

1 **A daily sunshine duration (SD) dataset in China from** 2 **Himawari AHI imagery (2016-2023)**

3 Zhanhao Zhang^{1,2}, Shibo Fang¹, Jiahao Han¹

4 ¹State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing
5 100081, China

6 ²College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, 100049,
7 China

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

9

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

23

24 **1. Introduction**

25 Solar radiation is a major driver of photosynthesis and evapotranspiration, plays an indispensable
26 role in regulating temperature and supporting agricultural production, and has effects on photovoltaic
27 power generation, making it critical to the Earth's ecosystem and to productive human life (Yu et al.,
28 2022; Feng et al., 2021). The solar radiation measured by radiation observatory can accurately predict
29 solar radiation potential and participate in climate change and agricultural production model. Nonetheless,

30 the existing radiation data in China is not validated through terrestrial observations due to the limited
31 number of less than 200 stations in mainland China (Liang et al., 2006; Zhang et al., 2015) for the
32 expensive upkeep of terrestrial radiation measuring devices, making precise tracking of high
33 spatiotemporal solar radiation over time difficult (Zhang et al., 2017; Chukwujindu et al., 2017).

34 Sunshine duration (SD) is a readily available and cost-effective indicator for monitoring the global
35 radiation resources, and the variability of which is determined by a combination of regional factors as
36 well as the solar constant, cloud cover, water vapor, and atmospheric pollutants. SD is a key parameter
37 of solar power potential forecasting (Liu et al., 2022; Qin et al., 2023), climate change assessment and
38 agricultural production (Ghanghermeh et al., 2022), in addition, some researchers have found that
39 changes in SD also affect the probability of human diseases (Chang et al., 2022; Gu et al., 2019). The
40 SD measured from conventional meteorological observation has the advantages of long observation time,
41 good continuity, high spatial density and reliability, which is considered the best alternative to solar
42 radiation (Xia, 2010). Accurate inversion of SD is therefore an important reference for agricultural
43 production, solar resource utilization and global climate change analysis. The Ångström-PreScott model
44 (Ångström, 1924) is the dominant and most widely used model based on SD and solar radiation. The
45 quadratic and cubic forms of the Ångström-PreScott model have been improved and applied to different
46 meteorological conditions (Rietveld, 1978; Bahel et al., 1987; Chen et al., 2004; Wu et al., 2007; Liu et
47 al., 2012; Ampratwum et al., 1999; Elagib et al 2000).

48 Studies on SD have mostly been based on limited ground stations (Vivar et al., 2014; Fan et al.,
49 2018; Yao et al., 2018), while SD is affected by atmospheric conditions, and it is difficult for a single
50 station to represent this over a large area, so there is a great need for a high-resolution SD data based on
51 satellite remote sensing for studies on solar radiation. Currently, geostationary and polar-orbiting satellite
52 data are widely used for high spatiotemporal resolution ground information tracking, and the Advanced
53 Himawari Imager (AHI) instrument, carried on board the new generation of geostationary satellites -
54 Himawari-8 and 9, has been widely used for the estimating radiation indicators different time scales
55 (Damiani et al, 2018; Hou et al., 2020; Letu et al., 2020; Tana et al., 2023). However, there are always
56 biases in the AHI radiation data and those inverted radiation indicators due to less ground measured
57 stations for validation and the susceptibility of remote sensing data to cloud and aerosols, while SD
58 reflects both solar radiation and cloud conditions and thus is well suited for inversion using remote
59 sensing radiation data, we can take advantage of the high spatiotemporal resolution of AHI to estimate

60 SD.

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

66

67 **2. Data and method**

68 **2.1 Remote sensing data**

69 The geostationary meteorological satellites, Himawari, was launched on 7 October 2014 from the
70 Japan Meteorological Agency (JMA) in Tanegashima, Japan, with its hypocenter located at 0.0°N and
71 140.7°E, approximately 35,800 km above the land surface. In comparison with other geostationary
72 satellites, Himawari AHI exhibits superior temporal and spatial resolution, reflection band sensitivity and
73 accuracy (Zhang et al., 2016). The AHI from Himawari-8 and 9 has 16 spectral channels covering the
74 visible to infrared range, with wavelengths ranging from 0.47 μm to 13.3 μm , providing a wealth of
75 spectral information (Bessho et al., 2016; Kim et al., 2018; Yu et al., 2019). The temporal and spatial
76 resolution of the land surface products provided by Himawari AHI is 10 minutes and 5 km respectively,
77 which is important for understanding the spatiotemporal variations on short time scales (Sawada et al.,
78 2019).

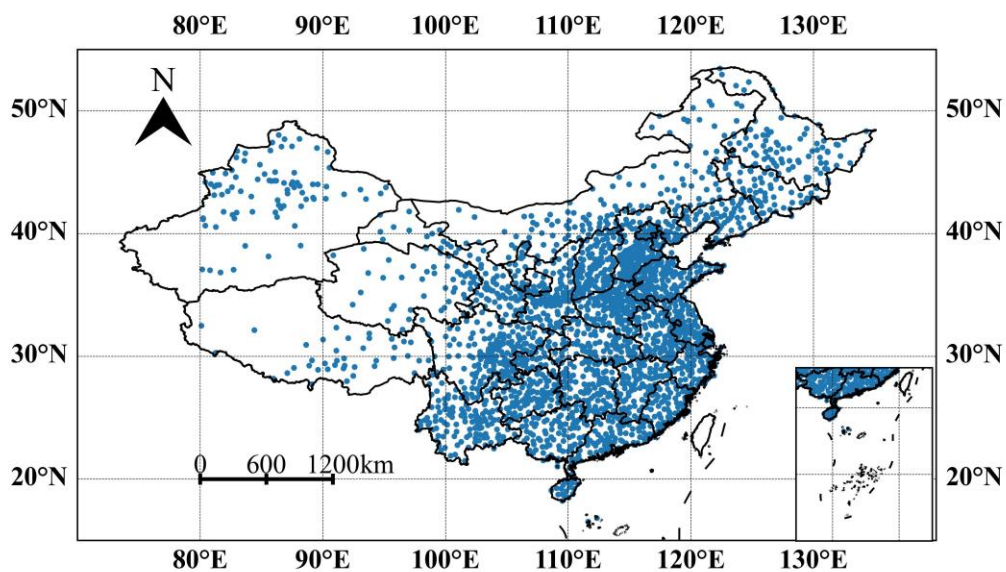
79 In this study, the Himawari AHI level 3 hourly shortwave radiation (5 km resolution) data from 1
80 January 2016 to 31 December 2023 was used for SD dataset construction, which calculated by plane-
81 parallel theory and considered the top of atmosphere (TOA) radiation by difference between the 300-
82 3000 nm solar shortwave band and reflected solar radiation by the atmosphere/land surface (Frouin et
83 al., 2007). This approach assumes that the effects of clouds and clear atmosphere can be decoupled,
84 which proved to be effective (Dedieu et al., 1987; Frouin and Rachel, 1995). In the event of a one-hour
85 interval being absent from the imagery, linear interpolation is conducted on each pixel of the missing
86 imagery based on the time series. In instances where the imagery is absent for a period exceeding one
87 hour, the day in question is excluded. We calculate the daily average shortwave radiation in China based
88 on China Standard Time (CST) using this hourly AHI shortwave radiation data.

89 The MCD19A2 is a MODIS Terra and Aqua combined multi-angle Implementation of Atmospheric

90 Correction (MAIAC) Land AOD gridded Level 2 product produced daily at 1 km pixel resolution, which
91 corrected for atmospheric gases and aerosols using a new MAIAC algorithm that is based on a time series
92 analysis and a combination of pixel- and image-based processing (Lyapustin et al., 2022). In this study,
93 the AOD at 550 nm in MCD19A2 from 2016 to 2023 were collected using Google Earth Engine (GEE)
94 (Gorelick et al., 2017).

95 2.2 Ground Measurements data

96 The ground measurements in CMA from 1 January 2016 to 31 December 2023 used to perform SD
97 estimation. The spatial coverage of Himawari covers 2380 CMA automatic meteorological stations in
98 China. The CMA performs quality control of the data, including spatiotemporal consistency checks and
99 manual corrections and adjustments before releasing the meteorological data (Moradi, 2009; Tang et al.,
100 2010). Although the quality of the ground-based measurements should have been controlled before
101 acquisition, there was still a need for a more stringent check on the quality of the data based on the
102 methodology of daily meteorological data reconstruction from CMA (Zhang et al., 2015). Figure 1 shows
103 the spatial distribution of 2380 meteorological. In this study, daily SD, vapor pressure (VAP), temperature,
104 wind speed and precipitation from the CMA automatic meteorological stations were used to fit and
105 validate the grid-dataset as well as to analyze the factors influencing the estimated performance,
106 respectively, and March-May was classified as spring, June-August as summer, September-November as
107 autumn and December-February as winter.



108

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

111

112 2.3 Model overview

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

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

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

$$R_a = 37.6 d_r (\omega_s \sin \phi \sin \delta + \cos \phi \cos \delta \sin \omega_s) \quad (2)$$

$$d_r = 1 + 0.033 \cos\left(\frac{2\pi}{365} \text{DOY}\right) \quad (3)$$

$$\delta = 0.4093 \sin\left(\frac{2\pi}{365} \text{DOY} - 1.39\right) \quad (4)$$

$$\omega_s = \arccos(-\tan \phi \tan \delta) \quad (5)$$

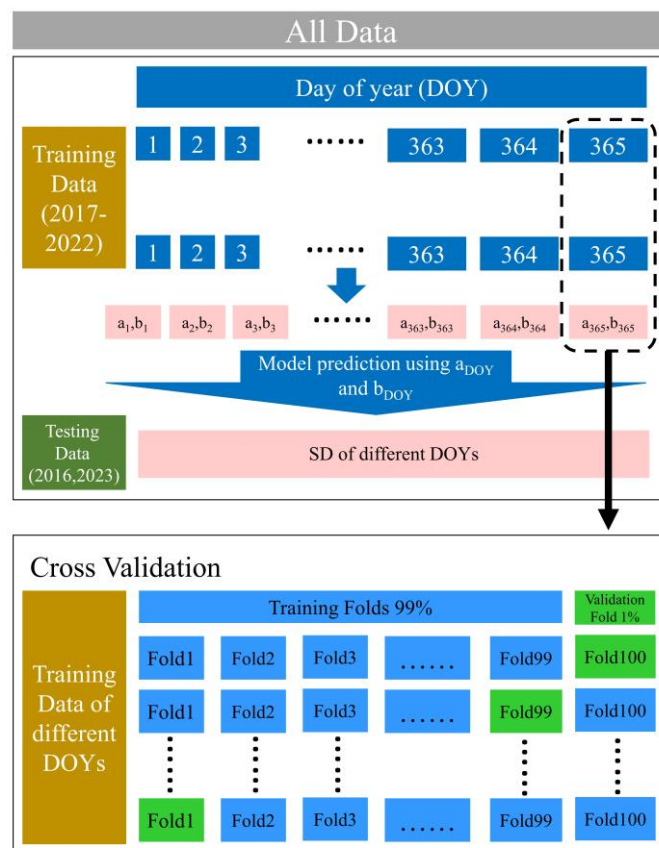
$$N = \frac{24}{\pi} \omega_s \quad (6)$$

120 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,
121 δ is the inclination angle of the sun, and DOY is the days of a year. We considered Himawari AHI level
122 3 hourly shortwave radiation as the R_s in this model, and SD of ground-based observation as a
123 validation of n , and the parameters a and b of Ångström-Prescott model were fitted using the least-squares
124 method.

125 2.4 Validation

126 We divided the original data into a training set (more than 5×10^6 grid cells during 2017-2022) and
127 a testing set (2016 and 2023). In order to identify the best Ångström-Prescott model and its corresponding

128 parameters, the performance of the Ångström-Prescott model on the training set (2017-2022) was
 129 evaluated using a 100-fold cross-validation (CV) approach, using a DOY-based CV strategy. In each
 130 iteration of each DOY, 99 folds were used as the training set and the remaining folds as the validation
 131 set, and the training and validation process was repeated 100 times to obtain the best model parameters
 132 a and b for each DOY. In addition, the 2016 and 2023 ground-based SD data were used as the test data
 133 to evaluate the generalization capability of the best model parameters a and b at each DOY. The specific
 134 process is shown in Figure 2. Pearson correlation coefficient (R) and root mean square error (RMSE)
 135 were calculated to evaluate the performance of the model.



136

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

138 **2.5 Methods of spatiotemporal variation analysis**

139 Empirical orthogonal function (EOF) decomposition is a significant technique used to investigate
 140 the geographical and temporal fluctuations in meteorological characteristics (Zhou et al., 2021). The
 141 variable field can be decomposed into two parts: a spatial function that remains constant across time and
 142 a temporal function that changes exclusively with time, thus the primary spatial and temporal variations
 143 of which are evident in the area with a significant contribution to the variance. The spatial function

144 component comprises several mutually independent and orthogonal spatial modes, also considered as
145 eigenvectors. The temporal function part consists of the projection of the spatial modes in time, which is
146 represented by the time coefficients. We used EOF to analyze spatiotemporal variations of the established
147 SD dataset in China, then the original variable field information and spatial coefficients is concentrated
148 in the first few modes.

149

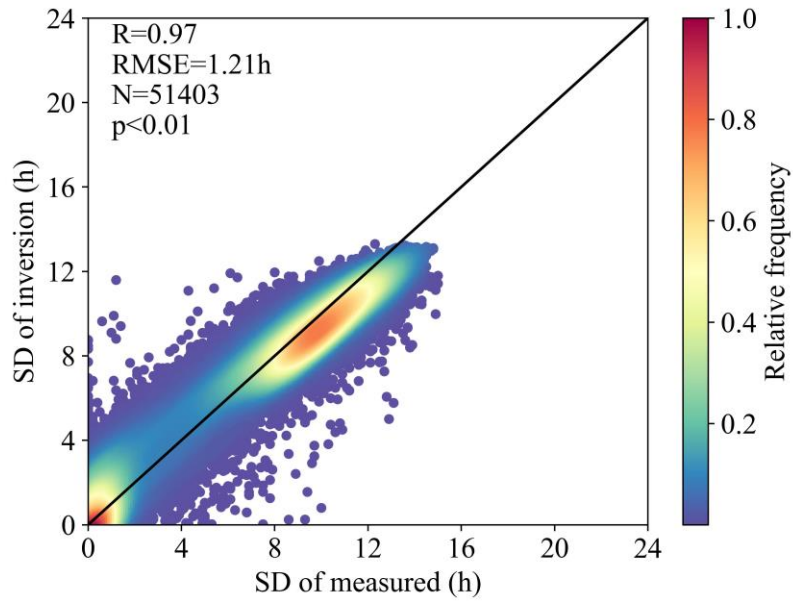
150 **3. Results**

151 **3.1 Evaluation of the training data**

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

163 Figure 5 shows the optimal Ångström-Prescott model parameters a and b at different DOYs. The
164 parameter a has an upward parabolic trend with DOY, with a local maximum value of 0.22 at DOY =
165 306 and a local minimum value of 0.13 at DOY = 351. Parameter b showed a significant "W"-shaped
166 variation with DOY, with a local maximum value of 0.74 at DOY = 146 and two local minimum values
167 of 0.66 and 0.63 at DOY = 99 and 351. In general, parameters a and b of Ångström-Prescott model are
168 characterized by more pronounced seasonal variations. Figure 6 shows the variation of the training set
169 evaluation indicator (R and RMSE) with DOY. More than half of the DOYs had R values greater than
170 the overall R value in Figure 3, but there were still 134 days with R values less than 0.97 and a minimum
171 value of 0.94 at DOY = 193. Meanwhile more than half of the DOYs have RMSE values less than the
172 overall RMSE values in Figure 3, but there are still 157 days with R values less than 1.2h, and again
173 there is a maximum value of 2.1h for RMSE at DOY = 193. The evaluation indicator for the training set

174 were not characterized by significant seasonal variations.

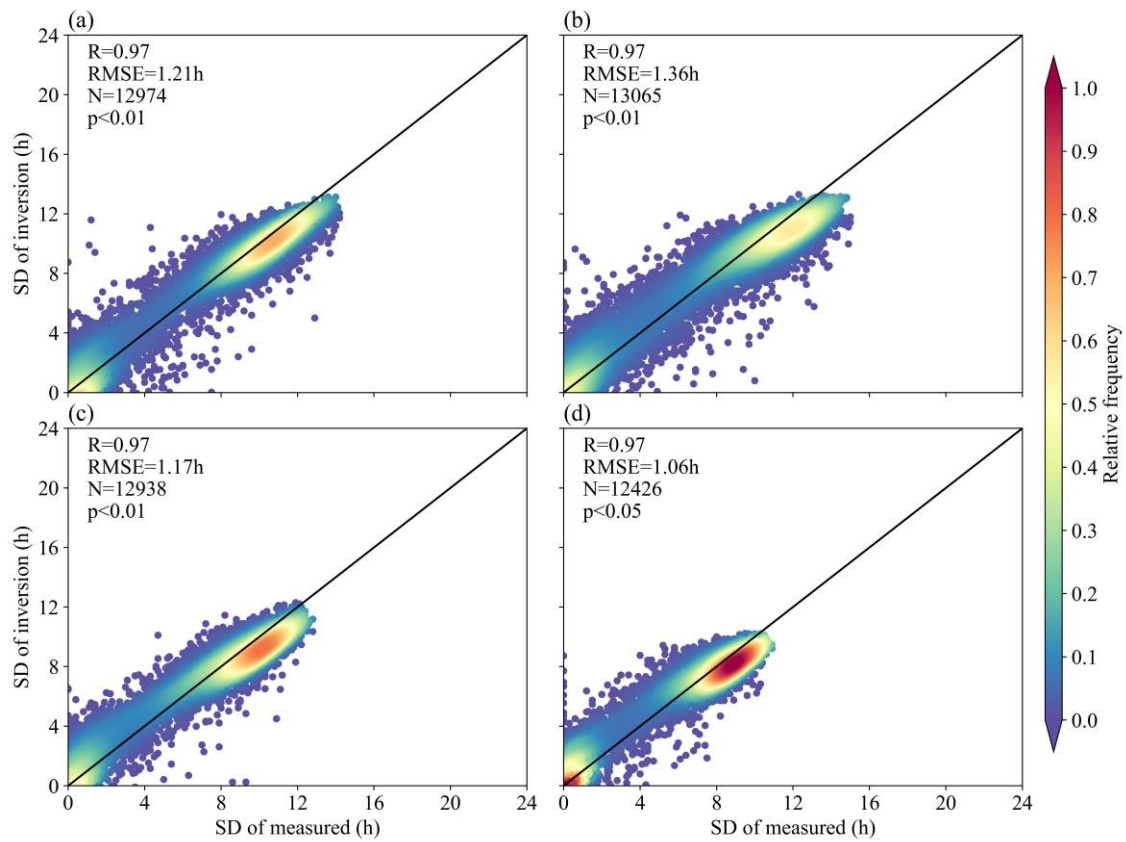


175

176

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

177

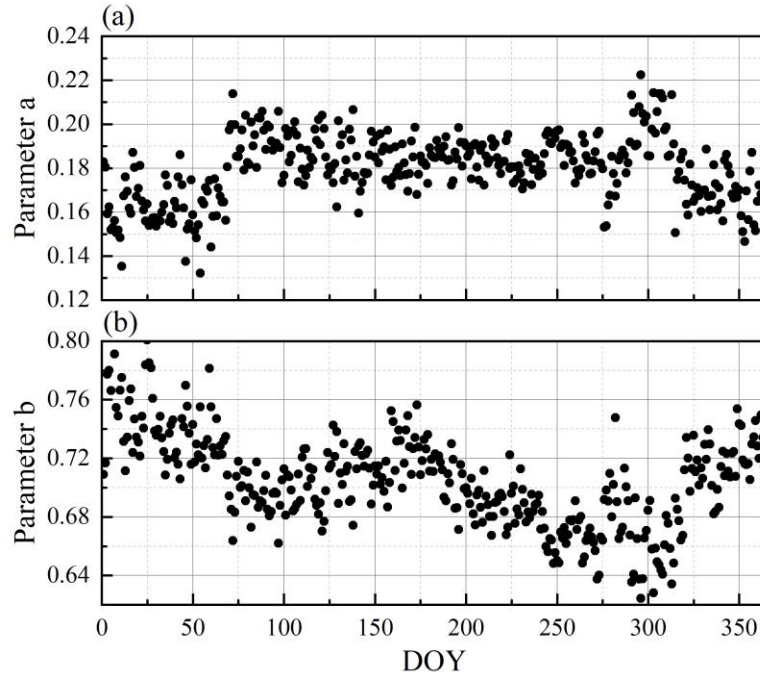


178

179 Figure 4. Estimation results of the CV sampling method in training set from different seasons ((a)

180

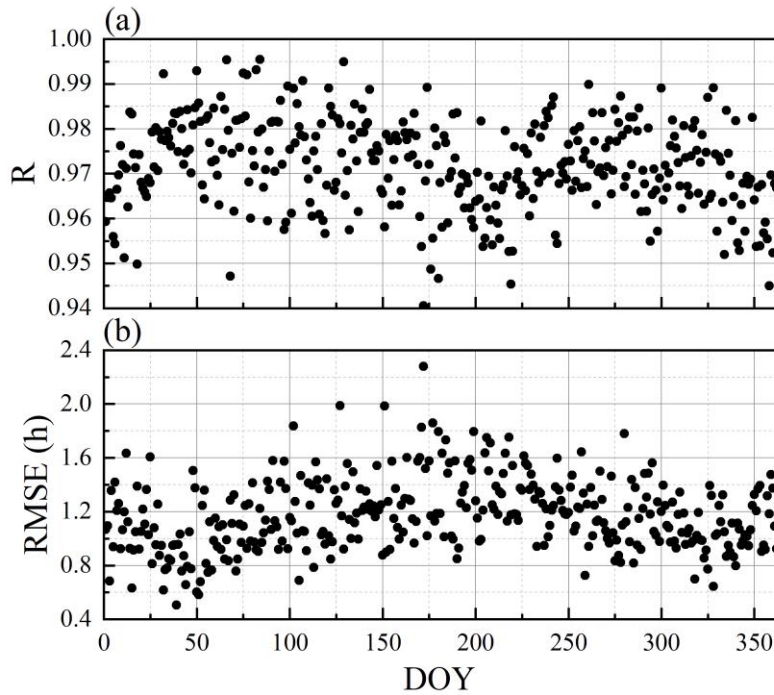
spring, (b) summer, (c) autumn, (d) winter).



181

182

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



183

184

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

185

different DOYs.

186

187 3.2 Evaluation of the testing data

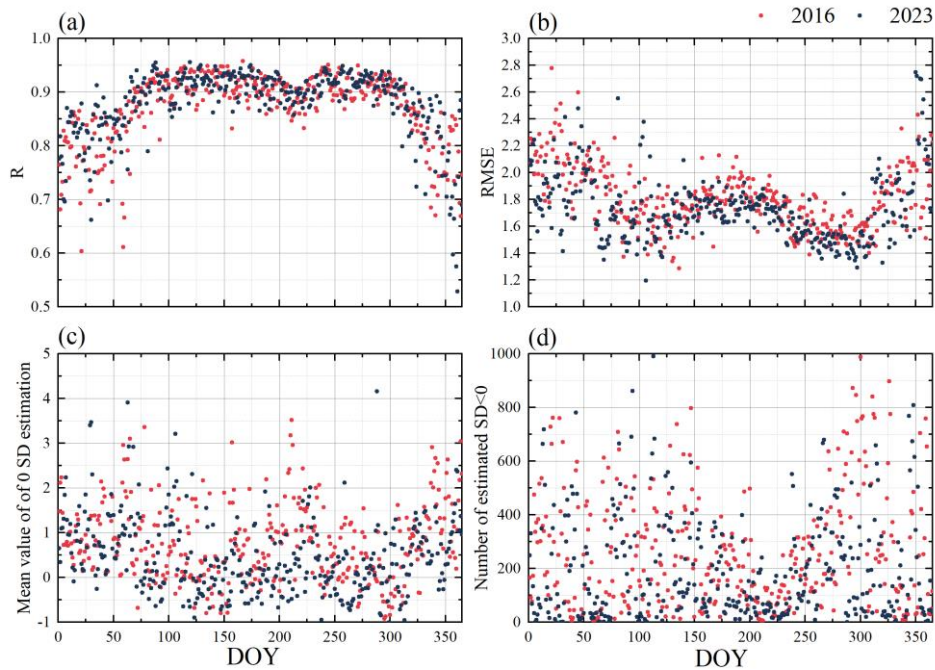
188

The different evaluation indicator for the test set (2016 and 2023) are given in Figure 7, respectively.

189

Figure 7(a) shows the R of 2016 and 2023, with the trends in these two years are essentially the identical,

190 with an "M" shape. The average R value for 2016 is 0.88, which is generally consistent with 2023. The
 191 minimum R value of 0.52 in 2023 (DOY=361) was lower than that of 0.60 in 2016 (DOY=21), but both
 192 occurred in winter. The trend of RMSE values for 2016 and 2023 is opposite to the R value, with the
 193 maximum and minimum RMSE values occurring in 2023 at 2.77 (DOY=355) and 1.19 (DOY=106),
 194 respectively. Figures 7(c) and (d) show the estimated performance of the 0 SD (no sunshine for the whole
 195 day) for the CMA meteorological stations in 2016 and 2023. Figure 7(c) shows the estimated mean values
 196 of 0 SD for different DOYs in 2016 and 2023, where the mean value in 2023 (0.49h) is smaller than in
 197 2016 (0.75h), with the maximum and minimum mean values still occurring in 2023 at 3.42 (DOY=211)
 198 and -0.75 (DOY=134), respectively. Figure 7(d) gives the number of estimated SD less than 0 for
 199 different DOYs in 2016 and 2023, of which there were more average daily estimated SDs less than 0 in
 200 2016 than in 2023, at 267/day, with the lowest value also occurring in 2016, at 997 for DOY=294. The
 201 bias in the 0SD estimation is linked to the over- and under-representation of its number. Changing all
 202 estimated SD less than 0 to 0 resulted in an improvement in their estimated performance (Figure 8), with
 203 2016 having a greater improvement than 2023 and having the greatest improvement with DOY=285.
 204



205

206

Figure 7. Estimated performance in testing set.

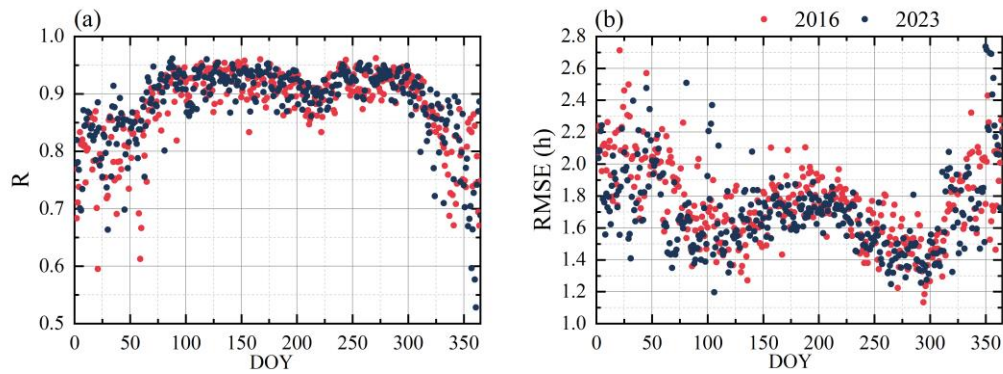
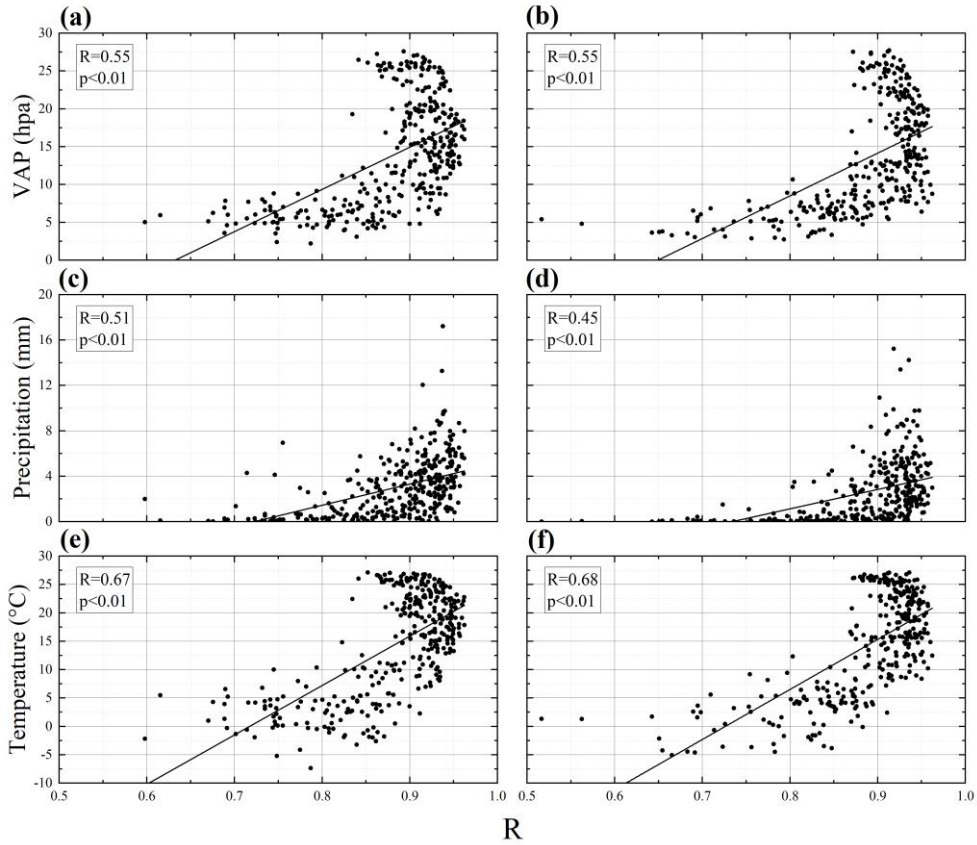


Figure 8. Estimated performance by changing all estimated SD less than 0 to 0 in testing set.

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 R values in Figure 8. The R values 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 estimation in 2016 and 2023 respectively, along with the annual average AOD and wind speed. On an annual scale, site and estimated SD are in better consistency in eastern and northern China, while both years have higher estimates in eastern China and lower estimates in northwestern and northeastern China, comparing the impact factors, higher wind speed and lower AOD in these areas both affected the SD estimation.



225

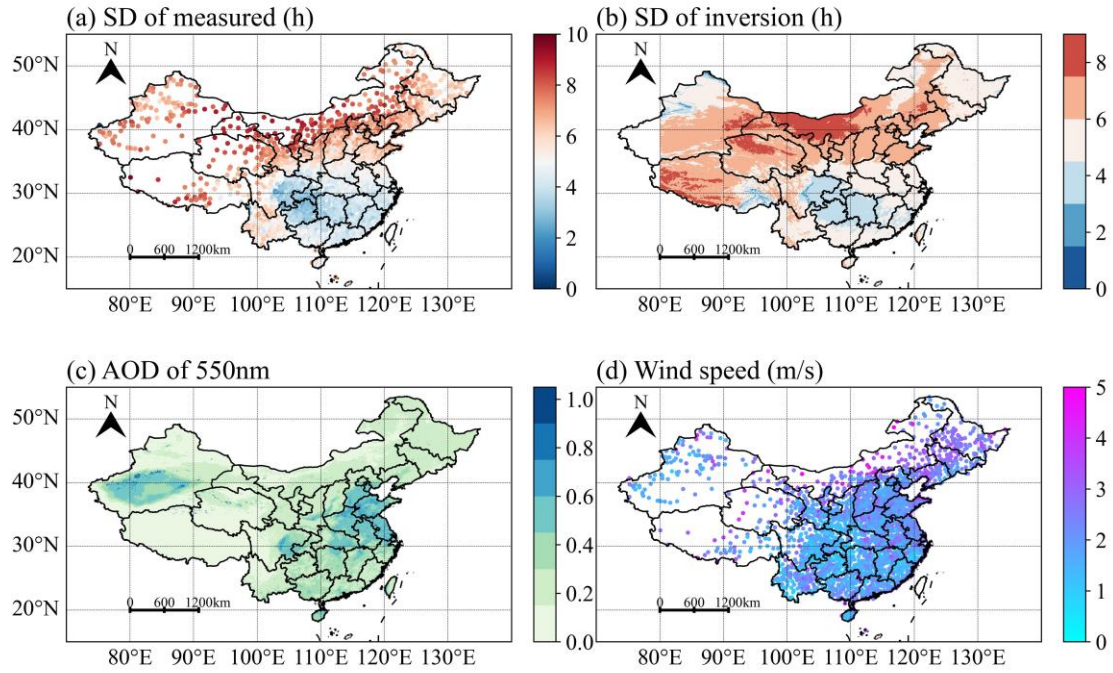
226 Figure 9. R values and different environmental factors (VAP (a, b), Precipitation (c, d), Temperature (e,
 227 f) correlations in 2016 (a, c and e) and 2023 (b, d and f).

228

229 **Table1.** Correlation coefficients between estimated performance and influencing factors in different
 230 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** |

231



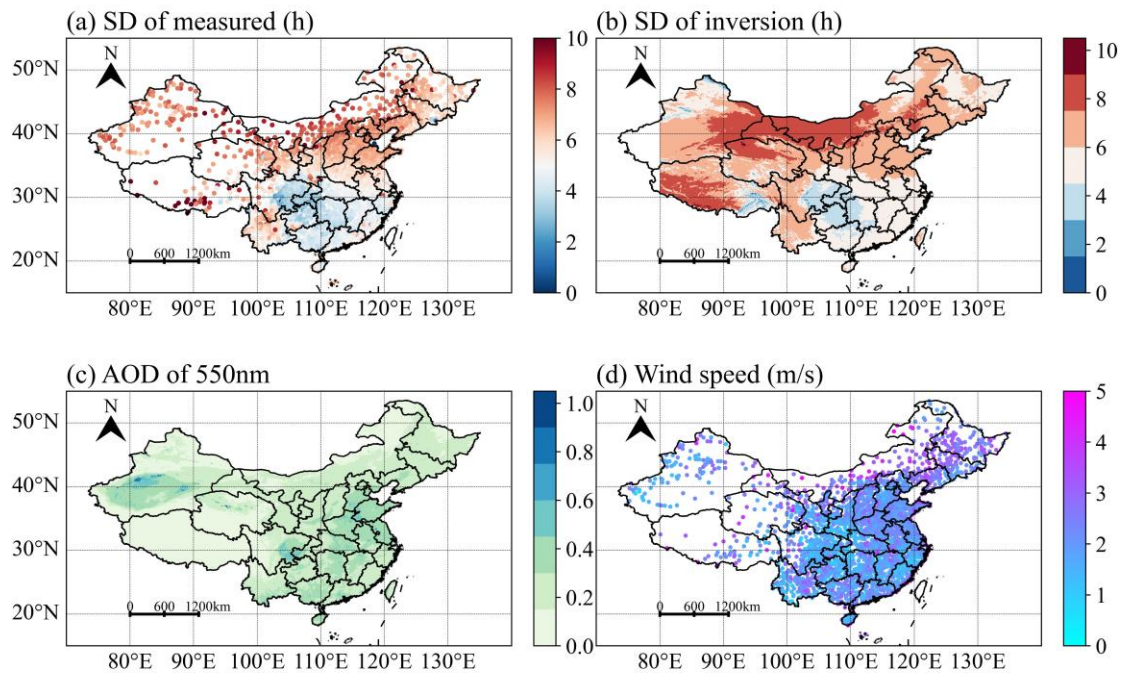
232

233

Figure 10. Comparison of annual average ground measurement (a) and Himawari (b) SD in 2016,

234

giving annual average AOD of 550nm (c) and the wind speed (d).



235

236

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

237

238 3.3 Effect of different environmental factors on SD estimation

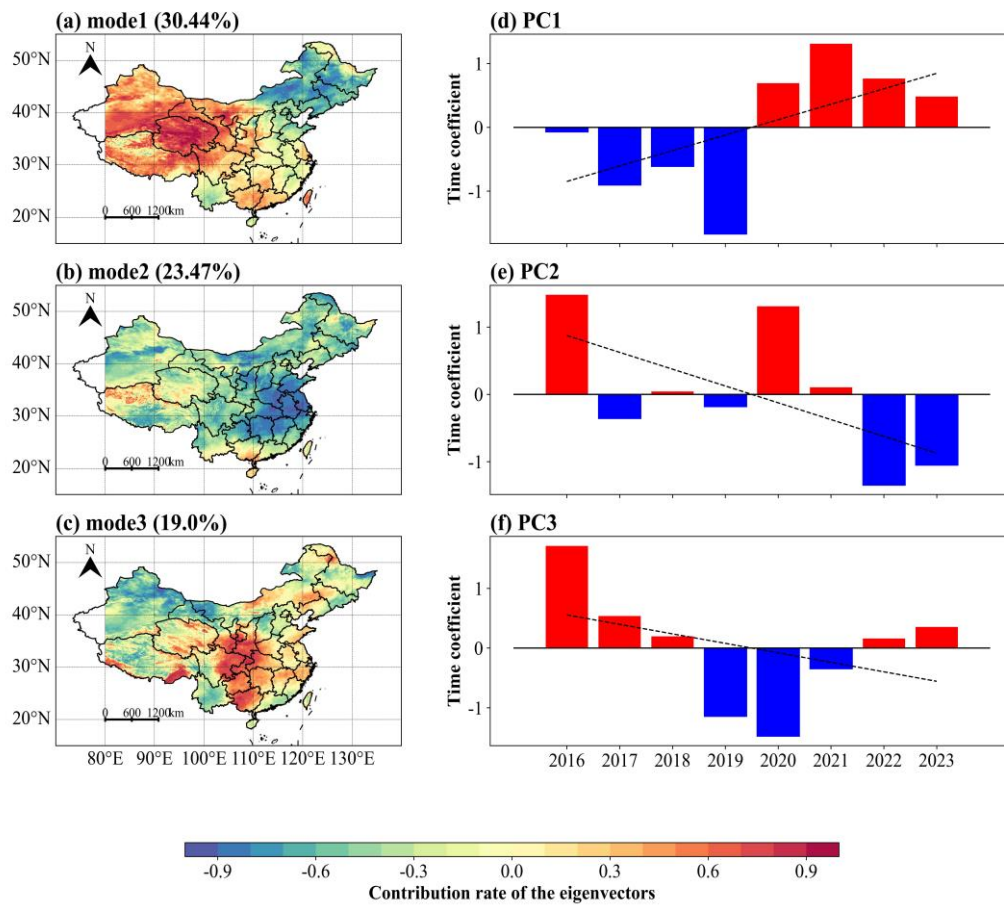
239

EOF analysis of mean annual SD grid-data in China from 2015-2023, the spatial variance

240

contribution rate of the eigenvectors in the first three EOF modes are shown in Figure 12, where the

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



253

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

255

three modes of SD.

256

257 **4. Discussion**

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

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

280 It is worth noting that there is a bias in the validation of the training and test data, where there is an
281 overestimation at OSD (Figure 3), may be the strong light in almost most of the area under a DOY leads
282 to Ångström-Prescott model larger parameters and over-estimation of a very small portion of the image
283 elements that contain aerosols, clouds and even precipitation. In addition, it also occurred in the test data
284 that the estimated SD was less than 0 (Figure 7 cd), because the thicker clouds, atmospheric aerosols and
285 water vapor in majority of the area on that day did not have much effect on the ground-based SD

286 instrument (the atmospheric longwave radiation contained in the direct radiation was not affected), but
287 had a significant effect on the AHI shortwave radiation data, resulting in SD less than 0. After changing
288 the image elements with SD less than 0 to 0, the validation results are still substantial (Figure 8),
289 indicating that this part of radiation is essentially less than the threshold for SD observations (120 W/m^2).
290 In conclusion, as our approach is carried out based on time series, it is unavoidable that we will encounter
291 input data that are not sensitive to different sky conditions. In the future, the use of relevant physical
292 precipitation models will be considered to simulate the precipitation process at different times of the day
293 based on the radiation data. This will enable us to estimate SD, and this aspect of the Ångström-Prescott
294 model will be improved subsequently.

295 We found that temperature, moisture conditions, wind speed and atmospheric pollutants all
296 influence the SD estimation, with temperature having the greatest effect in temporal variation and wind
297 speed having a stronger effect in spatial variation compared with AOD. However, we believe that the
298 effects of these environmental factors are not independent, but are the result of interaction (Tang et al.,
299 2022). In densely populated and economically developed areas (eastern and southern China), where
300 pollutant levels are higher and increased wind speed accelerates their dispersion, this regulatory
301 mechanism is enhanced with increasing pollutants (O'Dowd et al., 1993; Wang et al., 2014). An increase
302 or decrease in wind speed affects the rate of diffusion of water vapor and pollutants in the air, which in
303 turn affects atmospheric transparency and ultimately the SD estimation. However, the effect of
304 temperature on SD estimation in this study are not consistent with some previous studies (Tang et al.,
305 2022; Feng et al., 2019; Ren et al., 2017), which suggests that the relationship between SD and
306 temperature and relative humidity is complex and needs to be further determined in future studies.

307 The EOF method analysis of mean annual SD declare that it decreases from western to northeast
308 China, which is consistent with the Tang et al. (2022) and Xiong et al. (2020), suggesting that the pattern
309 of industrial development between western to eastern China is affecting radiation levels to some extent.
310 The time coefficients of EOF show that there is a certain degree of increase in SD in recent years, which
311 is closed to long-term SD analysis from Tang et al. (2022). This trend may be related to global climate
312 change (Josefsson and Landelius, 2000), because of the variation in wind speeds due to global warming
313 has resulted in decreased cloud dissipation across mainland China (Xiong et al., 2020). In addition, the
314 decrease in human activities in recent years (Liu et al., 2020) has also contributed to a weakening of the
315 urban rain island effect and aerosols (Glantz et al., 2006), and it appears that the latter factor is more

316 influential from this study. However short-term reductions in human activity cannot become the norm,
317 and sunshine duration are bound to fluctuating changes due to the acceleration of the hydrological cycle.

318

319 **5. Data availability**

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

321 **6. Conclusion**

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

330

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

334

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

337

338 **Financial support.** This research was supported by the National Key Research and Development
339 Program of China (grant no. 2023YFE0122200), the National Nature Sciences Foundation (grant no.
340 42075193).

341

342 **Reference**

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

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

348 Ångström A.: Solar and terrestrial radiation. Report to the international commission for solar research on
349 actinometric investigations of sola and atmospheric radiation. Q J Roy Meteor Soc., 50:121–6.

350 <https://doi.org/10.1002/qj.49705021008>, 1924.

351 Bahel, V., Bakhsh, H. and Srinivasan, R.: A correlation for estimation of global solar radiation. *Energy*,
352 12, 131-135. [https://doi.org/10.1016/0360-5442\(87\)90117-4](https://doi.org/10.1016/0360-5442(87)90117-4), 1987.

353 Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., Kumagai, Y., Miyakawa, T., Murata,
354 H., Ohno, T., Okuyama, A., Oyama, R., Sasaki, Y., Shimazu, Y., Shimoji, K., Sumida, Y., Suzuki, M.,
355 Taniguchi, H., Tsuchiyama, H., Uesawa, D., Yokota, H. and Yoshida, R.: An introduction to Himawari-
356 8/9—Japan's new-generation geostationary meteorological satellites. *J. Meteorol. Soc. Japan., Ser. II*,
357 94(2), 151-183. <https://doi.org/10.2151/JMSJ.2016-009>, 2016.

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

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

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

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

370 Dedieu, G., P. Y. Deschamps, and Y. H. Kerr.: Satellite estimation of solar irradiance at the surface of the
371 earth and of surface albedo using a physical model applied to Meteosat Data. *J. Appl. Meteorol. Climatol.*,
372 26.1: 79-87. [https://doi.org/10.1175/1520-0450\(1987\)026<0079:SEOSIA>2.0.CO;2](https://doi.org/10.1175/1520-0450(1987)026<0079:SEOSIA>2.0.CO;2), 1987.

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

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

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

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

383 Frouin, R. and Murakami, H.: Estimating photosynthetically available radiation at the ocean surface from
384 ADEOS-II global imager data. *J Oceanogr.*, 63, 493-503. <https://doi.org/10.1007/S10872-007-0044-3>,
385 2007.

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

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

392 Glantz, P., Nilsson, D.E. and Hoyningen-Huene, W.V. Estimating a relationship between aerosol optical
393 thickness and surface wind speed over the ocean. *Atmos. Chem. Phys.*, 6, 11621-11651.

394 <https://doi.org/10.5194/ACPD-6-11621-2006>, 2006.

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

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

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

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

407 Kim, B., Lee, K., Jee, J. and Zo, I.: Retrieval of outgoing longwave radiation at top-of-atmosphere using
408 Himawari-8 AHI data. *Remote Sens Environ.*, 204, 498-508. <https://doi.org/10.1016/J.RSE.2017.10.006>,
409 2018.

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

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

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

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

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

426 Liu, Q., Sha, D., Liu, W., Houser, P.R., Zhang, L., Hou, R., Lan, H., Flynn, C., Lu, M., Hu, T. and Yang,
427 C. Spatiotemporal Patterns of COVID-19 Impact on Human Activities and Environment in Mainland
428 China Using Nighttime Light and Air Quality Data. *Remote. Sens.*, 12, 1576.
429 <https://doi.org/10.3390/rs12101576>, 2020.

430 Liu, X., Mei, X., Li, Y., Wang, Q., Zhang, Y. and Porter, J. R.: Variation in reference crop
431 evapotranspiration caused by the Ångström–Prescott coefficient: Locally calibrated versus the FAO
432 recommended. *Agric Water Manag.*, 96(7), 1137-1145. <https://doi.org/10.1016/J.AGWAT.2009.03.005>,
433 2009.

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

437 Moradi, I.: Quality control of global solar radiation using sunshine duration hours. *Energy*, 34, 1-6.

438 <https://doi.org/10.1016/J.ENERGY.2008.09.006>, 2009

439 O'Dowd, C. D. and Smith, M. H.: Physicochemical properties of aerosols over the northeast Atlantic:
440 Evidence for wind-speed-related submicron sea-salt aerosol production. *J. Geophys. Res. Atmos.*, 98(D1),
441 1137-1149. <https://doi.org/10.1029/92JD02302>, 1993.

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

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

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

450 Rietveld, M. R.: A new method for estimating the regression coefficients in the formula relating solar
451 radiation to sunshine. *Agric. Meteorol.*, 19(2-3), 243-252. [https://doi.org/10.1016/0002-1571\(78\)90014-](https://doi.org/10.1016/0002-1571(78)90014-6)
452 [6](https://doi.org/10.1016/0002-1571(78)90014-6), 1978.

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

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

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

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

465 Vivar, M., Fuentes, M., Norton, M., Makrides, G. and Bustamante, I.D.: Estimation of sunshine duration
466 from the global irradiance measured by a photovoltaic silicon solar cell. *Renew. Sust. Energ. Rev.*, 36,
467 26-33. <https://doi.org/10.1016/J.RSER.2014.04.045>, 2014.

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

471 Wu, G., Liu, Y. and Wang, T.: Methods and strategy for modeling daily global solar radiation with
472 measured meteorological data—A case study in Nanchang station, China. *Energy Convers. Manage.*, 48(9),
473 2447-2452. <https://doi.org/10.1016/J.ENCONMAN.2007.04.011>, 2007.

474 Xia, X.: Spatiotemporal changes in sunshine duration and cloud amount as well as their relationship in
475 China during 1954–2005. *J. Geophys. Res. Atmos.*, 115(D7). <https://doi.org/10.1029/2009JD012879>,
476 2010.

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

480 Yao, W., Zhang, C., Wang, X., Zhang, Z., Li, X. and Di, H.: A new correlation between global solar
481 radiation and the quality of sunshine duration in China. *Energy Convers. Manage.*, 164, 579-587.

482 <https://doi.org/10.1016/J.ENCONMAN.2018.03.037>, 2018.

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

486 Yu, Y., Shi, J., Wang, T., Letu, H., Yuan, P., Zhou, W. and Hu, L.: Evaluation of the Himawari-8
487 Shortwave Downward Radiation (SWDR) Product and its Comparison With the CERES-SYN, MERRA-
488 2, and ERA-Interim Datasets. *IEEE J Sel Top Appl Earth Obs Remote Sens*, 12, 519-532.
489 <https://doi.org/10.1109/JSTARS.2018.2851965>, 2019.

490 Zhang, J., Zhao, L., Deng, S., Xu, W. and Zhang, Y.: A critical review of the models used to estimate
491 solar radiation. *Renew. Sust. Energ. Rev.*, 70, 314-329. <https://doi.org/10.1016/J.RSER.2016.11.124>,
492 2017.

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

496 Zhang, X., Liang, S., Wild, M. and Jiang, B.: Analysis of surface incident shortwave radiation from four
497 satellite products. *Remote Sens Environ.*, 165, 186-202. <https://doi.org/10.1016/J.RSE.2015.05.015>,
498 2015.

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

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