



Estimation of monthly sunshine duration using satellite derived cloud data

Kazım Kaba¹ · Erdem Erdi² · Mutlu Avcı³ · H. Mustafa Kandırmaz⁴

Received: 7 March 2022 / Accepted: 1 April 2024
© The Author(s) 2024

Abstract

Sunshine duration (SD) is one of the critical meteorological parameters used in different fields of application such as climate, renewable energy and agriculture. In this respect, determination and/or estimation of the temporal and spatial variability of SD is critical. Meteorological satellite data/products can be used for estimating SD and in constructing their maps due to their frequent observation of large areas at once. In this study, a multilayer perceptron type artificial neural network model was built to estimate the monthly mean SD for Türkiye using the EUMETSAT CM SAF (Satellite Application Facility on Climate Monitoring) CFC (Cloud Fractional Coverage) and CTY (Cloud Type) data, GMTED2010 (Global Multi-resolution Terrain Elevation Data) data, month number and daylength. The datasets of 45 stations, spanning nine years (2005–2013), were used for training the model and 12 stations for testing and validating the simulated values. We have compared the results of our model with the ground-measured values for the whole period under consideration and the root mean square error (RMSE), mean absolute error (MAE), mean bias error (MBE) and the coefficient of determination (R^2) were found as 0.7803 h, 0.6206 h, 0.1751 h and 0.9387, respectively. It has been shown that using the new generation cloud products such as CFC and CTY, elevation data such as GMTED2010 and daylength, it is possible to predict the SD for regions under the coverage of the satellite, in case no measurement is possible or may be unreliable, without needing any measured meteorological data.

1 Introduction

Sunshine duration (SD) is a precious climatic parameter because it is directly or indirectly used in many studies and applications. For example, SD is the most frequently used meteorological data for estimating the global solar radiation (GSR) (Badescu 1999; Li et al. 2011; Trnka et al. 2005). The

long time and short time measurements of SD are needed to detect climatic changes (Liao et al. 2021; Sanchez-Lorenzo et al. 2008; Stanhill and Cohen 2005; Wood and Harrison 2011; You et al. 2010; Matuszko and Węglarczyk 2015; Sanchez-Lorenzo and Wild 2012). SD data also widely employed in the studies of agriculture, hydrology, human health and tourism (Ahn et al. 2021; Akgün et al. 2021; Brown 2013; Hu et al. 2021; Liu et al. 2021; Mieczkowski 1985; Wang et al. 2021; Zhang et al. 2021). All these studies require accurate information about the spatial and temporal distribution of SD.

SD is defined as the length of time for which direct solar irradiance exceeds the level of 120 W/m^2 (WMO: World Meteorological Organization). Up until now, five different methods, namely, burn, pyr heliometric, pyronometric, contrast and scanning methods, have been used for measuring the value of SD in meteorological stations (Baumgartner et al. 2018). The burning method, first introduced by J.F.

✉ Kazım Kaba
kazimkaba@atauni.edu.tr

¹ Astronomy and Space Sciences, Atatürk University, Erzurum 25240, Türkiye

² Data Services User Support Office, EUMETSAT, Darmstadt, Germany

³ Department of Biomedical Engineering, Çukurova University, Adana, Türkiye

⁴ Department of Physics, Çukurova University, Adana, Türkiye

Campbell in 1853 and later modified by G. G. Stokes, is the most commonly used method and this type of measuring device is known as Campbell-Stokes (CS) recorder. An extensive discussion on the CS recorder and its historical development could be found in the study by Sánchez-Lorenzo et al. (2013). SD measurements have been routinely performed over many years in many parts of the world, for example, for the last 160 years for some parts of Europe and since 1890 for Japan. Recently, new generation instruments having automatic sensors have been designed and loaded into some stations but currently they are located in a limited number of stations. Due to its importance, studies on the estimation of SD have been growing rapidly and up until now, many studies have already been reported in literature. For example, Rangarajan et al. (1984) calculated the SD values from a 10-year mean cloud cover data using an empirical relationship with an accuracy of about 4–7%. Mean monthly SD for different latitudes was estimated with a RMSE% error in between 7% and 18% by Tejeda and Vargas (1996). Essa and Etman (2004) computed SD using cloud cover data for stations in Cairo, Bahtim and Sedi-Barrani in Egypt with the standard error of estimate (SEE) changing from 0.198 h to 0.844 h. El-Metwally (2005) proposed a nonlinear model which was based on cloud cover fraction and maximum and minimum temperatures to predict the relative SD for Egypt and it was shown that MBE% and RMSE% values changed from -0.2% to -13.3% and from 2.3 to 14.5%, respectively. Matzarakis and Katsoulis (2006) tried estimating the spatial and temporal distribution of bright sunshine hours over Greece using the percentage of land cover around each station (radius of 20 km), distance of each station from the nearest coast, height above sea level for each station location, latitude of each station and longitude of each station. The correlation coefficient (R) and RMSE were calculated as 0.87 and 9.90 h, 0.58 and 6.15 h, 0.89, and 4.69 h, 0.86 and 6.22 h, and 0.84 and 5.33 h, for winter, spring, summer, and autumn, respectively, for annual sunshine. Robaa (2008) derived three empirical formulae to estimate the relative SD using the cloud data for Egypt. It was shown that relative percentage error, mean percentage error, MBE and RMSE changed from -7.2698% to $+3.7908\%$, -0.6240% to $+0.8069\%$, -0.0053 to $+0.0070$ and 0.0046 to 0.0160 , respectively.

However, some authors have used satellite data and/or its products for estimating SD that provide us with almost continuous spatial coverage of the clouds over large areas. For instance, Kandirmaz (2006) used a statistical relationship between the daily mean cloud cover index and relative SD to derive daily global SD using METEOSAT First Generation (MFG) data. Shamim et al. (2012) improved Kandirmaz's model by including snow cover information, sun and satellite angles, and a trend connection factor for

seasons in the computation of the cloud cover index and obtained better results. Good (2010) also proposed a simple method that uses 15-minute time series of cloud type data from METEOSAT Second Generation (MSG) to compute daily SD for the United Kingdom. A time series of daily SD maps was compiled for Belgium and Luxembourg by combining in situ SD measurements with high resolution ancillary data derived from METEOSAT First Generation (MFG) satellite images (Journée et al. 2013). Kandirmaz and Kaba (2014) used MODIS (Moderate Resolution Imaging Spectroradiometer) images and derived a new quadratic correlation between the cloud cover index and relative SD to estimate the SD for nine stations in Türkiye. In the study by Kothe et al. (2013), they applied the method used by Good (2010) to predict SD over Europe. Bartoszek et al. (2021) proposed a correlation between the average areal totals of SD and changes in the amount of cloud cover, circulation types, and atmospheric optical depth (satellite data obtained from MODIS) for evaluating temporal and spatial trends in SD in Poland.

Recently, many machine learning algorithms have been frequently employed for forecasting purposes in different studies and they yielded highly accurate results compared to conventional models (Haykin 1994; Werbos 1988). Although numerous machine learning studies for solar radiation estimation is available in literature, unfortunately there is only a few about SD estimation. The first study that implemented a machine learning approach for estimating SD was conducted by Mohandes and Rehman (2013). They used Particle Swarm Optimization (PSO) and Support Vector Machine (SVM) algorithms to predict the SD for Saudi Arabia using meteorological parameters; maximum possible SD, extra-terrestrial solar radiation, latitude, longitude, altitude, and month number. Rahimikhoob (2014) examined the potential for the use of artificial neural networks (ANNs) to assess SD using air temperature and humidity data for Sistan and Baluchestan provinces in Iran. Kandirmaz et al. (2014) introduced an ANN approach for estimating monthly mean daily values of global SD for Türkiye using a climatic variable (cloud cover) and two geographical variables (day-length and month). Kaba et al. (2017) used linear, polynomial, and radial basis function (RBF) kernels of SVM models and meteorological parameters to predict daily SD for 14 stations in Türkiye and it is concluded that SVM with the RBF model is suitable for predicting the daily SD.

In this study using EUMETSAT METEOSAT based CM SAF CFC and CTY products, elevation data of the GMTED2010 (Global Multi-resolution Terrain Elevation Data) digital elevation model, month number and day-length as inputs, a multilayer perceptron (MLP) type artificial neural network model was proposed for estimating the monthly mean SD. The constructed model was trained using

a dataset of 45 stations and tested with ground-truth data of 12 stations in Türkiye. A monthly mean SD map for each month was constructed for 2014 and its distribution across the country was also discussed.

2 Material and method

2.1 Study area and datasets

This study was conducted over Türkiye, which is geographically located at the southwestern extremity of Asia and at the south-eastern extremity of Europe (36° and 42° N and 26° and 45° E) and has an area of 783,562 km². Because of its irregular topography, climate of Türkiye is diverse and it includes mainly seven sub regions, namely, Marmara, Black Sea, Mediterranean, Aegean, Eastern Anatolian, South-eastern Anatolian, and Central Anatolian. The average SD was determined as 7.2 h/day and 6.94 h/day between years 1966 and 1982 and 1988 and 2017, respectively (<http://www.eie.gov.tr/>; <https://www.mgm.gov.tr/>). Total annual average precipitation was approximately calculated as 527.61 mm for the years 1979–2019 (Bulut and Sakalli 2021). The annual mean temperature varies from 3.6 °C to 20.1 °C from region to region across the country (Deniz et al. 2011). The climate of the Black Sea area is wet and warm in summer and cold and rainy in winter and the average values of air temperature are about 23 °C for summer and 7 °C for winter. The Black Sea coast receives the greatest amount of rainfall, about 2200 mm annually. The Mediterranean and a substantial part of the Aegean coasts have mild and rainy winters and hot and moderately dry summers. The Marmara Region is surrounded on two sides by the Black Sea and Aegean Sea and shows characteristics of the Mediterranean and Black Sea and Continental climate and it has an average temperature of 4 °C in winter and 27 °C in summer. On the other side the Central Anatolia has a semi-arid climate with cold, snowy winters and hot, dry summers and the South-Eastern Anatolian region generally has generally mild spring and autumn and hot and dry summers. Eastern Anatolian, which is the largest region occupying 21% of the total area of the country, climate is similar to the desert climate and this region shows the hot-dry climate zones with a great temperature difference between day and night (Yılmaz 2007).

There is a high correlation between SD and some meteorological variables such as cloud cover, temperature, precipitation, relative humidity, wind speed, and astronomical variables, or a combination of meteorological and relevant astronomical variables (Kaba et al. 2017; Kaiser and Qian 2002; Rahimikhoob 2014). In fact, relevant astronomical variables can be directly calculated using the mathematical relationships and temperature, precipitation, relative

humidity and wind speed can be accurately measured at almost all meteorological stations. But it should be mentioned that the most important parameter affecting SD is the cloud cover (Matuszko 2012). Unfortunately, finding reliable cloud data for any location could be a problem because cloud cover and cloud types over the sky are classically determined by a trained meteorologist and obviously it is a subjective work. In addition, SD values should be predicted for regions where no direct measurement is possible or measurements are unreliable or missing. This is generally done using interpolation techniques that consider the values of the nearest stations. However, the density of stations for some regions may not be sufficient and uniform and furthermore the region under consideration may have very different climatic conditions than those of the nearest stations. In such cases, large errors may occur during the estimation process. To remove such deficiencies, researchers need new models that can estimate SD with high accuracy with minimum number of measured ground data. Taking into account this reality, we tried estimating monthly mean SD using only satellite data (CM SAF cloud cover and cloud type data), digital elevation data (GMTED2010 data), daylength and month number.

The satellite - based cloud parameters of CM SAF used in this study are derived from the EUMETSAT NWC SAF (Support to Nowcasting and Very Short-Range Forecasting SAF) project. The general aim of the NWC SAF project is to provide algorithms and software which can be used to generate operational products to ensure the optimum use of meteorological satellite data in nowcasting and very short-range forecasting by targeted users. The cloud type (CT) product provides detailed cloud analyses in the METEOSAT SEVIRI (Spinning Enhanced Visible and InfraRed Imager) instrument with a pixel resolution of 3 km by 3 km. The CT product (see first and second columns in Table 1) contains information on the major cloud classes as follows: fractional clouds, semi-transparent clouds, high, medium and low clouds (including fog) for all pixels identified as cloudy in a scene. CTY product of CM SAF (see first and third columns of Table 1) gives information about the cloud type that are originally produced from NWC SAF CT data. The sum of the percentages of the five cloud classes is equal to 100. To find the absolute ratios of these cloud averages in the CTY pixel, these values are multiplied by the value of the same pixel in the CM SAF CFC (cloud fractional cover) product, which is the product of the total cloud coverage ratio. The CTY products have been produced by CM SAF since September 2005 and are provided free of charge. CTY data could not be produced for 10 months between March 2012 and December 2012 and therefore, this period was omitted in the study. In each CFC and CTY file, there are 24 data sets representing each hour of the relevant month (such

Table 1 NWC SAF GEO CT and CM SAF CTY products

Class	CT Definition	CTY Definition
0	Non-processed	Not used
1	No contamination by snow/ice covered surface, no contamination by clouds; but contamination by thin dust/volcanic clouds not checked	No cloud
2	No contamination by snow/ice covered surface, no contamination by clouds; but contamination by thin dust/volcanic clouds not checked	
3	Land contaminated by snow	
4	Sea contaminated by snow/ice	
5	Very low clouds	Very low clouds
6	Very low and stratiform clouds	
7	Low and cumuliform clouds	
8	Low and stratiform clouds	
9	Mid-level clouds	Mid-level clouds
10	Mid-level and stratiform clouds	
11	High opaque and cumuliform clouds	High clouds
12	Very high clouds	
13	Very high opaque and cumuliform clouds	
14	Very high opaque and stratiform clouds	
15	High semi-transparent thin clouds	High semi-transparent clouds
16	High semi-transparent meanly thick clouds	
17	High semi-transparent thick clouds	
18	High semi-transparent above low or medium clouds	
19	Fractional clouds	Fractional clouds
20	Undefined cloud	Not used

as 00:45, 01:45, 02:45 GMT). In the study, data sets corresponding to 09:45, 11:45, and 13:45 local times were used to estimate the targeted sunshine duration. Morning and evening hours, which are close to sunrise and sunset times, are not particularly used. This is due to the performance limitations associated with these products. In the related document, it is stated that these products may have incorrect classifications in cloud type information obtained due to the inability to calculate the reflection values correctly due to the high solar zenith angle at sunrise and sunset times. The amount of cloud given in these 3 h for five different cloud classes was collected separately for each class and divided by three to obtain averages. Note that detailed information about the NWC SAF CT and CM SAF CFC and CTY can be found on the website <https://www.nwcsaf.org/> and https://www.cmsaf.eu/EN/Home/home_node.html.

An input to our model is the altitude of the area under consideration. Lu et al. (2011) showed that the surface altitude is an important factor for estimating the solar radiation in large areas with varied terrain. Since there is a strong relationship between SD and solar radiation, the same logic should be valid and applied for estimating SD. Altitude data were obtained from the GMTED2010 digital elevation model data. The U.S. Geological Survey (USGS) developed

a global topographic elevation model, GTOPO30, with a horizontal resolution of 30 arc-seconds for the entire Earth and is used for various purposes such as climatological, hydrological and geomorphological military applications. It is then improved to a new product that is GMTED2010, providing a new level of global topographic data. GMTED2010 has three separate resolutions; (horizontal post spacings) of 30 arc-seconds (about 1 km), 15 arc-seconds (about 500 m), and 7.5 arc seconds (about 250 m). It also provides global coverage of all land areas from latitudes 84° N to 56° S for most products, and coverage from 84° N to 90° S for several products. It has an advantage over GTOPO30 because it has new raster elevation products, which are available at each resolution. The new elevation products have been produced as follows: mean elevation, maximum elevation, minimum elevation, median elevation, standard deviation of elevation, systematic subsample, and breakline emphasis (<https://pubs.usgs.gov/of/2011/1073/pdf/of2011-1073.pdf>).

Other input data, the daylength, which must find the maximum possible sunshine hours in a day and can be calculated for any location by using the following relation given in Eqs. (1–3) (Duffie and Beckman 2013) where ω_s and δ are the solar hour and zenith angle, ϕ is latitude of the location in the range between -90 and 90 degrees, and J is the number of days of the year starting from the first of January.

$$SD = 24 * \frac{\omega_s}{\pi} \quad (1)$$

$$\omega_s = \arccos(-\tan \phi \tan \delta) \quad (2)$$

$$\delta = 0.409 * \sin\left(\frac{2\pi J}{365} - 1.39\right) \quad (3)$$

The nine-year SD data set belonging to 57 stations was collected by Turkish State Meteorological Service (TSMS), which is in charge for calibration and maintenance of the devices installed in the official stations. These stations are geographically distributed over almost the entire country and thus one can assume that they reflect all the different climatic characteristics of Türkiye. Geographical distributions of the stations over the country and climatic zones are given in Fig. 1. The WMO code, name, latitude, longitude, altitude and mean values of the SD of the selected stations belonging to the study period are given in Table 2. As shown in Table 2, measured SD values are gradually change between the 4.28 h and 8.22 h, which really implies that the country has climatologically and geographically different regions.

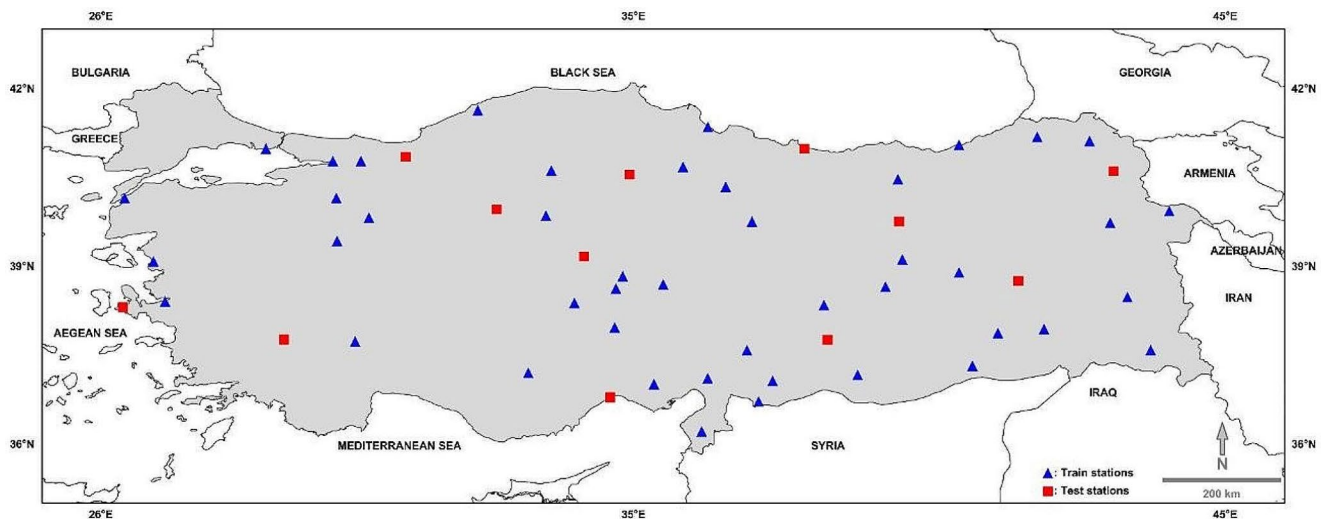


Fig. 1 The geographic distribution of 57 stations over Türkiye. Blue triangle shows the stations used for training and red square shows the stations used for testing in the model

2.2 Artificial neural networks

Artificial Neural Networks are the mathematical modeling approaches for human neurological systems to obtain the advantages of human thinking mechanism into the computation environment (McCulloch and Pitts 1943). The network topology of an artificial neural network is a limited connection and interaction model of artificial neurons or artificial computing unit elements. Artificial neural networks may be cyclic or acyclic. Generally, they have layered connection approaches to realize human neural system like actions. The widely used artificial neural networks are the feedforward neural networks of the layered structures. In a feed forward neural network, layers are sequential stages with one or more neurons operating simultaneously. A parallel computation task is done by the neurons of a layer. After the completion of the computation by a layer neuron, the computational outputs are applied to the next layer as input. This operation occurs sequentially from the input layer to the output layer in a feed forward artificial neural network. The layers between the input and output layers are called hidden layers. The hidden layers may have one or more sequential layers. Since the input layer elements are used only for holding the inputs without any operation, they are not considered neurons. The input layer is not considered an operational layer. The computational tasks are done by hidden and output layers. Learning of a neural network means the change or optimization of the computational parameters of the neurons. These parameters are called weights and biases. Weights and biases represent the synaptic connection strength and threshold values of a biological neuron, respectively. Linear or nonlinear activation functions are used to obtain each neuron output after multiplication and

summation operations with weight and bias values. Learning tasks can be of supervised, unsupervised and reinforcement learning types. In supervised learning, in addition to input data, the corresponding target output data is also given to train the neural network. Both regression and classification tasks can be done by a supervised learning neural network. The unsupervised learning neural networks do not consider the target values into consideration. Instead, they group the input data into clusters. The only task done by the unsupervised neural networks is the clustering. The reinforcement learning contains reward/punishment mechanism additionally. The direction of computation is sensed according to the rewards. The supervised neural networks may have one-pass or repetitive learning approaches for both regression and classification tasks. In the repetitive supervised learning, in addition to forward calculation from input to output, backward calculation from output to input also occurs for output error minimization. This approach is called error back propagation. The weight and bias value corrections are done through the error back propagation phase. In this work, one of the most well-known supervised learning artificial neural networks called Multi-Layer Perceptron (MLP) is used for the estimation task. It is trained by the error back propagation method. The implementation is performed in the Python environment with the PyBrain library toolbox (Schaul et al. 2010).

An MLPNN (MLP Neural Networks) topology includes an input layer, one or more hidden layers, and an output layer. There is no specific method to specify the optimum neuron and hidden layer number of the network. Instead, the number of hidden layers and neurons is determined by trial and error.

Table 2 The WMO code, name, latitude, longitude, altitude and mean SD of the stations (shaded areas or last 12 rows indicate the testing stations)

WMO code	Name	Sub-region	Latitude (Degree)	Longitude (Degree)	Altitude (m)	Mean sunshine duration over study period (h)
17,020	Bartın	Black Sea	41.624	32.366	33	5.8228
17,030	Samsun	Black Sea	41.343	36.255	4	5.6245
17,040	Rize	Black Sea	41.040	40.499	3	4.2821
17,045	Artvin	Black Sea	41.175	41.819	625	4.8873
17,085	Amasya	Black Sea	40.667	35.835	409	5.8089
17,086	Tokat	Black Sea	40.331	36.557	611	6.3198
17,088	Gümüşhane	Black Sea	40.460	39.465	1216	5.3203
17,046	Ardahan	East Anatolian	41.106	42.706	1827	5.7939
17,099	Agrı	East Anatolian	39.725	43.052	1646	6.0820
17,100	İğdır	East Anatolian	39.922	44.052	856	6.1902
17,165	Tunceli	East Anatolian	39.106	39.541	981	7.0290
17,172	Van	East Anatolian	38.469	43.346	1675	8.0282
17,199	Malatya	East Anatolian	38.337	38.217	950	7.6442
17,201	Elazığ	East Anatolian	38.644	39.256	989	5.7910
17,203	Bingöl	East Anatolian	38.885	40.500	1177	6.1380
17,080	Çankiri	Central Anatolian	40.608	33.610	751	6.2983
17,285	Hakkari	East Anatolian	37.575	43.739	1727	7.5956
17,090	Sivas	Central Anatolian	39.743	37.002	1294	6.8971
17,123	Eskişehir	Central Anatolian	39.812	30.528	787	6.6836
17,135	Kırıkkale	Central Anatolian	39.843	33.518	751	7.0613
17,140	Yozgat	Central Anatolian	38.820	34.816	1301	7.2129
17,192	Aksaray	Central Anatolian	38.370	33.999	970	7.3949
17,193	Nevşehir	Central Anatolian	38.616	34.702	1260	7.1521
17,196	Kayseri	Central Anatolian	38.687	35.500	1094	6.7291
17,246	Karaman	Central Anatolian	37.193	33.220	1018	8.0388
17,250	Niğde	Central Anatolian	37.959	34.679	1195	7.4548
17,155	Kütahya	Aegean	39.417	29.989	969	5.9649
17,220	İzmir	Aegean	38.395	27.082	29	8.0132
17,066	Kocaeli	Marmara	40.766	29.917	74	5.6423
17,069	Sakarya	Marmara	40.767	30.393	30	5.4321
17,112	Çanakkale	Marmara	40.141	26.399	6	6.6528
17,120	Bilecik	Marmara	40.141	29.977	539	6.6527
17,636	Florya	Marmara	40.976	28.786	37	6.8247
17,180	Dikili	Mediterranean	39.073	26.888	3	8.0850
17,238	Burdur	Mediterranean	37.722	30.294	957	7.3186
17,255	Kahramanmaraş	Mediterranean	37.576	36.915	572	6.4989
17,351	Adana	Mediterranean	37.004	35.344	23	7.3341
17,355	Osmaniye	Mediterranean	37.102	36.253	94	7.7574
17,372	Hatay	Mediterranean	36.205	36.151	104	7.1919
17,210	Siirt	South-eastern Anatolian	37.932	41.935	895	7.1030
17,261	Gaziantep	South-eastern Anatolian	37.058	37.351	854	6.5252
17,262	Kilis	South-eastern Anatolian	36.708	37.112	640	7.7401
17,270	Şanlıurfa	South-eastern Anatolian	37.160	38.786	550	7.3887
17,275	Mardin	South-eastern Anatolian	37.310	40.728	1040	8.2177
17,282	Batman	South-eastern Anatolian	37.863	41.156	610	7.5111
17,084	Çorum	Black Sea	40.546	34.936	776	5.8135
17,072	Düzce	Black Sea	40.843	31.149	146	5.1778
17,033	Ordu	Black Sea	40.984	37.886	5	4.3789
17,160	Kırşehir	Central Anatolian	39.164	34.159	1007	7.3122
17,129	Etimesgut	Central Anatolian	39.956	32.685	806	6.8711
17,221	Çesme	Aegean	38.304	26.372	5	7.7101
17,237	Denizli	Aegean	37.762	29.092	425	6.9044
17,097	Kars	East Anatolian	40.604	43.107	1777	6.4145
17,204	Muş	East Anatolian	38.751	41.502	1322	6.3944

Table 2 (continued)

WMO code	Name	Sub-region	Latitude (Degree)	Longitude (Degree)	Altitude (m)	Mean sunshine duration over study period (h)
17,094	Erzincan	East Anatolian	39.752	39.486	1216	6.3533
17,340	Mersin	Mediterranean	36.781	34.603	7	7.5636
17,265	Adiyaman	South-eastern Anatolian	37.755	38.277	672	6.8146

MLPNN with error backpropagation supervised learning provides the approximation of input-output mappings of multivariate, non-linear functions. After the training of MLPNN, mapping among inputs and outputs is obtained. In the backpropagation learning phase, the learning originates from the output neurons by considering the error values. The difference between the desired output and estimated output is called an error. It is continuously calculated for each iteration and backpropagated through the network either batch or stochastically. The backpropagation process modifies the weight and bias parameters with a particular learning rate (α) and/or momentum (β) term.

Fig. 2 shows the general structure of the MLPNN with one hidden layer. As the figure depicted, inputs are processed through the hidden layers, and the output is formed by a nonlinear function.

The input layer involves linear combinations of dimensional inputs:

$$net_j^1 = \sum_{i=0}^d w_{ji}^{(1)} x_i + b \tag{4}$$

where $j = 1, 2, 3, \dots, M$. The quantities net_j^1 are called weighted sums, the parameters $w_{ji}^{(1)}$ are the weights, and b is the bias. The superscript 1/ indicates that this is the first layer of the network. $w_{ji}^{(1)}$ expresses the weight vector from i^{th} element in the input layer to j^{th} element in the hidden layer. Each net calculation is applied to a linear or nonlinear activation function $h()$, typically a logarithmic sigmoid, given in (5).

$$u_j = h\left(net_j^1\right) = \frac{1}{1 + e^{-net_j^1}} \tag{5}$$

where u_j are the outputs of hidden layers. The most popular activation functions are pure linear, tangent hyperbolic sigmoid, logarithmic sigmoid and Gaussian functions. In the second layer, the outputs of the hidden layer neurons are linearly combined to obtain the inputs of the K output units by using (6).

$$net_k^{(2)} = \sum_{j=0}^M w_{kj}^{(2)} u_j + b \tag{6}$$

where $k = 1, 2, 3, \dots, K$. This operation is done by the second layer of the neural network parameterized by weights $w_{kj}^{(2)}$. The output neurons were transformed using an activation function. Typically, logarithmic sigmoid function can be used as given in (7).

$$y_k = g\left(net_k^{(2)}\right) = \frac{1}{1 + e^{-net_k^{(2)}}} \tag{7}$$

These equations could be combined to give the overall equation that describes the forward propagation through the network, as in (8).

$$y_k = g\left(\sum_{j=0}^M w_{kj}^{(2)} h\left(\sum_{i=0}^d w_{ji}^{(1)} x_i\right)\right) \tag{8}$$

In the standard backpropagation learning algorithm weight corrections can be defined by using (9).

$$\Delta w_{kj}^{\tau+1} = -\eta \frac{\partial E}{\partial w_{kj}} + \mu \Delta w_{kj}^{\tau} \tag{9}$$

where $\Delta w_{kj}^{\tau+1}$ and Δw_{kj}^{τ} are defined as the updated weight and instant weight of τ^{th} iteration, respectively, while η is the learning rate, μ is the momentum factor, and E express the square of the error in stochastic mode or mean square error in batch mode. The backpropagation procedure is repeated until the mean square error (MSE) of the system converges to a target error value or computational bounds are reached. To realize the backpropagation algorithm, the estimated y must be compared with the desired output t . The error is obtained as in (10).

$$e = t - y \tag{10}$$

Where e is the instant error that is defined as the difference between the target and estimated output. In the standard gradient descent-based backpropagation learning algorithm weight corrections can be defined by using (9).

The error function can be obtained by summing over a training set of N examples as given in (11) and (12).

$$E = \sum_{n=1}^N E^n \tag{11}$$

$$E^n = \frac{1}{2} \sum_{k=1}^K (t_k^n - y_k^n)^2 \tag{12}$$

where the y_k^n parameter could be determined using (14). The derivative of E with respect to the hidden layer to output layer weights w_{1j} can be written using the chain rule of differentiation, as given in (13).

$$\frac{\partial E}{\partial w_{1j}} = \frac{\partial E}{\partial net_1^{(2)}} \frac{\partial net_1^{(2)}}{\partial w_{1j}} \tag{13}$$

Finally, the error signal is found, as defined in (14).

$$\delta_k^{(2)} = \frac{\partial E}{\partial y_k^n} \frac{\partial y_k^n}{\partial net_1^{(2)}} = (t_k^n - y_k^n) f'(net_1^{(2)}) \tag{14}$$

Where f' is first derivative of the activation function. Similarly, the derivative of E with respect to input layer to hidden layer weights w_{ji} must be calculated using (15).

$$\delta_j^n = \frac{\partial E}{\partial net_j^{(1)}} = \sum_{k=1}^{10} \frac{\partial E}{\partial net_k^{(2)}} \frac{\partial net_k^{(2)}}{\partial net_j^{(1)}} = \sum_{k=1}^{10} \delta_k^{(2)} \frac{\partial net_k^{(2)}}{\partial net_j^{(1)}} \tag{15}$$

Since $net_k^{(2)}$ depends on u_j as indicated in (15), by using the chain rule of differentiation, as given in (16),

$$\frac{\partial net_k^{(2)}}{\partial net_j^{(1)}} = \frac{\partial net_k^{(2)}}{\partial u_j} \frac{\partial u_j}{\partial net_j^{(1)}} = w_{1j} f'(net_j^{(1)}) \tag{16}$$

Thus, substituting (16) into (15), the backpropagation of error equation can be obtained as given in (17).

$$\delta_j^n = f'(net_j^{(1)}) \sum_{k=1}^{10} \delta_k^{(2)} w_{1j} \tag{17}$$

The derivatives of the input-to-hidden weights can be calculated using (18).

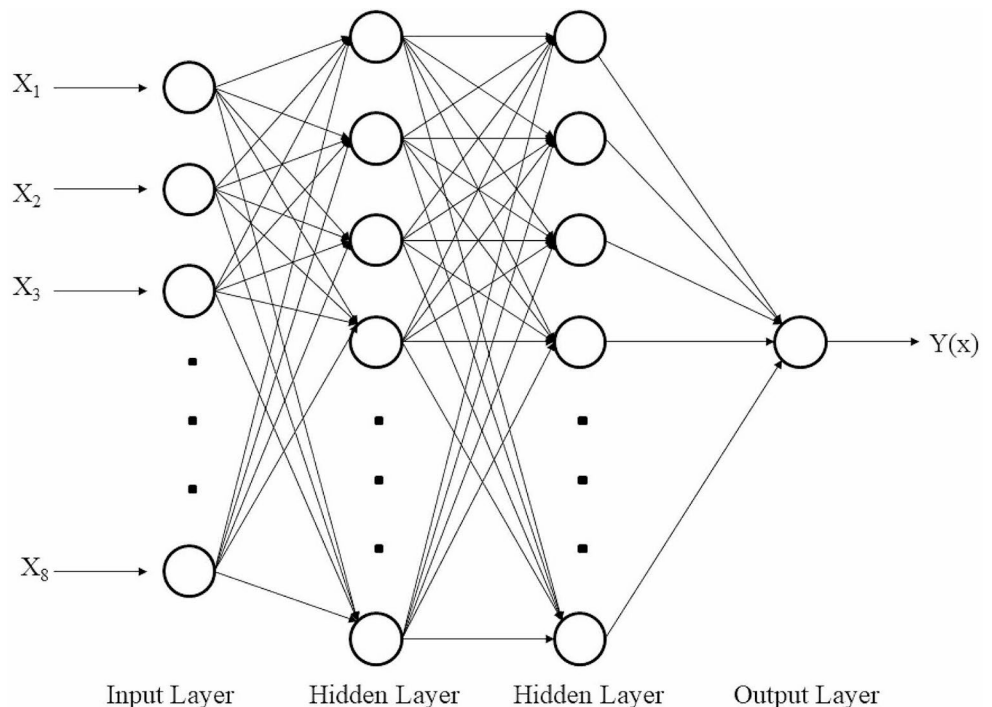
$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial net_j^{(1)}} \frac{\partial net_j^{(1)}}{\partial w_{ji}} = \delta_j x_i \tag{18}$$

This approach can be applied recursively to further hidden layers. The weight vectors and bias parameters of the neurons can be updated until all target errors or the maximum number of iterations are achieved.

As mentioned earlier, the error signal propagates back to the previous layer. Then, the weight and bias parameters associated with each iteration are updated. The backpropagation procedure is repeated until the mean square error (MSE) of the system approaches the target error, and mapping is done between inputs and outputs.

In this work the MLPNN with 8-9-9-1 topology as shown in Fig. 2 is developed and trained. The MLPNN has eight inputs, two hidden layers with nine neurons in each and an output. The hidden layers execute logarithmic sigmoid activation function, where the output has pure linear one. The inputs of the MLPNN are.

Fig. 2 The ANN structure used in this study



- X_1 : Altitude of the Station.
- X_2 : Month number.
- X_3 : Daylength.
- X_4 : Ratio of fractional clouds.
- X_5 : Ratio of high semi-transparent cloud.
- X_6 : Ratio of high opaque clouds.
- X_7 : Ratio of mid-level clouds.
- X_8 : Ratio of low-level clouds.

This dataset is separated into three groups for training, validation and test. Training, validation and test data have 70%, 5%, and 20% of the overall data. The validation data set to detect overfitting is decided as 5%. The performance of the trained MLPNN is determined according to the test data. Performance criteria and the estimation accuracy are measured according to the statistical indices given in the next part of this work.

2.3 Model evaluation

In this study, we used four types of statistical indicators for checking the model accuracy namely mean bias error (MBE), mean absolute error (MAE), root mean square error (RMSE), mean bias error and coefficient of determination (R^2). The mathematical expressions for these statistical indices are given as follows;

$$MBE = \frac{\sum_{i=1}^n (E_i - O_i)}{n} \tag{19}$$

$$MAE = \frac{\sum_{i=1}^n |E_i - O_i|}{n} \tag{20}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (E_i - O_i)^2}{n}} \tag{21}$$

$$R^2 = \frac{[\sum_{i=1}^n (E_i - \bar{E})(O_i - \bar{O})]^2}{\sum_{i=1}^n (E_i - \bar{E})^2 \sum_{i=1}^n (O_i - \bar{O})^2} \tag{22}$$

Here n is equal to the total number of sample data, E_i is the value obtained from the model, O_i is the measured value, \bar{E} is the average of the model results while \bar{O} is the average of the measured values. R^2 is used to give information about the relationship between the dependent and independent variables in a regression analysis and it changes between 0 and 1. The MBE, MAE and RMSE values indicate the measure of differences between measured and estimated values and thus ideal values of these three indicators are equal to 0.

3 Results and discussion

SD over any surface on the earth is strongly related to clouds as well as to solar radiation (Fox 1961; Kim and Ramanathan 2008). Cloud cover generally prevent incoming solar radiation and sometimes they cause an increase, due to reflection/backscattering and multi-scattering of short-wave radiation, solar radiation at the surface. It is much easier to estimate the SD and GSR for a clear sky than those for an overcast and a partly overcast sky. That is information about the total fraction and type of cloud over any area is crucial for determining and estimating the SD. In this study, we used CM SAF cloud products for estimating SD over Türkiye. The MLP, which is one of the most popular and practical architectures of ANN, was employed order to simulate SD values by using CM SAF CFC and CTY products, GMTED2010 digital elevation model, month number and daylength as input. The building of the MLP model was done by means of PyBrain Library (Schaul et al. 2010). PhyBrain is a modular machine learning library for Python and it presents algorithms for supervised and unsupervised learning (For more information about PhyBrain see the web site www.pybrain.org). The model used here had four layers, namely, input layer, two hidden layers and the output layer. Previous studies have shown that such topology could be suitable for solving similar real-world problems (Piotrowski et al. 2015; Quej et al. 2017). The multi-layer feed-forward network is the type of network used in this study. Many combinations of parameters were examined, and the optimum results were obtained using eight inputs with nine neurons in hidden layers and a single neuron in the output layer. For developing the model, we used 70% of the data for training and the rest of 30% was used for testing (25%) and (5%) validation purposes. The sigmoid activation function and linear activation function were used for hidden layers and the output layer, respectively. All input data were normalized to a range 0 and 1 and to get the MLP output directly as SD. In the input layer has a total of eight parameters, namely, the elevation of the station, the month number, mean daylength, ratio of fractional cloud, the ratio of high semi-transparent cloud, the ratio of high opaque cloud, ratio of mid-level cloud and ratio of low-level clouds were used, and the network trained 25 times until the error between the observed and the predicted value reached a significantly low level.

The performance of our model was tested using the four statistical indices namely, MBE, MAE, RMSE and R^2 . A scatter diagram that shows the estimated values versus observed values of SD was plotted for each test station and all diagrams are illustrated in Fig. 3. Computed values of the statistical indices for all test stations are summarized in Table 3. The scatter plot of predicted SD values against to

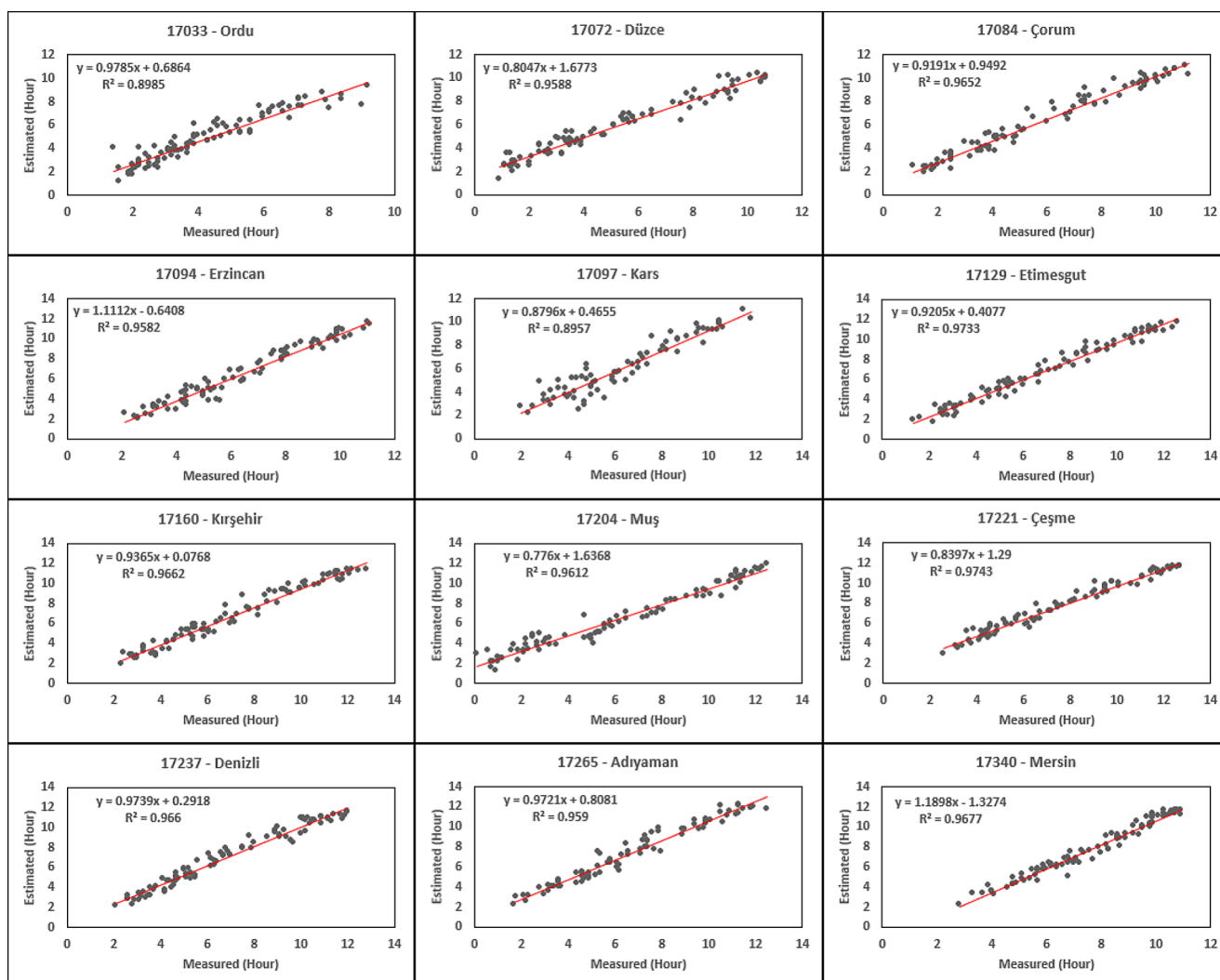


Fig. 3 Plots of estimated versus observed values of SD for all test stations

Table 3 MBE, MAE, RMSE and R^2 values of each test station (** indicates the best result for the considered column while * indicates the worst result)

WMO Code	Name	Sub region	Estimated SD (h)	Measured SD (h)	MBE (h)	MAE (h)	RMSE (h)	R^2
17,084	Çorum	Black Sea	6.2925	5.8135	0.4790	0.5923	0.7446	0.9652
17,072	Düzce	Black Sea	5.8441	5.1778	0.6663*	0.8893*	1.0247*	0.9588
17,033	Ordu	Black Sea	4.9712	4.3789	0.5923	0.6979	0.8737	0.8985
17,160	Kirsehir	Central Anatolian	6.9244	7.3122	-0.3878	0.5578	0.8262	0.9662
17,129	Etimesgut	Central Anatolian	6.7328	6.8711	-0.1383	0.4544	0.5671	0.9733
17,097	Kars	East Anatolian	6.1075	6.4145	-0.3070	0.7129	0.8826	0.8957*
17,204	Muş	East Anatolian	6.5989	6.3944	0.2045	0.8484	0.6820	0.9612
17,094	Erzincan	East Anatolian	6.4188	6.3533	0.0655	0.5320	0.6412	0.9582
17,221	Çesme	Egean	7.7641	7.7101	0.0540**	0.5334	0.6501	0.9743**
17,237	Denizli	Egean	7.0164	6.9044	0.1120	0.4495**	0.5471**	0.9660
17,340	Mersin	Mediterranean	7.6717	7.5636	0.1081	0.5166	0.6309	0.9677
17,265	Adıyaman	Southeastern Anatolian	7.4327	6.8146	0.6181	0.6790	0.8531	0.9590

observed values for all stations for the whole study period is given in Fig. 4.

As can be deduced from Table 3; Fig. 3 that our model yielded nearly similar SD values when compared against the measured values and thus very low MBE, MAE and RMSE and high R^2 values were calculated almost for all test stations. The MBE values varied between -0.3878 h (Kırşehir station) and 0.6663 h (Düzce Station). Underestimation was dominant at Kırşehir, Kars, and Etimesgut stations and overestimation was introduced for other nine stations. The MAE and RMSE values were found to be less than 0.9000 h and 1.1000 h, respectively, and the R^2 values were greater than 0.8900 for all test stations. The overall results of RMSE, MAE, MBE and R^2 were calculated as 0.7803 h, 0.6206 h, 0.1751 h and 0.9387 , respectively (see Fig. 4).

The lowest MAE was found for Denizli station with a value of 0.4495 h. The highest MAE value was calculated for Düzce station as 0.8893 h and the second highest was found for Muş station as 0.8484 h. The MAE values of seven stations, namely Denizli, Etimesgut, Mersin, Erzincan, Çeşme, Kırşehir and Çorum, were found to be very close to each other and their values changed from 0.4495 h to 0.5923 h. Adıyaman, Ordu, and Kars stations produced higher MAE values than these seven stations having the values of 0.6790 h, 0.6979 h, and 0.7129 h, respectively. The lowest and the second lowest RMSE values were found for the Denizli and Etimesgut stations at 0.5471 h and 0.5671 h, respectively. The RMSE values belonging to the Mersin, Erzincan, Çeşme and Muş stations were less than 0.7000 h and these stations had yielded nearly the same RMSE values, 0.6309 h, 0.6412 h, 0.6501 h and 0.6820 h, respectively. Kırşehir, Adıyaman, Ordu, and Kars stations produced a bit higher RMSE values, varying between 0.8000 h and 0.9000 h, than Mersin, Erzincan, Çeşme and Muş stations.

The highest RMSE value was calculated for Düzce as 1.0247 h. However, although the highest values of R^2 were found at Cesme and Etimesgut stations as 0.9743 and 0.9733 , respectively, the R^2 values of the other stations were not far away from these results and varied between 0.9677 and 0.9582 , except for stations Ordu and Kars. The lowest and the second lowest values of R^2 were found for Kars and Ordu stations as 0.8957 and 0.8985 . It was obviously seen that considerably high values of R^2 were obtained for almost all stations. Scattering diagrams also indicated that the model provides good agreement with ground data even for low SD values in which one may expect high estimation errors. According to the results given in Table 3 the model yielded the best results for Çesme and Denizli stations and the worst results for the Düzce and Kars stations. The model generally worked better for stations where their measured SD values were higher than the others.

To show and analyze the spatial and temporal distribution of SD across the country, we reconstructed monthly mean SD maps for 2014 (Fig. 5) and calculated the average monthly mean SD values of Türkiye for the study period (Table 4). As it is evident from the maps included in Fig. 5; Table 4. Türkiye had lower SD values for December (first month of winter in Northern Hemisphere) having a value of 2.8287 h for 2014 year. It then gradually increased in the upcoming months, except March. The average monthly mean SD for January, February, March and April were found as 3.7231 h, 6.2655 h, 5.5981 h and 6.9024 h, respectively. Then, it is sharply increased for the months from April to May (7.7732 h) and from May to June (9.6643 h). The highest SD values were obtained for July (second month of the summer in Northern Hemisphere) having an average value of 10.6825 h and the second highest value was obtained for August and the average value for this month was calculated

Fig. 4 The scatter plot of the model predicted SD values against the observed values for all stations for the entire study period

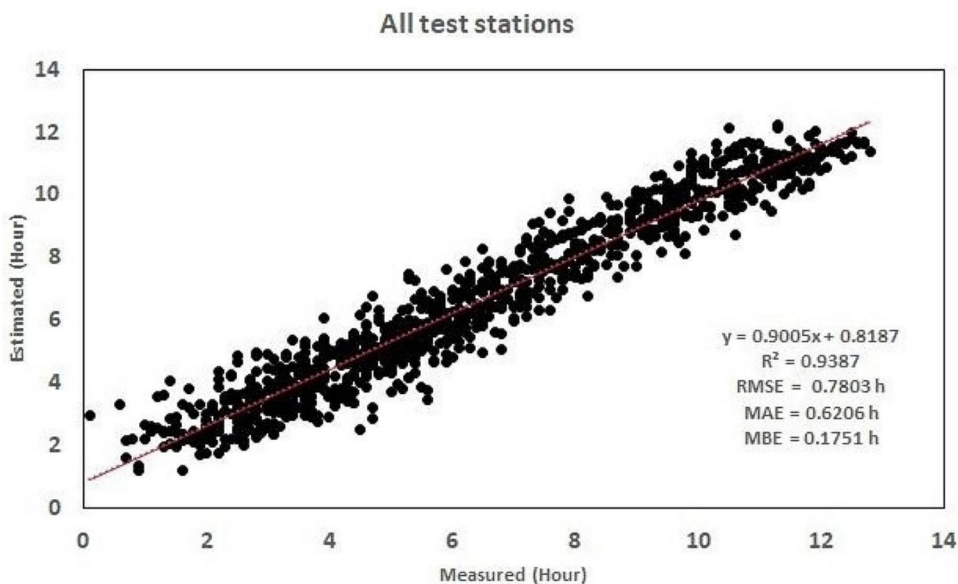


Fig. 5 Reconstructed/Estimated monthly mean SD maps of Türkiye corresponding to the 12 months of 2014 (from January to December)

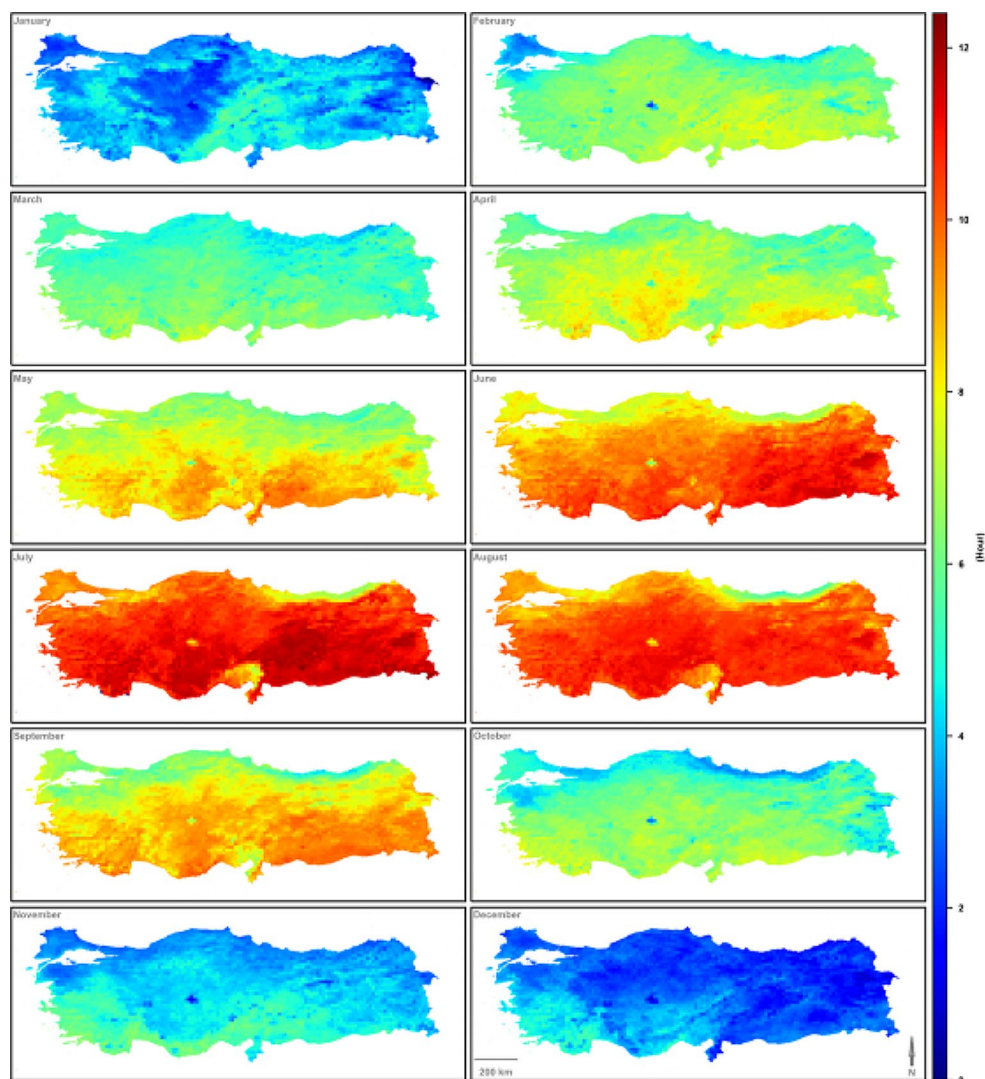


Table 4 Estimated monthly mean SD values of Türkiye (ND represents that no data is provided)

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Month										
January	ND	3.4447	4.4049	4.7187	3.4143	2.3534	3.6495	3.0611	2.7852	3.7231
February	ND	3.6885	4.6794	5.6037	2.6710	3.0582	4.9622	5.0450	3.9508	6.2655
March	ND	5.0938	5.6445	5.2117	4.5060	5.3099	6.0067	ND	5.0611	5.5981
April	ND	6.3445	6.9110	6.1252	7.0563	7.0565	5.6807	ND	7.3252	6.9024
May	ND	8.4403	8.5751	8.7878	8.8087	8.6539	7.7866	ND	8.4101	7.7732
June	ND	10.4345	10.4136	10.6667	9.9654	9.2385	9.9294	ND	10.2330	9.6643
July	ND	10.3579	11.0366	10.9435	10.1439	10.5148	10.8658	ND	10.2608	10.6825
August	ND	10.4231	10.3363	10.2944	10.1661	10.4976	10.3685	ND	10.2685	10.1627
September	8.6637	8.9252	9.7828	8.2213	7.9608	8.7202	9.2128	ND	8.9278	8.2740
October	6.4846	4.3102	6.5521	6.6652	6.6707	5.1026	6.4490	ND	7.6302	5.9173
November	3.3864	5.4531	4.3822	5.1650	4.7330	6.1078	5.1394	ND	5.0821	4.2600
December	3.1454	4.9863	3.8089	3.6697	2.6712	3.0415	3.9504	ND	4.1223	2.8287

as 10.4226 h. The values were then gradually decreased in the upcoming months.

The lower SD values were detected for the regions especially located inside the Black Sea Region. This was actually

expected because the Black Sea region is the cloudiest part of the country that receives the greatest rainfall amount. The lower SD values were also found for the Marmara region because a significant part, especially the northern part, of

this region is affected by the Black Sea climate. The SD values of the East Anatolian region were generally higher than those in the Black Sea and Marmara regions and nearly the same as those in the Central Anatolian region. The highest SD and second highest SD values were observed for South-eastern Anatolian and Mediterranean regions. These regions had relatively higher SD values than the Black Sea and Marmara regions and slightly higher than the East Anatolian region, Central Anatolian and Aegean regions in the order of appearance. It was deduced that, in contrast to cloudiness, SD values were generally increased from the north of the country, higher latitude, to the south of the country, lower latitudes (see Fig. 5; Table 2). These regional observations were nearly consistent with the data obtained from the meteorological stations of the General Directorate of Renewable Energy (YEGM, formerly EIE).

We also constructed a mean annual SD map spanning the years 2005–2014, and it was compared with a mean annual SD map that is the average of 20 years of meteorological data from 1988 to 2017. These two maps are given in Fig. 6a and b, respectively. As it can be seen from the figures, these two maps resemble each other closely and have almost similar SD distribution over the country although different models are used and the time coverage are not the same. However, a deeper look reveals that there are some differences between these two maps. The first difference between the maps seems to come from the fact that the SD values are distributed homogeneously around the meteorological stations so that map of TSMS has a much smoother surface than our satellite-derived map and hence, changes in values can be easily followed during transitions from one region to another, however, this is not the case in reality. For example, although the cloud cover ratio around İskenderun Gulf region was higher than the values measured by meteorological stations located around this region (Hatay, Adana, Osmaniye) we could not observe such reality in the map of TSMS. The same situation could also be observed at Eastern Black Sea Region, Marmara Region and some other parts of the country. These can be seen by comparing the long-term mean cloud cover map (1991–2015), given in Fig. 6c (Kaba and Yeşilyaprak 2021), which were constructed from the CM SAF CFC with the SD map of TSMS. Actually, such a problem is expected because SD map was constructed using an interpolation method in which values were computed by considering the values of the nearest stations. That is, the amount of cloud cover in any meteorological station could not represent the amount of cloud cover in the vicinity of the station because the formation of clouds can be influenced by factors such as humidity, landscapes, and wind. The second major difference is observed in the values of SD over the lakes. It was seen that generally incorrect SD values were produced by the satellite model because the spectral

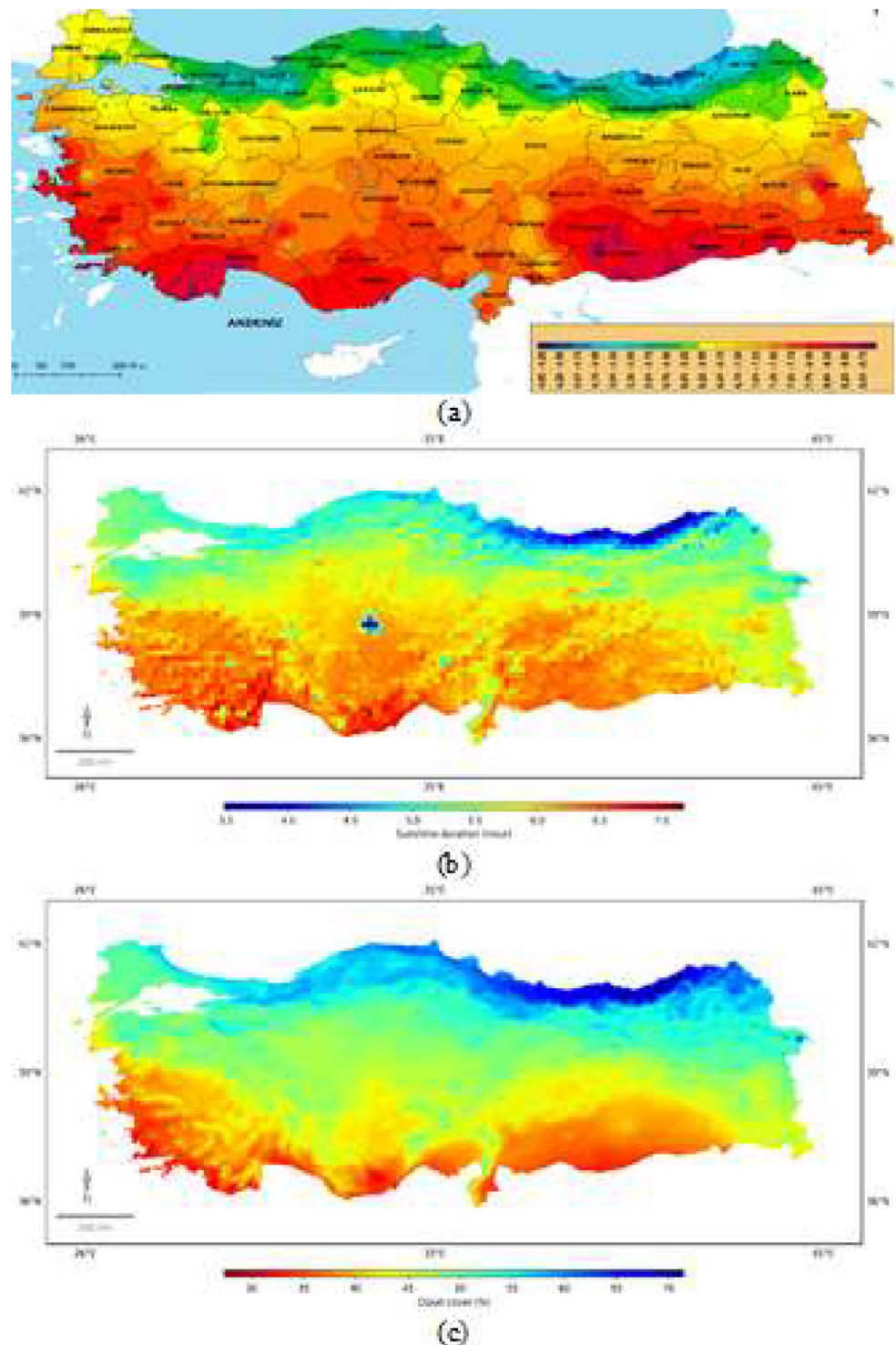
reflectance of the surface water is different from its environment. This type of error occurred for the Van Lake, which is the largest lake of the country and Tuz Lake, the second largest. Different SD values were calculated for these two lakes since the contents of the water and water depths are different. In this regard, one can claim that the values of SD TSMS are more accurate than the satellite estimated results for such surfaces. We also discussed our long-time monthly mean values belonging to years from 2005 to 2014 with the long-time monthly mean values from 1988 to 2017 of TSMS in Fig. 7. It is seen that there is a minimal difference between these two results and they are nearly the same. Finally, we analyzed annual mean values obtained from the model and TMSM through years from 2006 to 2012 and from 2013 to 2014 using Fig. 8. It is once again seen that the simulated results and observed values are very close to each other. Note that we neglected the years 2005 and 2012 because the satellite data were incomplete for those years (no data for 8 months in 2005 and 10 months for 2012) as mentioned before. The maximum difference is observed for 2007, where the model overestimated the TSMS with a value of 0.1606 h. The minimum difference observed for 2009 was 0.0360 h. The minimum yearly mean daily SD value for the TSMS was observed for 2009 with a value of 6.6000 h, while the maximum was found for 2013 a value of 7.1400 h. On the other side the minimum and maximum values of the model were found as 6.5640 h for 2009 and 7.2106 h for 2007, respectively. The maximum change in annual variation of SD between consecutive years for TSMS and the model is calculated as 6.3830% and 8.4863% between years 2008 and 2009, respectively.

The results of the model used in this study may be affected possibly negatively or positively from various sources of error. For example, error could be introduced while deriving satellite products and some measured values of SD could be unreliable. We assumed that the data were correct and no data were discarded or corrected in any way.

4 Conclusions

Since SD is a crucial data for many applications, its spatial and temporal distribution should be correctly estimated for regions where there is a lack of measurements or measurements are unreliable. A typical solution to this type problem could be found using only satellite-derived data or products in the estimation process instead of recorded values of SD or relevant measured meteorological data. Considering into account this fact, we tried deriving a satellite-based model for estimating the monthly mean SD for every region in Türkiye. To do this, satellite cloud products, digital elevation data and the astronomical data, daylength and month

Fig. 6 (a) Long-term (1988–2017) mean map constructed by the State Meteorological Service of Türkiye (<https://mgm.gov.tr/kurumici/turkiye-guneslenme-suresi.aspx>) (b) Estimated long-term mean SD map of Türkiye corresponding to the 2005–2014 (c) long-term mean cloud cover map (1991–2015) created from CM SAF CFC product (Kaba and Yeşilyaprak 2021)



number were used as inputs of the MLP type ANN model to get the SD as output. Data belonging to 45 stations were used for training the model and while 12 stations for testing and validation purposes. The simulated values of SD have indicated that our model yielded superb results because very high R^2 and very low MBE, MAE, and RMSE values computed at almost all test stations. The overall values

of R^2 , MBE, MAE, and RMSE were computed as 0.9387, 0.1751 h, 0.6206 h and 0.7803 h, respectively, for all stations. After the validation process, we also produced spatially continuous SD maps for each month of the study period (2005–2014). The following outcomes can be derived from the constructed SD maps;

Fig. 7 Comparison of long-term monthly mean values of SD for Türkiye

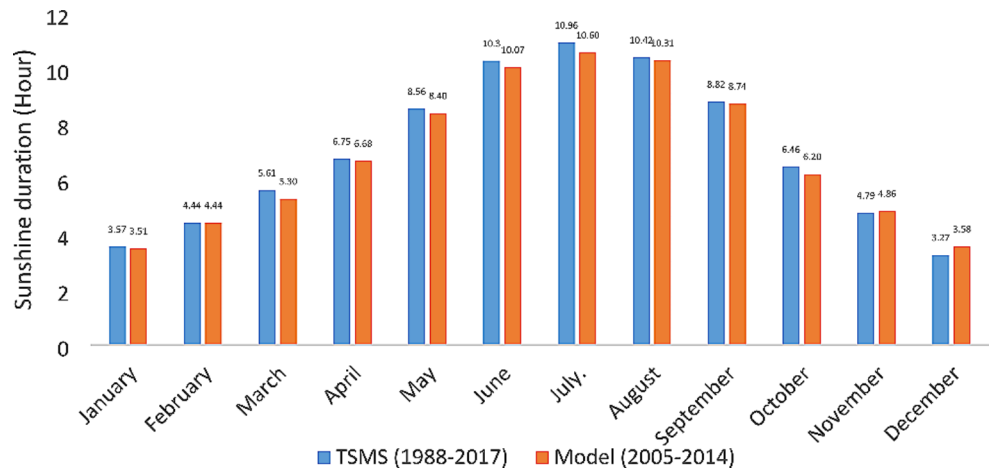
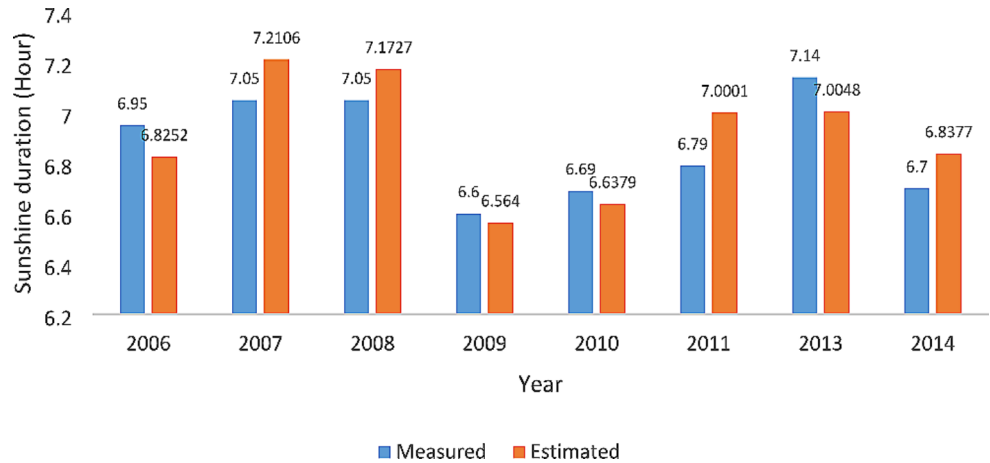


Fig. 8 The annual mean SD values of Türkiye were obtained from the model results and measured values



- The satellite-derived annual SD map and TSMS annual SD map (constructed interpolating SD values of stations) seemed nearly similar although data were not obtained simultaneously. This indicates that the annual variation of SD through the years is minimal for the dates from 1988 to 2014.
- The average monthly mean value of SD for Türkiye for the study period (2005–2014) and for 2014 was calculated as 6.7924 h and 6.8377, respectively.
- The highest SD values were obtained for July 2007 having an average value of 11.0366 h.
- The lowest SD values were obtained for January 2010, having an average value of 2.3534 h.
- The northeast part of the country (Black Sea and northern part of Marmara Region) was found to have lower SD values than the southeast part (Southern East Anatolian, Mediterranean, and southern part of Aegean region). That is, SD values generally increase as one moves from a higher latitude to lower latitude and decreased from lower altitude to a higher altitude.

is, under the coverage of the satellite of the interest, can be successfully estimated using a machine learning model, which uses satellite-derived cloud data, elevation data, day-length and ground measured meteorological data as inputs. Therefore, spatially continuous SD maps could be produced without using interpolation techniques.

Acknowledgements The authors would like to thank the Ç. U. BAP for the support and Turkish State Meteorological Service (TSMS), The European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) Climate Monitoring Satellite Application Facility (CM SAF) project and The U.S. Geological Survey (USGS) for providing the data used in the study.

Author contributions Kazım Kaba contributed to the preparation of materials such as maps, tables and plots, the writing manuscript and editing of the article. Erdem Erdi provided for collecting data, the application of the method, the production of results and the preparation of materials such as maps, tables and graphics. Mutlu Avcı contributed to the selection of the method, the implementation of the method and the writing of the method section. H. Mustafa Kandırmaz contributed to the writing process by designing the subject, method and format of the article.

Funding This study was supported by Çukurova University (Ç. U.) Scientific Research Projects (BAP) Unit (Project No: FYL-2014-3286).

Thus, results of this study have shown that SD (monthly mean SD for the present study) values over any area, which

Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data availability The python script of the ANN model developed in this study and Türkiye's average monthly sunshine duration data, which are the estimation results made with this model, can be accessed at <https://github.com/kkaba46/SunshineDuration>. The model's input data, cloud data, is available https://www.cmsaf.eu/EN/Home/home_node.html and DEM data is available <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-multi-resolution-terrain-elevation>.

Code availability Codes are available on request.

Declarations

Ethical approval The authors confirm that this article is an original research.

Consent to participate The authors confirm that this article has not been previously published in any journal.

Consent for publication The authors have agreed to submit this manuscript in its current form for publication in the journal.

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahn JB, Kim YH, Shim KM et al (2021) Climatic yield potential of Japonica-type rice in the Korean Peninsula under RCP scenarios using the ensemble of multi-GCM and multi-RCM chains. *Int J Climatol* 41:E1287–E302. <https://doi.org/10.1002/joc.6767>
- Akgün N, Açıkgöz M, Çelebi U et al (2021) The effect of weather variables on the severity, duration and frequency of headache attacks in the cases of episodic migraine and episodic tension-type headache. *Turk J Med Sci*. <https://doi.org/10.3906/sag-2004-66>
- Badescu V (1999) Correlations to estimate monthly mean daily solar global irradiation: application to Romania. *Energy* 24(10):883–893. [https://doi.org/10.1016/S0360-5442\(99\)00027-4](https://doi.org/10.1016/S0360-5442(99)00027-4)
- Bartoszek K, Matuszko D, Węglarczyk S (2021) Trends in sunshine duration in Poland (1971–2018). *Int J Climatol* 41(1):73–91. <https://doi.org/10.1002/joc.6609>
- Baumgartner D, Pötzi W, Freislich H et al (2018) A comparison of long-term parallel measurements of sunshine duration obtained with a Campbell-Stokes sunshine recorder and two automated sunshine sensors. *Theoret Appl Climatol* 133(1):263–275. <https://doi.org/10.1007/s00704-017-2159-9>
- Brown I (2013) Influence of seasonal weather and climate variability on crop yields in Scotland. *Int J Biometeorol* 57(4):605–614. <https://doi.org/10.1007/s00484-012-0588-9>
- Bulut U, Sakalli A (2021) Impacts of climate change and distribution of precipitation on hydroelectric power generation in Turkey. Paper presented at the IOP Conference Series: Materials Science and Engineering
- Deniz A, Toros H, Incecik S (2011) Spatial variations of climate indices in Turkey. *Int J Climatol* 31(3):394–403. <https://doi.org/10.1002/joc.2081>
- Duffie JA, Beckman WA (2013) *Solar engineering of thermal processes*. Wiley
- El-Metwally M (2005) Sunshine and global solar radiation estimation at different sites in Egypt. *J Atmos Solar Terr Phys* 67(14):1331–1342. <https://doi.org/10.1016/j.jastp.2005.04.004>
- Essa KS, Etman SM (2004) On the relation between cloud cover amount and sunshine duration. *Meteorol Atmos Phys* 87(4):235–240. <https://doi.org/10.1007/s00703-003-0046-7>
- Fox RL (1961) Sunshine-cloudiness relationships in the United States. *Mon Weather Rev* 89(12):543–548. <https://doi.org/10.1175/1520-0493>
- Good E (2010) Estimating daily sunshine duration over the UK from geostationary satellite data. *Weather* 65(12):324–328. <https://doi.org/10.1002/wea.619>
- Haykin S (1994) *Neural Networks: A Comprehensive Foundation*, (MacMillan, New York)
- Hu C, Kang P, Jaffe DA et al (2021) Understanding the impact of meteorology on ozone in 334 cities of China. *Atmos Environ* 248:118221. <https://doi.org/10.1016/j.atmosenv.2021.118221>
- Journée M, Demain C, Bertrand C (2013) Sunshine duration climate maps of Belgium and Luxembourg based on Meteosat and in-situ observations. *Adv Sci Res* 10(1):15–19. <https://doi.org/10.5194/asr-10-15-2013>
- Kaba K, Yeşilyaprak C (2021) CM SAF CFC Bulut Verisinin Doğruluk Testi ve Doğu Anadolu Gözlemevi (DAG) Yerleşkesi için Analizi. *J Adv Res Nat Appl Sci* 7(3):304–318. <https://doi.org/10.28979/jarnas.871585>
- Kaba K, Kandirmaz HM, Avci M (2017) Estimation of daily sunshine duration using support vector machines. *Int J Green Energy* 14(4):430–441. <https://doi.org/10.1080/15435075.2016.1265971>
- Kaiser DP, Qian Y (2002) Decreasing trends in sunshine duration over China for 1954–1998: indication of increased haze pollution? *Geophys Res Lett* 29(21):38–31. <https://doi.org/10.1029/2002GL016057>
- Kandirmaz HM (2006) A model for the estimation of the daily global sunshine duration from meteorological geostationary satellite data. *Int J Remote Sens* 27(22):5061–5071. <https://doi.org/10.1080/01431160600840960>
- Kandirmaz HM, Kaba K (2014) Estimation of daily sunshine duration from Terra and Aqua Modis data. *Advances in Meteorology* 2014. <https://doi.org/10.1155/2014/613267>
- Kandirmaz HM, Kaba K, Avci M (2014) Estimation of monthly sunshine duration in Turkey using artificial neural networks. *Int J Photoenergy* 2014. <https://doi.org/10.1155/2014/680596>
- Kim D, Ramanathan V (2008) Solar radiation budget and radiative forcing due to aerosols and clouds. *J Geophys Research: Atmos* 113(D2). <https://doi.org/10.1029/2007JD008434>
- Kothe S, Good E, Obregón A et al (2013) Satellite-based sunshine duration for Europe. *Remote Sens* 5(6):2943–2972. <https://doi.org/10.3390/rs5062943>
- Li H, Ma W, Lian Y et al (2011) Global solar radiation estimation with sunshine duration in Tibet, China. *Renewable Energy* 36(11):3141–3145. <https://doi.org/10.1016/j.renene.2011.03.019>
- Liao Y, Wang Z, Xiong J et al (2021) Dimming in the Pearl River Delta of China and the major influencing factors. *Climate Res* 82:161–176. <https://doi.org/10.3354/cr01626>

- Liu B, Liang M, Huang Z et al (2021) Duration–severity–area characteristics of drought events in eastern China determined using a three-dimensional clustering method. *Int J Climatol* 41:E3065–E84. <https://doi.org/10.1002/joc.6904>
- Lu N, Qin J, Yang K et al (2011) A simple and efficient algorithm to estimate daily global solar radiation from geostationary satellite data. *Energy* 36(5):3179–3188. <https://doi.org/10.1016/j.energy.2011.03.007>
- Matuszko D (2012) Influence of cloudiness on sunshine duration. *Int J Climatol* 32(10):1527–1536. <https://doi.org/10.1002/joc.2370>
- Matuszko D, Węglarczyk S (2015) Relationship between sunshine duration and air temperature and contemporary global warming. *Int J Climatol* 35(12):3640–3653. <https://doi.org/10.1002/joc.4238>
- Matzarakis A, Katsoulis V (2006) Sunshine duration hours over the Greek region. *Theoret Appl Climatol* 83(1):107–120. <https://doi.org/10.1007/s00704-005-0158-8>
- McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5(4):115–133. <https://doi.org/10.1007/BF02478259>
- Mieczkowski Z (1985) The tourism climatic index: a method of evaluating world climates for tourism. *Can Geographer/Le Géographe Canadien* 29(3):220–233. <https://doi.org/10.1111/j.1541-0064.1985.TB00365.X>
- Mohandes MA, Rehman S (2013) Estimation of sunshine duration in Saudi Arabia. *J Renew Sustain Energy* 5(3):033128. <https://doi.org/10.1063/1.4811284>
- Piotrowski AP, Napiorkowski MJ, Napiorkowski JJ et al (2015) Comparing various artificial neural network types for water temperature prediction in rivers. *J Hydrol* 529. <https://doi.org/10.1016/j.jhydrol.2015.07.044>. :302–15
- Quej VH, Almorox J, Arnaldo JA et al (2017) ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. *J Atmos Solar Terr Phys* 155:62–70. <https://doi.org/10.1016/j.jastp.2017.02.002>
- Rahimikhooob A (2014) Estimating sunshine duration from other climatic data by artificial neural network for ET 0 estimation in an arid environment. *Theoret Appl Climatol* 118(1):1–8. <https://doi.org/10.1007/s00704-013-1047-1>
- Rangarajan S, Swaminathan M, Mani A (1984) Computation of solar radiation from observations of cloud cover. *Sol Energy* 32(4):553–556. [https://doi.org/10.1016/0038-092X\(84\)90270-6](https://doi.org/10.1016/0038-092X(84)90270-6)
- Robaa S (2008) Evaluation of sunshine duration from cloud data in Egypt. *Energy* 33(5):785–795. <https://doi.org/10.1016/j.energy.2007.12.001>
- Sanchez-Lorenzo A, Wild M (2012) Decadal variations in estimated surface solar radiation over Switzerland since the late 19th century. *Atmos Chem Phys* 12(18):8635–8644. <https://doi.org/10.5194/acp-12-8635-2012>
- Sanchez-Lorenzo A, Calbó J, Martin-Vide J (2008) Spatial and temporal trends in sunshine duration over Western Europe (1938–2004). *J Clim* 21(22):6089–6098. <https://doi.org/10.1175/2008JCLI2442.1>
- Sánchez-Lorenzo A, Calbó J, Wild M et al (2013) New insights into the history of the Campbell-Stokes sunshine recorder. <https://doi.org/10.1002/wea.2130>
- Schaul T, Bayer J, Wierstra D et al (2010) PyBrain J Mach Learn Res 11(ARTICLE):743–746
- Shamim MA, Remesan R, Han D-w et al (2012) An improved technique for global daily sunshine duration estimation using satellite imagery. *J Zhejiang Univ Sci A* 13(9):717–722. <https://doi.org/10.1631/jzus.A1100292>
- Stanhill G, Cohen S (2005) Solar radiation changes in the United States during the twentieth century: evidence from sunshine duration measurements. *J Clim* 18(10):1503–1512. <https://doi.org/10.1175/JCLI3354.1>
- Tejeda A, Vargas A (1996) A correlation between visual observations and instrumental records of cloudiness in Mexico. *Geofisica Int* 35(4)
- Trnka M, Žalud Z, Eitzinger J et al (2005) Global solar radiation in central European lowlands estimated by various empirical formulae. *Agric for Meteorol* 131(1–2):54–76. <https://doi.org/10.1016/j.agrformet.2005.05.002>
- Wang C, Shi X, Liu J et al (2021) Interdecadal variation of potato climate suitability in China. *Agric Ecosyst Environ* 310:107293. <https://doi.org/10.1016/j.agee.2020.107293>
- Werbos PJ (1988) Generalization of backpropagation with application to a recurrent gas market model. *Neural Netw* 1(4):339–356. [https://doi.org/10.1016/0893-6080\(88\)90007-X](https://doi.org/10.1016/0893-6080(88)90007-X)
- Wood CR, Harrison RG (2011) Scorch marks from the sky. *Weather* 66(2):39–41. <https://doi.org/10.1002/wea.657>
- Yılmaz Z (2007) Evaluation of energy efficient design strategies for different climatic zones: comparison of thermal performance of buildings in temperate-humid and hot-dry climate. *Energy Build* 39(3):306–316. <https://doi.org/10.1016/j.enbuild.2006.08.004>
- You Q, Kang S, Flügel W-A et al (2010) From brightening to dimming in sunshine duration over the eastern and central tibetan Plateau (1961–2005). *Theoret Appl Climatol* 101(3):445–457. <https://doi.org/10.1007/s00704-009-0231-9>
- Zhang H, Sun R, Peng D et al (2021) Spatiotemporal Dynamics of Net Primary Productivity in China’s Urban lands during 1982–2015. *Remote Sens* 13(3):400. <https://doi.org/10.3390/rs13030400>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.