DOI:10.28974/idojaras.2023.3.2

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 127, No. 3, July – September, 2023, pp. 285–298

Estimating the sunshine duration using multiple linear regression in Kocaeli, Turkey

Mine Tulin Zateroglu

University of Cukurova
Vocational School of Technical Sciences
(01350) Adana/Turkey

Author E-mail: mtzateroglu@cu.edu.tr

(Manuscript received in final form September 27, 2022)

Abstract— This study aims to estimate and evaluate the characteristic behavior of sunshine duration for long-term records. Sunshine duration and other climate variables such as cloudiness, precipitation, relative humidity, etc., have been measured in meteorological stations for a long time all over the world. But in some cases, such as missing data or unavailable station, the estimation of sunshine duration play a crucial role. Statistical models can be used to predict the sunshine duration over climate variables. To evaluate the behavior of sunshine duration, several climate variables were analyzed for different time scales. The data used in this study were collected from a ground-based meteorological station. In the first, all data were arranged according to different time scales as monthly, seasonal, and annual average values. Prediction models were constructed for each time scale. This study used multiple linear regression (MLR) to build the models and the Pearson correlation analysis to determine the relations between the climate elements. The created models for estimating sunshine duration were validated as well. According to the results, MLR can be utilized and recommended for the prediction of the sunshine duration over climate variables.

Key-words: sunshine duration, estimating, climate variables, Pearson correlation analysis, MLR

1. Introduction

Sunshine duration, which is a key element for solar radiation, has been widely studied over the recent years. The energy that comes from solar radiation is clean and environmentally friendly. Sunshine duration is highly related to solar radiation via the Angström-Prescott model that is utilized for predicting the amount of daily global solar radiation.

The measurement of sunshine duration is made pointwise at meteorological stations. In places, where there is no station and no measurement can be made, the sunshine duration values are estimated using different methods such as regression analysis and interpolation techniques. Sunshine duration is correlated to climate variables in the atmospheric environment. Some researchers used climate variables to determine their relationship to sunshine duration and to estimate the value of sunshine duration. An empirical formula was presented to estimate the sunshine duration using the cloud amount data (*Reddy*, 1974). Similarly, *Stanghellini* (1981) developed an empirical formula to predict the monthly sunshine duration via daily mean cloudiness values. *Chagnon* (1981) determined that a high amount of cirrus-type clouds originating from jet planes caused a decrease in sunshine duration. Sunshine duration is negatively related to cloudiness (*Angell et al*., 1984; *Essa* and *Etman*, 2004; *Hoyt*, 1977; *Palle* and *Butler*, 2001; *Robaa*, 2008; *Sanchez-Lorenzo et al*., 2009; *Weber*, 1994; *You et al*., 2010). Besides, similarly to cloudiness, relative humidity was also found as negatively correlated to sunshine duration (*Aksoy*, 1999; *Yang et al*., 2009a; *You et al*., 2010; *Zateroglu*, 2021a). Furthermore, sunshine duration was declared that positively correlated with wind speed (*Yang et al*., 2009a, 2009b). Additionally, *Sanchez-Lorenzo et al*. (2009) expressed a positive relationship between sunshine duration and atmospheric pressure. Some studies were shown that sunshine duration was negatively related to precipitation (*Yang et al*., 2009a; *You et al*., 2010; *Zateroglu*, 2021a). *Aksoy* (1999) evaluated the changes in sunshine duration over the changes in other climate parameters for Ankara in Turkey. *Yildirim et al*. (2013) investigated the trends of observed sunshine duration data and found a decrease in sunshine duration due to anthropogenic air pollution. Furthermore, air pollutants such as particulate matter and sulfur dioxide influence the sunshine duration and are associated with environmental parameters in the atmospheric periphery over urban areas (*Zateroglu*, 2021b, 2021c and 2022). Additionally, particulate matter and sulfur dioxide decrease the amount of sunshine duration (*Zateroglu*, 2021a). Also, the Pasquill–Gifford–Turner (PGT) scheme, which is used in predicting vertical and horizontal dispersion of a plume in air pollution models, considers the steady-state atmospheric conditions such as the quantity of solar radiation, fractional cloudiness, horizontal surface wind speed, and also vertical temperature gradient (*USEPA*, 1993; *Venkatram*, 1996). The air pollutants reflect and scatter solar radiation and then reduce the surface temperature. Sunshine duration is related to solar radiation with the Ansgtröm-

Prescott formula which is used for estimating the global solar radiation; so sunshine duration is associated with air pollutants (*Zateroglu*, 2022).

Sunshine duration is measured at meteorological stations but anyway, in some cases, measurements cannot be done due to some conditions such as remote areas, geographical problems, and not existing or insufficient stations. Different methods have been used to predict climate elements in climatological studies. Linear regression analysis was used to estimate sunshine duration (*Stanghellini*, 1981). This method was preferred in terms of compatibility with climate data, ease of operation, and efficient outcomes. Sunshine duration data were obtained from the Campbell-Stokes instrument. This equipment records the sunshine data by burning a specific card upon which sun rays were focused via a glass sphere of the sunshine recorder (*WMO*, 1996).

The main purpose of the present study was to gain the prediction models for sunshine duration over a statistical approach. Several climate variables were used as variables in building models. This study focuses on the 1961–2010 period. Monthly mean values of daily climate elements were taken for estimating the sunshine duration measured by a ground-based meteorological station. Furthermore, the accuracy of the empirical models obtained from the statistical analysis was evaluated via validation parameters of the regression. Finally, performance indices were implemented for the suitability of the prediction models. The findings were discussed and interpreted over the results of validation indicators.

2. Study area and data

This study was conducted for Kocaeli in the northwestern region of Turkey. The province is located between 29°22'E–30°21'E longitudes, 40°31'N– 41°13'N latitudes at 76 m altitude. There are intensive industrial activities and transportation facilities in the area. The population of the urban area is continually increasing as a result of developing industrialization. In the province, a temperate climate prevails on the Izmit Gulf coasts and the Black Sea coast, and a harsher climate prevails in the mountainous areas. It can be said that the climate of Kocaeli constitutes a transition between the Mediterranean climate and the Black Sea climate. In the city center, summers are hot and less rainy, and winters are rainy, snowy, and cold from time to time. There are some differences between the climate of Kocaeli's coasts facing the Black Sea and the coasts facing the Izmit Gulf. While sometimes sweltering heat is experienced on the gulf coasts in summer, the Black Sea coasts are cooler. According to the long-term records, the annual mean maximum air temperature is 19.8 °C, the annual mean minimum air temperature is 10.8 °C, the annual average monthly precipitation is 815.2 mm, annual average rainy days is 150.1, the number of annual mean sunshine duration is 5.7 hour, the annual average relative humidity is 71.7%, and the annual mean wind speed is 1.6 m/s. The average annual precipitation on the Black Sea coast exceeds 1,000 mm. This amount decreases in the south; and falls below 800 mm (784.6 mm). On the slopes of the mountains facing the gulf, the climate is similar to the Black Sea coast. The amount of precipitation is also different in this section. Winds blow from the north and northeast in winter and from the northeast in summer.

In the present study, daily values of climate elements such as sunshine duration (SD), cloudiness (CLD), relative humidity (RHUM), wind speed (WS), precipitation (PREC), evaporation (EVAP), atmospheric pressure (PRES), minimum air temperature (TMIN), and maximum air temperature (TMAX) were obtained from the ground-based observation station at Izmit/Kocaeli. The measurements were realized by the Turkish State Meteorological Service (TSMS). Monthly average values of climate variables were computed over daily data. The values of relative sunshine duration were calculated for prediction models. The arranged data were analyzed by using a statistical approach, then the obtained statistical models were validated.

3. Methods

The relative sunshine duration *RSD* is defined as the ratio of the measured (S) and maximum possible daily (S_o) sunshine duration and its value varies between 0 and 1. S_o is calculated using the formula as follows (*Duffie* and *Beckman*, 1991; *Goswami,* 2015; *Kalogirou*, 2014):

$$S_0 = \left(\frac{2}{15}\right) cos^{-1}(-tan\delta tan\varphi) , \quad (1)$$

$$\delta = 23.45 * sin\left(\frac{360}{365}(284 + d)\right) , \quad (2)$$

where φ is the latitude angle ($-90 \leq \varphi \leq +90$) and depends on the location of interest, δ defines the solar declination angle between the equatorial plane and incoming solar rays, and d determines the number of days of the year (begins from January 1).

The Pearson correlation coefficient is a scale which denotes the strength of the linear correlation between variables. This metric determines not only the quantity but also the direction of the relation. It is computed as follows:

$$r(s_i, s_j) = \frac{Cov(s_i, s_j)}{\sigma(s_i)\sigma(s_j)} , \quad (3)$$

where s_i and s_j are the measured values of the two climate variables, $\sigma(s_i)$ and $\sigma(s_j)$ are the standard deviations, and $Cov(s_i, s_j)$ denotes the covariance of s_i

and s_j. $r(s_i, s_j)$ demonstrates the correlation coefficient between the *RSD* and either climate variable. According to value of the Pearson correlation coefficient, the relations are categorized as low (value in 0.0–0.49), moderate (value in 0.5–0.69) and high (value in 0.7–1.0).

Regression-based techniques have been utilized in the estimation studies of climate parameters. To build the models for the climate and the other atmospheric elements, the multiple linear regression (MLR) method is commonly preferred for estimation among the several statistical methods. In this statistical approach, the method processes the dataset that fit the normal distribution. Climate data is appropriate for this analysis. MLR represents the relationships between dependent (response) and independent (predictor or explanatory) variables. In this study, *RSD* is the dependent variable, the other climate elements are the independent variables. The method reveals the number of changes in *RSD* as a percentage that is explained by other climate variables. The relationship between dependent and independent variables obtained from MLR analysis is defined as a mathematical model:

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots\dots\dots + a_rX_r + \mathcal{E} \,, \tag{4}$$

where Y expresses the dependent variable which consists of a measured data matrix with dimension (n x 1), n is the number of measurements, $X_1, X_2, \dots\dots, X_r$ define the independent variables, X determines the measured data matrix with dimension ($n \times r$), r denotes the number of independent variables, a_o denotes the constant, $a_1, a_2, \dots\dots, a_r$ demonstrate the regression coefficients, a is the coefficient matrix with dimension ($r \times 1$), $\mathcal{E}$ determines the predicted error term. To minimize the error term, the values of constant and regression coefficients are computed by the least squares method utilizing the coefficient matrix a, which is determined by formula $a = (X^TX)^{-1}(X^TY)$. X^T is the transpose of matrix X. The significance levels of constant and regression coefficients are determined over the t value and F distribution. In MLR analysis, the accuracy of the models is judged by two indicators named the coefficient of determination (R^2) and the standard error of estimation (*SEE*). R^2 denotes a measure of how well the predicted model fits the data used. It is expressed as the percent value changes from 0 to 1. SEE gives the amount of difference between actual and estimated values.

$$R^2 = 1 - \frac{\sum(P_k - \overline{M_k})^2}{\sum(M_k - \overline{M_k})^2} \,, \tag{5}$$

$$SEE = \sqrt{\frac{\sum(M_k - P_k)^2}{n-2}} \,, \tag{6}$$

where P_k and M_k determine the predicted and measured values, respectively, and n is the number of measurements. The level of the confidence interval was taken into account as 95% in constructing the empirical models.

To be evaluated on the same scale, all climate data were standardized before constructing the figures. The values of climate variables were transformed to normalized values (vary between 0 and 1) concerning the following formula:

$$I_i = \frac{I_j - I_{min}}{I_{max} - I_{min}} \quad , \tag{7}$$

where I_j expresses the observed value, I_i determines the normalized value of I_j. I_{min} is the minimum value and I_{max} is the maximum value of the related dataset.

To verify the suitability of the prediction, some error terms are applied to the built models. The widely used performance indices, the root mean square error (RMSE), mean bias error (MBE), mean absolute error (MAE), percentage mean absolute error (MAPE), normalized mean square error (NMSE), fractional bias (FB), and index of agreement (IOA) were utilized to interpret the accuracy of the twelve models for months. These seven indices were calculated by the following formulas:

$$RMSE = \sqrt{\frac{\sum_{k=1}^{n}(P_k - M_k)^2}{n}}, \tag{8}$$

$$NMSE = \frac{\overline{(M_k - P_k)^2}}{\overline{M_k} * \overline{P_k}} \quad , \tag{9}$$

$$MBE = \frac{\sum_{k=1}^{n}(P_k - M_k)}{n} \quad , \tag{10}$$

$$MAE = \frac{\sum_{k=1}^{n}|P_k - M_k|}{n}, \tag{11}$$

$$MAPE = \frac{100}{n}\sum_{k=1}^{n}\frac{|P_k - M_k|}{M_k} \quad , \tag{12}$$

$$FB = \frac{(\overline{M_k} - \overline{P_k})}{0.5 * (\overline{M_k} + \overline{P_k})}, \tag{13}$$

$$IOA = 1 - \frac{\sum_{k=}^{n}(P_k - M_k)^2}{\sum_{k=1}^{n}(|P_k - \bar{M}_k| + |M_k - \bar{M}_k|)^2} \quad , \tag{14}$$

where P_k, and M_k denote the predicted and measured values respectively, $\overline{M_k}$ determines the mean value of the measured values, and n defines the number of measurements.

The performance indices shown in Eqs. (8)-(14) may be used to evaluate the efficiency of the prediction models. Commonly, as much as small (i.e., close to zero) for the value of RMSE, MBE, MAE, MAPE, NMSE, and FB, but as far as big (i.e., near 1) for the value of IOA are acceptable for the success of the predictions. The values of FB are limited by -2 to +2. Positive values of FB describe a constructed model under-estimation, and the negative values describe a constructed model over-estimation.

To determine the distribution of climate variables, one-sample Kolmogorov-Smirnov test was implemented on a dataset for the normality test. The SPSS (Statistical Package for Social Science) package program was utilized to examine the statistical analysis.

4. Results and discussion

Pearson correlation analysis was used to reveal the statistically significant correlations between relative sunshine duration and climate variables. The values shown in *Table 1* were expressed in the directions such as positive and negative, and magnitudes of the relations for all months, i.e., from December to November (DE, JA, FE, MR, AP, MA, JN, JL, AU, SE, OC, NO). For EVAP, there were no available (NA) data for the months JA, FE, and MR. According to *Table 1*, RSD was correlated with CLD, RHUM, and PREC negatively, whereas it was correlated with WS, EVAP, PRES, TMIN, and TMAX positively despite the ignored exception cases. The statistically significant correlation coefficients were represented in bold in *Table 1*. In all months, CLD, RHUM, and PREC (except MR) have significant correlations with RSD on moderate and high levels for CLD, and low and moderate levels for RHUM and PREC. RSD was associated with WS and TMAX on weak levels, and with PRES on weak and moderate levels.

Table 1. Pearson correlation matrix for RSD

Time scale	CLD	RHUM	WS	PREC	EVAP	PRES	TMIN	TMAX
DE	**-0.674****	**-0.443****	-0.01	**-0.334***	0.075	**0.500****	-0.093	-0.009
JA	**-0.725****	**-0.336***	0.002	**-0.500****	NA	**0.517****	-0.069	-0.239
FE	**-0.749****	**-0.415****	-0.125	**-0.570****	NA	0.251	0.214	0.277
MR	**-0.834****	**-0.620****	-0.154	-0.014	NA	0.245	0.071	**0.349***
AP	**-0.844****	**-0.596****	0.021	**-0.500****	**0.437****	**0.406****	0.119	0.197
MA	**-0.796****	**-0.679****	0.114	**-0.527****	**0.772****	0.176	0.028	**0.281***
JN	**-0.775****	**-0.682****	0.194	**-0.387****	**0.608****	0.26	-0.095	0.278
JL	**-0.718****	**-0.602****	0.117	**-0.493****	**0.583****	0.023	0.043	0.163
AU	**-0.866****	**-0.671****	**0.305***	**-0.517****	**0.679****	-0.087	0.103	**0.449****
SE	**-0.808****	**-0.482****	0.245	**-0.488****	**0.735****	-0.02	0.128	0.242
OC	**-0.784****	**-0.631****	**0.352***	**-0.463****	**0.777****	-0.042	0.279	**0.427****
NO	**-0.831****	**-0.617****	0.064	**-0.534****	**0.412***	0.001	0.279	**0.329***

* Correlation is significant at the 0.05 level ($p<0.05$)
** Correlation is significant at the 0.01 level ($p<0.01$)

Besides, the variations of monthly correlation coefficients between RSD and each climate parameter were plotted in *Fig. 1*. It was seen that the Pearson correlation coefficients for CLD and RHUM varied similarly. Additionally, the changes in the correlation coefficient for TMIN and TMAX were shown as nearly close characteristics. The correlation coefficients of EVAP suddenly began to decrease from two months, MA and OC. For PREC, the coefficient sharply has a minimum value in MR. The highest statistically significant correlations were found for CLD in AU as high, RHUM in JN as moderate, WS in OC as low, PREC in FE as moderate, EVAP in OC as high, PRES in JA as moderate, TMAX in AU as low.

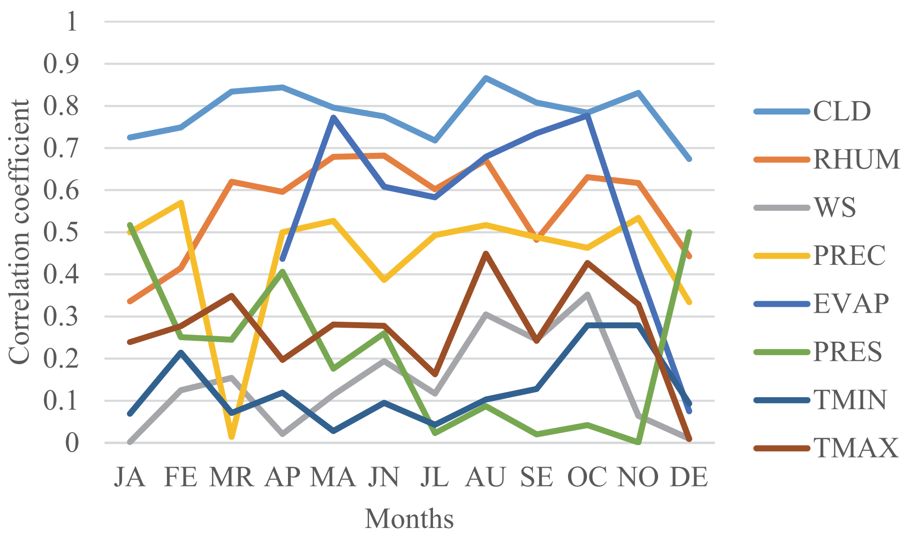


Fig. 1. Variations of correlation coefficients between RSD and each climate variable.

The monthly variations of climate variables were represented as normalized values in *Fig.2 (a-h)*. The normalized values of related climate variables were given below the table. As seen in *Fig.2. a, b*, RSD has high values in JN, JL, and AU and low values in DE, JA, and FE contrary to CLD and RHUM. RSD was changed reversely with CLD and RHUM. When RSD increased/decreased, CLD and RHUM decreased/increased. PREC and PRES behaved similarly against RSD as changed oppositely (*Fig. 2. c, d*).

The characteristics of monthly TMAX and TMIN had close behavior to the changes in RSD (*Fig.2. e, f*). Although some monthly data were missing, the variations in EVAP was completely compatible with RSD (*Fig. 2. g*). According to *Fig. 2 h*, the variation of WS showed different features compared to RSD. The amount of change in the increasing and decreasing behaviours of WS was not in the same portion as in RSD.

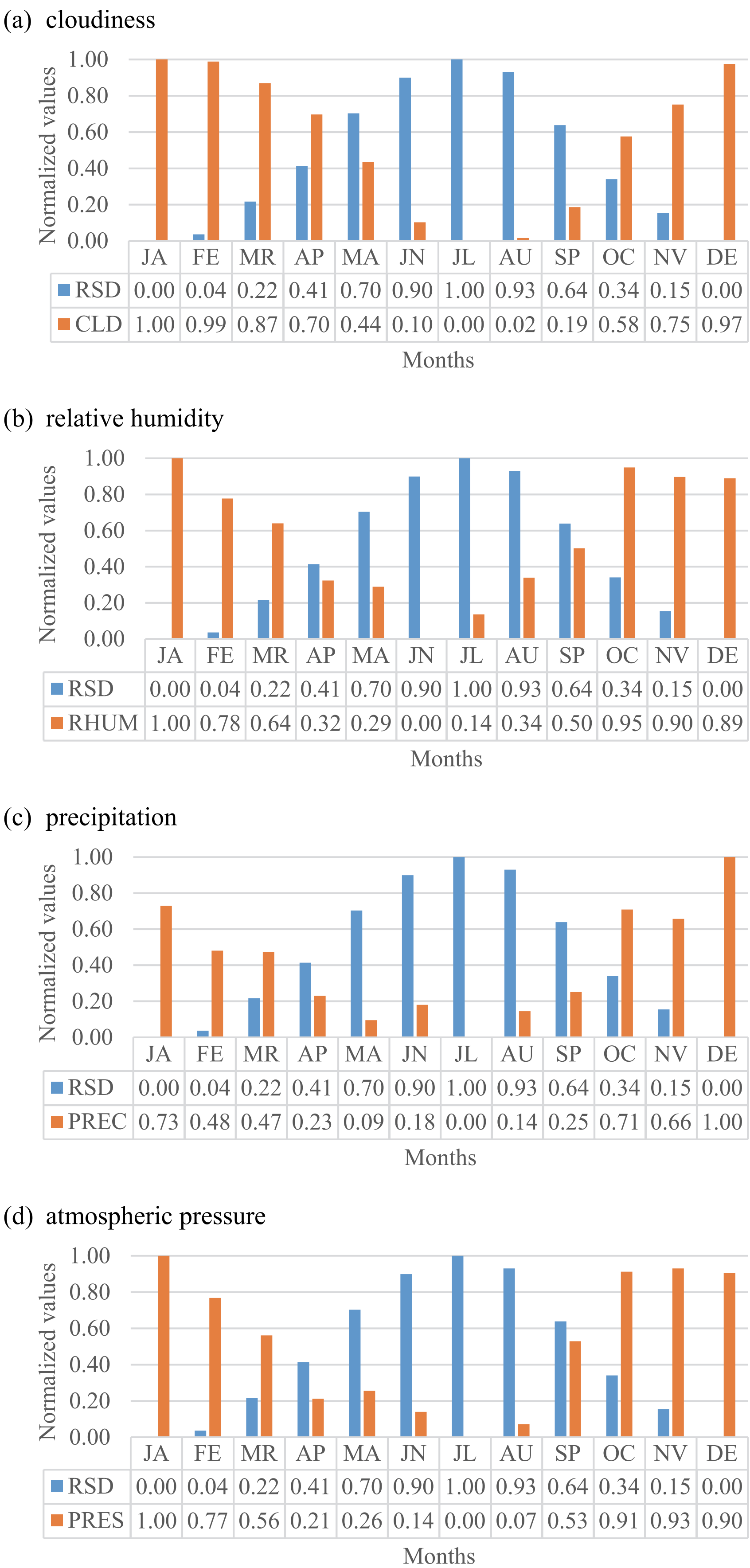
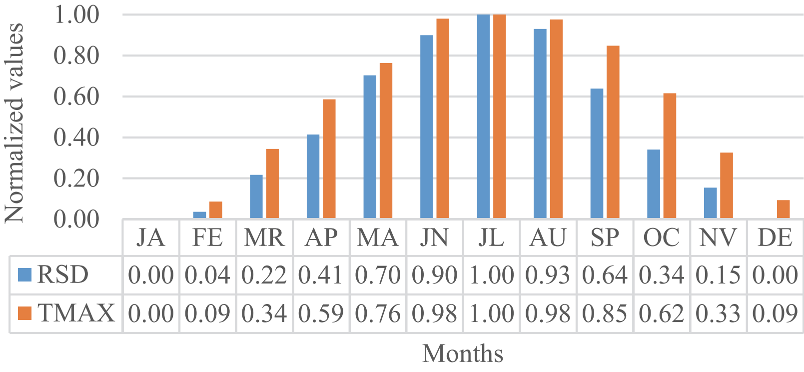


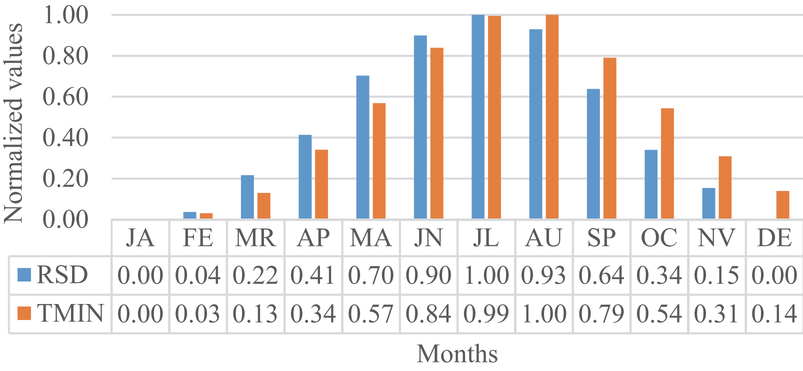
Fig. 2. Monthly normalized values for the RSD and the climate variables.

(e) maximum air temperature



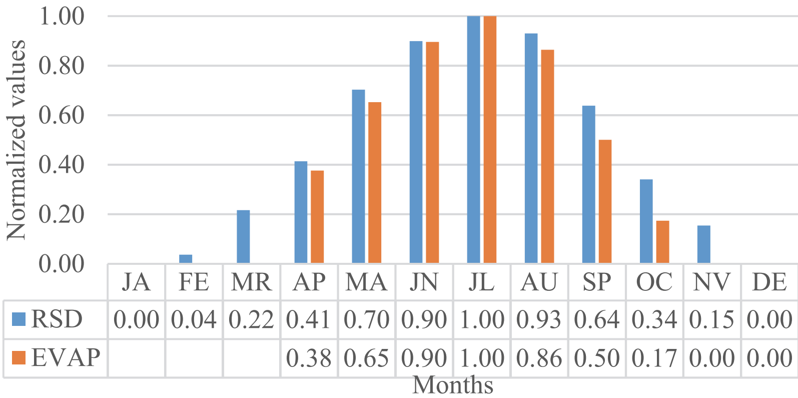
	JA	FE	MR	AP	MA	JN	JL	AU	SP	OC	NV	DE
RSD	0.00	0.04	0.22	0.41	0.70	0.90	1.00	0.93	0.64	0.34	0.15	0.00
TMAX	0.00	0.09	0.34	0.59	0.76	0.98	1.00	0.98	0.85	0.62	0.33	0.09

(f) minimum air temperature



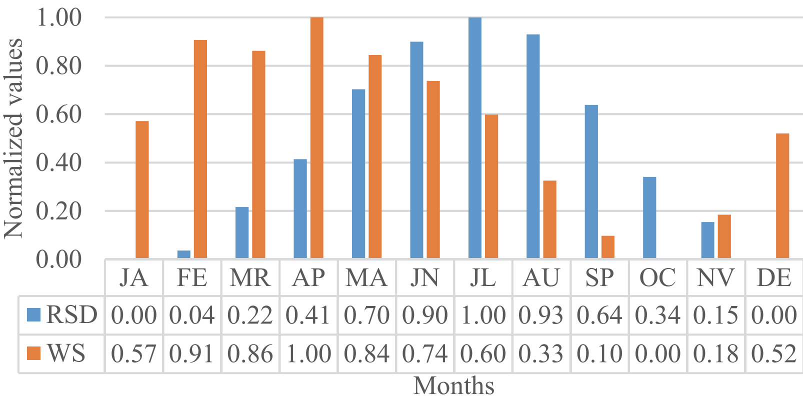
	JA	FE	MR	AP	MA	JN	JL	AU	SP	OC	NV	DE
RSD	0.00	0.04	0.22	0.41	0.70	0.90	1.00	0.93	0.64	0.34	0.15	0.00
TMIN	0.00	0.03	0.13	0.34	0.57	0.84	0.99	1.00	0.79	0.54	0.31	0.14

(g) evaporation



	JA	FE	MR	AP	MA	JN	JL	AU	SP	OC	NV	DE
RSD	0.00	0.04	0.22	0.41	0.70	0.90	1.00	0.93	0.64	0.34	0.15	0.00
EVAP				0.38	0.65	0.90	1.00	0.86	0.50	0.17	0.00	0.00

(h) wind speed



	JA	FE	MR	AP	MA	JN	JL	AU	SP	OC	NV	DE
RSD	0.00	0.04	0.22	0.41	0.70	0.90	1.00	0.93	0.64	0.34	0.15	0.00
WS	0.57	0.91	0.86	1.00	0.84	0.74	0.60	0.33	0.10	0.00	0.18	0.52

Fig. 2. Continue

To obtain a quantitative prediction of RSD, MLR analysis was performed. From this analysis, twelve empirical models were derived to estimate the RSD over several climate variables (*Table 2*). To specify the best predictors, the stepwise regression technique, which is one of the MLR methods, was implemented on the dataset. The superiority of this method was that the estimation of RSD was gained by statistically significant climate variables in the analysis. The order of the variables written in the prediction models also indicated their order of importance. Concerning *Table 2*, the best predictor for all months was CLD except for OC and EVAP. The sequel of sequencing could be expressed for the winter months as WS and PREC for DE; and PREC and WS for JA and FE. In MR, CLD was accompanied by RHUM and PRES. EVAP was in second-order in AP and MA, subsequently for TMIN and RHUM, respectively. For the remaining months, RHUM, PREC, TMIN, and EVAP were the second-order predictors that explain the RSD.

Table 2. Prediction models for RSD

Month	Model	R	R2	AdjR^2	SEE
JA	0.6397-0.0581*CLD-0.0007*PREC+0.0491*WS	0.842	0.709	0.685	0.04454
FE	0.7576-0.0696*CLD-0.0011*PREC+0.0465*WS	0.902	0.814	0.800	0.04232
MR	-4.0243-0.0496*CLD-0.0062*RHUM+0.0051*PRES	0.885	0.784	0.767	0.03772
AP	0.61-0.0514*CLD+0.0014*EVAP-0.0076*TMIN	0.892	0.795	0.775	0.03546
MA	0.7458-0.0479*CLD+0.0015*EVAP-0.0028*RHUM	0.913	0.833	0.817	0.03237
JN	1.0879-0.0539*CLD-0.0044*RHUM	0.878	0.772	0.758	0.03254
JL	0.8644-0.0618*CLD-0.0006*PREC	0.820	0.672	0.653	0.04030
AU	1.1272-0.0889*CLD-0.0109*TMIN	0.886	0.785	0.769	0.02824
SP	0.5268-0.0447*CLD+0.0022*EVAP	0.873	0.762	0.748	0.03622
OC	0.4889+0.0031*EVAP-0.0462*CLD	0.906	0.821	0.810	0.0358
NV	1.2586-0.0585*CLD-0.0074*RHUM	0.921	0.849	0.838	0.03866
DE	-7.5471-0.0597*CLD+0.0677*WS+0.008*PRES	0.798	0.637	0.609	0.04640

To achieve the accuracy of the prediction models, the value of R^2 is the best scale to indicate the success of the linear regression models. Mostly, R^2 and adjusted (Adj) R^2 were evaluated together to interpret the models. If the two values were close to each other, it was declared that the constructed model was appropriate. According to *Table 2*, R^2 and AdjR^2 values were nearly close to each other, so all models could be expressed as appropriate models. The calculated R^2 values were obtained as generally high levels except for JL and DE as moderate ones. For all time scales, SEE values were gained as small values desired for the suitable models.

To confirm of the concluded empirical models, the monthly mean values of RSD were computed for the period mentioned above. For either month, the predicted RSD value was calculated over the related climate elements by using the deduced model. Predicted and measured values of RSD were compared by implementing the performance indices (*Table 3*). The lower the RMSE, MBE, MAE, MAPE, NMSE, and FB and the higher the IOA, the smaller the model error and the preferable the estimation performance. The performance indices showed that the obtained regression models were appropriate to estimate the RSD for given time scales.

*Table 3***.** Statistical indicators

Term	RMSE	MBE	MAE	MAPE	NMSE	FB	IOA
JA	0.06218	0.01591	0.04369	9.17987	0.05805	0.04387	0.82250
FE	0.06297	0.02162	0.03555	6.64318	0.04567	0.11522	0.88391
MR	0.06187	0.01492	0.04013	7.27445	0.03354	0.04414	0.82524
AP	0.04978	-0.00845	0.03684	3.21047	0.01450	-0.00259	0.86526
MA	0.04064	0.00443	0.03166	2.34710	0.00641	-0.00486	0.91476
JN	0.03373	0.00291	0.02714	1.84488	0.00320	-0.01214	0.90923
JL	0.04618	-0.00643	0.03628	1.98324	0.00531	-0.02072	0.84372
AU	0.04553	0.00776	0.03206	1.93990	0.00488	-0.00332	0.89146
SP	0.048	-0.00819	0.03694	2.30729	0.00718	-0.03572	0.86043
OC	0.06071	0.00194	0.03976	3.88605	0.02006	-0.04139	0.84605
NV	0.05497	0.0000021	0.04073	4.67938	0.02273	-0.01660	0.89559
DE	0.05748	-0.02059	0.04595	7.41340	0.05761	-0.15991	0.77577

5. Conclusıon

In this study, the statistical modeling of the sunshine duration was implemented because of its importance in many applications, especially in predicting solar radiation. This study presents an analysis of the estimation of the monthly mean sunshine duration. The MLR analysis method is widely used in estimating the climate variables. Correlation analysis expresses the relations as strength and direction between RSD and climate elements mentioned as CLD, RHUM, WS, PREC, EVAP, TMIN, and TMAX. RSD was correlated with CLD, RHUM, and PREC negatively, while with WS, EVAP, PRES, TMIN, and TMAX positively. The level of strength of the relations differed as weak, moderate, and high order according to the time scales. Further, the constructed prediction models were obtained as compatible with the climate elements. Climate variables selected in the models explained the RSD successfully. Additionally, some cases may affect the accuracy of prediction models. For instance. measurements cannot be made accurately due to equipment calibration problems or climatic conditions. RHUM or/and PREC can affect the sunshine recorder and the sensitivity of the equipment. This may cause inaccurate measurements of sunshine duration. Besides, it should be noted that the amount of CLD is measured as visual observations by the observer, and this may cause measurement errors. Finally, the results for the statistical indicators demonstrated that the MLR method can be used for estimating the sunshine duration data for a specific location accurately.

Furthermore, according to the Pasquill-Gifford-Turner protocol, solar radiation. vertical air temperature gradient, CLD, and WS are highly associated with air pollutants; so air pollutants may affect the sunshine duration and can cause variations in its quantity because of the relationship between sunshine duration and solar radiation.

Additonally, the North Atlantic Oscillation (NAO) affects the region of the Black Sea. The NAO influences climate parameters, especially rainfall and air temperature. Therefore, because of the interactions between sunshine duration and other climate elements. NAO may influence the amount of sunshine duration in Kocaeli.

References

Aksoy, B., 1999: Analysis of changes in sunshine duration data for Ankara. Turkey. *Theor. Appl.Climatol.64* (3-4). 229–237. https://doi.org/10.1007/s007040050125

Angell, J.K., Korshover, J., and *Cotton, G.F.,* 1984: Variation in United States Cloudiness and Sunshine. 1950-82. *J. Appl. Meteorol.23.* 752–761. https://doi.org/10.1175/1520-0450(1984)023<0752:VIUSCA>2.0.CO;2

Changnon, S.A., 1981: Midwestern Cloud. Sunshine and Temperature Trends since 1901: Possible Evidence of Jet Contrail Effects. *J. Appl.Meteorol. 20*. 496–508. https://doi.org/10.1175/1520-0450(1981)020<0496:MCSATT>2.0.CO;2

Duffie, J.A. and *Beckman W.A*., 1991: Solar Engineering of Thermal Processes. John Wiley & Sons. New York. NY. USA. 2nd edition.

Essa, K.S. and *Etman. M.S.,* 2004: On the Relation Between Cloud Cover Amount and Sunshine Duration. *Meteorol. Atmos. Phys. 87*. 235–240. https://doi.org/10.1007/s00703-003-0046-7

Goswami, D.Y., 2015: Principles of Solar Engineering. CRC Press. Taylor and Francis Group. Boca Raton. London. New York. 3rd edition.

Hoyt, D.V., 1977: Percent of Possible Sunshine and the Total Cloud Cover. *Month.Weather Rev. 105*, 648–652. https://doi.org/10.1175/1520-0493(1977)105<0648:POPSAT>2.0.CO;2

Kalogirou, S.A., 2014: Solar Energy Engineering Processes and Systems. Elsevier Inc. Amsterdam. Netherlands. 2nd edition. 815 pp.

Palle, E. and *Butler, C.J.,* 2001: Sunshine Records From Ireland: Cloud Factors and Possible Links to Solar Activity and Cosmic Rays. *Int. J. Climatol. 21*. 709–729. https://doi.org/10.102/joc.657

Reddy, S.J., 1974: An empirical method for estimating sunshine from total cloud amount. *Solar Energy 15*(4), 281–285. https://doi.org/10.1016/0038-092X(74)90017-6

Robaa, S.M., 2008: Evaluation of Sunshine Duration From Cloud Data in Egypt. *Energy 33*, 785–795. https://doi.org/10.1016/j.energy.2007.12.001

Sanchez-Lorenzo. A.. Calbo. J.. Brunettı. M.. and *Deser. C..* 2009: Dimming / Brightenning over the Iberian Peninsula: Trends in Sunshine Duration and Cloud Cover and Their Relations with Atmospheric Circulation. *J. Geophys. Res. 114*. https://doi.org/10.1029/2208JD011394

Stanghellini, C.A., 1981: Simple Method for Evaluating Sunshine Duration by Cloudiness Observations. *J. Appl. Meteorol. 20*, 320–323. https://doi.org/10.1175/1520-0450(1981)020<0320:ASMFES>2.0.CO;2

USEPA, 1993: Selection Criteria for Mathematical Models Used in Exposure Assessments: Atmospheric Dispersion Models. United States Environmental Protection Agency. Office of Research and Development Washington. DC 20460. EPA/600/8-91/038.

Venkatram, A., 1996: An examination of the Pasquill-Gifford-Turner dispersion scheme. *Atmos. Environ.30*(8). 1283–1290. https://doi.org/10.1016/1352-2310(95)00367-3

Weber, G.R., 1994: On the Seasonal Variation of Local Relationships Between Temperature. Temperature Range. Sunshine and Cloudiness. *Theor. Appl. Climatol. 50*(1–2). 15–22. https://doi.org/10.1007/BF00864898

WMO, 1996: Guide to meteorological instruments and methods of observations. 6th ed. WMO. No.8. Geneva.

Yang, Y.H., Zhao. N., Hao. X.H., and *Li. C.Q.,* 2009a: Decreasing trend of sunshine hours and related driving forces in North China. *Theor. Appl. Climatol. 97*, 91–98. https://doi.org/10.1007/s00704-008-0049-x

Yang, Y., Zhao. N., Hu, Y., and *Zhou, X.,* 2009b: Effect of Wind Speed on Sunshine Hours in Three Cities in Northern China. *Climate Research 39*. 149–157. https://doi.org/10.3354/cr00820

Yildirim, U., Yilmaz, I.O., and *Akinoglu, B.G.,* 2013: Trend analysis of 41 years of sunshine duration data for Turkey. *Turkish J. Engineer. Environ. Sci. 37*. 286–305. https://doi.org/10.3906/muh-1301-11

You, Q., Kang, S., Flugel, W.A., Lorenzo, A.S., Yan, Y., Huang, J., and *Vide, J.M.,* 2010: From Brightening to Dimming in Sunshine Duration over the Eastern and Central Tibetan Plateau (1961–2005). *Theor. Appl. Climatol. 101*, 445–457. https://doi.org/10.1007/s00704-009-0231-9

Zateroglu, M.T., 2021a: Statistical Models For Sunshine Duration Related To Precipitation and Relative Humidity. *Eur. J.f Sci. Technol. 29*. 208–213. https://doi.org/10.31590/ejosat.1022962

Zateroglu, M.T., 2021b: Assessment of the effects of air pollution parameters on sunshine duration in six cities in Turkey. *Fresenius Environ. Bull. 30*(02A). 2251–2269.

Zateroglu, M.T., 2021c: The Role of Climate Factors on Air Pollutants (PM_{10} and SO_2). *Fresenius Environ. Bull. 30*, 12029–12036.

Zateroglu, M.T., 2022: Modelling The Air Quality Index For Bolu. Turkey. *Carpathian J. Earth Environ. Sci. 17,* 119 –130. https://doi.org/10.26471/cjees/2022/017/206