A daily sunshine duration (SD) dataset in China from Himawari AHI imagery (2016-2023)

Zhanhao Zhang[1,2], Shibo Fang[1], Jiahao Han[1]

[1]State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China

[2]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China

Correspondence to: Shibo Fang (sbfang0110@163.com)

Abstract. Monitoring global radiation resources relies on sunshine duration (SD) as a significant indication, but there is a scarcity of research that have examined high-resolution SD data. This study established a daily 5-km SD dataset in China from 2016 to 2023 using Himawari's Advanced Himawari Imager (AHI) Level 3 shortwave radiation fitted with the Ångström-Prescott model based on time series. We used ground-measured SD at 2380 Chinese Meteorological Administration (CMA) stations to verify the accuracy of SD dataset. The results of the testing set indicated that the average correlation coefficient (R) between the SD from estimation and the ground-measurement is 0.88. We investigated the effects of wind speed, vapor pressure (VAP), precipitation and aerosol optical depth (AOD) on the estimated performance of SD, and the results showed that temperature had the greatest effect on SD estimation. We also found that both too low AOD and too high wind speed ~~also~~ affected the SD estimation on the average annual scale. This high-resolution SD data can provide important support for accurate radiation resource assessment in China. The SD dataset is freely accessible at https://doi.org/10.57760/sciencedb.10276 (Zhang et al., 2024).

1. Introduction

Solar radiation is a major driver of photosynthesis and evapotranspiration, plays an indispensable role in regulating temperature and supporting agricultural production, and ~~also~~ has effects on photovoltaic power generation, making it critical to the Earth's ecosystem and to productive human life (Yu et al., 2022; Feng et al., 2021). ~~Because of the high cost of using~~The solar radiation measured by radiation observatory can accurately predict solar radiation potential and ~~maintaining ground radiation~~

~~measuring instruments, which are fewer~~participate in climate change and agricultural production model. Nonetheless, the existing radiation data in China is not validated through terrestrial observations due to the limited number of less than 200 stations in mainland China ~~and unevenly distributed over short time spans, there are lacking or unavailable long-term solar radiation data in most areas~~ (Liang et al., 2006; Zhang et al., 2015~~). Therefore, it is~~) for the expensive upkeep of terrestrial radiation measuring devices, making precise tracking of high spatiotemporal solar radiation over time difficult ~~to accurately verify the estimated long-term and high-precision solar radiation indicators with information provided by ground radiation measurement compared with conventional meteorological measurement~~ (Zhang et al., 2017; Chukwujindu et al., 2017).

Sunshine duration (SD) is a readily available and cost-effective indicator for monitoring the ~~condition of~~ global radiation resources, and the variability of which is determined by a combination of regional factors as well as the solar constant, cloud cover, water vapor, and atmospheric pollutants. SD is a key ~~element~~parameter of solar ~~radiation that affects many areas of human life, such as tourism activities, planning~~ power ~~plants~~potential forecasting (Liu et al., 2022; Qin et al., 2023), climate change assessment and agricultural production (Ghanghermeh et al., ~~2022~~2022), in addition, some researchers have found that changes in SD also affect the probability of human diseases (Chang et al., 2022; Gu et al., 2019). The SD measured from conventional meteorological observation has the advantages of long observation time, good continuity, high spatial density and ~~high~~ reliability, ~~and~~which is considered the best alternative to solar radiation (Xia, 2010). Accurate inversion of SD is therefore an important reference for agricultural production, solar resource utilization and global climate change analysis. The Ångström-Prescott model (~~Angstrom, 2007~~Ångström, 1924) is the dominant and most widely used model based on SD and solar radiation. The quadratic and cubic forms of the Ångström-Prescott model have been improved ~~by researchers~~ and applied to different meteorological conditions (Rietveld, 1978; Bahel et al., 1987; Chen et al., 2004; Wu et al., 2007; Liu et al., 2012~~), and other forms of the model (e.g., logarithmic and exponential) have also been proposed and applied worldwide to estimated SD or solar radiation (~~; Ampratwum et al., 1999; Elagib et al 2000).

Studies on SD ~~estimation~~ have mostly been based on limited ground stations (Vivar et al., 2014; Fan et al., 2018; Yao et al., 2018), while SD is affected by atmospheric conditions, and it is difficult for a single station to represent this over a large area, so there is a great need for a high-resolution SD data based on satellite remote sensing for studies on solar radiation. ~~The~~Currently, geostationary and polarorbiting satellite data are widely used for high spatiotemporal resolution ground information tracking, and the Advanced Himawari Imager (AHI) instrument, carried on board the new generation of geostationary satellites -Himawari-8 and 9~~ satellite, can be considered to observe and invert solar~~, has been widely used for the estimating radiation ~~indicator.~~indicators different time scales (Damiani et al, 2018; Hou et al., 2020; Letu et al., 2020; Tana et al., 2023). However, ~~despite~~there are always biases in the ~~release of a short-wave~~AHI radiation ~~product by Himawari, the product does not adequately consider the effect of aerosols on solar~~data and those inverted radiation ~~under clear sky, nor does it consider~~indicators due to less ground measured stations for validation and the ~~effect of different~~ susceptibility of remote sensing data to cloud ~~phases on~~and aerosols, while SD reflects both solar radiation ~~under cloudy~~and cloud conditions~~,~~ and thus ~~the accuracy of solar~~is well suited for inversion using remote sensing radiation ~~estimated under heavy aerosol-polluted backgrounds or cloudy sky conditions is limited.~~data, we can take advantage of the high spatiotemporal resolution of AHI to estimate SD.

In this study, we generate a daily SD dataset in China at a spatial resolution of 5-km using Himawari AHI L3 shortwave radiation data from 2015 to 2023 fitted with Ångström-Prescott model at different days of year (DOY). We validated and assessed the accuracy of the daily SD data by the ground-measured SD and other meteorological data (Wind speed, vapor pressure (VAP) and precipitation) at 2380 Chinese Meteorological Administration (CMA) stations, as well as the aerosol optical depth (AOD) from MODIS.

2. Data and method

2.1 Remote sensing data

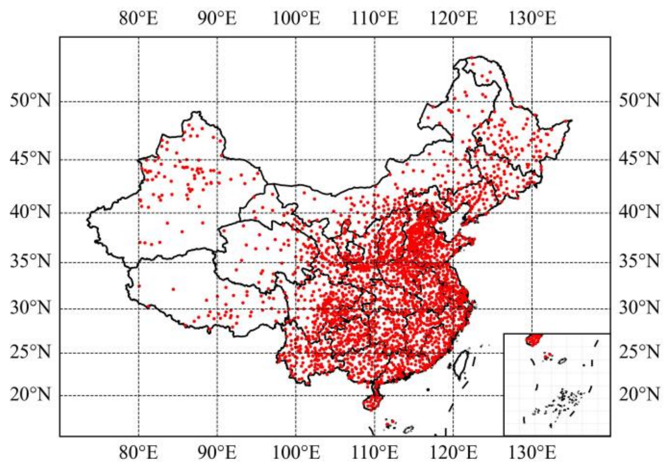
The geostationary meteorological satellites, Himawari, was launched on 7 October 2014 from the Japan Meteorological Agency (JMA) in Tane Ashima, Japan, with its hypocenter located at 0.0◦N and 140.7◦E, approximately 35,800 km above the land surface. In comparison with other geostationary satellites, Himawari AHI exhibits superior temporal and spatial resolution, reflection band sensitivity and accuracy (Zhang et al., 2016). The AHI from Himawari-8 and 9 has 16 spectral channels covering the visible to infrared range, with wavelengths ranging from 0.47 μm to 13.3 μm, providing a wealth of spectral information (Bessho et al., 2016; Kim et al., 2018; Yu et al., 2019). The temporal and spatial resolution of the land surface products provided by Himawari AHI is 10 minutes and 5 km respectively, which is important for understanding the spatiotemporal variations on short time scales (Sawada et al., 2019).

In this study, the Himawari AHI level 3 hourly shortwave radiation (5 km resolution) data from 1 January 2016 to 31 December 2023 was used for SD dataset construction, which calculated by ~~considering~~plane-parallel theory and considered the top of atmosphere (TOA) radiation by difference between the 300-3000 nm ~~incident~~ solar ~~flux absorbed~~shortwave band and reflected solar radiation by the atmosphere ~~and the solar flux reflected back to space by the atmosphere and the~~ /land surface (Frouin et al., 2007). ~~For imagery with a missing~~ This approach assumes that the effects of clouds and clear atmosphere can be decoupled, which proved to be effective (Dedieu et al., 1987; Frouin and Rachel, 1995). In the event of a one-hour interval ~~of one hour in a day~~being absent from the imagery, linear interpolation is ~~performed~~conducted on each pixel of the missing imagery based on the time series~~, and for~~. In instances where the imagery ~~missing~~is absent for ~~more than~~a period exceeding one hour, the day in question is excluded. We calculate the daily average shortwave radiation in China based on China Standard Time (CST) using this hourly AHI shortwave radiation data.

The MCD19A2 is a MODIS Terra and Aqua combined multi-angle Implementation of Atmospheric Correction (MAIAC) Land AOD gridded Level 2 product produced daily at 1 km pixel resolution, which corrected for atmospheric gases and aerosols using a new MAIAC algorithm that is based on a time series analysis and a combination of pixel- and image-based processing (Lyapustin et al., 2022). In this study, the ~~daily, monthly and annual~~ AOD at 550 nm in MCD19A2 from 2016 to 2023 were collected using Google Earth Engine (GEE) (Gorelick et al., 2017).

2.2 Ground Measurements data

The ground measurements in CMA from 1 ~~July 2015~~January 2016 to 31 December 2023 used to perform SD estimation. The spatial coverage of Himawari covers 2380 CMA automatic meteorological stations in China. The CMA performs quality control of the data, including spatiotemporal consistency checks and manual corrections and adjustments before releasing the meteorological data (Moradi, 2009; Tang et al., 2010). Although the quality of the ground-based measurements should have been controlled before acquisition, there was still a need for a more stringent check on the quality of the data based on the methodology of daily meteorological data reconstruction from CMA (Zhang et al., 2015). Figure 1 shows the spatial distribution of 2380 meteorological. In this study, daily SD, vapor pressure (VAP), temperature, wind speed and precipitation from the CMA automatic meteorological stations were used to fit and validate the grid-dataset as well as to analyze the factors influencing the estimated performance, respectively~~. In this study,~~, and March-May was classified as spring, June-August as summer, September-November as autumn and December-February as winter.



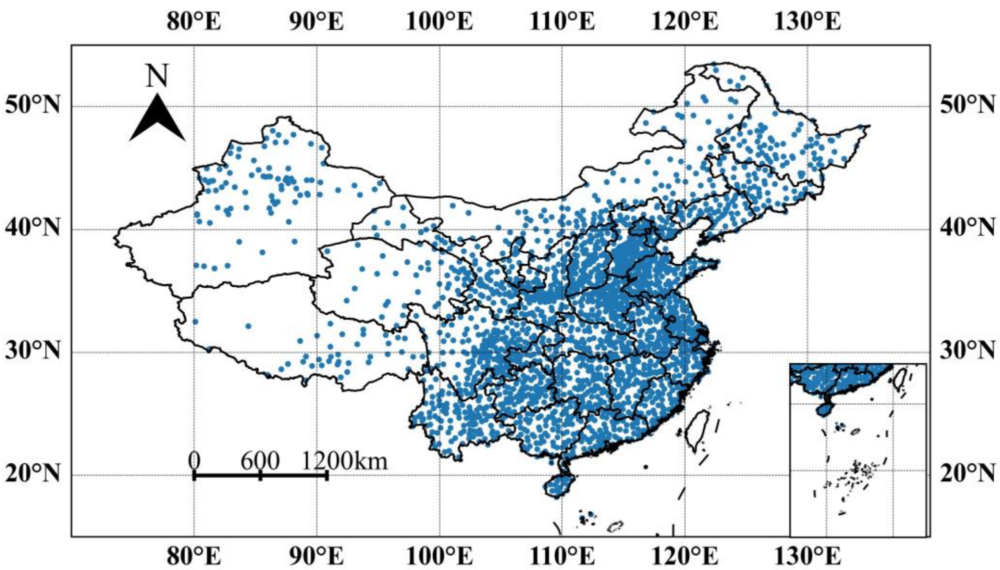


Figure 1. Spatial distribution of the 2380 automatic meteorological stations of the China Meteorological Administration (CMA).

2.3 Model overview

The Ångström-Prescott model is an empirical model which based on the relationship between SD and solar radiation, and is widely used in meteorology and agricultural science. The model was proposed by Ångström based on~~ the basis of~~ total solar radiation on clear days and improved by Prescott on the basis of astronomical radiation (~~Angstrom, 2007~~Ångström, 1924) with the following equations:

$$R_s=(a+b\frac{n}{N})R_a \tag{1}$$

where R_s is the total solar radiation reaching the surface, R_a is the astronomical radiation, a and b are empirical coefficients, n is the actual SD, and N is the maximum SD available. R_a and N counts are calculated with reference to Liu et al. (2009):

$$R_a=37.6d_r(\omega_s\sin\varphi\sin\delta+\cos\varphi\cos\delta\sin\omega_s) \tag{2}$$

$$d_r=1+0.033\cos(\frac{2\pi}{365}DOY) \tag{3}$$

$$\delta=0.4093\sin(\frac{2\pi}{365}DOY-1.39) \tag{4}$$

$$\omega_s=\arccos(-\tan\varphi\tan\delta) \tag{5}$$

$$N=\frac{24}{\pi}\omega_s \tag{6}$$

where d_r is the eccentricity of the Earth's orbit around the Sun, ω_s is the angle at sunset, φ is the latitude, δ is the ~~angle of~~ inclination angle of the sun, and DOY is the days of a year. We considered Himawari AHI level 3 hourly shortwave radiation as the R_s in this model, and SD of ground-based observation as a validation of n, and the parameters a and b of Ångström-Prescott model were fitted using the least-squares method.

2.4 Validation

We divided the original data into a training set (more than 5×10^6 grid cells during 2017-2022) and a testing set (2016 and 2023). In order to identify the best Ångström-Prescott model and its corresponding parameters, the performance of the Ångström-Prescott model on the training set (2017-2022) was evaluated using a 100-fold cross-validation (CV) approach, using a DOY-based CV strategy. In each iteration of each DOY, 99 folds were used as the training set and the remaining folds as the validation set, and the training and validation process was repeated 100 times to obtain the best model parameters a and b for each DOY. In addition, the 2016 and 2023 ground-based SD data were used as the test data to evaluate the generalization capability of the best model parameters a and b at each DOY. The specific process is shown in Figure 2. ~~pearson~~Pearson correlation coefficient (R) and root mean square error (RMSE) were calculated to evaluate the performance of the model.

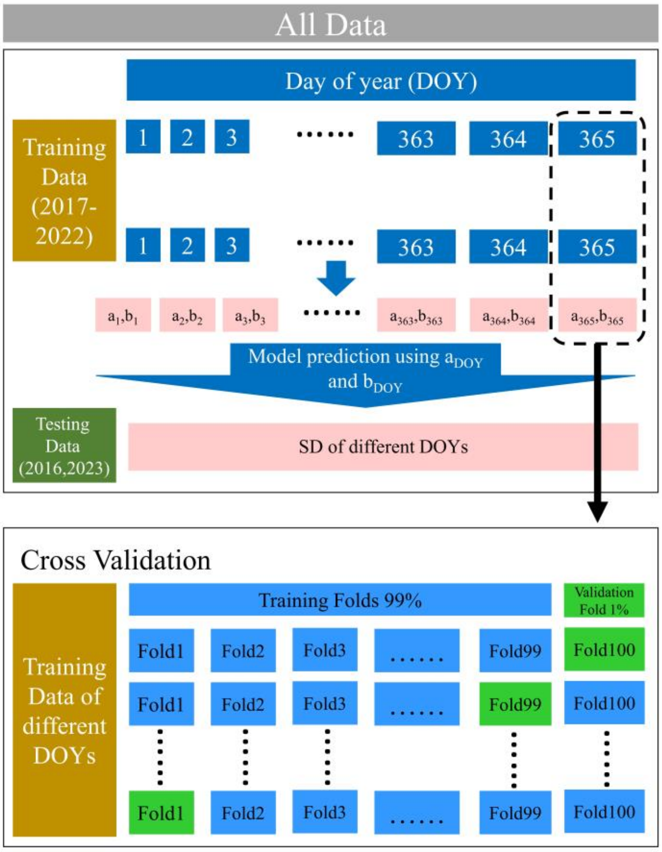


Figure2. Detailed process of model cross-validation and testing.

2.5 Methods of spatiotemporal variation analysis

Empirical orthogonal function (EOF) decomposition is a significant technique used to investigate the geographical and temporal fluctuations in meteorological characteristics (Zhou et al., 2021). The variable field can be decomposed into two parts: a spatial function that remains constant across time and a temporal function that changes exclusively with time, thus the primary spatial and temporal variations of which are evident in the area with a significant contribution to the variance. The spatial function component comprises several mutually independent and orthogonal spatial modes, also considered as eigenvectors. The temporal function part consists of the projection of the spatial modes in time, which is represented by the time coefficients. We used EOF to analyze spatiotemporal variations of the established SD dataset in China, then the original variable field information and spatial coefficients is concentrated in the first few modes.

3. Results

3.1 Evaluation of the training data

Figure 3 shows the estimation results of the CV sampling method for all DOYs in the training set (N=68806), an R value of 0.9695 was obtained for the entire training set, with a corresponding RMSE value of 1.2h. The measured and inverted SD converge to the 1:1 trend line, but overestimation occurs in the dense region around 10h. Figure 4 discusses the inverse performance of the different seasons in the training set separately. The SD is significantly higher in spring and summer than in autumn and winter, which is more concentrated in the 0h and 10h regions in winter. From Figure 4 it can be seen that in spring the highest R value is 0.9747 and RMSE value is 1.18h, while in winter the lowest RMSE value is 1.13h. However, in summer the highest RMSE value is 1.3h, and it is obvious that the estimation in summer performs the worst when the measured SD is 0h. The measured and inverted SD in spring most converge to the 1:1 trendline, while overestimation of which occurs in the dense region around 10h in winter.

Figure 5 shows the optimal Ångström-Prescott model parameters a and b at different DOYs. The parameter a has an upward parabolic trend with DOY, with a local maximum value of 0.22 at DOY = 306 and a local minimum value of 0.13 at DOY = 351. Parameter b showed a significant "W"-shaped variation with DOY, with a local maximum value of 0.74 at DOY = 146 and two local minimum values of 0.66 and 0.63 at DOY = 99 and 351. In general, parameters a and b of Ångström-Prescott model are characterized by more pronounced seasonal variations. Figure 6 shows the variation of the training set evaluation indicator (R and RMSE) with DOY. More than half of the DOYs had R values greater than the overall R value in Figure 3, but there were still 134 days with R values less than 0.97 and a minimum value of 0.94 at DOY = 193. Meanwhile more than half of the DOYs have RMSE values less than the overall RMSE values in Figure 3, but there are still 157 days with R values less than 1.2h, and again there is a maximum value of 2.1h for RMSE at DOY = 193. The evaluation indicator for the training set were not characterized by significant seasonal variations.

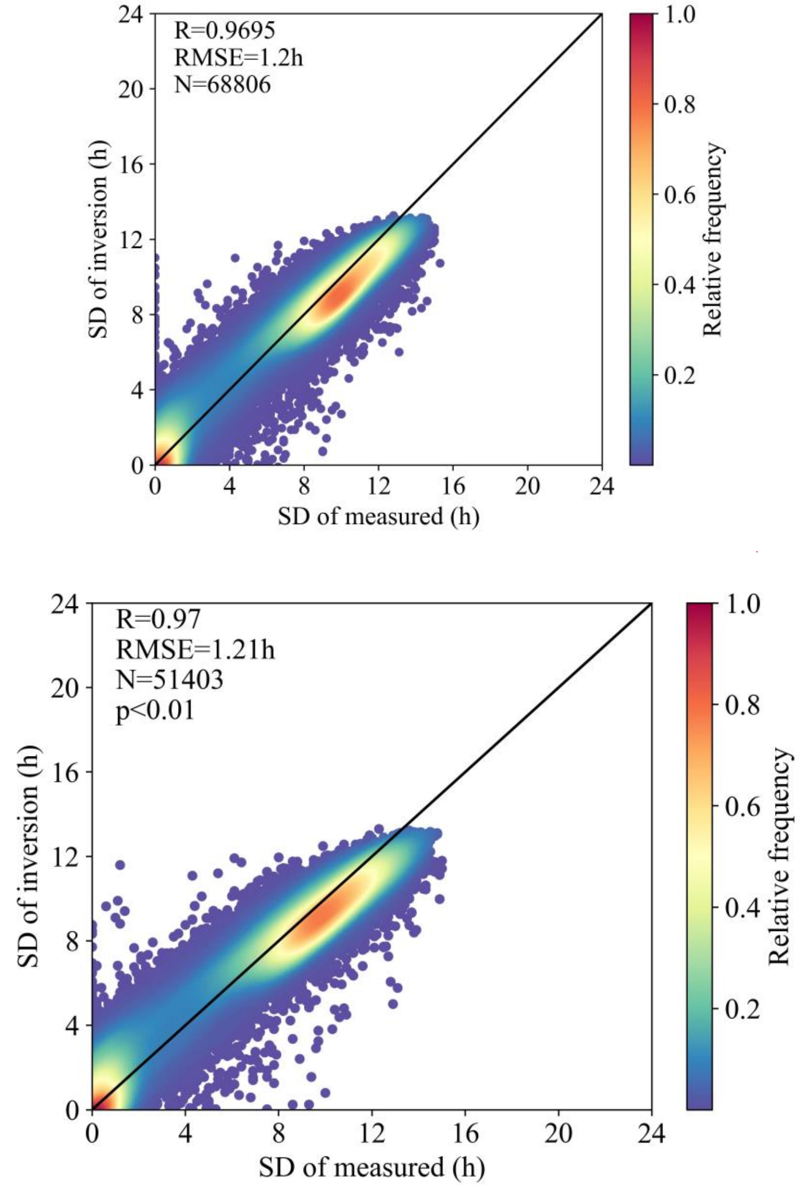
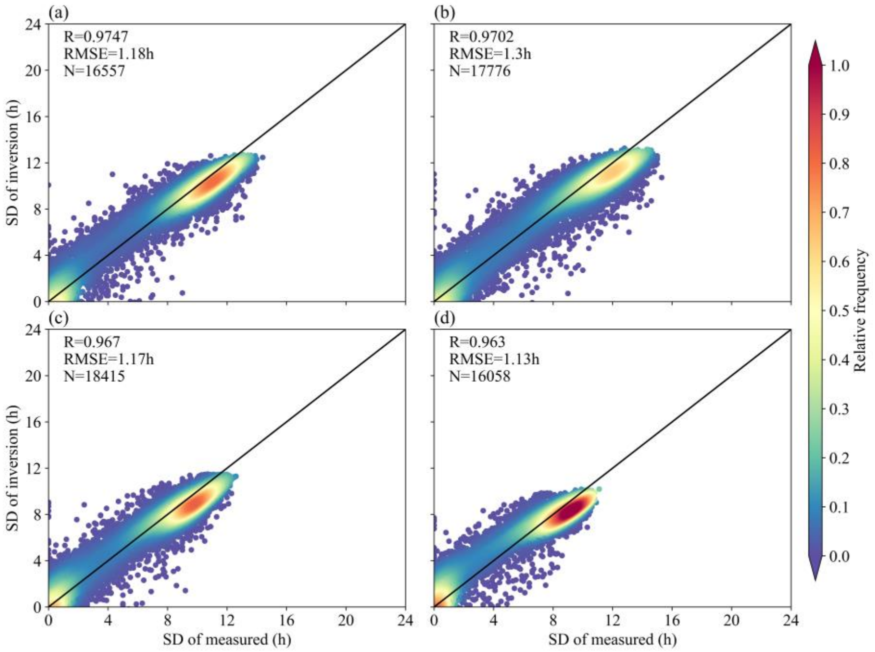


Figure 3. Estimation results of the CV sampling method in training set



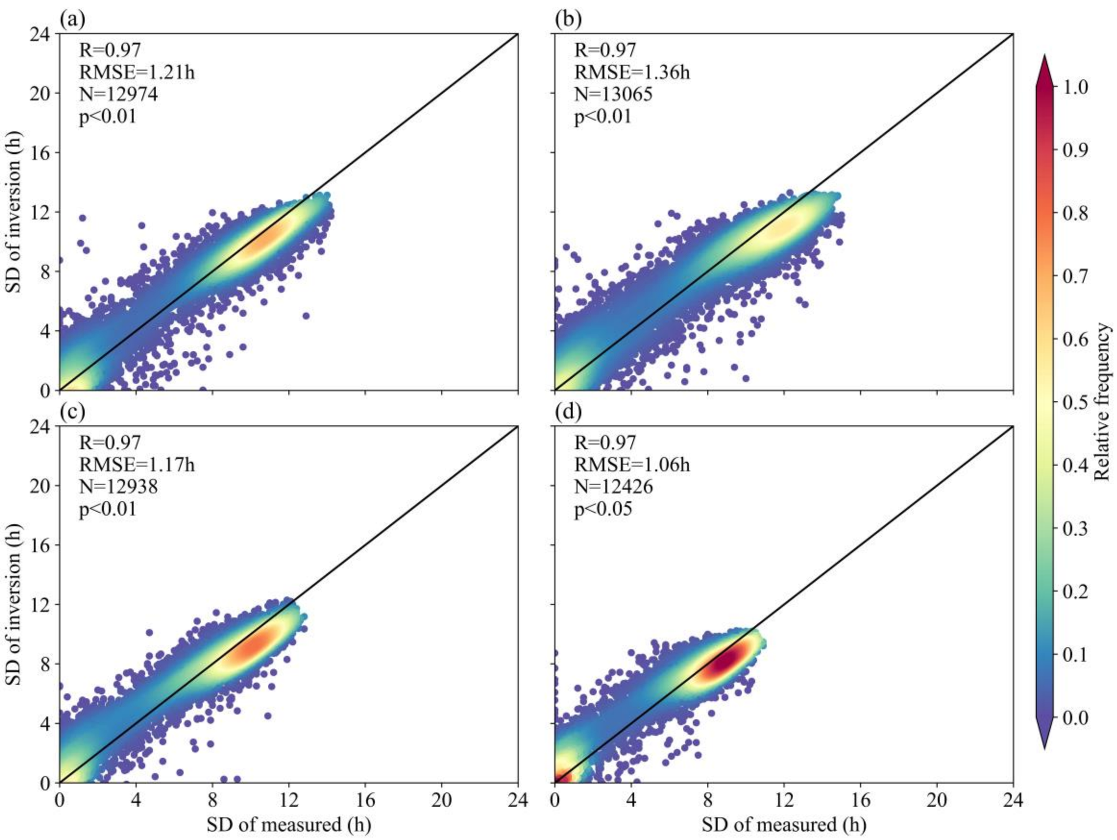


Figure 4. Estimation results of the CV sampling method in training set from different seasons ((a) spring, (b) summer, (c) autumn, (d) winter).

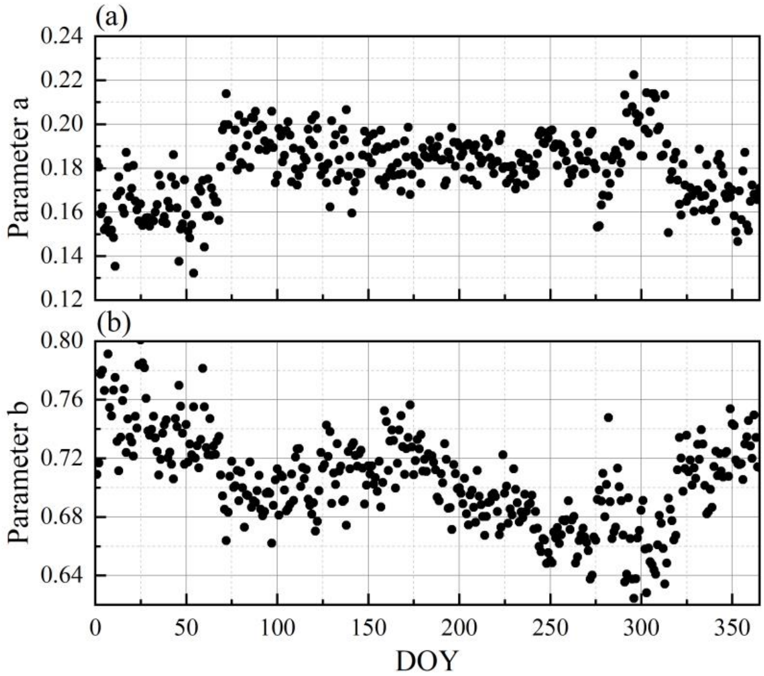


Figure 5. The a and b coefficients of Ångström-Prescott model for different DOYs.

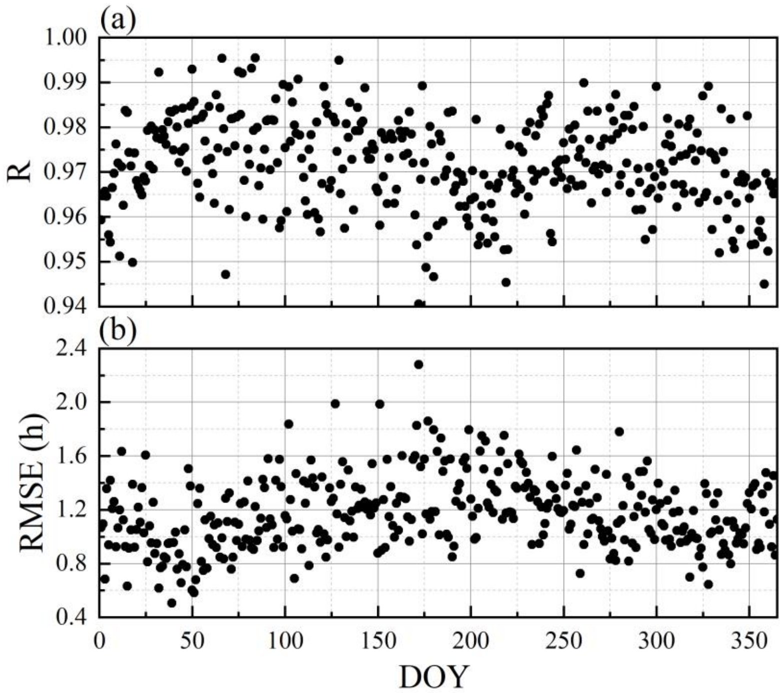


Figure 6. The correlation coefficients (R) (a) and RMSE (b) of CV sampling method in training set for different DOYs.

3.2 Evaluation of the testing data

The different evaluation indicator for the test set (2016 and 2023) are given in Figure 7, respectively. Figure 7(a) shows the R of 2016 and 2023, with the trends in these two years are essentially the identical, with an "M" shape. The average R value for 2016 is 0.88, which is generally consistent with 2023. The minimum R value of 0.52 in 2023 (DOY=361) was lower than that of 0.60 in 2016 (DOY=21), but both occurred in winter. The trend of RMSE values for 2016 and 2023 is opposite to the R value, with the maximum and minimum RMSE values occurring in 2023 at 2.77 (DOY=355) and 1.19 (DOY=106), respectively. Figures 7(c) and (d) show the estimated performance of the 0 SD (no sunshine for the whole day) for the CMA meteorological stations in 2016 and 2023. Figure 7(c) shows the estimated mean values of 0 SD for different DOYs in 2016 and 2023, where the mean value in 2023 (0.49h) is smaller than in 2016 (0.75h), with the maximum and minimum mean values still occurring in 2023 at 3.42 (DOY=211) and -0.75 (DOY=134), respectively. Figure 7(d) gives the number of estimated SD less than 0 for different DOYs in 2016 and 2023, of which there were more average daily estimated SDs less than 0 in 2016 than in 2023, at 267/day, with the lowest value also occurring in 2016, at 997 for DOY=294. ~~It can be seen that the~~The bias in the 0SD estimation is linked to the over- and under-representation of its number. Changing all estimated SD less than 0 to 0 resulted in an improvement in their estimated performance (Figure 8), with 2016 having a greater improvement than 2023 and having the greatest improvement with DOY=285.

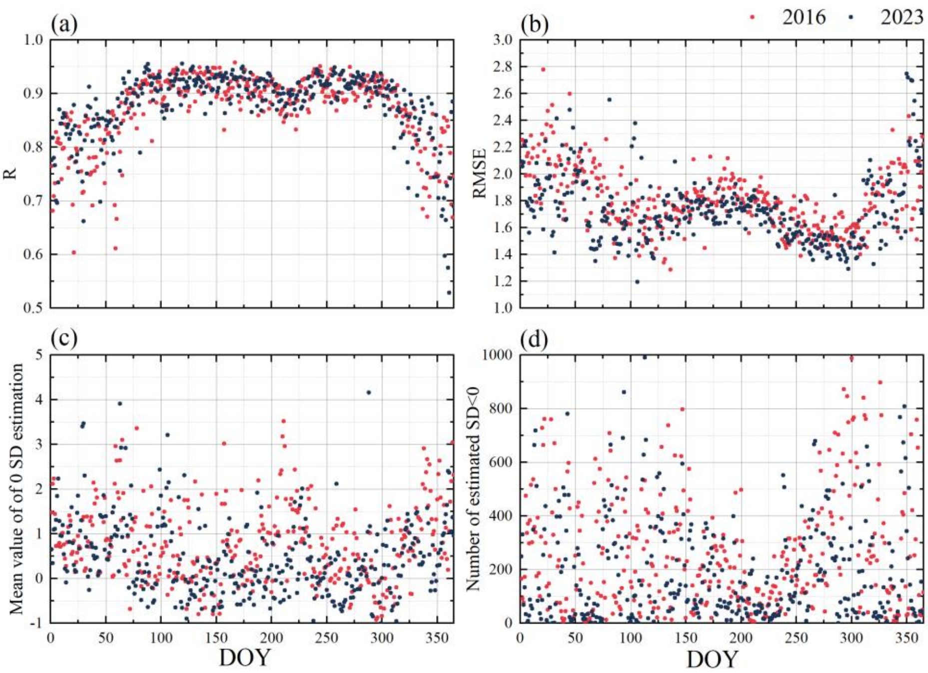


Figure 7. Estimated performance in testing set.

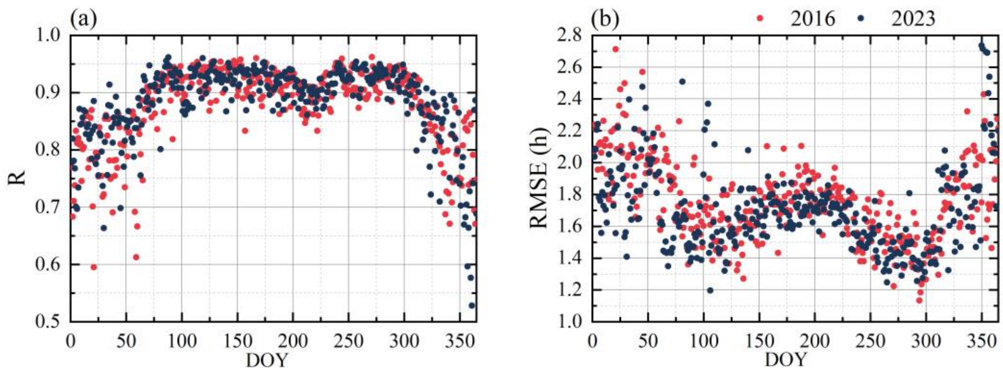
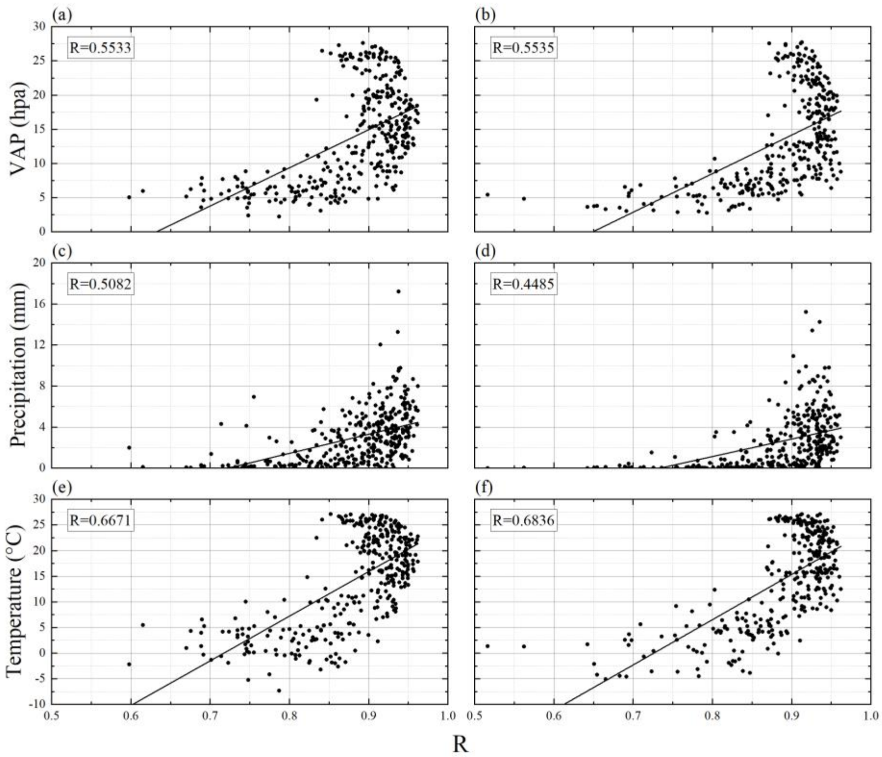


Figure 8. Estimated performance by changing all estimated SD less than 0 to 0 in testing set.

3.3 Effect of different environmental factors on SD estimation

Figure 9 shows the effect of national daily average VAP, precipitation, and temperature (based on CMA meteorological stations) on ~~estimated performance.~~R values in Figure 8. The R values~~ (changing all estimated SD less than 0 to 0 in Figure 8)~~ is exponentially related to both VAP and precipitation, and VAP has a greater effect on R than precipitation. Meanwhile the estimated performance in 2016 is more affected by moisture conditions. Temperature has the greatest impact on R, with 2023 being affected to a greater extent than 2016 (Figure 9 (e, f)). The influences on SD estimation are discussed by distinguishing the different seasons (Table 1), with VAP, precipitation and temperature having the greatest influence on R values in autumn and the least in winter. It is worth noting that R in summer were negatively correlated with VAP and temperature.

Figure 10 and 11 shows the annual average SD form CMA meteorological station and Himawari ~~estimated SD for 28 September~~ estimation in 2016 ~~(DOY=271, R=0.95),~~and 2023 respectively, along with the annual average AOD and wind speed~~ at that moment. The consistency of sites~~. On an annual scale, site and estimated SD ~~is stronger~~are in ~~northwest, north~~better consistency in eastern and ~~northeast~~northern China, while ~~overestimation occurs~~both years have higher estimates in eastern China~~.~~ ~~From Figure 10 (c, d), it can be found that~~ and lower estimates in northwestern and northeastern China, comparing the ~~excessively low AOD and high~~impact factors, higher wind speed ~~in East China affect~~and lower AOD in these areas both affected the ~~estimation of~~ SD estimation.



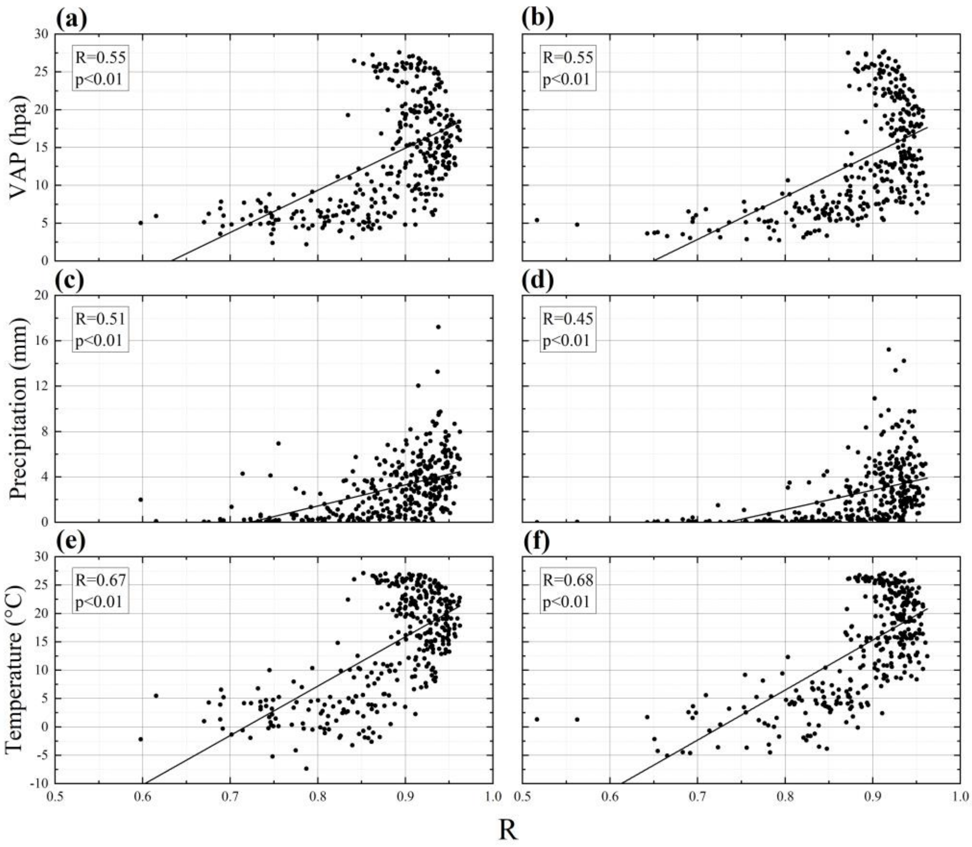
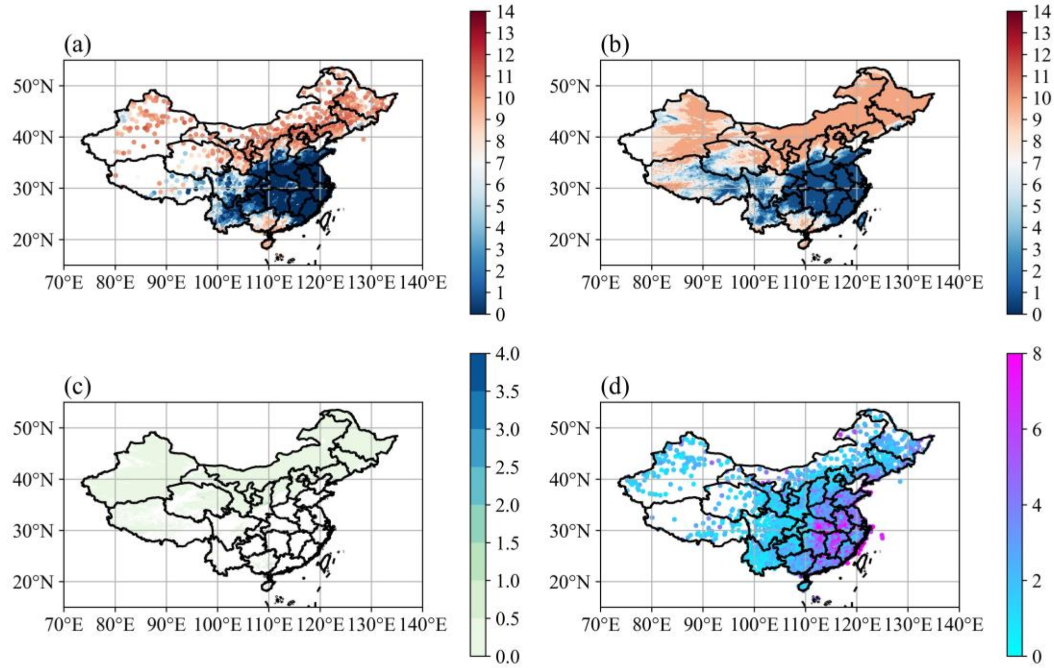


Figure 9. ~~Estimated performance (~~R values~~)~~ and different environmental factors (VAP (a, b), Precipitation (c, d), Temperature (e, f)) correlations in 2016 (a, c and e) and 2023 (b, d and f).

Table1. Correlation coefficients between estimated performance and influencing factors in different seasons (* and ** refer to passing the $p < 0.05$ and $p < 0.01$ significance tests, respectively)

Time	Influencing Factors		
	VAP	Precipitation	Temperature
Spring	0.29*	0.43**	0.31*
Summer	-0.56*	0.28*	-0.53**
Autumn	0.59**	0.46**	0.62**
Winter	0.28*	0.26**	0.22**



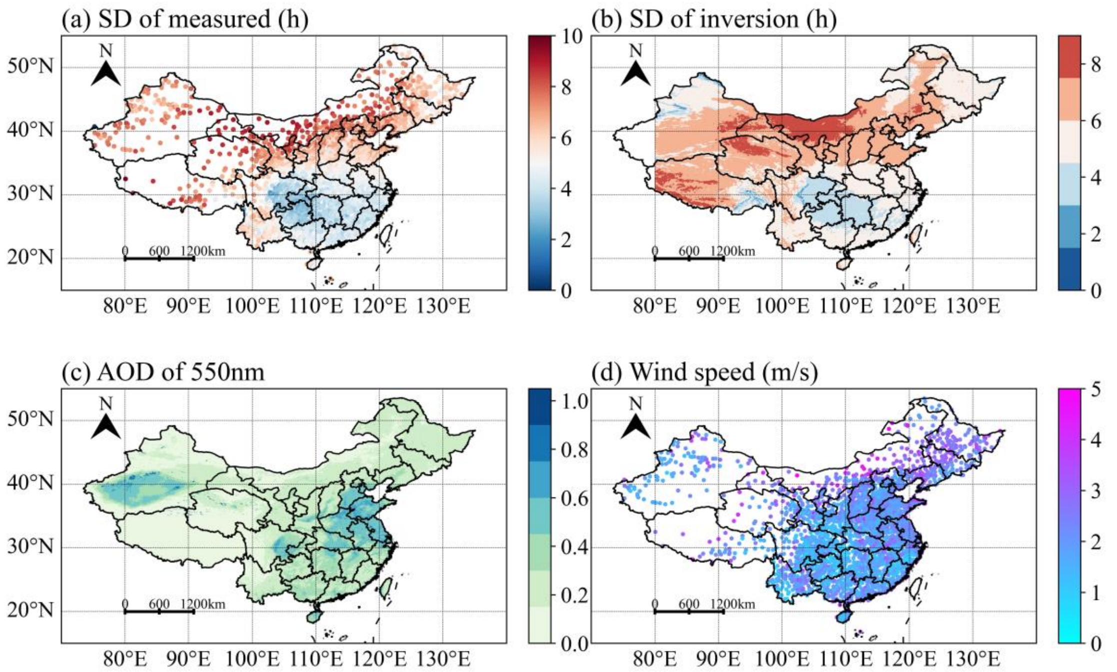


Figure 10. Comparison of annual average ground measurement (a) and Himawari (b) SD ~~on 28 September~~in 2016, giving ~~daily~~annual average AOD of 550nm (c) and the wind speed (d).

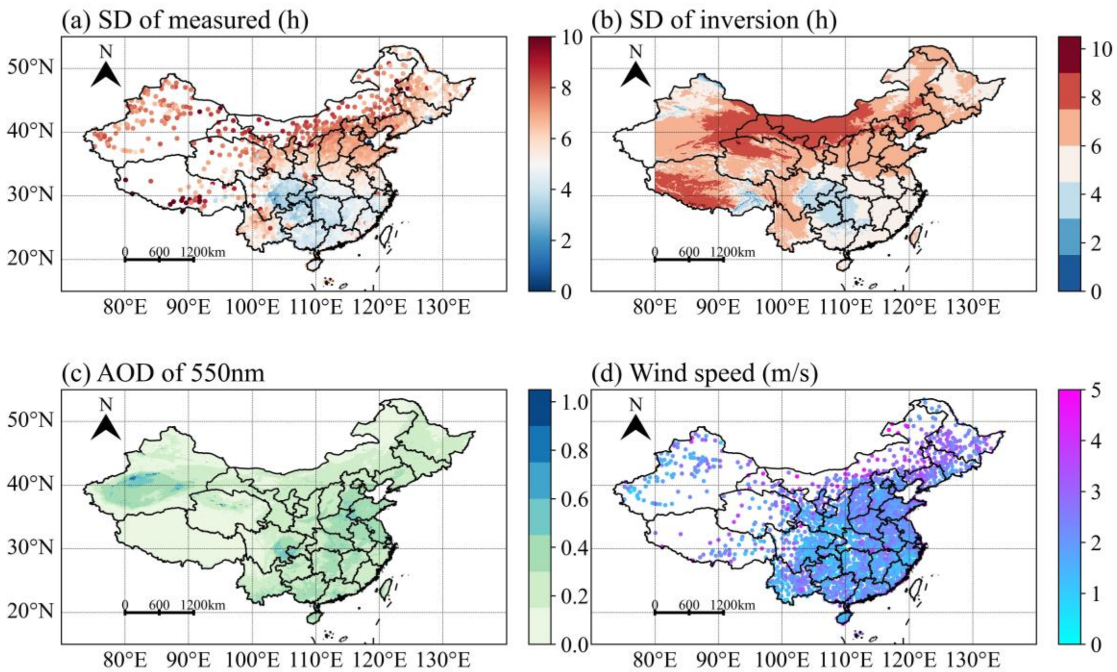
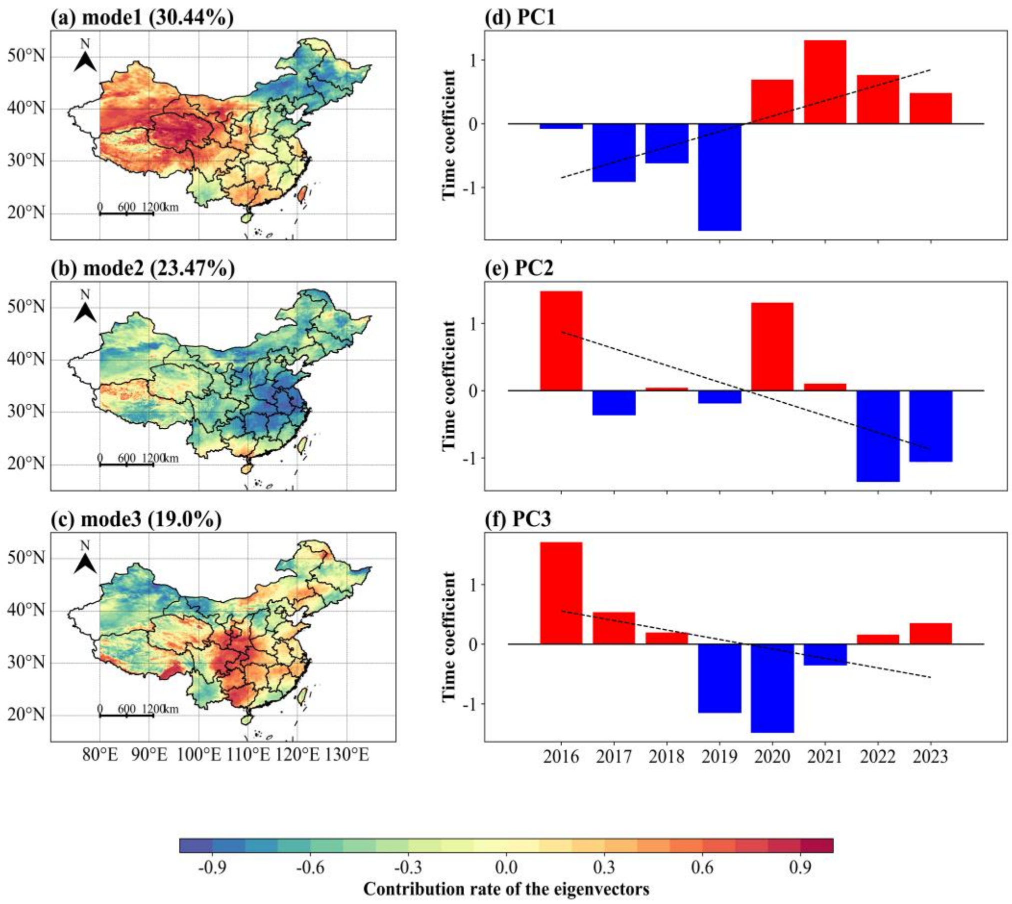


Figure 11. Same as Figure 10, but in 2023.

3.3 Effect of different environmental factors on SD estimation

EOF analysis of mean annual SD grid-data in China from 2015-2023, the spatial variance contribution rate of the eigenvectors in the first three EOF modes are shown in Figure 12, where the explained variance of each mode is 30.44%, 23.47% and 19.0%, respectively, with a cumulative variance contribution of about 72.91%. The variance contribution rate of mode 1 eigenvectors in Figure 12a surpasses that of other models, making it the predominant spatial distribution in China. The mode 1 decreases from western to eastern China, the northwest China exhibits extremely low values, but there are exceptions in Yunnan Province. The mode 2 (Figure 12b) exhibits a dipolar-type of distribution decreasing from the southern to northeast China, and the mode 3 shows a tri-pole distribution decreasing from central China to sides. Generally, it can be concluded that the SD decreases from western to northern China. Figure 12def shows the time coefficients of SD from the first three models in China, the SD time coefficients of the mode 1 (Figure 12d) shows an increasing trend from 2016 to 2023, with the minimum time coefficient in 2019 and maximum time coefficient in 2021. It can be seen from Figure 12ef that the SD time coefficients of the mode 2 and 3 show a decreasing trend, and both are positive in 2016 and negative in 2019.



Figures 12. Distribution of eigenvectors contribution rate (a-c) and time coefficients (d-f) for the first

three modes of SD.

4. Discussion

There is no explicit remote sensing inversion model for SD as its observation is founded upon the accumulation of radiation. Consequently, SD datasets were constructed through the spatial interpolation, which results in the absence of SD datasets that are released with high spatiotemporal resolution. In this study, a 5km-resolution SD dataset in China from 2016 to 2023 has been established based on time series using Himawari imagery fitted with Ångström-Prescott model, which previous studies have not been conducted.

The time series~~-~~ based Ångström-Prescott model was used to invert the SD in China, setting the coefficients of a and b to fixed values for the whole region at different DOYs~~. However,~~, while the suggested coefficients in this study are not comparable with the calibrated coefficients for other regions. Previous studies on the Ångström-Prescott model have confirmed that it is a reliable tool for estimating solar energy in practical applications, with no significant dependence of its accuracy on latitude (Paulescu ~~rt~~et al., 2016). It has also been confirmed that the model's accuracy has a strong dependence on and season (Liu et al., 2023) according to the results of the present study (Figure 4-~~7~~8), the cause of which can be attributed to differences in the length of day and night in different seasons. This work not only forms a more accurate evaluation standard for the level of radiation received on the ground, but also provides a better support for the radiation estimation ~~of surface short-wave radiation~~ in the future~~ by using the established Ångström-Prescott model~~, and more conventional meteorological stations will be established in the future to validate and improve the Ångström-Prescott model based on time-series. A fact that cannot be ignored is that the number of meteorological observation stations in southwestern China (especially in the Tibetan Plateau Region) is small and spatially distributed unevenly, and the snow in the plateau seriously affects the judgement of the reflectance data from the Himawari imagery, and we will consider the input of the land cover characteristics as the climatological data in the following to improve this poor performance.

~~The 0 SD accounts for~~ It is worth noting that there is a ~~certain proportion~~bias in the validation of the training and test data, ~~and the~~ where there is an overestimation at 0SD (Figure 3), may be the strong light in almost most of the area under a DOY leads to Ångström-Prescott model ~~still needs to be improved~~ larger parameters and ~~optimized in determining this situation (Figure 7 c, d), which is presumed to be~~

~~due to the low impact of cloudy~~over-estimation of a very small portion of the image elements that contain aerosols, clouds and ~~rainy days~~even precipitation. In addition, it also occurred in the test data that the estimated SD was less than 0 (Figure 7 cd), because the thicker clouds, atmospheric aerosols and water vapor in majority of the area on ~~the~~ that day did not have much effect on the ground-based SD instrument (the atmospheric longwave radiation contained in the direct radiation was not affected), but had a significant effect on the AHI shortwave radiation data, resulting in SD less than 0. After changing the image elements with SD less than 0 to 0, the validation results are still substantial (Figure 8), indicating that this part of radiation is essentially less than the threshold for SD observations~~, resulting in the low sensitivity of the shortwave bands to the SD estimation. Subsequently~~ (120 W/m^2). In conclusion, as our approach is carried out based on time series, it is unavoidable that we will encounter input data that are not sensitive to different sky conditions. In the future, the use of relevant physical precipitation models will be considered to simulate the precipitation process at different times of the day based on the ~~radiometric~~radiation data ~~before proceeding~~. This will enable us to estimate SD, and this aspect of the Ångström-Prescott model will be improved subsequently.

~~In this study we~~ We found that temperature, moisture conditions, wind speed and atmospheric pollutants all ~~have an effect on~~influence the SD estimation, with temperature having the greatest effect in temporal variation and wind speed having a stronger effect in spatial variation compared with AOD. However, we believe that the effects of these environmental factors are not independent, but are the result of interaction (Tang et al., 2022). In densely populated and economically developed areas (eastern and southern China), where pollutant levels are higher and increased wind speed accelerates their dispersion, this regulatory mechanism is enhanced with increasing pollutants (O'Dowd et al., 1993; Wang et al., 2014). An increase or decrease in wind speed affects the rate of diffusion of water vapor and pollutants in the air, which in turn affects atmospheric transparency and ultimately the SD estimation. However, ~~the results of~~ the effect of temperature on SD estimation in this study are not consistent with some previous studies (Tang et al., 2022; Feng et al., 2019; Ren et al., 2017), which suggests that the relationship between SD and temperature and relative humidity is complex and needs to be further determined in future studies.

The EOF method analysis of mean annual SD declare that it decreases from western to northeast China, which is consistent with the Tang et al. (2022) and Xiong et al. (2020), suggesting that the pattern of industrial development between western to eastern China is affecting radiation levels to some extent.

The time coefficients of EOF show that there is a certain degree of increase in SD in recent years, which is closed to long-term SD analysis from Tang et al. (2022). This trend may be related to global climate change (Josefsson and Landelius, 2000), because of the variation in wind speeds due to global warming has resulted in decreased cloud dissipation across mainland China (Xiong et al., 2020). In addition, the decrease in human activities in recent years (Liu et al., 2020) has also contributed to a weakening of the urban rain island effect and aerosols (Glantz et al., 2006), and it appears that the latter factor is more influential from this study. However short-term reductions in human activity cannot become the norm, and sunshine duration are bound to fluctuating changes due to the acceleration of the hydrological cycle.

5. Data availability

The SD dataset is freely accessible at https://doi.org/10.57760/sciencedb.10276 (Zhang et al., 2024).

6. Conclusion

We have introduced a newly developed high-resolution dataset, which provides SD in China for the period 2016–2023. We calculated daily SD by Himawari Level 3 shortwave radiation fitted with the Ångström-Prescott model based on time series, and used ground-measured SD to evaluate the estimation performance. The validation of testing data from ground-measured SD gave favorable results, with R values greater than 0.5 and an average of 0.88 for all days in 2016 and 2023. We also found that temperature and wind speed dominate the Ångström-Prescott model estimating SD. A future direction for this study would be to divide the Chinese regions into suitable areas to independently estimate and synthesize a more accurate daily SD dataset in China.

Author contributions. ZZ and SF designed and organized the paper. ZZ and JH prepared the related materials and ran the dataset. ZZ evaluated the accuracy of the dataset. All authors discussed the results and commented on the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Financial support. This research was supported by the National Key Research and Development Program of China (grant no. 2023YFE0122200), the National Nature Sciences Foundation (grant no. 42075193).

Reference

Ampratwum, D. B. and Dorvlo, A. S.: Estimation of solar radiation from the number of sunshine hours. Appl. Energy, 63(3), 161-167. https://doi.org/10.1016/S0306-2619(99)00025-2, 1999.

~~Angstrom,~~Ångström A.: Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of ~~solar~~sola and atmospheric radiation. Q J ~~R Meteorol~~Roy Meteor Soc~~,~~., 50~~,~~:121~~-126~~–6. ~~https://doi.org/10.1002/QJ.49705021008~~https://doi.org/10.1002/qj.49705021008, ~~2007~~1924.

Ångström A.: Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of sola and atmospheric radiation. Q J Roy Meteor Soc., 50:121–6. https://doi.org/10.1002/qj.49705021008, 1924.

Bahel, V., Bakhsh, H. and Srinivasan, R.: A correlation for estimation of global solar radiation. Energy, 12, 131-135. https://doi.org/10.1016/0360-5442(87)90117-4, 1987.

Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., Kumagai, Y., Miyakawa, T., Murata, H., Ohno, T., Okuyama, A., Oyama, R., Sasaki, Y., Shimazu, Y., Shimoji, K., Sumida, Y., Suzuki, M., Taniguchi, H., Tsuchiyama, H., Uesawa, D., Yokota, H. and Yoshida, R.: An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. J. Meteorol. Soc. Japan,. Ser. II, 94(2), 151-183. https://doi.org/10.2151/JMSJ.2016-009, 2016.

Chang, Z., Chen, Y., Zhao, Y., Fu, J., Liu, Y., Tang, S., Han, Y. and Fan, Z.: Association of sunshine duration with acute myocardial infarction hospital admissions in Beijing, China: A time-series analysis within-summer. The Science of the total environment, 154528. https://doi.org/10.1016/10.1016/j.scitotenv.2022.154528, 2022.

Chen, R., Ersi, K., Yang, J., Lu, S. and Zhao, W.: Validation of five global radiation models with measured daily data in China. Energy Convers. Manage., 45, 1759-1769. https://doi.org/10.1016/J.ENCONMAN.2003.09.019, 2004.

Chukwujindu, N.S.: A comprehensive review of empirical models for estimating global solar radiation in Africa. Renew. Sust. Energ. Rev., 78, 955-995. https://doi.org/10.1016/J.RSER.2017.04.101, 2017.

Damiani, A., Irie, H., Horio, T., Takamura, T., Khatri, P., Takenaka, H., Nagao, T.M., Nakajima, T.Y.: and Cordero, R.R. Evaluation of Himawari-8 surface downwelling solar radiation by SKYNET observations. Atmos. Meas. Tech. Discuss, 1-28. https://doi.org/10.5194/AMT-2017-440, 2018.

Dedieu, G., P. Y. Deschamps, and Y. H. Kerr.: Satellite estimation of solar irradiance at the surface of the earth and of surface albedo using a physical model applied to Metcosat Data. J. Appl. Meteorol. Climatol., 26.1: 79-87. https://doi.org/10.1175/1520-0450(1987)026<0079:SEOSIA>2.0.CO;2, 1987.

Elagib, N.A. and Mansell, M.G.: New approaches for estimating global solar radiation across Sudan. Energy Convers. Manage., 41, 419-434. https://doi.org/10.1016/S0196-8904(99)00123-5, 2000.

Fan, J., Wang, X., Wu, L., Zhang, F., Bai, H., Lu, X. and Xiang, Y.: New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: A case study in South China. Energy Convers. Manage., 156, 618-625. https://doi.org/10.1016/J.ENCONMAN.2017.11.085, 2018.

Feng, Y., Zhang, X., Jia, Y., Cui, N., Hao, W., Li, H. and Gong, D.: High-resolution assessment of solar radiation and energy potential in China. Energy Convers. Manage., 240, 114265. https://doi.org/10.1016/j.atmosenv.2022.119286, 2021.

Feng, Z., Guo, B., Ren, S. and Li, Y.: Reduction in sunshine duration and related factors over mainland China during 1961–2016. Energies, 12(24), 4718. https://doi.org/10.3390/en12244718, 2019.

Frouin, R. and Murakami, H.: Estimating photosynthetically available radiation at the ocean surface from

ADEOS-II global imager data. J Oceanogr., 63, 493-503. https://doi.org/10.1007/S10872-007-0044-3, 2007.

Frouin, Robert, and Rachel T. Pinker.: Estimating photosynthetically active radiation (PAR) at the earth's surface from satellite observations. Remote Sens Environ., 51.1: 98-107. https://doi.org/10.1016/0034-4257(94)00068-X, 1995.

Ghanghermeh, A., Roshan, G. and Halabian, A.: Projecting spatiotemporal variations of sunshine duration with regards to climate change in Iran as a step towards clean energy. Sustain. Energy Technol. Assess., 53, 102630. https://doi.org/10.1016/j.seta.2022.102630, 2022.

Glantz, P., Nilsson, D.E. and Hoyningen-Huene, W.V. Estimating a relationship between aerosol optical thickness and surface wind speed over the ocean. Atmos. Chem. Phys., 6, 11621-11651. https://doi.org/10.5194/ACPD-6-11621-2006, 2006.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens Environ., 202, 18-27. https://doi.org/10.1016/J.RSE.2017.06.031, 2017.

Gu, S., Huang, R., Yang, J., Sun, S., Xu, Y., Zhang, R., Wang, Y., Lu, B., He, T., Wang, A., Bian, G. and Wang, Q.: Exposure-lag-response association between sunlight and schizophrenia in Ningbo, China. Environ. Pollut, 247, 285-292. https://doi.org/10.1016/j.envpol.2018.12.023, 2019.

Hou, N., Zhang, X., Zhang, W., Wei, Y., Jia, K., Yao, Y., Jiang, B. and Cheng, J.: Estimation of Surface Downward Shortwave Radiation over China from Himawari-8 AHI Data Based on Random Forest. Remote. Sens., 12, 181. https://doi.org/10.3390/rs12010181, 2023.

Josefsson, W. and Landelius, T. Effect of clouds on UV irradiance: As estimated from cloud amount, cloud type, precipitation, global radiation and sunshine duration. J. Geophys., 105, 4927-4935. https://doi.org/10.1029/1999JD900255, 2000.

Kim, B., Lee, K., Jee, J. and Zo, I.: Retrieval of outgoing longwave radiation at top-of-atmosphere using Himawari-8 AHI data. Remote Sens Environ., 204, 498-508. https://doi.org/10.1016/J.RSE.2017.10.006, 2018.

Letu, H., Yang, K., Nakajima, T.Y., Ishimoto, H., Nagao, T.M., Riedi, J.C., Baran, A.J., Ma, R., Wang, T., Shang, H., Khatri, P., Chen, L., Shi, C. and Shi, J.: High-resolution retrieval of cloud microphysical properties and surface solar radiation using Himawari-8/AHI next-generation geostationary satellite. Remote Sens Environ., 239, 111583. https://doi.org/10.1016/j.rse.2019.111583, 2020.

Liang, S., Zheng, T., Liu, R., Fang, H., Tsay, S. and Running, S.W.: Estimation of incident photosynthetically active radiation from Moderate Resolution Imaging Spectrometer data. J. Geophys. Res., 111. https://doi.org/10.1029/2005JD006730, 2006.

Liu, F., Wang, X., Sun, F. and Wang, H.: Correct and remap solar radiation and photovoltaic power in China based on machine learning models. Appl. Energy, 312, 118775. https://doi.org/10.1016/j.apenergy.2022.118775, 2022.

Liu, J., Liu, J., Linderholm, H.W., Chen, D.L., Yu, Q., Wu, D. and Haginoya, S.: Observation and calculation of the solar radiation on the Tibetan Plateau. Energy Convers. Manage., 57, 23-32. https://doi.org/10.1016/J.ENCONMAN.2011.12.007, 2012.

Liu, J., Shen, Y., Zhou, G., Liu, D., Yu, Q. and Du, J.: Calibrating the Ångström–Prescott Model with Solar Radiation Data Collected over Long and Short Periods of Time over the Tibetan Plateau. Energies. https://doi.org/10.3390/en16207093, 2023.

Liu, Q., Sha, D., Liu, W., Houser, P.R., Zhang, L., Hou, R., Lan, H., Flynn, C., Lu, M., Hu, T. and Yang, C. Spatiotemporal Patterns of COVID-19 Impact on Human Activities and Environment in Mainland

China Using Nighttime Light and Air Quality Data. Remote. Sens., 12, 1576. https://doi.org/10.3390/rs12101576, 2020.

Liu, X., Mei, X., Li, Y., Wang, Q., Zhang, Y. and Porter, J. R.: Variation in reference crop evapotranspiration caused by the Ångström–Prescott coefficient: Locally calibrated versus the FAO recommended. Agric Water Manag., 96(7), 1137-1145. https://doi.org/10.1016/J.AGWAT.2009.03.005, 2009.

Lyapustin, A. and Wang, Y.: MODIS/Terra+ Aqua Land Aerosol Optical Depth Daily L2G Global 1km SIN Grid V061 [Data set]. Accessed 2022-03-05 from. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MCD19A2.061, 2022.

Moradi, I.: Quality control of global solar radiation using sunshine duration hours. Energy, 34, 1-6. https://doi.org/10.1016/J.ENERGY.2008.09.006, 2009

O'Dowd, C. D. and Smith, M. H.: Physicochemical properties of aerosols over the northeast Atlantic: Evidence for wind-speed-related submicron sea-salt aerosol production. J. Geophys. Res. Atmos., 98(D1), 1137-1149. https://doi.org/10.1029/92JD02302, 1993.

Paulescu, M., Stefu, N., Calinoiu, D., Paulescu, E., Pop, N., Boată, R. and Mares, O.: Ångström–Prescott equation: Physical basis, empirical models and sensitivity analysis. Renew. Sust. Energ. Rev., 62, 495-506. https://doi.org/10.1016/J.RSER.2016.04.012, 2016.

Qin, S., Liu, Z., Qiu, R., Luo, Y., Wu, J., Zhang, B., Wu, L. and Agathokleous, E.: Short–term global solar radiation forecasting based on an improved method for sunshine duration prediction and public weather forecasts. Appl. Energy, 343, 121205. https://doi.org/10.1016/j.apenergy.2023.121205, 2023.

Ren, J., Lei, X., Zhang, Y., Wang, M., and Xiang, L.: Sunshine duration variability in haihe river basin, China, during 1966–2015. Water, 9(10), 770. https://doi.org/10.3390/W9100770, 2017.

Rietveld, M. R.: A new method for estimating the regression coefficients in the formula relating solar radiation to sunshine. Agric. Meteorol., 19(2-3), 243-252. https://doi.org/10.1016/0002-1571(78)90014-6, 1978.

Sawada, Y., Okamoto, K., Kunii, M. and Miyoshi, T.: Assimilating Every-10-minute Himawari-8 Infrared Radiances to Improve Convective Predictability. J. Geophys. Res. Atmos., 124, 2546 - 2561. https://doi.org/10.1029/2018JD029643, 2019.

Tana, G., Ri, X., Shi, C., Ma, R., Letu, H., Xu, J. and Shi, J.: Retrieval of cloud microphysical properties from Himawari-8/AHI infrared channels and its application in surface shortwave downward radiation estimation in the sun glint region. Remote Sens Environ., 290, 113548. https://doi.org/10.1016/j.rse.2023.113548, 2023.

Tang, C., Zhu, Y., Wei, Y., Zhao, F., Wu, X. and Tian, X.: Spatiotemporal characteristics and influencing factors of sunshine duration in China from 1970 to 2019. Atmosphere, 13(12), 2015. https://doi.org/10.3390/atmos13122015, 2022.

Tang, W., Yang, K., He, J. and Qin, J.: Quality control and estimation of global solar radiation in China. Solar Energy, 84, 466-475. https://doi.org/10.1016/J.SOLENER.2010.01.006, 2010.

Vivar, M., Fuentes, M., Norton, M., Makrides, G. and Bustamante, I.D.: Estimation of sunshine duration from the global irradiance measured by a photovoltaic silicon solar cell. Renew. Sust. Energ. Rev., 36, 26-33. https://doi.org/10.1016/J.RSER.2014.04.045, 2014.

Wang, Y. W., Yang, Y. H., Zhou, X. Y., Zhao, N. and Zhang, J. H.: Air pollution is pushing wind speed into a regulator of surface solar irradiance in China. Environ. Res. Lett., 9(5), 054004. https://doi.org/10.1088/1748-9326/9/5/054004, 2014.

Wu, G., Liu, Y. and Wang, T.: Methods and strategy for modeling daily global solar radiation with measured meteorological data–A case study in Nanchang station, China. Energy Convers. Manage., 48(9), 2447-2452. https://doi.org/10.1016/J.ENCONMAN.2007.04.011, 2007.

Xia, X.: Spatiotemporal changes in sunshine duration and cloud amount as well as their relationship in China during 1954–2005. J. Geophys. Res. Atmos., 115(D7). https://doi.org/10.1029/2009JD012879, 2010.

Xiong, J., Wang, Z., Lai, C., Liao, Y. and Wu, X.: Spatiotemporal variability of sunshine duration and influential climatic factors in mainland China during 1959–2017. Int. J. Climatol., 40, 6282 - 6300. https://doi.org/10.1002/joc.6580, 2020.

Yao, W., Zhang, C., Wang, X., Zhang, Z., Li, X. and Di, H.: A new correlation between global solar radiation and the quality of sunshine duration in China. Energy Convers. Manage., 164, 579-587. https://doi.org/10.1016/J.ENCONMAN.2018.03.037, 2018.

Yu, L., Zhang, M., Wang, L., Qin, W., Jiang, D. and Li, J.: Variability of surface solar radiation under clear skies over Qinghai-Tibet Plateau: Role of aerosols and water vapor. Atmos. Environ., 287, 119286, https://doi.org/10.1016/j.atmosenv.2022.119286, 2022.

Yu, Y., Shi, J., Wang, T., Letu, H., Yuan, P., Zhou, W. and Hu, L.: Evaluation of the Himawari-8 Shortwave Downward Radiation (SWDR) Product and its Comparison With the CERES-SYN, MERRA-2, and ERA-Interim Datasets. IEEE J Sel Top Appl Earth Obs Remote Sens, 12, 519-532. https://doi.org/10.1109/JSTARS.2018.2851965, 2019.

Zhang, J., Zhao, L., Deng, S., Xu, W. and Zhang, Y.: A critical review of the models used to estimate solar radiation. Renew. Sust. Energ. Rev., 70, 314-329. https://doi.org/10.1016/J.RSER.2016.11.124, 2017.

Zhang, P., Guo, Q., Chen, B. and Feng, X.: The Chinese Next-Generation Geostationary Meteorological Satellite FY-4 Compared with the Japanese Himawari-8/9 Satellites. Adv. Meteorol. Sci. Technol., (1), 4. https://doi.org/10.3969/j.issn.2095-1973.2016.01.010, 2016. (in chinese)

Zhang, X., Liang, S., Wild, M. and Jiang, B.: Analysis of surface incident shortwave radiation from four satellite products. Remote Sens Environ., 165, 186-202. https://doi.org/10.1016/J.RSE.2015.05.015, 2015.

Zhang, Z., Fang, S. and Han, J.: A daily sunshine duration (SD) dataset in China from Himawari AHI imagery (2016-2023) (~~V1~~V2), https://doi.org/10.57760/sciencedb.10276, 2024.

Zhou, Y., Yu, D., Yang, Q., Pan, S., Gai, Y., Cheng, W., Liu, X. and Tang, S.: Variations of Water Transparency and Impact Factors in the Bohai and Yellow Seas from Satellite Observations. Remote. Sens., 13, 514. https://doi.org/10.3390/rs13030514, 2021