https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

Open Access Earth System Science Data Discussions

A 10 km daily-level ultraviolet radiation predicting dataset based on machine learning models in China from 2005 to 2020

Yichen Jiang[1], Su Shi[1], Xinyue Li[1], Chang Xu[1], Haidong Kan[1, 2], Bo Hu[3, *], Xia Meng[1, 2, **]

[1]School of Public Health, Key Laboratory of Public Health Safety of the Ministry of Education and Key Laboratory of Health Technology Assessment of the Ministry of Health, Fudan University, Shanghai, 200032, China

[2]Shanghai Key Laboratory of Meteorology and Health IRDR International Center of Excellence on Risk Interconnectivity and Governance on Weather/Climate Extremes Impact and Public Health WMO/IGAC MAP-AQ Asian Office Shanghai, Fudan University, Shanghai, China

[3]State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, People's Republic of China

Correspondence to: Xia Meng (mengxia@fudan.edu.cn); Bo Hu (hub@post.iap.ac.cn)

Abstract. Ultraviolet (UV) radiation is closely related to health, but limited measurements hindered further investigation of its health effects in China. Machine learning algorithm has been widely used in predicting environmental factors with high accuracy, but limited studies have done for UV radiation. This study aimed to develop UV radiation prediction model based on random forest method, and predict UV radiation at daily level and 10 km resolution in mainland China in 2005–2020. A random forest model was employed to predict UV radiation by integrating ground UV radiation measurements from monitoring stations and multiple predictors, such as UV radiation data from satellite. Missing data of satellite-based UV radiation was filled by three-day moving average method. The model's performance was evaluated through multiple cross-validation (CV) methods. The overall R^2 (root mean square error, RMSE) between measured and predicted UV radiation from model development and model 10-fold CV was 0.97 (15.64 W m^{-2}) and 0.83 (37.44 W m^{-2}) at daily level, respectively. The model with OMI EDD performed higher predicting accuracy than the one without it. Based on predictions of UV radiation at daily level and 10 km spatial resolution and nearly 100% spatiotemporal coverage, we found UV radiation increased by 4.20% while $PM_{2.5}$ levels decreased by 48.51% and O_3 levels rose by 22.70% in 2013–2020, suggesting a potential correlation among these environmental factors. Uneven spatial distribution of UV radiation was found to be associated with factors such as latitude, elevation, meteorological factors and seasons. The eastern areas of China posed higher risk with both high population density and UV radiation intensity. Based on machine learning algorithm, this study generated a gridded dataset characterized by relatively high precision and extensive spatiotemporal coverage of UV radiation, which demonstrates the spatiotemporal variability of UV radiation levels in China and can facilitate health-related research in the future. This dataset is currently freely available at https://doi.org/10.5281/zenodo.10884591 (Jiang et al., 2024).

1 Introduction

UV radiation stands as one of the most crucial environmental factors closely related to human health (Brenner and Hearing,

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

2008; Narayanan et al., 2010). Previous studies have confirmed hazardous effects of UV radiation on skin cancer (Griffin et al., 2023; Vienneau et al., 2017) and reported inconsistent results regarding the adverse effects of UV radiation on eye diseases (Lagreze et al., 2017; Tian et al., 2018; Wolffsohn et al., 2022) or health benefits under moderate UV radiation (Boscoe and Schymura, 2006; Vopham et al., 2017; Swaminathan et al., 2019). More studies are on call to ascertain the effects of UV radiation on human health. However, lacking exposure data of UV radiation with high accuracy is one of the reasons that hinders such health investigations.

The exposure assessment methods used in previous health studies of UV radiation mainly include the following ones. Firstly, UV index is a frequently used proxy of UV radiation in epidemiological studies (Thayer, 2014; Marson et al., 2021; Walls et al., 2013). It predicts UV radiation levels on a scale ranging from 1 to 11+. Although the UV index is easy to interpret, converting continuous measurements of UV radiation to the UV index results in the loss of numerical information. Secondly, satellite remote sensing data are also used to estimate UV radiation exposure. For example, erythemal UV irradiance from the Total Ozone Mapping Spectrometer (TOMS), despite stands as one of the initial instruments for evaluating the UV radiation backscattered by the Earth's atmospheric layers, it exhibits lower spatial resolution of 50 km×50 km, and it has limited accuracy (Boscoe and Schymura, 2006; Mohr et al., 2008; Lin et al., 2012; Zhou et al., 2019). Additionally, erythemally daily dose (J/m^2) retrieved from the Ozone Monitoring Instrument (OMI EDD) could also be utilized to evaluate UV radiation exposure level with higher spatiotemporal resolution and was employed in the United States to represent ground UV radiation levels and identify hotspots for skin cancer (Zhou et al., 2019; Deng et al., 2021). However, missing values of OMI EDD data are non-random and have been increased since 2008 that poses a challenge to the accuracy of exposure assessment in epidemiological studies (Mcpeters et al., 2015). Thirdly, in some studies, participants were required to wear personal dosimeters for individual measurements (Stump et al., 2023; Grandahl et al., 2018). Although the data quality is high, the cost is substantial, making it difficult to be applied in large-population studies. Therefore, UV radiation data with higher accuracy and spatiotemporal resolution are needed to support further exposure assessment.

With the enrichment of data resource and improvement of computing powers, machine learning algorithm can integrate data from multiple sources to predict environmental factors with high quality (Chen et al., 2021; Zhu et al., 2022; Liu et al., 2022), but limited applications in predicting UV radiation have been done. Models used in UV radiation prediction in earlier years were empirical model or statistical model (González-Rodríguez et al., 2022; Vopham et al., 2016; Pei and He, 2019; Liu et al., 2017). In recent years, limited studies started to employ machine learning algorithm to predict UV radiation in China (Wu et al., 2022; Qin et al., 2020). The spatiotemporal resolution of predictions of one study was relatively low (0.50°×0.625°), and the other one produced UV radiation predictions with significant missingness due to missing data of one predictor (aerosol optical depth from satellite) that may lead to seasonal bias of UV radiation assessment. In addition, these studies did not include direct measurements of UV radiation from satellite, such as OMI EDD, which has been proven to be an effective predictor for

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

UV radiation evaluation (Zhou et al., 2019; Deng et al., 2021). Satellite-based measurements can be used as one of the "real" measurements of UV radiation, which can help constrain overfitting of model in spatiotemporal extrapolation. Overall, further studies are highly needed to add more evidence in model development of UV radiation with advanced algorithms and comprehensive predictors.

Hence, this study aimed to develop a random forest model, one of the machine learning algorithms, to predict UV radiation in mainland China at a daily level and a spatial resolution of 10 km in 2005–2020. Multiple predictors, including satellite-based UV radiation, UV radiation simulations and parameters form re-analysis meteorological dataset, and other variables, were included in model development. What's more, missingness of satellite-based UV radiation were filled to improve spatial coverage of final UV radiation predictions. Finally, based on the predictions with relatively high spatiotemporal resolution and a long time period, temporal and spatial trends as well as hotspots of UV radiation were identified in mainland China.

2 Data and methods

2.1 Data

2.1.1 Ground UV radiation measurements

The Chinese Ecosystem Research Network (CERN) has been observing UV radiation since 2004 (Liu et al., 2017). The monitoring data is accessible online via http://www.cern.ac.cn/. Hourly monitoring data on UV radiation from 40 ground-based stations between 2005 and 2015 and 36 ground-based stations between 2016 and 2020 were collected from CERN (Fig. 1). Those stations cover eight ecological land cover types across China, including urban, agricultural, grassland, forest, lakes, bays, wetlands, and deserts. Daily UV radiation values were calculated by adding up 24 hourly UV radiation values per day. Days with continuous 2 hourly missing or unavailable UV radiation values were excluded.

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

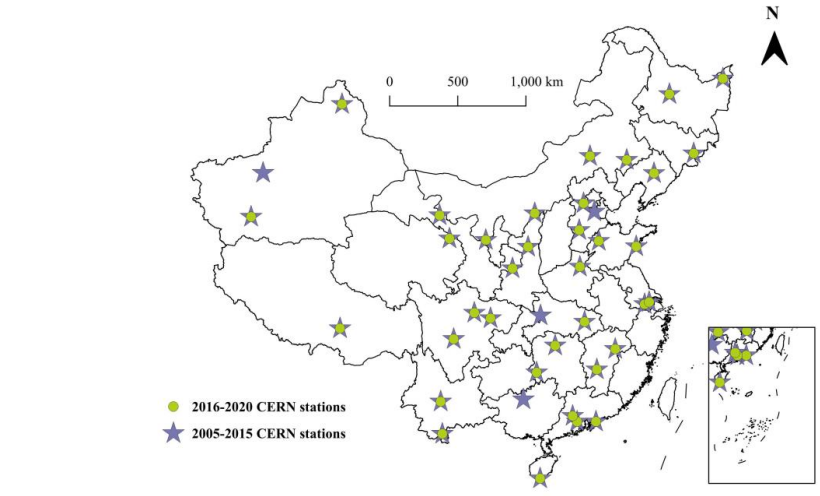


Figure 1. Spatial distributions of CERN stations monitoring UV radiation in China in 2005–2020.

2.1.2 Predictors directly related to UV radiation

In this study, the Level-2 OMI EDD (v.003) data were utilized as the main predictor for UV radiation, which has a temporal resolution of daily level and a spatial resolution of 0.25°×0.25° (Zhou et al., 2019). OMI EDD represents the overall amount of UV radiation that can cause sunburn during a day. The other predictor was downward UV radiation at the surface from the fifth generation European Center for Medium-Range Weather Forecasts Reanalysis on single levels (ERA-5 UV), with a temporal resolution of hourly level and a spatial resolution of 0.25°×0.25° (https://cds.climate.copernicus.eu/). The daily ERA-5 UV was obtained by adding the data over 24 hours for each day. OMI EDD and ERA-5 UV with spatial resolution of 0.25°×0.25° were interpolated to the 10 km grid cells employing the inverse distance weighted (IDW) method.

2.1.3 Meteorological parameters

Meteorological parameters may affect UV radiation were extracted from multiple ERA-5 products (https://cds.climate.copernicus.eu/) according to previous studies (Dieste-Velasco et al., 2023; Hu et al., 2010). Among them, total cloud cover, total column water vapour, and forecast albedo were extracted from single-level ERA-5 product with a temporal resolution of hourly level and a spatial resolution of 0.25° × 0.25°; and relative humidity were extracted from pressure-level ERA-5 product at 1000 hPa with a temporal resolution of hourly level and a spatial resolution of 0.25° × 0.25°. Total precipitation and temperature at 2m were extracted from ERA5-Land product with a temporal resolution of hourly level and a spatial resolution of 0.1° × 0.1°. Regarding temporal resolution, the hourly data were converted to daily mean data by averaging the 24-hour data for each day. Concerning spatial resolution, IDW method was used to interpolate the meteorological parameters to 10 km grid cells.

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

2.1.4 Other predictor variables

Other predictor variables that were incorporated including elevation, solar zenith angle (SZA), ground ozone (O_3) concentrations, and aerosol optical depth (AOD) that could affect UV radiation levels according to previous studies (Santos et al., 2011; Habte et al., 2019). Elevation data were derived from the Advanced Spaceborne Thermal Emission and Radiometer (ASTER) Global Digital Elevation Map (GDEM), which has a spatial resolution of 30 m (https://asterweb.jpl.nasa.gov/GDEM.asp). SZA data were obtained from Aqua (MYD06_L2) with a temporal resolution of daily level and a spatial resolution of 5 km (https://search.earthdata.nasa.gov). O_3 data is the maximum daily 8 h average (MDA8) O_3 concentrations predicted based on a random forest model at a daily level and spatial resolution of 1 km × 1 km in China (Meng et al., 2022). This study also included AOD data from the Multi-Angle Implementation of Atmospheric Correction (MAIAC AOD) algorithm based on the Moderate Resolution Imaging Spectroradiometer (MODIS) with a temporal resolution of daily level and a spatial resolution of 1 km (Shi et al., 2022; Meng et al., 2021). MAIAC AOD values for cloud contamination or land covered by snow were cleaned based on quality assurance (QA) flags. For comparing long-term trend of UV radiation and air pollution, fine particulate matter ($PM_{2.5}$) data was also included in this study, which was predicted using a random forest model with a daily level and spatial resolution of 1 km × 1 km in China (Meng et al., 2021; Shi et al., 2023a; Shi et al., 2023b). Elevation and SZA was spatially joined and averaged into 10 km grid cells. O_3 and MAIAC AOD were obtained by matching 1 km grid cells with 10 km grid cells, and then calculating the mean value of the data within 10 km grid cells.

2.2 Methods

2.2.1 Model development

In recent years, machine learning algorithms have been widely used to predict environmental factors due to their flexibility and excellent data processing capabilities (Corrêa, 2023; Wu et al., 2022). This study utilized the random forest, one of machine learning algorithms, to develop a model for predicting UV radiation in China during 2005–2020. The dependent variable was the daily ground measured UV radiation, while the independent variables included OMI EDD, ERA-5 UV, elevation, SZA, O_3, MAIAC AOD as well as meteorological parameters including total cloud cover, relative humidity, total column water vapour, forecast albedo, total precipitation, and temperature at 2 m. Random forest improves the overall prediction performance by building multiple decision trees and combining their results (Breiman, 2001). It uses bootstrap sampling, which is to draw different subsamples from the original dataset with replacement, as the training data for each decision tree. During the training process, each decision tree makes predictions for the input data, and the final result of the random forest is obtained by averaging the predictions from all trees. Importance ranking of all predictors is a method to evaluate the contribution of different predictors in a model, and it is also one of the advantages of random forest. The importance of a predictor is measured

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

by randomly permuting its values and comparing the accuracy drop between the predictions before and after the permutation. Model development was implemented using the "Rborist" package in R version 3.6.3.

OMI EDD serves as a measurement of UV radiation from satellite, but it has non-random missing values due to cloud cover and technological issue of OMI since 2008, with an averaged missing rate of 23.04% (3.03%–35.29%) during all years in the study period (Table A1). To address this issue, a three-day moving average method was used to fill the gap when the OMI EDD was missing, which reduced the missing rate of OMI EDD to 0.62%.

2.2.2 Model validation

The model performance was tested through overall 10-fold cross-validation (CV), temporal 10-fold CV, spatial 10-fold CV and by year 10-fold CV, which was a stricter temporal CV. Overall 10-fold CV was achieved by randomly dividing the dataset into ten parts, with nine parts were used to train a random forest model and one part was used as a test dataset for predictions. This process was replicated ten times, and the measurements were compared with the corresponding predictions. Temporal 10-fold CV could be used to evaluate the performance of the model in temporal extrapolation. Each time, 90% of days was randomly selected and data on these days were used to develop a training model to predict UV radiation for the remaining 10% of dates, and this process was repeated ten times. In a similar manner, spatial 10-fold CV could be used to evaluate the performance of the model in spatial extrapolation. Each time, 90% of sites data was randomly selected to develop a training model to predict UV radiation for the remaining 10% of sites, and the procedure was repeated ten times. In contrast to previous studies that generally only employed a temporal 10-fold CV, this study performed by year 10-fold CV simultaneously, which took out an entire year of data as testing data each time (Meng et al., 2021). This process was repeated 16 times, given the availability of 16 years of UV radiation data for model development. To indicate the precision of the prediction, the study calculated the regression R^2, CV R^2 and root mean square error (RMSE, defined as the square root of the average of the squared differences between the predictions and measurements) between UV radiation measurements and predictions.

3 Results

3.1 Description of UV radiation measurements

Table 1 summarizes the statistical descriptions of averaged daily mean for UV radiation measurements from CERN spanning the years 2005 to 2020. The mean annual value from 2005 to 2020 of UV radiation at monitoring stations was 168.40 W m^{-2}, with a standard deviation of 91.39 W m^{-2}. During the 16-year period, the minimum level of 155.46 W m^{-2} was recorded in 2010, while the maximum UV radiation level of 190.10 W m^{-2} was recorded in 2020 with an increase of 22.28% compared to 2010. The UV radiation level fluctuated between 2005 and 2012, but the overall trend was relatively stable. From 2013 to

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

160 2020, there was a clearly increasing trend in UV radiation, which increased by 18.66% during this period.

Table 1. Statistical descriptions of UV radiation measurements from ground monitoring stations in CERN in China during 2005–2020

Year	Mean (W m^{-2})	Standard deviation (W m^{-2})	P25 (W m^{-2})	Median (W m^{-2})	P75 (W m^{-2})
2005	160.62	81.07	94.35	153.57	160.62
2006	158.34	80.56	94.20	149.90	214.90
2007	159.54	82.99	91.81	150.41	220.21
2008	162.39	83.09	93.49	153.60	223.16
2009	159.64	82.65	91.46	152.20	222.60
2010	155.46	81.73	88.56	144.91	215.80
2011	160.95	84.37	90.11	152.60	223.50
2012	159.65	85.38	88.75	153.60	221.80
2013	160.21	82.87	92.00	149.93	221.50
2014	160.87	82.41	94.06	152.90	221.50
2015	170.96	91.32	96.66	162.70	238.20
2016	175.66	96.84	97.72	162.75	248.00
2017	180.90	109.28	100.90	168.40	254.60
2018	187.00	103.48	102.00	176.30	262.00

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

2019	189.80	104.63	103.90	178.60	265.70
2020	190.10	105.01	104.10	177.20	266.90
2005–2020	168.40	91.39	94.80	158.10	232.80

3.2 Model performance

This study first compared levels between indicators of UV radiation and measurements of UV radiation. The results indicated an R^2 of 0.65 between ERA-5 UV and UV radiation measurements and an R^2 of 0.55 between OMI EDD and UV radiation measurements in 2005–2020, respectively, indicating that both simulated and satellite remote-sensed UV radiation data could moderately represent ground UV radiation levels.

The overall R^2 (RMSE) of model development between measured and predicted UV radiation were 0.97 (15.64 W m^{-2}) at daily level. Figure 2 displays the scatter density plots between measurements and CV predictions of UV radiation at daily level, including overall CV (a), spatial CV (b), temporal CV (c) and by year CV (d). From the density scatter plots, it could be seen that most of the measured-predicted pairs from CV fell on the 1:1 line, indicating a relatively high consistency between measurements and the CV predictions. The CV R^2 (RMSE) values between measured and predicted UV radiation were 0.83 (37.44 W m^{-2}) for overall CV, 0.75 (45.56 W m^{-2}) for spatial CV, 0.83 (37.48 W m^{-2}) for temporal CV, and 0.82 (38.86 W m^{-2}) for by year CV at daily level, and 0.91 (21.01 W m^{-2}), 0.81 (31.14 W m^{-2}), 0.91 (21.05 W m^{-2}), 0.89 (22.90 W m^{-2}) at monthly level for overall, spatial, temporal and by year CV, respectively. Figure 3 shows the temporal trend of monthly average values for predicted and measured UV radiation at monitoring stations from 2005 to 2020, which also indicated a high consistency between them, although the predictions tended to overestimate UV radiation when it was low and underestimate UV radiation when it was high.

Figure A1 shows the importance ranking of all predictors produced by the random forest model. ERA-5 UV, OMI EDD and MAIAC AOD were the top important predictors in predicting UV radiation among all predictive variables. Figure A2 illustrates that, with other predictors held constant, the inclusion of OMI EDD as a predictor in the model yielded an overall CV R^2 (RMSE) of 0.83 (37.44 W m^{-2}), compared to 0.81 (39.18 W m^{-2}) when OMI EDD is not included.

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

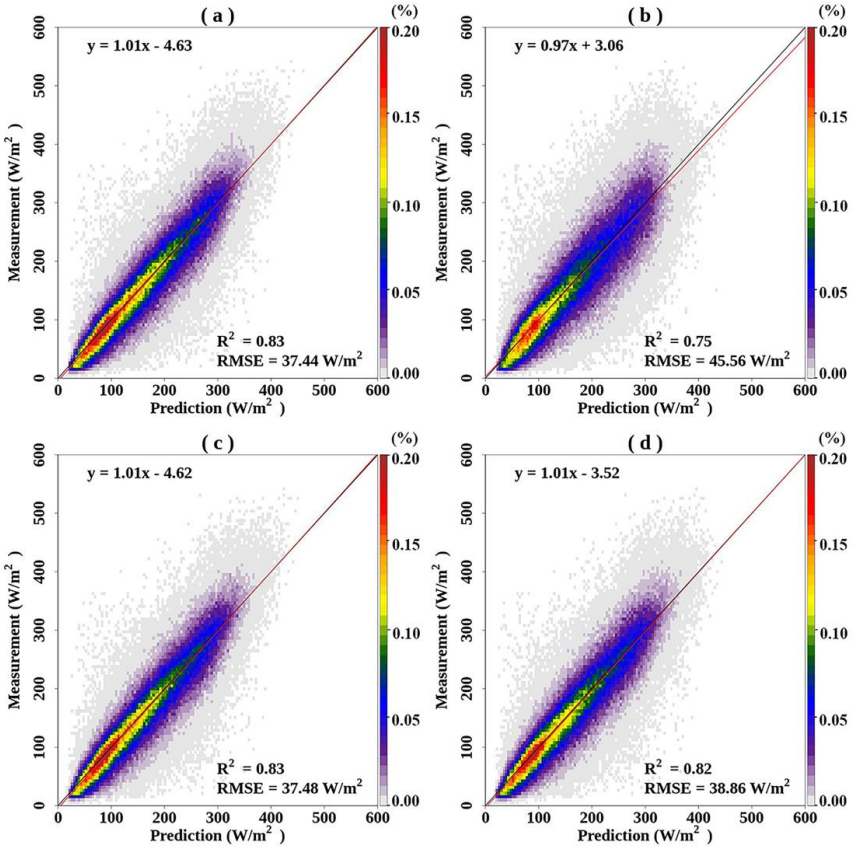


Figure 2. Density scatter plots and linear regressions between measurements and predictions of UV radiation at daily level based on random forest model during 2005–2020. Overall CV(a); spatial CV(b); temporal CV(c); by-year CV(d).

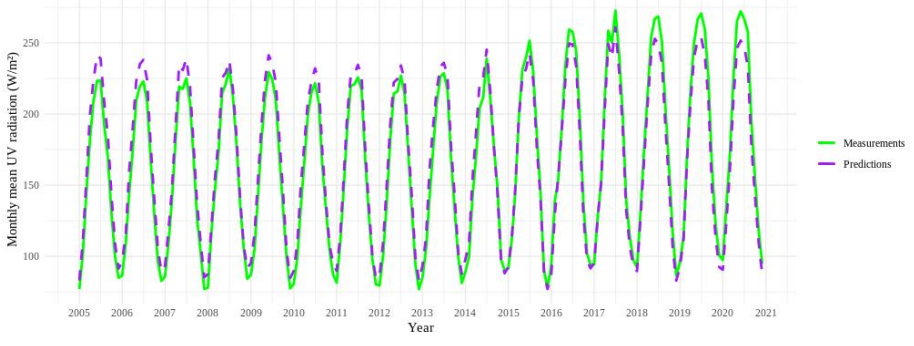


Figure 3. Time series plot of monthly mean UV radiation for measurements (green line) and predictions (purple dash) at monitoring stations during 2005–2020.

3.3 Spatiotemporal distributions of UV radiation based on predictions

The spatial distribution of annual average UV radiation based on predictions from 2005 to 2020 is illustrated in Fig. A3 for

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

each year and in Fig. 4 for the averaged values during 2005–2020, indicating uneven spatial distribution of UV radiation in China associated with factors such as latitude, elevation (Fig. A4), and meteorological factors. On one hand, UV radiation was stronger in southern region with lower latitude than in northern region with higher latitude. For example, in subregions labeled G of Fig. 4, located at the southernmost latitude in mainland China (~18° N), the UV radiation value was 205.86 W m^{-2} that was 1.46 times than the UV radiation in subregions labeled A, situated at the northernmost latitude in China (~50° N). On the other hand, UV radiation was higher in western region with higher elevation than in region with lower elevation, for example, the subregions labeled C of Fig. 4 with averaged elevation of 4730 m had the highest UV radiation level of 228.36 W m^{-2} that was 1.50 times than subregions labeled E with averaged elevation of 5 m. But due to the influence of climatic factors, the relationship between UV radiation and latitude as well as elevation may vary in some regions. For example, subregions labeled D and F of Fig. 4 shared similar elevations and latitudes, but UV radiation at F was 152.14 W m^{-2} that was 14.29% higher than that at D. Figure A5 shows the population density, indicating that although subregions labeled C had highest UV radiation in China, but its population was sparse; while southeastern coastal areas of China had both dense population and relatively strong UV radiation and therefore had a relatively higher population exposure risk.

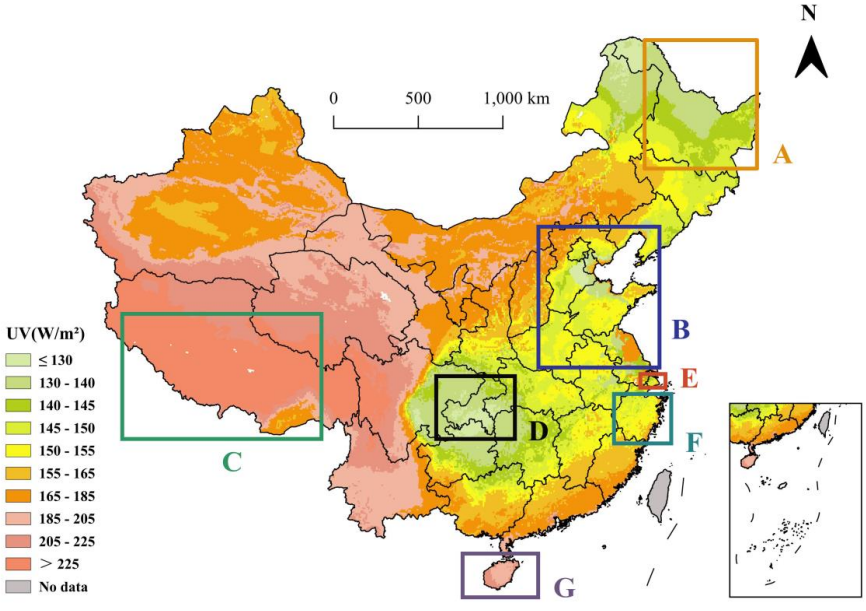


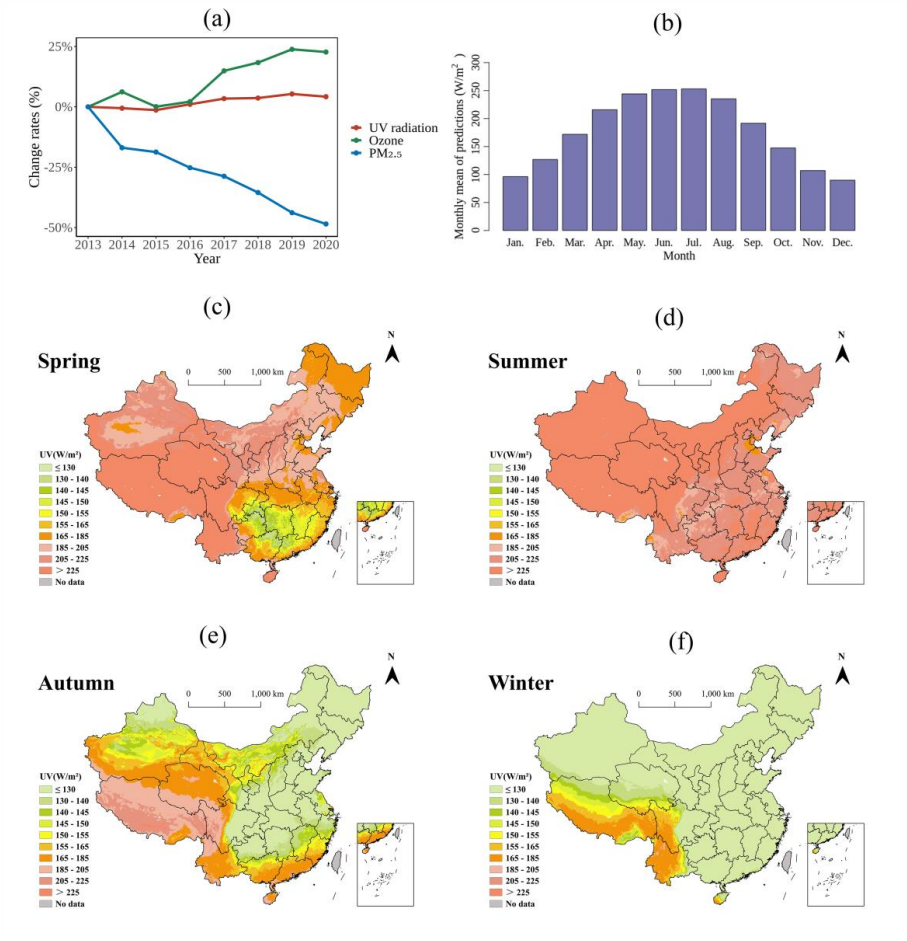
Figure 4. Spatial distribution of averaged annual-mean UV radiation during 2005–2020. The boxes represent: Heilongjiang Province (A); North China Plain (B); Tibet Autonomous Region (C); Chongqing City (D); Shanghai City (E); Zhejiang Province (F); and Hainan Province (G).

The inter-annual and intra-annual trends of UV radiation are shown in Fig. 5. For long-term temporal trend, UV radiation experienced slight fluctuations from 2005 to 2014 but remained relatively stable, and then increased since 2015. Figure 5a

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

depicts the trends in the changes of UV radiation, O_3, and $PM_{2.5}$ across the China mainland from 2013 to 2020, showing that $PM_{2.5}$ demonstrated a prominent downward trend while both UV radiation and O_3 exhibited noticeable upward trends during the period. In comparison to values in 2013, UV radiation increased by 4.20% in 2020, rising from 176.68 W m^{-2} to 184.10 W m^{-2} nationwide, O_3 experienced a 22.70% increase, while $PM_{2.5}$ decreased by 48.51%. In addition, it can be seen from Fig. A3 that North China Plain (labeled as B in Fig. 4) increased the most significantly since 2015, with UV radiation rising by 7.13% from 2013 to 2020, which is 1.70 times the nationwide UV radiation growth rate. For the intra-annual variation, UV radiation exhibited a clear seasonal trend, with significantly higher levels during the summer than in winter. It was highest in July with an average value of 253.02 W m^{-2} in 2005–2020, and then gradually decreases, reaching the lowest in December with an average of 89.81 W m^{-2}. Additionally, Fig. 3(c)–(f) illustrate the varying spatial trends of UV radiation across different seasons. In spring, the intensity of UV radiation in northern regions surpassed that in most southern areas. During summer, the UV radiation across mainland China consistently exceeded 162 W m^{-2}. The spatial distribution of UV radiation intensity was primarily affected by elevation and latitude in autumn. In winter, with the exception of specific areas in western China, the UV radiation levels in other regions remained below 140 W m^{-2}.

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024



225 Figure 5. Inter-annual and intra-annual variation of UV radiation based on predictions in mainland China. Annual change rates of UV radiation, O_3, $PM_{2.5}$ in mainland China from 2013 to 2020 (a); the averaged monthly mean UV radiation in mainland China in 2005–2020 (b); average seasonal mean UV radiation in mainland China in 2005–2020 in spring (c), summer (d), autumn (e), and winter (f).

4 Discussion

230 This study developed a random forest model using a variety of predictors to predict daily UV radiation with relatively high accuracy, resolution and spatiotemporal coverage in mainland China. Based on predictions generated from the model, the temporal and spatial characteristics were identified. A gradual rise in UV radiation in recent years were found and uneven spatial distribution was explored.

This study predicted UV radiation based on machine learning algorithm at daily level and 10 km spatial resolution with nearly
235 full coverage in China with multiple predictors including satellite and simulated UV radiation data. The R^2 (RMSE) between measured and predicted UV radiation was 0.97 (15.64 W m^{-2}) for model development and 0.83 (37.44 W m^{-2}) for overall 10-

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

fold CV at daily level. Compared with other environmental factors affecting population health, such as air pollution, limited studies developed model for UV radiation, and most of them conducted in the United States and Europe with statistical model such as regression analysis and Area-to-point residual kriging (Feister et al., 2008; Junk et al., 2007; Pei and He, 2019; Vopham et al., 2016). In recent years, several studies started to employ machine learning methods for predicting UV radiation (Wu et al., 2022; Zhao and He, 2022). In previous studies R^2 between measured and predicted UV radiation for model development ranged from 0.92 to 0.98 (Liu et al., 2017; Zhao and He, 2022; Qin et al., 2020), which were comparable with ours. Several studies developed models to predict UV radiation in China, but they did not investigate the roles of satellite UV radiation measurements in model performance. UV radiation data from satellite has been proven as one effective variable for evaluating exposure levels and identifying hotspots of skin cancer risks in other countries (Zhou et al., 2019; Kennedy et al., 2021). Satellite-sourced UV radiation data, such as OMI EDD, could be recognized as one kind of direct measurements of UV radiation from satellite, providing "real values" to constrain UV radiation predictions during spatial extrapolation (Gholamnia et al., 2021). Including OMI EDD in UV radiation model improved the predicting accuracy by approximate 2% compared with the model without it in this study. Additionally, this study filled the missing values of OMI EDD data to make the spatiotemporal coverage of UV radiation predictions close to 100%, which was higher than previous studies that predicted UV radiation at 724 conventional meteorological stations in China or that did not deal with the missing values in UV radiation predictions caused by incomplete predictor variables such as AOD data from remote sensing (Wu et al., 2022; Liu et al., 2017). The gridded UV predictions with nearly full spatiotemporal coverage could provide more comprehensive and flexible support for exposure assessment in health studies regarding with exposure windows and geographic locations.

The results indicated that UV radiation is unevenly distributed throughout China, with high exposure areas primarily located in the southwest and health risk hotspots primarily located in the eastern region. The spatial distribution of UV radiation correlated closely with elevation, latitude, and climatic factors. Higher elevation resulted in stronger UV radiation primarily due to the thinner atmosphere, which means less UV radiation is absorbed or scattered by the atmosphere (Blumthaler et al., 1997). UV radiation intensity also increased with decreasing latitude, primarily because regions at low latitudes have a smaller SZA (Holzle and Honigsmann, 2005). The spatial distribution of UV radiation in autumn can effectively reflected its correlation with elevation and latitude. Simultaneously, meteorological factors like cloud cover impact UV radiation intensity as they can absorb and scatter UV radiation (Dieste-Velasco et al., 2023). For example, the higher cloud cover and humidity in subregions labeled D resulted in higher UV radiation at F compared to D, despite their similar elevations and latitudes (Fig. 4). In spring, owing to factors such as air currents, the southern regions were subjected to increased precipitation, resulting in elevated cloud cover and humidity (Yao et al., 2017). Consequently, this phenomenon may resulted in lower UV radiation intensity in southern regions as compared to the relatively arid northern regions. Besides natural factors, population distribution should be considered for identifying health risk hotspots. Although UV radiation levels were medium high in southeast coastal

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

regions, the population health effects due to UV radiation should not be ignored for the high population density there. The threshold for health effects of UV radiation for population is still unclear and there is not standard of atmospheric UV radiation so far, which need the support from further epidemiological studies. The UV radiation predictions in this study covers entire geographical areas of mainland China, providing exposure data to support health studies in different regions and further identify the health risk hotspots of UV radiation exposure in China.

UV radiation levels exhibited both seasonal and long-term temporal trends. The seasonal pattern showed the strongest UV radiation in summer and the lowest in winter. This observed pattern might be linked to variations in daylight hours and alterations in SZA over the year (Liu et al., 2017). Specifically, our findings demonstrated an increase trend in UV radiation since 2015, accompanied by a decrease in $PM_{2.5}$ and an increase in O_3, suggesting a potential correlation between UV radiation levels and air pollution. The decrease in $PM_{2.5}$ might contribute to the increase in UV radiation, as $PM_{2.5}$ has the ability to absorb and reflect UV radiation (Madronich et al., 2023; Gao et al., 2013). Meanwhile, UV radiation plays a crucial role in the production of surface O_3, as ground-level O_3 primarily originates from photochemical reactions (Meng et al., 2022). Chinese government launched and implemented a series of nationwide policies to decrease air pollution levels, including the Action Plan of Air Pollution Prevention and Control (APPC-AP) in 2013 and Three-year (2018–2020) Action Plan for Cleaner Air in 2017. By virtue of these policies, concentrations of several criteria air pollutants especially $PM_{2.5}$ have dropped significantly in China since 2013. Therefore, alongside the decrease of $PM_{2.5}$, there is a simultaneous need to enhance public awareness of UV radiation protection.

The relatively small number of UV radiation monitoring stations, amounting to 40 stations employed for model development across the national landscape, may influence the extrapolation performance of the model. The UV monitoring stations were distributed in different geographic locations with multiple land cover types, which could help validate the model performance in spatial extrapolation. On the other hand, the spatial CV was conducted, which only slightly decreases compared to the overall CV, showing the relatively higher accuracy of spatial extrapolation.

5 Data availability

The UV radiation gridded dataset across mainland China in 2005–2020 is currently freely available at https://doi.org/10.5281/zenodo.10884591 (Jiang et al., 2024).

6 Conclusion

This study established a machine learning model for predicting daily UV radiation levels at a 10 km × 10 km spatial resolution across mainland China for 16 years, and model with satellite-sourced UV radiation measurements performed higher predicting

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

accuracy than the one without such predictor. Based on the high-resolution and coverage predictions, a gradual rise in UV radiation in recent years and uneven spatial distribution were found in China. This study provides modeling method and exposure data of UV radiation to support exposure assessment for future epidemiological studies and identification of exposure risk and health risk hotspots of UV radiation for Chinese population.

Appendix A: Additional figures and tables

Table A1. The by-year missing rates of erythemally daily dose retrieved from the Ozone Monitoring Instrument in mainland China during 2005–2020.

year	Missing rate before gap-filling
2005	3.03%
2006	3.53%
2007	3.38%
2008	5.69%
2009	20.33%
2010	30.28%
2011	33.59%
2012	21.80%
2013	24.24%
2014	28.20%
2015	31.95%
2016	35.29%
2017	32.78%

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

2018	32.19%
2019	32.12%
2020	30.34%
2005-2020	23.04%

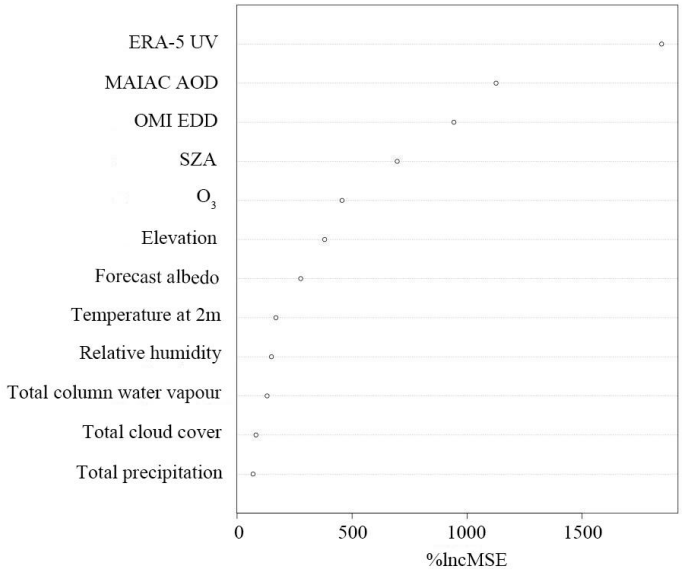


Figure A1. Ranking of importance for predictor variables in UV radiation prediction model. Note: downward UV radiation at the surface from the fifth generation European Center for Medium-Range Weather Forecasts Reanalysis (ERA-5 UV), aerosol optical depth data from the Multi-Angle Implementation of Atmospheric Correction (MAIAC AOD), erythemally daily dose retrieved from the Ozone Monitoring Instrument (OMI EDD), solar zenith angle (SZA).

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

Open Access Earth System Science Data Discussions

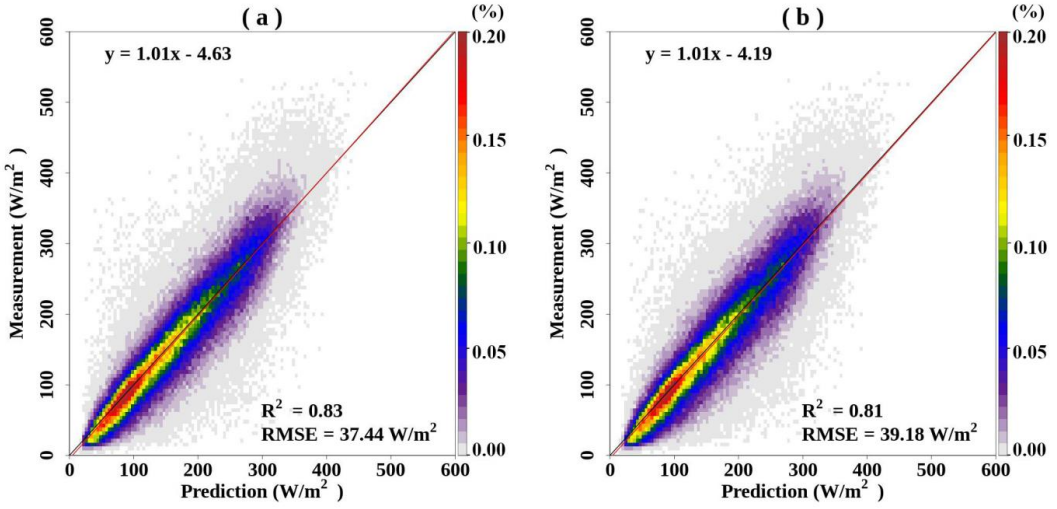


Figure A2. Density scatter plots and linear regressions between measurements and predictions of UV radiation at daily level based on random forest model during 2005–2020. With erythemally daily dose retrieved from the Ozone Monitoring Instrument (a); without erythemally daily dose retrieved from the Ozone Monitoring Instrument (b).

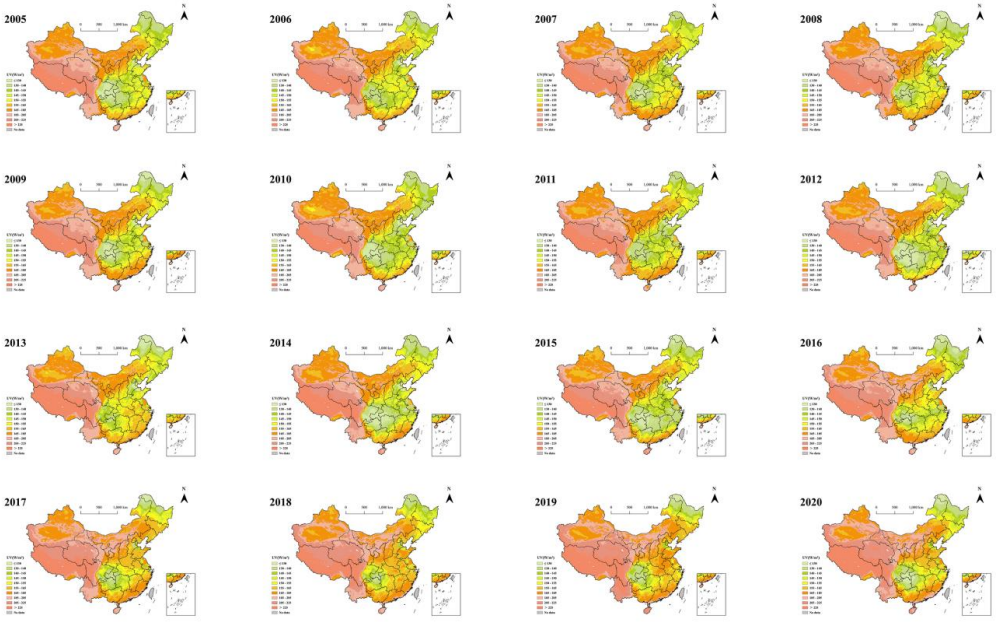


Figure A3. Spatial distributions of UV radiation based on predictions at annual level from 2005 to 2020.

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

Open Access Earth System Science Data Discussions

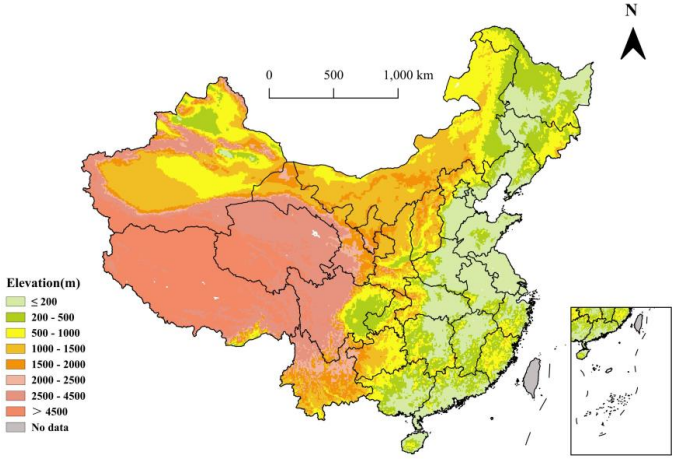


Figure A4. Spatial distribution of elevation in mainland China.

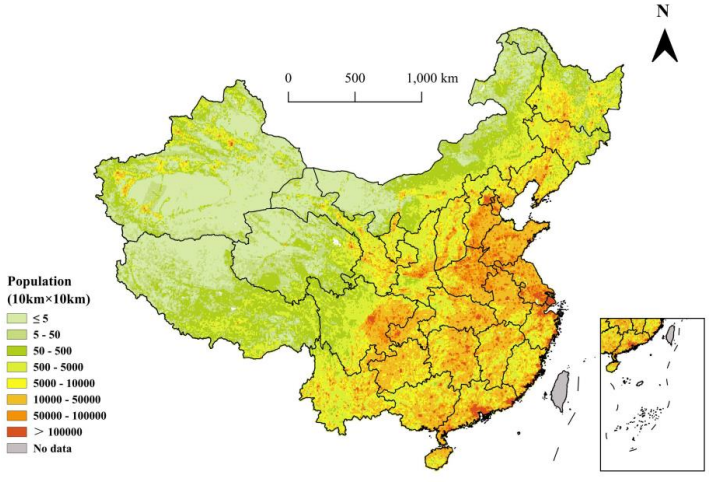


Figure A5. Spatial distribution of population in mainland China in 2020

Author contributions

Yichen Jiang: Conceptualization, Data curation, Methodology, Software, Writing- Original draft preparation, Writing – review & editing. Su Shi: Data curation, Software, Validation. Xinyue Li: Data curation, Software. Chang Xu: Data curation, Software. Haidong Kan: Funding acquisition, Writing – review & editing. Bo Hu: Resources, Funding acquisition, Writing – review & editing. Xia Meng: Conceptualization, Resources, Funding acquisition, Supervision, Writing – review & editing.

Competing interests

We declare that we have no conflict of interest.

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

Acknowledgements

This work was supported by the National Key Research and Development Program of China (No. 2023YFC3708304, 2022YFC3700705); National Natural Science Foundation of China (82030103).

References

Blumthaler, M., Ambach, W., and Ellinger, R.: Increase in solar UV radiation with altitude, Journal of Photochemistry and Photobiology B: Biology, 39, 130-134, https://doi.org/10.1016/s1011-1344(96)00018-8, 1997.

Boscoe, F. P. and Schymura, M. J.: Solar ultraviolet-B exposure and cancer incidence and mortality in the United States, 1993-2002, BMC Cancer, 6, 264, https://doi.org/10.1186/1471-2407-6-264, 2006.

Breiman, L.: Random Forests, Machine Learning, 45, 5-32, https://doi.org/10.1023/A:1010933404324, 2001.

Brenner, M. and Hearing, V. J.: The protective role of melanin against UV damage in human skin, Photochem Photobiol, 84, 539-549, https://doi.org/10.1111/j.1751-1097.2007.00226.x, 2008.

Chen, Y., Liang, S., Ma, H., Li, B., He, T., and Wang, Q.: An all-sky 1 km daily land surface air temperature product over mainland China for 2003–2019 from MODIS and ancillary data, Earth System Science Data, 13, 4241-4261, https://doi.org/10.5194/essd-13-4241-2021, 2021.

Corrêa, M. d. P.: UVBoost: An erythemal weighted ultraviolet radiation estimator based on a machine learning gradient boosting algorithm, Journal of Quantitative Spectroscopy and Radiative Transfer, 298, https://doi.org/10.1016/j.jqsrt.2023.108490, 2023.

Deng, Y., Yang, D., Yu, J. M., Xu, J. X., Hua, H., Chen, R. T., Wang, N., Ou, F. R., Liu, R. X., Wu, B., and Liu, Y.: The Association of Socioeconomic Status with the Burden of Cataract-related Blindness and the Effect of Ultraviolet Radiation Exposure: An Ecological Study, Biomed Environ Sci, 34, 101-109, https://doi.org/10.3967/bes2021.015, 2021.

Dieste-Velasco, M. I., García-Rodríguez, S., García-Rodríguez, A., Díez-Mediavilla, M., and Alonso-Tristán, C.: Modeling Horizontal Ultraviolet Irradiance for All Sky Conditions by Using Artificial Neural Networks and Regression Models, Applied Sciences, 13, https://doi.org/10.3390/app13031473, 2023.

Feister, U., Junk, J., Woldt, M., Bais, A., Helbig, A., Janouch, M., Josefsson, W., Kazantzidis, A., Lindfors, A., Outer, P. N. d., and Slaper, H.: Long-term solar UV radiation reconstructed by ANN modelling with emphasis on spatial characteristics of input data, Atmospheric Chemistry and Physics, 8, 3107–3118, https://doi.org/10.5194/acp-8-3107-2008, 2008.

Gao, Z., Gao, W., and Chang, N.-B.: Spatial Statistical Analyses of Global Trends of Ultraviolet B Fluxes in the Continental United States, GIScience & Remote Sensing, 49, 735-754, https://doi.org/10.2747/1548-1603.49.5.735, 2013.

Gholamnia, R., Abtahi, M., Dobaradaran, S., Koolivand, A., Jorfi, S., Khaloo, S. S., Bagheri, A., Vaziri, M. H., Atabaki, Y., Alhouei, F., and Saeedi, R.: Spatiotemporal analysis of solar ultraviolet radiation based on Ozone Monitoring Instrument dataset in Iran, 2005-2019, Environ Pollut, 287, 117643, https://doi.org/10.1016/j.envpol.2021.117643, 2021.

González-Rodríguez, L., Rodríguez-López, L., Jiménez, J., Rosas, J., García, W., Duran-Llacer, I., de Oliveira, A. P., and Barja, B.: Spatio-temporal estimations of ultraviolet erythemal radiation in Central Chile, Air Quality, Atmosphere & Health, 15, 837-852, https://doi.org/10.1007/s11869-022-01195-y, 2022.

Grandahl, K., Eriksen, P., Ibler, K. S., Bonde, J. P., and Mortensen, O. S.: Measurements of Solar Ultraviolet Radiation Exposure at Work and at Leisure in Danish Workers, Photochem Photobiol, 94, 807-814, https://doi.org/10.1111/php.12920, 2018.

Griffin, G. K., Booth, C. A. G., Togami, K., Chung, S. S., Ssozi, D., Verga, J. A., Bouyssou, J. M., Lee, Y. S., Shanmugam, V., Hornick, J. L., LeBoeuf, N. R., Morgan, E. A., Bernstein, B. E., Hovestadt, V., van Galen, P., and Lane, A. A.: Ultraviolet radiation shapes dendritic cell leukaemia transformation in the skin, Nature, 618, 834-841, https://doi.org/10.1038/s41586-023-06156-8, 2023.

Habte, A., Sengupta, M., Gueymard, C. A., Narasappa, R., Rosseler, O., and Burns, D. M.: Estimating Ultraviolet Radiation

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

From Global Horizontal Irradiance, IEEE Journal of Photovoltaics, 9, 139-146, https://doi.org/10.1109/jphotov.2018.2871780, 2019.

Holzle, E. and Honigsmann, H.: UV-radiation-Sources, Wavelength, Environment, J Dtsch Dermatol Ges, 3 Suppl 2, S3-10, https://doi.org/10.1111/j.1610-0387.2005.04392.x, 2005.

Hu, B., Wang, Y., and Liu, G.: Variation characteristics of ultraviolet radiation derived from measurement and reconstruction in Beijing, China, Tellus B: Chemical and Physical Meteorology, 62, 100-108, https://doi.org/10.1111/j.1600-0889.2010.00452.x, 2010.

Jiang, Y., Shi, S., Li, X., Xu, C., Kan, H., Hu, B., & Meng, X.: A database of 10 km Ultraviolet Radiation Product over mainland China: 2005-2020, Zenodo [data set], https://doi.org/10.5281/zenodo.10884591,2024.

Junk, J., Feister, U., and Helbig, A.: Reconstruction of daily solar UV irradiation from 1893 to 2002 in Potsdam, Germany, International Journal of Biometeorology, 51, 505-512, https://doi.org/10.1007/s00484-007-0089-4, 2007.

Kennedy, C., Liu, Y., Meng, X., Strosnider, H., Waller, L. A., and Zhou, Y.: Developing indices to identify hotspots of skin cancer vulnerability among the Non-Hispanic White population in the United States, Ann Epidemiol, 59, 64-71, https://doi.org/10.1016/j.annepidem.2021.04.004, 2021.

Lagreze, W. A., Joachimsen, L., and Schaeffel, F.: [Current recommendations for deceleration of myopia progression], Ophthalmologe, 114, 24-29, https://doi.org/10.1007/s00347-016-0346-1, 2017.

Lin, S. W., Wheeler, D. C., Park, Y., Cahoon, E. K., Hollenbeck, A. R., Freedman, D. M., and Abnet, C. C.: Prospective study of ultraviolet radiation exposure and risk of cancer in the United States, Int J Cancer, 131, E1015-1023, https://doi.org/10.1002/ijc.27619, 2012.

Liu, H., Hu, B., Zhang, L., Zhao, X. J., Shang, K. Z., Wang, Y. S., and Wang, J.: Ultraviolet radiation over China: Spatial distribution and trends, Renewable and Sustainable Energy Reviews, 76, 1371-1383, https://doi.org/10.1016/j.rser.2017.03.102, 2017.

Liu, S., Geng, G., Xiao, Q., Zheng, Y., Liu, X., Cheng, J., and Zhang, Q.: Tracking Daily Concentrations of PM(2.5) Chemical Composition in China since 2000, Environ Sci Technol, 56, 16517-16527, https://doi.org/10.1021/acs.est.2c06510, 2022.

Madronich, S., Sulzberger, B., Longstreth, J. D., Schikowski, T., Andersen, M. P. S., Solomon, K. R., and Wilson, S. R.: Changes in tropospheric air quality related to the protection of stratospheric ozone in a changing climate, Photochem Photobiol Sci, 22, 1129-1176, https://doi.org/10.1007/s43630-023-00369-6, 2023.

Marson, J. W., Litchman, G. H., and Rigel, D. S.: The magnitude of increased United States melanoma incidence attributable to ground-level ultraviolet radiation intensity trends, J Am Acad Dermatol, 84, 1734-1735, https://doi.org/10.1016/j.jaad.2020.08.100, 2021.

McPeters, R. D., Frith, S., and Labow, G. J.: OMI total column ozone: extending the long-term data record, Atmospheric Measurement Techniques, 8, 4845-4850, https://doi.org/10.5194/amt-8-4845-2015, 2015.

Meng, X., Liu, C., Zhang, L., Wang, W., Stowell, J., Kan, H., and Liu, Y.: Estimating PM(2.5) concentrations in Northeastern China with full spatiotemporal coverage, 2005-2016, Remote Sens Environ, 253, https://doi.org/10.1016/j.rse.2020.112203, 2021.

Meng, X., Wang, W., Shi, S., Zhu, S., Wang, P., Chen, R., Xiao, Q., Xue, T., Geng, G., Zhang, Q., Kan, H., and Zhang, H.: Evaluating the spatiotemporal ozone characteristics with high-resolution predictions in mainland China, 2013-2019, Environ Pollut, 299, 118865, https://doi.org/10.1016/j.envpol.2022.118865, 2022.

Mohr, S. B., Garland, C. F., Gorham, E. D., Grant, W. B., and Garland, F. C.: Relationship between low ultraviolet B irradiance and higher breast cancer risk in 107 countries, Breast J, 14, 255-260, https://doi.org/10.1111/j.1524-4741.2008.00571.x, 2008.

Narayanan, D. L., Saladi, R. N., and Fox, J. L.: Ultraviolet radiation and skin cancer, Int J Dermatol, 49, 978-986, https://doi.org/10.1111/j.1365-4632.2010.04474.x, 2010.

Pei, C. and He, T.: UV RADIATION ESTIMATION IN THE UNITED STATES USING MODIS DATA, IEEE International Symposium on Geoscience and Remote Sensing, 1880-1883, https://doi.org/10.1109/IGARSS.2019.8900659, 2019.

Qin, W., Wang, L., Wei, J., Hu, B., and Liang, X.: A novel efficient broadband model to derive daily surface solar Ultraviolet radiation (0.280-0.400 mum), Sci Total Environ, 735, 139513, https://doi.org/10.1016/j.scitotenv.2020.139513, 2020.

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

Santos, J. B., Villán, D. M., and Castrillo, A. d. M.: Analysis and cloudiness influence on UV total irradiation, International Journal of Climatology, 31, 451-460, https://doi.org/10.1002/joc.2072, 2011.

Shi, S., Wang, W., Li, X., Hang, Y., Lei, J., Kan, H., and Meng, X.: Optimizing modeling windows to better capture the long-term variation of PM(2.5) concentrations in China during 2005-2019, Sci Total Environ, 854, 158624, https://doi.org/10.1016/j.scitotenv.2022.158624, 2022.

Shi, S., Wang, W., Li, X., Hang, Y., Lei, J., Kan, H., and Meng, X.: Optimizing modeling windows to better capture the long-term variation of PM(2.5) concentrations in China during 2005-2019, Sci Total Environ, 854, 158624, https://doi.org/10.1016/j.scitotenv.2022.158624, 2023a.

Shi, S., Wang, W., Li, X., Xu, C., Lei, J., Jiang, Y., Zhang, L., He, C., Xue, T., Chen, R., Kan, H., and Meng, X.: Evolution in disparity of PM2.5 pollution in China, Eco-Environment & Health, 2, 257-263, https://doi.org/10.1016/j.eehl.2023.08.007, 2023b.

Stump, T. K., Fastner, S., Jo, Y., Chipman, J., Haaland, B., Nagelhout, E. S., Wankier, A. P., Lensink, R., Zhu, A., Parsons, B., Grossman, D., and Wu, Y. P.: Objectively-Assessed Ultraviolet Radiation Exposure and Sunburn Occurrence, Int J Environ Res Public Health, 20, https://doi.org/10.3390/ijerph20075234, 2023.

Swaminathan, A., Harrison, S. L., Ketheesan, N., van den Boogaard, C. H. A., Dear, K., Allen, M., Hart, P. H., Cook, M., and Lucas, R. M.: Exposure to Solar UVR Suppresses Cell-Mediated Immunization Responses in Humans: The Australian Ultraviolet Radiation and Immunity Study, J Invest Dermatol, 139, 1545-1553 e1546, https://doi.org/10.1016/j.jid.2018.12.025, 2019.

Thayer, Z. M.: The vitamin D hypothesis revisited: race-based disparities in birth outcomes in the United States and ultraviolet light availability, Am J Epidemiol, 179, 947-955, https://doi.org/10.1093/aje/kwu023, 2014.

Tian, X., Zhang, B., Jia, Y., Wang, C., and Li, Q.: Retinal changes following rapid ascent to a high-altitude environment, Eye (Lond), 32, 370-374, https://doi.org/10.1038/eye.2017.195, 2018.

Vienneau, D., de Hoogh, K., Hauri, D., Vicedo-Cabrera, A. M., Schindler, C., Huss, A., Roosli, M., and Group, S. N. C. S.: Effects of Radon and UV Exposure on Skin Cancer Mortality in Switzerland, Environ Health Perspect, 125, 067009, https://doi.org/10.1289/EHP825, 2017.

VoPham, T., Bertrand, K. A., Yuan, J. M., Tamimi, R. M., Hart, J. E., and Laden, F.: Ambient ultraviolet radiation exposure and hepatocellular carcinoma incidence in the United States, Environ Health, 16, 89, https://doi.org/10.1186/s12940-017-0299-0, 2017.

VoPham, T., Hart, J. E., Bertrand, K. A., Sun, Z., Tamimi, R. M., and Laden, F.: Spatiotemporal exposure modeling of ambient erythemal ultraviolet radiation, Environ Health, 15, 111, https://doi.org/10.1186/s12940-016-0197-x, 2016.

Walls, A. C., Han, J., Li, T., and Qureshi, A. A.: Host risk factors, ultraviolet index of residence, and incident malignant melanoma in situ among US women and men, Am J Epidemiol, 177, 997-1005, https://doi.org/10.1093/aje/kws335, 2013.

Wolffsohn, J. S., Dhallu, S., Aujla, M., Laughton, D., Tempany, K., Powell, D., Gifford, K., Gifford, P., Wan, K., Cho, P., Stahl, U., and Woods, J.: International multi-centre study of potential benefits of ultraviolet radiation protection using contact lenses, Cont Lens Anterior Eye, 45, 101593, https://doi.org/10.1016/j.clae.2022.101593, 2022.

Wu, J., Qin, W., Wang, L., Hu, B., Song, Y., and Zhang, M.: Mapping clear-sky surface solar ultraviolet radiation in China at 1 km spatial resolution using Machine Learning technique and Google Earth Engine, Atmospheric Environment, 286, https://doi.org/10.1016/j.atmosenv.2022.119219, 2022.

Yao, S., Jiang, D., and Fan, G.: Seasonality of Precipitation over China, J Sciences(in Chinese), 46, 1191-1203, https://doi.org/10.3878/j.issn.1006-9895.1703.16233, 2017.

Zhao, R. and He, T.: Estimation of 1-km Resolution All-Sky Instantaneous Erythemal UV-B with MODIS Data Based on a Deep Learning Method, Remote Sensing, 14, https://doi.org/10.3390/rs14020384, 2022.

Zhou, Y., Meng, X., Belle, J. H., Zhang, H., Kennedy, C., Al-Hamdan, M. Z., Wang, J., and Liu, Y.: Compilation and spatio-temporal analysis of publicly available total solar and UV irradiance data in the contiguous United States, Environ Pollut, 253, 130-140, https://doi.org/10.1016/j.envpol.2019.06.074, 2019.

Zhu, Q., Bi, J., Liu, X., Li, S., Wang, W., Zhao, Y., and Liu, Y.: Satellite-Based Long-Term Spatiotemporal Patterns of Surface Ozone Concentrations in China: 2005-2019, Environ Health Perspect, 130, 27004, https://doi.org/10.1289/EHP9406,

https://doi.org/10.5194/essd-2024-111
Preprint. Discussion started: 15 May 2024

465 2022.