Bulgarian Journal of Meteorology and Hydrology

Национален институт по метеорология и хидрология

National Institute of Meteorology and Hydrology

Bul. J. Meteo & Hydro 30/2 (2026)

Influence of Saharan dust on measured direct solar radiation at ground surface over Bulgaria

Vanya Maneva[1*], Anton Petrov[2]

[1]National Institute of Meteorology and Hydrology, 66 Tsarigradsko Shose blvd., 1784 Sofia, Bulgaria
[2]BULATSA, 1 Brussels blvd., 1540 Sofia, Bulgaria

Abstract: This study investigates the influence of Saharan dust transport on measured direct solar radiation at ground surface over Bulgaria. A daily merged dataset was compiled for the period 1 July 2020 – 30 June 2025 for three stations — Sofia, Sandanski and Ahtopol — combining ground-based pyranometer observations, daily dust aerosol optical depth at 550 nm ($DAOD_{550}$) from the Copernicus Climate Data Store and total cloud cover derived from ERA5 hourly reanalysis data. To reduce the influence of solar geometry and the annual cycle, direct radiation was normalized by daily extraterrestrial radiation on a horizontal surface. The statistical analysis was performed using multiple linear regression models with heteroscedasticity-robust HC3 standard errors, complemented by log-transformed response models and spline-based generalized additive models as sensitivity and robustness checks. The results show that cloud cover is the dominant controlling factor for normalized direct radiation, while $DAOD_{550}$ retains a negative effect after adjustment for cloud cover, station and residual seasonality in the principal linear model. The dust-related attenuation signal is strongest and most robust in the nearly clear-sky subset, indicating that cloud variability masks part of the dust effect in the full dataset. The additional nonlinear analyses further support this interpretation by showing that the $DAOD_{550}$ effect becomes clearer and more structurally consistent when cloud influence is reduced. These findings suggest that Saharan dust episodes are associated with measurable attenuation of direct solar radiation at ground surface over Bulgaria.

Keywords: Saharan dust, solar radiation, DAOD, cloud cover, pyranometer

1. INTRODUCTION

Solar radiation is a fundamental meteorological element and a key component of the atmospheric energy balance. It plays a central role in climate dynamics, land–atmosphere interactions, hydrological processes and solar energy applications. Variations in the amount of direct solar radiation reaching the surface influence not only the local radiation regime, but also the partitioning of sensible and latent heat fluxes, surface temperature behavior and the efficiency of photovoltaic and solar-thermal systems.

During its passage through the atmosphere, solar radiation is attenuated by scattering, absorption and reflection caused by clouds and atmospheric particles (IPCC, 2023). Among atmospheric aerosols, mineral dust transported from desert regions has a particularly strong potential to modify radiative transfer, especially the direct component of solar radiation reaching the surface (Yu et al., 2006; Kok et al., 2023). This effect is physically plausible because mineral dust alters atmospheric transparency through extinction processes that depend on particle concentration, size distribution, composition and vertical distribution. In observational datasets, however, the radiative influence of dust is often difficult to quantify because it is superimposed on other atmospheric controls that are usually stronger and more variable.

Saharan dust outbreaks are a recurrent feature over Southeastern Europe and the Balkans. Over Bulgaria, these events are associated with increased aerosol optical depth, reduced atmospheric transparency and visible impacts on air quality and sky conditions. Recent work has documented both the climatology of Saharan dust over Bulgaria and the characteristics of individual dust episodes over Sofia and the surrounding region (Ilieva et al., 2025; Peshev et al., 2022). From the point of view of surface radiation, such outbreaks are of particular interest because they may attenuate the direct solar beam even under otherwise favorable radiative conditions.

A major methodological difficulty in quantifying the dust effect on solar radiation is the need to separate it from the influence of cloud cover. Cloudiness is usually the dominant controlling factor for the attenuation of direct solar radiation, and therefore dust-related variability may remain partially masked unless it is explicitly accounted for in the statistical framework. Additional complications arise from station-dependent differences, residual seasonality and the fact that aerosol and cloud data are commonly available as gridded products rather than strictly local measurements. For this reason, the observed relationship between dust loading and direct radiation cannot be interpreted adequately on the basis of simple bivariate inspection alone.

The present study addresses this problem by combining daily pyranometer observations from three Bulgarian stations with satellite-derived dust aerosol optical depth at 550 nm ($DAOD_{550}$) and ERA5 total cloud cover. Direct radiation is normalized by daily extraterrestrial radiation on a horizontal surface in order to reduce the influence of solar geometry and seasonal variation in day length. The statistical framework includes a main multiple linear regression model, a nearly clear-sky subset model, and additional sensitivity checks based on log-transformed response models and generalized additive models.

The aim of the study is to assess the relationship between dust loading, represented by $DAOD_{550}$, and measured direct solar radiation over Bulgaria while simultaneously accounting for cloud cover, station-dependent differences and seasonality. For the analyzed period, the variability of $DAOD_{550}$ is interpreted as being dominated to a large extent by Saharan dust transport episodes. The working hypothesis is that increased dust loading is associated with attenuation of direct solar radiation, but that this relationship becomes more clearly detectable when the masking influence of cloud cover is reduced. The results presented here should be regarded as a preliminary quantitative assessment of this relationship, intended to establish a basis for a more comprehensive investigation of the dust–radiation interaction in future work.

2. DATA AND METHODOLOGY

2.1. Study area and ground-based radiation data

The study covers the period from 1 July 2020 to 30 June 2025 and includes three stations in Bulgaria: Sofia (42.653835°N, 23.382884°E), Sandanski (41.55°N, 23.2672°E), and Ahtopol (42.0839°N, 27.9522°E). The station locations are shown in Figure 1.

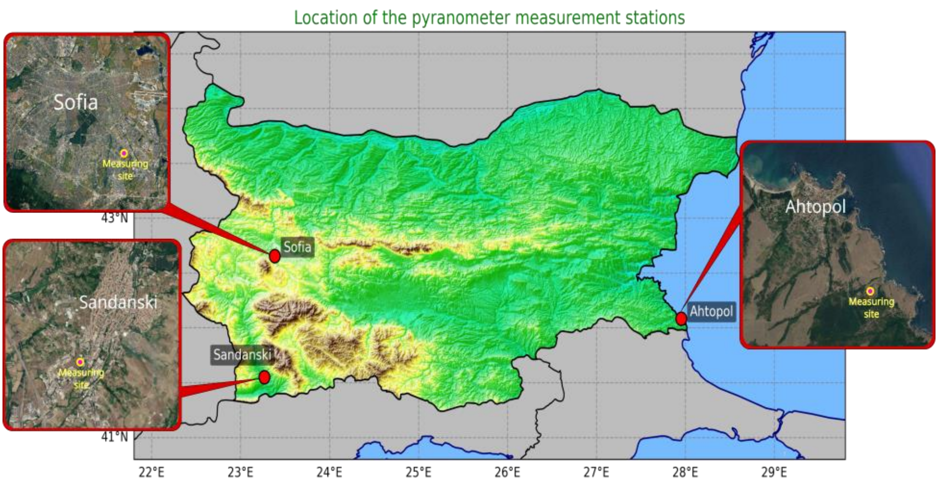


Fig. 1. Location of the study stations in Bulgaria.

Direct solar radiation and sunshine duration data were obtained from pyranometer observations at the respective stations. The original daily files were transformed into tabular format, the calendar fields were combined into a unified datetime index, and missing values flagged in the source files were converted into standard missing values for subsequent processing and merging.

The three stations were selected to represent different geographic settings within Bulgaria, including inland, southwestern and coastal conditions, thus allowing the analysis to capture both common and site-specific features of the relationship between dust load, cloud cover and direct solar radiation.

2.2. Dust aerosol and cloud cover data

Dust aerosol optical depth at 550 nm ($DAOD_{550}$) was obtained from the Copernicus Climate Data Store (dataset name: *Aerosol properties gridded data from 1995 to present derived from satellite observations*). In the study period, the variability of $DAOD_{550}$ was interpreted as being dominated primarily by episodes of Saharan dust transport, and therefore this variable was used as an indicator of dust loading in the atmosphere.

Total cloud cover data were obtained from the Copernicus Climate Data Store (dataset name: *ERA5 hourly data on single levels from 1940 to present*). The hourly total cloud cover field was used to derive a daily cloudiness metric representative of the daylight period.

For each station, both $DAOD_{550}$ and total cloud cover values were assigned from the grid cell containing the station location. This approach ensured spatial consistency between the station position and the corresponding gridded satellite and reanalysis data.

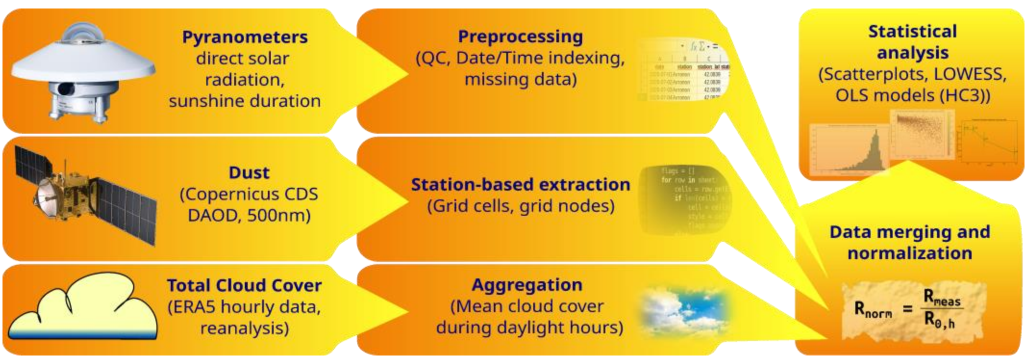


Fig. 2. Schematic workflow of the data processing and analysis, including station observations, satellite-derived $DAOD_{550}$, ERA5 total cloud cover, daily aggregation, data merging and normalization.

The station observations, satellite-derived $DAOD_{550}$ and ERA5 cloud cover data were merged into a common daily dataset by date and station. Hourly ERA5 total cloud cover was aggregated to daily daylight mean values, and the resulting series were combined with the station and satellite data into a unified database. The overall workflow of data extraction, pre-processing, normalization and analysis is shown in Figure 2.

2.3. Data processing and normalization

To preserve the full daily structure of the time series and to facilitate merging between the different data sources, complete daily sequences were retained for the whole study period, while missing files or missing values were left explicitly as missing. The three data sources were then merged by date and station into a common daily dataset.

The dependent variable used in the regression analysis was the normalized direct solar radiation R_{norm}, defined as the ratio between measured daily direct radiation and daily extraterrestrial radiation on a horizontal surface:

$$R_{norm} = \frac{R_{dir}}{R_{0h}} \tag{1}$$

where R_{dir} is the measured daily direct solar radiation (MJ m^{-2} day^{-1}), and R_{0h} is the daily extraterrestrial radiation on a horizontal surface (MJ m^{-2} day^{-1}). This normalization reduces the influence of seasonal changes in solar elevation and day length and allows a more consistent comparison across seasons and stations. The daily extraterrestrial radiation was computed using the standard FAO formulation (Allen et al., 1998):

$$R_{0h} = \left(\frac{24 \cdot 3600}{\pi}\right) G_{sc} d_r [\omega_s sin\phi sin\delta + cos\phi cos\delta sin\omega_s] \tag{2}$$

where G_{sc} is the solar constant (W m^{-2}), d_r is the inverse relative Earth–Sun distance factor (dimensionless), ϕ is the latitude (rad), δ is the solar declination (rad), ω_s is the sunset hour angle (rad), and J is the day of the year.

This normalization was introduced to improve comparability between seasons and stations and to allow the statistical analysis to focus more directly on the effects of dust loading and cloud cover.

2.4. Statistical methods

The statistical analysis was performed using a sequence of linear regression models estimated by ordinary least squares (OLS). Heteroscedasticity-robust HC3 standard errors were used for all models in order to obtain more reliable uncertainty estimates under possible non-constant residual variance (White, 1980; MacKinnon and White, 1985). The general form of the model is:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_i \tag{3}$$

The OLS estimator is given by:

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{4}$$

where $\hat{\beta}$ is the vector of estimated regression coefficients, $\boldsymbol{X}$ is the design matrix, $\boldsymbol{X}^T$ is its transpose, and $\boldsymbol{y}$ is the vector of observed values of the dependent variable.

Residual seasonality was represented by harmonic terms based on the day of the year, J:

$$S_1 = sin\left(\frac{2\pi J}{365.25}\right), S_2 = cos\left(\frac{2\pi J}{365.25}\right) \quad (5)$$

These harmonic terms were included in the regression models in order to account for the remaining smooth annual cycle after radiation normalization.

A broader set of five model specifications was tested during the exploratory stage of the study. However, in the present paper only the two principal models are presented, namely the main multiple linear regression model and the nearly clear-sky model, as these are the most directly relevant to the objective of isolating the dust effect on direct solar radiation. The full comparison of all five model formulations, including intermediate and interaction-based specifications, will be addressed in a separate, more detailed study.

The main multiple linear regression model was defined as:

$$R_{norm,i} = \beta_0 + \beta_1 DAOD_{550,i} + \beta_2 TCC_i + \gamma_{s(i)} + \beta_3 S_1 + \beta_4 S_2 + \varepsilon i \quad (6)$$

where $R_{norm,i}$ is the normalized direct solar radiation, $DAOD_{550,i}$ is the dust aerosol optical depth at 550 nm, TCC_i is the daylight mean total cloud cover, $\gamma_{s(i)}$ denotes the fixed effect for station, and ε_i is the residual term. Station fixed effects ($\gamma_{s(i)}$) were included to account for persistent site-specific differences not explicitly represented by the other predictors. These terms absorb systematic between-station differences related to geographical setting, local relief, coastal or inland influence, and other unobserved characteristics affecting the radiation regime.

In addition, a nearly clear-sky model was estimated for the subset of days satisfying the condition TCC $\leq$ 0.20, i.e. the daily mean total cloud cover is less than 20%. This model has the form:

$$R_{norm,i} = \beta_0 + \beta_1 DAOD_{550,i} + \gamma_{s(i)} + \beta_2 S_1 + \beta_3 S_2 + \varepsilon i \quad (7)$$

The clear-sky subset was introduced in order to assess the attenuation effect of dust on direct solar radiation under conditions of minimal cloud influence.

As an additional sensitivity check, the regression analysis was repeated using the natural logarithm of normalized direct solar radiation, $ln(R_{norm})$, as the response variable instead of R_{norm} itself. This transformation was introduced because R_{norm} is strictly non-negative and may exhibit asymmetric variability. The same predictor structure was retained as in the original multiple linear regression framework, including $DAOD_{550}$, daylight mean total cloud cover, station fixed effects and harmonic seasonal terms:

$$ln(R_{norm,i}) = \beta_0 + \beta_1 DAOD_{550,i} + \beta_2 TCC_i + \gamma_{s(i)} + \beta_3 S_1 + \beta_4 S_2 + \varepsilon i \quad (8)$$

for the full-sample log-transformed model, and

$$ln(R_{norm,i}) = \beta_0 + \beta_1 DAOD_{550,i} + \gamma_{s(i)} + \beta_2 S_1 + \beta_3 S_2 + \varepsilon i \quad (9)$$

for the nearly clear-sky log-transformed model.

In addition, a nonlinear robustness check was performed using generalized additive models (GAMs) with spline-based smooth functions. In these models, the dependence of the response variable on selected predictors is represented by smooth functions rather than by fixed linear slopes, allowing the data to determine whether the underlying relationships are linear or nonlinear:

$$ln(R_{norm,i}) = \beta_0 + f_1(DAOD_{550,i}) + f_2(TCC_i) + \gamma_{s(i)} + \beta_1 S_1 + \beta_2 S_2 + \varepsilon i (10)$$

$$ln(R_{norm,i}) = \beta_0 + f_1(DAOD_{550,i}) + \gamma_{s(i)} + \beta_1 S_1 + \beta_2 S_2 + \varepsilon i \quad (11)$$

where $ln(R_{norm,i})$ is the natural logarithm of normalized direct solar radiation for observation i, β_0 is the intercept, β_1, β_2, β_3 and β_4 are regression coefficients, $\gamma_{s(i)}$ denotes the fixed effect for station, ε_i is the residual term, and f_1 and f_2 are smooth spline functions estimated from the data. These additional models were used as sensitivity and robustness checks in order to determine whether the inferred dust effect remains stable under more flexible model structures.

3. RESULTS AND DISCUSSION

3.1. General characteristics of the dataset

The merged daily dataset combines pyranometer observations, satellite-derived $DAOD_{550}$ and ERA5 total cloud cover for the period from 1 July 2020 to 30 June 2025 for the three study stations: Sofia, Sandanski and Ahtopol. After quality control, preprocessing and temporal matching, the dataset provided sufficient coverage for regression-based statistical analysis, despite the presence of some missing values in the station, satellite and reanalysis records.

The exploratory analysis confirmed that direct solar radiation is strongly related to sunshine duration. The scatterplot of daily direct radiation versus sunshine duration shows a pronounced positive but nonlinear relationship (Figure 3). LOWESS smoothing indicates a change in curvature at sunshine duration values of approximately 7 h, suggesting that the increase in direct solar radiation is not uniform across the entire range of sunshine duration. This result is physically consistent with the combined effects of solar geometry, cloud conditions and atmospheric transparency, and supports the need for a multiple regression framework when assessing the specific role of dust loading, although this descriptive relationship is distinct from the regression models used to estimate the dust effect.

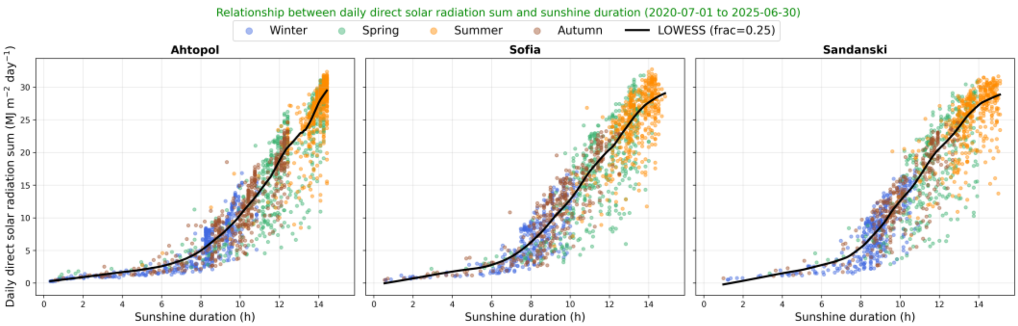


Fig. 3. Scatterplots showing the relationship between daily direct solar radiation sum (MJ m-2 day-1) and sunshine duration (h) at Ahtopol, Sofia and Sandanski during the period 1 July 2020 – 30 June 2025. The black curve represents LOWESS smoothing (frac = 0.25).

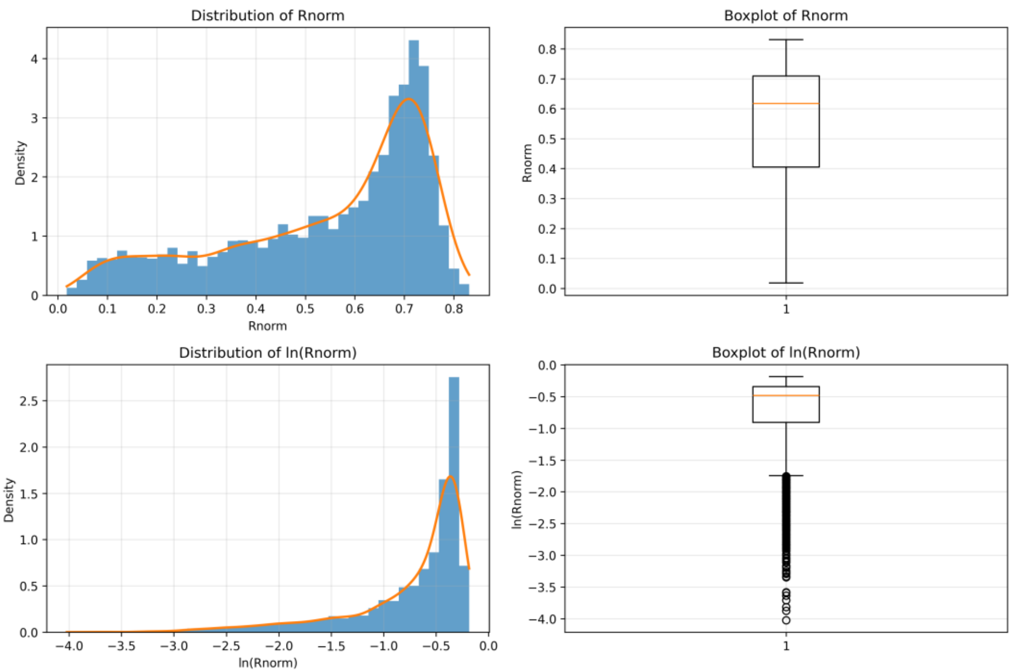


Fig. 4. Histogram and boxplot representations of normalized direct solar radiation (R_{norm}) and its logarithmic transform, $ln(R_{norm})$, for the merged daily dataset

The empirical distributions of R_{norm} and $ln(R_{norm})$ are shown in Figure 4. The logarithmic transformation does not eliminate the asymmetry of the response variable, but it changes the scale and compresses the upper range of the distribution. For this reason, $ln(R_{norm})$ was not treated as a normalization device, but rather as an alternative specification used to test the robustness of the estimated dust effect.

At the same time, the raw relationship between $DAOD_{550}$ and normalized direct radiation appears visually weak in simple scatterplots. This is not unexpected. The effect of dust is superimposed on much stronger sources of variability, most notably cloud cover, as well as site-specific differences and residual seasonality. Therefore, the dust signal is expected to emerge more clearly only after the major confounding factors are explicitly controlled for in the regression analysis.

3.2. Main multiple linear regression model

The principal statistical result of the study is provided by the main multiple linear regression model, which includes $DAOD_{550}$, daylight mean total cloud cover, station fixed effects and harmonic seasonal terms. In this model, $DAOD_{550}$ retains a negative and statistically significant coefficient ($\beta = -0.0804$, $p = 0.002$), indicating that, all else being equal, higher dust loading is associated with lower normalized direct solar radiation. At the same time, cloud cover has a strong negative and highly significant effect ($\beta = -0.2933$, $p < 0.001$), confirming its dominant role in controlling the direct radiation regime (Table 1).

The coefficient of determination for the model is $R^2 = 0.364$, with $n = 2944$, indicating that a substantial part of the variability in normalized direct radiation is explained by the combined influence of cloud cover, dust loading, station-related differences and residual seasonal structure. The station fixed effects are also statistically significant, which suggests that the three stations differ systematically in their baseline radiation regime even after accounting for dust, cloud cover and seasonality. This is consistent with the different geographic settings of the stations, including inland, southwestern and coastal environments.

The harmonic seasonal terms are likewise significant, indicating that some smooth annual variability remains present even after radiation normalization. This remaining seasonality may reflect additional atmospheric and environmental controls not explicitly represented in the model, such as changes in background aerosol conditions, humidity regime, or other seasonal differences in atmospheric transmission.

Type II ANOVA results for the main multiple linear regression model are summarized in Table 2 and provide a supplementary descriptive assessment of the relative contribution of the predictors. These results are consistent with the interpretation that cloud cover represents the dominant contribution in the full-sample linear formulation, while $DAOD_{550}$ retains an independent negative association with normalized direct solar radiation.

Overall, the main multiple linear regression model shows that cloud cover is the strongest controlling factor, but also that the dust effect does not disappear once cloudiness, location and seasonality are taken into account. This is a key result of the present study.

Table 1. Coefficients of the main multiple linear regression model for normalized direct solar radiation as a function of $DAOD_{550}$, daylight mean total cloud cover, station effects and harmonic seasonal terms. *Note*: Station fixed effects were included in the model but are not shown. Ahtopol was used as the reference station.

Variable	Coefficient (β)	Robust SE (HC3)	p-value
Intercept	0.7355	0.0040	<0.001
$DAOD_{550}$	-0.0804	0.0260	0.002
Daylight mean total cloud cover	-0.2933	0.0090	<0.001
sin(J)	0.0164	0.0030	<0.001
cos(J)	-0.0201	0.0030	<0.001
Model summary: n = 2944, R^2 = 0.364.			

Table 2. Type II ANOVA results for the main multiple linear regression model. The table provides a supplementary descriptive assessment of the relative contribution of $DAOD_{550}$, daylight mean total cloud cover, station effects and harmonic seasonal terms to the explained variability in normalized direct solar radiation.

Term	Sum of squares	df	F-value	p-value
$DAOD_{550}$	0.0571	1	5.60	0.018
Daylight mean total cloud cover	24.801	1	2433.23	<0.001
C(station)	0.439	2	21.54	<0.001
sin(J)	0.405	1	39.75	<0.001
cos(J)	0.974	1	95.59	<0.001
Residual	33.086	3246	—	—

3.3. Residual diagnostics of the principal OLS models

For the main multiple linear regression model (Figure 5), the residual diagnostic plots indicate that the fitted linear structure captures the central tendency of the data reasonably well, although it does not fully account for the entire variability of normalized direct solar radiation. The residuals are more tightly clustered around zero at higher fitted values and become more dispersed toward the lower fitted range, suggesting some non-uniform residual variance. At the same time, the residual distribution remains approximately symmetric and centered near zero, while the Q-Q plot shows good agreement with the normal reference line in the central part of the distribution, with departures mainly in the tails.

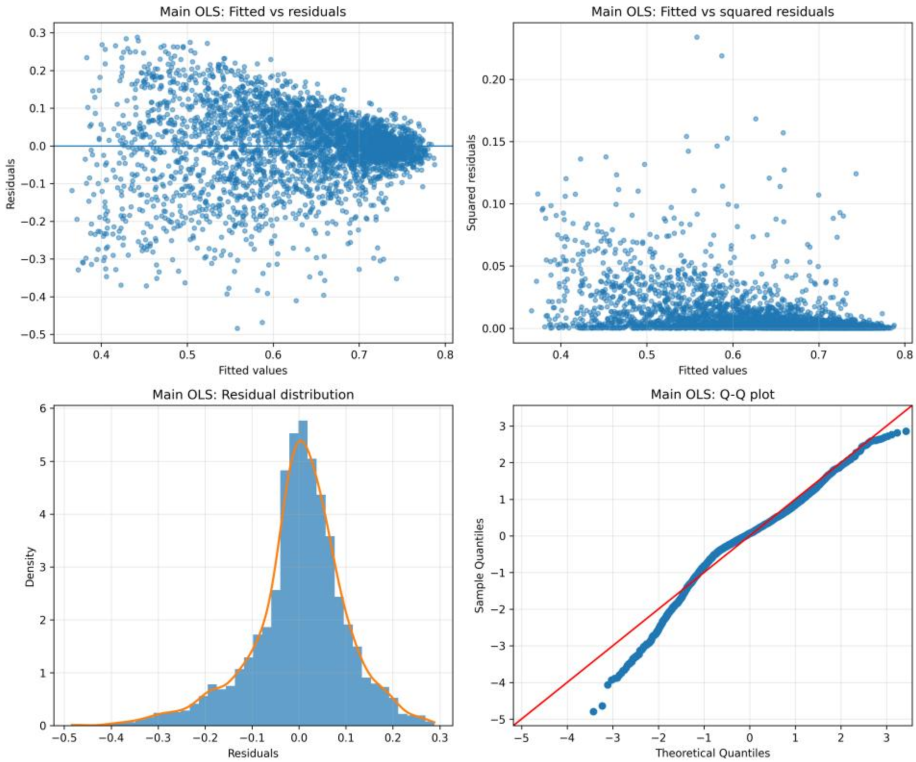


Fig. 5. Residual diagnostic plots for the main multiple linear regression model: fitted values versus residuals, fitted values versus squared residuals, histogram of residuals, and normal Q-Q plot.

For the nearly clear-sky OLS model (Figure 6), the residual diagnostic plots show that the residuals remain broadly concentrated around zero within the main fitted range, although a slight extension toward more negative residual values is apparent. The residual distribution is moderately asymmetric, with a longer lower tail, while the Q-Q plot indicates reasonable agreement with the normal reference line in the central part of the distribution and clearer departures in the lower tail. Overall, these diagnostics suggest that the nearly clear-sky linear model captures the main directional structure of the relationship, although some non-random variability remains in the residuals.

The residual ACF and PACF plots for the main multiple linear regression model (Figure 7) suggest the presence of weak short-lag autocorrelation, most clearly at lag 1. Beyond the first few lags, the autocorrelation values remain small, indicating limited residual persistence. These diagnostics imply that the residuals are not fully independent in time, but that the remaining dependence is moderate and consistent with the observational nature of daily atmospheric radiation data.

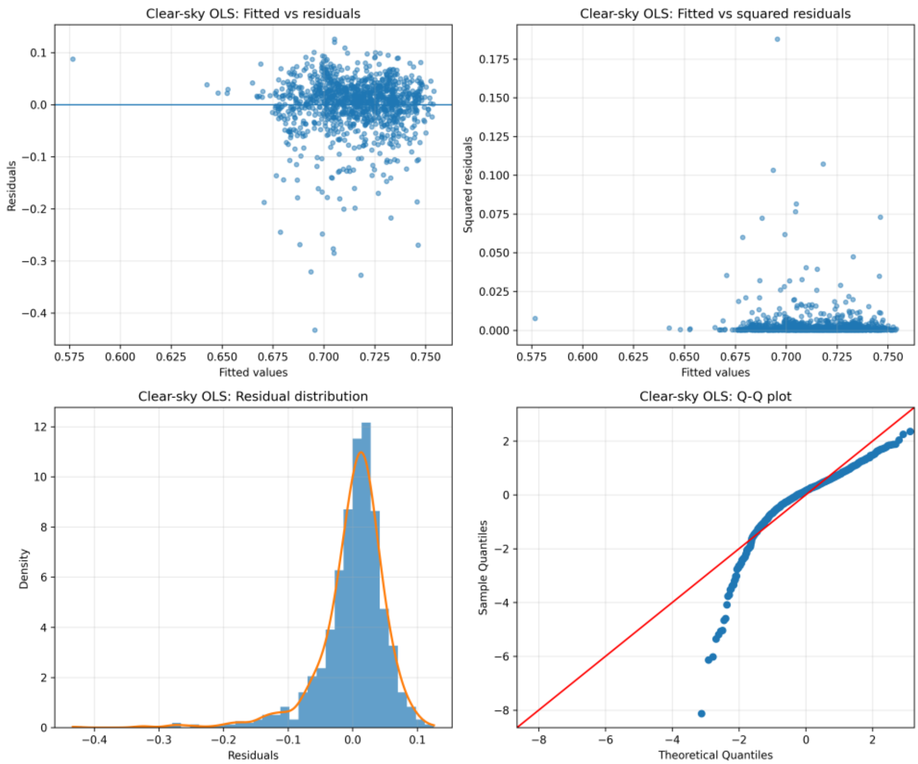


Fig. 6. Residual diagnostic plots for the nearly clear-sky OLS model: fitted values versus residuals, fitted values versus squared residuals, histogram of residuals, and normal Q-Q plot.

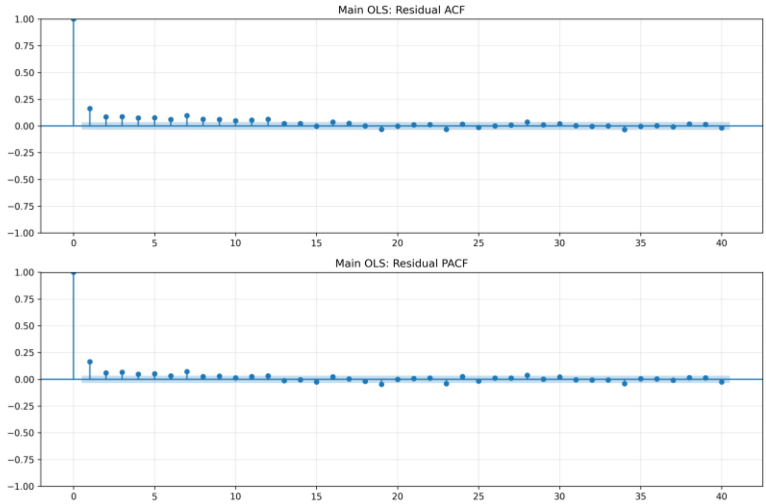


Fig. 7. Autocorrelation function (ACF) and partial autocorrelation function (PACF) of residuals from the main multiple linear regression model.

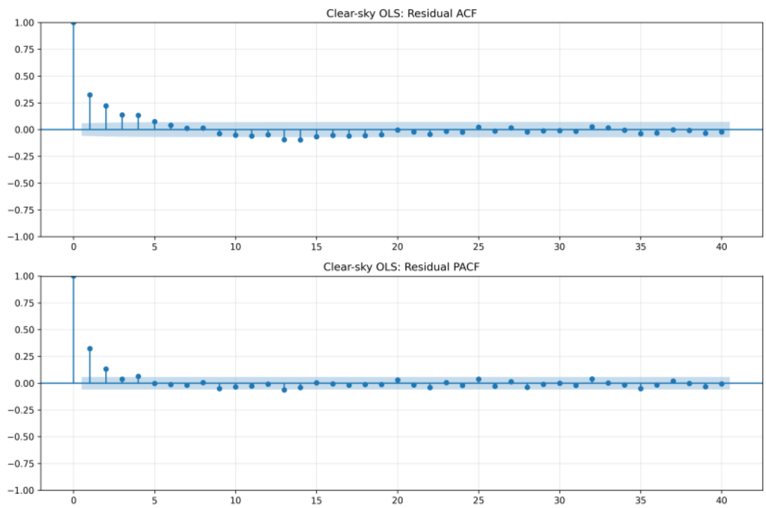


Fig.8. Autocorrelation function (ACF) and partial autocorrelation function (PACF) of residuals from the nearly clear-sky OLS model.

For the nearly clear-sky OLS model, the residual ACF and PACF plots indicate the presence of short-lag autocorrelation, most clearly at lag 1 and more weakly at lag 2. Beyond the first few lags, the residual dependence decreases rapidly and remains limited. These diagnostics suggest that the residuals are not fully independent in time, but that the remaining persistence is concentrated at short lags and is consistent with the temporal continuity of atmospheric conditions under nearly cloud-free regimes.

3.4. Nearly clear-sky subset

To isolate the dust effect under conditions of minimal cloud influence, a second model was estimated using only days with daylight mean total cloud cover not exceeding 0.20 (20%). This nearly clear-sky subset allows a more direct assessment of the attenuation effect of dust on direct solar radiation, with cloud-related variability substantially reduced.

Under these conditions, the coefficient of $DAOD_{550}$ becomes more negative in absolute value and remains highly significant ($\beta = -0.1935$, $p < 0.001$), with $R^2 = 0.117$ and $n = 1118$ (Table 3). The same tendency is evident in the clear-sky binned representation shown in Fig. 9, where mean normalized direct solar radiation decreases with increasing $DAOD_{550}$. Although the explained variance is lower than in the main multiple linear regression model, this is expected because the subset excludes a major source of variability, namely cloud cover, and focuses on a more specific radiative regime.

The stronger negative coefficient in the nearly clear-sky subset suggests that part of the dust signal in the full dataset is masked by cloud variability. Once cloud influence is minimized, the attenuation effect of dust on direct solar radiation becomes more clearly detectable. This result is physically plausible, since under nearly cloud-free conditions the reduction in direct radiation is more directly linked to atmospheric turbidity and aerosol loading.

The persistence of a statistically significant negative $DAOD_{550}$ coefficient in the nearly clear-sky subset provides strong support for the interpretation that Saharan dust contributes independently to the attenuation of direct solar radiation at the surface. In this sense, the clear-sky analysis does not merely confirm the result of the main model, but strengthens it.

Table 3. Coefficients of the nearly clear-sky OLS model for normalized direct solar radiation estimated for days with daylight mean total cloud cover not exceeding 0.20. Note: Station fixed effects were included in the model but are not shown. Ahtopol was used as the reference station.

Variable	Coefficient (β)	Robust SE (HC3)	p-value
Intercept	0.7180	0.0040	<0.001
$DAOD_{550}$	-0.1935	0.0360	<0.001
sin(J)	0.0185	0.0020	<0.001
cos(J)	-0.0183	0.0030	<0.001
Subset condition: daylight mean total cloud cover ≤ 0.20.			

Variable	Coefficient (β)	Robust SE (HC3)	p-value
Model summary: n = 1118, $R^2 = 0.117$.			

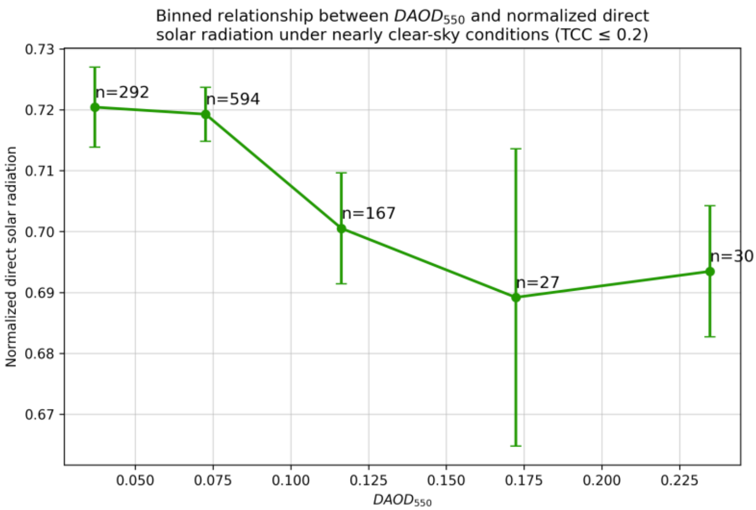


Fig. 9. Binned relationship between $DAOD_{550}$ and normalized direct solar radiation under nearly clear-sky conditions (daylight mean total cloud cover ≤ 0.20). Points represent mean values within predefined $DAOD_{550}$ intervals, and vertical bars denote 95% confidence intervals.

3.5. Sensitivity analysis with log-transformed and nonlinear models

As an additional sensitivity test, the regression analysis was repeated using $ln(R_{norm})$ as the response variable instead of Rnorm. The logarithmic transformation was not introduced as a normalization device. Rather, it was used as an alternative modeling scale that changes the scale of the response, compresses the upper range of Rnorm, and provides an additional sensitivity check for the estimated dust effect. The results of the additional sensitivity and robustness checks are summarized in Table 4.

In the full-sample specification, the coefficient of $DAOD_{550}$ remained negative, but it was no longer statistically significant. This indicates that, when the full variability associated with cloudiness is retained in the dataset, the estimated dust effect is more sensitive to the statistical specification of the model. In contrast, for the nearly clear-sky subset the coefficient of $DAOD_{550}$ remained negative and highly significant in the log-transformed model. This supports the interpretation that the attenuation signal associated with dust loading is more stable under conditions of minimal cloud influence. Thus, while the full-sample dust effect appears partially dependent on the exact response specification, the nearly clear-sky result proves more robust.

Table 4. Comparison of the estimated dust effect under alternative linear, log-transformed and spline-based model specifications. The table summarizes the direction, statistical stability and relative robustness of the $DAOD_{550}$ effect across the tested formulations.

Model specification	Response variable	Sample	Dust effect representation	Inference	Model performance
OLS multiple linear model	R_{norm}	Full sample	-0.0804	$p = 0.002$	$R^2 = 0.364$
OLS multiple linear model	R_{norm}	Nearly clear-sky subset	-0.1935	$p < 0.001$	$R^2 = 0.117$
Log-transformed OLS model	$\ln(R_{norm})$	Full sample	-0.0720	$p = 0.252$	$R^2 = 0.403$
Log-transformed OLS model	$\ln(R_{norm})$	Nearly clear-sky subset	-0.2761	$p < 0.001$	$R^2 = 0.101$
GAM / spline model	$\ln(R_{norm})$	Full sample	nonlinear smooth term for $DAOD_{550}$	modest additional contribution	pseudo-R^2 = 0.448
GAM / spline model	$\ln(R_{norm})$	Nearly clear-sky subset	nonlinear smooth term for $DAOD_{550}$	more pronounced contribution	pseudo-R^2 = 0.106

Note: The full-sample GAM was estimated with smooth terms for $DAOD_{550}$ and daylight mean total cloud cover, while the nearly clear-sky GAM included a smooth term for $DAOD_{550}$ only. Station effects and harmonic seasonal terms were retained as linear components in both cases.

Type II ANOVA results for the log-transformed full-sample model are summarized in Table 5. These results support the view that the full-sample dust effect remains directionally consistent after logarithmic transformation of the response, even though its statistical significance becomes weaker.

Table 5. Type II ANOVA results for the log-transformed full-sample model. The table provides a supplementary descriptive assessment of the relative contribution of $DAOD_{550}$, daylight mean total cloud cover, station effects and harmonic seasonal terms after logarithmic transformation of the response variable.

Term	Sum of squares	df	F-value	p-value
C(station)	2.4699	2	21.19	<0.001
$DAOD_{550}$	0.0692	1	1.19	0.276

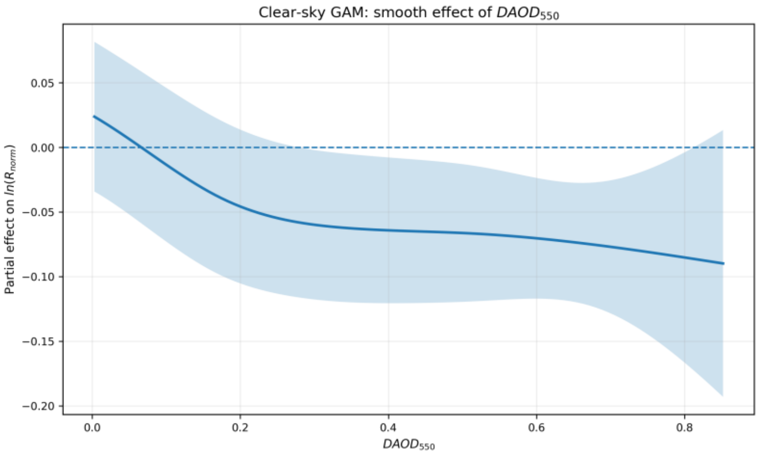
Term	Sum of squares	df	F-value	p-value
Daylight mean total cloud cover	101.8459	1	1747.38	<0.001
sin(J)	1.6440	1	28.21	<0.001
cos(J)	6.9830	1	119.81	<0.001
Residual	189.1928	3246	—	—

A further robustness check was performed using spline-based generalized additive models. The estimated spline smooths are shown in Figure10. In the full sample, the smooth effect of daylight mean total cloud cover is strongly negative and decreases monotonically across the observed range, indicating that cloudiness remains the dominant control on ln(Rnorm). The decline becomes steeper at higher cloud-cover values, particularly beyond approximately 0.8, which suggests stronger attenuation of direct solar radiation under very cloudy conditions.

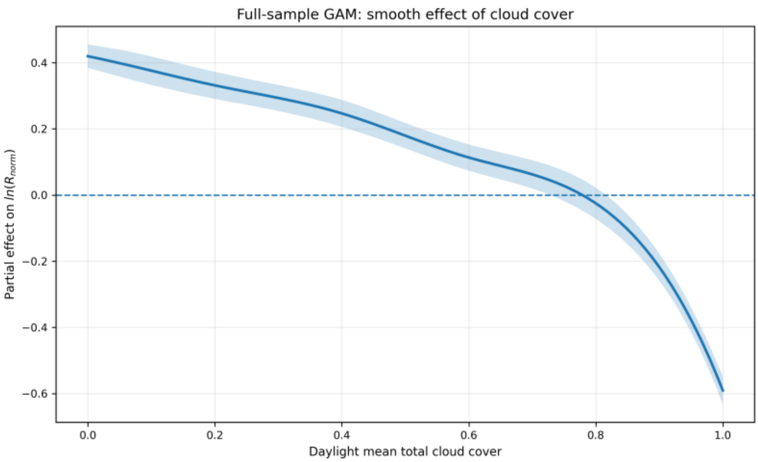
By contrast, the full-sample smooth effect of $DAOD_{550}$ remains negative throughout the analyzed range but is not monotonic. The partial effect becomes more negative up to intermediate-to-high dust loading and then weakens again at the upper end of the $DAOD_{550}$ range. This indicates that, in the full dataset, the dust effect is weaker and more structurally sensitive than the cloud-cover effect.

For the nearly clear-sky subset, the smooth effect of $DAOD_{550}$ decreases monotonically, with progressively lower partial effects on $ln(R_{norm})$ as dust loading increases. This result is physically consistent with attenuation of direct solar radiation by atmospheric dust under conditions of minimal cloud interference and supports the interpretation that the dust signal becomes more clearly expressed once cloud variability is reduced.

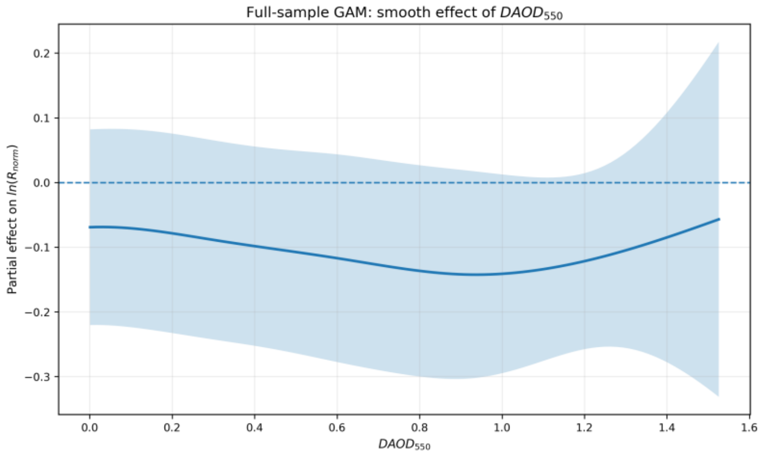
Taken together, the additional sensitivity analyses indicate that the negative association between $DAOD_{550}$ and normalized direct solar radiation is most robust in the nearly clear-sky subset, whereas in the full dataset the dust signal is more strongly affected by model specification and by the dominant contribution of cloud cover.


Clear-sky GAM: smooth effect of $DAOD_{550}$
Partial effect on $ln(R_{norm})$
0.05
0.00
−0.05
−0.10
−0.15
−0.20
0.0
0.2
0.4
0.6
0.8
$DAOD_{550}$

(a)


Full-sample GAM: smooth effect of cloud cover
Partial effect on $ln(R_{norm})$
0.4
0.2
0.0
−0.2
−0.4
−0.6
0.0
0.2
0.4
0.6
0.8
1.0
Daylight mean total cloud cover

(b)



(c)

Fig. 10. Estimated spline smooths from the GAM formulations: (a) smooth effect of $DAOD_{550}$ in the nearly clear-sky subset, (b) smooth effect of daylight mean total cloud cover in the full sample, and (c) smooth effect of $DAOD_{550}$ in the full sample. Shaded bands denote 95% confidence intervals.

3.6. Interpretation of the dust effect

Taken together, the original linear models and the additional sensitivity analyses provide a consistent overall interpretation. In the full dataset, cloud cover remains the dominant control on normalized direct solar radiation, while the estimated dust effect is more sensitive to model specification. In contrast, in the nearly clear-sky subset the negative effect of $DAOD_{550}$ remains detectable under the original linear, log-transformed and spline-based formulations. This pattern supports the interpretation that Saharan dust contributes independently to the attenuation of direct solar radiation at the surface, but that its statistical expression is most robust under conditions of minimal cloud influence.

4. CONCLUSIONS

This study examined the influence of Saharan dust on measured direct solar radiation at the surface in Bulgaria using a merged daily dataset for the period 1 July 2020 – 30 June 2025 for the stations Sofia, Sandanski and Ahtopol. The analysis combined pyranometer observations, satellite-derived $DAOD_{550}$ and ERA5 total cloud cover, with direct solar radiation normalized by daily extraterrestrial radiation on a horizontal surface in order to reduce the influence of solar geometry and day length.

The results show that cloud cover is the dominant control on normalized direct solar radiation, which is physically expected. Nevertheless, the dust-related signal does not disappear after accounting for cloudiness, station-related differences and residual seasonality. In the main multiple linear regression model, $DAOD_{550}$ retains a negative and statistically significant coefficient, indicating that increased dust loading is associated with reduced direct solar radiation at the surface. In the nearly clear-sky subset, the corresponding coefficient becomes stronger in absolute value, which suggests that part of the dust signal in the full dataset is masked by cloud variability.

The additional sensitivity analyses refine this interpretation. In the full dataset, the sign of the $DAOD_{550}$ effect remains negative, but its statistical significance becomes more sensitive to the specification of the response variable. In contrast, in the nearly clear-sky subset the negative $DAOD_{550}$ effect remains robust under both log-transformed and spline-based formulations. The GAM smooths further indicate that cloud cover exerts a strong monotonic attenuation effect in the full sample, whereas the dust-related attenuation signal becomes more clearly expressed and structurally more consistent when cloud influence is minimized.

Overall, the study supports the interpretation that Saharan dust transport contributes independently to measurable attenuation of direct solar radiation over Bulgaria, but that its statistical expression is clearest under conditions of minimal cloud interference. The results should be regarded as preliminary and the present analysis as a focused observational-statistical assessment rather than a complete treatment of the problem. Further work will extend and refine the proposed framework through the incorporation of additional predictors, alternative time-series formulations and broader spatial coverage, with the aim of improving the robustness, accuracy and physical interpretation of the estimated dust–radiation relationship.

5. ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the National Institute of Meteorology and Hydrology (NIMH) for providing the opportunity to present and publish the results of this study. They also gratefully acknowledge the colleagues whose support, assistance and professional guidance contributed to the realization of this work.

REFERENCES

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M. (1998) Crop evapotranspiration: Guidelines for computing crop water requirements, FAO Irrigation and Drainage Paper 56, Food and Agriculture Organization of the United Nations, Rome.

Copernicus Climate Change Service, Climate Data Store, (2019): Aerosol properties gridded data from 1995 to present derived from satellite observations. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.239d815c (Accessed on 09-Mar-2026).

Copernicus Climate Change Service, Climate Data Store, (2023): ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.adbb2d47 (Accessed on 09-Mar-2026).

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D. and Thépaut, J.-N. (2020) The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146(730), 1999-2049.

Ilieva, R., Stoev, K. and Guerova, G. (2025) Climatology and circulation classification of Saharan dust over Bulgaria, Journal of Atmospheric and Solar-Terrestrial Physics, 267, 106403.

IPCC (2023) *Climate Change 2023: Synthesis Report*. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, IPCC, Geneva, 184 pp.

Kok, J. F., Adebiyi, A. A., Albani, S., Balkanski, Y., Checa-Garcia, R., Chin, M., Colarco, P. R., Hamilton, D. S., Ito, A., Klose, M., Mahowald, N. M., Miller, R. L., Obiso, V., García-Pando, C. P., Rocha-Lima, A., and co-authors (2023) Mineral dust aerosol impacts on global climate and climate change, *Nature Reviews Earth & Environment*, 4, 71–86.

MacKinnon, J. G. and White, H. (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties, Journal of Econometrics, 29(3), 305-325.

Peshev, Z., Deleva, A., Vulkova, L. and Dreischuh, T. (2022) Large-scale Saharan dust episode in April 2019: Study of desert aerosol loads over Sofia, Bulgaria, using remote sensing, in situ, and modeling resources, Atmosphere, 13(6), 981.

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica, 48(4), 817-838.

Yu, H., Kaufman, Y. J., Chin, M., Feingold, G., Remer, L. A., Anderson, T. L., Balkanski, Y., Bellouin, N., Boucher, O., Christopher, S., DeCola, P., Kahn, R., Koch, D., Loeb, N., Reddy, M. S., Schulz, M., Takemura, T. and Zhou, M. (2006) A review of measurement-based assessments of the aerosol direct radiative effect and forcing, Atmospheric Chemistry and Physics, 6, 613–666.