

Global trends in downward surface solar radiation from spatial interpolated ground observations during 1961-2019

Yuan, M., Leirvik, T. & Wild, M.

Published in: Journal of Climate
DOI: 10.1175/JCLI-D-21-0165.1

Available online: 09 Nov 2021

Citation:

Yuan, M., Leirvik, T. & Wild, M. (2021). Global trends in downward surface solar radiation from spatial interpolated ground observations during 1961-2019. *Journal of Climate*, 34(23), 9501-9521. doi: 10.1175/JCLI-D-21-0165.1

© Copyright 2021 American Meteorological Society (AMS). For permission to reuse any portion of this work, please contact permissions@ametsoc.org. Any use of material in this work that is determined to be “fair use” under Section 107 of the U.S. Copyright Act (17 U.S. Code § 107) or that satisfies the conditions specified in Section 108 of the U.S. Copyright Act (17 USC § 108) does not require the AMS’s permission. Republication, systematic reproduction, posting in electronic form, such as on a website or in a searchable database, or other uses of this material, except as exempted by the above statement, requires written permission or a license from the AMS. All AMS journals and monograph publications are registered with the Copyright Clearance Center (<https://www.copyright.com>). Additional details are provided in the AMS Copyright Policy statement, available on the AMS website (<https://www.ametsoc.org/PUBSCopyrightPolicy>).

Global Trends in Downward Surface Solar Radiation from Spatial Interpolated Ground Observations during 1961–2019

MENGHAN YUAN,^a THOMAS LEIRVIK,^a AND MARTIN WILD^b

^a Graduate School of Business, Nord University, Bodø, Norway

^b Institute for Atmospheric and Climate Science, ETH Zurich, Zürich, Switzerland

(Manuscript received 26 February 2021, in final form 14 July 2021)

ABSTRACT: Downward surface solar radiation (SSR) is a crucial component of the global energy balance, affecting temperature and the hydrological cycle profoundly, and it provides crucial information about climate change. Many studies have examined SSR trends; however, they have often concentrated on specific regions due to limited spatial coverage of ground-based observation stations. To overcome this spatial limitation, this study performs a spatial interpolation based on a machine learning method, random forest, to interpolate monthly SSR anomalies using a number of climatic variables (various temperature indices, cloud coverage, etc.), time-point indicators (years and months of SSR observations), and geographical characteristics of locations (latitude, longitude, etc.). The predictors that provide the largest explanatory power for interannual variability are *diurnal temperature range* and *cloud coverage*. The output of the spatial interpolation is a $0.5^\circ \times 0.5^\circ$ monthly gridded dataset of SSR anomalies with complete land coverage over the period 1961–2019, which is used afterward in a comprehensive trend analysis for (i) each continent separately and (ii) the entire globe. The continental-level analysis reveals the major contributors to the global dimming and brightening. In particular, the global dimming before the 1980s is primarily dominated by negative trends in Asia and North America, whereas Europe and Oceania have been the two largest contributors to the brightening after 1982 and up until 2019.

KEYWORDS: Climate change; Radiative fluxes; Interannual variability; Machine learning

1. Introduction

Surface solar radiation is a crucial climate variable and a main constituent of the global energy balance, playing an important role in temperature change and the hydrological cycle (see, e.g., Budyko 1969; Liepert et al. 2004; Pfeifroth et al. 2018; Obryk et al. 2018). The positive trend in downward surface solar radiation since the 1980s in combination with increasing greenhouse gases leads to an intensification of the land-based hydrological cycle (Wild et al. 2008; Wild and Liepert 2010; Wild 2016). Moreover, it has profound impacts on various aspects of the society and economy, especially on agriculture. For example, crop yields could be significantly influenced not only by increases or decreases in solar radiation through enhancing or weakening of photosynthesis, but also indirectly by the resulting temperature change from solar radiation variations (Greenwald et al. 2006; Roderick and Farquhar 2012; Gupta et al. 2017; Proctor et al. 2018).

To analyze the drivers and economic impacts of climate change, it is of critical importance to have an understanding of surface solar radiation (SSR), in terms of its trends, levels, and

variations. Ground-based observations are believed to be the most reliable long-term data source for solar radiation, and have been used in many climate studies to monitor its evolution (see, e.g., Wild et al. 2005; Sanchez-Lorenzo et al. 2015, 2017; Pfeifroth et al. 2018; Parding et al. 2016). Despite their reliability as compared to other sources, one of the main drawbacks of ground-based measurements is their limited temporal and spatial coverage. For a start, extensive SSR observations have a relatively short history of only a few decades; they were not widely available until the 1960s and have a time-lag effect due to the time-consuming process of data collecting and homogenizing. As for spatial coverage, climate stations tend to be concentrated in regions that can provide the financial and technical support to maintain the devices. Therefore it is essential to extrapolate the available observations, in the dimension of both space and time, thereby enabling a more comprehensive overview that better represents all areas with continuous time series. The method that aims to fill gaps in spatial datasets is called spatial interpolation. Conventional spatial interpolation methods such as inverse distance weighting, kriging, splines, etc., have seen extensive applications in various climate processes (see, e.g., Collins 1995; Erxleben et al. 2002; Scudiero et al. 2016). This study contributes to and expands the existing literature by applying a novel machine learning method to interpolate a station observation dataset of SSR.

Machine learning methods have seen an increasing number of applications in spatial interpolation and shown effectiveness in reproducing and predicting climate variables with high accuracy and low uncertainty (see, e.g., Jiang 2008; Sun et al. 2016; Zhou et al. 2017). Qin et al. (2019) compared the performance of four physically deterministic models with eight

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-21-0165.s1>.

Corresponding author: Menghan Yuan, menghan.yuan@nord.no

DOI: 10.1175/JCLI-D-21-0165.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

TABLE 1. Summary of predictors. All climatic variables in the CRU are on a monthly basis.

Variable category	Symbol	Definition
Climatic variables	cld	Monthly average cloud cover (%)
	dtr	Monthly average diurnal temperature range (°C)
	frs	No. of days with ground frost in a month
	pre	Monthly total precipitation data (mm month ⁻¹)
	tmn	Monthly average minimum temperature (°C)
	tmp	Monthly average mean temperature (°C)
	tmx	Monthly average maximum temperature (°C)
	vap	Monthly average vapor pressure (hPa)
	wet	No. of rainy days in a month
Geographical variables	lat	Latitude of the location
	lon	Longitude of the location
	alt	Terrain altitude in m
	urban	1 if urban, 0 if rural
Temporal variables	year	Year of the observation/estimation
	mon	Month of the observation/estimation

machine learning models in an application of reproducing data of photosynthetically active radiation and reported out-performance of machine learning methods. The existing literature is mostly focused on simulating regional patterns of solar radiation; for example, Zhou et al. focused on the downward solar radiation over the United States whereas Qin et al. focused on the photosynthetically active radiation over China. In this study we aim for a comprehensive study of SSR on a global land scale; therefore, it is essential that the selected method should be able to cope with a large quantity of data. Among a wide range of machine learning approaches, *random*

forest has exceptional advantages in handling a large number of explanatory variables and in its capacity for processing large datasets due to its computational efficiency (Firth et al. 2005; Myoung et al. 2020). The study of Leirvik and Yuan (2021) compared the performance of random forest with those of seven other conventional (deterministic) spatial interpolation methods in an application of predicting global SSR, and showed a profound advantage of random forest in terms of prediction accuracy and performance stability.

In this paper, we apply random forest to a global SSR dataset, the dataset of the Global Energy Balance Archive

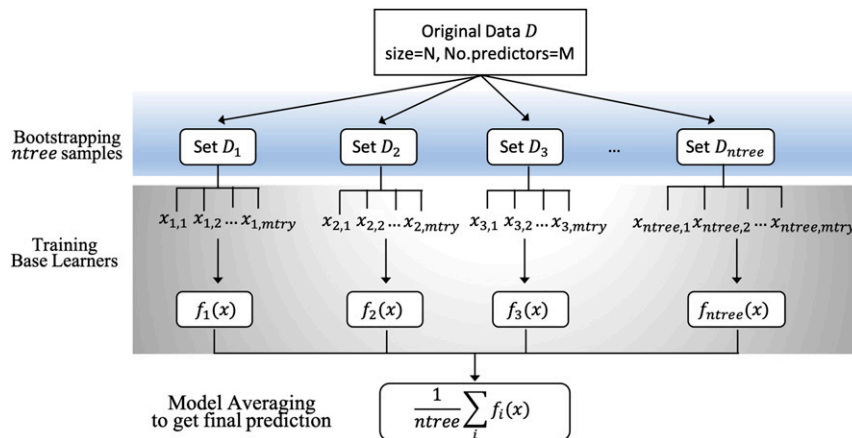


FIG. 1. Flowchart of random forest. The training algorithm of random forest applies the general technique of bootstrap aggregating to base learners. We start with a dataset D with $N = 328\,222$ observations and $M = 15$ features. The first step is to generate $ntree = 700$ random samples of size N with replacement. That is, each sample D_i , $i = 1, \dots, ntree$ has the same size as the population, and we call it a base tree learner. We train a regression tree $f_i(x)$ on D_i for $i = 1, \dots, ntree$. Each regression tree takes $mtry = 8$ ($mtry < M$) features as predictors, which could have different members for individual trees. For example, we choose 8 predictors randomly out of the total 15 predictors for D_1 ; another 8 predictors are chosen for D_2 . The two groups of predictors for D_1 and D_2 share some common variables, and they also have some distinct ones. In this way, variable randomness is added to the model. The final prediction is the average of predictions of all base learners.

TABLE 2. Number of SSR stations and monthly mean observations from GEBA in each continent.

Continent	Abbr ^a	No. of stations ^b	No. of obs ^c
South America	SA	134	13 362
Oceania	OC	76	13 549
Africa	AF	234	33 238
North America	NrA	214	37 427
Asia	AS	312	98 912
Europe	EU	516	131 734
Total		1486	328 222

^a Abbreviation for continent names.

^b Number of stations on the continent.

^c Number of monthly observations on the continent.

(GEBA). A total of 15 variables are selected as predictors for SSR, including nine climatic variables (various temperature indices, cloud cover, frost days, etc.), geographical coordinates (longitude and latitude), altitude, urbanization, and temporal indicators that indicate the time point (year and month) of the observed/simulated SSR record (refer to Table 1 for a detailed definition of the predictors). In the current study, we focus on the decadal long-term trends of SSR, which is embedded with strong seasonal intra-annual variability in the overall variations. To reduce seasonal variability, we train our model on SSR anomalies in the dataset, and the model is then applied to interpolate values at unsampled locations. The result is a $0.5^\circ \times 0.5^\circ$ monthly gridded dataset, conforming with the resolution and time step of the input Climate Research Unit (CRU) dataset. The constructed dataset provides SSR estimations with complete global land coverage and temporal coverage of 59 years (1961–2019). Based on this dataset, remote land areas such as Africa and Siberia that were rarely investigated before are made accessible for investigation of their long-term trends.

Trends could vary, or even reverse, during a multidecadal time period, making it highly important to detect potential breakpoints over the whole period. In fact, not only do the long-term trends reach far back in the past of interest, but it is also of critical importance to identify the most recent sustainable trends up until now. Particularly, global SSR experienced a widespread reduction (on the order of $3\text{--}9 \text{ W m}^{-2}$ decade⁻¹) from the 1950s to the 1980s, followed by an increase (on the order of $1\text{--}4 \text{ W m}^{-2}$ decade⁻¹), referred to respectively as global dimming and global brightening (Liepert and Kukla 1997; Stanhill and Cohen 2001; Stanhill 2005; Wild et al. 2005; Wild 2009; Gilgen et al. 2009; Wild 2012; Pfeifroth et al. 2018). Trend reversals (positive trends) have been shown in observational studies. A widespread trend reversal in the SSR records was first reported by Wild et al. (2005), while previous studies primarily pointed to the dimming. Regional brightening was documented in areas including Europe, North America, and Japan (e.g., Liepert 2002; Pinker et al. 2005; Parding et al. 2016; Tanaka et al. 2016). Aggregated series over a region reveal the overall interannual variation and trends; however, subregional trends are neutralized, or masked, by the aggregation if they are of opposite signs, which is often the case on many continents (Romanou

TABLE 3. Error measures for RF simulations. Units of MAE and RMSE are W m^{-2} .

Continent	MAE.ano	RMSE.ano	R.Squared.ano
Europe	7.90	12.03	0.56
North America	8.37	12.13	0.49
Oceania	9.39	13.60	0.43
Asia	10.48	14.92	0.51
Africa	11.50	15.74	0.34
South America	11.87	16.55	0.59
Global avg	9.92	14.16	0.49

et al. 2007; Pfeifroth et al. 2018). Therefore, the spatial distribution based on global grid boxes at the resolution of 0.5° is investigated in this study, and the results show significant spatial diversity.

The paper is organized as follows: section 2 describes the datasets and methods used in the spatial interpolation, model performance and trend analyses based on the constructed dataset are discussed in section 3, and section 4 is a discussion of the results found in the paper.

2. Datasets and method

a. Method

Random forest (RF) is a decision tree technique for regression and classification (Breiman 2001). In contrast to conventional decision tree methods, random forest constructs a forest of decision trees that operates as a predicting ensemble whose prediction accuracy is higher than that of any individual tree. Randomness in the RF is the distinctive characteristic that makes it one of the most powerful and widely used machine learning methods in recent applications (see e.g., Sun et al. 2016; Zhou et al. 2017; Xu et al. 2018).

A flowchart of random forest can be seen in Fig. 1. Randomness is added to the model through two steps. First, the RF uses bootstrapping to generate *n*tree training sets consisting of individual decision trees. Note that each decision tree is of the same size as the population but allows for replacement. The process is often known as bagging or bootstrapping. A single decision tree could be sensitive to training samples, as small changes in the training data could result in fluctuations in the tree structure. However, through the process of bagging, random forest takes the average of a forest of individual trees as the final prediction. This reduces the model's sensitivity to training data and helps to improve performance stability by reducing variance and avoiding overfitting in the algorithm. Second, random forest incorporates feature randomness such that each tree in the forest only corresponds to a random subset of independent variables. In contrast to conventional decision trees, all of which are trained on the same group of predictors and split sequentially at the most separation among observations, random forest allows variations of predictors among trees, which enables higher utilization for various combinations of regressors in parallel and therefore more diversity in tree structure. The parameters that control the two types of randomness are the number of bootstraps (i.e., the number of

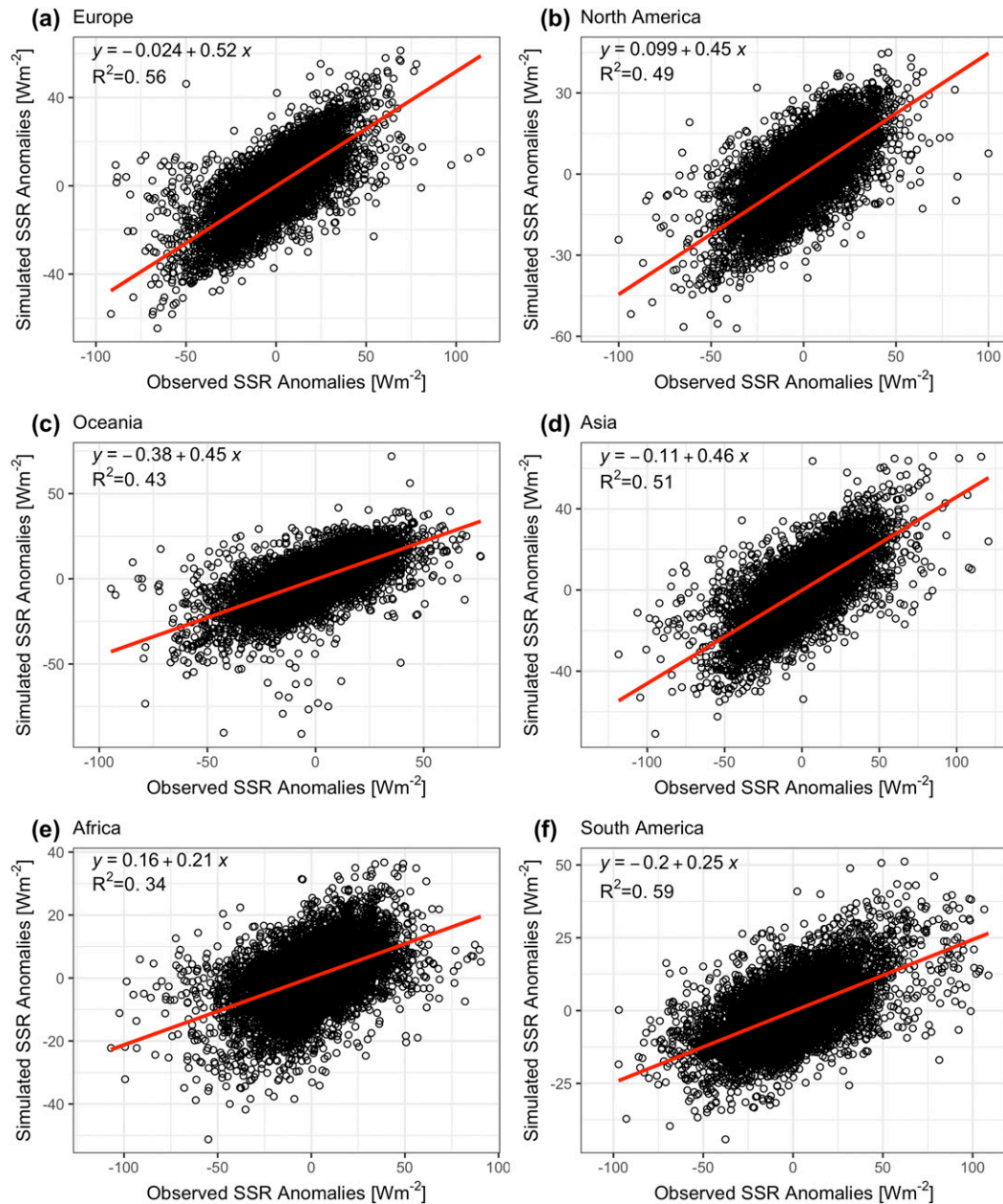


FIG. 2. Simulated against observed monthly anomalies of SSR. The red line is obtained by regressing simulated SSR on observed SSR, displayed together with the corresponding equation and R^2 (the coefficient of determination of the regression).

independent samples used to train the model—*ntree*) and the degree of feature randomness (i.e., the number of candidate independent variables to split at in each node—*mtry*). Through a parameter tuning process, the number of bootstraps is set to 700 and the degree of feature randomness is set to 8 (see Text S2 in the online supplemental material for details).

A range of variables are used as predictors for SSR variation, including climatic variables that are closely related to variations of SSR and geographical characteristics of locations, together with temporal stamps of the records. [Table 1](#)

summarizes the symbols and detailed definitions of all predictors. Overall SSR variability has two profound components: seasonal variations and interannual long-term variations. In this study we focus on investigating the latter ones, which are commonly represented by the linear trends of deseasonalized SSR series, or SSR anomalies. In other words, for each observation station, monthly SSR values are subtracted by the monthly averages over the period during which the station exists. All predictors are deseasonalized accordingly with respect to the same reference period as given by the available

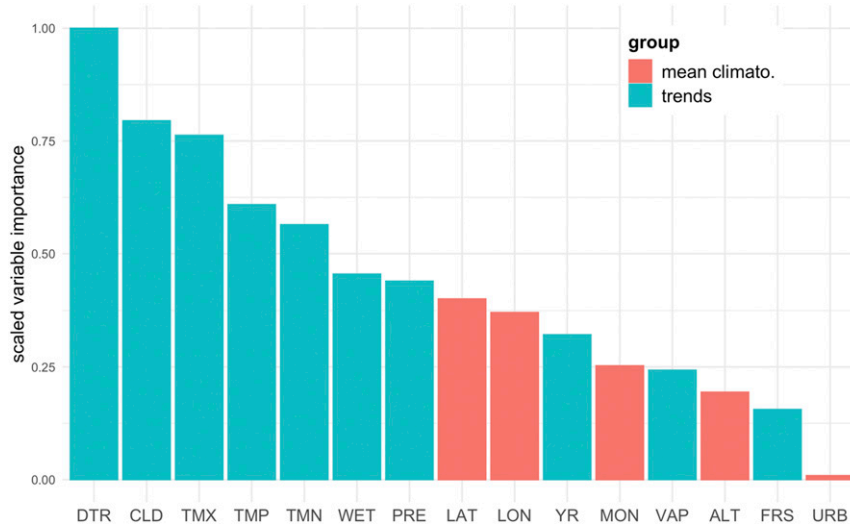


FIG. 3. Random forest permutation variable importance. All importance values are scaled by the largest importance. The variables are classified into two groups: the climate variables that govern the SSR trends (in turquoise) and the static variables (e.g., latitude) and month (a proxy of TOA) that determine the mean climatologies of SSR and have no influence on the trends (in coral).

SSR data. By using climate anomalies rather than the absolute magnitudes of the climate observations, the seasonal and latitudinal variabilities are separated from the overall variability, such that only the long-term variability is fed into the model. To evaluate the model performance, three error metrics are calculated: mean absolute error (MAE), root-mean-square error (RMSE), and R squared (R^2). They are given by the following equations:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{z}_i - z_i|, \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{z}_i - z_i)^2},$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{z}_i - z_i)^2}{\sum_{i=1}^n (z_i - \bar{z}_i)^2},$$

where n is the total number of predictions, \hat{z}_i are predicted values for station i , and z_i are observed values for station i . No percentage error measures are used because they tend to be biased toward very large values if SSR observations are close to zero, which could be common in winter for high-latitude areas.

b. Datasets

Ground-based all-sky SSR observations were obtained from the Global Energy Balance Archive (GEBA; Wild et al. 2017). The GEBA dataset contains monthly SSR values from approximately 1500 observation stations scattered on all continents except for Antarctica. Europe accounts for more than one-third of the worldwide stations (516 out of 1486 global

stations), making it the most extensively and intensively covered continent by radiation stations (refer to Fig. S1 in the online supplemental material for a global station distribution map and Table 2 herein for a summary of the total numbers of stations and monthly observations on each continent). On the other hand, taking into account the broad area of South America and its limited number of observation stations, South America has the sparsest coverage of observation stations. The dataset has an unparalleled temporal coverage, which extends from the early 1950s until 2013.

GEBA collects shortwave irradiance observations directly from pyranometer measurements, the quality of which could be affected by a few factors, for instance, the calibration procedure of pyranometer windows and the random error of single pyranometer readings. Gilgen et al. (1998) estimated the relative random error (root-mean-square error divided by mean) of the downward solar radiation values in GEBA at 5% for a monthly mean and 2% for a yearly mean. This means that the measurement errors in GEBA are negligible and it can therefore be a reliable data source for climate research. This dataset has been previously examined for temporal homogeneity by Sanchez-Lorenzo et al. (2013) and widely used in the literature since the 1990s (see, e.g., Ohmura and Gilgen 1993; Arking 1996; Liepert 2002; Nabat et al. 2014; Wang et al. 2014; Cherian et al. 2014; He et al. 2018).

The climatic variables used as predictors for SSR are available from the Climate Research Unit time series data version 4.04 (CRU-TS v.4.04; Harris et al. 2020). The CRU dataset provides high-resolution ($0.5^\circ \times 0.5^\circ$) gridded data of monthly observations for a range of meteorological variables over the period 1901–2019 (Table 1). The CRU dataset is interpolated from extensive networks of weather station observations and homogenized by sophisticated techniques, and the data quality

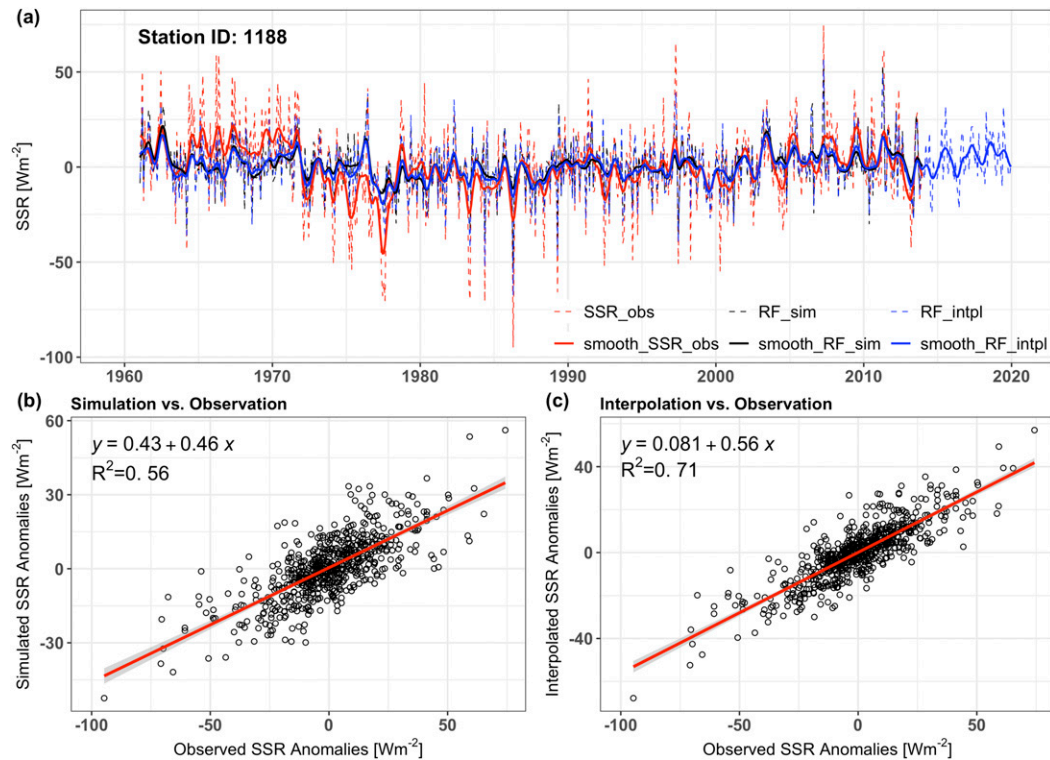


FIG. 4. Model evaluation for one station in GEBA: station 1188, Locarno-Monti, Switzerland. (a) Monthly SSR anomaly series for observations (red dashed line), simulations (black dashed line), and interpolations (blue dashed line), together with their corresponding smoothed series (solid lines) using a 12-month Gaussian kernel. The series are expressed as anomalies from the 1961–2013 mean. Observation and simulation series exist only for the period 1961–2013; interpolation series extends until 2019 from the random forest model predictions. (b) Scatterplot of simulated vs observed series. The red line is obtained by regressing simulations on observations. Regression equation and R^2 are shown for the regression line. (c) As in (b), but for the interpolated vs observed series.

is of a high standard. A $0.5^\circ \times 0.5^\circ$ global altitude map is available from the U.S. National Geophysical Data Center (NGDC et al. 1995). The Global Rural–Urban Mapping Project (GRUMP) dataset V1 (CIESIN 2004) provides a 30 arc-second urban extents grid for all land areas except Antarctica and parts of the Greenland ice sheet. Each grid is classified as either rural or urban based on 1995 data. Although urban extents have changed considerably over time, unfortunately we do not have urban extent data that are updated continuously. To merge the GRUMP with the CRU data, each $0.5^\circ \times 0.5^\circ$ grid cell was obtained as the mean of the 3600 GRUMP values contained in the cell.

The training dataset is obtained by collocating the GEBA stations with corresponding grid boxes in the gridded datasets, such that for each station in the GEBA, a range of predictor variables as well as the SSR anomalies are matched by month. In terms of data quality control, we deleted observations with abnormally low W m^{-2} SSR values in summer times, which are clearly measurement errors since SSR should be at high levels during that time of the year. Moreover, because the RF uses bootstrapping to generate a large set of independent samples of training data, the algorithm is specifically designed to be relatively insensitive to sporadic outliers in the training data.

For the interpolation, all three gridded datasets (CRU, NGDC, and GRUMP) are merged together and provide the input variables for the trained RF model. The interpolation data cover the period 1961–2019, with the end year decided by the extension of the CRU dataset.

3. Results

a. Estimation and evaluation

A continent-by-continent tenfold cross-validation (CV) was implemented on the training dataset, such that each continent's distinctive characteristics are sufficiently accounted for. Otherwise, if a global universal model is trained, the trained model would be biased toward the SSR dynamics of the continent with the most concentrated observations (i.e., Europe). Tenfold CV means to partition the training dataset into 10 equal-size subsamples. Of the 10 subsamples, a single subsample is retained as the validation data for testing the model, and the remaining 9 subsamples are used as the data to train the model. The cross-validation is repeated 10 times, such that each of the 10 subsamples is used once as validation data. Combining together simulations for all 10 validation subsets generates a complete out-of-sample simulation (i.e., the data

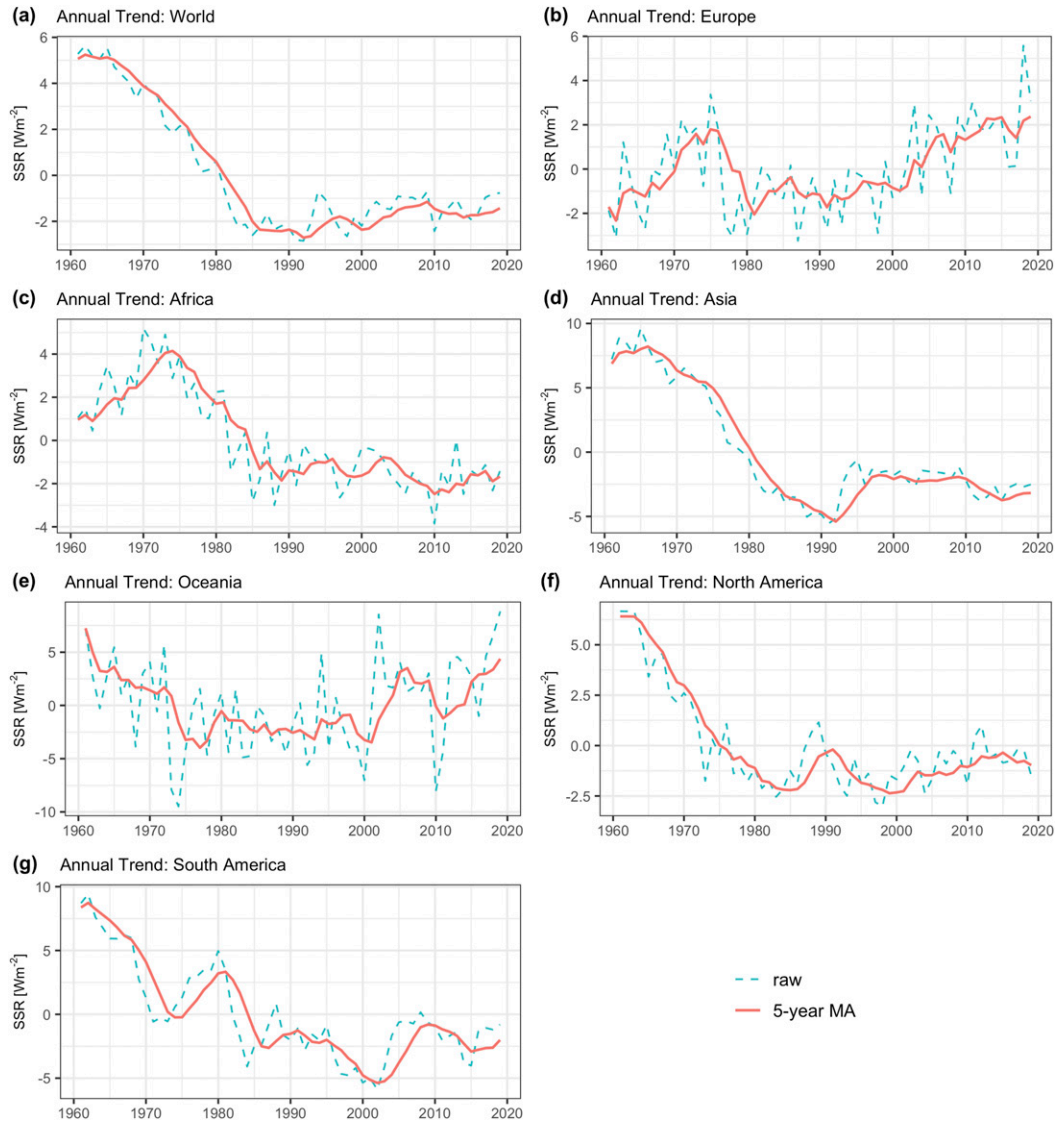


FIG. 5. Annual anomalies over the period 1961–2019 for each continent and the world. The raw series is shown by the green dashed line, and the 5-yr moving average (MA) series is shown by the coral solid line. When the number of years is less than 5, a partial moving average is used for the available years up until the time point.

used for validation are not used in training) for each observation record.

Table 3 shows the error measures of the predicted monthly SSR anomalies against observations for each continent. Since this study is focused on evaluating multidecadal trends of SSR, stations existing for less than 15 years are excluded from the performance evaluation to alleviate deviations brought by stations that only existed briefly. The mean absolute errors vary from 7.90 W m^{-2} (Europe) to 11.87 W m^{-2} (South America), with a global average of 9.92 W m^{-2} . The model accuracies revealed from mean absolute errors and from root-mean-square errors coincide among continents—that is, the continent associated with the lowest mean absolute error also has the lowest root-mean-square error, and equivalently so for the continent with the largest mean absolute error. The R^2 range

from the lowest 0.34 (Africa) to the highest 0.59 (South America), with a global average of 0.49. This indicates that our RF model on average captures approximately half of the global SSR interannual variations. The R^2 values are generally in line with MAE and RMSE for continent evaluation. It is worth noticing that although the bias indicated by MAE and RMSE is large for South America, the R^2 shows that almost 60% of annual variability is captured by the continental model. The large values of MAE and RMSE may be due to the spread between predicted and observed SSR anomalies for SSR anomalies larger than 50 W m^{-2} (Fig. 2).

RF is a data-intensive machine learning approach that learns model features solely based on input data and does not rely on any presumptions about model structure or specifications, the

TABLE 4. Decadal linear trends for annual SSR averaged over individual continents and the global land surfaces. After significant breakpoints are detected, the linear trend of the continental average time series is estimated by least squares for each segment. The first row for each continent or the world shows the trend over the whole period; subperiods split by the breakpoints are shown in the following rows. For instance, the global average trend shows a negative slope of $-1.15 \text{ W m}^{-2} \text{ decade}^{-1}$ over 1961–2019 and one breakpoint in 1982 is significantly detected. Split by 1982, a negative trend of $-3.07 \text{ W m}^{-2} \text{ decade}^{-1}$ is reported over the first segment and a positive trend of $+0.33 \text{ W m}^{-2} \text{ decade}^{-1}$ is reported over the latter segment 1982–2019.

Continent	Segment	Slope ^a	Slope std dev	<i>t</i> value	Pval	Pval.symbol ^b
World	1961–2019	−1.152	0.131	−8.812	0.000	***
	1961–81	−3.068	0.172	−17.874	0.000	***
	1982–2019	0.328	0.079	4.164	0.000	***
Europe	1961–2019	0.466 ^c	0.138	3.383	0.001	**
	1961–76	2.791	0.759	3.675	0.002	**
	1977–2019	1.169	0.169	6.935	0.000	***
Africa	1961–2019	−0.985	0.112	−8.772	0.000	***
	1961–69	1.991	1.175	1.694	0.134	
	1970–81	−3.255	0.695	−4.681	0.001	***
	1982–2019	−0.221	0.139	−1.587	0.121	
Asia	1961–2019	−1.920	0.215	−8.929	0.000	***
	1961–92	−5.383	0.231	−23.287	0.000	***
	1993–2019	−0.643	0.165	−3.908	0.001	***
Oceania	1961–2019	0.339	0.327	1.039	0.303	
	1961–72	−1.440	2.660	−0.542	0.600	
	1973–2019	1.718	0.395	4.353	0.000	***
North America	1961–2019	−0.826	0.150	−5.521	0.000	***
	1961–77	−5.068	0.483	−10.496	0.000	***
	1978–2019	0.249	0.121	2.051	0.047	*
South America	1961–2019	−1.529	0.199	−7.685	0.000	***
	1961–2003	−2.919	0.243	−12.003	0.000	***
	2004–19	−0.756	0.594	−1.272	0.224	

^a Slope unit: $\text{W m}^{-2} \text{ decade}^{-1}$.

^b Significance symbol representation: *** indicates $p < 0.001$, ** for $p < 0.01$, * for $p \leq 0.05$, · for $p \leq 0.1$, and no symbol if $p > 0.1$.

^c The linear trend is smaller than in both segments due to the large jump in 1976 (refer to Fig. 5b).

performance of which is highly dependent on input training data, from perspectives of both quality and quantity. Note that the RF model was trained for per continent; that is, data characteristics for each continent affect its performance. Specifically, possible reasons that lead to the lower RF performance in South America and Africa include, but are not limited to, the relatively smaller amount of data available and/or the lack of homogeneity in the continents' data. The scatterplots of simulated against observed anomalies provide further graphical confirmation of this (Fig. 2). The continents with better performance show highly clustered points alongside the regression line, for instance, Europe and North America (Figs. 2a,b), in contrast to Africa and South America (Figs. 2e,f), which show more scattered points away from the regression line.

From the permutation variable importance analysis (Fig. 3), we see that the most important variable in determining long-term SSR variability is *diurnal temperature range*. *Cloud coverage* is the second most important variable, followed by the temperature indices (maximum, average, and minimum monthly temperature). Note that the temperature indices and diurnal temperature range could be closely correlated;

however, this does not affect the tree structure of the random forest given its nonparametric nature, which does not depend on any functional form and therefore has no problem of collinearity.

Since we trained the model on SSR anomalies, which are deseasonalized series and are therefore unaffected by seasonal variability, the seasonal indicator *month* is unsurprisingly found to be a minor factor influencing SSR long-term variability. Precipitation affects SSR long-term trends as well, while vapor pressure and frost days provide little explanatory power for SSR interannual variability. We categorize these meteorological variables as trends determining variables. The other group of variables includes geographical coordinates of observations (latitude and longitude), months of observations, altitudes, and urbanization indicators. We name this group as mean climatology predictors because they decide average climatic characteristics of locations. Given that the target variable is SSR anomalies, the mean values are zero for every location. Not surprisingly, the mean climatology predictors are less important, if not entirely irrelevant, compared to the meteorological predictors that govern the SSR long-term trends.

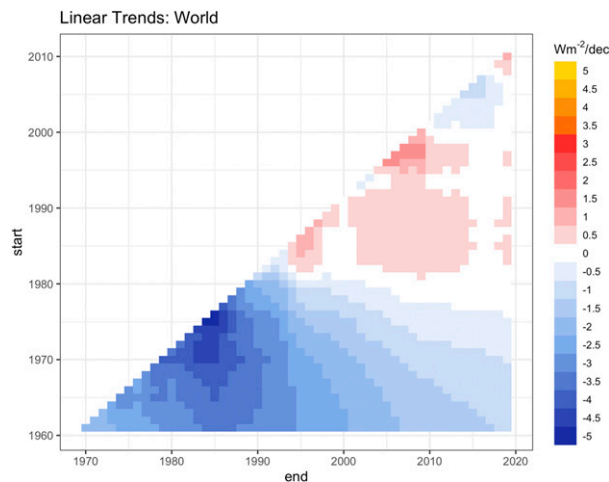


FIG. 6. Decadal trend raster estimated on the global average SSR anomalies series shown in Fig. 5a. The y axis denotes the start years, and the x axis denotes the end years of the periods considered in the calculation of the trends. The minimum trend length is 10 years.

b. Interpolated SSR dataset trend analysis

The tenfold cross-validation shows the RF model's capability in out-of-sample estimation of SSR. To make use of the SSR information in the GEBA dataset exhaustively, all available SSR observations (instead of 9 out of 10 subsamples used in the tenfold CV) are used to train the RF model per continent. The trained RF models are then applied to the interpolation data. The result of the interpolation is a $0.5^\circ \times 0.5^\circ$ gridded monthly SSR dataset over the period 1961–2019 with complete coverage of global land areas. The resolution and the time span are determined by the input climatic variables from the CRU datasets. Note that the GEBA (training) dataset only has a substantial amount of SSR anomalies until 2013. By extending the prediction period up until 2019, we assume the prediction relationship between SSR and the predictors remains the same both prior to and after 2013.

A station in the GEBA dataset (station 1188) is given as an example to illustrate the interpolation procedure. The station is located at Locarno-Monti (46.17°N , 8.78°E), Switzerland, and is one of the longest-standing radiation stations in Europe. We started with observations (shown by the red lines in Fig. 4a) from the GEBA dataset. A tenfold CV was implemented on the data for Europe and thereby generating out-of-sample estimations corresponding to the observation records. The estimation series are shown by the black lines. Then all available data on Europe were used to train an RF model that was later used in the interpolation for each grid box in Europe. By extracting the values in the grid box in which station 1188 is located, interpolation series are obtained (shown by the blue lines). We see that the interpolation series approximate the observations more precisely than the simulations. This is because the interpolation is an in-sample forecast that uses 10% more data than the

simulations that are based on the tenfold CV. By comparing the blue lines with the red lines, we see that the interpolation series are able to capture the observed SSR variation with reasonable accuracy, indicating the robustness of the RF model and the reliability of the generated interpolation series. Figures 4b and 4c show scatterplots for simulations against observations and interpolations against observations, respectively. The scatterplot for interpolations has more concentrated points alongside the regression line and has a larger value of R^2 (0.71 vs 0.56 for Fig. 4c vs Fig. 4b). The improvement of the interpolation precision demonstrates the reliability of using the random forest model to predict long-term SSR variability.

1) TREND OVERVIEW OF THE ENTIRE PERIOD

Annual anomalies for global land areas and each continent are shown in Fig. 5. To account for the areas of grid boxes reducing with increasing latitude, the global/continental average SSR are calculated as the weighted averages of values for all grid boxes in the defined areas, weighted by the *cosine* of latitudes. The global average SSR exhibits rapid dimming trends from 1961 until the mid-1980s, followed by a moderate reversal. Asia and North America manifest highly similar trends with the world average trends, reaching the lowest level around the late 1980s or early 1990s, then a mild reversal occurs and the continents enter into brightening periods. Oceania exhibited significant dimming trends until the mid-1970s; between then and the early 2000s, SSR stagnates and shows no clear trends. Entering into the twenty-first century, profound brightening occurs. Significant dimming is observed in South America until the early 2000s. Africa experienced a short period of brightening from 1960 to 1970, and profound dimming trends followed until 1990, after which SSR is stable and shows no obvious trends. Europe shows overall generally brightening trends over the whole period 1961–2019, except for a short decline between 1975 and 1980. Significant brightening has been observed from the 1980s up until now. From the continental average trends and global average trends, we observe that except for Europe, all continents experienced a relatively long (more than 10 years) dimming period. Asia and North America are the two largest contributors to the global dimming trends, while Europe and Oceania are important drivers of the brightening trends.

Although the global trend reversal takes place in the mid-1980s, each continent shows different breakpoints in terms of trend reversal. The fact that continents show strikingly different features with respect to their historical trends makes it of interest to detect structural changes in the long-term SSR trends for each continent separately. Breakpoints were detected based on moving sums (MOSUMS) of recursive and least squares residuals for annual mean values, which, despite losing monthly temporal details, enables an investigation of interannual variation and time series segments (Forkel et al. 2013). Time series segment and structural change detection is widely used in climate research (see, e.g., Cró and Martins 2017; Adedoyin et al. 2020; Vu et al. 2019). Readers are referred to Holben (1986) and Bai and Perron (1998, 2003) for detailed implementation of the algorithm.

Any breakpoints with the Mann-Kendall trend test significant at 5% are identified and trends for the separated segments are

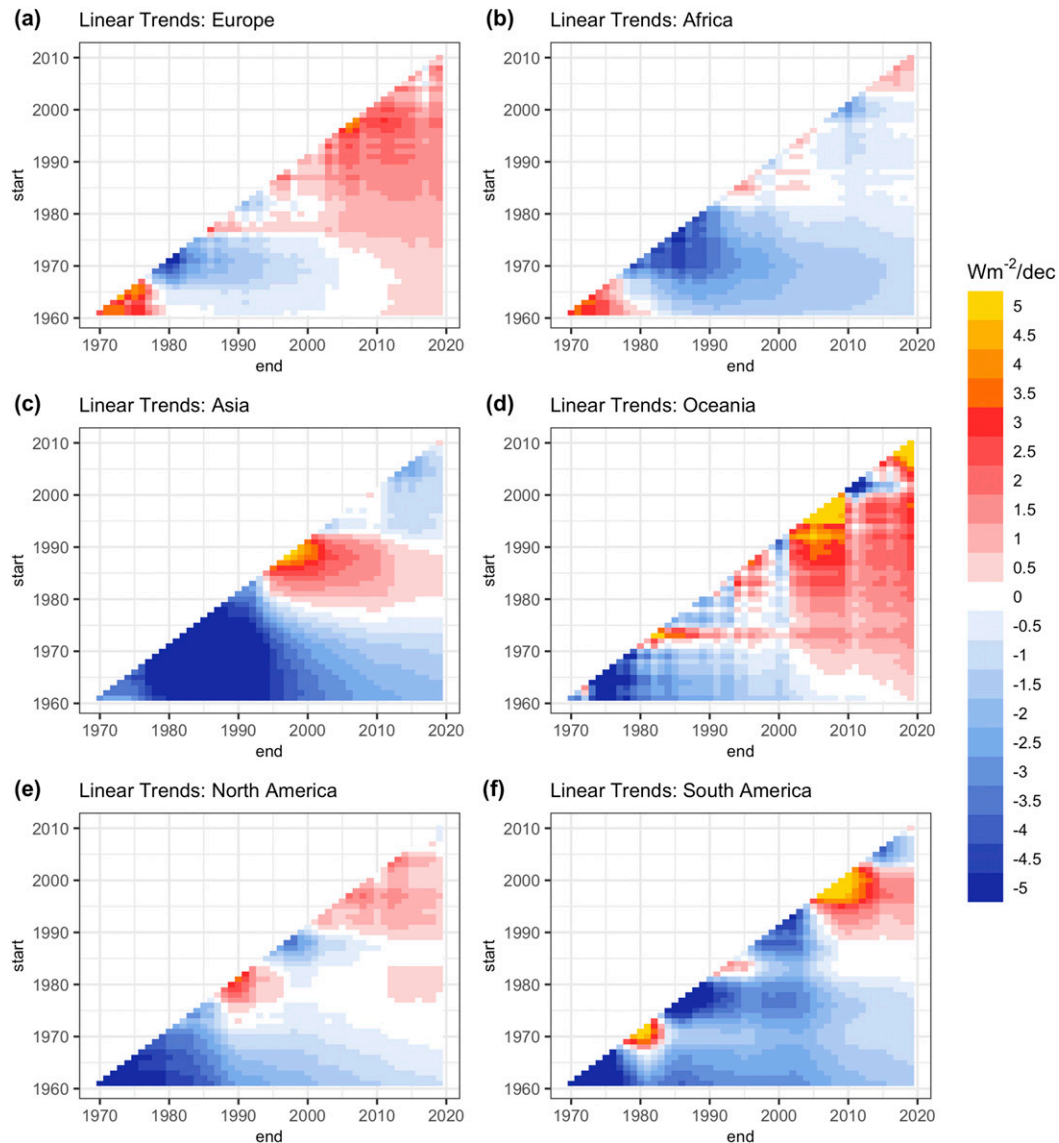


FIG. 7. As in Fig. 6, but showing decadal trend rasters for continents.

reported in Table 4. Asia and North America, with decadal linear trends of -5.38 and $-5.07 \text{ W m}^{-2} \text{ decade}^{-1}$, respectively, are the two largest contributors to the global dimming, whereas the global brightening is mostly attributed to the increasing trends in Europe and Oceania. Focusing on the latest trends after the breakpoints, most continents show significant positive trends (Europe, Oceania, and North America) or nonsignificant negative trends (Africa and South America). The only continent with significant negative trends in recent periods is Asia. Nevertheless, the negative trends in Asia have been significantly alleviated from -5.38 to $-0.63 \text{ W m}^{-2} \text{ decade}^{-1}$ since 1993. It is worth noting that aside from Africa, which has two breakpoints (1969 and 1981), only one breakpoint is detected for each continent. In what follows, this paper will investigate in more detail the periods after the latest detected breakpoints for each continent.

Given the large variability of SSR, the linear trends could be significantly affected by a different choice of start and end time. To avoid this bias, a running-trend estimation was implemented on the annual global average series for all possible segments equal to or longer than 10 years. The global decadal trend raster (Fig. 6) corroborates the annual anomaly series (Fig. 5a). The most negative trends are found between 1970 and 1980 (referring to start points and thereafter in the description of trend rasters), in contrast to the most positive trends between 1995 and 2000. After entering into the twenty-first century, the trends fade away. The continental running-trend rasters (Fig. 7) show that Europe is mostly dominated by positive long-term trends. Significant negative trends are observed in Asia from 1961 to 1980, which is followed by a cluster of positive trends from 1985 to 1990. The most recent SSR data show slightly negative trends in Asia. North America experiences a transition from dimming

TABLE 5. Decadal linear trends for seasonal average SSR averaged over individual continents and the global land surfaces over the periods from the respective latest detected breakpoints until 2019.

Continent	Seasons	Slope ^a	Slope std dev	<i>t</i> value	Pval	Pval.symbol ^b
World (1982–)	DJF	0.060	0.108	0.558	0.580	
	MAM	0.520	0.133	3.909	0.000	***
	JJA	0.431	0.130	3.323	0.002	**
	SON	0.232	0.110	2.118	0.041	*
Europe (1977–)	DJF	−0.166	0.142	−1.168	0.249	
	MAM	1.743	0.356	4.898	0.000	***
	JJA	2.535	0.405	6.256	0.000	***
	SON	0.587	0.261	2.247	0.030	*
Africa (1982–)	DJF	−0.156	0.274	−0.571	0.571	
	MAM	0.002	0.207	0.008	0.993	
	JJA	−0.576	0.214	−2.688	0.011	*
	SON	−0.267	0.207	−1.291	0.205	
Asia (1993–)	DJF	−0.648	0.252	−2.568	0.017	*
	MAM	−0.454	0.304	−1.492	0.148	
	JJA	−0.465	0.321	−1.447	0.160	
	JJA	−0.465	0.321	−1.447	0.160	
Oceania (1973–)	DJF	1.351	0.734	1.841	0.072	.
	MAM	1.388	0.719	1.930	0.060	.
	JJA	1.508	0.623	2.421	0.020	*
	SON	2.329	0.591	3.943	0.000	***
North America (1978–)	DJF	0.057	0.202	0.282	0.780	
	MAM	0.553	0.198	2.795	0.008	**
	JJA	0.214	0.195	1.100	0.278	
	SON	0.163	0.223	0.729	0.470	
South America (2004–)	DJF	−0.040	0.856	−0.046	0.964	
	MAM	−1.166	0.811	−1.438	0.172	
	JJA	−0.813	0.871	−0.933	0.366	
	SON	−1.138	0.843	−1.350	0.198	

^a Slope unit: $\text{W m}^{-2} \text{decade}^{-1}$.

^b Significance symbol representation: refer to Table 4.

to brightening around 1990. South America shows mainly dimming trends, with a cluster of brightening trends from 1995 to 2000. SSR trends in Oceania are highly volatile and show a lack of stability; no prevalent signs are observed.

The results shed light on how SSR trends evolve over time. However, because the data are aggregated over areas, either globally or continentally, opposite trends cancel each other out when calculating the area averages, resulting in less distinct trends overall. Therefore, the regional distribution of trends with $0.5^\circ \times 0.5^\circ$ resolution has also been investigated. In what follows, decadal trends of annual and seasonal SSR, both on a global and continental scale, are discussed based on their respective latest detected breakpoints as starting points. Furthermore, the spatial distribution of the trends is visualized in annual and seasonal maps.

2) LATEST SUSTAINABLE TREND

Based on the structural breaks in trends, the latest sustainable trends (i.e., the period after the latest detected breakpoints until 2019) for the globe and for each continent are inspected in this section. Their spatial and seasonal patterns are presented as trend maps for each season and the entire year. Seasons are defined as follows: winter [December–February

(DJF)], spring [March–May (MAM)], summer [June–August (JJA)], and autumn [September–November (SON)]. Seasonal trends after the detected breakpoints for the globe and each continent are reported in Table 5, which will be elaborated in the following sections.

(i) World

The global brightening starts from 1982 and continues onward, at a moderate rate of $+0.33 \text{ W m}^{-2} \text{decade}^{-1}$, only about one-tenth of the dimming rate before ($-3.07 \text{ W m}^{-2} \text{decade}^{-1}$). Note that the worldwide trend reversal point in 1982 is detected based on the worldwide annual SSR anomalies. Although the global average SSR has shown a positive trend since 1982, on a regional level significant dimming is observed in certain regions, particularly in the eastern United States, South Asia, and the Pacific island countries (Fig. 8a). Africa shows dimming-neutral trends in general, with slight brightening observed in eastern and northern Africa. On the other hand, widespread brightening is observed in Europe, northern Asia, Oceania, and South America. North America shows a blend of dimming and brightening trends. Coastal areas in the United States show negative trends, whereas the inland United

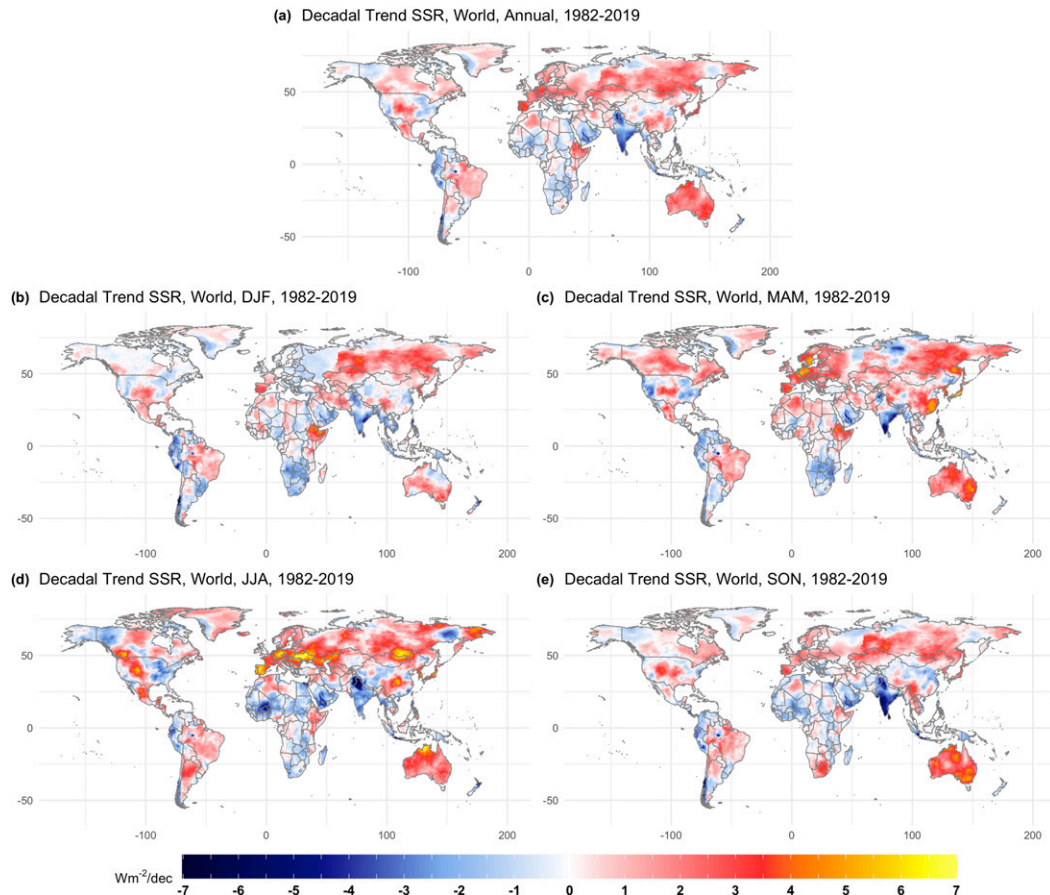


FIG. 8. Linear trends of the annual and seasonal average SSR over the globe during the period 1982–2019 (after the breakpoint in 1981). The values are estimated for each $0.5^\circ \times 0.5^\circ$ grid and expressed as $\text{W m}^{-2} \text{decade}^{-1}$. (a) Decadal trends for annual average SSR. (b)–(e) Decadal trends for individual seasons. The seasons are defined as Northern Hemispheric spring (MAM), summer (JJA), autumn (SON), and winter (DJF, dated according to the year of January and February).

States shows positive trends. Canada and Greenland show generally widespread positive trends of SSR.

The seasonal maps show that Northern Hemispheric spring and summer demonstrate the largest trend variability (Figs. 8c,d); that is, both the strongest positive and negative trends are observed during these periods. On the other hand, Northern Hemispheric winter shows modest trends; both positive and negative trends are diminished as expected due to the lower absolute SSR values in Northern Hemispheric winter in the extratropics. Note that the decadal trends are squeezed into a range of -7 to $7 \text{ W m}^{-2} \text{decade}^{-1}$, which means that any extreme values falling out of the range are assigned as $-7 \text{ W m}^{-2} \text{decade}^{-1}$ if a value is smaller than $-7 \text{ W m}^{-2} \text{decade}^{-1}$, and $7 \text{ W m}^{-2} \text{decade}^{-1}$ if a value is larger than $7 \text{ W m}^{-2} \text{decade}^{-1}$.¹ It is also noteworthy that dimming

trends are observed throughout the year in India, with the most negative trends observed in autumn. In the Southern Hemisphere the strongest trends are observed in the winter (JJA) and spring (SON), showing widespread positive trends in South America and Oceania.

(ii) Europe

Recent long-term trends in Europe show an increase of average SSR at the rate of $1.17 \text{ W m}^{-2} \text{decade}^{-1}$ starting from the breakpoint in 1977 and up until recently (Table 4). At a seasonal scale, except for winter, all the other seasons show significant positive average trends, with summer having the strongest rate at $2.54 \text{ W m}^{-2} \text{decade}^{-1}$ (Table 5).

The spatial distribution of annual mean SSR shows a widespread increase with the most positive area located in the central European domain (CED) (Fig. 9a), defined as land areas from 45° to 55°N and from 10°W to 35°E , namely Germany, southern Poland, Slovakia, and Ukraine. The mean annual trend for the CED is $2.18 \text{ W m}^{-2} \text{decade}^{-1}$, which is about $1.01 \text{ W m}^{-2} \text{decade}^{-1}$ higher than the annual trend for the entire continent. In fact, a larger difference

¹ As we will see in the forthcoming result (Fig. 15), the largest concentration of the trends falls within $[-4, 4] \text{ W m}^{-2}$, i.e., the upper and lower 1.5 standard deviations from the mean, making it reasonable to visualize values in a truncated range of $[-7, 7] \text{ W m}^{-2} \text{decade}^{-1}$ in order to better see the small absolute values.

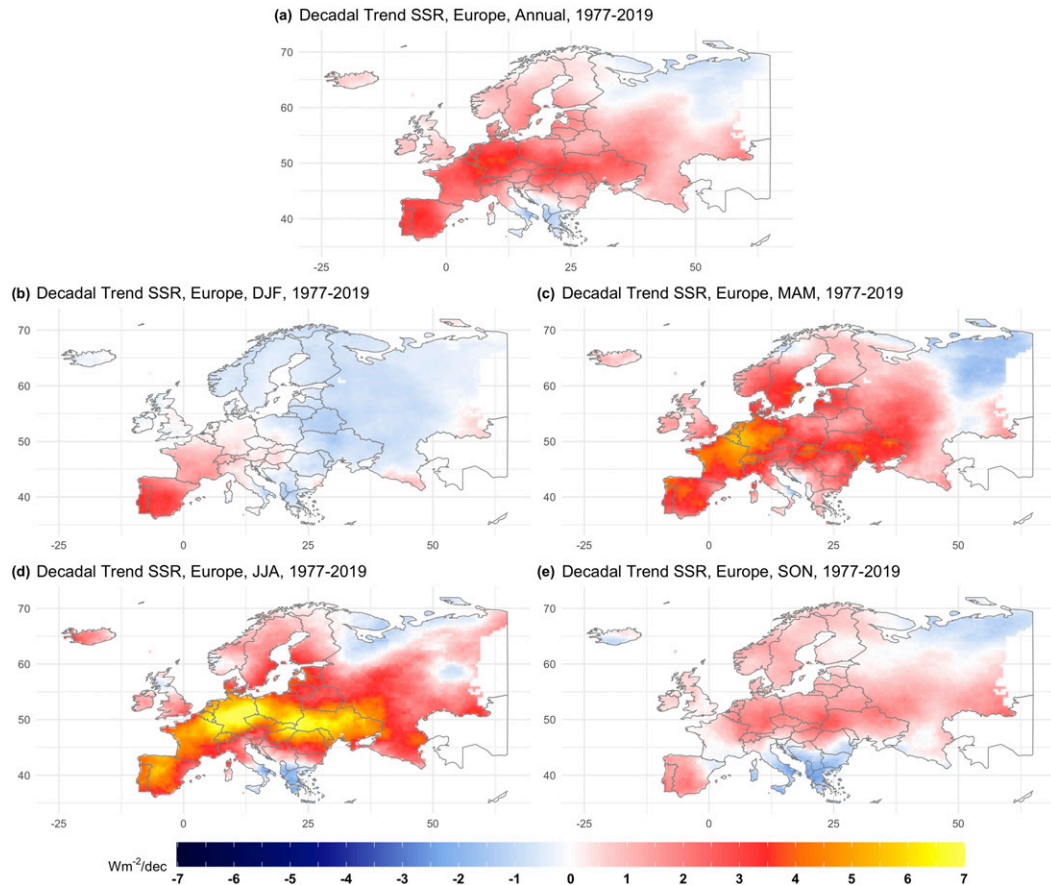


FIG. 9. Linear trends of the annual and seasonal average SSR in Europe during 1977–2019. Unit: $\text{W m}^{-2} \text{decade}^{-1}$.

of $2.35 \text{ W m}^{-2} \text{decade}^{-1}$ (4.59 vs $2.24 \text{ W m}^{-2} \text{decade}^{-1}$ for decadal trends over CED and the whole European land area, respectively) is observed if only summer (JJA) is considered. The spatial pattern is in agreement with Sanchez-Lorenzo et al. (2017), who reported higher rates of annual and seasonal SSR for the CED as compared to the entire continent. In particular, they documented a $1.0 \text{ W m}^{-2} \text{decade}^{-1}$ higher rate for the CED as compared to the whole continent for the annual series over the period 1983–2010, in contrast to a larger difference of $2.2 \text{ W m}^{-2} \text{decade}^{-1}$ for the summer counterpart. It is worth noting that in summer, the trends in the Mediterranean area are actually negative, in contrast to the general positive trends for the rest of the continent. In particular, the Mediterranean area shows negative trends throughout the year except during spring. The negative trends are also reported in Sanchez-Lorenzo et al. (2017). Furthermore, our study shows no significant European-average trends in winter; however, at a regional level, the CED shows slightly negative trends, and the positive trends observed in Spain and France are also noteworthy. The continuous increase in the winter series in Spain is also reported by Sanchez-Lorenzo et al. (2013).

(iii) Africa

Our results show that Africa has exhibited widespread dimming trends in SSR since the breakpoint in 1982 at a moderate

rate of $-0.22 \text{ W m}^{-2} \text{decade}^{-1}$ (Table 4; Fig. 5c). However, the rate is statistically insignificant. Central eastern and northern Africa show weak brightening trends, while southern Africa shows slightly dimming trends. The results are in line with Gilgen et al. (1998, 2009), who also reported negative decadal trends of SSR in southern Africa based on direct observations.

On a seasonal basis, significant negative trends are found in the Northern Hemispheric summer (JJA) at $-0.58 \text{ W m}^{-2} \text{decade}^{-1}$, which is confirmed by the predominantly dimming trends for the grid boxes in Africa (Figs. 10b–e; Table 5). The other seasons demonstrate no significant continental average trends. However, for certain subregions, significant trends exist. For instance, central eastern Africa shows primarily brightening throughout the year, and southern Africa shows widespread negative trends for all seasons except for spring (SON).

(iv) Asia

Large spatial variability is observed in Asia, and we see both dimming and brightening. Continental average SSR shows statistically significant negative trends for annual SSR at $-0.64 \text{ W m}^{-2} \text{decade}^{-1}$ after the breakpoint in 1993 (Table 4). Significant dimming is observed in the Middle East (e.g., Saudi

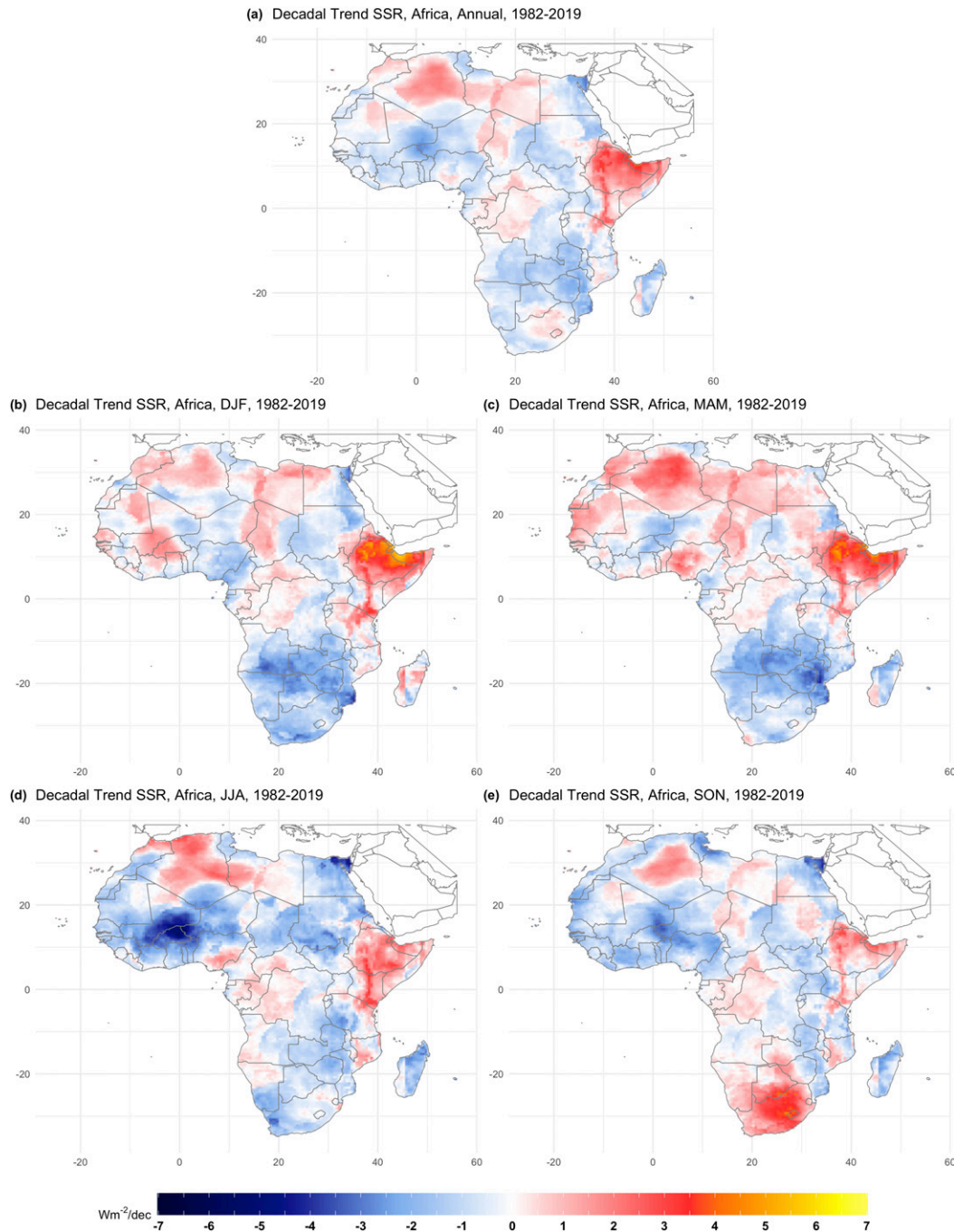


FIG. 10. Linear trends of the annual and seasonal average SSR in Africa during 1982–2019. Unit: $\text{W m}^{-2} \text{decade}^{-1}$.

Arabia and Yemen) and South Asia (India, Pakistan, etc.). India shows persistent dimming throughout the year, with the most negative trends observed in summer. Significant positive trends are observed in central Russia, Japan, North Korea, South Korea, southwestern China, etc. (Fig. 5a).

On a continental scale, the most negative trends take place in autumn, with a value of $-0.96 \text{ W m}^{-2} \text{decade}^{-1}$; winter shows a weaker negative trend at $-0.65 \text{ W m}^{-2} \text{decade}^{-1}$

(Table 5). Negative trends are also observed in spring and summer; however, these are statistically insignificant. The nonsignificant trends corroborate the findings of Yang et al. (2018), who indicated that the trend reversal, or jump, in the early 1990s in China reported by previous studies is falsely exaggerated because of instrument and operational changes. A series of procedures were performed by Yang et al. to obtain a homogenized series; based on these, they concluded

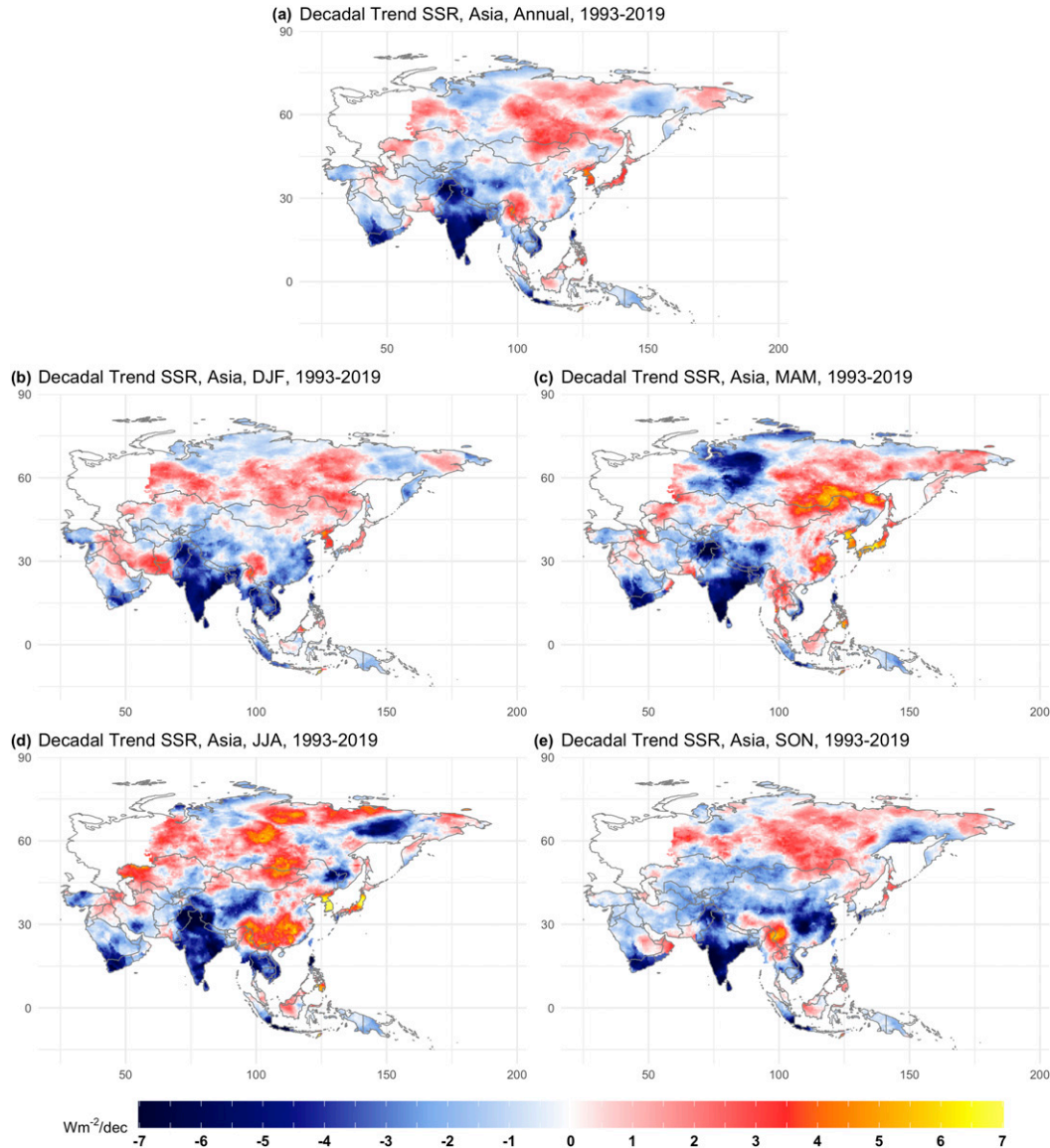


FIG. 11. Linear trends of the annual and seasonal average SSR in Asia during 1994–2019. Unit: $\text{W m}^{-2} \text{decade}^{-1}$.

that no significant trends are found in China over the period 1990–2016. Moreover, the interpolated dataset of the current paper to some extent mitigates the jump reported in the original observations by Moseid et al. (2020). In southeastern China, trends vary greatly from season to season. More specifically, significant negative trends are observed in autumn and winter, and substantial positive trends in spring and summer (Figs. 11b–e).

(v) Oceania

Caution needs to be taken for users to interpret the historic SSR trends in Oceania prior to 1988; the original data for Australia have been artificially detrended, as the meteorological service there was concerned that the instruments were drifting (Wild et al. 2005). Therefore, the trends prior to 1988, especially

the period from 1975 to 1980, might be flatter than they are supposed to be in reality (Fig. 5e).

Since northern Oceania (north of 18°S) has very few observation stations, the continental model has a bias when predicting SSR trends for the northern fringes. Therefore, we use a global model trained on worldwide observations to simulate SSR in Oceania. The latest sustainable SSR annual mean trend for Oceania is $1.72 \text{ W m}^{-2} \text{decade}^{-1}$ starting from 1973, which is the strongest among all continents (Table 4; Fig. 5e). The continent shows a high extent of homogeneity (Fig. 12). In particular, unanimous positive annual average trends are observed in Australia, with the western coast being more positive than the eastern coast. New Zealand shows a tendency of slightly negative trends. The seasonal maps show highly consistent positive trends; specifically, all

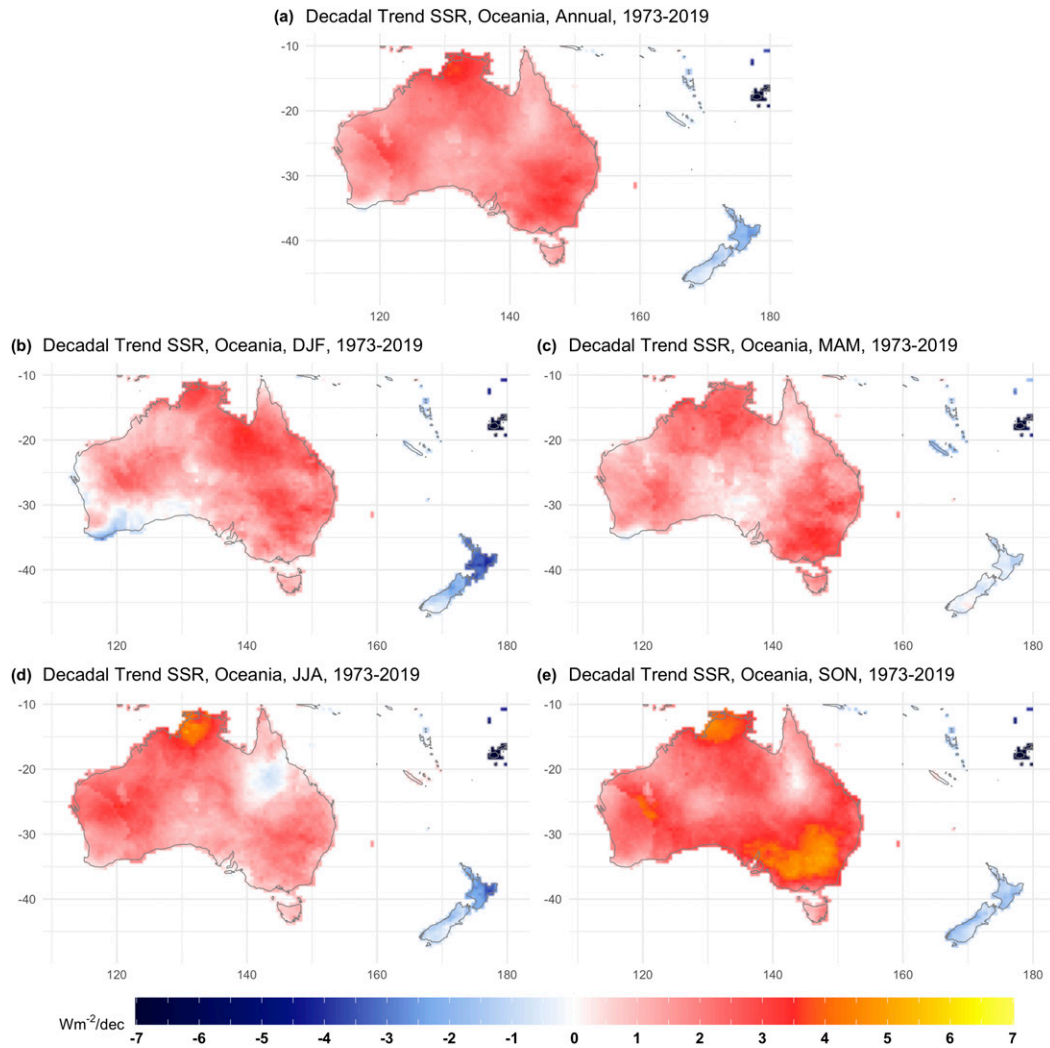


FIG. 12. Linear trends of the annual and seasonal average SSR in Oceania during 1973–2019. Unit: $\text{W m}^{-2} \text{decade}^{-1}$.

seasons show statistically significant positive continental average trends. The strongest continental average trend is found in spring (SON) at $2.33 \text{ W m}^{-2} \text{decade}^{-1}$ and the weakest continental average trend is found in summer (DJF) at $1.35 \text{ W m}^{-2} \text{decade}^{-1}$ (Table 5).

(vi) *North America*

North America has highly diverging trends; similar to Asia, both positive and negative trends are observed over the period 1978–2019, resulting in overall weak positive continental average trends at $0.25 \text{ W m}^{-2} \text{decade}^{-1}$ (Table 5; Fig. 5f). Slightly negative trends are observed in the eastern United States throughout the year, while the central United States shows positive trends (Fig. 13). Similar patterns were also reported by Gilgen et al. (2009), who include brightening in the central United States and dimming in the eastern United States. Moreover, our results of decadal trends are in line with the findings of Gilgen et al. (2009), who reported trend reversals for several observation sites in the United States. Canada

shows widespread brightening for annual SSR; the most positive trends are observed in spring, whereas winter shows prevalent neutral trends. Mexico shows primarily positive trends of SSR throughout the year.

(vii) *South America*

South America shows strikingly bipolar trends within the continent after the breakpoint in 2004; both highly negative and positive trends coexist (Fig. 14). Significant positive trends are observed in the northern fringes, while drastically negative trends are observed in the southern areas (20° to 40°S). Because the two opposite trend signs neutralize each other, the continental average SSR shows only slightly negative yet statistically insignificant trends (Table 4; Fig. 5g).

On a seasonal basis, northern South America shows persistently positive trends, with the strongest trends observed in summer (DJF) and autumn (MAM). Conversely, the southern areas show persistent negative trends of SSR, with the strongest trends found in autumn (MAM) and spring (SON). Since

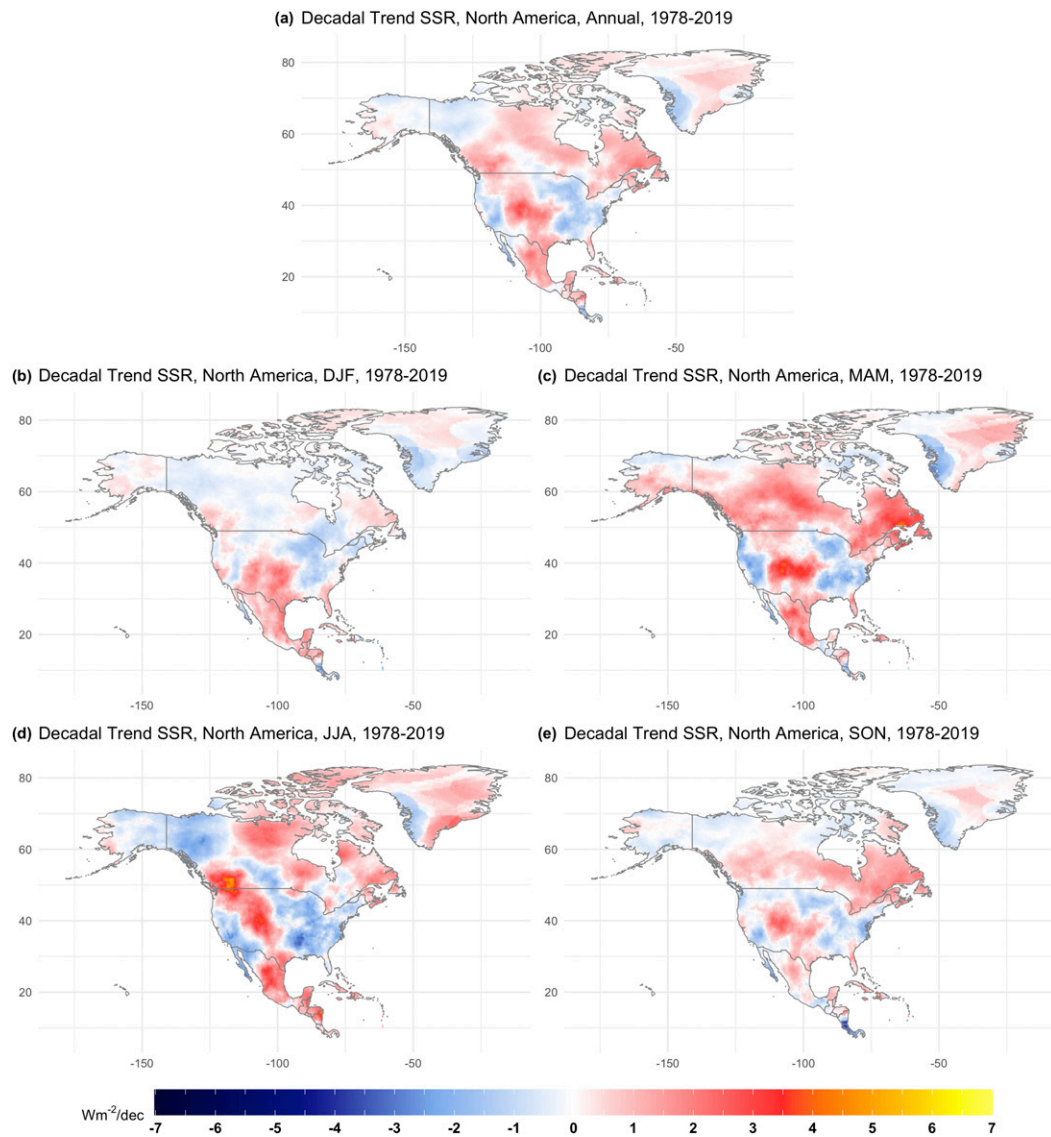


FIG. 13. Linear trends of the annual and seasonal average SSR in North America during 1978–2019. Unit: $\text{W m}^{-2} \text{decade}^{-1}$.

trends of drastic contrast coexists within the continent, continental average SSR shows no significant trends for all seasons (Table 5).

4. Discussion

This study interpolates a ground-based solar radiation observation dataset and constructs a dataset with complete global land coverage at high resolution covering nearly six decades. This enables a temporal and spatial trend analysis at the continental and the regional level, making it possible to obtain a comprehensive quantification of the long-term trends for the data-scarce continents (e.g., Africa and South America). Rather than relying principally on qualitative graphic visualization, a structural breakpoint detection algorithm based on moving sums (MOSUMS test) is applied

and provides quantitative identification of any structural changes in the long-term SSR anomalies. The results show that the global dimming–brightening transition takes place around 1982 on a global average level, which is in line with existing literature based on station observations. The two divided segments show a decrease over 1961–81 and an increase over 1982–2019 at -3.07 and $+0.33 \text{ W m}^{-2} \text{decade}^{-1}$, respectively. We further examined SSR trends on a continental scale. We found that the largest contributors to the global dimming are Asia and North America, while the two most important drivers for global brightening are Europe and Oceania.

To obtain an overview of the distribution of decadal trends for grid boxes within each continent, Fig. 15 shows box plots for annual trends as well as for seasonal trends for all grid boxes per continent. A longer box and a longer distance between whiskers indicate larger variability and more heterogeneity among grid

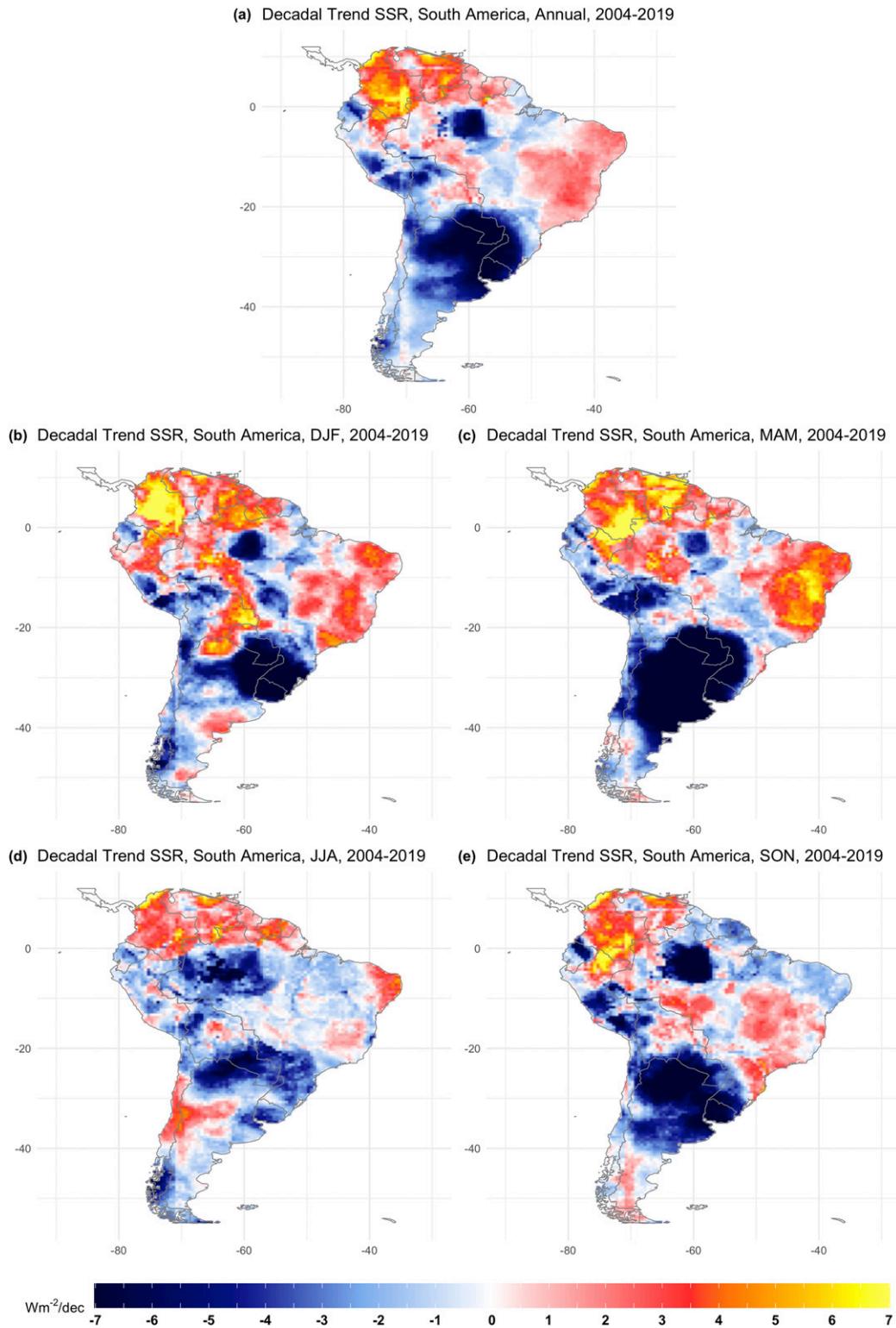


FIG. 14. Linear trends of the annual and seasonal average SSR in South America during 2004–19. Unit: $\text{W m}^{-2} \text{decade}^{-1}$.

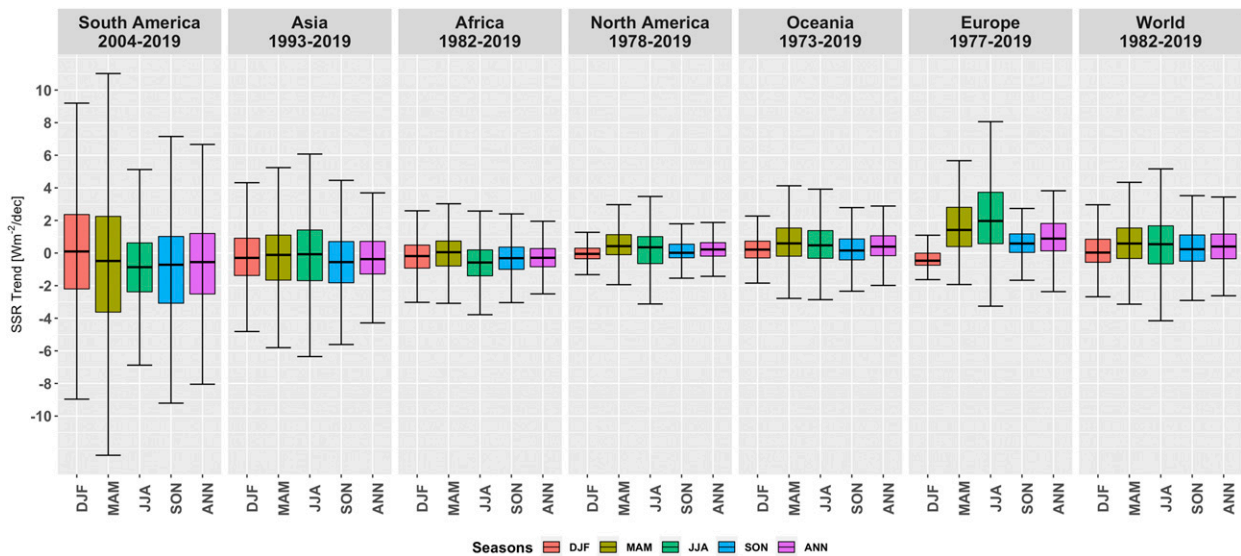


FIG. 15. Box plots for seasonal and annual SSR trends for individual continents and the global land surfaces. The box plots are statistics calculated based on trends at all individual grid boxes within each continent over the periods starting from the last respective breakpoints and up until 2019. Five values are shown for each box plot from bottom to top: lower whisker, first quartile (Q_1), median (Q_2), third quartile (Q_3), and upper whisker. The upper whisker is defined as the smaller of the maximum value or the third quartile plus 1.5 times the interquartile range ($IQR = Q_3 - Q_1$); similarly, the lower whisker is defined as the larger of the minimum value or the first quartile minus the 1.5 times the interquartile range. Continent panels are aligned in ascending order by the median of annual SSR trends from the left to the right, together with the world plot at the right. Seasonal (DJF, MAM, JJA, SON) and annual (ANN) trends are shown for each continent.

boxes on a continent. We observe that most boxes are generally symmetric, meaning the distribution is approximately normal. The box plots also indicate that South America and Asia show large variability, which can be concluded from their wider ranges between the upper and lower whiskers. Summer SSR in Europe also shows large variability. On a seasonal basis, the largest variability generally occurs in summer, and the smallest in winter. The decadal variability for individual seasons varies greatly among the continents.

The variable importance analysis shows that the long-term variability of SSR is largely explained by diurnal temperature range, cloud coverage, and other temperature indices. Moreover, maximum temperature provides the most explanation among maximum, average, and minimum monthly temperature. This is reasonable as daily maximum temperature is largely driven by the solar insolation, while the minimum temperature (at night when the sun is absent) is more driven by the greenhouse capacity of the atmosphere. Meteorological variables are the most important predictors for decadal variations of SSR. On the other hand, geographical variables (e.g., coordinates, altitude) contribute little to SSR predictions.

This study provides an observation-based dataset of SSR anomalies with complete spatial and temporal coverage. It completes the global surface solar radiation trend analysis by offering the possibility to analyze continents that have rarely been examined previously. The quantification of long-term trends is of significant importance to solar energy deployment (Müller et al. 2014). Last but not least, this study has the potential to be used for comparison with climate model simulations. It could serve as a benchmark reference for the evaluation and calibration of global or regional climate models. Other potential applications could be

climate change monitoring and a strengthened understanding of its closely related atmospheric or hydrological processes, such as temperature changes and aerosol dispersion.

Acknowledgments. This research is funded by the Norwegian Research Council, under the project of “Climate Change Modelling and Prediction of Economic Impact” (Grant 281071). The Global Energy Balance Archive (GEBA) is co-funded by the Federal Office of Meteorology and Climatology MeteoSwiss within the framework of GCOS Switzerland. Global dimming and brightening research at ETH Zurich is funded by the Swiss National Science Foundation (Grant 200020_188601).

Data availability statement. All data used in the study are from publicly accessible databases. GEBA data are available from <https://geba.ethz.ch/>. The CRU dataset can be downloaded from <https://crudata.uea.ac.uk/cru/data/hrg/>; elevation data are available from <http://research.jisao.washington.edu/data/>; rural-urban mapping data are available from <https://sedac.ciesin.columbia.edu/data/collection/grump-v1>. The interpolated SSR dataset is available from the corresponding author on request.

REFERENCES

- Adedoyin, F., I. Ozturk, I. Abubakar, T. Kumeka, O. Folarin, and F. V. Bekun, 2020: Structural breaks in CO₂ emissions: Are they caused by climate change protests or other factors? *J. Environ. Manage.*, **266**, 110628, <https://doi.org/10.1016/j.jenvman.2020.110628>.
- Arking, A., 1996: Absorption of solar energy in the atmosphere: Discrepancy between model and observations. *Science*, **273**, 779–782, <https://doi.org/10.1126/science.273.5276.779>.

- Bai, J., and P. Perron, 1998: Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**, 47–78, <https://doi.org/10.2307/2998540>.
- , and —, 2003: Computation and analysis of multiple structural change models. *J. Appl. Econ.*, **18** (1), 1–22, <https://doi.org/10.1002/jae.659>.
- Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- Budyko, M. I., 1969: The effect of solar radiation variations on the climate of the Earth. *Tellus*, **21**, 611–619, <https://doi.org/10.3402/tellusa.v21i5.10109>.
- Cherian, R., J. Quaas, M. Salzmann, and M. Wild, 2014: Pollution trends over Europe constrain global aerosol forcing as simulated by climate models. *Geophys. Res. Lett.*, **41**, 2176–2181, <https://doi.org/10.1002/2013GL058715>.
- CIESIN, 2004: Global Rural–Urban Mapping Project (GRUMP), alpha version: Urban extents grids. Center for International Earth Science Information Network (CIESIN), Columbia University, International Food Policy Research Institute, accessed 2019, <https://sedac.ciesin.columbia.edu/data/set/grump-v1-urban-extents/>.
- Collins, F. C., 1995: A comparison of spatial interpolation techniques in temperature estimation. Ph.D. thesis, Virginia Tech., 234 pp.
- Cró, S., and A. M. Martins, 2017: Structural breaks in international tourism demand: Are they caused by crises or disasters? *Tour. Manage.*, **63**, 3–9, <https://doi.org/10.1016/j.tourman.2017.05.009>.
- Erxleben, J., K. Elder, and R. Davis, 2002: Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains. *Hydrol. Processes*, **16**, 3627–3649, <https://doi.org/10.1002/hyp.1239>.
- Firth, L., M. L. Hazelton, and E. P. Campbell, 2005: Predicting the onset of Australian winter rainfall by nonlinear classification. *J. Climate*, **18**, 772–781, <https://doi.org/10.1175/JCLI-3291.1>.
- Forkel, M., N. Carvalhais, J. Verbesselt, M. Mahecha, C. Neigh, and M. Reichstein, 2013: Trend change detection in NDVI time series: Effects of inter-annual variability and methodology. *Remote Sens.*, **5**, 2113–2144, <https://doi.org/10.3390/rs5052113>.
- Gilgen, H., M. Wild, and A. Ohmura, 1998: Means and trends of shortwave irradiance at the surface estimated from Global Energy Balance Archive data. *J. Climate*, **11**, 2042–2061, [https://doi.org/10.1175/1520-0442\(1998\)011<2042:MATOSI>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<2042:MATOSI>2.0.CO;2).
- , A. Roesch, M. Wild, and A. Ohmura, 2009: Decadal changes in shortwave irradiance at the surface in the period from 1960 to 2000 estimated from Global Energy Balance Archive data. *J. Geophys. Res.*, **114**, D00D08, <https://doi.org/10.1029/2008JD011383>.
- Greenwald, R., M. H. Bergin, J. Xu, D. Cohan, G. Hoogenboom, and W. L. Chameides, 2006: The influence of aerosols on crop production: A study using the CERES crop model. *Agric. Syst.*, **89**, 390–413, <https://doi.org/10.1016/j.agry.2005.10.004>.
- Gupta, R., E. Somanathan, and S. Dey, 2017: Global warming and local air pollution have reduced wheat yields in India. *Climatic Change*, **140**, 593–604, <https://doi.org/10.1007/s10584-016-1878-8>.
- Harris, I., T. J. Osborn, P. Jones, and D. Lister, 2020: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data*, **7**, 109, <https://doi.org/10.1038/s41597-020-0453-3>.
- He, Y., K. Wang, C. Zhong, and M. Wild, 2018: A revisit of global dimming and brightening based on the sunshine duration. *Geophys. Res. Lett.*, **45**, 4281–4289, <https://doi.org/10.1029/2018GL077424>.
- Holben, B. N., 1986: Characteristics of maximum-value composite images from temporal AVHRR data. *Int. J. Remote Sens.*, **7**, 1417–1434, <https://doi.org/10.1080/01431168608948945>.
- Jiang, Y., 2008: Prediction of monthly mean daily diffuse solar radiation using artificial neural networks and comparison with other empirical models. *Energy Policy*, **36**, 3833–3837, <https://doi.org/10.1016/j.enpol.2008.06.030>.
- Leirvik, T., and M. Yuan, 2021: A machine learning technique for spatial interpolation of solar radiation observations. *Earth Space Sci.*, **8**, e2020EA001527, <https://doi.org/10.1029/2020EA001527>.
- Liepert, B. G., 2002: Observed reductions of surface solar radiation at sites in the United States and worldwide from 1961 to 1990. *Geophys. Res. Lett.*, **29**, 1421, <https://doi.org/10.1029/2002GL014910>.
- , and G. J. Kukla, 1997: Decline in global solar radiation with increased horizontal visibility in Germany between 1964 and 1990. *J. Climate*, **10**, 2391–2401, [https://doi.org/10.1175/1520-0442\(1997\)010<2391:DIGSRW>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<2391:DIGSRW>2.0.CO;2).
- , J. Feichter, U. Lohmann, and E. Roeckner, 2004: Can aerosols spin down the water cycle in a warmer and moister world? *Geophys. Res. Lett.*, **31**, L06207, <https://doi.org/10.1029/2003GL019060>.
- Moseid, K. O., and Coauthors, 2020: Bias in CMIP6 models as compared to observed regional dimming and brightening. *Atmos. Chem. Phys.*, **20**, 16 023–16 040, <https://doi.org/10.5194/acp-20-16023-2020>.
- Müller, B., M. Wild, A. Driesse, and K. Behrens, 2014: Rethinking solar resource assessments in the context of global dimming and brightening. *Sol. Energy*, **99**, 272–282, <https://doi.org/10.1016/j.solener.2013.11.013>.
- Myoung, B., J. Rhee, and C. Yoo, 2020: Long-lead predictions of warm season droughts in South Korea using North Atlantic SST. *J. Climate*, **33**, 4659–4677, <https://doi.org/10.1175/JCLI-D-19-0082.1>.
- Nabat, P., S. Somot, M. Mallet, A. Sanchez-Lorenzo, and M. Wild, 2014: Contribution of anthropogenic sulfate aerosols to the changing Euro-Mediterranean climate since 1980. *Geophys. Res. Lett.*, **41**, 5605–5611, <https://doi.org/10.1002/2014GL060798>.
- NGDC, and Coauthors, 1995: TerrainBase, Global 5 Arc-minute Ocean Depth and Land Elevation from the U.S. National Geophysical Data Center (NGDC). Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, accessed 2019, <https://doi.org/10.5065/E08M-4482>.
- Obryk, M. K., A. G. Fountain, P. T. Doran, W. B. Lyons, and R. Eastman, 2018: Drivers of solar radiation variability in the McMurdo Dry Valleys, Antarctica. *Sci. Rep.*, **8**, 5002, <https://doi.org/10.1038/s41598-018-23390-7>.
- Ohmura, A., and H. Gilgen, 1993: Re-evaluation of the global energy balance. *Interactions between Global Climate Subsystems: The Legacy of Hann, Meteor. Monogr.*, Amer. Geophys. Union, Vol. 75, 93–110, <https://doi.org/10.1029/GM075p0093>.
- Parding, K. M., B. G. Liepert, L. M. Hinkelman, T. P. Ackerman, K. F. Dagestad, and J. A. Olseth, 2016: Influence of synoptic weather patterns on solar irradiance variability in northern Europe. *J. Climate*, **29**, 4229–4250, <https://doi.org/10.1175/JCLI-D-15-0476.1>.
- Pfeifroth, U., A. Sanchez-Lorenzo, V. Manara, J. Trentmann, and R. Hollmann, 2018: Trends and variability of surface solar radiation in Europe based on surface- and satellite-based data records. *J. Geophys. Res. Atmos.*, **123**, 1735–1754, <https://doi.org/10.1002/2017JD027418>.

- Pinker, R. T., B. Zhang, and E. G. Dutton, 2005: Do satellites detect trends in surface solar radiation? *Science*, **308**, 850–854, <https://doi.org/10.1126/science.1103159>.
- Proctor, J., S. Hsiang, J. Burney, M. Burke, and W. Schlenker, 2018: Estimating global agricultural effects of geo-engineering using volcanic eruptions. *Nature*, **560**, 480–483, <https://doi.org/10.1038/s41586-018-0417-3>.
- Qin, W., L. Wang, M. Zhang, Z. Niu, M. Luo, A. Lin, and B. Hu, 2019: First effort at constructing a high-density photosynthetically active radiation dataset during 1961–2014 in China. *J. Climate*, **32**, 2761–2780, <https://doi.org/10.1175/JCLI-D-18-0590.1>.
- Roderick, M. L., and G. D. Farquhar, 2012: Geoengineering: Hazy, cool and well fed? *Nat. Climate Change*, **2**, 76–77, <https://doi.org/10.1038/nclimate1395>.
- Romanou, A., B. Liepert, G. A. Schmidt, W. B. Rossow, R. A. Ruedy, and Y. Zhang, 2007: 20th century changes in surface solar irradiance in simulations and observations. *Geophys. Res. Lett.*, **34**, 5713, <https://doi.org/10.1029/2006GL028356>.
- Sanchez-Lorenzo, A., J. Calbó, and M. Wild, 2013: Global and diffuse solar radiation in Spain: Building a homogeneous dataset and assessing their trends. *Global Planet. Change*, **100**, 343–352, <https://doi.org/10.1016/j.gloplacha.2012.11.010>.
- , M. Wild, M. Brunetti, J. A. Guijarro, M. Z. Hakuba, J. Calbó, S. Mystakidis, and B. Bartok, 2015: Reassessment and update of long-term trends in downward surface shortwave radiation over Europe (1939–2012). *J. Geophys. Res. Atmos.*, **120**, 9555–9569, <https://doi.org/10.1002/2015JD023321>.
- , A. Enriquez-Alonso, M. Wild, J. Trentmann, S. M. Vicente-Serrano, A. Sanchez-Romero, R. Posselt, and M. Z. Hakuba, 2017: Trends in downward surface solar radiation from satellites and ground observations over Europe during 1983–2010. *Remote Sens. Environ.*, **189**, 108–117, <https://doi.org/10.1016/j.rse.2016.11.018>.
- Scudiero, E., D. L. Corwin, F. Morari, R. G. Anderson, and T. H. Skaggs, 2016: Spatial interpolation quality assessment for soil sensor transect datasets. *Comput. Electron. Agric.*, **123**, 74–79, <https://doi.org/10.1016/j.compag.2016.02.016>.
- Stanhill, G., 2005: Global dimming: A new aspect of climate change. *Weather*, **60**, 11–14, <https://doi.org/10.1256/wea.210.03>.
- , and S. Cohen, 2001: Global dimming: A review of the evidence for a widespread and significant reduction in global radiation with discussion of its probable causes and possible agricultural consequences. *Agric. For. Meteorol.*, **107**, 255–278, [https://doi.org/10.1016/S0168-1923\(00\)00241-0](https://doi.org/10.1016/S0168-1923(00)00241-0).
- Sun, H., D. Gui, B. Yan, Y. Liu, W. Liao, Y. Zhu, C. Lu, and N. Zhao, 2016: Assessing the potential of random forest method for estimating solar radiation using air pollution index. *Energy Convers. Manage.*, **119**, 121–129, <https://doi.org/10.1016/j.enconman.2016.04.051>.
- Tanaka, K., A. Ohmura, D. Folini, M. Wild, and N. Ohkawara, 2016: Is global dimming and brightening in Japan limited to urban areas? *Atmos. Chem. Phys.*, **16**, 13 969–14 001, <https://doi.org/10.5194/acp-16-13969-2016>.
- Vu, T. T., J. Kiesel, B. Guse, and N. Fohrer, 2019: Analysis of the occurrence, robustness and characteristics of abrupt changes in streamflow time series under future climate change. *Climate Risk Manage.*, **26**, 100198, <https://doi.org/10.1016/j.crm.2019.100198>.
- Wang, K., Q. Ma, X. Wang, and M. Wild, 2014: Urban impacts on mean and trend of surface incident solar radiation. *Geophys. Res. Lett.*, **41**, 4664–4668, <https://doi.org/10.1002/2014GL060201>.
- Wild, M., 2009: Global dimming and brightening: A review. *J. Geophys. Res.*, **114**, D00D16, <https://doi.org/10.1029/2008JD011470>.
- , 2012: Enlightening global dimming and brightening. *Bull. Amer. Meteor. Soc.*, **93**, 27–37, <https://doi.org/10.1175/BAMS-D-11-00074.1>.
- , 2016: Decadal changes in radiative fluxes at land and ocean surfaces and their relevance for global warming. *Wiley Interdiscip. Rev.: Climate Change*, **7**, 91–107, <https://doi.org/10.1002/wcc.372>.
- , and B. Liepert, 2010: The Earth radiation balance as driver of the global hydrological cycle. *Environ. Res. Lett.*, **5**, 025203, <https://doi.org/10.1088/1748-9326/5/2/025203>.
- , and Coauthors, 2005: From dimming to brightening: Decadal changes in solar radiation at Earth's surface. *Science*, **308**, 847–850, <https://doi.org/10.1126/science.1103215>.
- , J. Grieser, and C. Schär, 2008: Combined surface solar brightening and increasing greenhouse effect support recent intensification of the global land-based hydrological cycle. *Geophys. Res. Lett.*, **35**, L17706, <https://doi.org/10.1029/2008GL034842>.
- , A. Ohmura, C. Schär, G. Müller, D. Folini, M. Schwarz, M. Z. Hakuba, and A. Sanchez-Lorenzo, 2017: The Global Energy Balance Archive (GEBA) version 2017: A database for worldwide measured surface energy fluxes. *Earth Syst. Sci. Data*, **9**, 601–613, <https://doi.org/10.5194/essd-9-601-2017>.
- Xu, T., and Coauthors, 2018: Evaluating different machine learning methods for upscaling evapotranspiration from flux towers to the regional scale. *J. Geophys. Res. Atmos.*, **123**, 8674–8690, <https://doi.org/10.1029/2018JD028447>.
- Yang, S., X. L. Wang, and M. Wild, 2018: Homogenization and trend analysis of the 1958–2016 in situ surface solar radiation records in China. *J. Climate*, **31**, 4529–4541, <https://doi.org/10.1175/JCLI-D-17-0891.1>.
- Zhou, Q., A. Flores, N. F. Glenn, R. Walters, and B. Han, 2017: A machine learning approach to estimation of downward solar radiation from satellite-derived data products: An application over a semi-arid ecosystem in the U.S. *PLOS ONE*, **12**, e0180239, <https://doi.org/10.1371/journal.pone.0180239>.

