \frac{\sum(ET_o - ET_p)^2}{\sum(ET_o - \overline{ET})^2} \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n}\sum(ET_o - ET_p)^2} \quad (10)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|ET_o - ET_p| \quad (11)$$

260

261 Where N is the number of data, ET_o is the observed evaporation values calculated by the SEBAL
262 algorithm and ET_p is the M5 decision tree model estimated evapotranspiration.

263

264 **3. Results**

265

266 **3.1. M5 decision tree**

267 For evapotranspiration estimation derivation M5 decision tree was used instead of SEBAL
268 physical-based model. By using decision tree for evapotranspiration estimation of two month,
269 different equations were extracted (Appendix1). The equations were obtained by using Albedo,
270 emissivity and Normalized difference vegetation index (NDWI) as model inputs. These three

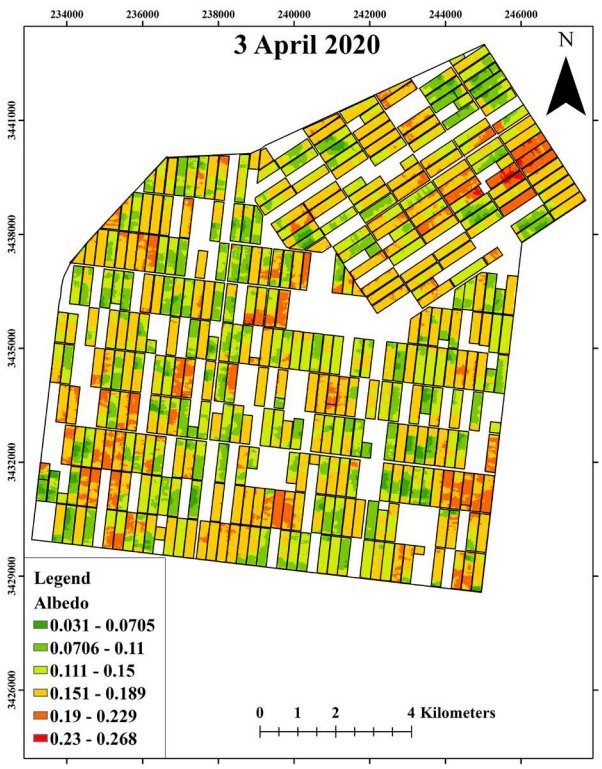
input parameters were calculated from 6 June 2020 to 24 July 2020 and were considered as input for each day of evapotranspiration estimation.

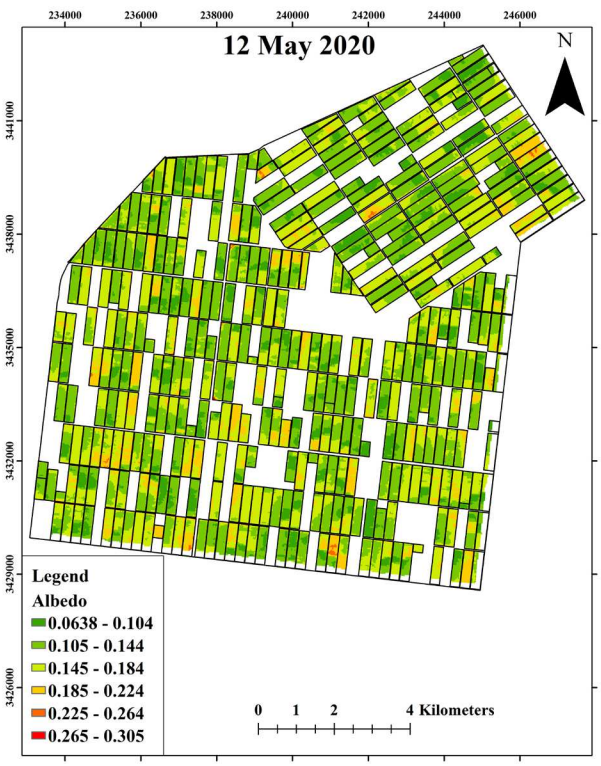
3.1.1. Inputs

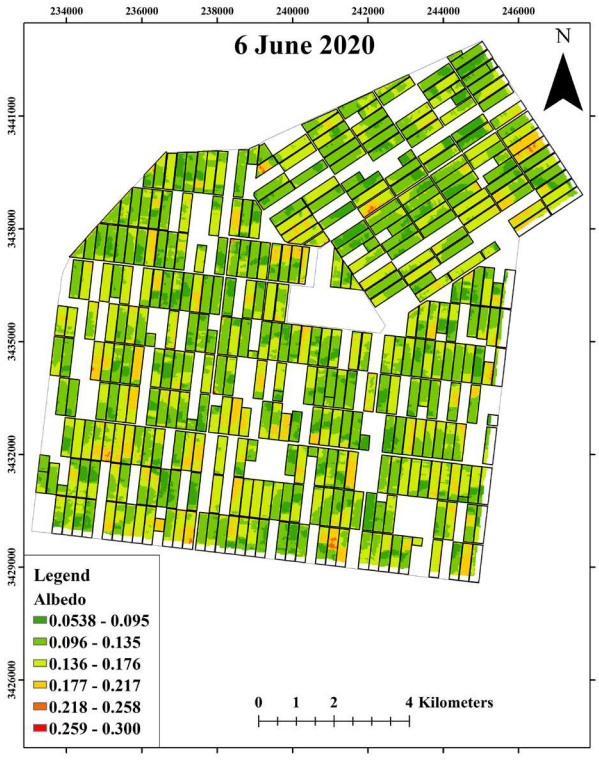
3.1.1.1. Albedo

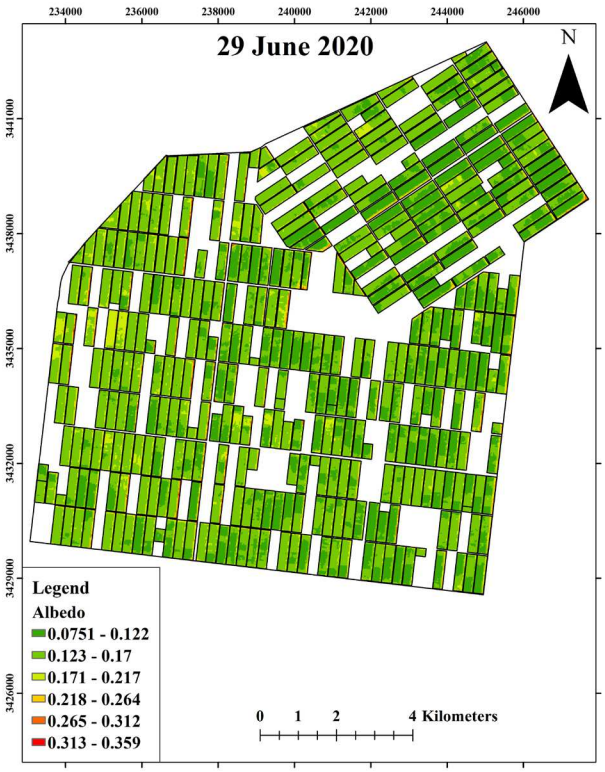
Albedo is the reflectance of solar energy from the earth's surface. Albedo is said to determine the amount of shortwave radiation to be absorbed by surfaces (Vargo et al., 2013). When initial absorption of shortwave radiation is limited, it impacts the longwave energy radiated by the earth's surface, as well as energy availability for evapotranspiration and energy to be converted to sensible heat. Ice, barren land, and sand have higher albedo, which causes higher reflectance of shortwave radiation (Vargo et al., 2013). Vegetative land cover has a lower albedo in comparison to barren land.

In this study albedo is one of model input parameters which was calculated for each image in the study period. In the first growth stage of sugarcane, vegetative cover is very low and the plant is not developed and canopy cover did not reach to its highest development. Figure 4 shows the albedo images as input parameters of the decision tree for the study duration. In satellite images canopy cover area is much lower than soil area at the first growth stage (figure 4, 3 April 2020, 12 May 2020).


3 April 2020
N
234000 236000 238000 240000 242000 244000 246000
3441000 3438000 3435000 3432000 3429000 3426000
Legend
Albedo
0.031 - 0.0705
0.0706 - 0.11
0.111 - 0.15
0.151 - 0.189
0.19 - 0.229
0.23 - 0.268
0 1 2 4 Kilometers


12 May 2020
N
234000 236000 238000 240000 242000 244000 246000
3441000 3438000 3435000 3432000 3429000 3426000
Legend
Albedo
0.0638 - 0.104
0.105 - 0.144
0.145 - 0.184
0.185 - 0.224
0.225 - 0.264
0.265 - 0.305
0 1 2 4 Kilometers


6 June 2020
N
234000 236000 238000 240000 242000 244000 246000
3441000 3438000 3435000 3432000 3429000 3426000
Legend
Albedo
0.0538 - 0.095
0.096 - 0.135
0.136 - 0.176
0.177 - 0.217
0.218 - 0.258
0.259 - 0.300
0 1 2 4 Kilometers


29 June 2020
N
234000 236000 238000 240000 242000 244000 246000
3441000 3438000 3435000 3432000 3429000 3426000
Legend
Albedo
0.0751 - 0.122
0.123 - 0.17
0.171 - 0.217
0.218 - 0.264
0.265 - 0.312
0.313 - 0.359
0 1 2 4 Kilometers

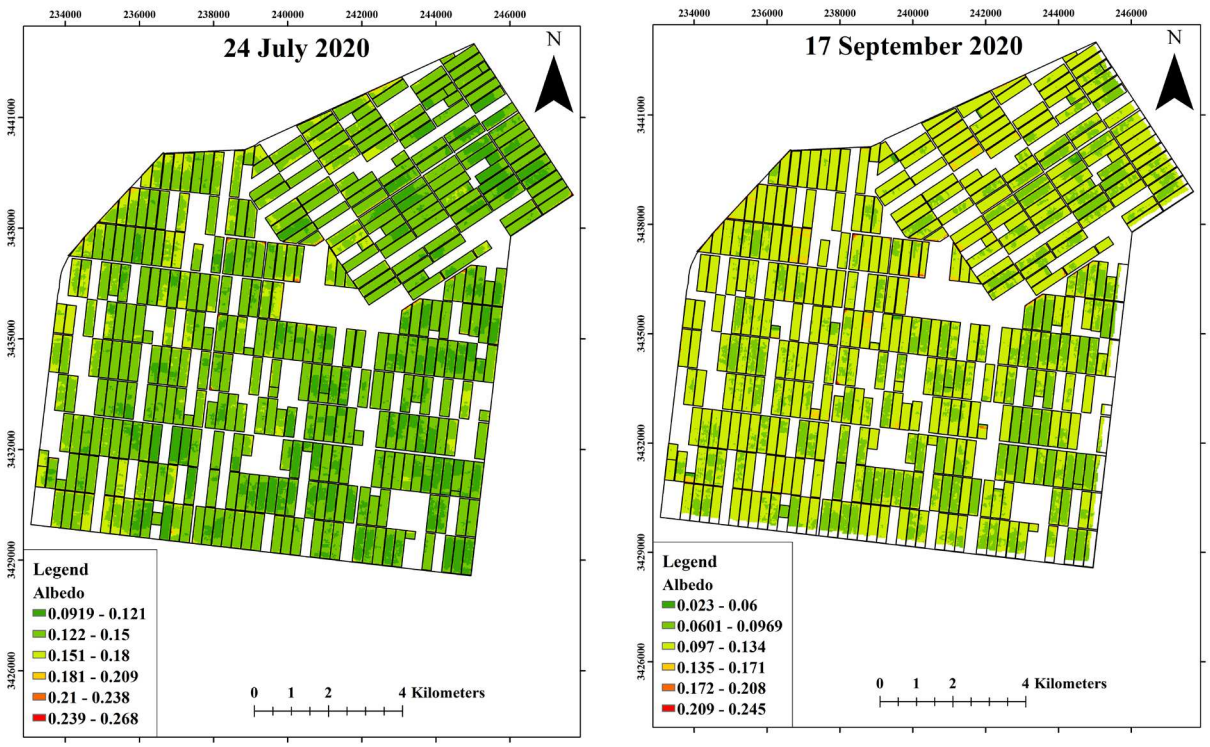


Fig. 4. Albedo input images for M5 decision tree

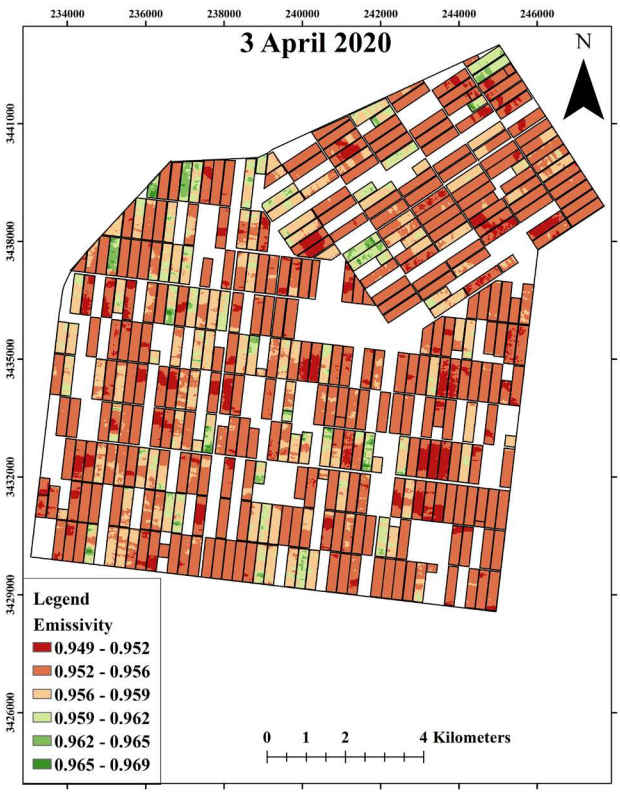
At the end stage of sugarcane growth stage (24 July 2020, 17 September 2020) by developing canopy cover of sugarcane most of the sun light absorbed by plant and less of the sun light reflected to the atmosphere, therefore albedo amount is decreases. Calculated albedo images from 3 April till 17 September 2020 makes clear that by developing and growth of sugarcane plant amount of albedo decreases and the canopy cover can affect the reflection of the sun light. Calculated albedo images for input of M5 decision tree indicates that this parameter have an acceptable spatial variability for evaluating evapotranspiration.

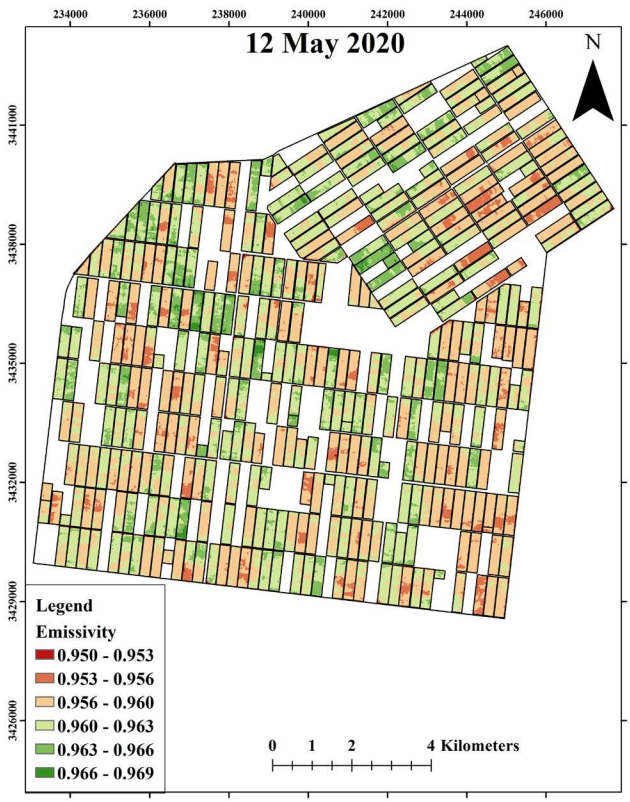
3.1.1.2. Emissivity

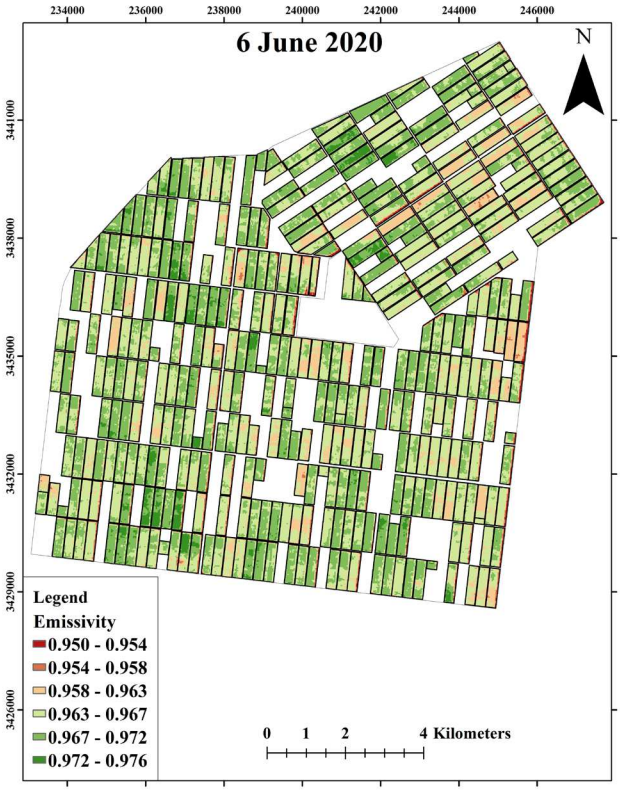
The emissivity of the surface of a material refers to the effectiveness of the surface in emitting energy as thermal radiation (electromagnetic radiation with wavelength depending on the temperature). Emissivity is mathematically defined as the ratio of the thermal radiation from the surface to the radiation from an ideal black surface at the same temperature; the value varies from 0 to 1. For C/SiC, the emissivity at 1600°C is ~0.7, which is high (Alfano et al., 2009).

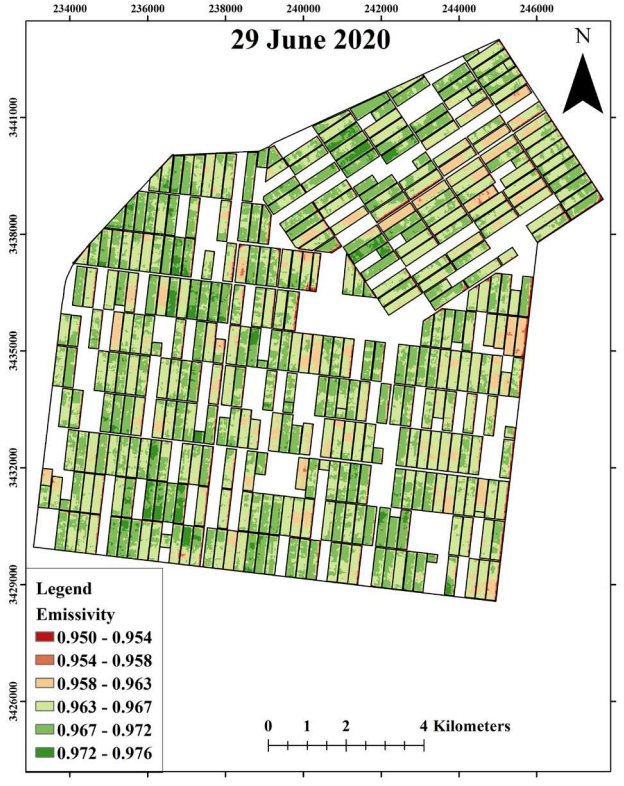
Emissivity as one input of the decision tree model related indirectly with Land Surface Temperature (LST). Due to the increase of LST in the study period of sugarcane growth can make the emissivity to increase.

Figure 5 shows the emissivity input parameter images for M5 decision tree model. Figure 5 indicates that at the first growth stage of sugarcane (figure5, 3 April 2020 and 12 May 2020) in which the LST is low can make the emissivity to decrease considering less canopy cover and more bare soil.


3 April 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
Emissivity
0.949 - 0.952
0.952 - 0.956
0.956 - 0.959
0.959 - 0.962
0.962 - 0.965
0.965 - 0.969
0 1 2 4 Kilometers


12 May 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
Emissivity
0.950 - 0.953
0.953 - 0.956
0.956 - 0.960
0.960 - 0.963
0.963 - 0.966
0.966 - 0.969
0 1 2 4 Kilometers


6 June 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
Emissivity
0.950 - 0.954
0.954 - 0.958
0.958 - 0.963
0.963 - 0.967
0.967 - 0.972
0.972 - 0.976
0 1 2 4 Kilometers


29 June 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
Emissivity
0.950 - 0.954
0.954 - 0.958
0.958 - 0.963
0.963 - 0.967
0.967 - 0.972
0.972 - 0.976
0 1 2 4 Kilometers

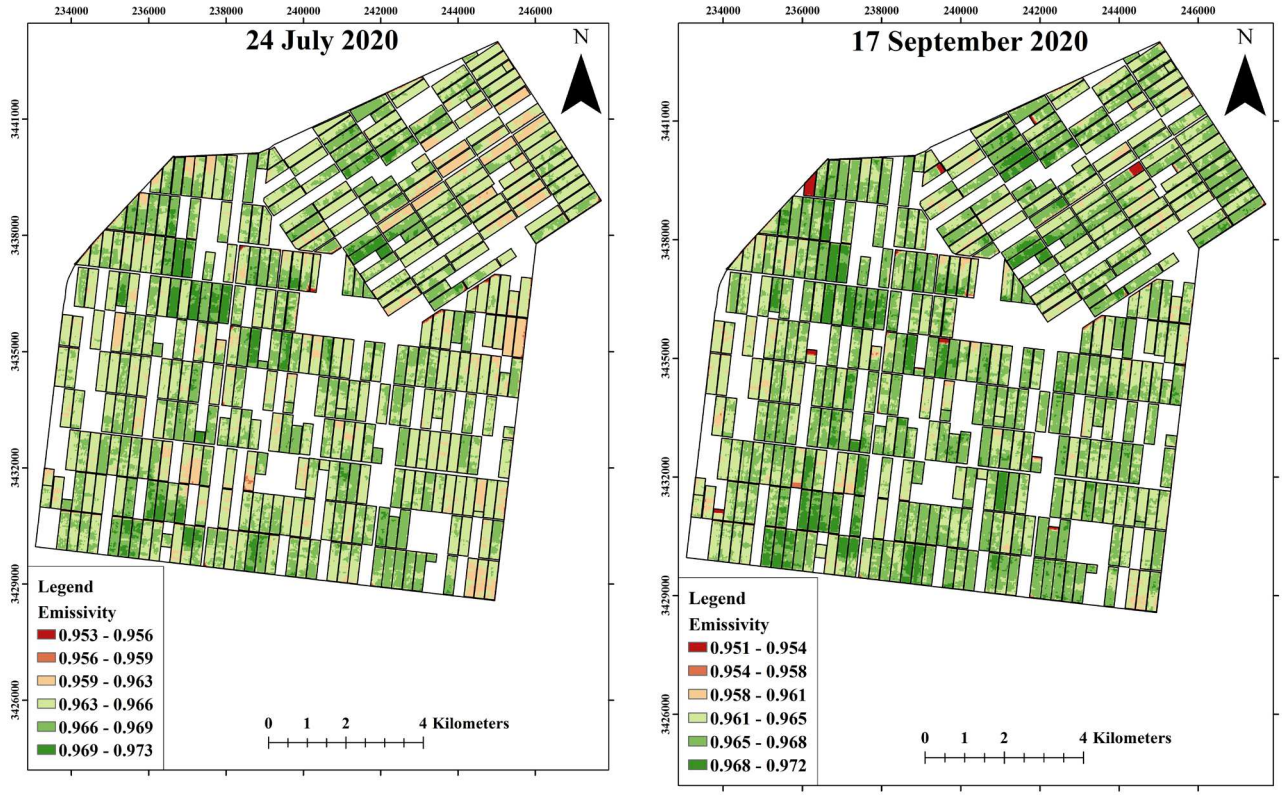


Fig. 5. Emissivity input images for M5 decision tree

In the end of sugarcane growth stage, the temperature increases which also makes the emissivity to increase. Figure 5 makes clear that emissivity has suitable spatial variability which can make the evaluation of evapotranspiration by using M5 decision tree with acceptable accuracy.

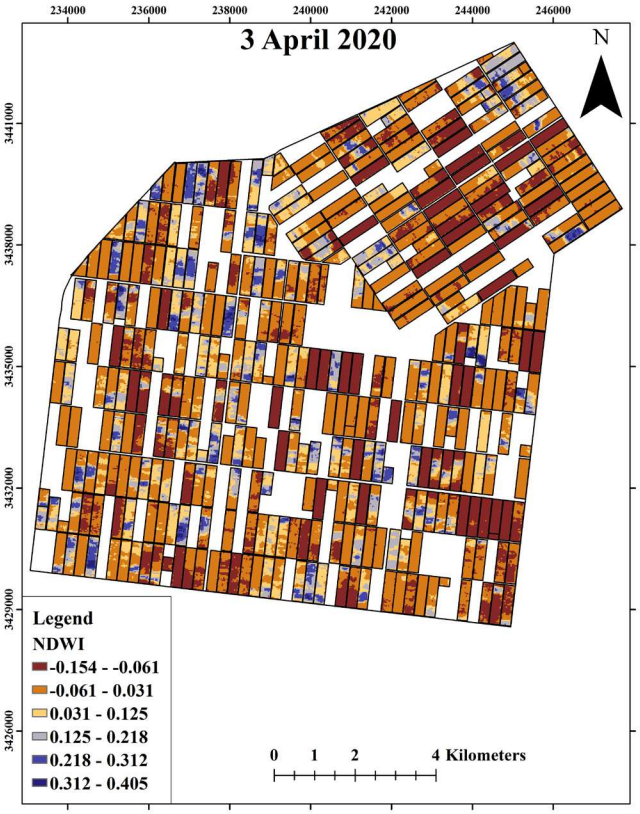
3.1.1.3. NDWI

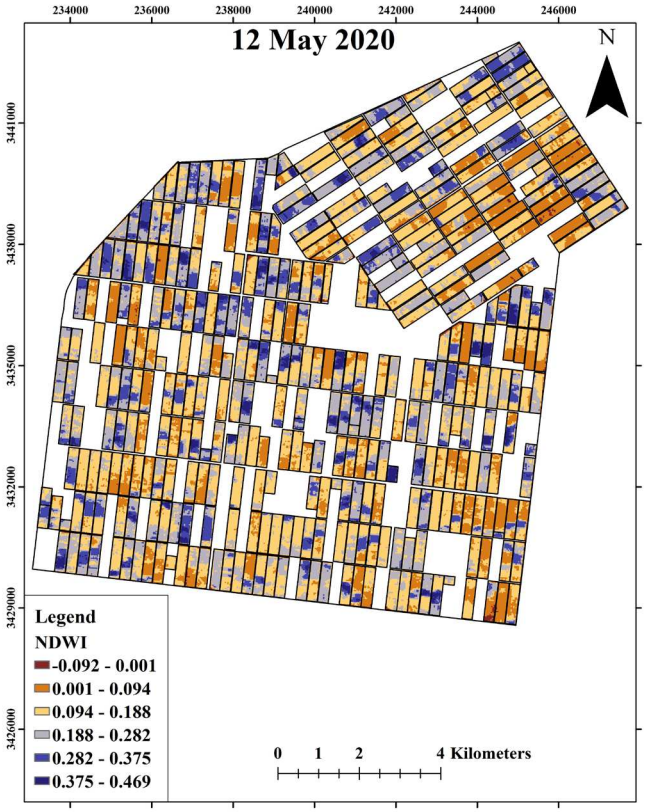
The Normalized Difference Water Index (NDWI) (Gao, 1996) is a satellite-derived index from the Near-Infrared (NIR) and Short Wave Infrared (SWIR) channels. The SWIR reflectance reflects changes in both the vegetation water content and the spongy mesophyll structure in vegetation canopies, while the NIR reflectance is affected by leaf internal structure and leaf dry matter content but not by water content. The combination of the NIR with the SWIR removes

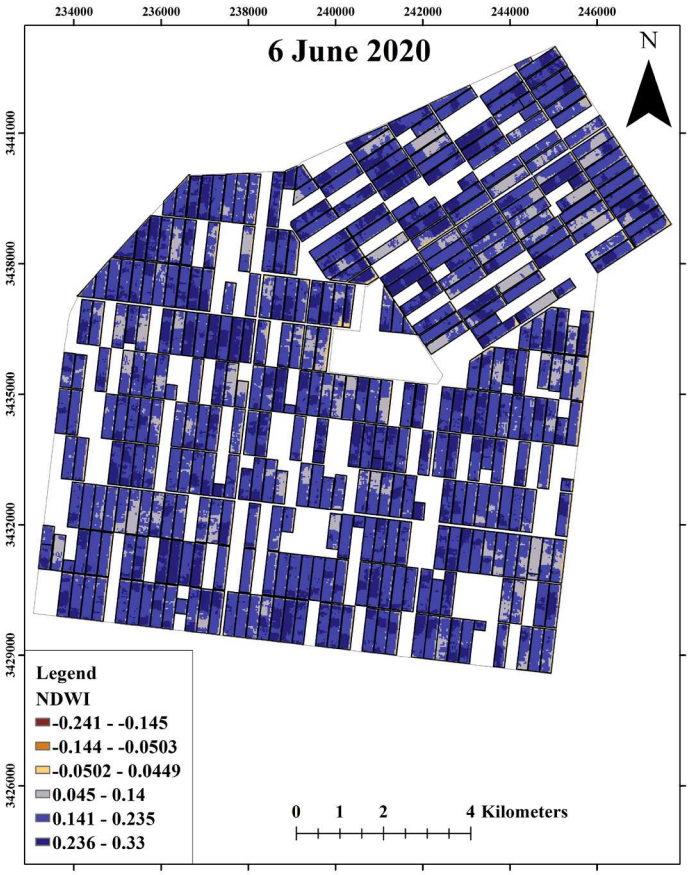
variations induced by leaf internal structure and leaf dry matter content, improving the accuracy in retrieving the vegetation water content (Ceccato et al. 2001). The amount of water available in the internal leaf structure largely controls the spectral reflectance in the SWIR interval of the electromagnetic spectrum. SWIR reflectance is therefore negatively related to leaf water content (Tucker, 1980). Its usefulness for drought monitoring and early warning has been demonstrated in different studies (Gu et al., 2007; Ceccato et al., 2002). It is computed using the near infrared (NIR) and the short wave infrared (SWIR) reflectance (Eq.8), which makes it sensitive to changes in liquid water content and in spongy mesophyll of vegetation canopies (Gao, 1996 ; Ceccato et al., 2001).

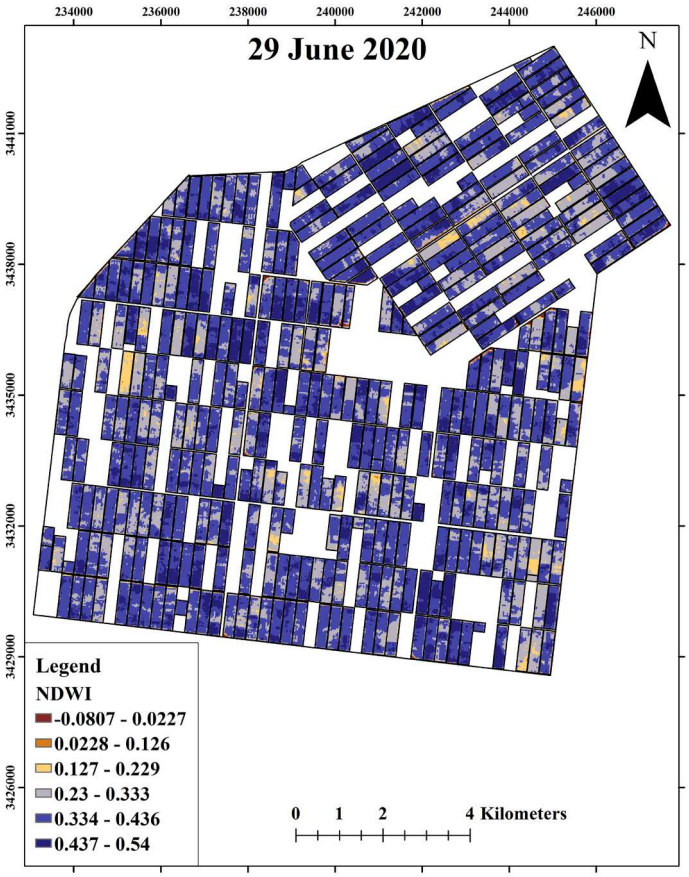
Figure 6 shows the NDWI input images for M5 decision tree. NDWI is considered as wetness index of sugarcane. According to figure 6 NDWI varies in different growth stages and has wide range of variability. In figure 6, 3 April 2020 and 12 May 2020 maximum and minimum amount of NDWI differs spatially but due to early growth stage of the plant most of the cultivated area are dry according to less irrigation water consumption.

Plant wetness status has effect on evapotranspiration, where plant with less wetness is under water deficit stress and has less evapotranspiration and plant with high wetness index has high evapotranspiration and this NDWI variability in different growth stages can obtain high accuracy for evapotranspiration evaluation with data mining.


3 April 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
NDWI
-0.154 - -0.061
-0.061 - 0.031
0.031 - 0.125
0.125 - 0.218
0.218 - 0.312
0.312 - 0.405
0 1 2 4 Kilometers


12 May 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
NDWI
-0.092 - 0.001
0.001 - 0.094
0.094 - 0.188
0.188 - 0.282
0.282 - 0.375
0.375 - 0.469
0 1 2 4 Kilometers


6 June 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
NDWI
-0.241 - -0.145
-0.144 - -0.0503
-0.0502 - 0.0449
0.045 - 0.14
0.141 - 0.235
0.236 - 0.33
0 1 2 4 Kilometers


29 June 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
NDWI
-0.0807 - 0.0227
0.0228 - 0.126
0.127 - 0.229
0.23 - 0.333
0.334 - 0.436
0.437 - 0.54
0 1 2 4 Kilometers

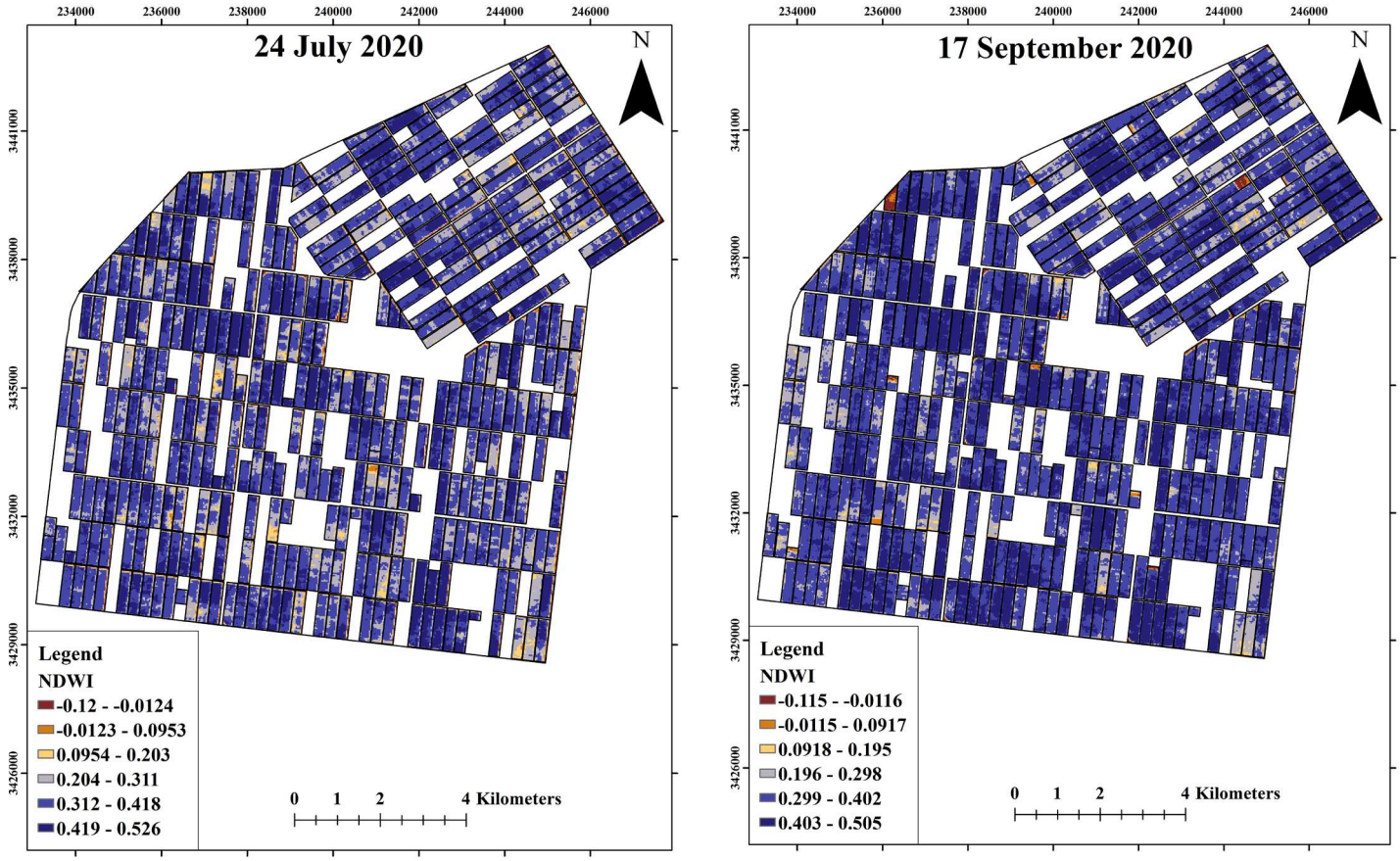


Fig. 6. NDWI input images for M5 decision tree

Also Figure 6 makes clear that in the end sugar cane growth stages most of the cultivated farms have high NDWI index due to the peak of irrigation water requirement of sugarcane.

3.1.2. Output

The values in parentheses under each label in the leaves indicate the number of segments resulting from the corresponding threshold. The second value indicates the number of times a misclassification occurred (Vieira et al., 2012).

Figure 7 shows the decision tree for the evapotranspiration from 3 April 2020 to 17 September 2020. 45 different equations were extracted with a Correlation coefficient of 0.9947, Mean absolute error and root mean squared error of 0.4101 and 0.5705, respectively.

By using fewer input parameters including albedo, emissivity, and NDWI, many evapotranspiration equations were extracted. Figure 4 reveals that the albedo input variable was located at the top of a decision tree which makes clear that albedo has high importance in evapotranspiration estimation based on this decision tree. By considering the geographical location of the study area, there is a high amount of receiving light in this area and the albedo input variable was considered as absorbed light. Therefore, M5 decision tree divisions show that the amount of absorbed light in this area has an important role in the evapotranspiration process which most of the decision tree divisions were based on the albedo. NDWI variable is considered as plant moisture which has an important role in evapotranspiration calculations after albedo variable. This shows that the plant moisture has an important role in extracting the decision tree equations beside the absorbed lights.

The emissivity variable was considered as the diffused light has less importance in the decision tree divisions which by considering the geographical location of the study area it shows that most of the received light was absorbed than diffused.

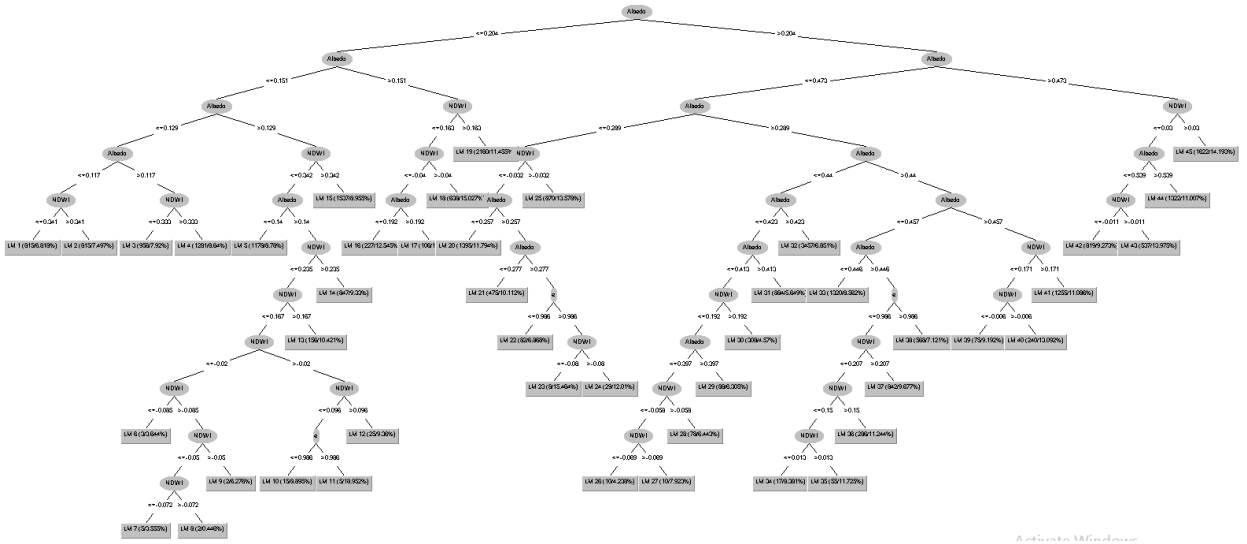


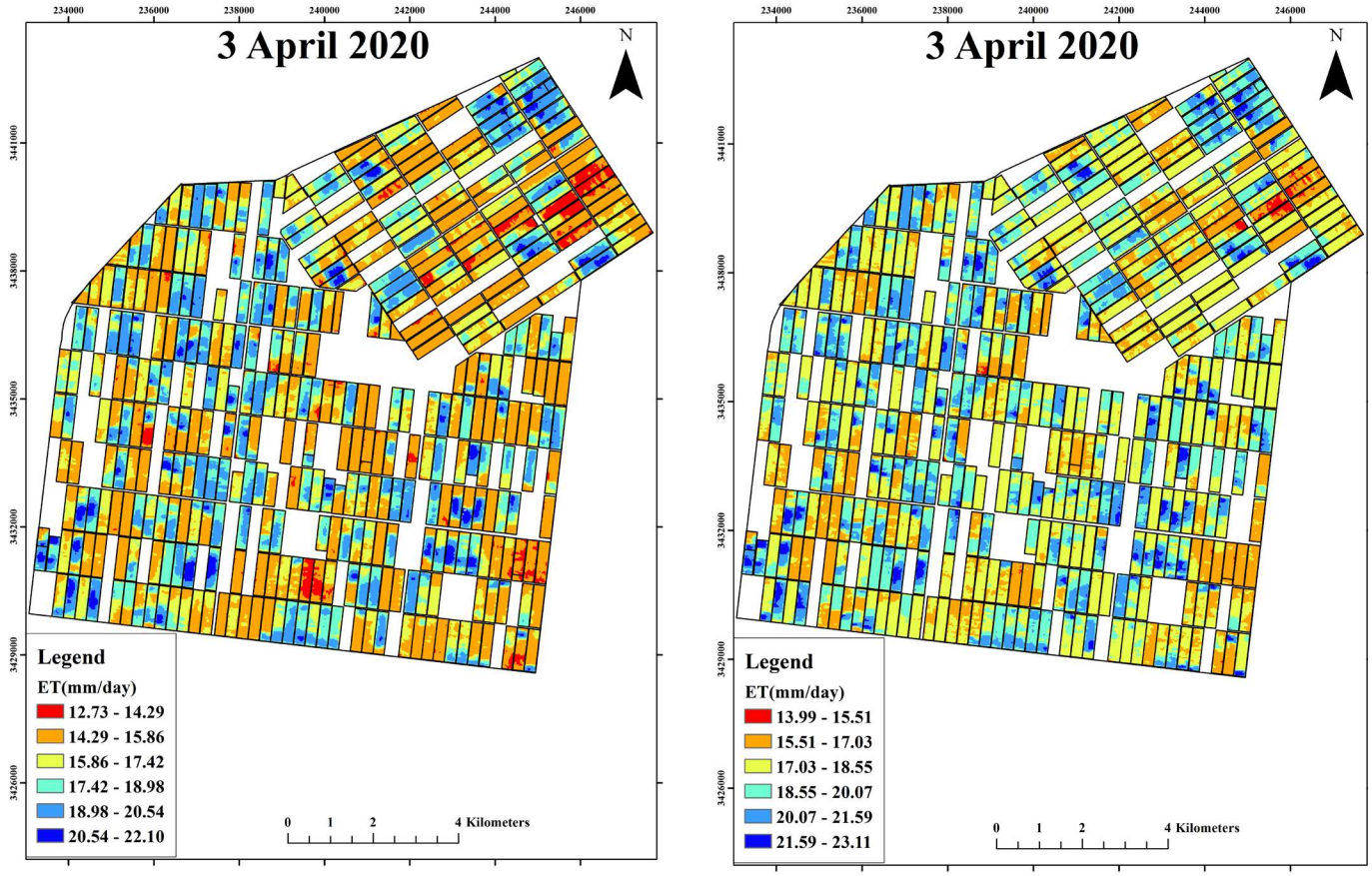
Fig. 7. Decision tree for the evapotranspiration estimation

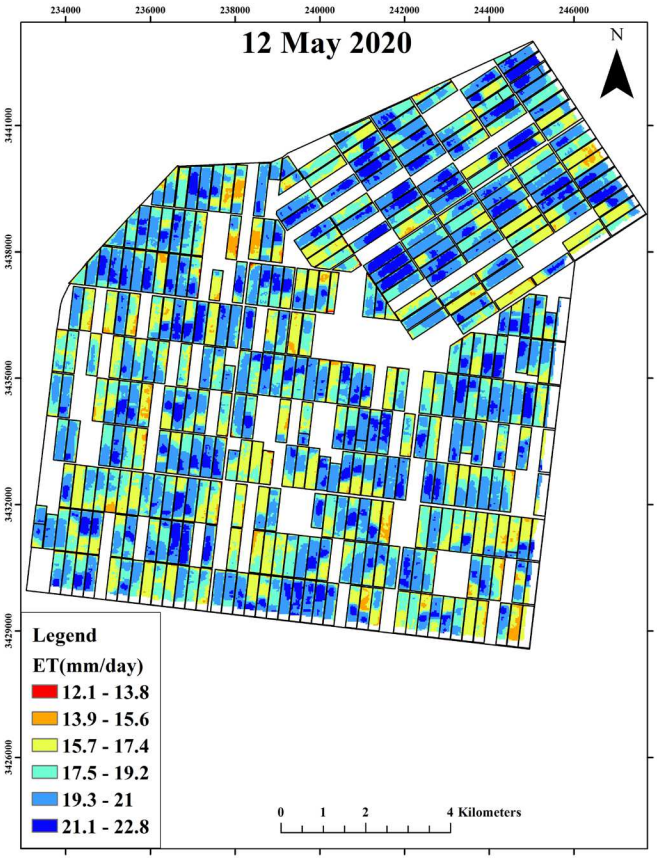
The extracted equations by using M5 decision tree and python scripts in the Arc Map environment are presented as Appendix1 at the end of this article.

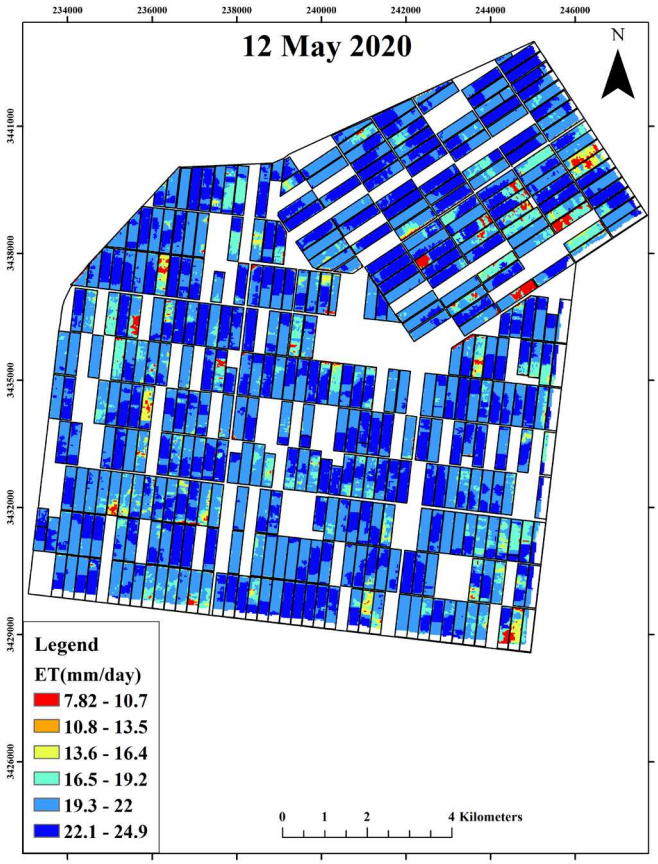
3.2. Combining M5 and GIS

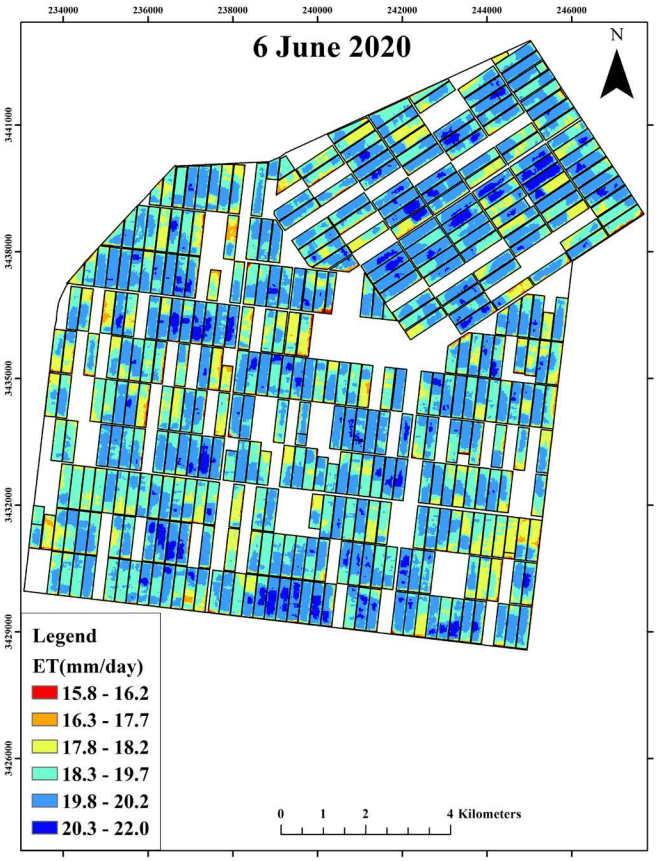
After extracting the most suitable equations from the M5 decision tree model, they were applied by using python scripts for faster and more accurate calculating. The equations were obtained by using evapotranspiration from 3 April 2020 to 17 September 2020 and were applied on input variables of the mentioned period to find if the extracted equations have acceptable model performance. Figure 8 shows the evapotranspiration map calculated by the SEBAL algorithm and M5 decision tree for 3 April 2020 to 17 September 2020. According to figure 8 column a shows the evapotranspiration calculated by using SEBAL algorithm and column b shows the evapotranspiration maps derived by using M5 decision tree algorithm.

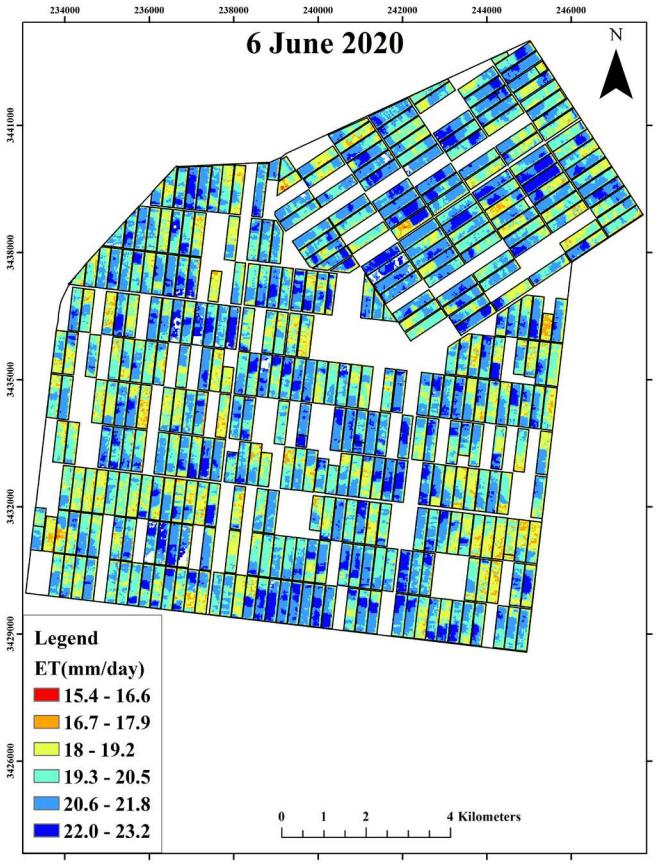
(a) **(b)**

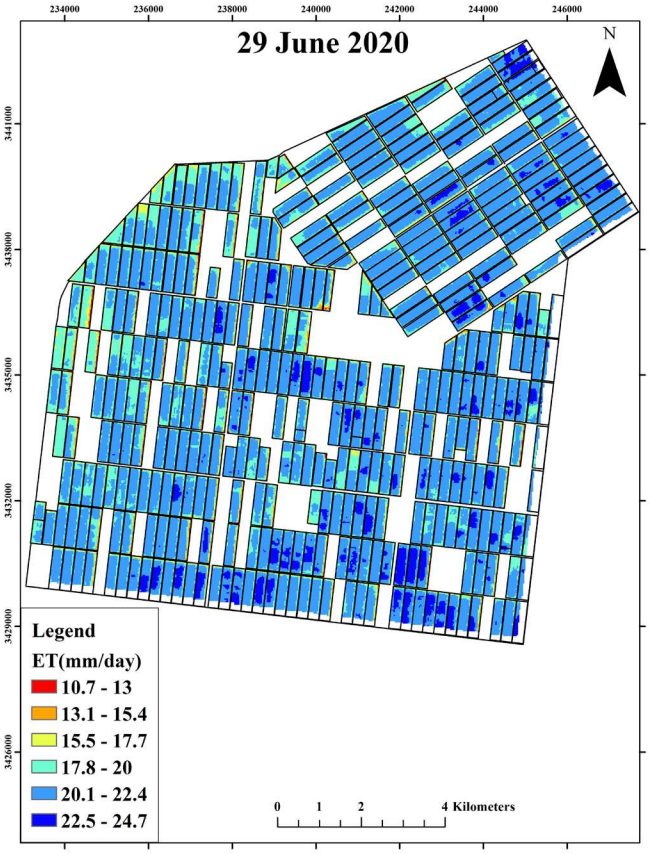

3 April 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
ET(mm/day)
12.73 - 14.29
14.29 - 15.86
15.86 - 17.42
17.42 - 18.98
18.98 - 20.54
20.54 - 22.10
0 1 2 4 Kilometers
3 April 2020
N
Legend
ET(mm/day)
13.99 - 15.51
15.51 - 17.03
17.03 - 18.55
18.55 - 20.07
20.07 - 21.59
21.59 - 23.11
0 1 2 4 Kilometers

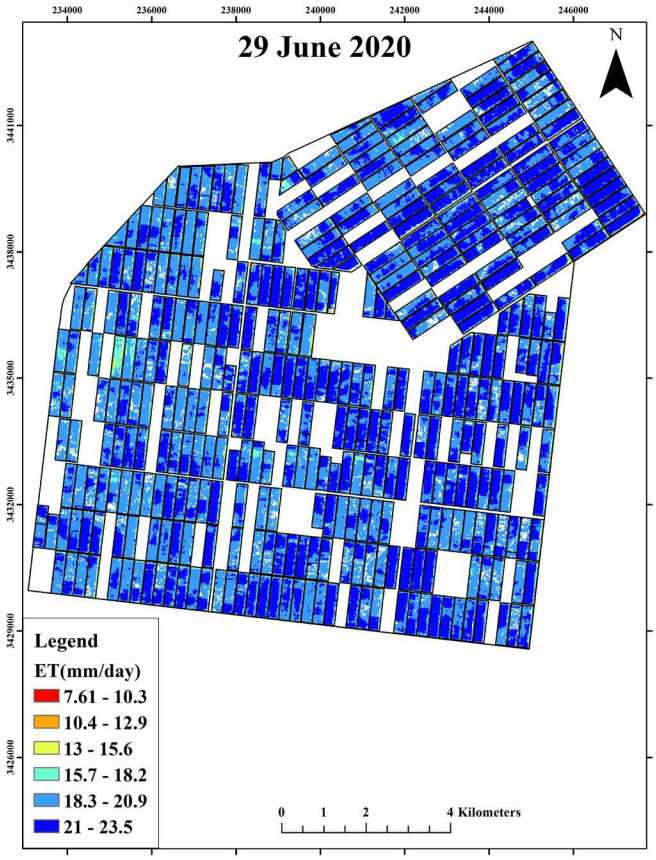

12 May 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
ET(mm/day)
12.1 - 13.8
13.9 - 15.6
15.7 - 17.4
17.5 - 19.2
19.3 - 21
21.1 - 22.8
0 1 2 4 Kilometers

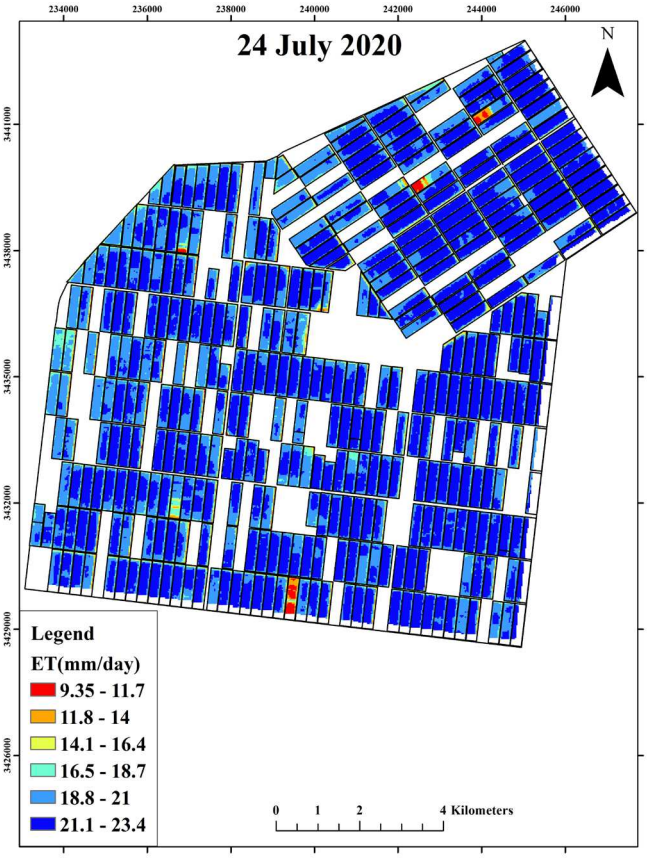

12 May 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
ET(mm/day)
7.82 - 10.7
10.8 - 13.5
13.6 - 16.4
16.5 - 19.2
19.3 - 22
22.1 - 24.9
0 1 2 4 Kilometers

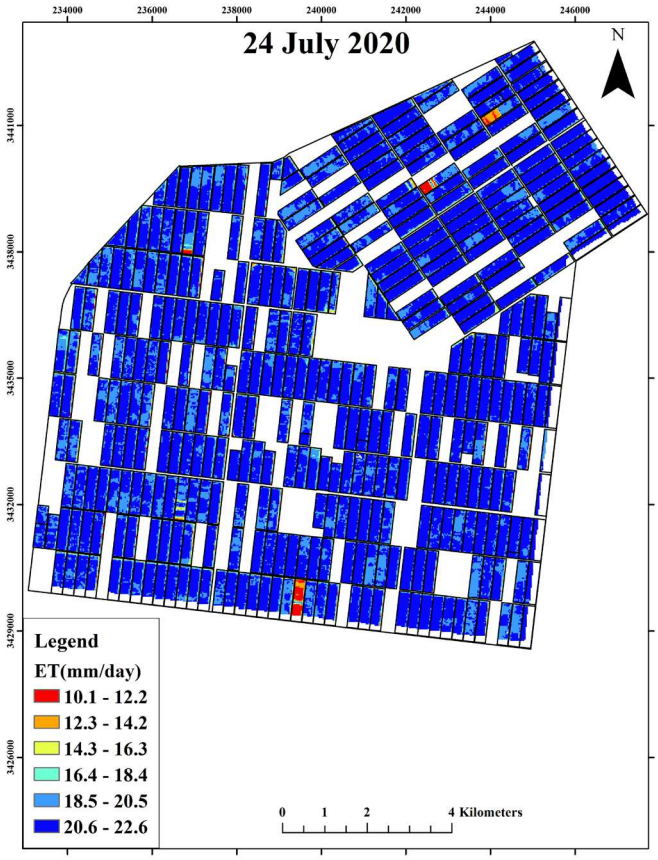

6 June 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
ET(mm/day)
15.8 - 16.2
16.3 - 17.7
17.8 - 18.2
18.3 - 19.7
19.8 - 20.2
20.3 - 22.0
0 1 2 4 Kilometers


6 June 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
ET(mm/day)
15.4 - 16.6
16.7 - 17.9
18 - 19.2
19.3 - 20.5
20.6 - 21.8
22.0 - 23.2
0 1 2 4 Kilometers


29 June 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
ET(mm/day)
10.7 - 13
13.1 - 15.4
15.5 - 17.7
17.8 - 20
20.1 - 22.4
22.5 - 24.7
0 1 2 4 Kilometers


29 June 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
ET(mm/day)
7.61 - 10.3
10.4 - 12.9
13 - 15.6
15.7 - 18.2
18.3 - 20.9
21 - 23.5
0 1 2 4 Kilometers


24 July 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
ET(mm/day)
9.35 - 11.7
11.8 - 14
14.1 - 16.4
16.5 - 18.7
18.8 - 21
21.1 - 23.4
0 1 2 4 Kilometers


24 July 2020
N
234000
236000
238000
240000
242000
244000
246000
3441000
3438000
3435000
3432000
3429000
3426000
Legend
ET(mm/day)
10.1 - 12.2
12.3 - 14.2
14.3 - 16.3
16.4 - 18.4
18.5 - 20.5
20.6 - 22.6
0 1 2 4 Kilometers

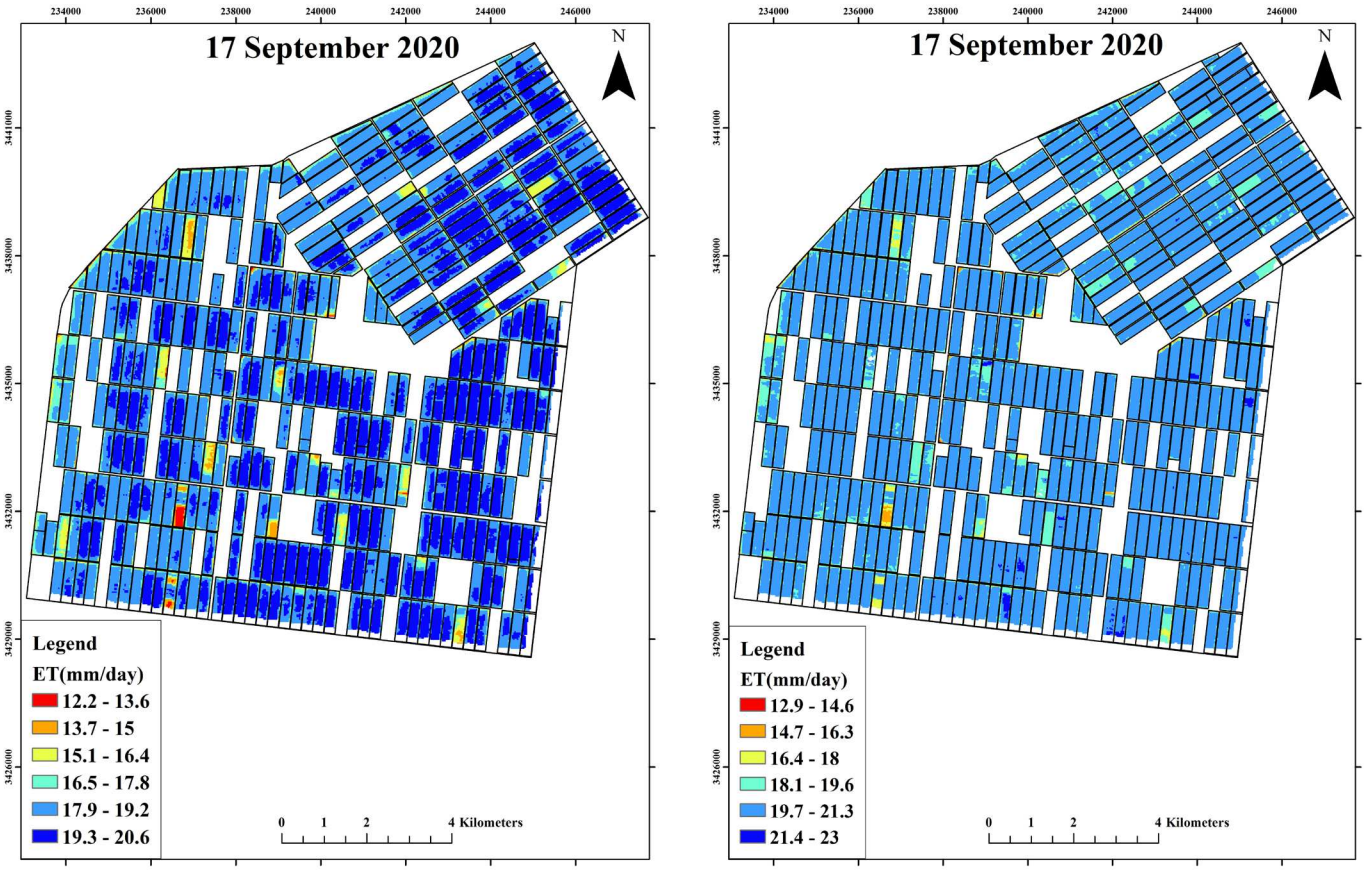
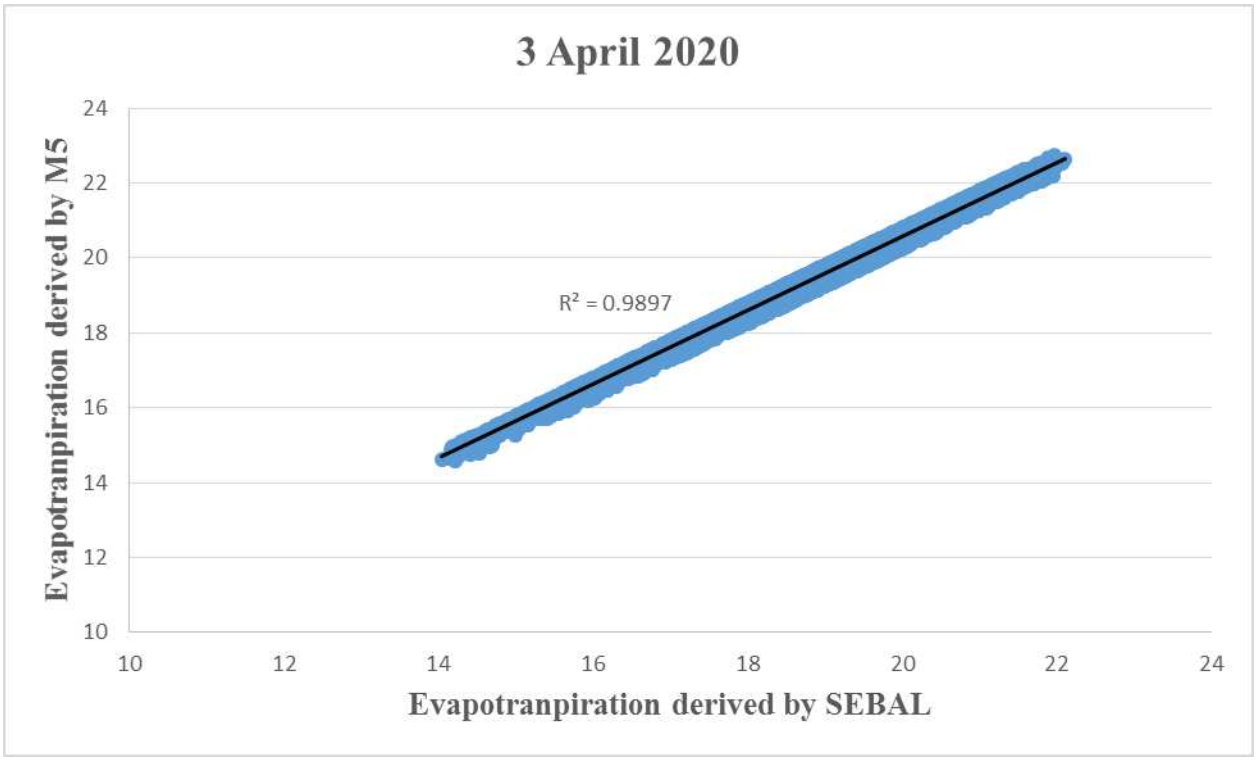
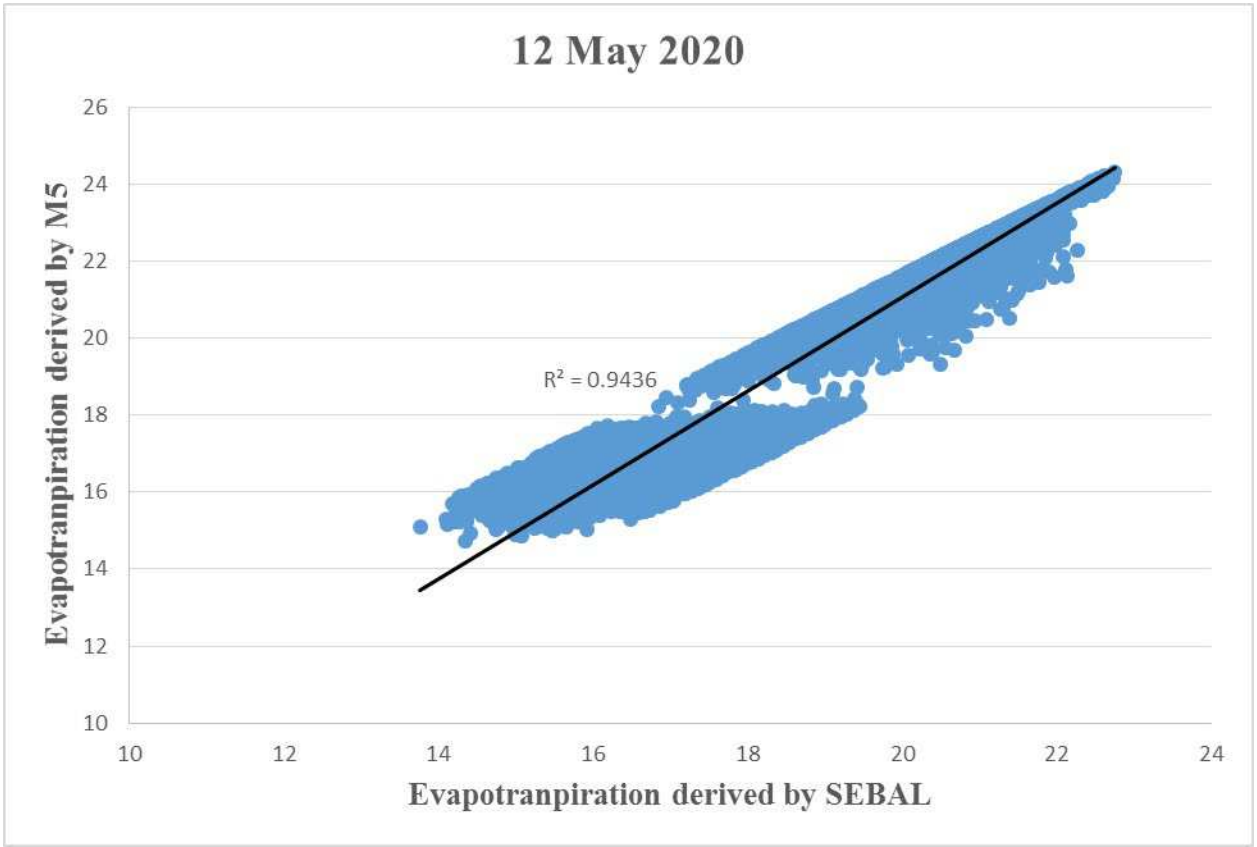
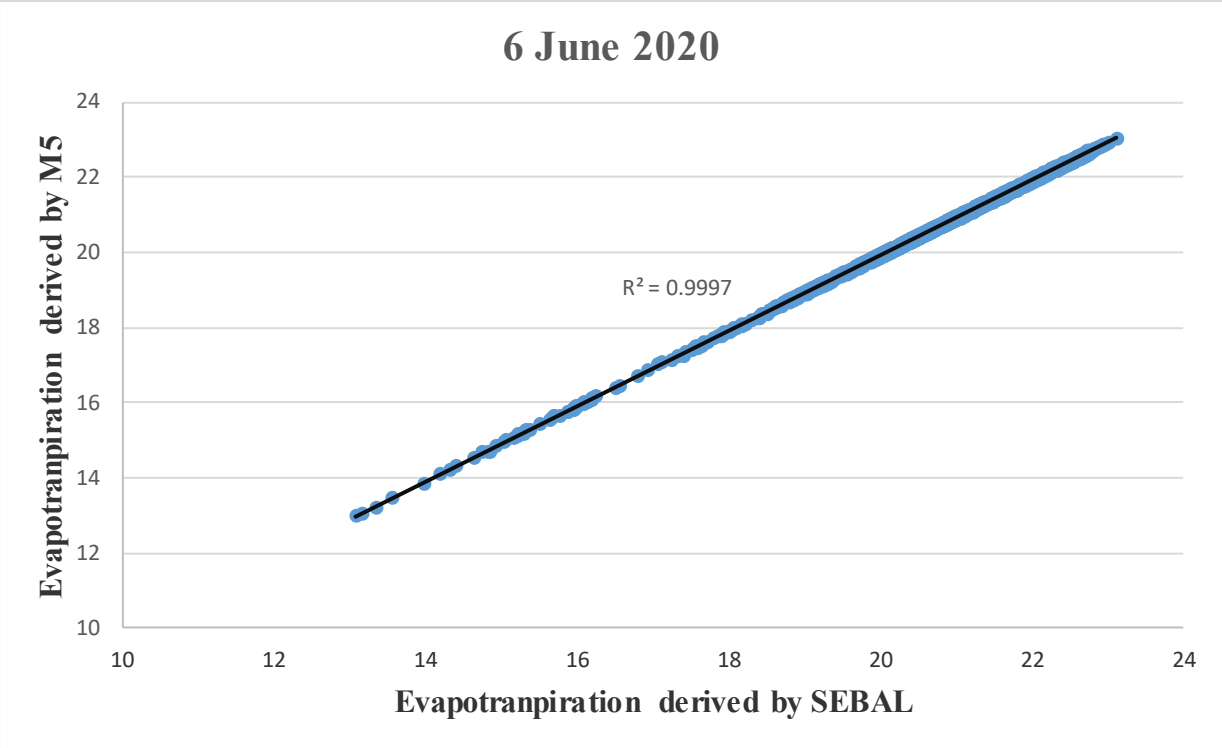


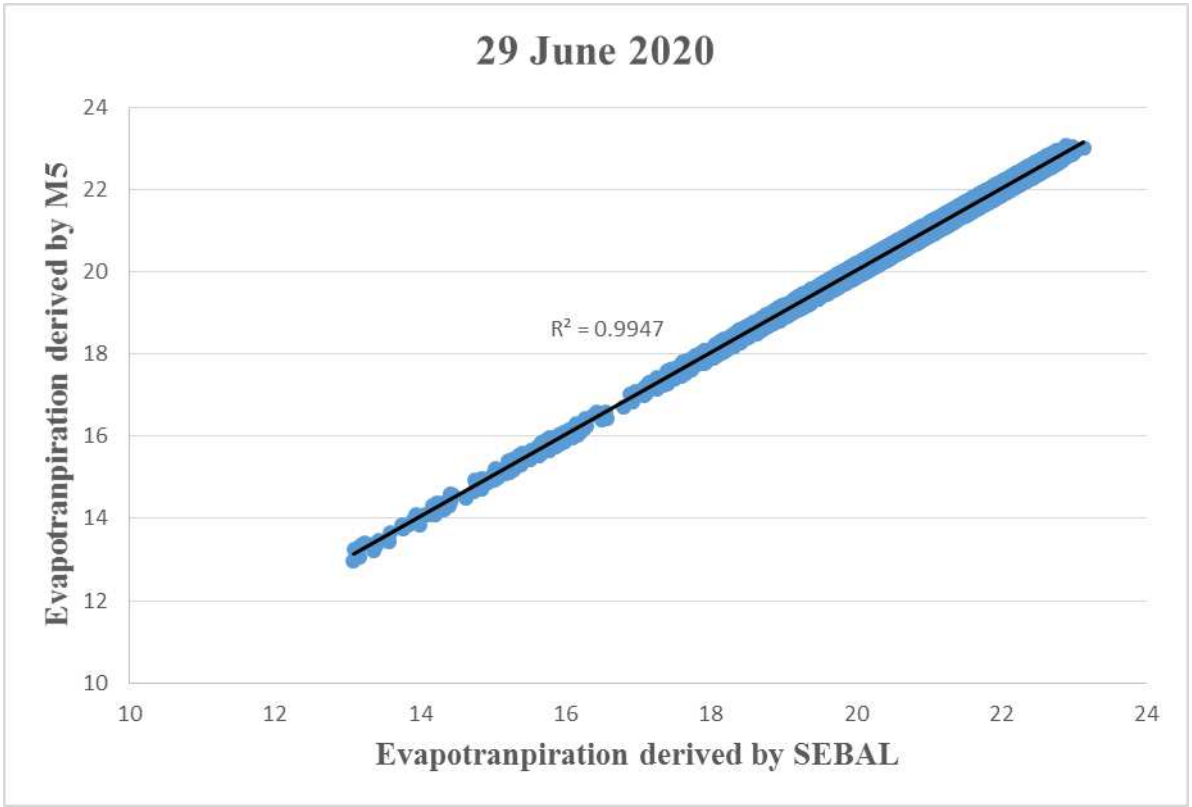
Fig. 8. Evapotranspiration map calculated by SEBAL algorithm for column (a) and evapotranspiration map calculated by M5 decision tree for column (b)

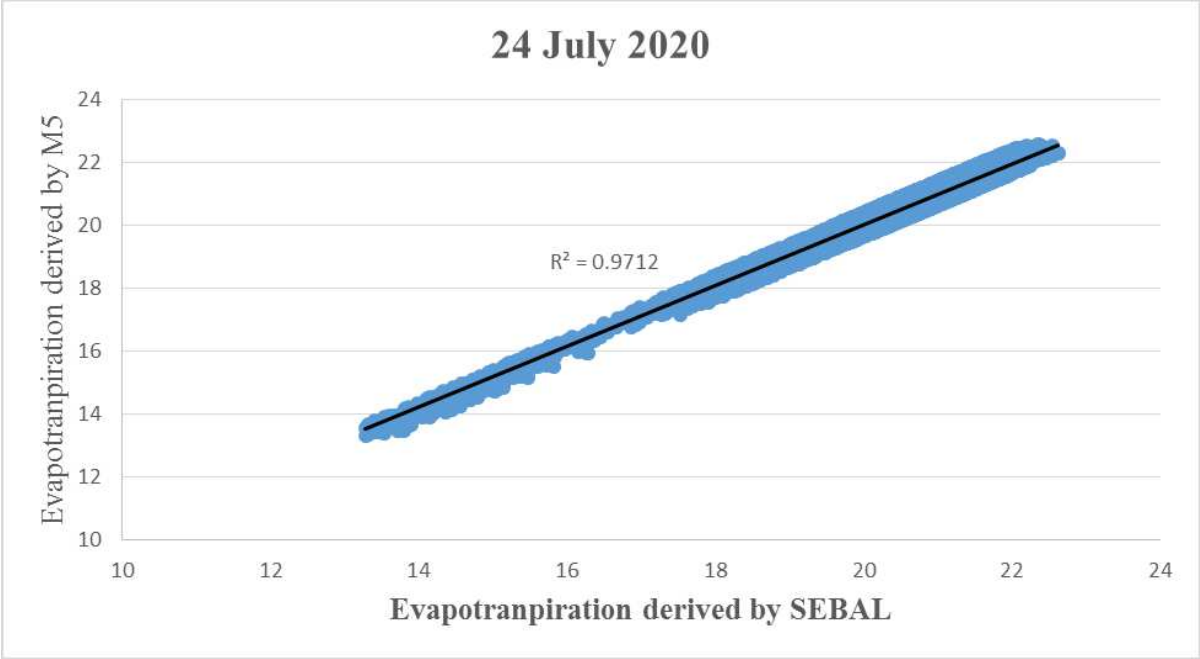
Figure 8 shows the results of the comparison between the SEBAL algorithm and M5 decision tree for this two month study duration. Table 1 shows statistical coefficients for 3 April 2020 to 17 September 2020. According to figure 8 and table 1 by comparing the obtained results for these two months, it could be possible to calculate evapotranspiration with fewer input. Mathematical models can be used instead of physically-based models with acceptable accuracy.

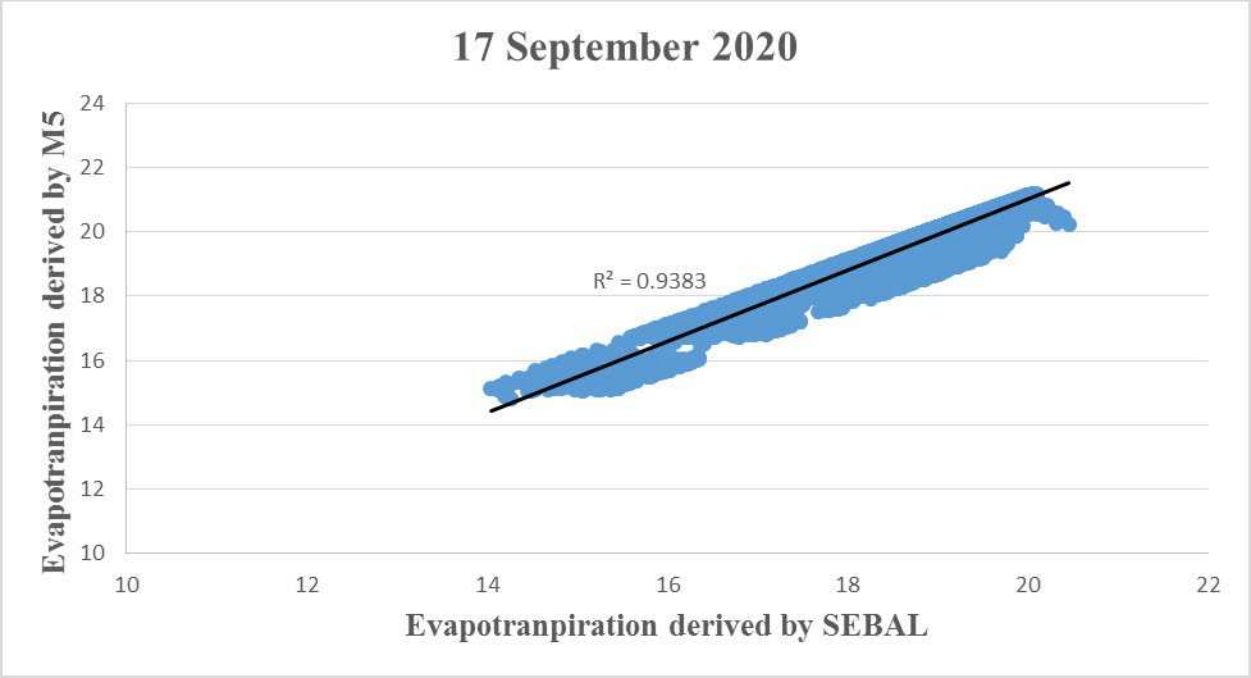

3 April 2020
Evapotranpiration derived by M5
Evapotranpiration derived by SEBAL
R² = 0.9897
24
22
20
18
16
14
12
10
10
12
14
16
18
20
22
24


12 May 2020
Evapotranpiration derived by M5
Evapotranpiration derived by SEBAL
R² = 0.9436
26
24
22
20
18
16
14
12
10
10
12
14
16
18
20
22
24


6 June 2020
Evapotranpiration derived by M5
Evapotranpiration derived by SEBAL
R² = 0.9997
24
22
20
18
16
14
12
10
10
12
14
16
18
20
22
24


29 June 2020
Evapotranpiration derived by M5
Evapotranpiration derived by SEBAL
R² = 0.9947
24
22
20
18
16
14
12
10
10
12
14
16
18
20
22
24





408

409 **Fig. 9.** Comparing the results between the SEBAL algorithm and M5 decision tree

410

411 According to Table 1, the calculated evapotranspiration for 6 and 29 June 2020 has obtained

412 better results than 17 September 2020. This is makes clear that the developed decision tree model

413 can evaluate evapotranspiration in mid-season better than in the first and last stage of the plant
414 growth season.

415

416 **Table 1.** Statistical coefficients for evapotranspiration obtained by M5 decision tree

date	R^2	RMSE	MAE
3 April 2020	0.9897	0.0215	0.0568
12 May 2020	0.9436	0.0178	0.0232
6 June 2020	0.9997	0.0541	0.0587
29 June 2020	0.9947	0.0152	0.0337
24 July 2020	0.9712	0.0881	0.0921
17 September 2020	0.9383	0.0775	0.0350

417

418

419 **4. Discussion**

420

421 Evapotranspiration calculations used for plant water requirements and is a complicated process
422 which depends on so many parameters. Climatological parameters are needed for
423 evapotranspiration calculations and must be measured concisely. By increasing the number of
424 weather stations in a study region, evapotranspiration will be calculated with higher accuracy.
425 Also there are many different algorithms for calculating evapotranspiration in remote sensing
426 and choosing the right algorithm could be a challenge itself. These algorithms have so many
427 different input parameters and this issue could affect the evapotranspiration calculations
428 accuracy. The main aim of this study was to calculate evapotranspiration by using Landsat8
429 satellite images and integrate them with data mining to decrease the number of input parameters
430 for calculating evapotranspiration calculation and also to enhance the resolution of obtained
431 evapotranspiration map from 30 × 100 to 30 × 30. The three input parameters for the developed
432 model were albedo, emissivity and NDWI. In the last section these three parameters were
433 calculated for the study period. According to the input images all three input parameters contain

of acceptable variability and for better performance of data mining the input parameters should have a suitable tolerance spatially and temporally.

The albedo input parameter or reflectivity represent the reflected light from the surface. The reflected light from plant surface is much lesser than the soil surface. In the study region albedo changes temporally and spatially in the cultivated area. In the first stage of plant growth most of the light reflected to the atmosphere due to less developed canopy cover. By considering that the plant need the light for photosynthesis and absorbs the light for this process, the absorbed light effects on evapotranspiration and other life cycle processes of the plant. The albedo images makes clear that during the growth season of the sugarcane the amount of absorbed light increases, also the evapotranspiration increased in the study duration. This is indicate that the absorbed light can effect on the evapotranspiration process.

Emissivity is the other input parameter of the decision tree model which related to LST and can represent the temperature of the land surface cover because temperature of the environment can effect on the plant evapotranspiration. Lansat8 satellite has a thermal band for calculating the land surface temperature with resolution of 100m × 100m, so by using the calculated emissivity instead of land surface temperature this resolution enhanced to 30m × 30m and without using the thermal band and by using data mining decision tree method the resolution of evapotranspiration map increases.

Wetness status of the plant and the environmental tensions can effect on evapotranspiration process. NDWI can represent the water and wetness status of the sugarcane. This water index differ spatially and temporally during the study season which can effect on the evapotranspiration evaluation.

By using albedo as reflected light, emissivity representing the cover temperature and NDWI as the water status of sugarcane, the evapotranspiration evaluated by using decision tree which in the tree divisions started from the albedo input parameter and the albedo is the root of the decision tree. By considering the location of the study area, the absorbed light has a significance role in evapotranspiration evaluation.

Evapotranspiration calculated by using decision tree did not have a significance difference with the evapotranspiration with calculated by using SEBAL algorithm. So by using data mining and less input parameters evapotranspiration evaluated with acceptable accuracy and the resolution of the evapotranspiration map enhanced to a higher resolution.

Also the main equation for calculating evapotranspiration in SEBAL algorithm is: $\lambda ET = Rn - G - H$, which Rn is the net radiation (W/m^2), G is the soil heat flux (W/m^2) and H is the sensible heat flux (W/m^2). In this case albedo is considered as reflected light, emissivity is the land cover temperature and NDWI is the water status of the land cover and the sugarcane. The main equation obtained by the decision tree is: $ET = (a \times NDWI) - (b \times Albedo) \pm (c \times emissivity)$. This equation makes clear that evapotranspiration can be evaluated by using less input parameters and indicate that water status of the plant can effect on evapotranspiration, when the water status is suitable evapotranspiration increases and life mechanisms of the plant enhances. Also by increasing the absorbed light, albedo decreases, hence the absorbed light has direct effect on the plant evapotranspiration. The emissivity depends on land surface cover. Surfaces with soil cover have lower emissivity comparing with plant cover surfaces. In plant cover surfaces with high emissivity, evapotranspiration increases and soil surfaces with lower emissivity, evapotranspiration decreases. So the main obtained equation conforms with the study region

conditions and it is suggested that apply this method to other case studies to find out if this method could be used in a wide region.

5. Conclusion

This study discovered that by using less input parameters and selecting the right parameters, evapotranspiration could be evaluated by using decision tree method and obtain acceptable results. This is makes clear that indirect parameters related to evapotranspiration can be considered as the main input parameters. The emissivity can represent the land canopy temperature and due to low resolution of thermal band of Landsat satellite, the resolution of evaluated evapotranspiration map enhanced to a higher resolution. Also the main equation for calculating evapotranspiration in SEBAL algorithm is: $\lambda ET = Rn - G - H$, which Rn is the net radiation (W/m^2), G is the soil heat flux (W/m^2) and H is the sensible heat flux (W/m^2). In this case albedo is considered as reflected light, emissivity is the land cover temperature and NDWI is the water status of the land cover and the sugarcane. The obtained decision tree equations (Appendix 1) show that the evapotranspiration could be calculated as: $ET = (a \times NDWI) - (b \times Albedo) \pm (c \times emissivity)$. This equation shows that evapotranspiration can be calculated by using this three simple satellite parameters and by subtracting plant water status from reflected light and addition or subtraction of emissivity (depend on land cover condition) with acceptable accuracy.

6.Acknowledgments

We are grateful to the Research Council of Shahid Chamran University of Ahvaz for financial support (GN:SCU.WI98.281).

Ethical Approval

Not applicable

Consent to participate

Consent was obtained from all individual participants included in the study.

Consent to publish

The participant has consented to the submission of the case report to the journal.

Author contribution

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [Lamya Neissi], [Mona Golabi], [Mohammad Albaji] and [AbdAli Naseri]. The first draft of the manuscript was written by [Lamya Neissi] and [Mohammad Albaji] and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This study was funded by "Shahid Chamran University of Ahvaz".

Competing Interests

The authors declare that there are no competing interests.

Availability of data and materials

Data will be made available on request.

Appendix 1. Results of M5 decision tree by using WEKA

=== Run information ===

Scheme: weka.classifiers.trees.M5P -M 4.0

Relation: mor_seb98all

Instances: 27425

Attributes: 4

527 ET

528 e

529 Albedo

530 NDWI

531 Test mode: split 66.0% train, remainder test

532

533 === Classifier model (full training set) ===

534

535 M5 pruned model tree:

536 (using smoothed linear models)

537

538 Albedo <= 0.204 :

539 | Albedo <= 0.151 :

540 | | Albedo <= 0.129 :

541 | | | Albedo <= 0.117 :

542 | | | | NDWI <= 0.341 : LM1 (815/6.818%)

543 | | | | NDWI > 0.341 : LM2 (815/7.497%)

544 | | | Albedo > 0.117 :

545 | | | | NDWI <= 0.333 : LM3 (956/7.92%)

546 | | | | NDWI > 0.333 : LM4 (1281/8.64%)

547 | | Albedo > 0.129 :

548 | | | NDWI <= 0.342 :

549 | | | | Albedo <= 0.14 : LM5 (1178/8.78%)

550 | | | | Albedo > 0.14 :

551 | | | | | NDWI <= 0.235 :

552 | | | | | | NDWI <= 0.167 :

```
553     |  |  |  |  |  |  |  NDWI <= -0.02 :
554     |  |  |  |  |  |  |  |  NDWI <= -0.085 : LM6 (3/3.644%)
555     |  |  |  |  |  |  |  |  NDWI >  -0.085 :
556     |  |  |  |  |  |  |  |  |  NDWI <= -0.05 :
557     |  |  |  |  |  |  |  |  |  |  NDWI <= -0.072 : LM7 (5/3.555%)
558     |  |  |  |  |  |  |  |  |  |  NDWI >  -0.072 : LM8 (2/0.446%)
559     |  |  |  |  |  |  |  |  |  NDWI >  -0.05 : LM9 (2/6.276%)
560     |  |  |  |  |  |  |  NDWI >  -0.02 :
561     |  |  |  |  |  |  |  |  NDWI <= 0.096 :
562     |  |  |  |  |  |  |  |  |  e <= 0.986 : LM10 (15/8.895%)
563     |  |  |  |  |  |  |  |  |  e >  0.986 : LM11 (5/18.952%)
564     |  |  |  |  |  |  |  |  NDWI >  0.096 : LM12 (25/9.36%)
565     |  |  |  |  |  |  NDWI >  0.167 : LM13 (156/10.421%)
566     |  |  |  |  |  NDWI >  0.235 : LM14 (847/9.33%)
567     |  |  |  NDWI >  0.342 : LM15 (1537/8.955%)
568     |  Albedo >  0.151 :
569     |  |  NDWI <= 0.163 :
570     |  |  |  NDWI <= -0.04 :
571     |  |  |  |  Albedo <= 0.192 : LM16 (227/12.545%)
572     |  |  |  |  Albedo >  0.192 : LM17 (106/15.817%)
573     |  |  |  NDWI >  -0.04 : LM18 (638/15.027%)
574     |  |  NDWI >  0.163 : LM19 (2160/11.455%)
575     Albedo >  0.204 :
576     |  Albedo <= 0.473 :
577     |  |  Albedo <= 0.289 :
578     |  |  |  NDWI <= -0.032 :
```

```
|  |  |  |  Albedo <= 0.257 : LM20 (1395/11.794%)
|  |  |  |  Albedo >  0.257 :
|  |  |  |  |  Albedo <= 0.277 : LM21 (475/10.112%)
|  |  |  |  |  Albedo >  0.277 :
|  |  |  |  |  |  e <= 0.986 : LM22 (82/6.868%)
|  |  |  |  |  |  e >  0.986 :
|  |  |  |  |  |  |  NDWI <= -0.08 : LM23 (8/15.464%)
|  |  |  |  |  |  |  NDWI >  -0.08 : LM24 (29/12.01%)
|  |  |  NDWI >  -0.032 : LM25 (870/13.578%)
|  |  Albedo >  0.289 :
|  |  |  Albedo <= 0.44 :
|  |  |  |  Albedo <= 0.423 :
|  |  |  |  |  Albedo <= 0.413 :
|  |  |  |  |  |  NDWI <= 0.192 :
|  |  |  |  |  |  |  Albedo <= 0.397 :
|  |  |  |  |  |  |  |  NDWI <= -0.058 :
|  |  |  |  |  |  |  |  |  NDWI <= -0.069 : LM26 (10/4.238%)
|  |  |  |  |  |  |  |  |  NDWI >  -0.069 : LM27 (10/7.923%)
|  |  |  |  |  |  |  |  NDWI >  -0.058 : LM28 (78/6.443%)
|  |  |  |  |  |  |  Albedo >  0.397 : LM29 (88/6.305%)
|  |  |  |  |  |  NDWI >  0.192 : LM30 (308/4.57%)
|  |  |  |  |  Albedo >  0.413 : LM31 (884/5.649%)
|  |  |  |  Albedo >  0.423 : LM32 (3457/6.851%)
|  |  |  Albedo >  0.44 :
|  |  |  |  Albedo <= 0.457 :
|  |  |  |  |  Albedo <= 0.446 : LM33 (1320/8.562%)
```

```
|  |  |  |  |  Albedo >  0.446 :
|  |  |  |  |  |  e <= 0.986 :
|  |  |  |  |  |  |  NDWI <= 0.207 :
|  |  |  |  |  |  |  |  NDWI <= 0.15 :
|  |  |  |  |  |  |  |  |  NDWI <= 0.013 : LM34 (17/8.381%)
|  |  |  |  |  |  |  |  |  NDWI >  0.013 : LM35 (55/11.725%)
|  |  |  |  |  |  |  |  NDWI >  0.15 : LM36 (286/11.244%)
|  |  |  |  |  |  |  NDWI >  0.207 : LM37 (842/9.677%)
|  |  |  |  |  |  e >  0.986 : LM38 (568/7.121%)
|  |  |  |  Albedo >  0.457 :
|  |  |  |  |  NDWI <= 0.171 :
|  |  |  |  |  |  NDWI <= -0.006 : LM39 (75/9.192%)
|  |  |  |  |  |  NDWI >  -0.006 : LM40 (240/13.092%)
|  |  |  |  |  NDWI >  0.171 : LM41 (1255/11.086%)
|  Albedo >  0.473 :
|  |  NDWI <= 0.03 :
|  |  |  Albedo <= 0.539 :
|  |  |  |  NDWI <= -0.011 : LM42 (819/9.273%)
|  |  |  |  NDWI >  -0.011 : LM43 (537/13.975%)
|  |  |  Albedo >  0.539 : LM44 (1322/11.007%)
|  |  NDWI >  0.03 : LM45 (1622/14.193%)

LM num: 1
ET = 62.4554 * e - 40.9371 * Albedo + 3.5279 * NDWI - 36.812
LM num: 2
ET = 69.1966 * e - 27.917 * Albedo + 5.3461 * NDWI - 45.4675
```

LM num: 3

ET = 0.4648 * e - 51.5918 * Albedo + 3.0445 * NDWI + 25.9698

LM num: 4

ET = 0.4648 * e - 50.1522 * Albedo + 6.1156 * NDWI + 24.8416

LM num: 5

ET = -31.1566 * e - 35.1227 * Albedo + 2.3982 * NDWI + 55.2954

LM num: 6

ET = -66.7867 * e - 80.9495 * Albedo + 15.0298 * NDWI + 96.6672

LM num: 7

ET = -66.7867 * e - 72.1864 * Albedo + 10.3921 * NDWI + 95.0887

LM num: 8

ET = -66.7867 * e - 73.3542 * Albedo + 9.4739 * NDWI + 95.1753

LM num: 9

ET = -66.7867 * e - 73.3542 * Albedo + 16.5157 * NDWI + 95.7477

LM num: 10

ET = -82426.9336 * e - 30.3958 * Albedo + 6.4124 * NDWI + 81296.754

LM num: 11

ET = -123607.0987 * e - 61.18 * Albedo + 6.0168 * NDWI + 121904.8793

LM num: 12

ET = -63.3351 * e - 30.3958 * Albedo + 3.0525 * NDWI + 85.8885

LM num: 13

ET = -8.246 * e - 39.8119 * Albedo + 0.8923 * NDWI + 33.5089

LM num: 14

ET = -0.4868 * e - 39.1115 * Albedo + 2.1969 * NDWI + 25.5099

LM num: 15

ET = -0.3901 * e - 41.7322 * Albedo + 6.8336 * NDWI + 24.2244

LM num: 16

ET = -83.1605 * e - 49.7803 * Albedo + 23.2742 * NDWI + 108.0703

LM num: 17

ET = -87.9133 * e - 8.3548 * Albedo + 2.6068 * NDWI + 102.9809

LM num: 18

ET = -25.3822 * e - 49.3935 * Albedo + 3.8507 * NDWI + 51.0969

LM num: 19

ET = 18.7019 * e - 42.39 * Albedo + 3.6584 * NDWI + 6.5261

LM num: 20

ET = -0.115 * e - 29.1976 * Albedo + 19.9745 * NDWI + 21.8428

LM num: 21

ET = -0.115 * e - 31.297 * Albedo + 19.9953 * NDWI + 22.261

LM num: 22

ET = -0.115 * e - 24.5096 * Albedo + 17.6309 * NDWI + 19.9878

LM num: 23

ET = -0.115 * e - 15.2947 * Albedo + 30.3915 * NDWI + 17.6678

LM num: 24

ET = -0.115 * e - 46.3871 * Albedo + 21.6422 * NDWI + 26.5724

LM num: 25

ET = -0.115 * e - 40.3798 * Albedo + 5.6154 * NDWI + 24.3707

LM num: 26

ET = 96664.8109 * e - 21.6461 * Albedo + 17.9463 * NDWI - 95292.9134

LM num: 27

ET = 4.8502 * e - 15.8725 * Albedo + 17.9463 * NDWI + 12.3265

LM num: 28

ET = 20.024 * e - 28.4493 * Albedo + 12.197 * NDWI + 0.9958

LM num: 29

ET = 0.7932 * e - 34.6546 * Albedo + 0.4622 * NDWI + 24.4417

LM num: 30

ET = 0.6011 * e - 1.2633 * Albedo + 0.3513 * NDWI + 11.2269

LM num: 31

ET = 0.1805 * e - 25.6532 * Albedo + 2.8913 * NDWI + 20.8266

LM num: 32

ET = -4.7214 * e - 37.1299 * Albedo + 1.8818 * NDWI + 30.7137

LM num: 33

ET = 0.393 * e - 33.092 * Albedo + 1.007 * NDWI + 24.0881

LM num: 34

ET = -15.7311 * e - 71.7375 * Albedo + 18.4013 * NDWI + 56.4027

LM num: 35

ET = -5.8303 * e - 12.8159 * Albedo + 3.2245 * NDWI + 20.7655

LM num: 36

ET = -200350.1463 * e - 2.409 * Albedo + 3.9492 * NDWI + 197556.0172

LM num: 37

ET = 1.4974 * e - 23.349 * Albedo + 0.0716 * NDWI + 18.7703

LM num: 38

ET = 168543.6701 * e - 32.0044 * Albedo - 1.6346 * NDWI - 166160.355

LM num: 39

ET = -2.7899 * e - 28.2191 * Albedo + 14.267 * NDWI + 23.4984

LM num: 40

ET = 1.3557 * e - 36.5223 * Albedo + 4.9728 * NDWI + 24.2026

LM num: 41

ET = 1.2782 * e - 41.2456 * Albedo + 0.0655 * NDWI + 27.2176

LM num: 42

ET = -0.1192 * e - 26.2217 * Albedo + 13.76 * NDWI + 19.92

LM num: 43

ET = -0.1192 * e - 26.8974 * Albedo + 14.2941 * NDWI + 20.6799

LM num: 44

ET = -0.1192 * e - 30.449 * Albedo + 21.5442 * NDWI + 22.4928

LM num: 45

ET = -0.1654 * e - 32.2564 * Albedo + 1.5478 * NDWI + 24.0679

Number of Rules : 45

Time taken to build model: 1.44 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.02 seconds

=== Summary ===

Correlation coefficient	0.9947
Mean absolute error	0.4101
Root mean squared error	0.5705
Relative absolute error	8.1455 %
Root relative squared error	10.2988 %
Total Number of Instances	9324

Reference

Alfano , D., Scatteia, L., Cantoni, S., Balat-Pichelin, M., 2009. Emissivity and catalycity measurements on SiC-coated carbon fibre reinforced silicon carbide composite. Journal of the European Ceramic Society. 29(10): 2045-2051.

737 Allen, R., Tasumi, M., Trezza, R., 2002. SEBAL (Surface Energy Balance Algorithms for Land)-Advanced Training
738 and User's Manual-Idaho Implementation, Version 1.0.

739 Anderson, M.C., Allen, R.G., Morse, A., Kustas, W.P., 2012. Use of Landsat thermal imagery in monitoring
740 evapotranspiration and managing water resources. Remote Sens. Environ. 122, 50–65.

741 Bastiaanssen, W., Menenti, M., Feddes, R., Holtslag, A., 1998. A remote sensing surface energy balance algorithm
742 for land (SEBAL). 1. Formulation. Jhyd. 212, 198–212.
743
744 Brutsaert, W., 1975. On a derivable formula for long-wave radiation from clear skies. Water Resour. Res. 11, 742–
745 744.
746
747 Carlson, T.N., Ripley, D.A., 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index.
748 Remote Sens. Environ. 62, 241–252.

749

750 Colaizzi, P.D., O'Shaughnessya, S.A., Evetta, S.R., Mounceb, R.B., 2017. Crop evapotranspiration calculation using
751 infrared thermometers aboard center pivots. Agri. Water Manag. 187, 173–189.

752

753 Diarraa, A., Jarlan, L., Er-Rakic, S., Le Pageb, M., Aouadec, G., Tavernierb, A., Bouletb, G., Ezzahard, J., Merlinb,
754 O., Khabbaa, S., 2017. Performance of the two-source energy budget (TSEB) model for the monitoring of
755 evapotranspiration over irrigated annual crops in North Africa. Agri. Water Manag. 193, 71–88.

756

757 Elnmer, A., Khadr, M., Kanae, S., Tawfik, A., 2019. Mapping daily and seasonally evapotranspiration using remote
758 sensing techniques over the Nile delta. Agri. Water Manag. 213, 682-692. ISSN 0378-3774,
759 https://doi.org/10.1016/j.agwat.2018.11.009.

760 Feng, H., Zou, B., 2019. A greening world enhances the surface-air temperature difference. Sci. Tot. Environ. 658,
761 385–394.

762 Gibert, K., Izquierdo, Joaquí., Sànchez-Marrè, M., Hamilton, S.H., Rodríguez-Roda, I., Holmes, G., 2018. Which
763 method to use? An assessment of data mining methods in Environmental Data Science. Environ. Model. Software.
764 110, 3-27. doi: https://doi.org/10.1016/j.envsoft.2018.09.021.
765
766 Gobbo, S., Lo Presti, S., Martello, M., Panunzi, L., Berti, A., Morari, F., 2019. Integrating SEBAL with in-Field
767 Crop Water Status Measurement for Precision Irrigation Applications—A Case Study. Remote Sens. 11, 2069.

768

769 Gordon, L.J., Steffen, W., Jönsson, B.F., Folke, C., Falkenmark, M., Johannessen, Å., 2005. Human modification of
770 global water vapor flows from the land surface. Proc. Natl. Acad. Sci. U. S. A. 102, 7612–7617.
771
772 Griend, A.A.V.d., Owe, M., 1993. On the relationship between thermal emissivity and the normalized difference
773 vegetation index for natural surfaces. Int. J. Remote Sens. 16, 1119–1131.

774

775 Jaferian, v., Toghraie, D., Pourfattah, F., Ali Akbari, O., Talebizadehsardari, P., 2019. Numerical investigation of
776 the effect of water/Al O nanofluid on heat transfer in trapezoidal, sinusoidal and stepped microchannels. Int. J. Num.
777 Methods for Heat & Fluid Flow. DOI: 10.1108/HFF-05-2019-0377.

778

779 Kamali, M.I., Nazari R., 2018. Determination of maize water requirement using remote sensing data and SEBAL
780 algorithm. Agri. Water Manag. 209, 197-205. ISSN 0378-3774, https://doi.org/10.1016/j.agwat.2018.07.035.
781
782 Kong, J., Hu, Y., Yang, L., Shan, Z., Wang, Y., 2019. Estimation of evapotranspiration for the blown-sand region in
783 the Ordos basin based on the SEBAL model. Int. J. Remote Sens. 40, 5-6, 1945-
784 1965, DOI: 10.1080/01431161.2018.1508919.
785
786 Mhawej, M., Elias, G., Nasrallah, A., Faour, G., 2020. Dynamic calibration for better SEBALI ET estimations:
787 Validations and recommendations. Agri. Water Manag. 230, 105955. ISSN 0378-
788 3774,https://doi.org/10.1016/j.agwat.2019.105955.
789
790 Mira, M., E. Valor, V. Caselles, E. Rubio, C. Coll, J. M. Galve, R. Niclos, J. M. Sanchez, Boluda, R., 2010. "Soil
791 Moisture Effect on Thermal Infrared (8-13-mu m) Emissivity." Ieee Transactions on Geosci. Remote Sens. 48(5),
792 2251-60. doi: Doi 10.1109/Tgrs.2009.2039143.
793
794 Nerry, F., Labed, J., and Stoll, M. P., 1988. Emissivity signatures in the thermal IR band for remote sensing:
795 calibration procedure and method of measurements. Appl. Optics. 27, 758–764.
796
797 Ochege, F.U., Luo, G., Obeta, M.C., Owusu, G., Duulatov, E., Cao, L., Nsengiyumva, J.B., 2019. Mapping
798 evapotranspiration variability over a complex oasis-desert ecosystem based on automated calibration of Landsat 7
799 ETM+ data in SEBAL. GISci. Remote Sens. 56(8), 1305-1332. DOI: 10.1080/15481603.2019.1643531.
800 Quinlan, J.R., 1992. Learning with continuous classes. In: Proceedings of Australian Joint Conference on Artificial
801 Intelligence (Singapore: World Scientific Press). 343–348.
802
803 Rahimikhoob, A., Asadi, M., Mashal, M., 2013. A comparison between conventional and M5 model tree methods
804 for converting pan evaporation to reference evapotranspiration for semi-arid region. Water Resour. Manag. 27(14),
805 4815– 4826.
806
807 Rechid, D., Raddatz, T.J., Jacob, D., 2009. Parameterization of snow-free land surface albedo as a function of
808 vegetation phenology based on MODIS data and applied in climate modelling. Theor. Appl. Climatol. 95, 245–255.
809
810 Salisbury, W., and D'ArÌ´a, D. M., 1992. Emissivity of terrestrial materials in the 8–14 mm atmospheric window.
811 Remote Sens. Environ. 42, 83–106.
812 Singh, R.K., Irmak, A., Irmak, S., Martin, D.L., 2008. Application of SEBAL model for mapping evapotranspiration
813 and estimating surface energy fluxes in south-central Nebraska. J. Irrig. Drain. Eng. 134(3), 273–285.
814
815 Sobrino, J.A., Jiménez-Muñoz, J.C., Paolini, L., 2004. Land surface temperature retrieval from LANDSAT TM5.
816 Remote Sensing of Environment. 90, 434–440.
817 Song, L., Liu, S., Kustas, W.P., Nieto, H., Sun, L., Xu, Z., Skaggs, T.H., Yang, Y., Ma, M., Xu, T., Tang, X., Li, Q.,
818 2018. Monitoring and validating spatially and temporally continuous daily evaporation and transpiration at river
819 basin scale. Remote Sens. Environ. 219, 72–88.
820 Staley, D.O., Jurica, G.M., 1972. Effective atmospheric emissivity under clear skies. J. Appl. Meteorol. 11, 349–
821 356.
822

Vieira, M.A., Formaggio, A.R., Rennó, C.D., Atzberger, C., Aguiar, D.A., Mello, M.P., 2012. Object Based Image Analysis and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. Remote Sens. Environ.123, 553–562.

Wang, Y., Witten, I.H., 1997. Induction of model trees for predicting continuous lasses. In: Proceedings of the Poster Papers of the European Conference on Machine Learning. University of Economics, Faculty of Informatics and Statistics, Prague.

Weng, Q., Karimi Firozjaei, M., Kiavarz, M., Alavipanah, S.K., Hamzeh, S., 2019. Normalizing land surface temperature for environmental parameters in mountainous and urban areas of a cold semi-arid climate. Sci. Tot. Environ. 650, 515–529.

Yilmaz, M.T., Hunt Jr., E.R., Jackson, T.J., 2008. Remote sensing of vegetation water content from equivalent water thickness using satellite imagery. Remote Sens. Environ. 112 (5), 2514–2522. http://www.sciencedirect.com/science/article/pii/ S0034425707004798.

Vargo, J., Habeeb, D., Stone, B., 2013. The importance of land cover change across urban-rural typologies for climate modeling. J. Environ. Manage. 114, 243-252.

Gao, B.-C. 1996. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment 58: 257-266.

Ceccato, P., Flasse, S., Tarantola, S., Jacquemond, S., and Gregoire, J.M. 2001. Detecting vegetation water content using reflectance in the optical domain. Remote Sensing of Environment 77: 22–33.

Tucker, C. J. 1980. Remote sensing of leaf water content in the near infrared. Remote Sensing of Environment 10: 23-32.

Gu, Y., Hunt, E., Wardlow, B., Basara, J.B., Brown, J.F., Verdin, J.P. 2008. Evaluation of MODIS NDVI and NDWI for vegetation drought monitoring using Oklahoma Mesonet soil moisture data. Geophysical Research Letters 35.

Figures

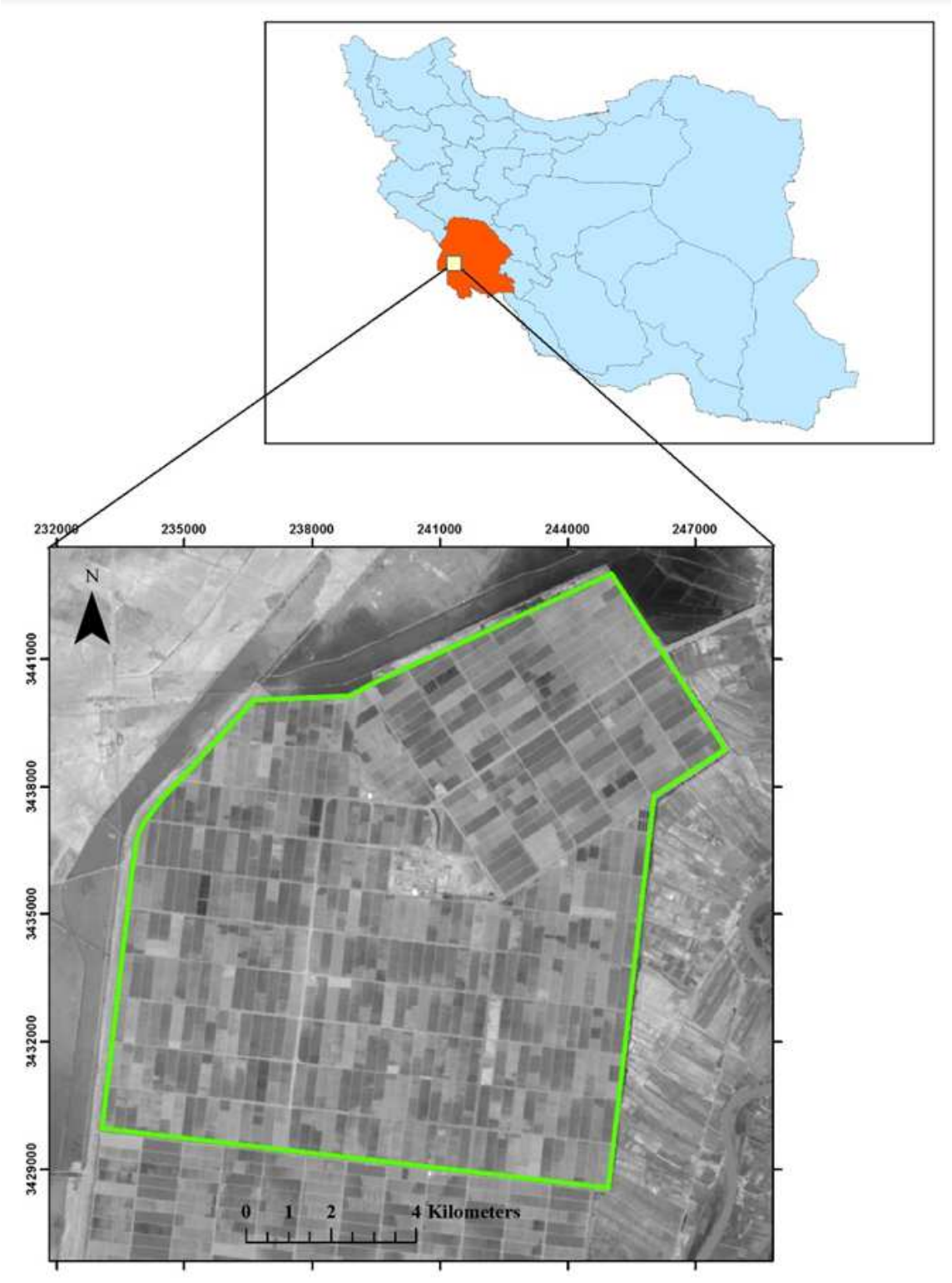


Figure 1

Amir-Kabir sugarcane Agro-Industry location area

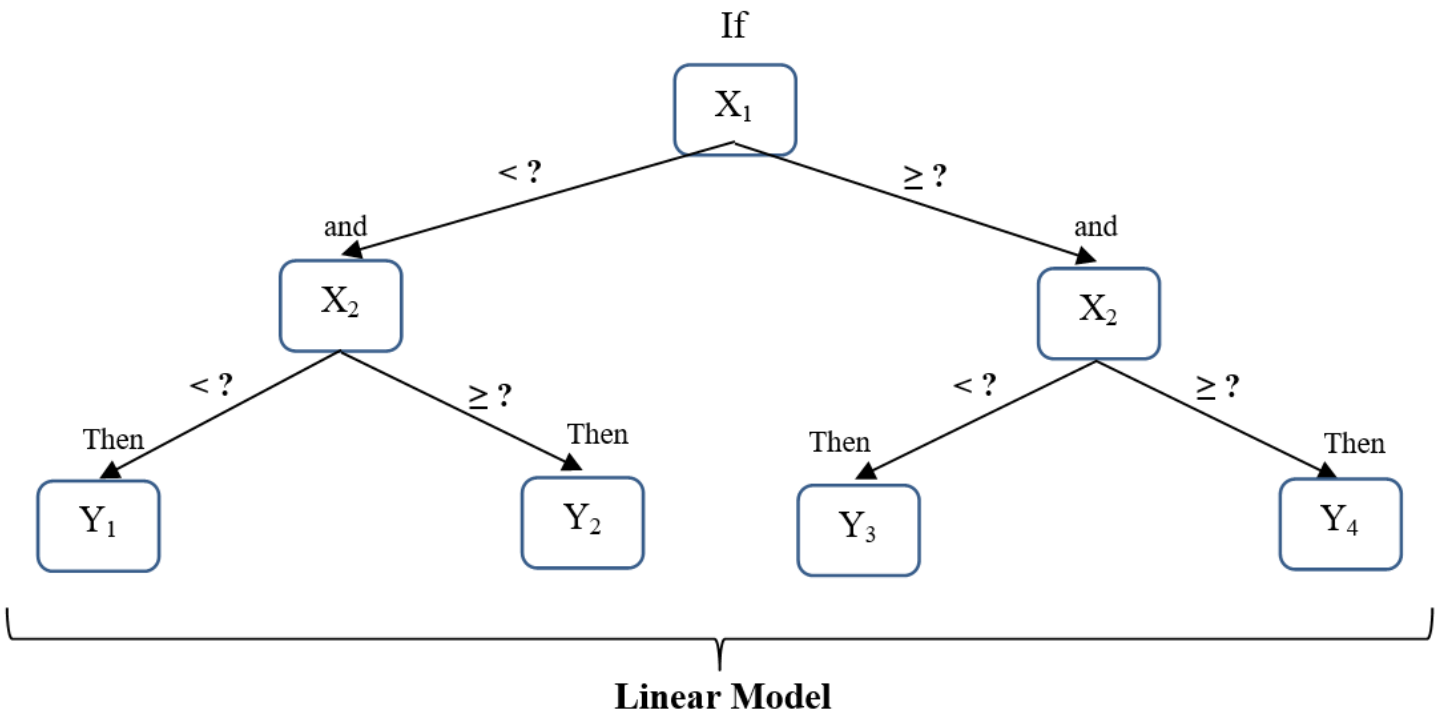


Figure 2

Structure of M5 decision tree (Models Y1–Y4 are linear regression models)

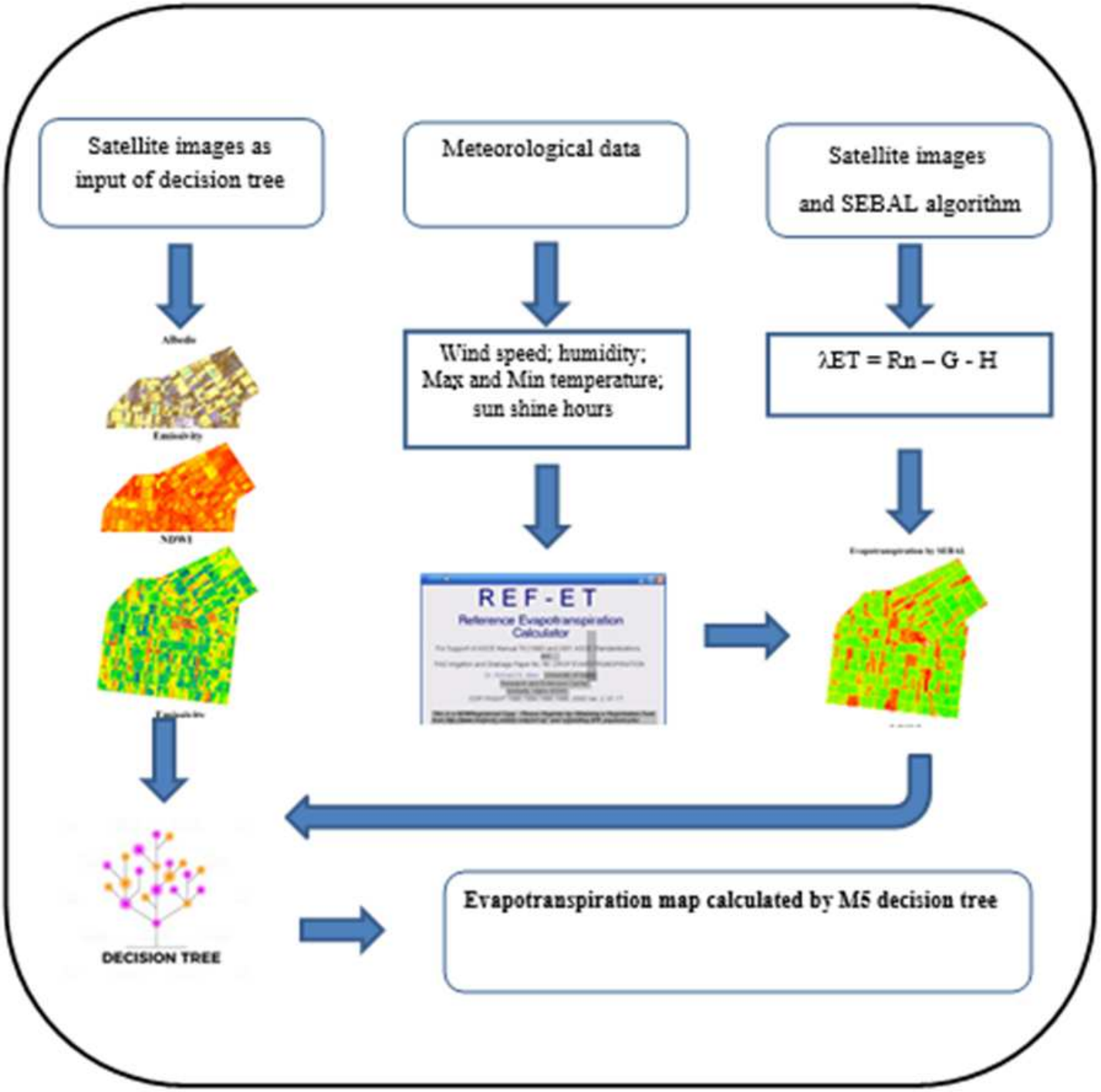


Figure 3

Flowchart of the M5 decision tree and SEBAL algorithm

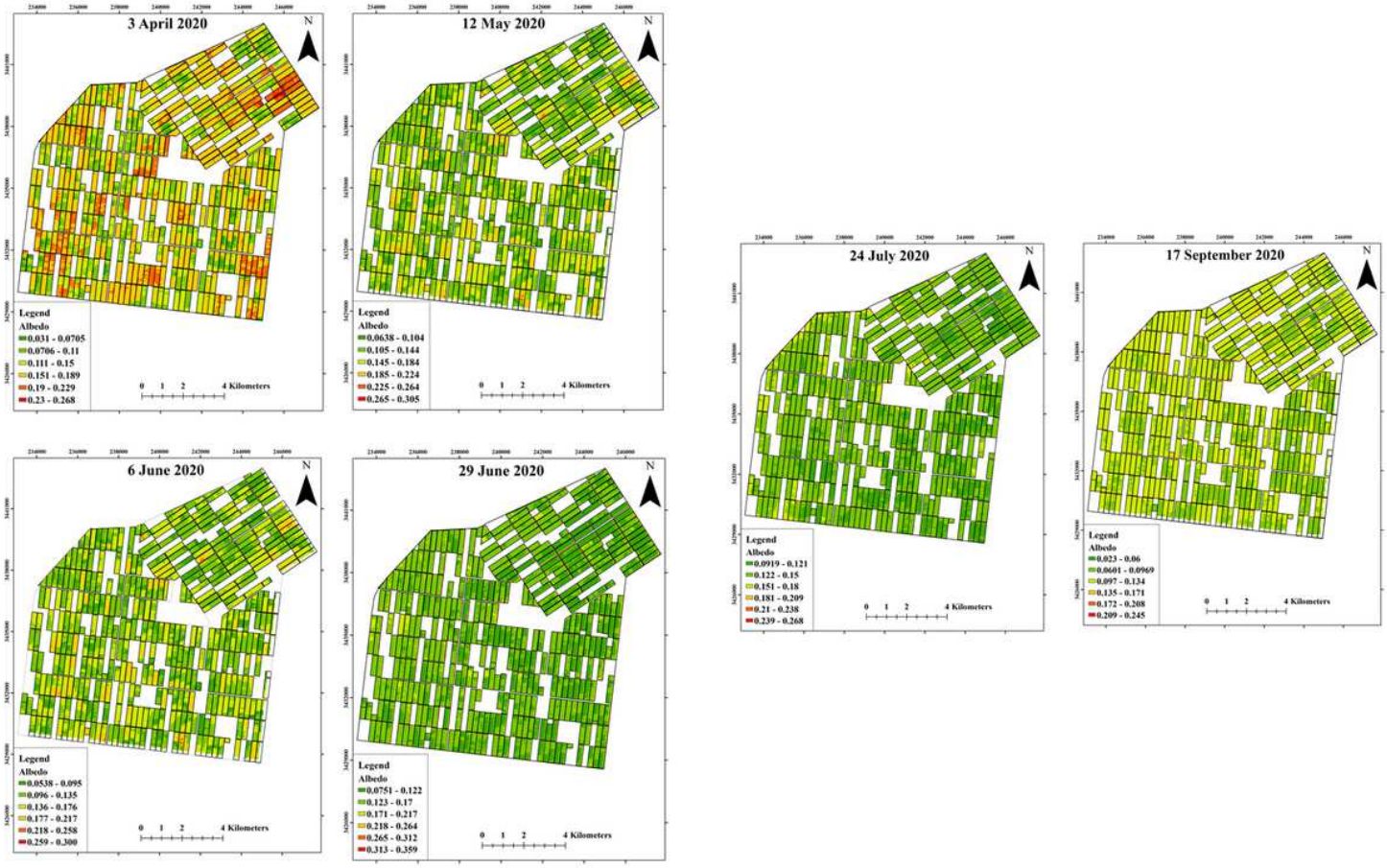


Figure 4

Albedo input images for M5 decision tree

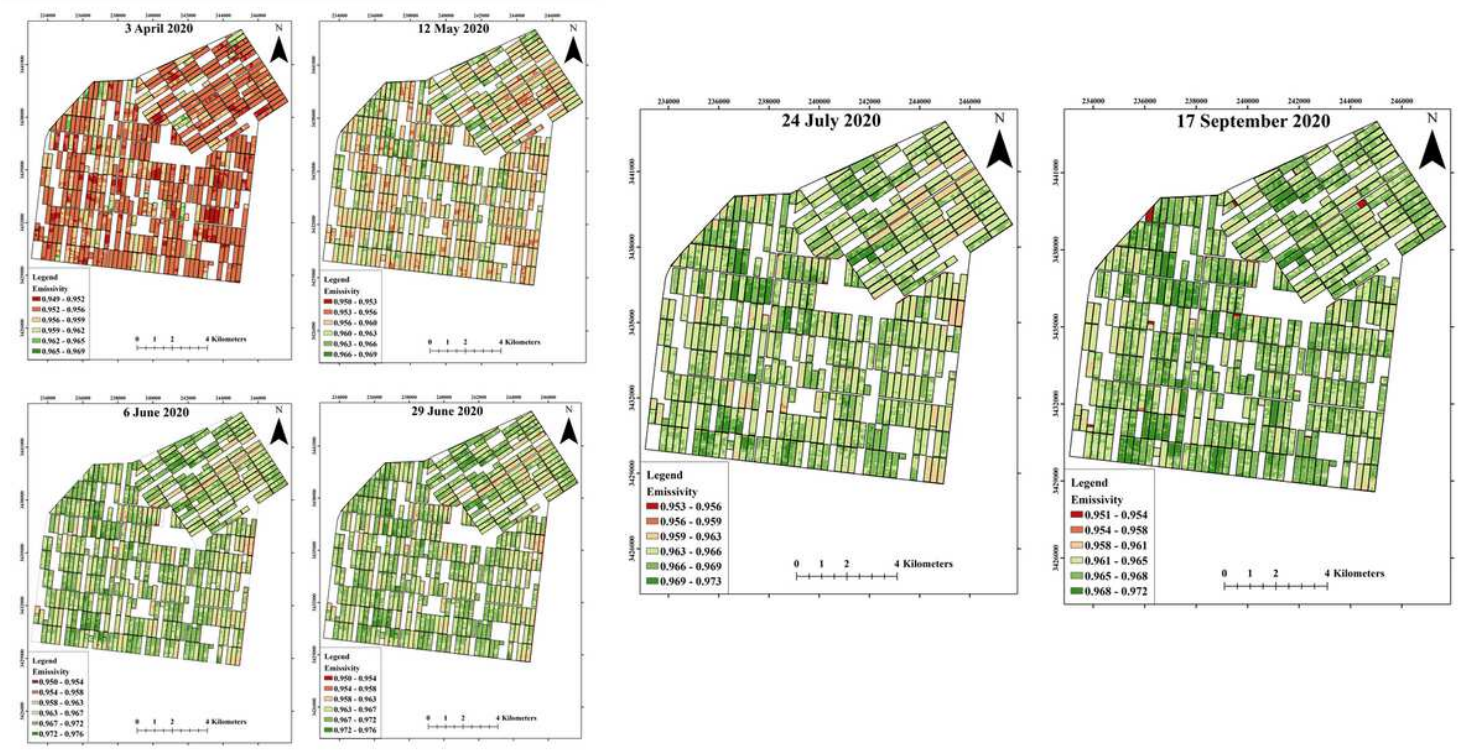


Figure 5

Emissivity input images for M5 decision tree

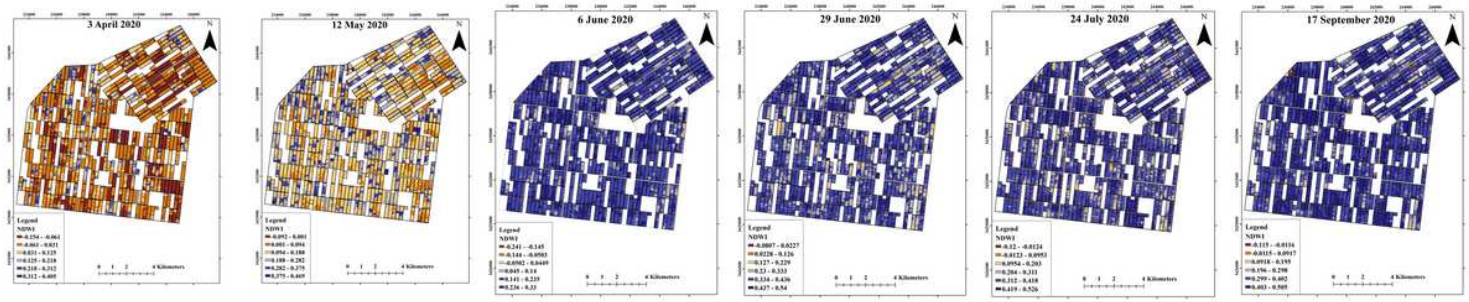


Figure 6

NDWI input images for M5 decision tree

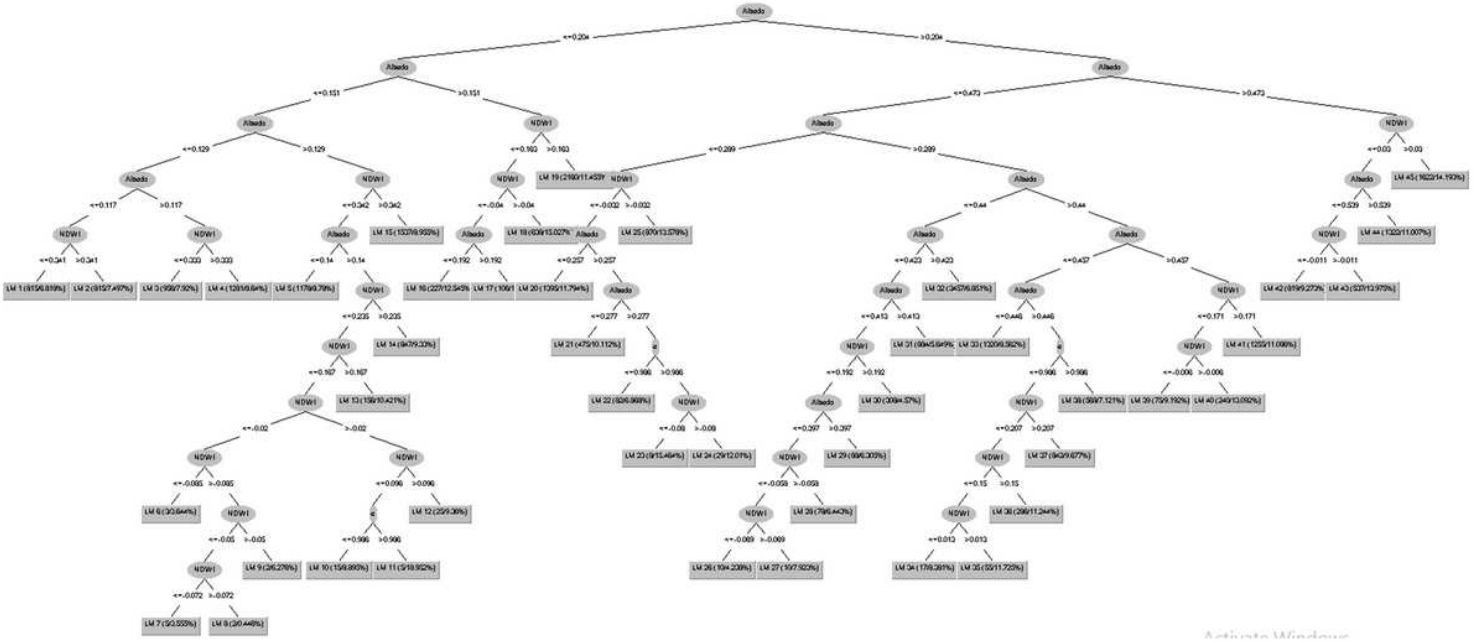


Figure 7

Decision tree for the evapotranspiration estimation

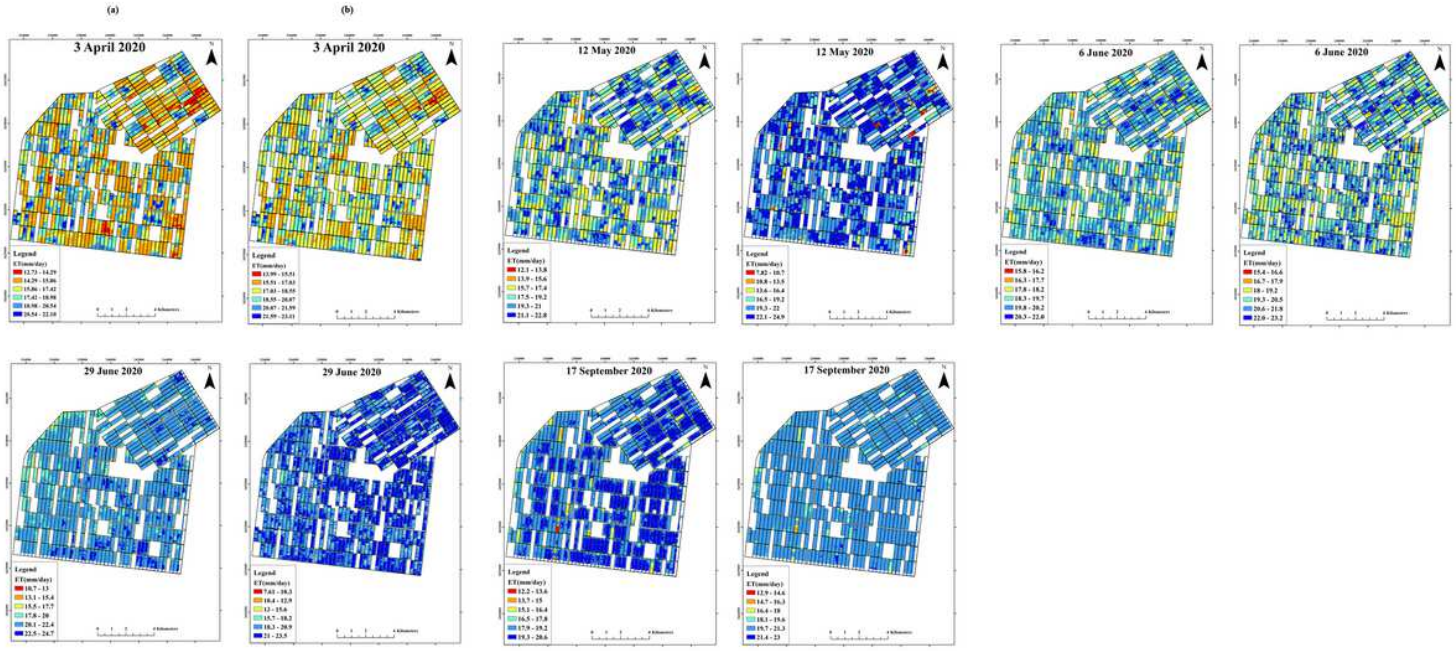


Figure 8

Evapotranspiration map calculated by SEBAL algorithm for column (a) and evapotranspiration map calculated by M5 decision tree for column (b)

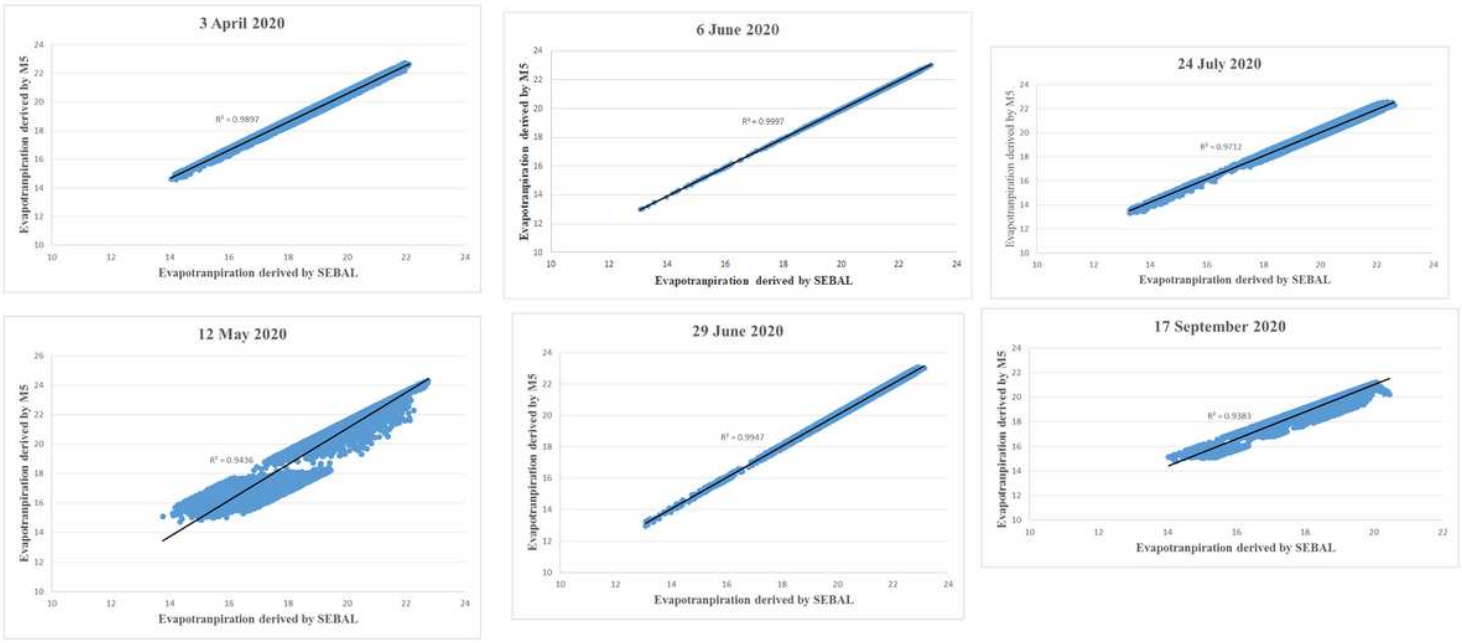


Figure 9

Comparing the results between the SEBAL algorithm and M5 decision tree

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- CombiningadecisiontreewithGIS.docx