DOWNSCALING AND EVALUATION OF EVAPOTRANSPIRATION USING REMOTELY SENSED DATA AND MACHINE LEARNING ALGORITHMS (STUDY AREA: MOGHAN PLAIN, IRAN)

L. Hossein Abadi[1], H. Aghighi[1], A. Matkan[1], A. Shakiba[1*]

[1]Remote Sensing and GIS Research Center, Faculty of Earth Sciences, Shahid Beheshti University, Tehran 653641255, Iran. (mypauk28@gmail.com)

Commission IV, WG IV/3

KEY WORDS: Landsat-8, Modis evapotranspiration product, Random forest regression, Support vector regression.

ABSTRACT:

Water balance estimation in arid and semi-arid areas is highly essential for water and irrigation management. In arid regions, around ninety percent of the rainfall that reaches the surface of earth is returned to the atmosphere by evaporation and transpiration process. Evapotranspiration (ET) estimation has been drastically improved by the help of cutting-edge technology of Remote Sensing (RS) and Machine Learning (ML) techniques. Satellite RS approaches can be advantageous in monitoring land surface processes over vast areas and different approaches have been advanced for assessing ET from moderate to low resolutions with the help of remotely-sensed data. This research demonstrated a MODIS 8-day (500m) ET downscaling technique in Moghan plain based on Landsat-8 indices (30m) and Random Forest Regression (RFR), Support Vector Regression (SVR) models. The outcome of this research showed that SVR outperformed RFR for both days. In SVR model, the accuracy assessment indices on the first and second days are respectively: RMSE= 9.28 and 8.65, rRSME= 27.85 and 63.71, MAE= 5.71 and 3.97. This study has illustrated the possibility of implementing ML methods for downscaling MODIS ET product considering their efficacy and relatively ease of execution. Nevertheless, our research has identified that the MODIS ET accuracy is the primary reason of the accuracy of the downscaled ET. Future research can investigate the utility of spatial-temporal fusion models with remotely-sensed data to ultimately improve the spatio-temporal resolution of downscaled ET maps.

1. INTRODUCTION

Water is the most important limiting factor in agriculture in Iran, and increasing the efficiency of its use in agricultural production is of the highest importance (Hejazizadeh et al., 2017). Since more than 72% of the country's water resources are wasted through evaporation and transpiration (Sharghi et al., 2010) estimating Evapotranspiration (ET), and applying proper management tools to reduce water losses in lands are vitally important. Therefore, in order to properly plan and manage irrigation, in many researches, accurate estimation of water requirement of crops and estimation of reference evapotranspiration are required (Golkar et al., 2008).

The ET process which is undeniably one of the most significant possesses in hydrology cycle, is influenced by lots of factors including sunlight, wind speed, topography, types of vegetation, relative humidity and air temperature. Accurate ET estimation is quite challenging due to the variety of parameters affecting this factor. ET measurement methods are based on various measuring principles. for instance, hydrological approaches such as lysimeter and eddy covariance system. Although these methods can measure ET accurately, they have various limitations. For example, as a consequence of the fast changes in the parameters which were mentioned earlier, such methods are unable to show a good accuracy of spatio-temporal changes in vast areas (Bindhu et al., 2013). The necessity of using models to estimate ET accurately and efficiently in vast fields is so important. During recent decades, with the increasing progress in Remote Sensing (RS) technology, various researchers have used RS methods to estimate ET and its spatio-temporal distribution.

The Moderate Resolution Imaging Spectroradiometer (MODIS) global evapotranspiration products (MOD16) are considered the primordial ET dataset with the interval of 8-day, monthly, and annual (Running et al., 2017). The products were built up and advanced using a Penman-Monteith method driven by MODIS derived surface albedo, Leaf Area Index (LAI), land cover, daily surface meteorological parameters, and the Fraction of absorbed Photosynthetically Active Radiation (FPAR) (Mu et al.,2011).

Although MODIS ET products provide an excellent opportunity to consistently monitor ET over large areas, 1 km spatial resolution is inaccurate and inapplicable for agriculture management purposes due to a high level of spatial heterogeneity in agriculture fields (Ke et al., 2016). In order to use MODIS ET products for agricultural purposes and water resource management, it is necessary to downscale ET products from coarser resolution to higher resolution imagery using various methods such as Machine Learning (ML) approaches. Considering that there is an inverse relationship between the spatial and the temporal resolution due to the limitations related to the design of sensors, obtaining an image with high spatial and temporal resolution is difficult. Particularly, we face a considerable amount of limitation in utilizing Landsat data to monitor and track changes in dynamic environment because of a 16-days temporal resolution and even the possibility of cloudiness of images with the spatial resolution of 30 meters is quite challenging.

The currently used approaches proposed by many researchers especially for downscaling ET involve generating ET from

MODIS and Landsat imagery, and by evaluating the link between the two ET products in order to obtain daily (Hong et al., 2011; Bhattarai et al., 2015), monthly (Singh et al.,2014) or seasonal (Bhattarai et al., 2015) ET at Landsat spatial resolution. The accuracy of downscaled ET is determined by MODIS and Landsat ET calculations which are mainly established by models such as Mapping Evapotranspiration at High Resolution with Internalized Calibration (METRIC) and The Surface Energy Balance Algorithm for Land (SEBAL).

In current study, we have used ML methods for downscaling MODIS ET. It should be noted that ML algorithms have been comprehensively practiced in downscaling land parameters such as latent heat flux (Kaheil et al., 2008) and soil moisture (Im et al., 2016). The main purpose of this research is generating more accurate downscaled 8-day ET maps at 30 m resolution. We investigated two ML methods including Random Forest Regression (RFR) and Support Vector Regression (SVR). With the intention of downscaling MODIS 1 km 8-day ET product (MOD16A2), eight 30-m Landsat-8 indices, have been used. Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), Land Surface Temperature (LST), Normalized Difference Moisture Index (NDMI), Normalized Difference Water Index (NDWI) Enhanced Vegetation Index (EVI), Normalized Difference Infrared Index—Landsat 8 OLI band 7 (NDIIb7), and Modified Soil Adjusted Vegetation Index (MSAVI), were derived from Landsat-8. Among these eight indices, LST has been implemented to estimate ET because it is highly sensitive to any humidity changes. Vegetation indices, which include NDVI, EVI, SAVI and MSAVI, are related to the condition of vegetation in different growth stages. Also, indices related to plant water content, such as NDWI, NDMI and NDIIb7, whose changes are reflected in the ET process, have been used.

Using vegetation indices for estimating actual ET especially in dry areas is quite beneficial. With the help of the information it provides monitoring and managing land and water in areas where lack in-situ data can be well-managed (Abbasi et al., 2021). Infrared and visible bands, which are used to calculate vegetation indices, have finer spatial resolution compared with the thermal band. Also, vegetation indices provide coherent information about plant growth conditions and physiological processes, which is so useful for estimating ET (Nagler et al., 2009).

NDVI, SAVI, and EVI indices are substantially used in many researches (Nagler et al., 2009; Glenn et al., 2010; Murray et al., 2009). For example, the NDVI index determines vegetation by measuring the difference between near-infrared and red bands. Considering the fact that vegetation has the highest absorption and the lowest reflectance in the red band and the lowest absorption and the highest reflectance in the near-infrared band (Glenn et al., 2011). Also, the NDWI index is pretty sensitive to any changes in the water content of leaves and it is a good indicator for water content in vegetation canopies and it is useful in detecting water stress in vegetation (Wang et al., 2008). Therefore, the purposes of this study were 1) to identify the most influential parameters (feature selection) using Random Forest (RF) algorithm 2) to downscale MODIS ET product by implementing ML algorithms using the most important Landsat-8 indices.

2. MAIN BODY

2.1 Material and Methods

2.1.1. Study area

The studying area is Agro-Industrial and Animal Husbandry Company (MAIAHC) which has been built since 1975 and is in the small part of the Moghan plain. Moghan plain is positioned in the north west of Iran and is almost 28,000 hectare's containing many irrigated and rained farmlands. Aras River, which is along the border of Iran and Azerbaijan, is responsible for irrigation of Company's lands including maize fields as well (Aghighi et al., 2018). This area is situated between latitude of 39.465∘N and latitude of 39.615∘N and longitude of 47.548∘E and longitude of 48.009∘E.
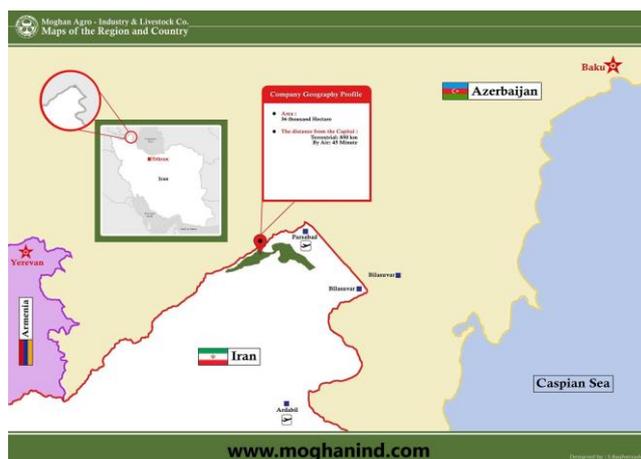
Figure 1. Location of the Moghan plain.
2.1.2. Image Pre-processing and Feature Selection

Two Landsat-8 surface reflectance tier 1 images with minimum cloud cover (<5% withing the Moghan plain) were acquired from Google Earth Engine, including 15-8-2017 and 31-8-2017. MODIS ET product (MOD16A2), includes global ET, potential ET (PET), potential LE (PLE) datasets, the quality control flag of ET (ET_QC), and latent heat flux (LE) for each 8-day interval at 500 m resolution (Mu et al., 2013). MOD16A2 ET is the sum of ET during each 8-day time periods within a year (mm/8 days, mm/5 days or mm/6 days for 361 composite data). In this research, 13-8-2017 and 29-8-2017 Landsat-8 image dates were chosen. ET and ET_QC layers were sectioned within the Moghan plain, and we have used the quality flag layer to eliminate low-grade ET. After that ET products were re-projected to be the exact projection system with Landsat-8 datasets (UTM WGS 1984).

At first, cloud and cloud shadows were masked by cloud Shadow BitMask and clouds BitMask. vegetation indices were derived from Landsat-8 images and surface reflectance of bands 1–7 were used to compute vegetation indices and a brightness temperature of band 10 was used to create LST map. All indices which were derived from Landsat-8 were aggregated and resampled to 480 m. MODIS ET product (MOD16A2) were also resampled from 500 m to 480 m. after that, by implementing RF algorithm a feature selection were done in order to select the most important variables in downscaling ET among all indices.

2.1.3. Machine learning-based downscaling models and Models assessment

Two ML algorithms, SVR and RFR were analysed in this research. SVR is the most frequently used type of Support

vector machines (SVMs) which has been suitable for regression problems especially in remote sensing application. There are Some principles governing SVM. One of them is that independent variables are framed to a higher dimensional space and under this circumstance by the help of kernel functions, nonlinear correlations are qualified to be linearly distinguishable. For maximizing the margin between different classes, the optimal linear separator is also needed (Russell, S et al., 2010). SVR follows the same principle as SVM and it generates real numbers as output given a set of independent input variables by defining a margin of tolerance. Another algorithm, which is numerously used in RS application, is RF and is based on a classification and regression trees (CART) technique. Comparing to the CART method, for improving model performance RF uses a bootstrap aggregating method (Breiman, L. 2001).

In order to optimize the number of indices, with the aim of choosing the most influential ones to build a model with an improved accuracy, ten-fold cross validation was used. by using the most influential indices as predictor variables and MODIS ET product as a response variable, the RFR and SVR prediction model were built and then applied to the 30 m Landsat most important indices in order to create downscaled ET 30 m product. To examine the accuracy of the models, three statistical criteria including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and relative RMSE (rRMSE) were used. Statistical analysis and modelling were conducted using R STUDIO and R version 4.0.5.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{y}_t - y_t \right| \qquad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i}^{n}(\hat{y}_t - y_t)^2}{n}} \qquad (2)$$

$$RRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(H_d^{i,m} - H_d^{i,s})^2}}{\sum_{i=1}^{n} H_d^{i,m}} \times 100 \qquad (3)$$

2.2. Results and Discussion

2.2.1. Feature selection results using random forest algorithm

In order to select the most important parameters, RF model with different mtry values was built and then by comparing the RMSEs, the mtry value with the smallest RMSE value was selected. The best models were chosen with the mtry which showed the lowest RMSE among other mtry values. For the first day the model was built with mtry=3 which had RMSE=9.64 and for the second day, the model was built with Mtry=2 with the RMSE = 8.51. Figure 2 shows the values of RMSE for various mtry values for each day.
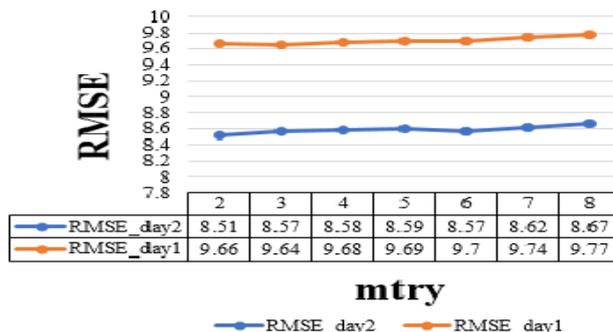
Figure 2. Choosing the most optimal mtry considering the RMSE value for both days.

Prediction models were built, and then the Landsat-8 indices were ranked based on the impact they had on building the models. Figure 3 shows Landsat-8 indices importance in building prediction model based on RF algorithm for each day. Figure 4 and 5 shows ML models performance on the first and second day, respectively.
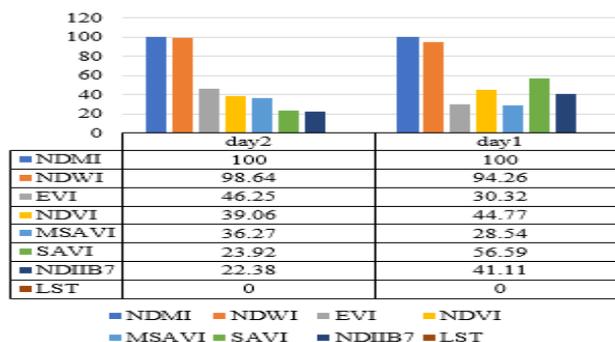
Figure 3. Ranking Landsat 8 indices in building model based on RF algorithm.
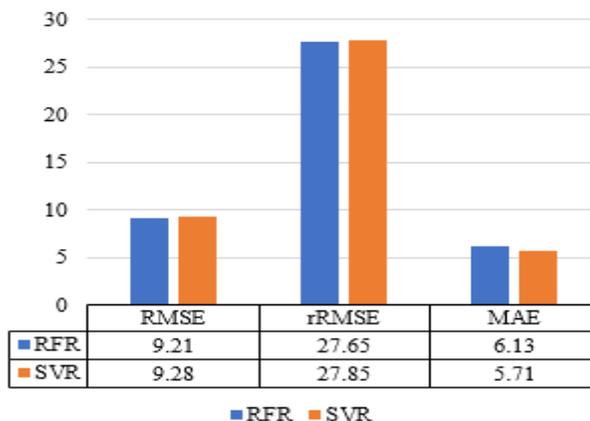
Figure 4. ML models performance on the first day.

First day
Second day
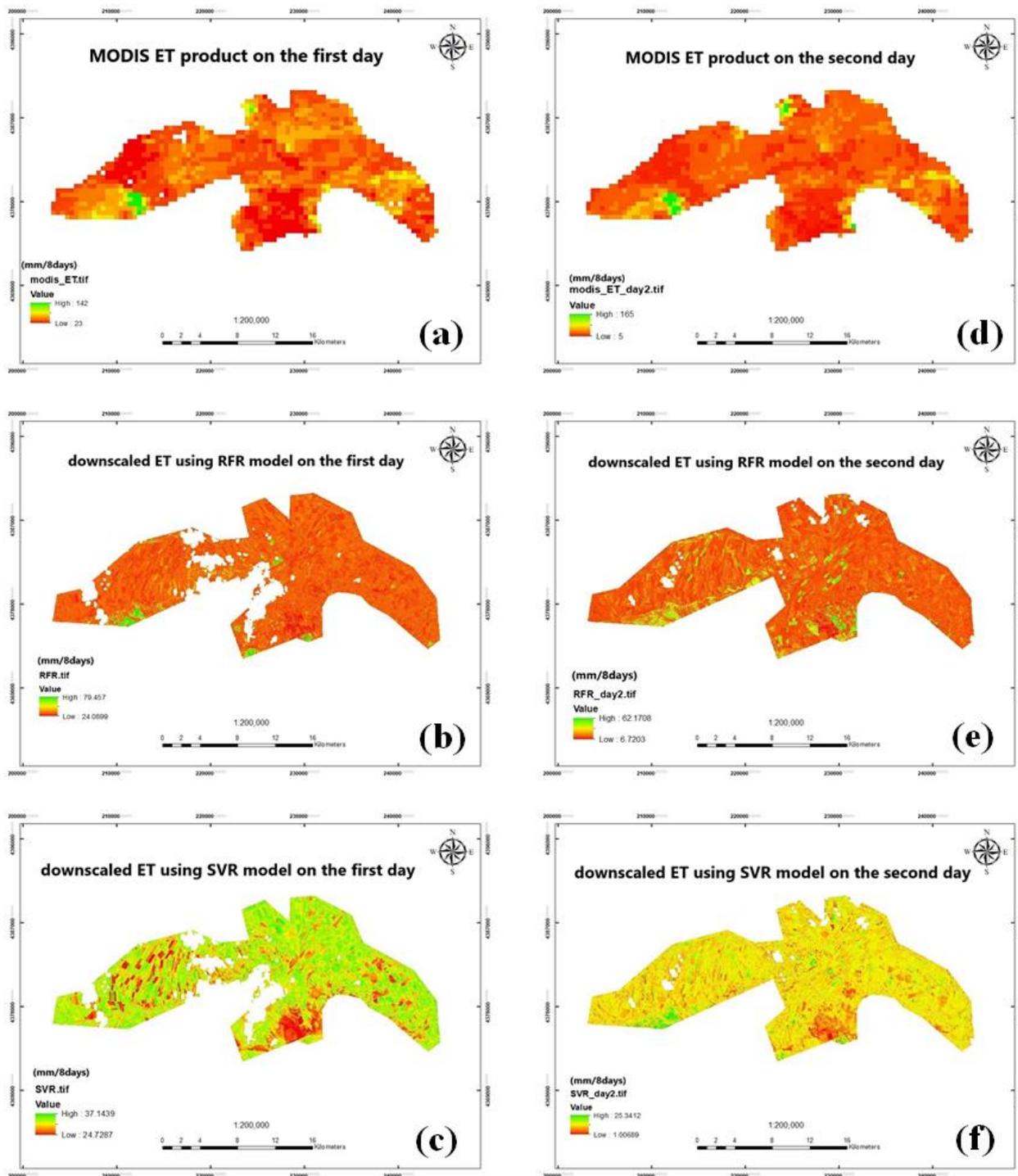
Figure 6. (a) and (d) MODIS 8-day ET for the first and second day, respectively, (b) and (e) Downscaled Landsat 30 m ET for the first and second day using RFR model, and (c) and (f) Downscaled Landsat 30 m ET for the first and second day using SVR model.

Ke et al. (2016) illustrated that indices including, EVI, NDVI, and SAVI which are known as vegetation growth-related were among the most predominant in 8-day ET modelling by RF algorithm. However, NDIIb7 and NDWI had moderately lesser significance. Also, LST which is an indicator of temperature had less significance than either vegetation greenness or wetness indicators. Also, two Landsat-8 indices related to water content, NDMI and NDWI, were identified as important parameters in modelling in the current study, and the importance of the NDWI parameter as an input for ET estimation was emphasized in the research of Morse et al (Mosre et al., 2021).
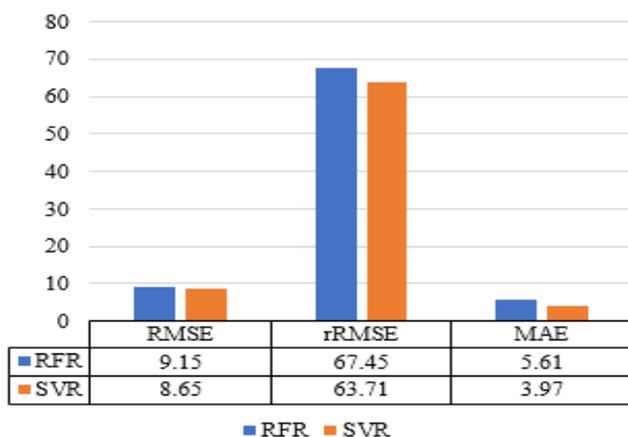
Figure 5. ML models performance on the second day.

2.2.2. Results obtained from two ML algorithms

Comparing these two models, SVR models resulted in better performances, for both dates. On the first day, RMSE and rRSME in the RFR model and the SVR model showed a slight difference, but the MAE value reported in the SVR model was lower than that in the RFR model. On the second day, the SVR model showed lower values of RMSE, rRSME and MAE compared to the RFR model. In general, the performance of the SVR performance is superior than the RFR model for both dates. Figure 3 and 4 presents RMSE, rRSME and MAE values for each day. In another study, Ke et al. (2016) used three machine learning-based methods for downscaling MODIS ET with the help of Landsat-8 data. their research showed that RF and Cubist model have executed more improved accuracies. Since agricultural sites contain numerous agricultural products and each product produces different amount of evapotranspiration, using downscaling method can be more accurate if applied on each specific product.

3. CONCLUSIONS

This research aims at suggesting a machine learning-based technique for downscaling MODIS ET product with the help of Landsat-8 data. Vegetation related indices, LST, and vegetation moisture-based indices were extracted from Landsat-8 datasets and then used as predictor variables and MODIS 8-day 500 m ET production as a response variable in order to build ML models. Moreover, feature selection using RF algorithm showed that on the first day NDMI, NDWI, SAVI, and NDVI indices were the most influential indices and they were used in prediction models for downscaling. However, on the second day NDMI, NDWI, EVI, and NDVI indices were the most influential indices. It demonstrates that vegetation related indicators and vegetation moisture-based indicators are highly influential among other indices and they were used to downscale the ET product. Between the two ML algorithms including SVR and RF tested, SVR which has excellent generalization capability, with high prediction accuracy resulted in better performance.

REFERENCES

Abbasi, N., Nouri, H., Didan, K., Barreto-Muñoz, A., Chavoshi Borujeni, S., Salemi, H., ... & Nagler, P. 2021. Estimating Actual Evapotranspiration over Croplands Using Vegetation Index Methods and Dynamic Harvested Area. *Remote Sensing*, *13*(24), 5167. doi.org/10.3390/rs13245167.

Aghighi, H., Azadbakht, M., Ashourloo, D., Shahrabi, H. S., & Radiom, S. 2018. Machine learning regression techniques for the silage maize yield prediction using time-series images of Landsat 8 OLI. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(12), 4563-4577. doi.org/10.1109/JSTARS.2018.2823361.

Bhattarai, N., Quackenbush, L. J., Dougherty, M., & Marzen, L. J. 2015. A simple Landsat–MODIS fusion approach for monitoring seasonal evapotranspiration at 30 m spatial resolution. *International Journal of Remote Sensing*, *36*(1), 115-143. doi.org/10.1080/01431161.2014.990645.

Bindhu, V. M., Narasimhan, B., & Sudheer, K. P. 2013. Development and verification of a non-linear disaggregation method (NL-DisTrad) to downscale MODIS land surface temperature to the spatial scale of Landsat thermal data to estimate evapotranspiration. Remote Sensing of Environment, 135, 118-129. doi.org/10.1016/j.rse.2013.03.023.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Glenn, E. P., Nagler, P. L., & Huete, A. R. 2010. Vegetation index methods for estimating evapotranspiration by remote sensing. *Surveys in Geophysics*, *31*(6), 531-555. doi.org/10.1007/s10712-010-9102-2.

Glenn, E. P., Neale, C. M., Hunsaker, D. J., & Nagler, P. L. 2011. Vegetation index-based crop coefficients to estimate evapotranspiration by remote sensing in agricultural and natural ecosystems. *Hydrological Processes*, *25*(26), 4050-4062. doi.org/10.1002/hyp.8392.

Golkar, H. Y. H., Kaveh, F., Ghahraman, B., & Sedghi, H. 2008. Time trend analysis monthly reference evapotranspiration by using of recommendation fao-penman-monteith method. www.sid.ir/paper/7844/en.

Hejazizadeh, Z., Fathollah, T. D., & Aligholi, S. 2017. The evaluation of agro climatic potentials of kermanshah and isfahan provinces in autumn cultivation of sugar beet. www.sid.ir/paper/102327/en.

Hong, S. H., Hendrickx, J. M., & Borchers, B. 2011. Down-scaling of SEBAL derived evapotranspiration maps from MODIS (250 m) to Landsat (30 m) scales. *International journal of remote sensing*, *32*(21), 6457-6477. doi.org/10.1080/01431161.2010.512929.

Im, J., Park, S., Rhee, J., Baik, J., & Choi, M. 2016. Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. *Environmental Earth Sciences*, *75*(15), 1-19. doi.org/10.1007/s12665-016-5917-6.

Kaheil, Y. H., Gill, M. K., McKee, M., Bastidas, L. A., & Rosero, E. 2008. Downscaling and assimilation of surface soil moisture using ground truth measurements. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(5), 1375-1384. doi.org/10.1109/TGRS.2008.916086.

Ke, Y., Im, J., Park, S., & Gong, H. 2016. Downscaling of MODIS One kilometer evapotranspiration using Landsat-8 data and machine learning approaches. *Remote Sensing*, *8*(3), 215. doi.org/10.3390/rs8030215.

Mosre, J., & Suárez, F. 2021. Actual evapotranspiration estimates in arid cold regions using machine learning algorithms with in situ and remote sensing data. *Water*, *13*(6), 870. doi.org/10.3390/w13060870.

Mu, Q., Zhao, M., & Running, S. W. 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote sensing of environment*, *115*(8), 1781-1800. doi.org/10.1016/j.rse.2011.02.019.

Mu, Q., Zhao, M., & Running, S. W. 2013. MODIS global terrestrial evapotranspiration (ET) product (NASA MOD16A2/A3). *Algorithm Theoretical Basis Document, Collection*, *5*, 600.

Murray, R. S., Nagler, P. L., Morino, K., & Glenn, E. P. 2009. An empirical algorithm for estimating agricultural and riparian evapotranspiration using MODIS Enhanced Vegetation Index and ground measurements of ET. II. Application to the Lower Colorado River, US. *Remote Sensing*, *1*(4), 1125-1138. doi.org/10.3390/rs1041125.

Nagler, P. L., Morino, K., Murray, R. S., Osterberg, J., & Glenn, E. P. 2009. An empirical algorithm for estimating agricultural and riparian evapotranspiration using MODIS enhanced vegetation index and ground measurements of ET. I. Description of method. *Remote Sensing*, *1*(4), 1273-1297. doi.org/10.3390/rs1041273.

Running, S. W., Mu, Q., Zhao, M., & Moreno, A. 2017. MODIS global terrestrial evapotranspiration (ET) product (NASA MOD16A2/A3) NASA earth observing system MODIS land algorithm. *NASA: Washington, DC, USA*. https://landweb.modaps.eosdis.nasa.gov/QA_WWW/forPage/user_guide/MOD16UsersGuide2016V1.52017May23.pdf

Russell, S., Norvig, P. 2010. Artificial Intelligence: A Modern Approach. Prentice Hall.

Sharghi, T., BARI, A. H., Asadi, M. A., & Kousari, M. R. 2010. Estimation of reference evapotranspiration by FAO-Penman-Monteith method and its zonation in Yazd province. www.sid.ir/paper/199770/en.

Singh, R. K., Senay, G. B., Velpuri, N. M., Bohms, S., & Verdin, J. P. 2014. On the downscaling of actual evapotranspiration maps based on combination of MODIS and Landsat-based actual evapotranspiration estimates. *Remote Sensing*, *6*(11), 10483-10509. doi.org/10.3390/rs61110483.

Wang, L., Qu, J. J., Hao, X., & Zhu, Q. 2008. Sensitivity studies of the moisture effects on MODIS SWIR reflectance and vegetation water indices. *International Journal of Remote Sensing*, *29*(24), 7065-7075. doi.org/10.1080/01431160802226034.