T.R.
ERCIYES UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE
DEPARTMENT OF CIVIL ENGINEERING

COMPARISON OF REGIONAL AND GENERAL MACHINE LEARNING MODELS FOR REFERENCE EVAPOTRANSPIRATION PREDICTION IN TURKEY

Prepared by
Yasser ZOUZOU

Supervisor
Assoc. Prof. Dr. Hatice ÇITAKOĞLU

MSc Thesis

July 2022
KAYSERI

T.R.
ERCIYES UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE
DEPARTMENT OF CIVIL ENGINEERING

COMPARISON OF REGIONAL AND GENERAL MACHINE LEARNING MODELS FOR REFERENCE EVAPOTRANSPIRATION PREDICTION IN TURKEY

(MSc Thesis)

Prepared by
Yasser ZOUZOU

Supervisor
Assoc. Prof. Dr. Hatice ÇITAKOĞLU

July 2022
KAYSERI

SCIENTIFIC ETHICS SUITABILITY

I declare that all information in this work was obtained in accordance with academic and ethical rules. Furthermore, I declare that all results and materials that are not at the essence of this work were quoted and referenced as required by those rules.

Yasser ZOUZOU

Signature

SUITABILITY FOR GUIDE

The MSc thesis titled **"Comparison of Regional and General Machine Learning Models for Reference Evapotranspiration Prediction in Turkey"** has been prepared in accordance with Erciyes University Graduate Education and Teaching Institute Thesis Proposal and Thesis Writing Guide.

Student	**Supervisor**
Yasser ZOUZOU	Assoc. Prof. Dr. Hatice ÇITAKOĞLU
Signature	Signature

Chairman of the Department of Civil Engineering

Prof. Dr. Ahmet Alper Öner

Signature

ACKNOWLEDGEMENTS

I am grateful to Allah for without His graces and blessings this thesis would not have been possible.

I would like to acknowledge and give my warmest thanks to my supervisor Assoc. Prof. Dr. Hatice Çıtakoğlu for her advice and guidance through all the stages of my thesis.

I would also like to give special thanks to my family for their unconditional support, without you none of this would have been possible.

Yasser Zouzou

July 2022, Kayseri

COMPARISON OF REGIONAL AND GENERAL MACHINE LEARNING MODELS FOR REFERENCE EVAPOTRANSPIRATION PREDICTION IN TURKEY

Yasser ZOUZOU

Erciyes University, Graduate School of Natural and Applied Sciences
MSc Thesis, July 2022
Supervisor: Assoc. Prof. Dr. Hatice ÇITAKOĞLU

ABSTRACT

Excessive freshwater usage in irrigation is an important issue in Turkey, which can be reduced by using evapotranspiration (ET) measurements to schedule irrigation based on crops' exact water need. The infeasibility of direct ET measurement has driven researchers to estimate its value from meteorological variables. In recent years, machine learning (ML) has been widely used for this purpose and showcased its potential in estimating reference evapotranspiration (ET_0) from limited variables. In this study, data collected from 165 weather stations in Turkey ranging between the years 1967 and 2020 was used to explore the spatial generalizability of ML models used for ET_0 estimation by comparing two modelling scenarios: (1) One model for the entire country (2) One model for each geographic region in Turkey. Three ML algorithms were used for this purpose: Support Vector Regression (SVR), Random Forests (RF), and Gaussian Process Regression (GPR). Furthermore, stations were split into train and test stations to allow for conducting a cross-station evaluation. The models were tested using 16 input variable combinations. In general, SVR and GPR models performed better than RF models, which were prone to overfit. For SVR and GPR models, the change in mean absolute error (MAE) between the two scenarios ranged between -16.3% and -0.4% and the change in root mean squared error (RMSE) was between -17.0% and 1.1%, where negative values indicate a lower prediction error in the regional models' scenario. Finally, a randomization test conducted on four of the input combinations showed that the reduction in prediction error when using regional models is statistically significant.

Keywords: Reference evapotranspiration, machine learning, cross-station evaluation, regional model, support vector regression, Gaussian processes, random forests

TÜRKİYE'DE REFERANS EVAPOTRANSPİRASYON TAHMİNİ İÇİN GENEL VE BÖLGESEL MAKİNE ÖĞRENME MODELLERİNİN KARŞILAŞTIRILMASI

Yasser ZOUZOU

Erciyes Üniversitesi, Fen Bilimleri Enstitüsü
Yüksek Lisans Tezi, Temmuz 2022
Danışman: Assoc. Prof. Dr. Hatice ÇITAKOĞLU

ÖZET

Sulamada aşırı su kullanımı Türkiye'de önemli bir sorundur ve bu sorun evapotranspirasyon (ET) ölçümleri kullanılarak bitkilerin su ihtiyacına göre sulama planlaması yapılarak azaltılabilir. Doğrudan ET ölçümünün pratik olmaması, araştırmacıları bu değeri meteorolojik değişkenlerden tahmin etmeye yöneltmiştir. Son yıllarda, makine öğrenmesi (MÖ) bu amaçla yaygın olarak kullanılmış ve sınırlı değişkenlerden referans evapotranspirasyonun (ET_0) tahmin edilmesindeki potansiyelini göstermiştir. Bu çalışmada, Türkiye'de bulunan 165 meteoroloji istasyonundan elde edilen ve 1967 ile 2020 yılları kapsayan veriler, iki modelleme senaryosunu karşılaştırarak ET_0 tahmini için kullanılan MÖ modellerinin mekansal genelleştirilebilirliğini araştırmak için kullanılmıştır: (1) Tüm ülke için bir model (2) Türkiye'deki her coğrafi bölge için bir model. Bu amaçla üç makine öğrenme algoritması kullanılmıştır: Destek Vektör Regresyonu (DVR), Rastgele Ormanlar (RO) ve Gauss Süreci Regresyonu (GSR). Ayrıca, istasyonlar arası değerlendirme gerçekleştirmek için istasyonlar eğitme ve test istasyonlarına ayrılmıştır. Modeller 16 girdi değişkeni kombinasyonu kullanılarak test edilmiştir. Genel olarak, DVR ve GSR modelleri, aşırı uyma eğilimli olan RO modellerinden daha iyi performans göstermiştir. DVR ve GSR modelleri için, iki senaryo arasındaki ortalama mutlak hata değişimi -%16,3 ile -%0,4 arasında ve kök ortalama kare hata değişimi -%17,0 ile %1,1 arasında değişmiştir; burada negatif değerler bölgesel modellerin senaryosunda daha düşük bir tahmin hatasına işaret etmektedir. Son olarak, girdi kombinasyonlarından dördü üzerinde yapılan bir rastgeleleştirme testi, bölgesel modeller kullanıldığında tahmin hatasındaki azalmanın istatistiksel olarak anlamlı olduğunu göstermiştir.

Anahtar kelimeler: Referans evapotranspirasyon, makine öğrenmesi, istasyonlar arası değerlendirme, bölgesel model, destek vektör regrasyonu, Gauss süreçleri, rastgele ormanlar

TABLE OF CONTENTS

COMPARISON OF REGIONAL AND GENERAL MACHINE LEARNING MODELS FOR REFERENCE EVAPOTRANSPIRATION PREDICTION IN TURKEY

SCIENTIFIC ETHICS SUITABILITY ii

SUITABILITY FOR GUIDE iii

ACCEPTANCE AND APPROVAL PAGE iv

ACKNOWLEDGEMENTS v

ABSTRACT vi

ÖZET vii

TABLE OF CONTENTS viii

LIST OF SYMBOLS AND ABBREVIATIONS x

LIST OF TABLES xii

LIST OF FIGURES xiii

INTRODUCTION 1

CHAPTER 1

GENERAL INFORMATION AND LITERATURE REVIEW

1.1 Importance of evapotranspiration 3

1.2 $ET0$ estimation using empirical equations and data driven models 5

1.3 Spatial variation of $ET0$ and its influencing factors 8

1.4 Generalizability of ML models for $ET0$ estimation 8

1.5 Thesis objective 10

CHAPTER 2

METHODOLOGY

2.1 Study area and data 11

2.2 Modelling methodology 16

2.3 FAO 56 PM equation 17
2.4 Algorithms 21
2.4.1 Support Vector Regression (SVR) 21
2.4.2 Gaussian Process Regression (GPR) 24
2.4.3 Random Forests (RF) 25
2.5 Evaluation Metrics 27
2.6 Statistical Significance Test 28

CHAPTER 3
RESULTS

3.1. Input combination comparison 32
3.2 Comparing general and regional models 37

CHAPTER 4
DISCUSSION AND CONCLUSIONS

DISCUSSION AND CONCLUSIONS 41
REFERENCES 44
CURRICULUM VITAE 49

LIST OF SYMBOLS AND ABBREVIATIONS

ANFIS : Adaptive Network Based Fuzzy Inference System

ANN : Artificial Neural Network

ET_0 : Reference evapotranspiration

FAO : The Food and Agriculture Organization

GPR : Gaussian Processes Regression

KDE : Kernel Density Estimate

MAE : Mean absolute error

ML : Machine learning

NSE : Nash–Sutcliffe model efficiency coefficient

RBF : Radial basis function

RF : Random Forests

RMSE : Root mean squared error

rRMSE : Relative root mean squared error

SVGPR : Sparse Variational Gaussian Processes

SVR : Support Vector Regression

α : Albedo

Δ : Slope of saturation vapor pressure temperature relationship

e_a : Actual vapor pressure

e_s : Mean saturation vapor pressure

$e°(T)$: Saturation vapor pressure at the air temperature T

G : Soil heat flux

P : Atmospheric pressure

RH : Relative humidity

R_s : Solar radiation

R_n : Net radiation

R_{ns} : Net solar radiation

R_{nl} : Net outgoing longwave radiation

R_{so} : Clear-sky radiation

R_a : Extraterrestrial radiation

T : Temperature

U_2 : Average wind speed at 2 m height

U_Z : Average wind speed at Z m height

Z : Elevation above sea level

LIST OF TABLES

Table 1. Comparison between the number of stations in the original dataset and the number of stations remaining after filtering the data 12
Table 2. Mean (standard deviation) of monthly averages of the meteorological variables 13
Table 3. Input combinations 17
Table 4. Performance ratings by NSE values 28
Table 5. MAE and RMSE of general and regional models 33
Table 6. rRMSE and NSE of general and regional models 34

LIST OF FIGURES

Fig. 1. Determining actual evapotranspiration from reference evapotranspiration. 4
Fig. 2. Station distribution and data availability cases.....6
Fig. 3. Clustered station distribution in Brazil 9
Fig. 4. Turkey's seven regions and station distribution.....11
Fig. 5. Variable distributions in regions compared to their distributions in the combined dataset.....16
Fig. 6. Number of train and test datapoints and stations per region.....17
Fig. 7. The regression hyperplane and its margin in a one-dimensional linear SVR case 22
Fig. 8. One-dimensional GP samples and test data predictions 25
Fig. 9. Box plot showing the distribution of the dataset sizes of the randomized clusters 31
Fig. 10. General and regional model metrics 36
Fig. 11. Change in prediction error between general and regional models.....37
Fig. 12. Train and test RMSE of general and regional models 38
Fig. 13. Change in RMSE between general and regional models per region 38
Fig. 14. (a) RMSE of general and regional models (b) Change in RMSE between general and regional models. 39
Fig. 15. Randomization test distributions and p-values 40

INTRODUCTION

The problem of freshwater scarcity is a worldwide concern that is being addressed by all governments. In order to tackle this problem, efforts are being made to preserve freshwater resources and to guarantee efficient water consumption in all sectors. In Turkey, agriculture accounts for the greatest freshwater withdrawal, comprising 77% of the total freshwater withdrawal in the country [1]. Therefore, ensuring efficient water usage in agriculture is a priority for freshwater preservation in Turkey.

Irrigation scheduling that considers crop water need reduces water consumption by ensuring that only the required amount of water is delivered to plants. Crop water need is computable from reference evapotranspiration (ET_0), which in turn is solely dependent on climate. However, the required climatic variables for ET_0 computation are not available in every area, hence the necessity of a reference evapotranspiration estimation methodology.

Researchers have been working on finding alternative empirical equations to estimate ET_0 from a few and easily obtainable climatic measurements. In recent years, machine learning (ML) has been widely used to estimate ET_0 using limited climatic variables. However, the spatial generalizability of ML models that estimate ET_0 has rarely been explored. The following questions help clarify the meaning of spatial generalizability: Would a single model for the entire world be able to estimate ET_0 in acceptable and uniform accuracy all around the world? Or is it necessary to create models for smaller regions to get decent and uniform accuracy?

This thesis addresses the spatial generalizability of ET_0 estimating ML models by comparing the performance of two types of models in Turkey: (1) A general model trained on data from all around Turkey (2) Regional models each trained on data from one of

Turkey's seven geographic regions, i.e., Mediterranean, Aegean, Marmara, Black Sea, Central Anatolia, Eastern Anatolia, and Southeastern Anatolia. In both scenarios, a cross-station assessment was conducted by training the models on data from a subset of the stations and testing them on the remaining stations.

CHAPTER 1

GENERAL INFORMATION AND LITERATURE REVIEW

1.1 Importance of evapotranspiration

The accelerating climate change and continuous human-induced pollution is severely affecting the useable freshwater resources around the world. In recent years, immense efforts have been focused on protecting freshwater resources and minimizing its consumption and wastage. This has driven governments to regulate freshwater withdrawal and consumption in order to maintain its resources.

Freshwater consumption in the agricultural sector in Turkey constitutes 77% of the total freshwater consumption in the country [1], rendering agriculture a primary sector in which water wastage must be minimized. The water used for irrigation is 2-3 times the actual requirement of plants in Turkey [2]. The reasons behind this excessive water usage include water losses in the irrigation system, low water prices, and lack of trained farmers. Besides the reestablishment of irrigation systems, providing systems to plan irrigation based on crop water demand, then increasing water prices accordingly will result in significant reduction in water consumption in agriculture.

Crops need water to compensate for evaporation and transpiration, therefore crop water need is also called evapotranspiration [3]. Evapotranspiration is defined as the total loss of moisture from land, which occurs through evaporation of free water and transpiration from plants [4]. Evapotranspiration measurements are used in climate change studies,

water resource evaluation, drought forecasting, and plant water need management [5]. Whereas climate change studies and drought forecasting only require sparse evapotranspiration measurements, using evapotranspiration as a driver of irrigation

planning necessitates knowing its value in specific locations. Furthermore, measuring actual evapotranspiration is only possible using lysimeters, which are limited to research use due to their high costs and special maintenance requirements [6]. Therefore, actual evapotranspiration is commonly estimated from reference evapotranspiration, which is only dependent on climatic factors. Reference evapotranspiration (ET_0) is the evapotranspiration from a hypothetical grass crop that is 0.12 m in height, has a surface resistance of 70 $s \cdot m^{-1}$, and an albedo of 0.23 [4]. Knowing the value of ET_0, actual evapotranspiration can be determined by adjusting ET_0 using coefficients that relate crop type and other conditions to ET_0 (Fig. 1).

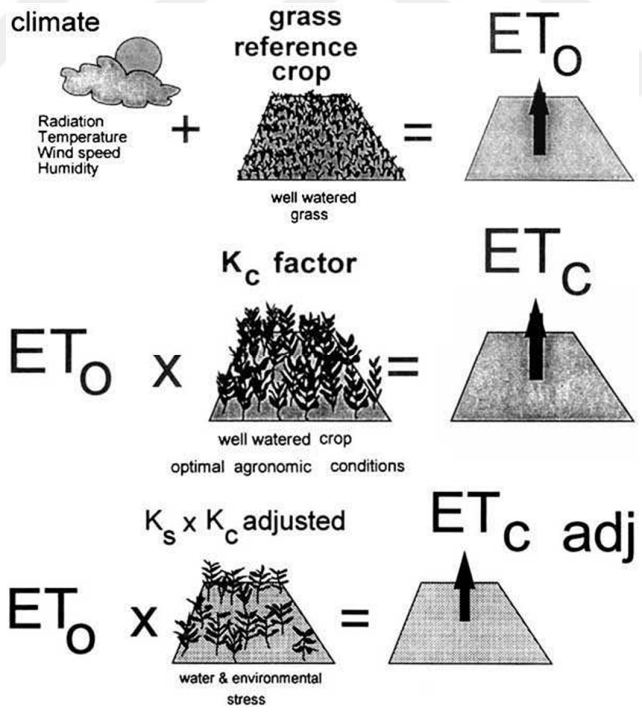

Fig. 1. Determining actual evapotranspiration from reference evapotranspiration [4].

Although several methods for ET_0 computation have existed, the Food and Agriculture Organization (FAO) Penman-Monteith method has been recommended as the sole

standard method that defines and computes reference evapotranspiration after an expert consultation held in 1990 [4]. FAO does not encourage the use of other ET_0 methods. Nevertheless, ET_0 computation using this equation is still not feasible in many areas because it requires meteorological measurements that are only available in weather stations. Accordingly, researchers have been working on further simplifying the estimation of ET_0 using less climatic variables, especially ones that are widely available like temperature and wind speed. These studies have used two approaches: (1) Empirical equations (2) Data driven models.

1.2 ET_0 estimation using empirical equations and data driven models

With the emergence of machine learning (ML), numerous studies have compared the performance of ML models to that of traditional empirical equations in ET_0 estimation. Supervised ML algorithms learn the relationship between input variables and target variables from sample data, and subsequently predict target variables from unseen input data [7]. The ability of ML to work with non-linear and high dimensional data makes it superior to traditional regression methods.

In their study comparing ML methods to empirical equations for ET_0 estimation in Turkey, Citakoglu et al. [8] concluded that Artificial Neural Network (ANN) and Adaptive Network Based Fuzzy Inference System (ANFIS) are both superior to empirical equations. A cross-station assessment was used in their study, where models were tested on data from stations other than the stations used to train these models. Wen et al. [9] used a dataset of 146 days of meteorological measurements from one site in Ejina basin, China to train and test Support Vector Regression (SVR) and ANN models. The two ML model predictions were compared to three empirical equations: Priestly-Taylor, Hargreaves, and Ritchie. Amongst all models and equations, the SVR model had the best accuracy. However, the SVR model was only marginally better than the empirical Priestly-Taylor equation, which can be attributed to the small size of the dataset, as ML algorithms require large datasets to produce good predictions.

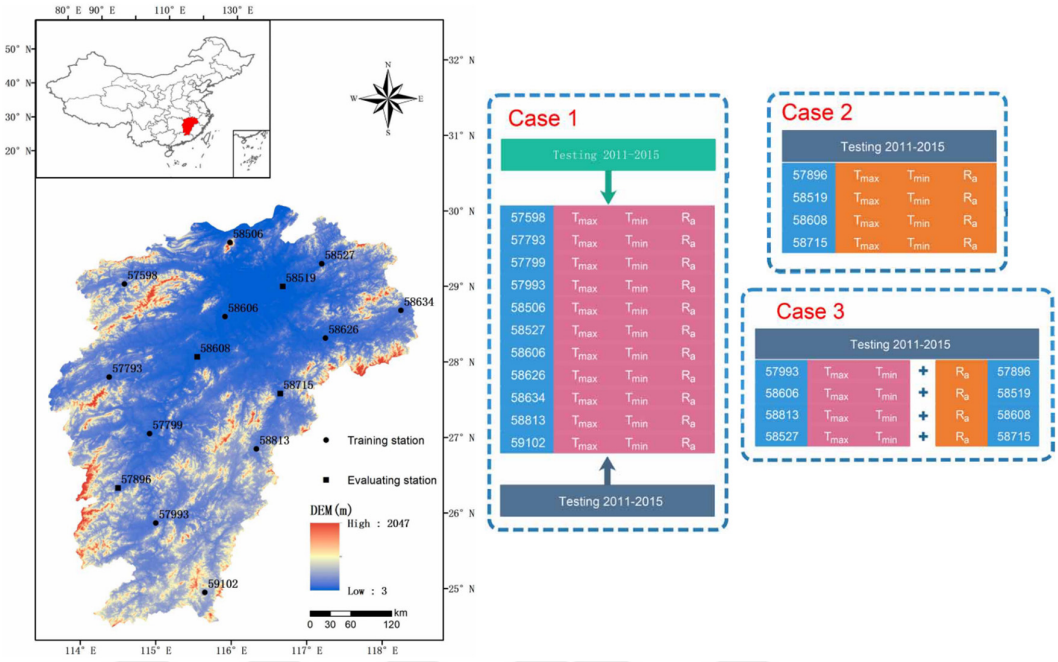

Fig. 2. Station distribution and data availability cases [10].

Wu et al. [10] addressed several realistic data availability scenarios in their study comparing eight ML models and one empirical equation (Hargreaves-Samani) using climatic data between the years 2001 and 2015 from 15 stations in the Jiangxi province, China. The authors first trained the models using maximum temperature, minimum temperature, and extraterrestrial radiation measured between the years 2001 and 2010 from 11 of the 15 stations. Then, these models were tested using data from the years 2011 to 2015 in 3 different scenarios (Fig. 2): (1) Using the data from the same 11 stations used for training (2) Using data from the 4 remaining stations (3) Using temperature data from 4 stations adjacent to the 4 remaining stations, assuming that temperature is not measured in these stations. While the first scenario serves as a benchmark, the second scenario represents a case where models are to be used in locations where only temperature is measured (extraterrestrial radiation is computable from location and time of year). Furthermore, the third scenario represents a case where the models are to be used in areas where climatic measurements are not available at all, therefore measurements from nearby stations are used. In the three scenarios, all ML models performed better than the empirical equation. However, the root mean squared error (RMSE) of the best performing

model in the first case increased by approximately 17% from 0.269 to 0.315. This change in prediction error emphasizes the importance of using a cross-station assessment of models.

Landeras G. et al. estimated ET0 in Ghana using ANN and GEP models along with three empirical equations: Hargreaves-Samani and two variants of the FAO 56 equation that fewer input variables [11]. The data used in this study was daily data collected from four stations between the years 2006 and 2012. Splitting was done by assigning some years for training/calibrating and the remaining years for testing. Furthermore, three modelling scenarios were explored for the ML models: (1) Using local temperature and relative humidity data (2) Using sunshine durations from a nearby station along with the local temperature and relative humidity data (3) Using complete data from a nearby station. In general, the ML models outperformed the empirical equations as expected. Comparing the first and second modelling scenarios, the improvement in predictions in the second scenario is highly dependent on the choice of the nearby stations. In some cases the second scenario resulted in worse predictions. This further emphasizes the importance of testing models on data from different stations, as the variation in data between nearby stations may be significant. Finally, the third scenario seems unrealistic due to the absence of target data from the studied station to train ML models.

The previous studies confirm the advantage of ML models over empirical equations in the task of ET_0 estimation. In these studies, different approaches have been used to test the ML models:

1 Using data from one or several weather station to train and test ML models, where train and test datasets are created by assigning certain years to each dataset (temporal split).

2 Using data from several stations to train and test ML models, where train and test datasets are created by assigning certain stations to each dataset (spatial split or cross-station).

In the context of using ML models to estimate ET_0 in areas where only few meteorological variables are measured, the second method is a better approach because

these models are to be used in areas where historical meteorological data sufficient to compute ET_0 is not available.

1.3 Spatial variation of ET_0 and its influencing factors

By clustering 75 stations using the regression coefficients obtained by regressing climatic variables to ET_0, Zhang et al. [12] studied the spatial variation of ET_0 in the Qinghai-Tibetan Plateau in China. The researchers found that ET_0 is mainly influenced by windspeed and solar radiation in the northern and southern regions of the plateau, respectively. In their study on the effect of measurement errors in climatic variables on the computed values of reference evapotranspiration, Yildirim et al. [13] concluded that ET_0 is sensitive to different variables in different regions of Turkey. ET_0 values in eastern and southeastern Anatolia were mostly affected by errors in wind speed, while being mainly affected by air temperature in central Anatolia, Marmara, and Aegean Sea regions. Cobaner et al. [14] indicated an improvement in ET_0 estimations made using the Hargreaves-Samani equation after calibrating its coefficients to the seven regions of Turkey. The previous studies emphasize the variation of factors that influence ET_0 values from a region to another, further justifying the question: To what degree can ET_0 estimating ML models generalize spatially?

1.4 Generalizability of ML models for ET_0 estimation

In a study comparing four data driven models to empirical equations using data from 29 stations in Iran, Shiri et al. [15] tested two modelling scenarios: (1) Local models trained and tested on individual stations (2) Pooled data models trained and tested on groups of stations clustered using the aridity index. The results show that local models outperform pooled models in the arid and semi-arid regions of this study. In the humid region, which includes only 4 stations, the pooled model had similar results to the local models. It is worth mentioning that even in the pooled data model case no stations were left out for testing.

Ferreira et al. [16] used all climatic measurements, including the computed ET_0, to cluster 203 weather stations (Fig. 3) using the K-means algorithm in one part of their study to estimate ET_0 using ANN and SVR models along with empirical equations. Individual models (regional models) were subsequently trained for each of the clusters, in addition

to a general model trained using all 203 stations. The authors observed an improvement in performance when using regional models in comparison to the general model. The authors split the data to train and test sets by assigning different years to each set, thus no stations were left out for testing the models. Furthermore, the clustering technique used is not scalable to real applications for two reasons: (1) ET_0 was used as a variable in the clustering algorithm, therefore determining the cluster of a new area in which these models are to be used is not possible due to the lack of ET_0 measurements there (2) The clusters created do not have a clear spatial distribution, which prohibits the identification of the cluster of a new area merely from its location.

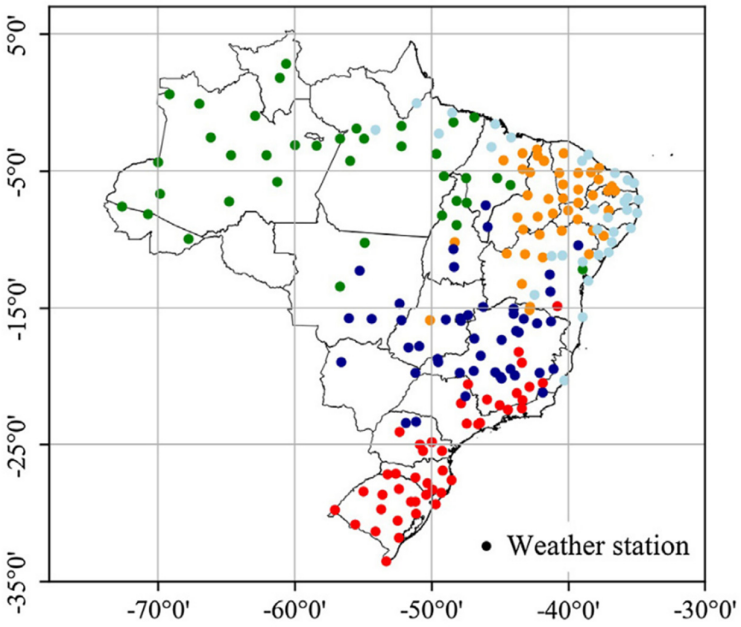

Fig. 3. Clustered station distribution in Brazil

In a study conducted on data from 9 stations in Brazil, Ferreira et al. [17] tested the generalizability of empirical equations, ANN, and multivariate adaptive regression splines (MARS) models. Two scenarios were studied: (1) Models were trained iteratively on data from all but 1 station and evaluated in the station left out (regional models) (2) Models were trained and tested on data from the same station (local models). The performance of local models was better than regional models as expected. However,

regional models were able to predict ET_0 with acceptable accuracy, except for a couple of stations where the deviation in prediction error between local and regional models was large. The authors attributed this deviation to the fact that the wind speed values in these stations are more spread out than in other stations. Furthermore, when the local models were tested on data from other stations, their performance was worse than that of the regional models. This finding further underlines the importance of training models on data from several stations for better generalization.

1.5 Thesis objective

All studies addressing the generalizability of ML models for ET_0 prediction, to the knowledge of the author, compare models trained and tested on the same stations to models trained on a number of stations and tested on different stations. Given the purpose of these models, which is to be used in areas in which the required variables for ET_0 computation is not available, models should be trained and tested on different stations to get realistic accuracy values. This type of model assessment is referred to in literature as cross-station assessment. Therefore, when examining the generalizability of models, both local and general models should be assessed using the cross-station method.

In an aim to address the generalizability of ET_0 estimating models within the guidelines mentioned earlier, this thesis compared general and regional models in Turkey using a cross-station assessment in both model scenarios. The regional models were trained on data from each of Turkey's seven geographic regions: Mediterranean, Aegean, Marmara, Black Sea, Central Anatolia, Eastern Anatolia, and Southeastern Anatolia. Three machine learning algorithms were used in this study: Support Vector Regression (SVR), Gaussian Processes Regression (GPR), and Random Forests (RF). The results of this thesis should provide insights on the spatial generalizability of ML models for estimating ET_0 in a climatically diverse country like Turkey.

CHAPTER 2

METHODOLOGY

2.1 Study area and data

Turkey stretches 1800 km between longitudes 26° and 45° E, and 600 km between latitudes 36° and 45° N, covering a total area of 780,000 km^2. The location of Turkey is in a transitional region between polar and tropical air masses [18]. The diverse topography in Turkey results in high climatic variance between different regions [19]. Although several studies have attempted clustering the climate zones of Turkey [19–21], the widely accepted climate zones still are the seven geographic regions defined by Erinc, S. [22], which were defined based on climatic, economic, and social factors. These regions are: Mediterranea, Aegean, Marmara, Black Sea, Central Anatolia, Eastern Anatolia, and Southeastern Anatolia (Fig. 4).

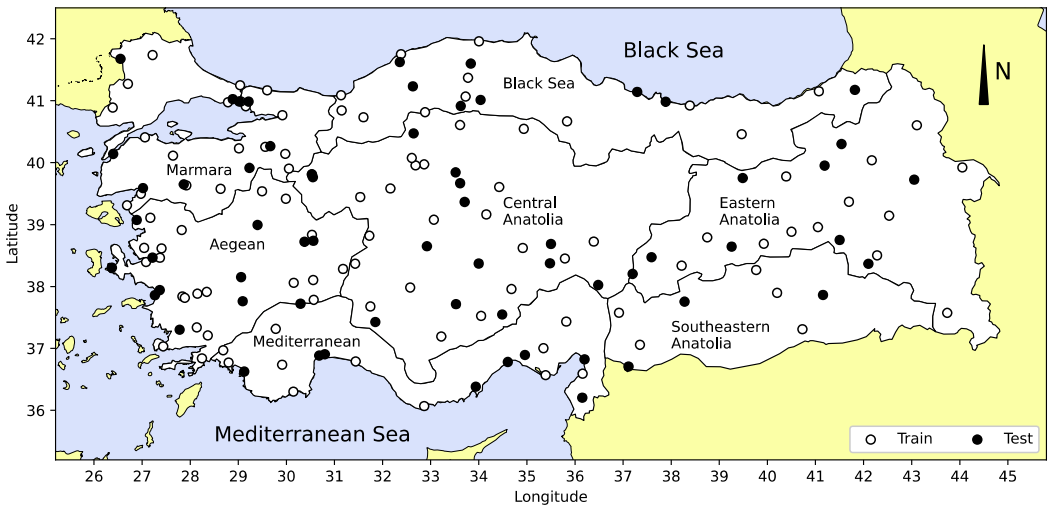

Fig. 4. Turkey's seven geographic regions and station distribution

Monthly averages of climatic variables were obtained from the Turkish State Meteorological Service. The measurement years, which vary from one station to another, range between 1967 and 2020. Stations that do not include the required variables for ET_0 computation were excluded. These variables are relative humidity, average wind speed, solar radiation, and temperature. Furthermore, years that do not include all 12 months measurements were also excluded to maintain a uniform number of months among all stations. After filtering the data, 44388 measurements from 165 stations were left in the dataset. A comparison between the number of stations in the original dataset and the number of stations remaining after filtering the data is shown in Table 1. The means and standard deviations of the monthly averages are presented in Table 2, where T is temperature ($°C$), RH is relative humidity (%), U_2 is average wind speed at 2m height ($m.s^{-1}$), R_s is solar radiation ($MJ.m^{-2}.day^{-1}$), and ET_0 is reference evapotranspiration ($mm.day^{-1}$) computed using the FAO 56-PM equation. The distributions of these variables are presented in Fig. 5 using a Kernel Density Estimate (KDE). The distributions of the train and test datasets of the combined data are almost identical, therefore only one line (the distribution of the whole dataset) was plotted to reduce clutter.

Table 1. Comparison between the number of stations in the original dataset and the number of stations remaining after filtering the data

	Black sea	Marmara	Central Anatolia	Eastern Anatolia	Aegean	Southeastern Anatolia	Mediter-ranean	All Data
Total number of stations	247	197	205	176	224	82	167	1298
Number of stations used	20	28	35	22	33	8	19	165
Percentage of stations used	8.1%	14.2%	17.1%	12.5%	14.7%	9.8%	11.4%	12.7%

Table 2. Mean (standard deviation) of monthly averages of the meteorological variables

		Black sea	Marmara	Central Anatolia	Eastern Anatolia	Aegean	Southeastern Anatolia	Mediter-ranean	All Data
Year	Train	1996 (13)	1995 (11)	1995 (12)	1995 (13)	1995 (11)	1994 (12)	1994 (11)	1995 (12)
	Test	1998 (9)	1993 (12)	1995 (12)	1995 (12)	1993 (11)	1995 (12)	1994 (11)	1995 (11)
T_{min}	Train	0.19 (8.46)	3.15 (7.75)	-1.92 (9.08)	-1.93 (11.61)	3.44 (8.52)	4.54 (9.19)	8.57 (7.80)	1.54 (9.77)
	Test	1.09 (8.67)	3.35 (7.78)	-2.01 (9.09)	-3.60 (11.49)	4.23 (7.92)	6.40 (8.33)	9.65 (7.35)	1.74 (9.93)
T_{max}	Train	25.23 (8.01)	26.75 (7.35)	24.96 (8.63)	22.82 (10.93)	27.70 (8.40)	27.97 (9.96)	29.31 (7.38)	26.03 (9.01)
	Test	25.17 (7.83)	26.35 (7.49)	24.63 (8.77)	22.71 (10.60)	27.64 (7.78)	29.05 (9.81)	29.60 (7.07)	25.96 (8.89)
RH_{avg}	Train	71.57 (9.67)	71.57 (8.09)	62.51 (11.65)	58.67 (14.88)	62.43 (10.42)	56.52 (15.97)	65.70 (8.97)	63.99 (12.48)
	Test	69.84 (9.06)	70.13 (9.74)	62.90 (11.78)	61.81 (13.59)	63.65 (10.22)	52.53 (14.80)	64.58 (7.91)	64.08 (11.94)
U_2	Train	1.34 (0.57)	1.79 (0.66)	1.70 (0.65)	1.23 (0.51)	1.63 (0.52)	1.53 (0.71)	1.47 (0.63)	1.54 (0.63)
	Test	1.27 (0.32)	1.98 (0.73)	1.77 (0.54)	1.34 (0.54)	1.61 (0.68)	1.54 (0.54)	1.67 (0.76)	1.61 (0.65)
R_s	Train	12.58 (6.25)	13.12 (6.35)	15.64 (6.63)	14.90 (6.24)	15.46 (6.57)	15.58 (6.82)	15.07 (5.98)	14.75 (6.51)
	Test	13.60 (6.35)	13.31 (6.62)	15.61 (6.66)	14.33 (6.30)	14.46 (6.25)	14.51 (6.32)	15.79 (6.16)	14.61 (6.46)
ET_0	Train	3.30 (1.52)	3.87 (1.59)	3.96 (1.86)	3.29 (1.83)	4.09 (1.86)	4.13 (2.37)	3.82 (1.56)	3.78 (1.82)
	Test	3.30 (1.56)	3.97 (1.66)	4.01 (1.85)	3.35 (1.81)	3.90 (1.71)	4.09 (2.11)	4.12 (1.71)	3.81 (1.79)

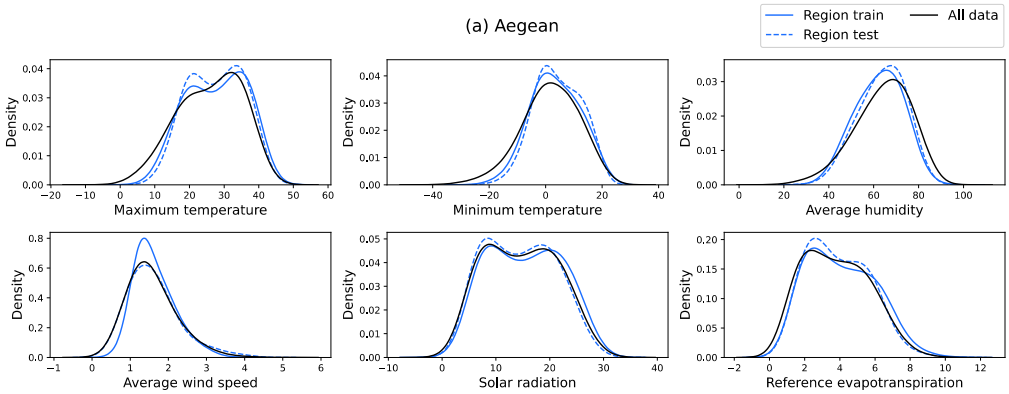

Fig. 5.a. Variable distributions in the Aegean region compared to their distributions in the combined dataset

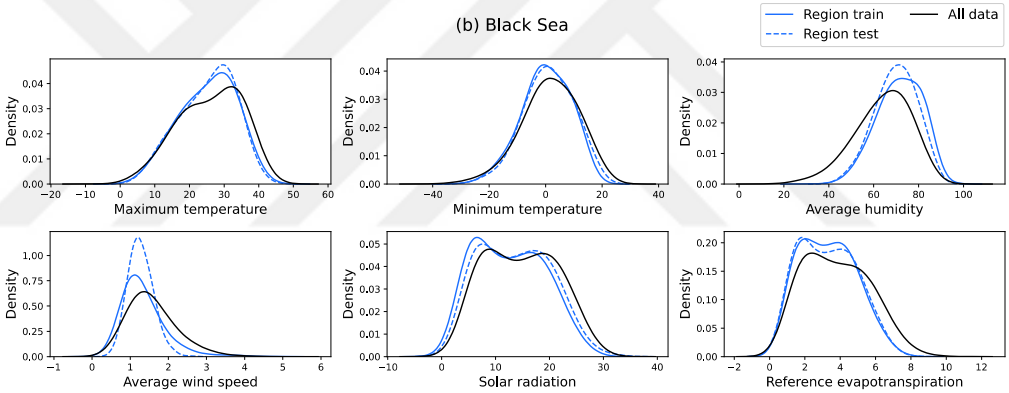

Fig. 5.b. Variable distributions in the Black Sea region compared to their distributions in the combined dataset

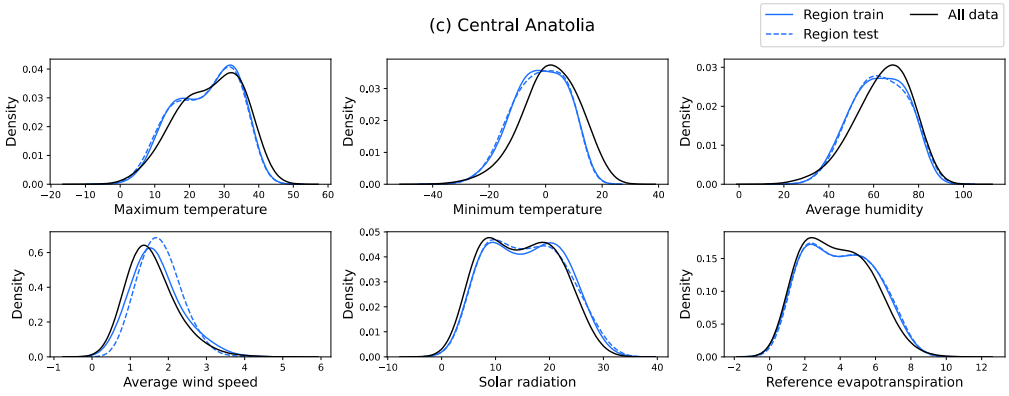

Fig. 5.c. Variable distributions in the Central Anatolia region compared to their distributions in the combined dataset

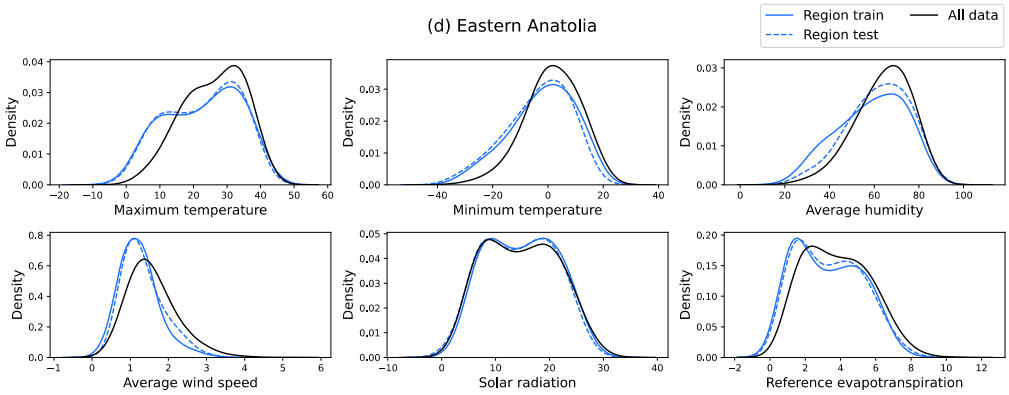

Fig. 5.d. Variable distributions in the Eastern Anatolia region compared to their distributions in the combined dataset

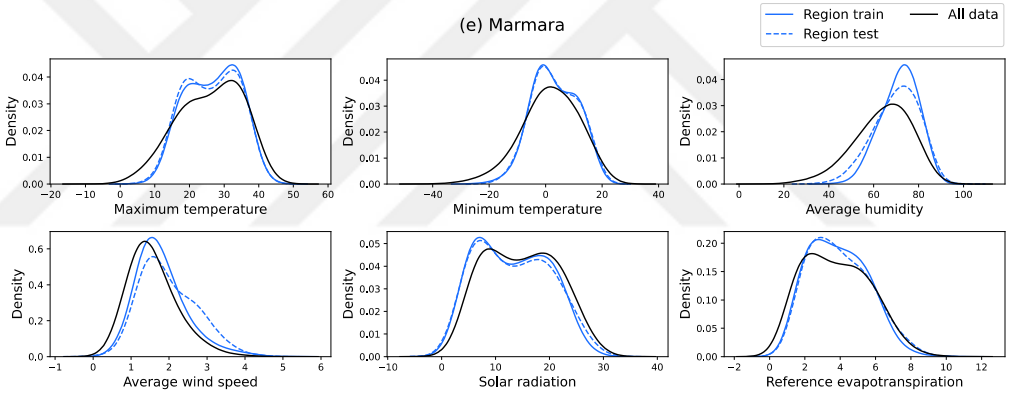

Fig. 5.e. Variable distributions in the Marmara region compared to their distributions in the combined dataset

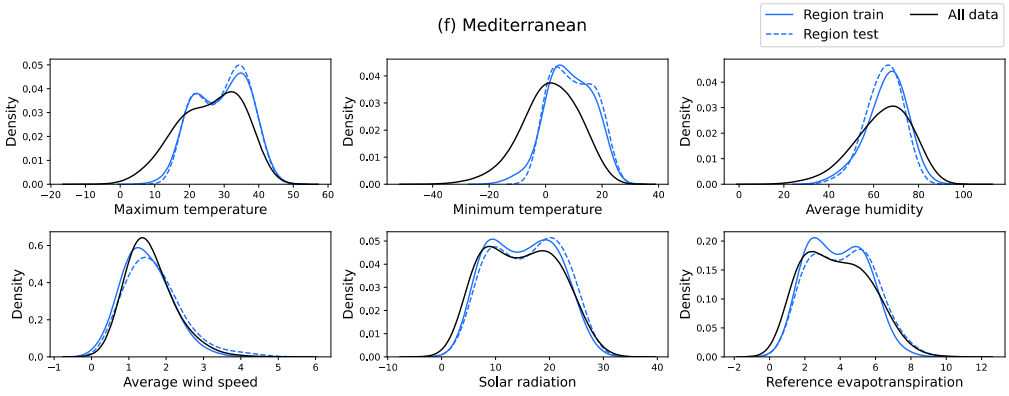

Fig. 5.f. Variable distributions in the Mediterranean region compared to their distributions in the combined dataset

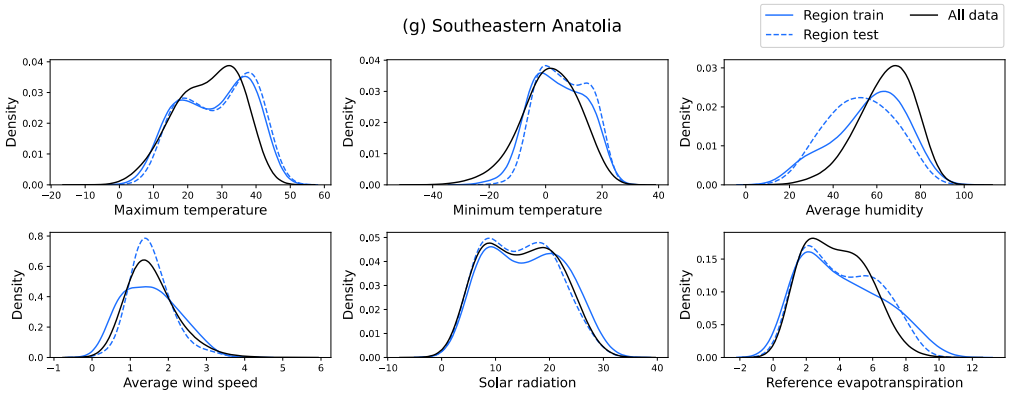

Fig. 5.g. Variable distributions in the Southeastern Anatolia region compared to their distributions in the combined dataset

2.2 Modelling methodology

In this study, two modelling schemes were compared: (1) A general model for all of Turkey (2) Regional models for each of the seven geographic regions. After assigning each station to its corresponding region, the stations were randomly assigned to train and test stations in each region to achieve a test ratio of approximately 40%. These stations were used to train and test a regional model for each region. The number of data points and number of stations in each region are presented in (Fig. 6). The general model was trained and tested using the train and test stations of all regions combined, with a total of 102 and 63 training and test stations, respectively. The test data ratio of the general model was 39.6%. The stations used to test both general and regional models were the same, thus eliminating the potential influence of data variation between general and regional models on the comparison between them. The models were created using 16 input combinations (Table 3) and 3 ML algorithms, which are described in section 2.4. The 16 input combinations represent different scenarios of available meteorological measurements in the locations where such models are to be used.

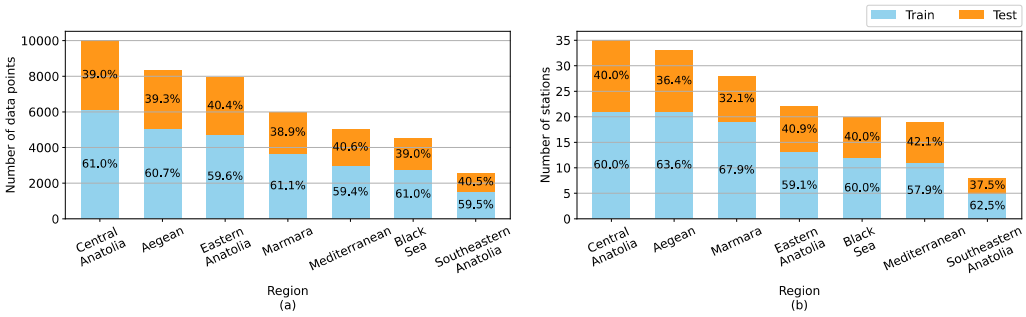

Fig. 6. Number of train and test datapoints and stations per region.

Table 3. Input combinations

No.	Input variables	No.	Input variables
1	T_{max}, U_2	9	R_s, U_2
2	T_{min}, U_2	10	R_s, RH_{max}
3	$T_{max}, U_2, Month$	11	R_s, RH_{avg}
4	T_{max}, RH_{max}	12	R_s, T_{max}
5	T_{max}, RH_{min}	13	$R_s, Month$
6	T_{max}, RH_{avg}	14	$T_{max}, U_2, Month, R_a$
7	$T_{max}, Month$	15	$T_{max}, U_2, Month, R_s$
8	RH_{max}, U_2	16	$T_{max}, U_2, Month, RH_{max}$

2.3 FAO 56 PM equation

The equation used in this study to compute reference evapotranspiration from climatic measurements is the FAO 56 Penman Monteith equation (Eq. 1), which is recommended by the Food and Agriculture Organization (FAO) [4]. The FAO 56 PM equation computes the evapotranspiration from a hypothetical grass crop that is 0.12 m in height, has a surface resistance of 70 $s \cdot m^{-1}$, and an albedo of 0.23.

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} U_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad (1)$$

Where:

ET_0: Reference evapotranspiration in $mm \cdot day^{-1}$

T: Mean air temperature in $°C$

Δ: Slope of saturation vapor pressure temperature relationship in $KPa \cdot °C^{-1}$

γ: Psychrometric constant in $KPa \cdot °C^{-1}$

e_s: Mean saturation vapor pressure in KPa

e_a: Actual vapor pressure in KPa

U_2: Average wind speed at 2m height in $m \cdot s^{-1}$

R_n: Net radiation in $MJ \cdot m^{-2} \cdot day^{-1}$

G: Soil heat flux in $MJ \cdot m^{-2} \cdot day^{-1}$

Average temperature is computed using Eq. 2, where T_{max} and T_{min} are the monthly averages of daily maximum and minimum temperatures (°C).

$$T = \frac{T_{max} + T_{min}}{2} \quad (2)$$

The slope of saturation vapor pressure curve (Δ), measured in $KPa \cdot °C^{-1}$, was computed using Eq. 3.

$$\Delta = \frac{4098 \times \left[0.6108 \times \exp\left(\frac{17.27 \times T}{T + 237.3}\right)\right]}{(T + 237.3)^2} \quad (3)$$

The psychrometric constant (γ) was computed using Eq. 4, where P is the atmospheric pressure (KPa) computed relative to the altitude using Eq. 5.

$$\gamma = 0.665 \times 10^{-3} P \quad (4)$$

$$P = 101.3\left(\frac{293 - 0.0065 \times Z}{293}\right) \quad (5)$$

Where Z is the elevation above sea level (m).

Mean saturation vapor pressure (e_s) was computed using Eq. 6, where $e°(T)$ is the saturation vapor pressure (KPa) at the air temperature T as computed in Eq. 7.

$$e_s = \frac{e°(T_{max}) + e°(T_{min})}{2} \quad (6)$$

$$e°(T) = 0.6108 \times \exp\left(\frac{17.27 \times T}{T + 237.3}\right) \quad (7)$$

Actual vapor pressure was computed using Eq. 8 as recommended in the FAO 56 paper. given that no information about the integrity of relative humidity measurements is available [4].

$$e_a = e°(T_{min})\frac{RH_{max}}{100} \quad (8)$$

The average wind speed in the weather stations used in this study are measured at 10m above the ground. Therefore, wind speed values were converted to the equivalent wind speed at 2m above the ground using Eq. 9.

$$U_2 = U_Z \frac{4.87}{\ln(67.8 \times Z - 5.42)} \quad (9)$$

Where U_Z is the wind speed ($m \cdot s^{-1}$) measured at a height Z meters above the ground.

Net radiation (Eq. 10) takes into account all incoming and outgoing shortwave and longwave radiations to represent the balance of the energy absorbed, emitted, and reflected by the earth surface [4].

$$R_n = R_{ns} - R_{nl} \quad (10)$$

Where R_{ns} is the net solar radiation in $MJ \cdot m^{-2} \cdot day^{-1}$ (Eq. 11) and R_{nl} is the net outgoing longwave radiation in $MJ \cdot m^{-2} \cdot day^{-1}$ (Eq. 12).

$$R_{ns} = (1 - \alpha) \cdot R_s \tag{11}$$

Where R_s is the measured solar radiation in $MJ \cdot m^{-2} \cdot day^{-1}$ and α is the albedo, which is the proportion of the solar radiation that is reflected by the earth surface. The albedo is dependent on the earth cover and is assumed 0.23 for the reference grass surface.

$$R_{nl} = \sigma \left[\frac{T_{max,K}^4 + T_{min,K}^4}{2}\right] \left(0.34 - 0.14\sqrt{e_a}\right) \left(1.35 \times \frac{R_s}{R_{so}} - 0.35\right) \tag{12}$$

Where:

σ: Stefan-Boltzmann constant ($4.903 \times 10^{-9}\ MJ \cdot K^{-4} \cdot m^{-2} \cdot day^{-1}$).

$T_{max,K}$, $T_{min,K}$: Maximum and minimum temperature, respectively, in Kelvin.

R_{so}: Clear-sky radiation in $MJ \cdot m^{-2} \cdot day^{-1}$, which is the solar radiation that would reach the earth in the same location and time of year when the sky is cloudless (Eq. 13).

$$R_{so} = (0.75 + 2 \times 10^{-5} \times Z\ ) \cdot R_a \tag{13}$$

Where Z is the elevation above sea level in m and R_a is the extraterrestrial radiation in $MJ \cdot m^{-2} \cdot day^{-1}$ (Eq. 14).

$$R_a = \frac{24 \times 60}{\pi} \times 0.082 d_r [\omega_s sin(\varphi) sin(\delta) + cos(\varphi) cos(\delta) sin(\omega_s)] \tag{14}$$

Where:

d_r: Inverse relative distance Earth-Sun (Eq. 15).

φ: Latitude in rad.

δ: Solar decimation in rad (Eq. 16).

ω_s: Sunset hour angle in rad (Eq. 17).

$$d_r = 1 + 0.033 \cos\left(\frac{2\pi}{365} J\right) \quad (15)$$

$$\delta = 0.409 \sin\left(\frac{2\pi}{365} J - 1.39\right) \quad (16)$$

$$\omega_s = arccos\left(-tan(\varphi) \cdot tan(\delta)\right) \quad (17)$$

The recommended equation in the FAO Irrigation and Drainage Paper No. 56 [4] for computing the soil heat flux is Eq. 18. In cases where the temperature of the following month is not available, Eq. 19 can be used. In this study, the recommended equation (Eq. 18) was used wherever possible, and Eq. 19 was used otherwise. Furthermore, when the previous month value was not available, for instance in the first month measurement of the first year in each station, the last month temperature measurement of the same year is used as an approximation of the previous month temperature value.

$$G_{month,i} = 0.07(T_{month,i+1} - T_{month,i-1}) \quad (18)$$

$$G_{month,i} = 0.14(T_{month,i} - T_{month,i-1}) \quad (19)$$

Where $T_{month,i}$ is the mean air temperature of month i in $°C$.

2.4 Algorithms

2.4.1 Support Vector Regression (SVR)

Support Vector Machines (SVM) were first introduced by Boser et al. [23] as a classifying algorithm that maximizes the margin between the decision hyperplane and the training points. A regression algorithm, Support Vector Regression (SVR), was then developed based on the concepts of SVM [24]. SVR uses a loss function that does not penalize errors between training points and the regression hyperplane that are less than ϵ, hence the name ϵ-insensitive loss function. This creates a tolerance margin around the regression hyperplane as shown in Fig. 7 for a one-dimensional linear SVR case. Errors outside this

margin, shown as ξ_i, ξ_i^*, are scaled by a regularization term (C) and penalized in the loss function. C is chosen to manage the trade-off between regularization and prediction error.

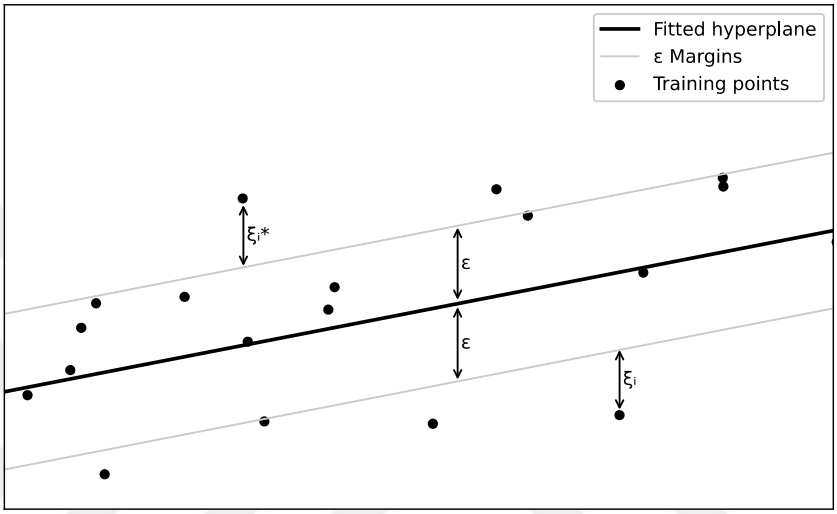

Fig. 7. The regression hyperplane and its margin in a one-dimensional linear SVR case [25].

$$y_i = f(x) = w^T x_i, \quad w, x \in \mathbb{R}^{d+1} \tag{20}$$

Eq. 20 shows the equation of the regression hyperplane in a linear case, where predictions (y) are written as a function of the d-dimensional input feature space (x).

$$Min\left\{ \frac{1}{2}\|w^2\| + C \sum_{i=1}^{N} (\xi_i + \xi_i^*) \right\} \tag{21}$$

$$\begin{cases} y_i - w^T x_i \leq \epsilon + \xi_i^* \\ w^T x_i - y_i \leq \epsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{22}$$

Eq. 21 represents the loss function of the SVR algorithm [25], where w is the vector normal to the regression hyperplane, C is the regularization factor, and ξ_i, ξ_i^* are the errors

outside the tolerance margin. The constraining conditions of the loss function are given in Eq. 22. By minimizing w in Eq. 21, the risk of overfitting is decreased by fitting a flatter hyperplane to the training data. The value of C determines how much weight is given to errors in the loss function. A small value for C will minimize the influence of errors and therefore result in a flatter regression hyperplane. In contrast, a large value for C will magnify the influence of errors and will result in a complex hyperplane that is prone to overfit the data and produce bad predictions from unseen data.

As described thus far, the SVR algorithm can work well in regressing variables with a linear relationship. However, in cases where the relationship is non-linear, a hyperplane that fits the data well enough cannot be found in the feature space. In this case, kernels are used to map the feature space to a higher dimensional space in which it is possible to fit a hyperplane to the data. In this study, the radial basis function (RBF) (Eq. 23), which is commonly used to model smooth and stationary functions [26], was used in modelling the reference evapotranspiration data.

$$K_{RBF}(x_i, x_j) = e^{-\gamma \cdot (x_i - x_j)^2} \tag{23}$$

The value of γ in the RBF function (Eq. 23) affects the degree of influence that each training point has on the others, which in turn affects the smoothness of the fitted hyperplane. As γ increases, the influence of each training data point on other points decreases. In this study, the values of the regularization term (C) and the RBF kernel parameter γ were chosen using a grid search. In a grid search, different combinations of these parameters are tested iteratively on the training data using cross-validation. In cross-validation, training data is split to k-folds, where k-1 folds are used to train a model and the remaining fold is used for testing. This process is repeated k times and the average test score of the k models is used to grade the performance of the studied combination of parameters.

Finally, all input features were scaled using Eq. 24 to avoid features with greater values dominating other features due to differences in scale.

$$x_{s,i} = \frac{x_i - \mu}{\sigma} \tag{24}$$

Where $x_{s,i}$ is the scaled feature, x_i is the original feature value, μ is the mean of the feature, and σ is the standard deviation of the feature. μ and σ were both computed from the training data to prevent any information leakage from the test data.

2.4.2 Gaussian Process Regression (GPR)

In Gaussian Process Regression (GPR), a posterior distribution over functions that maps the input features to the target variable is obtained by Bayesian inference from a prior distribution and the likelihood of the training data [26]. The prior distribution is a multivariate Gaussian distribution that models our prior beliefs about the predicted variable. The mean of the prior distribution is usually set to zero when the target variable can be centered around zero. The covariance matrix is defined using a kernel function that models the dependence between the data points based on their distances. The choice of the kernel function determines the shape of the resulting functions. In this study, the radial basis function (RBF) kernel (Eq. 25) was used, as it is suitable for modelling stationary and smooth functions.

$$K_{RBF}(x_i, x_j) = \sigma^2 \cdot e^{-\frac{1}{2}\left(\frac{\|x_i - x_j\|}{l}\right)^2} \tag{25}$$

Where $\|x_i - x_j\|$ is the Eucledean distance between points x_i and x_j. The choice of the length scale (l) of the RBF kernel function scales the distance between data points, therefore affecting the degree of influence between these points. Smaller l values reduce the influence of points on farther points, therefore resulting in a more rugged function. The variance (σ^2) uniformly scales the covariance between all points. The choice of the kernel parameters is done by maximizing the likelihood of the training data [27].

The mean and covariance of the posterior are subsequently obtained by Bayesian inference. The result is a distribution over functions, which translates to a Gaussian distribution describing each prediction, hence providing an uncertainty measure for the predictions. A one-dimensional Gaussian process is visualized in Fig. 8 showing the training points, test points, and function samples from the prior and the posterior. See [26] for a comprehensive description of Gaussian processes.

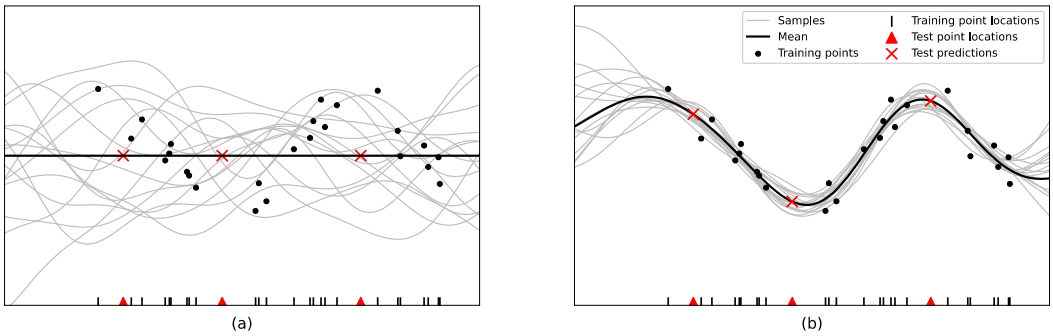

Fig. 8. One-dimensional GP samples and test data predictions [26] (a) Sampling from the prior (b) Sampling from the posterior

Maximizing the likelihood function using all training data requires the inversion of an $n \times n$ matrix, where n is the size of the training data, which makes it infeasible for large datasets. Sparse Variational Gaussian Processes (SVGPR) use a subset of the training data to optimize the model parameters, therefore permitting the use of GPR with large datasets [28]. Instead of defining the prior on all the training data (n), it is defined using a subset of the data (n_s) called the inducing variables, where $n_s < n$, assuming that the training data is dependent on the inducing variables. The SVGPR algorithm was used in this study.

Finally, all input features were standardized using Eq. 9 similarly to SVR.

2.4.3 Random Forests (RF)

The Random Forests (RF) regression algorithm, first introduced by Breiman [29], is an ensemble method that uses regression trees as its weak learners. A regression tree is a tree structure that starts with one node containing all the data, subsequently splitting every node into two nodes by choosing the split, out of all possible splits at this node, that minimizes a defined loss function [30]. At each node, a certain value (C) of one of the input variables (X_i) is chosen to create the split, where all data points that satisfy $X_i < C$ are assigned to one node and the remaining data points are assigned to the other node. Eq. 26 represents the loss function used to create a split in regression trees, where $Q_{L,R}$ is typically chosen as the squared residuals in the resulting left and right nodes (Eq. 27) [30].

$$Q_{split} = n_L Q_L + n_R Q_R \tag{26}$$

Where n_L and n_R are the number of data points assigned to the left and right nodes, respectively; and Q_L and Q_R are the squared residuals in the left and right nodes, respectively (Eq. 27).

$$Q = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}) \tag{27}$$

Where n is the number of data points in the concerning node, y_i is the ith target variable, and $\bar{y}$ is the average of all target variables in this node. Terminal nodes or leaves, i.e., nodes that are not split, are reached either when there is no more data to split, or when a predefined criterium such as the maximum height of a tree is met. A prediction is done by passing a new data point through the tree until it reaches one of the terminal nodes, where the prediction is the average of the target values in this terminal node.

In random forests, a prediction is made by averaging the predictions of all the regression trees in the random forest. The upper bound of the generalization error of a random forest (Eq. 28) is a function of the correlation between regression trees ($\bar{\rho}$) and their generalization errors ($PE^*(tree)$) [29].

$$PE^*(forest) \leq \bar{\rho} \cdot PE^*(tree) \tag{28}$$

Therefore, reasonable randomness must be introduced to the regression trees to minimize the correlation between them without significantly sacrificing their accuracies. For this reason, regression trees are trained on bootstrapped data, i.e., data sampled with replacement from the original data. Further randomness can be added to the trees by using a subset of the input features for node splits. However, using a small number of features for node splits in regression trees has been found to notably increase the tree errors, hence undermining the effect of randomization on the random forest error upper bound [29].

In this study, three RF parameters were tuned: the number of decision trees, minimum number of samples in the resulting nodes to allow for a split, and the size of the bootstrap sample used to train individual decision trees. These parameters were chosen by

iteratively testing diffferent combinations using a grid search and cross validation as explained in section 2.4.1.

2.5 Evaluation Metrics

In this study, four model performance evaluation metrics were used, two unitless metrics: relative root mean squared error (rRMSE) and the Nash–Sutcliffe model efficiency coefficient (NSE); and two metrics that have the same unit as the predicted target variable: root mean squared error (RMSE) and mean absolute error (MAE). RMSE (Eq. 29) values range between 0 for a perfect model and $+\infty$. rRMSE (Eq. 30) is a normalized version of RMSE and its value is a percentage of the average of the measured variable. Both RMSE and rRMSE use the squared residuals and therefore penalize larger errors. MAE (Eq. 31), which is the average of the absolute residuals, treats all errors equally. Finally, the NSE (Eq. 32) computes the ratio of the variance explained by the model to the total variance in the data, where a perfect model would have an NSE value of 1 and values less than zero indicate that the mean is a better predictor than the studied model [31]. Table 4 provides a classification of model performance based on the NSE metric [32].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{29}$$

$$rRMSE = \frac{RMSE}{\bar{O}} \tag{30}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \tag{31}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \quad (32)$$

Where P_i represents model predictions, O_i represents ET_0 observations, $\bar{P}$ is the average of model predictions, and $\bar{O}$ is the ET_0 observations average.

Table 4. Performance ratings by NSE values

Range of NSE	Performance
NSE ≤ 0.4	Unsatisfactory
0.40 < NSE ≤ 0.5	Acceptable
0.50 < NSE ≤ 0.65	Satisfactory
0.65 < NSE ≤ 0.75	Good
0.75 < NSE ≤ 1.0	Very good

2.6 Statistical Significance Test

Although the seven regions used in this study are the commonly accepted seven climatic regions of Turkey, Erinc, S. [21] did not rely merely on climatic factors to define these regions as mentioned earlier. Therefore, a statistical significance test was necessary to ensure that the observed difference in prediction accuracy between general and regional models is not obtainable by random assignment of stations to clusters. In other words, if this specific division of regions is of any importance in regards of ET_0 prediction using ML models, it would be hard to obtain similar improvements in prediction accuracy from other randomly defined clusters of stations. In this test, the null hypothesis states that using these specific regional models does not improve prediction accuracy over the general model; and the observed improvement is random and may also be obtained by random allocation of stations to seven clusters. The alternative hypothesis, on the other hand, states that using this setup results in higher prediction accuracy than the general model. The test was conducted on four of the input combinations, which were chosen based on their performance and limited climatic variable requirements. Due to the complexity of this test, it was not possible to use any of the traditional statistical significance tests. Therefore, a randomization test was conducted as described in the following steps:

1. Randomly assign stations to seven clusters
2. Split stations in each cluster to train and test stations by allocating approximately 40% of the data for testing.
3. Train an ML model for each of the seven clusters and one general model using the training data of all clusters combined.
4. Measure the change between the combined cluster models' RMSE and the general model RMSE (Eq. 33).

$$RMSE\ change\ \% = \frac{\left(RMSE_{cluster} - RMSE_{general}\right)}{RMSE_{general}} \tag{33}$$

5. Repeat steps 1 to 4 1000 times.
6. Compute the ratio of iterations that had a reduction in RMSE equal or greater than the observed RMSE reduction to get a one-sided p-value.

A similar test was used Park et al. [33] to evaluate the significance of each split in a hierarchical clustering dendrogram. At each node of the dendrogram, the clusters were permuted, and the intra-cluster variance of permuted clusters was computed and compared to the intra-cluster variance of the clusters obtained at this node to get a p-value that describes the significance of this split. The previous process was also used by Bruzzes to automatically return the significant clustering scheme from the results of a hierarchical clustering process [34].

The numbers of data points in the randomized clusters had to be close to that of the clusters used in this study (see Fig. 6) to ensure that the test simulates the conducted study. On that account, a 7th order ($k = 7$) Dirichlet distribution [35] was used to generate the number of data points per cluster in each iteration of the randomization test (Eq. 34-36).

$$f(y^k) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\prod_{i=1}^{k} y_i^{\alpha_i - 1} \tag{34}$$

$$\alpha_0 = \sum_{i=1}^{k} \alpha_i \tag{35}$$

$$E[y_i] = \frac{\alpha_i}{\alpha_0} \tag{36}$$

$$\alpha_i = \frac{\beta n_i}{N} \tag{37}$$

The seven parameters of the Dirichlet distribution (α_i) were set as defined in Eq. 37, where n_i is the number of data points in the corresponding region of the original study, N is the total size of the dataset, and β is the concentration parameter. The concentration parameter value determines how close the sampled values are to the expected value, which is computed using Eq.36. This choice of the Dirichlet parameter results in an expected value equal to $\frac{n_i}{N}$ for each value of the seven-dimensional vector, which is equal to the ratio of the cluster data to the whole dataset. Multiplying this vector by the dataset size results in a seven-dimensional vector, where each value represents the sampled number of datapoints in one of the seven clusters and all values sum up to the dataset size N. Using this method, a sample was drawn from this distribution in each iteration of the randomization test and stations were iteratively allocated to each cluster until the number of data points in the cluster reached the sampled number.

Fig. 9 shows a box plot of the number of datapoints in each of the seven randomized clusters, with the diamond marker referring to the number of datapoints in the region that this cluster is simulating. For example, cluster 1 simulates the region with the greatest number of datapoints, i.e., Central Anatolia.

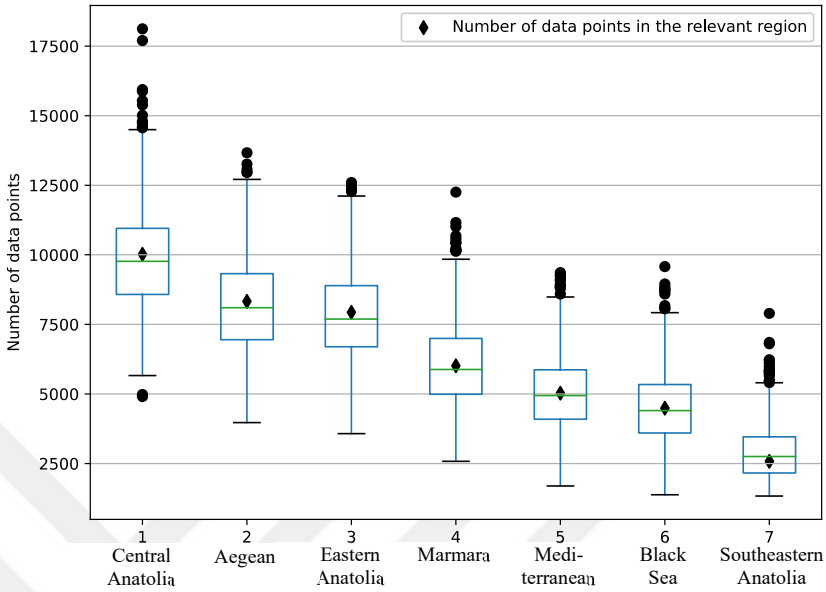

Fig. 9. Box plot showing the distribution of the dataset sizes of the randomized clusters

The significance level used in this test was a Bonferroni corrected significance level [36] that achieves a family-wise false positive rate of 5%. This corresponds to a cut-off value equal to $\frac{0.05}{4} = 0.0125$, where 4 is the number of tests conducted.

Finally, the designed randomization test requires training 8 models in each of the 1000 iterations. Using one of the three models used in this study (SVR, RF, or GPR) for this purpose was impractical given the long training time of these algorithms. Accordingly, a second order polynomial was opted for to train the models in the randomization test. This choice was justified by the similar behavior of the second order polynomial model to SVR and GPR models as demonstrated in section 3.2.

CHAPTER 3

RESULTS

3.1. Input combination comparison

The two processes that constitute evapotranspiration are mass transfer and energy balance [37]. In reference evapotranspiration, the surface properties are constant by definition. Under this condition, mass transfer is determined by vapor pressure deficit and aerodynamic resistance only, which in turn depend on relative humidity, wind speed, and partially on temperature as it affects saturation vapor pressure. Meanwhile energy balance is influenced by solar radiation and ambient temperature. The 16 input combinations used in this study will be compared in the following paragraph taking into consideration the variables that influence mass transfer and energy balance. The model performance metrics of the general and regional models are presented in Table 5 and Table 6. The metrics are also presented in Fig. 10 for easier comparison.

Table 5. MAE and RMSE of general and regional models

Input	Model	Metric					
		MAE			RMSE		
		GPR	RF	SVR	GPR	RF	SVR
C1	General	0.389	0.392	0.388	0.503	0.507	0.507
	Regional	0.352	0.364	0.352	0.456	0.476	0.463
C2	General	0.758	0.751	0.751	0.965	0.954	0.965
	Regional	0.650	0.654	0.649	0.846	0.842	0.851
C3	General	0.225	0.235	0.227	0.314	0.327	0.320
	Regional	0.206	0.244	0.208	0.288	0.337	0.293
C4	General	0.669	0.669	0.668	0.855	0.855	0.855
	Regional	0.647	0.645	0.642	0.834	0.831	0.828
C5	General	0.674	0.673	0.671	0.861	0.861	0.859
	Regional	0.657	0.650	0.653	0.843	0.835	0.839
C6	General	0.667	0.668	0.668	0.853	0.853	0.854
	Regional	0.654	0.653	0.654	0.839	0.836	0.840
C7	General	0.516	0.518	0.515	0.673	0.676	0.675
	Regional	0.496	0.501	0.493	0.665	0.673	0.663
C8	General	1.160	1.144	1.135	1.389	1.382	1.397
	Regional	1.155	1.146	1.128	1.399	1.390	1.403
C9	General	0.525	0.527	0.520	0.682	0.687	0.681
	Regional	0.517	0.518	0.511	0.689	0.683	0.679
C10	General	0.622	0.618	0.614	0.813	0.809	0.810
	Regional	0.576	0.582	0.574	0.759	0.768	0.762
C11	General	0.626	0.625	0.621	0.818	0.817	0.818
	Regional	0.577	0.582	0.577	0.757	0.765	0.760
C12	General	0.460	0.463	0.458	0.613	0.614	0.614
	Regional	0.438	0.446	0.438	0.588	0.599	0.593
C13	General	0.538	0.542	0.535	0.699	0.705	0.706
	Regional	0.486	0.502	0.488	0.642	0.664	0.653
C14	General	0.226	0.228	0.225	0.318	0.321	0.324
	Regional	0.207	0.236	0.210	0.298	0.332	0.305
C15	General	0.138	0.157	0.135	0.197	0.216	0.196
	Regional	0.115	0.180	0.115	0.163	0.267	0.163
C16	General	0.225	0.234	0.226	0.314	0.324	0.318
	Regional	0.209	0.243	0.210	0.299	0.335	0.298

Table 6. rRMSE and NSE of general and regional models

Input	Model	Metric					
		rRMSE			NSE		
		GPR	RF	SVR	GPR	RF	SVR
C1	General	0.132	0.133	0.133	0.921	0.920	0.920
	Regional	0.120	0.125	0.121	0.935	0.930	0.933
C2	General	0.253	0.250	0.253	0.711	0.717	0.711
	Regional	0.222	0.221	0.223	0.777	0.780	0.775
C3	General	0.082	0.086	0.084	0.969	0.967	0.968
	Regional	0.076	0.088	0.077	0.974	0.965	0.973
C4	General	0.224	0.224	0.224	0.773	0.773	0.773
	Regional	0.219	0.218	0.217	0.784	0.785	0.787
C5	General	0.226	0.226	0.225	0.770	0.770	0.771
	Regional	0.221	0.219	0.220	0.779	0.783	0.781
C6	General	0.224	0.224	0.224	0.774	0.774	0.773
	Regional	0.220	0.219	0.220	0.781	0.783	0.781
C7	General	0.177	0.177	0.177	0.859	0.858	0.858
	Regional	0.174	0.177	0.174	0.863	0.859	0.863
C8	General	0.364	0.362	0.366	0.401	0.407	0.394
	Regional	0.367	0.365	0.368	0.392	0.400	0.389
C9	General	0.179	0.180	0.179	0.856	0.853	0.856
	Regional	0.181	0.179	0.178	0.852	0.855	0.857
C10	General	0.213	0.212	0.212	0.794	0.796	0.796
	Regional	0.199	0.201	0.200	0.821	0.817	0.820
C11	General	0.215	0.214	0.215	0.792	0.793	0.792
	Regional	0.199	0.201	0.199	0.822	0.818	0.821
C12	General	0.161	0.161	0.161	0.883	0.883	0.883
	Regional	0.154	0.157	0.156	0.892	0.888	0.891
C13	General	0.183	0.185	0.185	0.848	0.846	0.845
	Regional	0.168	0.174	0.171	0.872	0.863	0.868
C14	General	0.083	0.084	0.085	0.969	0.968	0.967
	Regional	0.078	0.087	0.080	0.972	0.966	0.971
C15	General	0.052	0.057	0.051	0.988	0.985	0.988
	Regional	0.043	0.070	0.043	0.992	0.978	0.992
C16	General	0.082	0.085	0.083	0.969	0.967	0.969
	Regional	0.079	0.088	0.078	0.972	0.965	0.972

The results clearly show the superiority of energy-balance based models (C7, C12, and C13) to the mass-transfer based model (C8). Combination C1, which combines energy-balance (T_{max}) and mass-transfer (U_2) variables, predicts ET_0 with an rRMSE of approximately 13% and 12% using the general and regional models, respectively. The performance of C1 is notably better than C2 (rRMSE = 18%), in which R_s is used instead of T_{max}. This may be attributed to the fact that T_{max} carries information about both evapotranspiration processes, i.e., energy-balance and mass-transfer, as it influences saturation vapor pressure. In C3, the month of the year ($Month$) was added to C1, which greatly improved the models' prediction accuracy due to the information about extraterrestrial radiation contained in the month variable. Furthermore, comparing combinations C1 (T_{max} & U_2) and C2 (T_{min} & U_2), T_{max} predicted ET_0 with significantly higher accuracy than T_{min}, which was also noted in [38]. Comparing the two mass-transfer variables, RH and U_2, wind speed was clearly a better predictor than relative humidity (compare C1 (T_{max} & U_2) and C9 (R_s & U_2) to C6 (T_{max} & RH_{avg}) and C11 (R_s & RH_{avg}), respectively). Moreover, the performance of C16 (T_{max} & U_2 & $Month$ & RH_{max}) is almost identical to C3, which does not include RH_{max}. C15 was the best performing input combination, which is expected given that it includes 4 out of the 5 variables used in ET_0 computation. Overall, C3 (T_{max} & U_2 & $Month$) demonstrated the best balance between prediction error (rRMSE $\approx$ 8%) and the number of measurements required (temperature and wind speed). This model's NSE value of 0.974 renders it way above the "Very good" criterium of 0.75 stated in Table 4. This also applies to the combinations C7 and C13, both using one measured variable only, with NSE values of 0.863 and 0.872, respectively.

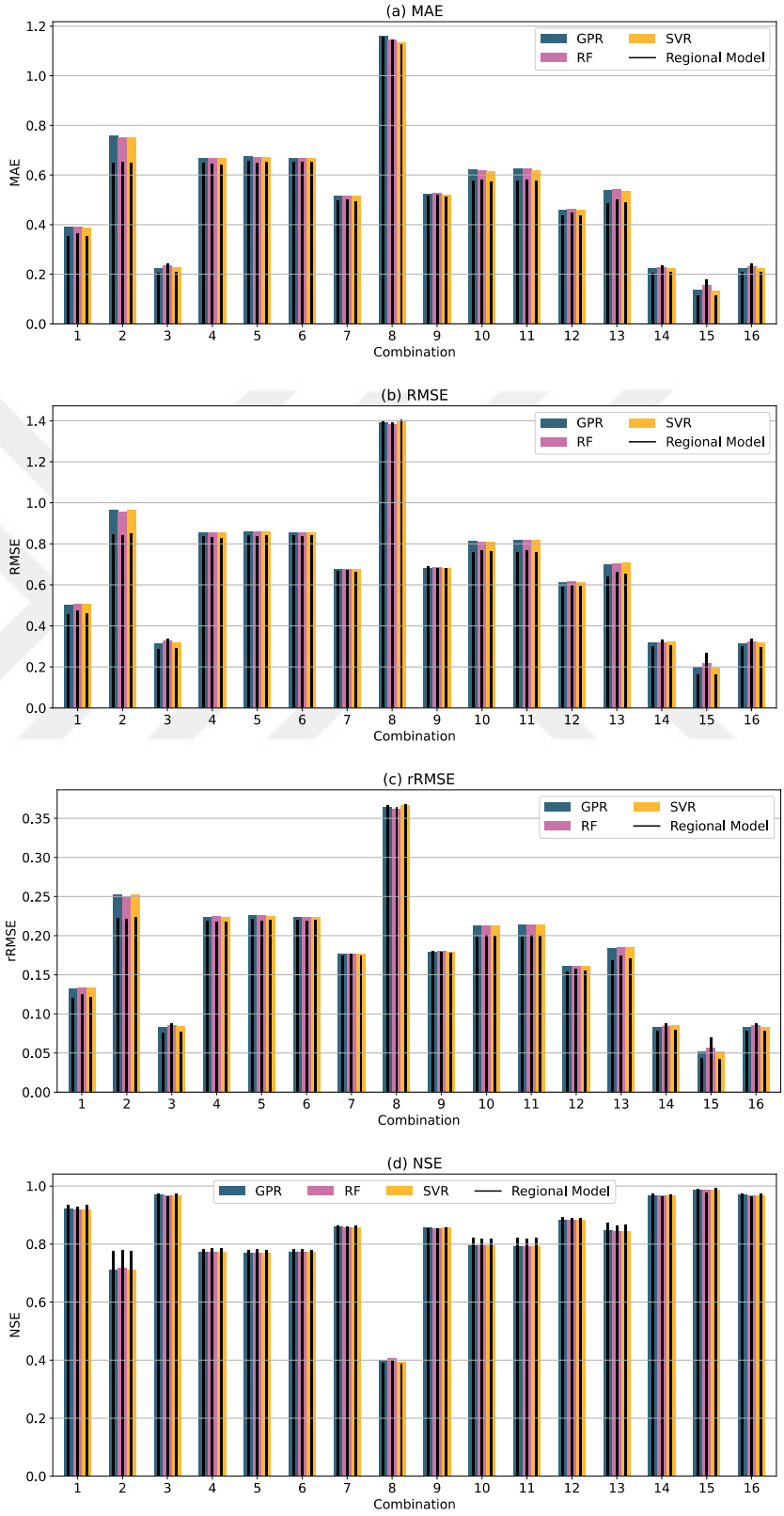

Fig. 10. General and regional model metrics (a) MAE (b) RMSE (c) rRMSE (d) NSE

3.2 Comparing general and regional models

The thin black bars in Fig. 10 show the performance metrics of the regional models of the corresponding algorithms and input combinations. Fig. 11 shows the percentage change in MAE and RMSE between general and regional models as computed in Eq. 33 for each input combination.

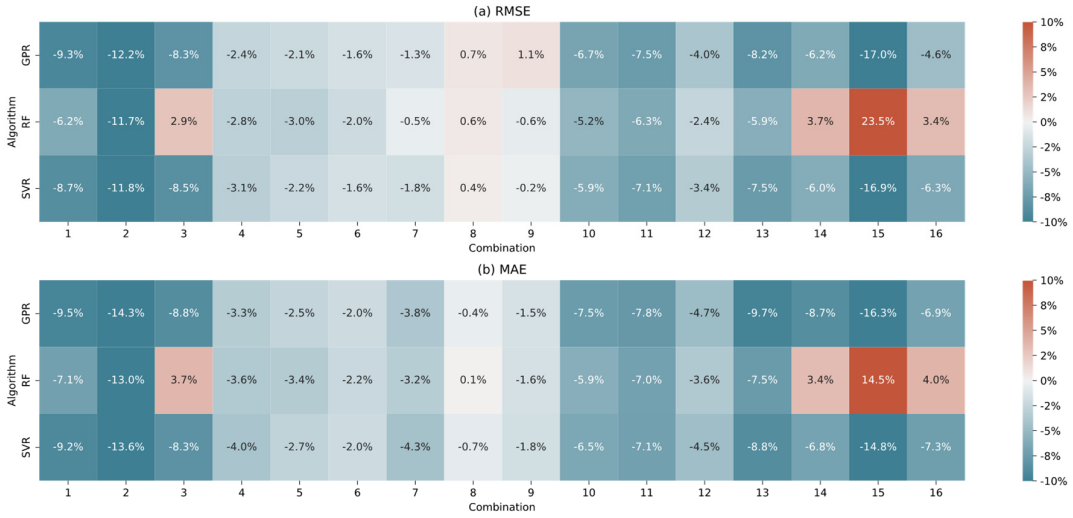

Fig. 11. Change in prediction error between general and regional models. Negative values indicate lower error in the regional model case. (a) % Change in RMSE (b) % Change in MAE.

SVR and GPR regional models performed better than the general models in almost all input combinations. On the other hand, RF regional model performances were worse than the general model in several combinations due to overfitting. Fig. 12 demonstrates the overfitting in RF in the four input combinations in which the regional RF models performed notably worse than the general model. It can be observed that the difference between train and test prediction error when using RF was larger than SVR and GPR models. This effect was more profound when using regional models.

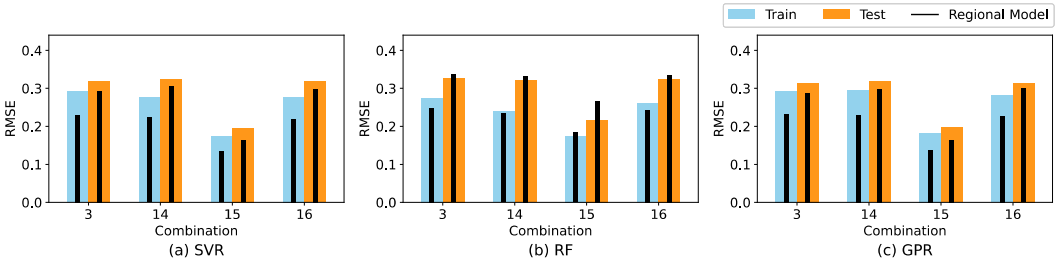

Fig. 12. Train and test RMSE of general and regional models. The colored bars represent the RMSE of the general model and the thin black bars represent the RMSE of the regional models. (a) SVR (b) RF (c) GPR

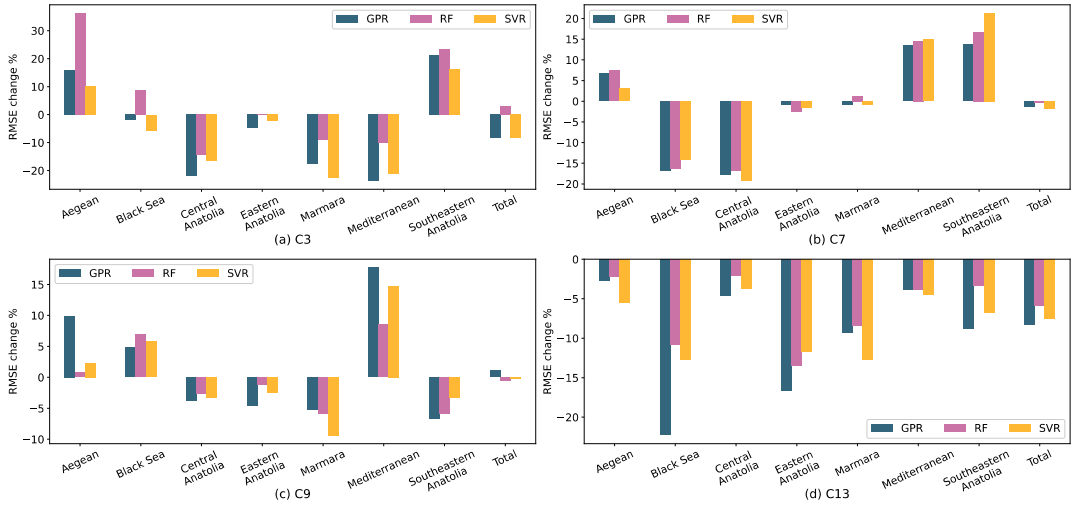

Fig. 13. Change in RMSE between general and regional models per region

Given that the purpose of research on predicting ET_0 using ML is to be able to estimate ET_0 using limited climatic measurements, the best performing combinations that use 1 or 2 measured variables are chosen for further discussion and significance testing. These combinations are: C3 (T_{max} & U_2 & $Month$), C7 (T_{max} & $Month$), C9 (R_s & U_2), and C13 (R_s & $Month$). These combinations cover different cases of climatic measurement availability. Fig. 13 shows the change in RMSE between general and regional models as a percentage of the general model RMSE (Eq. 33). Negative values indicate lower RMSE in the regional model case. It can be noticed that for each region, some combinations produce better predictions when using the regional model and other combinations are better when using the general model. This is dependent on a number of factors like the

size of the regional dataset (Fig. 6), the difference in distributions between regional and general datasets (Fig. 5), the difference in distribution between regional train and test datasets (Fig. 5), and overfitting when using the regional dataset. However, when looking at the combined performance, regional models perform better in the majority of input combinations.

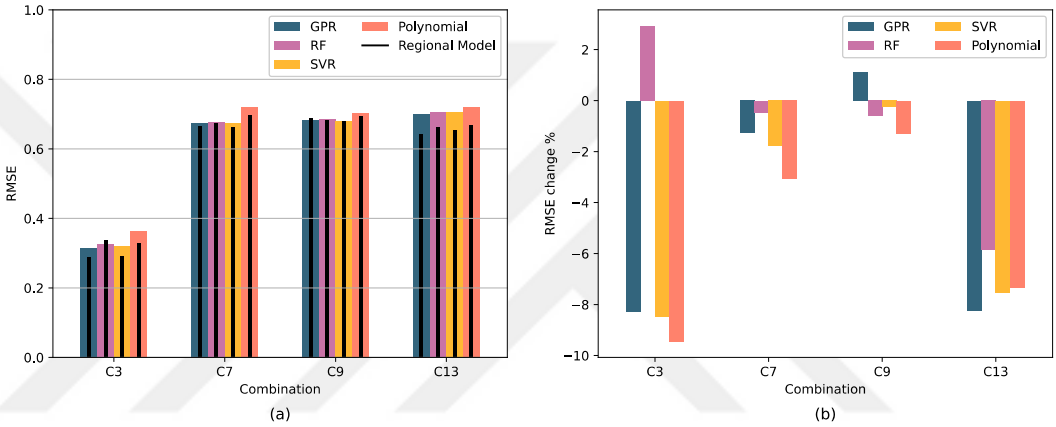

Fig. 14. (a) RMSE of general and regional models (b) Change in RMSE between general and regional models.

The significance of the improvement in prediction accuracy was tested by conducting a randomization test. As mentioned in section 2.6, the randomization test was conducted using a second-order polynomial regression model instead of the models used in this study. Fig. 14 shows the RMSE of the polynomial regression model and the RMSE change between general and regional models for the 4 chosen combinations. The prediction error of the polynomial regression model is expectedly higher than the models used in this study. However, the change in RMSE between general and regional model cases, which is the focus of the randomization test, is comparable to that of SVR and GPR models. Therefore, the polynomial model was used as an approximation of the other models in the randomization test.

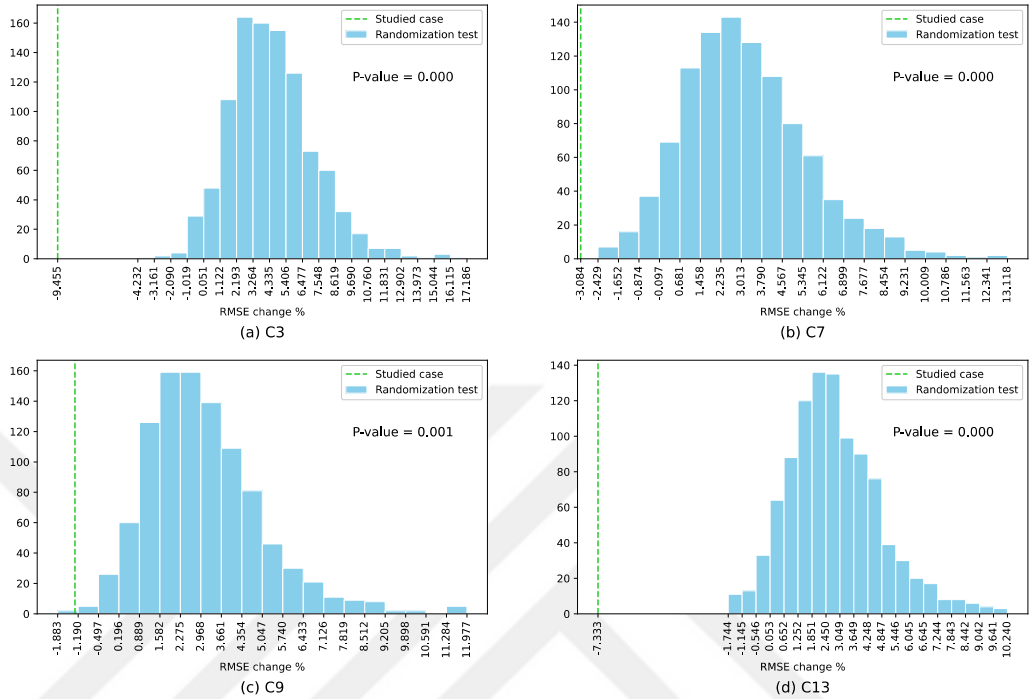

Fig. 15. Randomization test distributions and p-values

The results of the 4 randomization tests are presented in Fig. 15. All 4 combinations showed a statistically significant reduction in prediction error when using the regional models. However, from a practical perspective, the reduction in prediction error in combinations C7 and C9 (2%) is negligible in a real-case application.

CHAPTER 4

DISCUSSION AND CONCLUSIONS

In this study, two different approaches for modelling reference evapotranspiration in Turkey using machine learning models were compared. The two approaches used were regional and general ML models. The regional models were trained on data from each of the seven regions of Turkey, i.e., Mediterranean, Aegean, Marmara, Black Sea, Central Anatolia, Eastern Anatolia, and Southeastern Anatolia. The stations in each region were split to train and test stations to allow for a cross-station evaluation. The general model was trained and tested on the combined train and test stations of all the geographic regions. The comparison was made using 3 ML algorithms, SVR, RF, and GPR. Furthermore, 16 climatic variable combinations were tested for ET_0 estimation to account for different measurement availability cases. Finally, a randomization test was conducted to evaluate the statistical significance of the reduction in prediction error when using regional models in comparison to that of the general model.

In the general model case, all three algorithms performed very similarly, with SVR and GPR having a slight advantage over RF due to overfitting in some RF models. However, in the regional model case, the superiority of SVR and GPR models over RF was more pronounced due to the increased overfitting that RF exhibited on the smaller datasets of regional models. Between GPR and SVR, GPR showed a better performance in general, making it the best model out of the three algorithms used in this study. These findings conform with the conclusions made by Shabani et al. in their study on predicting pan evaporation using GPR, SVR, RF, and KNN [39]. In addition, the ability of the GPR algorithms to provide uncertainty estimates with the predictions makes it more suitable for real-life applications.

Regarding input combinations, solar radiation combined with the month of the year was the best predictor of ET_0 when using one measured variable. Sattari et al. also found the best single predictor of ET_0 to be sunshine duration, which is a proxy for solar radiation [38]. For combinations using two measured meteorological variables, the combination of maximum temperature, average wind speed, and month of year resulted in an rRMSE of about 7.5% and an NSE value of about 0.97 when using the regional model.

The observed change in prediction error between general and regional models using SVR and GPR algorithms ranged between -16.3% and -0.4% for MAE and between -17.0% and 1.1% for RMSE, where negative values indicate that the prediction error in the regional model case was lower. Furthermore, the randomization test, which was conducted on four input combination models, showed that the reduction in RMSE was statistically significant. These results confirm the findings of Yildirim et al. which stated that the sensitivity of computed ET_0 values to errors in the input variables differs from region to another in Turkey [13]. Moreover, this improvement in prediction accuracy when using regional models is similar to the improvement noted by Ferreira et al. when creating models using clustered stations [16]. However, the implication of this study is more realistic as the definition of geographic regions allows for assigning new locations to their relevant regions; whereas the clustering method used in the previous study requires the availability of ET_0 values in the new location.

The results of this study emphasize the importance of determining the ideal geographical extent in which ML models can be applied to estimate reference evapotranspiration using limited climatical variables. This is especially important in areas with considerable spatial climatic variation like Turkey. Using individual models for each region resulted in lower prediction error than using one model for the whole country. However, the improvement in prediction accuracy was not uniform among all regions. Although the prediction accuracy improved in total when using the regional models, some regions still performed better in the general model case. Further research on clustering stations for the purpose of ET_0 estimation is required for defining regions that suite this purpose more than the regions used in this study, which are not merely defined as climatic regions. Moreover, similar studies in other parts of the world are necessary to obtain a better understanding of the spatial generalizability of ET_0 estimating ML models. Furthermore, studies using

daily meteorological data can simulate the real application of ET_0 estimating models better than monthly data given the daily requirement of ET_0 values for irrigation scheduling. Finally, studies using actual evapotranspiration measurements will provide more realistic insights by excluding the noise introduced by estimating reference evapotranspiration using the FAO 56 PM equation.

REFERENCES

1. Arslan, O., 2020. Changes in Crop and Irrigation Water Requirements in Niğde. **International Scientific and Vocational Studies Journal**, **4** (1): 68–74.

2. Cakmak, B., Ucar, Y., Akuzum, T., 2007. Water Resources Management, Problems and Solutions for Turkey. **International Congress on River Basin Management**, **1** (1): 867–880.

3. Brouwer, C., & Heibloem, M., 1986. Irrigation water management: irrigation water needs. **Training manual**, **3**.

4. Allen, R. G., Pereira, L. S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. **FAO, Rome**, **300** (9):.

5. Zhao, L., Xia, J., Xu, C. yu, Wang, Z., Sobkowiak, L., Long, C., 2013. Evapotranspiration estimation methods in hydrological models. **Journal of Geographical Sciences**, **23** (2): 359–369.

6. Nicolás-Cuevas, J. A., Parras-Burgos, D., Soler-Méndez, M., Ruiz-Canales, A., Molina-Martínez, J. M., 2020. Removable weighing lysimeter for use in horticultural crops. **Applied Sciences**, **10** (14): 4865.

7. Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., Ramkumar, P. N., 2020. Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. **Current Reviews in Musculoskeletal Medicine**, **13** (1): 69–76.

8. Citakoglu, H., Cobaner, M., Haktanir, T., Kisi, O., 2014. Estimation of Monthly

   Mean Reference Evapotranspiration in Turkey. **Water Resources Management**, **28** (1): 99–113.

9. Wen, X., Si, J., He, Z., Wu, J., Shao, H., Yu, H., 2015. Support-Vector-Machine-Based Models for Modeling Daily Reference Evapotranspiration With Limited

Climatic Data in Extreme Arid Regions. **Water Resources Management**, **29** (9): 3195–3209.

10. Wu, L., Peng, Y., Fan, J., Wang, Y., 2019. Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on cross-station and synthetic data. **Hydrology Research**, **50** (6): 1730–1750.

11. Landeras, G., Bekoe, E., Ampofo, J., Logah, F., Diop, M., Cisse, M., Shiri, J., 2018. New alternatives for reference evapotranspiration estimation in West Africa using limited weather data and ancillary data supply strategies . **Theoretical and Applied Climatology**, **132** (3): 701–716.

12. Zhang, X., Ren, Y., Yin, Z. Y., Lin, Z., Zheng, D., 2009. Spatial and temporal variation patterns of reference evapotranspiration across the Qinghai-Tibetan Plateau during 1971-2004. **Journal of Geophysical Research Atmospheres**, **114** (D15):.

13. Yildirim, T., Wagle, P., Gowda, P. H., Mengu, G. P., 2021. Sensitivity of reference evapotranspiration to weather variables across seven regions of Turkey. **Agrosystems, Geosciences & Environment**, **4** (2):.

14. Cobaner, M., Citakoglu, H., Haktanir, T., Kisi, O., 2017. Modifying Hargreaves-Samani equation with meteorological variables for estimation of reference evapotranspiration in Turkey. **Hydrology Research**, **48** (2): 480–497.

15. Shiri, J., Nazemi, A. H., Sadraddini, A. A., Landeras, G., Kisi, O., Fard, A. F., Marti, P., 2014. Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran. **Computers and Electronics in Agriculture**, **108**: 230–241.

16. Ferreira, L. B., da Cunha, F. F., de Oliveira, R. A., Fernandes Filho, E. I., 2019. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – A new approach. **Journal of Hydrology**, **572**: 556–570.

17. Ferreira, L. B., Cunha, F. F. D. A., Silva, G. H. D. A., Campos, F. B., Dias, S. H.

B., Santos, J. E. O., 2021. Generalizability of machine learning models and empirical equations for the estimation of reference evapotranspiration from temperature in a semiarid region. **Anais da Academia Brasileira de Ciências**, **93**:.

18. Deniz, A., Toros, H., Incecik, S., 2011. Spatial variations of climate indices in Turkey. **International Journal of Climatology**, **31** (3): 394–403.

19. Iyigun, C., Türkeş, M., Batmaz, I., Yozgatligil, C., Purutçuoğlu, V., Koç, E. K., Öztürk, M. Z., 2013. Clustering current climate regions of Turkey by using a multivariate statistical method. **Theoretical and Applied Climatology**, **114** (1): 95–106.

20. Türkeş, M., Tatli, H., 2011. Use of the spectral clustering to determine coherent precipitation regions in Turkey for the period 1929-2007. **International Journal of Climatology**, **31** (14): 2055–2067.

21. Unal, Y., Kindap, T., Karaca, M., 2003. Redefining the climate zones of Turkey using cluster analysis. **International Journal of Climatology: A Journal of the Royal Meteorological Society**, **23** (9): 1045–1055.

22. Erinç, S. 1984. Climatology and its methods. Marine Science, Institute of Geography, Istanbul University Press: Istanbul, Turkey (in Turkish).

23. Boser, B. E., Guyon, I. M., Vapnik, V. N., 1992. A training algorithm for optimal margin classifiers, pp. 144–152. *Proceedings of the fifth annual workshop on Computational learning theory*.

24. Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., Vapnik, V., 1996. Support vector regression machines. **Advances in Neural Information Processing Systems**, **9**:.

25. Awad, M., Khanna, R., 2015. Support vector regression, pp. 67-80. *In*: Efficient Learning Machines (M. Awad, R. Khanna). Apress, Berkeley, CA.

26. Schulz, E., Speekenbrink, M., Krause, A., 2018. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. **Journal of Mathematical Psychology**, **85**: 1–16.

27. Williams, C. K. I., Rasmussen, C. E., 1995. Gaussian Processes for Regression. **Advances in Neural Information Processing Systems**, **8**:.

28. Hensman, J., Fusi, N., Lawrence, N. D., 2013. Gaussian Processes for Big Data. **arXiv preprint arXiv:1309.6835**,.

29. Breiman, L., 2001. Random Forests. **Machine learning**, **45** (1): 5–32.

30. Cutler, A., Cutler, D. R., Stevens, J. R., 2012. Random forests, pp. 157-175. *In*: Ensemble machine learning (A. Cutler, D. R. Cutler, J. R. Stevens,). Springer, Boston, MA.

31. Schäuble, H., Marinoni, O., Hinderer, M., 2008. A GIS-based method to calculate flow accumulation by considering dams and their specific operation time. **Computers and Geosciences**, **34** (6): 635–646.

32. Nash, J. E., Sutcliffe, J. V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. **Journal of hydrology, 10** (3): 282-290.

33. Park, P. J., Manjourides, J., Bonetti, M., Pagano, M., 2009. A permutation test for determining significance of clusters with applications to spatial and gene expression data. **Computational Statistics and Data Analysis**, **53** (12): 4290–4300.

34. Bruzzese, D., 2015. DESPOTA : DEndrogram Slicing through a PemutatiOn Test Approach. **Journal of classification**, **32** (2): 285–304.

35. Lin, J., 2016. On The Dirichlet Distribution. Queens University, Department of Mathematics and Statistics, Master's Thesis, Ontario, 68 p.

36. Ranstam, J., 2016. Multiple P-values and Bonferroni correction. **Osteoarthritis and Cartilage**, **24** (5): 763–764.

37. Guo, D., Westra, S., Maier, H. R., 2016. An R package for modelling actual, potential and reference evapotranspiration. **Environmental Modelling and Software**, **78**: 216–224.

38. Sattari, M. T., Apaydin, H., Band, S. S., Mosavi, A., Prasad, R., 2021. Comparative

analysis of kernel-based versus ANN and deep learning methods in monthly reference evapotranspiration estimation. **Hydrology and Earth System Sciences**, **25** (2): 603–618.

39. Shabani, S., Samadianfard, S., Sattari, M. T., Mosavi, A., Shamshirband, S., Kmet, T., Várkonyi-Kóczy, A. R., 2020. Modeling pan evaporation using Gaussian Process Regression K-Nearest Neighbors Random Forest and support vector machines; comparative analysis. **Atmosphere**, **11** (1): 66.

CURRICULUM VITAE

PERSONAL INFORMATION

Name: Yasser Zouzou

Nationality: Syria

EDUCATION

Degree	Institute	Graduation date
MSc	Erciyes University – Graduate School of Natural and Applied Sciences	2022
BSc	Damascus University – Civil Engineering	2018
High school	Al-Saadeh School - Damascus	2013

LANGUAGES

1. Arabic
2. English

PUBLICATIONS

1. Zouzou, Y., & Citakoglu, H., 2021. Reference evapotranspiration prediction from limited climatic variables using support vector machines and Gaussian processes. **Avrupa Bilim ve Teknoloji Dergisi**, **28**: 346-351.