Large Differences in Diffuse Solar Radiation ~~in~~ Among Current-Generation Reanalysis and Satellite-Derived Products

T. Chakraborty[a] & X. Lee[a]

[a] *School of the Environment, Yale University, New Haven, CT*

Corresponding author: TC Chakraborty, tc.chakraborty@yale.edu

ABSTRACT

Though the partitioning of shortwave radiation ($K_{\downarrow}$) at the surface into its diffuse ($K_{\downarrow,d}$) and direct beam ($K_{\downarrow,b}$) components is relevant for, among other things, the terrestrial energy and carbon budgets, there is a dearth of large-scale comparisons of this partitioning across reanalysis and satellite-derived products. Here we evaluate $K_{\downarrow}$, $K_{\downarrow,d}$, and $K_{\downarrow,b}$, as well as the diffuse fraction (k_d) of solar radiation in four current-generation reanalysis (NOAA-CIRES-DOE, NCEP/NCAR, MERRA-2, ERA5) datasets and one satellite-derived product (CERES) using ≈1400 site years of observations. Although the systematic positive biases in $K_{\downarrow}$ is consistent with previous studies, the biases in gridded $K_{\downarrow,d}$ and $K_{\downarrow,b}$ vary in direction and magnitude, both annually and across seasons. The inter-model variability in cloud cover strongly explains the biases in both $K_{\downarrow,d}$ and $K_{\downarrow,b}$. Over Europe and China, the long-term (10-year plus) trends in $K_{\downarrow,d}$ in the gridded products are noticeably differ from corresponding observations and the grid-averaged 35-year trends show an order of magnitude variability. In the MERRA-2 reanalysis, which includes both clouds and assimilated aerosols, the reduction in both clouds and aerosols reinforce each other to establish brightening trends over Europe, while the effect of increasing aerosols overwhelm the effect of decreasing cloud cover over China. The inter-model variability in k_d seen here (0.27 to 0.50 from CERES to MERRA-2) suggests substantial differences in shortwave parameterization schemes and their inputs in climate models and can contribute to inter-model variability in coupled simulations.

Based on these results, we call for systematic evaluations of $K_{\downarrow,d}$ and $K_{\downarrow,b}$ in CMIP6 models.

Significance Statement

The direction of sunlight can be changed due to particles and clouds in the air. This is known as diffuse light and it affects solar energy generation and plant growth. Here, we wanted to address a gap in previous studies and compare the diffuse light in global datasets. We find large differences between datasets, explained mostly by differing cloud amounts. When compared to measurements from the ground, we find that these differences exist for most sites and across seasons. The change in diffuse light over the last 35 years also varies widely between datasets. Our results call for larger-scale comparisons of diffuse light in all current-generation global models. Doing so can help us better predict future climate change.

1. Introduction

Solar radiation is a key driver of the Earth's climate system. During its transmission through the atmosphere, it is scattered and absorbed by aerosols, clouds, and gases. Solar radiation incident on the surface ($K_{\downarrow}$) consists of beam radiation ($K_{\downarrow,b}$) and diffuse radiation ($K_{\downarrow,d}$). The former follows the original path of the sunlight and the latter is the scattered component that deviates from that path. These components are not routinely measured at weather stations (Stephens et al. 2012). Instead, climatic and ecological studies and solar energy applications generally rely on gridded estimates from atmospheric models, including reanalysis products and global

climate models (GCMs). Owing to computational limitations, these models are run at relatively coarse resolutions (50 km to over 200 km). This leads to simplified versions of radiative transfer codes being implemented in these models, as well as differences in input parameters like clouds and aerosols to those codes, both of which influence the estimated radiation fields (Oreopoulos et al. 2012). Systematic biases exist in these model estimates. It is well-known that $K_{\downarrow}$ is overestimated by most atmospheric models due in large part to the underestimation of cloud cover (Markovic et al. 2009; Bosilovich et al. 2011; Kennedy et al. 2011; Zhao et al. 2013; Wild et al. 2015; Zhang et al. 2016). This overestimation will lead to surface warming (Chakraborty and Lee 2019) and also increase the energy returned to the atmosphere through heat and moisture fluxes, which may artificially strengthen the hydrological cycle (Wild et al. 1998).

$K_{\downarrow,d}$ remains a relatively understudied component of the Earth's radiation budget. Several studies have demonstrated enhanced carbon uptake and evaporative fraction at various scales with increasing $K_{\downarrow,d}$ (Knohl and Baldocchi 2008; Mercado et al. 2009; Yue and Unger 2017; Rap et al. 2018; Chakraborty et al. ~~Under Review~~2021). Thus, a better constraint on $K_{\downarrow,d}$ can improve our ability to predict the surface energy, water and carbon budgets. Accurate estimates of the direct/diffuse partitioning of $K_{\downarrow}$ are also important for solar energy applications, particularly concentrating solar power (Lee et al. 2016).

This study is concerned with biases in surface $K_{\downarrow,d}$ in retrospective analysis or reanalysis datasets, which assimilate observations of some variables to constrain

other modeled variables (Kalnay et al. 1996). These observationally constrained datasets represent our best estimates of the current and historical global climate system. Although site-level comparisons of radiation transfer codes that also consider $K_{\downarrow,d}$ have been performed in the past (Oreopoulos et al. 2012), how accurately reanalysis models simulate $K_{\downarrow,d}$ remains largely unknown at the global scale. To our best knowledge, only a couple of regional scale evaluations of $K_{\downarrow,d}$ from reanalysis data are available, both for the ERA5 reanalysis (Jiang et al. 2019b, 2020). The second activity of an ongoing inter-model comparative project called Radiative Forcing Intercomparison Project (RFMIP) requests modeling centers to provide broadband fluxes based on their radiative transfer codes, but does not explicitly require the partitioning of $K_{\downarrow}$ into $K_{\downarrow,d}$ and $K_{\downarrow,b}$ (Pincus et al. 2016).

The lessons learned about model biases in $K_{\downarrow}$ are not necessarily applicable to $K_{\downarrow,d}$, $K_{\downarrow,b}$ incident on the surface is controlled by the total extinction of a light beam as it transmits through the atmosphere, while $K_{\downarrow,d}$ is a function of the scattered sunlight (Liu and Jordan 1960). Thus, one can hypothetically fix the overestimation of $K_{\downarrow}$ in modeled products by increasing aerosols or clouds or through statistical bias-correction algorithms (Zhao et al. 2013), but with unknown individual biases in $K_{\downarrow,d}$ and $K_{\downarrow,b}$. Since aerosols and clouds are parameterized differently in different gridded products, including, but not limited to, prescriptions of cloud droplet size distribution, cloud overlap, and aerosol properties, our hypothesis is that the bias in $K_{\downarrow,d}$ and $K_{\downarrow,b}$ are less systematic in direction than that seen for $K_{\downarrow}$ in previous studies

Formatted: Not Superscript/ Subscript

and strongly controlled by the cloud and aerosol inputs (Wild et al. 2015). To test these hypotheses, the specific objectives of this study are:

(1) To perform a systematic evaluation of the monthly $K_{\downarrow,b}$, $K_{\downarrow}$ and $K_{\downarrow,d}$ in gridded data products, including five current-generation reanalysis datasets (NOAA-CIRES-OE, NCEP/NCAR, JRA-55, MERRA-2 and ERA5; only $K_{\downarrow}$ for JRA-55) and one satellite-derived product (CERES)

(2) To examine the differences in these variables between the gridded data products benchmarked against observations at the annual and seasonal time scales

(3) To evaluate the ability of the gridded products to capture long-term changes in these variables for Europe and China, two regions which have relatively high densities of ground-based observations and have several previous relevant studies to compare the results

(4) To discuss potential sources of biases and inter-model variability, particularly due to cloud cover and also atmospheric aerosols, in these datasets

2. Methods

a. Global Reanalysis Products

We used monthly gridded data from five reanalysis products: (1) NOAA-CIRES-DOE -- the Twentieth Century Reanalysis version 3 from National Oceanic and Atmospheric Administration (NOAA), Cooperative Institute for Research in Environmental Science (CIRES), and Department of Energy (DOE), (2) NCEP/NCAR -- the 50-year Reanalysis from National Centers for Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR), (3)

Formatted: Not Superscript/ Subscript

JRA-55 -- the Japanese 55-year Reanalysis, (4) MERRA-2 -- the Modern-Era Retrospective analysis for Research and Applications, version 2, and (5) ERA5 -- the Fifth Generation Reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF). They represent the latest iteration of the major global reanalyses for research and applications. Note that although the Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) is newer than the NCEP/NCAR reanalysis, it does not publicly archive $K_{\downarrow,\mathrm{d.}}$ Table 1 summarizes the important information about the products considered in the study. Short descriptions of the datasets are given below.

1. NOAA-CIRES-DOE

This reanalysis assimilates surface pressure observations to provide estimates of the historical climate state (Slivinski et al. 2019). In addition to improvements in the assimilation system, the latest version of the reanalysis includes a higher resolution forecast model, more assimilated pressure observations, and better representation of storm intensity. The radiative transfer model for shortwave in this reanalysis interacts with fractional cloud cover, modeled O_3, time varying CO_2, volcanic aerosols, and solar variations (Hou et al. 2002).

2. NCEP/NCAR

This reanalysis assimilates data from a wide variety of weather observation including pressure measurements over land, pressure, temperature, and specific humidity measurements over oceans, radiosonde profiles, temperature and wind data observed from aircraft, and satellite-derived cloud-tracked wind data (Kistler et al.

2001). The shortwave parameterization in this reanalysis is based on the work by Lacis and Hansen (1974).

3. JRA-55

The JRA-55 reanalysis version improves upon the previous JRA-25 product with an updated assimilation system, more ingested observations, a newer longwave radiation scheme, and higher resolution forecasts (Kobayashi et al. 2015). The shortwave radiation is parameterized considering random overlap of clouds, H_2O absorption based on Briegleb et al. (1992), O_2, O_3, and CO_2 absorption based on Freidenreich and Ramaswamy (1999), and assuming standard atmospheric aerosol profiles with optical depths adjusted using monthly aerosol climatology. It only archives gridded data for $K_{\downarrow}$.

4. MERRA-2

The MERRA-2 reanalysis is a recent global reanalysis product that assimilates bias-corrected satellite observations of aerosols and clear-sky irradiances (Randles et al. 2017). It also uses observed precipitation to force the land-surface model. The shortwave radiation scheme is based on Chou and Suarez (1999) and the latest version of the Goddard Earth Observing System (GEOS-5) assimilates newer satellite observations. The total aerosol optical depth (AOD) in MERRA-2 has been evaluated against independent observations (Buchard et al. 2017).

5. ERA5

The ERA5 reanalysis uses the recently developed Integrated Forecasting System to improve upon its predecessor (Hersbach et al. 2020). In addition to the finer

horizontal model resolution, ERA5 has consistent hourly outputs, improvements in the dynamical core, and a four-dimensional variational data assimilation system (like JRA-55). Standardized sets of long-term forcing for aerosols, greenhouse gases, and O_3 are taken from the World Climate Research Programme (WCRP) and the McRad scheme is used to parameterize radiation (Morcrette et al. 2008).

b. Satellite-Derived Estimates

In addition to the reanalysis products, we used the satellite-derived monthly gridded $K_{\downarrow}$ and $K_{\downarrow,d}$ data in the latest version of the Clouds and the Earth's Radiant Energy System (CERES) synoptic product, (CERES– ~~EBAF~~SYN1deg– Ed4.1; ~~Kato~~ Rutan et al. ~~2018~~2015). The dataset is well-constrained by observations due to direct measurements of the top of the atmosphere components and the use of aerosol and cloud observations from satellites, including those carrying the Moderate Resolution Imaging Spectroradiometer (MODIS), in the radiative transfer code.

c. Ground-Based Point Observations

The Global Energy Balance Archive (GEBA) is a repository of energy flux measurements at the Earth's surface (Gilgen and Ohmura 1999) and is the most comprehensive global database of observed mean monthly surface radiation components currently available. Here, we used the latest iteration of the database (Wild et al. 2017) after removing sites with missing data and applying the following quality control steps:

1. We considered only the observations not flagged as erroneous by the database's quality control procedure (thus, data with flags 5, 6, 7, and 8 were used)

2. Observations for which the monthly means were 0 W m^{-2} were not considered since they are primarily due to either instrument errors or during polar nights.
3. We removed observations if the diffuse fraction, $k_d = K_{\downarrow,d}/K_{\downarrow}$, exceeded one or was 0.
4. Although there are many more stations with $K_{\downarrow}$ measurements than $K_{\downarrow,d}$ measurements, to keep the number of sites consistent, we only considered those with simultaneous observations of $K_{\downarrow}$ and $K_{\downarrow,d}$ in most cases (except for examining long-term trends; see next subsection).

After data screening, we obtained 221 stations (distribution of stations shown in Fig. S1) with a total of 16589 site-months of data between 1980 and 2015. Only a few GEBA sites have direct measurements of $K_{\downarrow,b}$. For evaluating modeled $K_{\downarrow,b,}$ the observed $K_{\downarrow,b}$ was computed as the difference between $K_{\downarrow}$ and $K_{\downarrow,d}$.

d. Data Processing and Metrices for Evaluation

We extracted monthly $K_{\downarrow,d}$, $K_{\downarrow,b}$, and $K_{\downarrow}$ from the gridded datasets from the start of 1980 to the end of 2019. Only $K_{\downarrow}$ was extracted for JRA-55 since it does not publicly archive $K_{\downarrow,d}$ or $K_{\downarrow,b}$. For the overall evaluation against GEBA, all the grids overlapping the observational sites and months between 2001 to 2015 were used. This period is common to all the six datasets and is referred to as Common Period I. For cases where multiple sites were within one grid box, the same grid value was compared against each of those observations. For NCEP/NCAR, the lowest resolution dataset, roughly 12% of the sites share a common grid with another site, while only 2% of sites share a common grid when using the highest resolution

dataset (ERA5). Four metrics were used to evaluate the modeled data, including coefficient of determination (r^2), root mean square error (RMSE), mean bias error (MBE), and mean percentage error (MPE).

To examine inter-model variability at annual and seasonal scales, we chose Common Period I and used the CERES data as reference. To avoid mixing the seasonality of the two hemispheres, we only used sites and grids in the northern hemisphere when examining seasonality. Although point-based observations of surface radiation fields are not always comparable to gridded estimates due to how models represent clouds, at the monthly scale, these uncertainties are reduced (see Limitations subsection).

We restricted our trend analysis to Europe and China between 1980 and 2015. This period, termed here as Common Period II, is longer than the CERES data period but covered by all the five reanalysis products (Table 1). These two regions have more sites with continuous data coverage than other regions of the world. We calculated temporal trends for the sites with at least a total of 10 years of data (not necessarily contiguous years). Site-averaged time trends of the gridded model data were based on the same measurement years and the grids containing these GEBA sites. The threshold of 10 years, although somewhat arbitrary, was used, not to estimate the true 35-year trend but to examine whether the gridded products showed similar trends during the corresponding time-frame. Similar thresholds have been used in other studies that have examined long-term trends in $K_{\downarrow}$ (Yang et al. 2019; Schwarz et al. 2020). Since few of the sites have both observations of $K_{\downarrow,d}$ and $K_{\downarrow}$

236 that satisfy all the above criteria, we used a different subset of measurements for $K_{\downarrow}$
237 and $K_{\downarrow,d}$ over these regions. This left us with 28 (7) stations over Europe (China) for
238 examining $K_{\downarrow}$ trends, and 15 (5) sites over Europe (China) for $K_{\downarrow,d}$ trends. Before
239 using these stations, however, we also tested for change/breakpoints in the time
240 series data using the Standard Normal Homogeneity Test (Alexandersson 1986).
241 Considering only those stations that show no breakpoints at the 95% significance
242 level, we got 24 (4) stations over Europe (China) for $K_{\downarrow}$ and 8 (4) stations over
243 Europe (China) for $K_{\downarrow,d}$. We also calculated the grid-averaged modeled trend for the
244 entire period (1980-2015) using all the grids that fall within Europe and China.
245 Before finding the grids intersecting these two regions, the five reanalysis products
246 were re-gridded to 1◦ × 1◦ grids, the grid size of CERES, using bilinear
247 interpolation, which is appropriate due to the spatial continuity in these variables at
248 the annual time scale. This re-gridded data were also used to demonstrate grid-by-
249 grid difference in multi-year average values between the products (see Section 3b).
250 In all cases, the trends were based on the annual average value regressed against the
251 year of observation, with the statistical significance of the trends calculated. Finally,
252 we also estimated cloud cover, top of the atmosphere $K_{\downarrow}$ and AOD (from MERRA-
253 2) for Europe, China, and globally to examine the reasons for some of these biases.

254 **3. Results**

255 ***a. Overall Evaluation and Annual Inter-Comparisons***

256 The global mean $K_{\downarrow}$ varies from 185.4 W m^{-2} (CERES) to 205.3 W m^{-2}
257 (NCEP/NCAR) for Common Period I (2001-2015), based on all model grids (Table

1). In comparison, Wild et al. (2015) found a multi-model mean $K_{\downarrow}$ of 189.1 W m^{-2} based on 43 CMIP5 models for 2000-2004. For the grid-years that coincide with the GEBA observations, the modeled mean $K_{\downarrow}$ varies from 165.2 W m^{-2} (CERES) to 208.1 W m^{-2} (NCEP/NCAR), and the observed mean $K_{\downarrow}$ is 162.5 W m^{-2}. All the reanalysis datasets capture the seasonal (Fig. 3) and geographic distributions of the GEBA-observed $K_{\downarrow}$ relatively well, with the overall r^2 varying from 0.9 for NOAA-CIRES-DOE and NCEP/NCAR to 0.96 for CERES (Table 2). As expected, CERES performs better than all the reanalysis products, both in terms of variability (r^2 = 0.97) and bias (MBE=2.6 W m^{-2}).

The global mean $K_{\downarrow,b}$ varies from 82.8 W m^{-2} (CERES) to 132.8 W m^{-2} (MERRA-2) during Common Period I, based on all model grids. The sign of the error in $K_{\downarrow,b}$ is less consistent across the different products than the error in $K_{\downarrow}$ (Fig. 1). While NCEP/NCAR, MERRA-2, and ERA5 overestimate $K_{\downarrow,b}$ (MBE = 36.8, 39.9, and 17.4 W m^{-2}, respectively), NOAA-CIRES-DOE and CERES underestimate it (MPE = -4.3 and -16.8 W m^{-2}, respectively; Table 2). Among the reanalyses, ERA5 performs the best at capturing the global variability in $K_{\downarrow,b}$ (r^2 = 0.9), and NCEP/NCAR perform the worst ($r^2 \approx 0.73$).

The global mean $K_{\downarrow,d}$ varies from 52.8 W m^{-2} (MERRA-2) to 102.6 W m^{-2} (CERES), and diffuse fraction k_d varies from 0.28 (MERRA-2) to 0.55 (CERES) during Common Period I based on all model grids (Table 1). For the grid-years that coincide with the GEBA observations, k_d ranges from 0.28 (MERRA-2) to 0.55 (CERES), and the observed mean k_d for the quality-controlled GEBA dataset is 0.46.

NOAA-CIRES-DOE, NCEP/NCAR and CERES have positive biases in $K_{\downarrow,d}$ (MBE = 13.8, 8.77, and 19.5 W m^{-2}, respectively; Table 2), and MERRA-2 and ERA5 have negative biases (MBE = -21.4 and -9.9 W m^{-2}, respectively; Table 2). Bias errors in k_d depends on errors in $K_{\downarrow,d}$ and $K_{\downarrow}$. For MERRA-2 and ERA5, $K_{\downarrow}$ is positively biased, and $K_{\downarrow,d}$ is negatively biased (Table 2). Consequently, these two reanalyses underestimate k_d, with MERRA-2 performing the worst among the datasets, with an MBE of -0.18 for all sites. NCEP/NCAR underestimates k_d (MBE = -0.07) because it overestimates $K_{\downarrow}$ more (relatively speaking) than it overestimates $K_{\downarrow,d}$ (Table 2). Even though NCEP/NCAR and NOAA-CIRES-DOE show smaller MBE than MERRA-2, they do not capture the observed variability in k_d well (r^2 = 0.36 to 0.41). ERA5 captures the variability in k_d the best (r^2 = 0.72), even better than CERES (r^2 = 0.67). CERES overestimates k_d (MBE = 0.09) as it underestimates $K_{\downarrow,b}$ and overestimates $K_{\downarrow,d}$.

Figure 1 shows the scatter plots between gridded and observed k_d, $K_{\downarrow,d}$, and $K_{\downarrow,b}$ for all common GEBA site-months, with each data point representing a monthly mean and the color representing the density of data. Fig. S2 shows the scatter plots for total $K_{\downarrow}$. The scatter is a result of both natural (seasonal and geographic) variations and measurement and model errors. As discussed earlier, the gridded data show larger variability than observations, and the biases in gridded $K_{\downarrow,d}$ and $K_{\downarrow,b}$ across products is less systematic in sign than the consistent overestimation seen for $K_{\downarrow}$ (Fig. S1 and Table 2). This lack of consistency is evident in the scatter plot. For instance, although the line of best fit for the gridded $K_{\downarrow,b}$ data is ERA5 is almost

identical to the 1:1 line, the slope is only 0.7 for $K_{\downarrow,d}$ since the intercept. For CERES, the line of best fit is less than the 1:1 line for $K_{\downarrow,b}$, but is greater than 1:1 for $K_{\downarrow,d}$, demonstrating the underestimation of $K_{\downarrow,b}$ and overestimation of $K_{\downarrow,d}$ by this dataset with an intercept close to zero. In general, more scatter is seen for NOAA-CIRES-DOE and NCEP/NCAR data and the least for ERA5 and CERES. For NOAA-CIRES-DOE, the large scatter for k_d suggests that the dataset cannot adequately capture the spatiotemporal distribution of this variable.

b. Site-level Evaluation and Spatial Patterns

Figure 2a and 2b map the MBE in $K_{\downarrow}$ at individual GEBA sites for NCEP/NCAR and CERES, respectively, for Common Period I. Similarly, Figs 2c and 2d display the site-level MBE in NCEP/NCAR and CERES for $K_{\downarrow,d}$. Bias maps for the other data products can be found in Figure S3.

The site-level MBE patterns of $K_{\downarrow}$ and $K_{\downarrow,d}$ are consistent with the overall evaluation in the previous subsection. The reanalysis products show a positive $K_{\downarrow}$ bias for the majority of the GEBA sites (80.5% for NOAA-CIRES-DOE, 98.9% for NCEP/NCAR, 90% for JRA-55, 94.7% for MERRA-2, and 78.9% for ERA5). The NCEP/NCAR reanalysis has the highest $K_{\downarrow}$ MBE among the datasets considered, and ERA5 and CERES have low biases. MERRA-2 underestimates $K_{\downarrow,d}$ for almost all the sites (99.7%; Fig. S1f) and ERA5 underestimates $K_{\downarrow,d}$ for 91.9% of the sites (Fig. S3g). NOAA-CIRES-DOE, NCEP/NCAR, and CERES overestimate $K_{\downarrow,d}$ for 93.6%, 76.5%, and 98% of the sites, respectively (Figs 3c, S3e, and 3d).

Formatted: Not Superscript/ Subscript

Formatted: Not Superscript/ Subscript

Since CERES captures both the magnitude and variability of $K_{\downarrow}$ more accurately than the reanalyses (Table 2), here we use CERES as the reference to examine anomaly hot spots of the reanalysis products for Common Period I (Figs 3c and S4). All the reanalysis products show qualitatively similar positive biases from CERES over southern China and along the western coast of South America. NOAA-CIRES-DOE shows some of the largest localized anomalies; positive biases as much as 100 W m^{-2} are evident over eastern China. Overall, the differences are lower over Europe (0.0 ± 1.2 W m^{-2} for ERA5 to 46.6 ± 2.0 W m^{-2} for NCEP/NCAR; mean ± standard deviation) than over China (12.3 ± 1.4 W m^{-2} for NOAA-CIRES-DOE to 65.6 ± 2.1 W m^{-2} for NCEP/NCAR) for all the reanalysis products, a pattern consistent with site-level evaluations using GEBA observations (Figs 2 and S3). The closer value between CERES and the reanalyses over Europe could be due to stronger constraints on the energy budget due to more quality-assured assimilated meteorological observations over this region. For reference, the number of common GEBA stations over Europe for Common Period I is 93, while there are only 10 over China; with similar sampling biases expected for assimilated variables. Another potential factor is the influence of higher aerosol loading over China, which is not explicitly represented in most of these datasets.

c. Annual Cycle

Figure 3 compares the northern hemisphere seasonal patterns in $K_{\downarrow,d}$, $K_{\downarrow}$, and k_d among the datasets and the GEBA observations in the northern hemisphere, using the site-months common to the datasets and the GEBA observations for Common

Formatted: Not Superscript/ Subscript

345 Period I. The complete northern hemispheric mean seasonal patterns are given in
346 sub-figures 3b, 3d, and 3f using all the model grids. The GEBA observations are
347 skewed towards middle to high latitudes; thus Figs 3a and 3b shows a stronger $K_{\downarrow}$
348 and $K_{\downarrow,d}$ seasonality than Figs 3b and 3d. For instance, the inter-seasonal range of
349 $K_{\downarrow}$, or the difference between the monthly maximum and monthly minimum $K_{\downarrow}$ for
350 the average year, is 134.4 W m^{-2} in Fig. 3b and 178.8 W m^{-2} in Fig. 3a for CERES.
351 The data products generally capture the observed seasonality, showing much higher
352 $K_{\downarrow}$ and $K_{\downarrow,d}$ values in the summer than in the winter (Figs 3a and 3c). Among the
353 datasets, the inter-seasonal range in site-corresponding $K_{\downarrow}$ varies from 178.8 W m^{-2}
354 in CERES and ERA5 (181 W m^{-2} for GEBA) to 207.7 W m^{-2} in NCEP/NCAR
355 according to Fig. 3a.

356 Generally, there is a larger inconsistency in the $K_{\downarrow,d}$ seasonal variations than the
357 $K_{\downarrow}$ seasonal variations among the datasets. Particularly, the CERES data shows a
358 more pronounced $K_{\downarrow,d}$ seasonality (inter-seasonal range = 101 W m^{-2}) than the
359 GEBA observations (inter-seasonal range = 70 W m^{-2}) and the other data products
360 (average inter-seasonal range of 61.6 W m^{-2} for the other products; Fig. 3c).

361 Globally, the observed k_d is higher in winter and lower in the summer (Fig. 3e).
362 The inter-seasonal range in k_d varies substantially between the products, with
363 CERES showing the lowest range of 0.03 and NOAA-CIRES-DOE showing the
364 highest range of 0.22 (Fig. 3e). In comparison, the inter-seasonal range in the
365 corresponding GEBA observations is 0.13. The muted seasonality in CERES is
366 evidently driven by the stronger seasonality for $K_{\downarrow,d}$ in this dataset. Combining all

Formatted: Not Superscript/ Subscript

Formatted: Not Superscript/ Subscript

367 the gridded products, for the northern hemisphere, the spatially averaged inter-
368 seasonal range in k_d is only 0.05 (Fig. 3f), compared to 0.13 for the grids
369 corresponding to the GEBA observations; evidently due to the higher frequency of
370 GEBA observations in the higher latitudes.

371 ***d. Long-term Trends over Europe and China***

372 We analyze the long-term trends in $K_{\downarrow}$ and $K_{\downarrow,d}$ in Europe and China in two ways.
373 First, we calculate the trends using the reanalysis products for Common Period II
374 (1980-2015) and all grid cells in these two regions. The results are presented as solid
375 bars in Fig. 4 with the statistical significance of the trends noted. Over Europe,
376 NOAA-CIRES-DOE shows a slightly negative trend and the other four reanalysis
377 products show clearly positive trends in $K_{\downarrow}$, with the rate of change varying from -
378 0.07 W m^{-2} per decade in NOAA-CIRES-DOE to 2.02 W m^{-2} per decade in ERA5
379 (Fig. 4a). The average trend of the five products is 0.80 ± 0.74 W m^{-2} per decade
380 (mean ± standard deviation; here standard deviation indicates variation among the
381 five products). The regional mean $K_{\downarrow,d}$ shows an increasing trend according to
382 NCEP/NCAR and decreasing trends according to the other three products (NOAA-
383 CIRES-DOE, MERRA-2, and ERA5), with the rate of change ranging from 0.39 W
384 m^{-2} per decade in NCEP/NCAR to -1.6 W m^{-2} per decade in MERRA-2 (Fig. 4b),
385 with a four-product mean value of -0.86 ± 0.75 W m^{-2} per decade. JRA-55 does not
386 provide $K_{\downarrow,d}$ data.

387 Over China, the trends in $K_{\downarrow}$ are less consistent than those over Europe. Two
388 products (MERRA-2 and JRA-55) show decreasing trends, and three (NOAA-

389 CIRES-DOE, NCEP/NCAR, and ERA5) show increasing trends. The rate of change
390 varies from -0.73 W m^{-2} per decade (MERRA-2) to 1.76 W m^{-2} per decade
391 (NCEP/NCAR), giving a five-product mean of 0.41 ± 0.88 W m^{-2} per decade. In
392 contrast, all products show decreasing trends in $K_{\downarrow,d}$, giving a four-product mean rate
393 of change of -0.72 ± 0.39 W m^{-2} per decade (Fig. 4d).

394 Second, we analyze the time trends using the GEBA data and the reanalysis data
395 from the grids containing these GEBA sites and for the same measurement years.
396 Since the number of sites which fulfill all the quality-control criteria, including the
397 homogeneity test, are small (see Methods section), we stress that these do not
398 necessarily represent regional trends. Instead, we examine whether the gridded
399 products capture the observed trends for the corresponding periods. The trends for
400 the individual stations included for each region are represented by the circles in Fig.
401 4, with the overall mean and standard errors for these shown as hatched bars. For
402 62.5% (15 of 24) of the GEBA sites considered over Europe, we see a positive trend,
403 with an average increasing trend in $K_{\downarrow}$ (2.18 W m^{-2} per decade), which is consistent
404 with existing studies (Wild 2012, 2016; Schwarz et al. 2020). None of the reanalysis
405 products capture the direction of the mean brightening trend for the corresponding
406 grids and years, though the regressions between the observed and modeled trends in
407 $K_{\downarrow}$ are positive for ERA5 (r = 0.28) and MERRA-2 (r = 0.32). For $K_{\downarrow,d}$, only half of
408 the six GEBA sites show decreasing trends. Among the reanalysis products, only
409 ERA5 captures (weakly) the corresponding trends (r = 31).

Over China, 3 of the 4 GEBA sites show a brightening trend (Fig. 4c), with none of the corresponding reanalysis products capturing the variability of trends between the sites. On the contrary, for $K_{\downarrow,d}$, all the four considered GEBA sites show an increase over time, with all the reanalysis products other than NCEP/NCAR capturing the positive direction of the mean trend.

Several previous studies have examined the long-term trends in $K_{\downarrow}$ over Europe and China owing to the larger data coverage and strong temporal trends in these regions (Samukova et al. 2014; Lorenzo et al. 2015; Feng et al. 2018; Schwarz et al. 2020). Although the magnitude of the trends varies across studies depending on quality control of the data and the selection of the observation sites and the time periods of interest, most studies have found strong brightening over Europe and weak to negligible brightening over China since the 1980s. For Europe, Lorenzo et al. (2015) found an increasing trend of 3.2 W m^{-2} per decade for $K_{\downarrow}$ between 1986 and 2012. Similarly, Pfeifroth et al. (2018) found increasing trends between 1.9 W m^{-2} and 2.4 W m^{-2} per decade for 1983-2015. Most recently, Schwarz et al. (2020) found an increase in the $K_{\downarrow}$ absorbed by the surface at a rate of 1.7 W m^{-2} per decade in Europe for the 31-year period between 1985 and 2015. For the time periods corresponding to the three studies mentioned above, we calculate the five-product mean brightening trends of 0.63 W m^{-2} per decade, 0.68 W m^{-2} per decade, and 0.57 W m^{-2} per decade, respectively. Over China, strong brightening trends (by 10.6 W m^{-2} per decade) have been seen for clear-sky conditions between 2006-2018 (Yang et al. 2019). For all-sky conditions, the absorbed $K_{\downarrow}$ at the surface showed a

dimming trend of -0.7 W m^{-2} per decade between 1985 and 2009 and a brightening trend of 1.4 W m^{-2} per decade between 2011 and 2015 (Schwarz et al. 2020). We find a five-product mean increase in $K_{\downarrow}$ by 0.41 W m^{-2} per decade for Common Period II in China. The observed increase in $K_{\downarrow}$ for the subset of GEBA sites in China is not captured by the reanalyses over the corresponding sites. The overall regional brightening has also been found to be missing in unconstrained CMIP5 model simulations (Moseid et al. 2020).

Observational constraints on long-term trends in $K_{\downarrow,d}$ are much rarer, partly because of the lack of sufficient ground stations that measure this variable, as well as higher uncertainties in these measurements. For Europe, a couple of studies show decreasing trend in $K_{\downarrow,d}$ since the 1980s (Samukova et al. 2014; Wild et al. 2017). We also find a decreasing four-product mean trend of -0.86 W m^{-2} per decade for Common Period II. For China, there are more studies on long-term trends in $K_{\downarrow,d}$, generally showing a decrease in $K_{\downarrow,d}$ till the 1990s, followed by an increase till 2010 (Wang and Yang, 2014). For northern China, $K_{\downarrow,d}$ showed an increasing tendency from 1959 to 2016 according to a recent study (Feng et al. 2018), but a strong decreasing tendency for the Beijing and Shenyang stations, both in the northern China, according to another study (Wang et al. 2020). We find a decreasing four-product mean trend in $K_{\downarrow,d}$ of -2.93 (-0.72) W m^{-2} per decade for Common Period II.

e. Role of Clouds and Aerosols on Inter-Model Variability in Gridded Products

Previous studies show overestimation of $K_{\downarrow}$ in reanalysis datasets due to the underestimation of clouds (Zhao et al. 2013; Wild et al. 2015; Loeb et al. 2019).

Here, we separately analyze the correlation of $K_{\downarrow,d}$ and $K_{\downarrow,b}$ with percentage cloud cover at the global scale, both within the gridded products using annual averages and between the products using multi-year averages for Common Period I (Fig. 5). The datasets show similar spatial patterns in cloud cover (Fig. S5), but large differences in global mean values, ranging from 52.3% in NCEP/NCAR to 66.5% in CERES. In general, products with lower cloud cover have higher average $K_{\downarrow,b}$ and lower $K_{\downarrow,d}$, which makes sense mechanistically. The exception to this strong linear relationship (r^2 = 0.96 for $K_{\downarrow,b}$ and 0.92 for $K_{\downarrow,d}$; Fig. 5) is NCEP/NCAR, which has the lowest cloud cover, but not the lowest $K_{\downarrow,d}$ or the highest $K_{\downarrow,b}$. Note that NCEP/NCAR does have the highest $K_{\downarrow}$ (Table 1). Thus, the issue is the partitioning of $K_{\downarrow,d}$ in the product. The underestimation of $K_{\downarrow,b}$ and overestimation of $K_{\downarrow,d}$ in CERES may be due to the positive bias in not just percentage cloud cover (Kato et al. 2018), but the well-known systematic overestimation in MODIS-derived cloud droplet size (Painemal and Zuidema 2011). Larger particles lead to more forward scattering (Plass and Kattawar 1968), which ~~would~~ could contribute to the positive bias in ~~total~~ $K_{\downarrow,d}$ at the surface while simultaneously reducing $K_{\downarrow,b}$.

Although the relationships between annual cloud cover and annual $K_{\downarrow,b}$ (and $K_{\downarrow,d}$) for each gridded product are not consistently strong, we find the expected direction of sensitivity to cloud cover in all the datasets. The sensitivities are positive for $K_{\downarrow,d}$, ranging from 0.22 W m^{-2} per cloud cover percentage in MERRA-2 to 0.87 W m^{-2} per cloud cover percentage in NOAA-CIRES-DOE, and negative for $K_{\downarrow,b}$, ranging from -0.47 W m^{-2} per cloud cover percentage in MERRA-2 to -2.89 W

476 m^{-2} per cloud cover percentage in NOAA-CIRES-DOE. Overall, the collinearity
477 between cloud cover and $K_{\downarrow,b}$ is higher than for $K_{\downarrow,d.}$ The individual scatterplots
478 between $K_{\downarrow,d}$ (and $K_{\downarrow,b}$) and cloud cover percentage are also in Fig. S6. Although
479 there are large uncertainties in both cloud and aerosol representation in coarse-
480 gridded models, that the inter-product variability of these atmospheric constituents
481 controls the inter-product variability in the surface radiation fields is a reasonable
482 assumption. This is because the top of the atmosphere incoming $K_{\downarrow}$ has strong
483 theoretical constraints and varies between 340.3 W m^{-2} in NOAA-CIRES-DOE to
484 341.9 W m^{-2} in NCEP/NCAR for Common Period I. It is harder to separate the
485 relative importance of the individual constituents due to the structural and
486 parametric differences between these products. Cloud-radiation interactions depend
487 not only on aerial coverage of clouds, but also on cloud thickness and cloud optical
488 properties, usually represented by the cloud optical depth (COD). For instance, an
489 underestimation of cloud cover and an overestimation of COD can lead to a positive
490 bias in $K_{\downarrow,b}$ and a negative bias in $K_{\downarrow,d}$ with minimal impact on overall $K_{\downarrow}$, which is
491 seen in ERA5 (Table 2). In contrast, CERES, which has a much higher cloud cover
492 than ERA5 (Fig. 5), shows a negative bias in $K_{\downarrow,b}$ and a positive bias in $K_{\downarrow,d}$, which
493 may be due to either the larger cloud droplet size or underestimated COD or a
494 combination of both (Minnis et al. 2011). The overestimation of optically thick
495 clouds in models compared to satellite observations has been known for a while
496 (Zhang et al. 2005) but has not been used to specifically examine the differences in
497 direct/diffuse partitioning among models. Although COD is not

Formatted: Not Superscript/ Subscript

Formatted: Font: Not Bold, Not Superscript/

publicly archived in most of these reanalysis products, preventing such an analysis in the present study, accurately representing both overall cloud cover and COD might reduce this large variability in direct/diffuse partitioning across these products. The reanalysis products also have large differences in aerosol representation. The NOAA-CIRES-DOE and NCEP/NCAR reanalysis do not include tropospheric aerosols (although NOAA-CIRES-DOE has volcanic aerosols), ERA5 and JRA-55 consider aerosol climatology, and MERRA-2 includes time-varying assimilated aerosols and is the only one of these products that archive AOD. Thus, a similar analysis using all gridded products is not possible for the inter-product variability in aerosols.

We also examined the long-term trends in clouds and aerosols for Europe and China. Figure 6 (a to f) shows the correlation between the trends in cloud cover and the trends in $K_{\downarrow}$ and $K_{\downarrow,d}$ for Europe and China among the six datasets. Common Period I is used for CERES and Common Period II for the reanalysis products. For Europe in particular, these correlations are strong (r^2 = 0.92 for $K_{\downarrow,b}$ and 0.80 for $K_{\downarrow,d}$; Figs 6c and 6e), suggesting that the strength of the brightening over these regions in the gridded data is primarily a function of the trends in the modeled cloud cover. All the datasets other than JRA-55 show a decrease in cloud cover over this region between 1980 and 2015 (Common Period II). We also calculate the trend in AOD for Europe from the MERRA-2 data (Fig. 6g), showing a decreasing trend of 0.04 per decade during the same period. The decrease in aerosol for Europe has been previously seen using both observations and models (Yang et al. 2020). Since

aerosols generally increase $K_{\downarrow,d}$, keeping all other factors constant, this would explain the simultaneous decadal increase in $K_{\downarrow}$ and decrease in $K_{\downarrow,d}$ over Europe. For China, cloud cover decreases in most of the gridded products (other than JRA-55), with the magnitudes of change roughly half of that seen over Europe. The correlation between trends in cloud cover and trends in $K_{\downarrow}$ is relatively weak, though this is primarily driven by MERRA-2 being an outlier (r^2 increases to 0.88 if MERRA-2 is not used in this regression). Incidentally, only MERRA-2 assimilates observations of aerosols, showing an increase in AOD by 0.03 per decade over this region (Fig. 6h). Moreover, according to the grid-averaged trends, $K_{\downarrow,d}$ decreased during this period over China. This pattern could be due to the relative change in absorbing and scattering aerosols over the region during this time period. MERRA-2 data shows a stronger increase in absorbed AOD compared to scattered AOD over China during Common Period II, suggesting a relative enhancement in absorbing aerosols (Fig. 6h). Even though the absorbing component of total AOD in MERRA-2 is modeled, not assimilated, observations bear out the increase in absorbing aerosols over China during this period (Schwarz et al. 2020).

Since in addition to cloud cover, MERRA-2 assimilates gridded AOD, we can estimate the sensitivity of the trends in $K_{\downarrow}$ and $K_{\downarrow,d}$ due to the trends in cloud cover and aerosols by solving this system of two equations:

$$\frac{K_{\mathrm{R,Tr,Eu}}}{K_{\mathrm{R,80,Eu}}} = a\,\frac{\mathrm{CLD_{R,Tr,Eu}}}{\mathrm{CLD_{R,80,Eu}}} + b\,\frac{\mathrm{AOD_{R,Tr,Eu}}}{\mathrm{AOD_{R,80,Eu}}} \quad (1)$$

$$\frac{K_{\mathrm{R,Tr,Ch}}}{K_{\mathrm{R,80,Ch}}} = a\,\frac{\mathrm{CLD_{R,Tr,Ch}}}{\mathrm{CLD_{R,80,Ch}}} + b\,\frac{\mathrm{AOD_{R,Tr,Ch}}}{\mathrm{AOD_{R,80,Ch}}} \quad (2)$$

Here, subscripts $_{R,Tr,Eu}$ and $_{R,Tr,Ch}$ represent the trends in the MERRA-2 gridded products for Europe and China, respectively. The variables considered are the incoming radiation (K_R; either $K_\downarrow$ or $K_{\downarrow,d}$), the cloud cover (CLD), and AOD. Since these variables have different ranges, they are normalized by the value of the variable for the base year (1980; denoted by subscripts $_{R,80,Eu}$ and $_{R,80,Ch}$) to represent the fractional rates of change. For reference, $CLD_{R,80,Eu}$ and $CLD_{R,80,Ch}$ are 63.56% and 47.95%, respectively, while $AOD_{R,80,Eu}$ and $AOD_{R,80,Ch}$ are 0.26 and 0.18. Finally, a and b are the unitless coefficients that give the sensitivity of the fraction rate of change in the radiation components to the fractional rate of change in CLD and AOD, respectively. Simultaneously solving these two equations, we find that both a and b are negative for fractional rate of change in $K_\downarrow$ (a = -0.184, b = -0.037) and positive for the corresponding fractional rate of change in $K_{\downarrow,d}$ (a = 1.074, b = 0.124). These values make sense physically since an increase in aerosols and clouds tends to decrease $K_\downarrow$ and increase $K_{\downarrow,d}$. In terms of magnitude, clouds play a stronger role than aerosols; with the sensitivity being almost 9 times higher for clouds for $K_{\downarrow,d}$ and 5 times stronger for $K_\downarrow$. Over Europe, the effect of clouds and aerosols reinforce each other, with both decreasing, thereby increasing $K_\downarrow$ and decreasing $K_{\downarrow,d}$. Over China, the total effect of aerosols, controlled by both the lower sensitivity to aerosols and the much higher fractional rate of change in AOD, overwhelm the impact of clouds, with $K_\downarrow$ decreasing in spite of a decrease in cloud cover.

We use this framework to constrain the sensitivity of $K_\downarrow$ and $K_{\downarrow,d}$ to CLD and AOD in the MERRA-2 dataset because of its conceptual simplicity. For comparison,

563 we also used multi-linear regressions to examine the trend in $K_{\downarrow}$ and $K_{\downarrow,d}$ as a
564 function of grid-averaged CLD and AOD separately for Europe and China, given by
565 the equation:

566 $$\frac{K_R}{K_{R,80}} = \beta_0 + \beta_1 \frac{CLD_R}{CLD_{R,80}} + \beta_2 \frac{AOD_R}{AOD_{R,80}} \quad (3)$$

567 Here, the annual average values of $K_{\downarrow}$ and $K_{\downarrow,d}$ (K_R), CLD (CLD_R), and AOD
568 (AOD_R) are normalized by their 1980 values and β_0, β_1, and β_2 are the coefficients of
569 regression. Although the results are different over the two regions since they are not
570 mathematically constrained by the same sensitivity as done for Eqs 1 and 2, we get
571 the same signs and similar relative magnitudes of the regression coefficients. β_1 and
572 β_2 are always positive for $K_{\downarrow,d}$ and always negative for $K_{\downarrow}$. Clouds play a stronger
573 role in both Europe and China for $K_{\downarrow}$ (β_1/β_2 = 14.52 and 6.26, respectively) and $K_{\downarrow,d}$
574 (β_1/β_2 = 3.57 and 8.03, respectively).

575 **4. Discussion and Summary**

576 ***a. Comparison with Other Modeled and Satellite-Derived Estimates***

577 We are not aware of any formalized attempts to evaluate $K_{\downarrow,d}$ in current-generation
578 CMIP6 models or their previous iterations. Although operational GCMs may
579 sometimes lead reanalysis products in model development efforts (for instance,
580 frequently using prognostic aerosols instead of prescribed aerosol distributions),
581 many of the radiation codes and cloud parameterizations used to generate the
582 reanalysis products are also implemented in those models. Additionally, that GCMs
583 are run with less constraints on the atmospheric and surface variables than reanalysis

products suggests that there may also be wide disparities in the $K_{\downarrow,d}$ modeled by GCMs. We see evidence of this from two studies that have evaluated $K_{\downarrow,d}$ at larger scales. Mercado et al. (2009) used radiative transfer calculations to simulate $K_{\downarrow,b}$ and $K_{\downarrow,d}$ globally. Using a subset of GEBA observations over Europe, Germany, and China, they evaluated their modeled $K_{\downarrow}$ and k_d. For GEBA stations in Germany and Europe, they found an underestimation in $K_{\downarrow}$ and an overestimation in k_d. Over China, their model overestimated $K_{\downarrow}$, but correctly simulated k_d, suggesting an overestimation in $K_{\downarrow,d}$. More recently, Chakraborty et al. (~~Under Review~~2021) used the latest version (version 6) of the Community Atmosphere Model (CAM6; Gettelman et al. 2019) to simulate $K_{\downarrow,d}$ and $K_{\downarrow,b}$ and evaluated the modeled values using all available GEBA observations. CAM6 overestimated $K_{\downarrow}$ and underestimated $K_{\downarrow,d}$, leading to an MBE of -0.08 for k_d for all GEBA sites.

Our evaluation of the CERES dataset shows that, while CERES does a great job at capturing both the magnitude and variability in $K_{\downarrow}$ (Table 2), there are issues with the direct/diffuse partitioning. CERES overestimates $K_{\downarrow,d}$ and underestimates $K_{\downarrow,b}$, leading to an overestimation in k_d (roughly 0.09 for all GEBA sites; Table 2), potentially caused by higher cloud fraction and cloud droplet size in satellite-derived products (Painemal and Zuidema 2011; Kato et al. 2018). In this context, a few other satellite-derived $K_{\downarrow,d}$ products also warrant discussion. Recently, Jiang et al. (2020) evaluated the $K_{\downarrow,d}$ in a recent dataset (JiEA) created using a deep learning algorithm and geostationary satellite measurements (Jiang et al. 2019a). Using 39 observation sites over East Asia, they found much better performance of the JiEA product

compared to ERA5. Consistent with our results, ERA5 underestimated $K_{\downarrow,d}$ (MBE=-17.2 W m^{-2}; -1.2 W m^{-2} for JiEA) over their study area. The grid-area averaged k_d over East Asia was 0.42 for JiEA and 0.35 for ERA5. For China, which covers a large part of their study area, the grid-averaged k_d for the same time period (2007-2014) varies from only 0.27 in the MERRA-2 dataset to roughly double that value (0.56) in the CERES data. Over Europe and Africa, the Copernicus Atmosphere Monitoring Service (CAMS) provides $K_{\downarrow}$, $K_{\downarrow,d}$, and $K_{\downarrow,b}$ estimates every 15 minutes based on the Heliosat-4 method using Meteosat geostationary satellite observations (Qu et al. 2017). We estimated $K_{\downarrow,d}$ over the region of Europe ('AGATE') covered by these satellites for Common Period I and found the regional average $K_{\downarrow,d}$ to range from 52.8 W m^{-2} in MERRA-2 to 102.6 in CERES versus a value of 65.4 W m^{-2} in CAMS. Another recent study produced global datasets of $K_{\downarrow}$, total photosynthetically active radiation (PAR) and its diffuse component from 2000 to 2016 by combining a radiative transfer model with an artificial neural network trained using MODIS data (Ryu et al. 2018). They calculated a global average ratio of 0.41 for diffuse PAR to total PAR and 0.46 for total PAR to $K_{\downarrow}$.

Of the data products we consider, only MERRA-2, NOAA-CIRES-DOE, and NCEP/NCAR publicly archive the diffuse portion of PAR. For Common Period I, we find large differences in these estimates for the three datasets for both diffuse PAR to total PAR (0.37 for MERRA-2; 0.54 for NOAA-CIRES-DOE; 0.46 for NCEP/NCAR) and for total PAR to $K_{\downarrow}$ (0.44 for MERRA-2; 0.52 for NOAA-CIRES-DOE; 0.61 for NCEP/NCAR). In comparison, the diffuse PAR to total PAR

and total PAR to $K_{\downarrow}$ in the CAM simulations by Chakraborty et al. (~~Under Review~~2021) were 0.41 and 0.51, respectively.

b. Limitations

Point observations have been frequently used to compare against gridded estimates of surface radiation (Markovic et al. 2009; Zhao et al. 2013; Wild et al. 2015). However, radiation transfer calculations in GCMs and reanalyses are based on the plane-parallel approximation, the assumption of one-dimensional atmospheric grids with horizontal planes as the upper and lower bounds, for computational efficiency. The real atmosphere has 3D cloud structures, particularly relevant for cloud-radiation interactions. For instance, cloud side illumination is the interception of radiation due to the existence of cloud sides in the real atmosphere, which are not captured by their plane-parallel approximations; a major issue at high solar zenith angles (Schäfer et al. 2016). Similarly, for low zenith angles, cloud side leakage causes more radiation to pass through the edges of clouds and reach the surface, which would be blocked in a plane-parallel representation (Ham et al. 2014). The overall result of these two mechanisms is generally an underestimation in simulated $K_{\downarrow}$ even when the cloud fraction is correctly captured by the approximation (Okata et al. 2017). Thus, these two effects on their own cannot explain the systematic overestimation in $K_{\downarrow}$ we find in the gridded products (Table 2). Cloud sky leakage would normally lead to more forward scattering, and may thus increase $K_{\downarrow,d}$ in regions with low zenith angle, which we do find in the GEBA observations compared to the MERRA-2 and ERA5 datasets. The effect of cloud side

650 illumination, on the other hand, primarily blocks $K_{\downarrow,b}$ (Hogan and Shonk 2013),
651 which would overestimate $K_{\downarrow,b}$ in GCMs, which is seen in all reanalysis products
652 other than NOAA-CIRES-DOE (Table 2).

653 These problems are most serious at shorter time scales, as patchy clouds can
654 cause large fluctuations in the observations at individual sites. Thus, since the signs
655 of these 3D effects largely depend on zenith angle, the errors are reduced
656 substantially when using monthly means since it averages over the various zenith
657 angles (as was done here) and by combining the biases over multiple sites in a
658 region. Note that the bias errors found here may also be related to other aspects of
659 the 3D cloud structure, such as how overlap of clouds at various heights is
660 represented (Wang et al. 2016). However, the inter-model variability is not affected
661 by these issues since all the products considered use similar approximations. We
662 find that this variability for both $K_{\downarrow,d}$ and $K_{\downarrow,b}$ is strongly controlled by cloud fraction
663 (Fig. 4). Additional differences are also expected due to the shortwave
664 parameterizations used in these datasets that convert the cloud representations to the
665 radiances across wavelength channels. However, such an evaluation requires a
666 modeling setup that controls for the different inputs to the radiative transfer models
667 used in the gridded products and hyperspectral observations for validation (Aumann
668 et al. 2018).

669 A quantitative comparison of the long-term trends using observations requires
670 consistent data coverage. The GEBA data are not always appropriate for this
671 purpose because the trends in $K_{\downarrow,d}$ and $K_{\downarrow}$ (circles and hatched bars, Figure 4) are

derived from two different subsets of the data (there are more $K_{\downarrow}$ observations than $K_{\downarrow,d}$ –observations). Moreover, even within Common Period II, the data coverage changes over time. This lack of consistent data coverage is particularly relevant for China since many studies find a reversal of the trends somewhere between 1990 and 2000, potentially influenced by the instrument changes after 1993 (Wang and Yang, 2014). We try to account for potential breakpoints in the trends by testing for homogeneity of the time-series. However, this reduces the number of available stations substantially, particularly over China (Fig. 4). As such, although the inter-model variability in long-term trends in the gridded datasets illustrates the differences between these models, we advise caution when talking about the 'real' regional trends using GEBA observations, particularly for $K_{\downarrow,d}$ given the dearth of available observations. For China, one alternative is to use data from the China Meteorological Data Service Center (http://data.cma.cn). However, as seen in Wang et al. (2020), after testing for homogeneity, only 12 stations are available with long-term observations of both $K_{\downarrow}$ and $K_{\downarrow,d}$. An in-depth analysis of the influence of station and year range selection on these trends is in Schwarz et al. (2020), though they do not focus on $K_{\downarrow,d}$. Given that we find that the gridded data cannot generally capture either the direction or the variability in trends across the available GEBA sites for the corresponding time-periods, further work is necessary to evaluate long-term trends in $K_{\downarrow,d}$ across CMIP6 models with consolidated observational databases that include both regional and global networks.

c. Summary

694 We find large differences in $K_{\downarrow,d}$, $K_{\downarrow,b}$, and k_d across current-generation gridded
695 products. The variability is evident from the monthly to the annual scales and show
696 large biases from observational benchmarks. For 2001-2015, the range of variability
697 is 10.7 % for global mean $K_{\downarrow}$ (185.4 to 205.3 W m^{-2}), 60.4% for global mean $K_{\downarrow,b}$
698 (82.8 to 132.8 W m^{-2}), 94.3% for global mean $K_{\downarrow,d}$ (52.8 to 102.6 W m^{-2}), and
699 96.4% for global mean k_d (0.28 to 0.55). The variability between these products is
700 statistically explained by the biases in modeled cloud fraction. Long-term (1980-
701 2015) trends in the two variables also differ over Europe and China and are not
702 captured well by the gridded products. These inter-model differences in $K_{\downarrow,d}$ would
703 affect Earth system simulations, particularly relevant for surface climate and for
704 estimating solar energy potential. Thus, we suggest comprehensive comparisons of
705 simulated k_d in the CMIP6 models to better identify potential deficiencies in current-
706 generation atmosphere models.
707

Acknowledgments We acknowledge Computational & Information Systems Lab at the National Center for Atmospheric Research (NCAR), Yale Center for Earth Observation (YCEO), and Microsoft (through the AI for Earth grant) for providing computational resources. This research is supported in part by the US National Science Foundation (grant AGS1933630).

Data availability The CERES data were obtained from the NASA Langley Research Center CERES ordering tool (https://ceres.larc.nasa.gov/). The NOAA-CIRESDOE and NCEP/NCAR reanalysis datasets were downloaded from the PSL website (https://psl.noaa.gov/). The MERRA-2 reanalysis dataset can be found on NASA's website (https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/). The ERA5 reanalysis data were downloaded from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/). The JRA-55 reanalysis data were downloaded from NCAR's Research Data Archive (https://rda.ucar.edu/). The CAMS dataset for Europe was downloaded from http://www.soda-pro.com/.

References

Alexandersson, Hans. 1986: A homogeneity test applied to precipitation data. *Journal of climatology,* **6,** no. 6, 661-675.

Aumann, H. H., & Coauthors, 2018: Evaluation of radiative transfer models with clouds. *Journal of Geophysical Research: Atmospheres*, **123(11)**, 6142-6157.

Bosilovich, M. G., F. R. Robertson, & J. Chen, 2011: Global energy and water budgets in merra. *Journal of Climate*, **24 (22)**, 5721–5739.

Briegleb, B. P., 1992: Delta-eddington approximation for solar radiation in the ncar community climate model. *Journal of Geophysical Research: Atmospheres*, **97 (D7)**, 7603–7612.

Buchard, V., & Coauthors, 2017: The MERRA-2 aerosol reanalysis, 1980 onward. Part II: Evaluation and case studies. *Journal of Climate*, **30**, no. 17: 6851-6872.

Chakraborty, T., & X. Lee, 2019: Land cover regulates the spatial variability of temperature response to the direct radiative effect of aerosols. *Geophysical Research Letters*, **46 (15)**, 8995–9003.

Chakraborty, T., X. Lee, & D. Lawrence, ~~2020~~2021: Strong Local Evaporative Cooling Over Land Due to Atmospheric Aerosols~~Isolating the local temperature response to aerosols through the evaporative pathway~~. ~~*Under Review.*~~*Journal of Advances in Modeling Earth Systems*.

Chou, M.-D., & M. J. Suarez, 1999: A solar radiation parameterization for atmospheric studies. volume 15.

Freidenreich, S., & V. Ramaswamy, 1999: A new multiple-band solar radiative parameterization for general circulation models. *Journal of Geophysical Research: Atmospheres*, **104 (D24)**, 31389–31409.

Gettelman, A., & Coauthors, 2019: The whole atmosphere community climate model version 6 (waccm6). *Journal of Geophysical Research: Atmospheres*, **124 (23)**, 12380– 12403.

Gilgen, H., & A. Ohmura, 1999: The global energy balance archive. *Bulletin of the American Meteorological Society*, **80 (5)**, 831–850.

Ham, S. H., S., Kato, H. W., Barker, F. G., Rose, & S. Sun-Mack, 2014: Effects of 3-D clouds on atmospheric transmission of solar radiation: Cloud type dependencies inferred from A-train satellite data. *Journal of Geophysical Research: Atmospheres*, **119(2)**, 943-963.

Hersbach, H., and Coauthors, 2020: The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*.

Formatted: Font color: Auto

760 Hogan, R. J., & J. K. Shonk, 2013: Incorporating the effects of 3D radiative transfer
761 in the presence of clouds into two-stream multilayer radiation schemes. *Journal of*
762 *the atmospheric sciences*, **70(2)**, 708-724.

763 Hou, Y.-T., S. Moorthi, & K. Campana, 2002: Parameterization of solar radiation
764 transfer in the ncep models.

765 Jiang, H., N. Lu, J. Qin, W. Tang, & L. Yao, 2019a: A deep learning algorithm to
766 estimate hourly global solar radiation from geostationary satellite data. *Renewable*
767 *and Sustainable Energy Reviews*, **114**, 109327.

768 Jiang, H., Y. Yang, H. Wang, Y. Bai, & Y. Bai, 2020: Surface diffuse solar
769 radiation determined by reanalysis and satellite over east asia: Evaluation and
770 comparison. *Remote Sensing*, **12 (9)**, 1387.

771 Jiang, H., Y. Yang, Y. Bai, & H. Wang, 2019b: Evaluation of the total, direct, and
772 diffuse solar radiations from the era5 reanalysis data in china. *IEEE Geoscience and*
773 *Remote Sensing Letters*.

774 Kalnay, E., & Coauthors, 1996: The ncep/ncar 40-year reanalysis project. *Bulletin*
775 *of the American meteorological Society*, **77 (3)**, 437–472.

776 Kato, S., & Coauthors, 2018: Surface irradiances of edition 4.0 clouds and the
777 earth's radiant energy system (ceres) energy balanced and filled (ebaf) data product.
778 *Journal of Climate*, **31 (11)**, 4501–4527.

779 Kennedy, A. D., X. Dong, B. Xi, S. Xie, Y. Zhang, & J. Chen, 2011: A comparison
780 of merra and narr reanalyses with the doe arm sgp data. *Journal of Climate*, **24 (17)**,
781 4541–4557.

782 Kistler, R., & Coauthors, 2001: The ncep–ncar 50-year reanalysis: monthly means
783 cd-rom and documentation. *Bulletin of the American Meteorological society*, **82 (2)**,
784 247–268.

785 Knohl, A., & D. D. Baldocchi, 2008: Effects of diffuse radiation on canopy gas
786 exchange processes in a forest ecosystem. *Journal of Geophysical Research:*
787 *Biogeosciences*, **113 (G2)**.

788 Kobayashi, S., & Coauthors, 2015: The jra-55 reanalysis: General specifications
789 and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, **93**
790 **(1)**, 5–48.

791 Lacis, A. A., & J. Hansen, 1974: A parameterization for the absorption of solar
792 radiation in the earth's atmosphere. *Journal of the atmospheric sciences*, **31 (1)**,
793 118–133.

794 Lee, K.-T., & Coauthors, 2016: Concentrator photovoltaic module architectures
795 with capabilities for capture and conversion of full global solar radiation.
796 *Proceedings of the National Academy of Sciences*, **113 (51)**, E8210–E8218.

797 Liu, B. Y., & R. C. Jordan, 1960: The interrelationship and characteristic
798 distribution of direct, diffuse and total solar radiation. *Solar energy*, **4(3)**, 1-19.

Loeb, Norman G., & Coauthors, 2019: Decomposing Shortwave Top-of-Atmosphere and Surface Radiative Flux Variations in Terms of Surface and Atmospheric Contributions. *Journal of Climate*, **32.16**, 5003-5019.

Markovic, M., C. G. Jones, K. Winger, & D. Paquin, 2009: The surface radiation budget over north america: gridded data assessment and evaluation of regional climate models. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **29 (15)**, 2226–2240.

Mercado, L. M., N. Bellouin, S. Sitch, O. Boucher, C. Huntingford, M. Wild, & P. M. Cox, 2009: Impact of changes in diffuse radiation on the global land carbon sink. *Nature*, **458 (7241)**, 1014–1017.

Minnis, P., & Coauthors, 2011: CERES edition-2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data—Part II: Examples of average results and comparisons with other data. *IEEE Transactions on Geoscience and Remote Sensing*, **49(11),** 4401-4430.

Morcrette, J., H. W. Barker, J. Cole, M. J. Iacono, & R. Pincus, 2008: Impact of a new radiation package, mcrad, in the ecmwf integrated forecasting system. *Monthly weather review*, **136 (12)**, 4773–4798.

Moseid, & Coauthors, 2020: Bias in CMIP6 models compared to observed regional dimming and brightening trends (1961–2014), Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-1210, in review.

Formatted: Font: (Default) Times New Roma
Font color: Black, Pattern: Clear

Formatted: Font: (Default) Times New Roma
Font color: Black, Pattern: Clear

Formatted: Font: (Default) Times New Roma
Font color: Black, Pattern: Clear

819 Okata, M., T. Nakajima, K. Suzuki, T. Inoue, T. Y. Nakajima, & H. Okamoto,
820 2017: A study on radiative transfer effects in 3-D cloudy atmosphere using satellite
821 data. *Journal of Geophysical Research: Atmospheres*, **122(1)**, 443-468.

822 Oreopoulos, L., & Coauthors, 2012: The continual intercomparison of radiation
823 codes: Results from phase I. *Journal of Geophysical Research: Atmospheres*, **117**,
824 no. D6.

825 Painemal, D., & P. Zuidema, 2011: Assessment of MODIS cloud effective radius
826 and optical thickness retrievals over the Southeast Pacific with VOCALS-REx in
827 situ measurements. *Journal of Geophysical Research: Atmospheres*, **116.D24**.

828 Pfeifroth, U., A. Sanchez-Lorenzo, V. Manara, J. Trentmann, & R. Hollmann,
829 2018: Trends and Variability of Surface Solar Radiation in Europe Based On
830 Surface- and Satellite-Based Data Records. *J. Geophys. Res. Atmos.* **123**, 1735–
831 1754.

832 Pincus, R., P. M. Forster, & B. Stevens, 2016: The radiative forcing model
833 intercomparison project (rfmip): experimental protocol for cmip6. *Geoscientific*
834 *Model Development*, **9 (9)**, 3447–3460.

835 Plass, G. N., & G. W. Kattawar, 1968: Monte Carlo calculations of light scattering
836 from clouds. *Applied optics*, **7(3)**, 415-419.

837 Qu, Z., Oumbe, A., & Coauthors, 2017: Fast radiative transfer parameterisation for
838 assessing the surface solar irradiance: The Heliosat-4 method. *Meteorologische*
839 *Zeitschrift*, **26(1)**, 33-57.

840 Randles, C., & Coauthors, 2017: The MERRA-2 aerosol reanalysis, 1980 onward.
841 part I: System description and data assimilation evaluation. *Journal of Climate*, **30**
842 **(17)**, 6823– 6850.

843 Rap, A., & Coauthors, 2018: Enhanced global primary production by biogenic
844 aerosol via diffuse radiation fertilization. *Nature Geoscience*, **11 (9)**, 640–644.

845 Rutan, D. A., S. Kato, D. R. Doelling, F. G. Rose, L. T. Nguyen, T. E. Caldwell, &
846 N. G. Loeb, 2015: Ceres synoptic product: Methodology and validation of surface
847 radiant flux. *Journal of Atmospheric and Oceanic Technology*, **32 (6)**, 1121–1143.

848 Ryu, Y., C. Jiang, H. Kobayashi, & M. Detto, 2018: Modis-derived global land
849 products of shortwave radiation and diffuse and total photosynthetically active
850 radiation at 5 km resolution from 2000. *Remote Sensing of Environment*, **204**, 812–
851 825.

852 Saha, S., & Coauthors, 2010: The NCEP climate forecast system
853 reanalysis. *Bulletin of the American Meteorological Society*, **91(8)**, 1015-1058.

854 Samukova, E. A., E. V. Gorbarenko, & A. E. Erokhina, 2014: Long-term variations
855 of solar radiation in Europe. *Russian Meteorology and Hydrology* **39,** 514–520.

856 Sanchez-Lorenzo, A. & Coauthors, 2015: Reassessment and update of long-term
857 trends in downward surface shortwave radiation over Europe (1939–2012). *J.*
858 *Geophys. Res. Atmos.* **120**, 9555–9569.

859 Schäfer, S. A., R. J., Hogan, C., Klinger, J. C., Chiu, & B. Mayer, 2016:
860 Representing 3-D cloud radiation effects in two-stream schemes: 1. Longwave
861 considerations and effective cloud edge length. *Journal of Geophysical Research:*
862 *Atmospheres*, **121(14)**, 8567-8582.

863 Schwarz, M., D. Folini, S. Yang, R. P. Allan, & M. Wild, 2020: Changes in
864 atmospheric shortwave absorption as important driver of dimming and brightening.
865 *Nature Geoscience*, **13 (2)**, 110–115.

866 Slivinski, L. C., & Coauthors, 2019: Towards a more reliable historical reanalysis:
867 Improvements for version 3 of the twentieth century reanalysis system. *Quarterly*
868 *Journal of the Royal Meteorological Society*, **145 (724)**, 2876–2908.

869 Stephens, G. L., & Coauthors, 2012: An update on earth's energy balance in light of
870 the latest global observations. *Nature Geoscience*, **5 (10)**, 691–696.

871 Wang, X., Y., Liu, & Q. Bao, 2016: Impacts of cloud overlap assumptions on
872 radiative budgets and heating fields in convective regions. *Atmospheric*
873 *Research*, **167**, 89-99.

874 Wang, Y. W. & Y. H. Yang, 2014: China's dimming and brightening: evidence,
875 causes and hydrological implications. *Annales Geophysicae* 32, 41–55.

876 Wang, Y., S. Yang, A. Sanchez-Lorenzo, W. Yuan, & M. Wild, 2020: A revisit of
877 direct and diffuse solar radiation in china based on homogeneous surface
878 observations: Climatology, trends, and their probable causes. *Journal of*
879 *Geophysical Research: Atmospheres*, **125 (9)**, e2020JD032634.

880 Wild, M., 2012: Enlightening global dimming and brightening. *Bulletin of the*
881 *American Meteorological Society*, **93 (1)**, 27–37.

882 Wild, M., 2016: Decadal changes in radiative fluxes at land and ocean surfaces and
883 their relevance for global warming. *Wiley Interdisciplinary Reviews: Climate*
884 *Change*, **7 (1)**, 91–107.

885 Wild, M., A. Ohmura, C. Schär, G. Müller, D. Folini, M. Schwarz, & M. Z.
886 Hakuba, 2017: The global energy balance archive (geba) version 2017: A database
887 for worldwide measured surface energy fluxes. *Earth System Science Data*, **9 (2)**,
888 601–613.

889 Wild, M., A. Ohmura, H. Gilgen, & J.-J. Morcrette, 1998: The distribution of solar
890 energy at the earth's surface as calculated in the ecmwf re-analysis. *Geophysical*
891 *research letters*, **25 (23)**, 4373–4376.

892 Wild, M., & Coauthors, 2015: The energy balance over land and oceans: an
893 assessment based on direct observations and CMIP5 climate models. *Climate*
894 *Dynamics*, **44(11-12)**, 3393-3429.

Yang, S., X. L. Wang, X. L. & M. Wild, 2019: M. Causes of Dimming and Brightening in China Inferred from Homogenized Daily Clear-Sky and All-Sky in situ Surface Solar Radiation Records (1958–2016). *Journal of Climate* **32,** 5901–5913 (2019).

Yang, Y., Lou, S., Wang, H., Wang, P. and Liao, H., 2020: Trends and source apportionment of aerosols in Europe during 1980–2018 *Atmospheric Chemistry and Physics*, **20(4)**, pp.2579-2590.

Yue, X., & N. Unger, 2017: Aerosol optical depth thresholds as a tool to assess diffuse radiation fertilization of the land carbon uptake in china. *Atmos. Chem. Phys*, **17**, 1329– 1342.

Zhang, M. H., & Coauthors, 2005: Comparing clouds and their seasonal variations in 10 atmospheric general circulation models with satellite measurements. *Journal of Geophysical Research: Atmospheres*, **110(D15)**.

Zhang, X., S. Liang, G. Wang, Y. Yao, B. Jiang, & J. Cheng, 2016: Evaluation of the reanalysis surface incident shortwave radiation products from ncep, ecmwf, gsfc, and jma using satellite and surface observations. *Remote Sensing*, **8 (3)**, 225.

Zhao, L., X. Lee, & S. Liu, 2013: Correcting surface solar radiation of two data assimilation systems against fluxnet observations in north america. *Journal of Geophysical Research: Atmospheres*, **118 (17)**, 9552–9564.

Formatted: Font: (Default) Times New Roma
Font color: Black, Pattern: Clear

Formatted: Font: (Default) Times New Roma
Font color: Black, Pattern: Clear

Formatted: Font: (Default) Times New Roma
Font color: Black, Pattern: Clear

915

916 **Table 1**. Summary of the global gridded products considered in the present
917 study. The global means and interannual standard deviations of the variables of
918 interest for Common Period I (2001-2015) are also noted.

Data product	Spatial resolution	Temporal coverage	Reference	Global $K_{\downarrow}$ (W m^{-2})	Global $K_{\downarrow,d}$ (W m^{-2})	Global $K_{\downarrow,b}$ (W m^{-2})	Global k_d (unitless)
NOAA-CIRES-DOE	1◦ × 1◦	1836 - 2015	Slivinski et al. (2019)	192.9 ± 0.4	92.2 ± 0.2	100.7 ± 0.5	0.478 ± 0.002
NCEP/NCAR	1.85° × 1.85°	1948 - Present	Kistler et al. (2001)	205.3 ± 0.6	81 ± 0.3	124.3 ± 0.6	0.394 ± 0.002
JRA-55	0.562° × 0.562°	1958 - Present	Kobayashi et al. (2015)	189 ± 0.8	NA	NA	NA
MERRA-2	0.5° × 0.625°	1980 - Present	Randles et al. (2017)	185.6 ± 0.7	52.8 ± 0.4	132.8 ± 1	0.284 ± 0.003
ERA5	0.25° × 0.25°	1979 - Present	Hersbach et al. (2020)	187.9 ± 0.4	63.7 ± 0.1	124.2 ± 0.4	0.339 ± 0.001
CERES	1◦ × 1◦	2000 - 2019	Rutan et al. (2015)	185.4 ± 0.3	102.6 ± 0.7	82.8 ± 0	0.553 ± 0.004

919

Table 2. Evaluations of monthly mean incoming shortwave radiation ($K_{\downarrow}$), direct beam radiation ($K_{\downarrow,b}$), diffuse radiation ($K_{\downarrow,d}$), and diffuse fraction (k_d) at the surface against the common GEBA observations for Common Period I (2001-2015). Statistical summaries of the evaluations include the intercept and slope of the line of best fit, coefficient of determination (r^2), Mean Bias Error (MBE), and Mean Percentage Error (MPE). The sample size is 14155 in all cases.

	Slope	Intercept	*r2*	RMSE	MBE	MPE	Slope	Intercept	*r2*	RMSE	MBE	MPE
	Total shortwave radiation ($K_{\downarrow}$)						Direct beam radiation ($K_{\downarrow,b}$)					
NOAA-CIRES-DOE	1	8.82	0.9	30.16	9.52	5.9	0.89	6.15	0.75	34.66	-4.28	-4.4
NCEP/NCAR	1.01	43.91	0.9	53.45	45.56	28	0.89	47.59	0.73	51.43	36.79	38.2
JRA-55	0.98	18.25	0.93	26.71	15.27	9.4	NA	NA	NA	NA	NA	NA
MERRA-2	1.03	13.06	0.94	28.22	18.54	11.4	1.08	32.11	0.86	49.44	39.9	41.5
ERA5	0.99	9.72	0.96	18.04	7.51	4.6	0.99	18.15	0.9	27.17	17.38	18.1
CERES	0.98	6.45	0.97	15.72	2.65	1.6	0.8	2.7	0.88	28.32	-16.82	-17.5
	Diffuse radiation ($K_{\downarrow,d}$)						Diffuse fraction (k_d)					
NOAA-CIRES-DOE	0.94	17.68	0.75	21.67	13.8	20.8	0.7	0.21	0.41	0.16	0.08	16.4
NCEP/NCAR	0.73	26.91	0.66	20.01	8.77	13.2	0.41	0.2	0.36	0.14	-0.07	-14.5
MERRA-2	0.61	4.84	0.83	25.84	-21.36	-32.2	0.56	0.03	0.62	0.2	-0.18	-38.4
ERA5	0.7	9.86	0.86	16	-9.87	-14.9	0.72	0.05	0.72	0.11	-0.08	-17.5
CERES	1.28	0.89	0.86	26.61	19.47	29.4	0.76	0.2	0.67	0.13	0.09	19.8

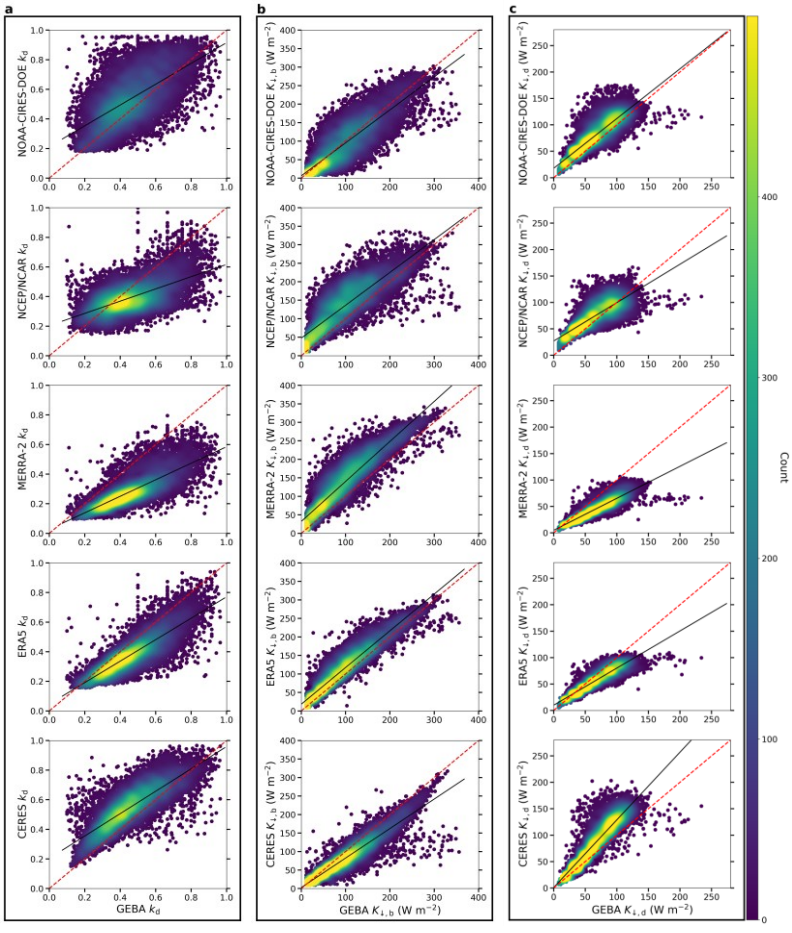

Fig. 1 Evaluations of monthly mean incoming **a** diffuse fraction (k_d), **b** beam radiation ($K_{\downarrow,b}$), and **c** diffuse radiation ($K_{\downarrow,d}$) -at the surface of gridded reanalysis and CERES products against the common GEBA observations for Common Period I (2001-2015). The red dashed lines represent the 1:1 relationship. Color indicates data density. Statistical summaries of the evaluations are in Table 2.

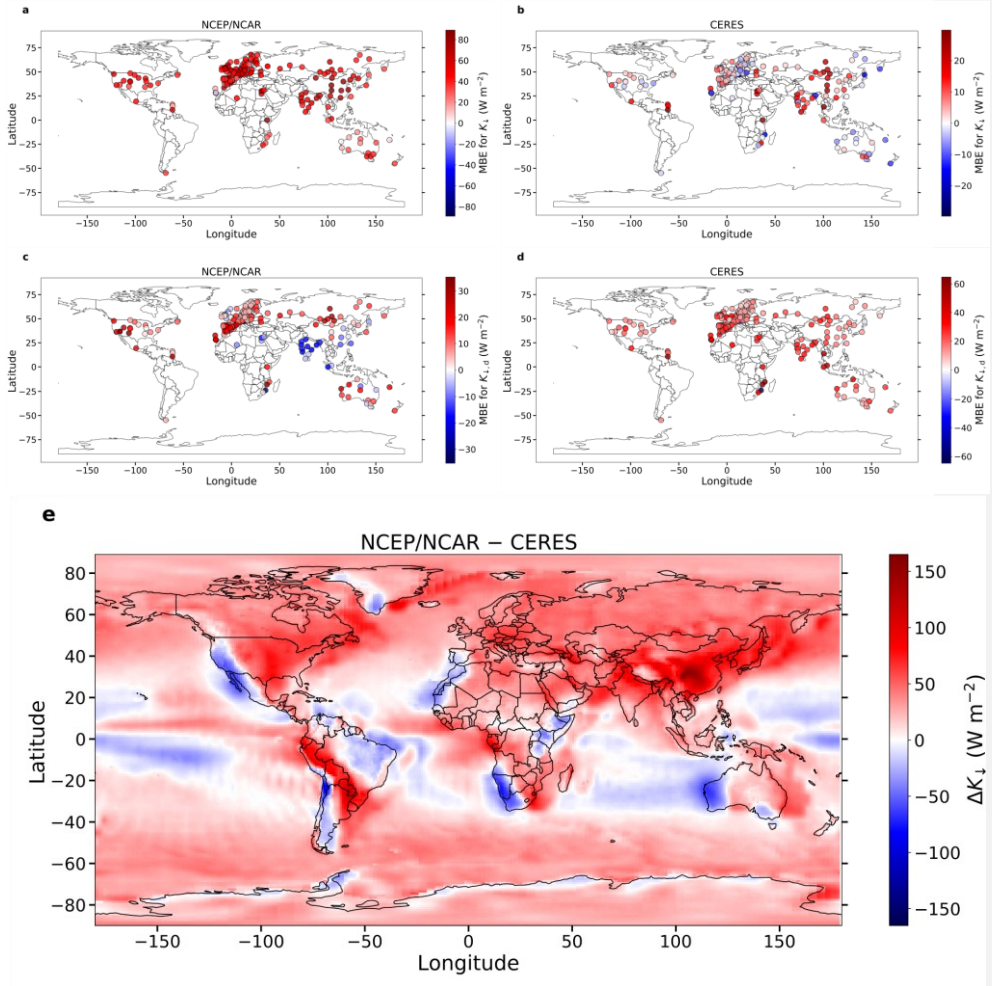

Fig. 2. Site-level mean bias error (MBE) in incoming shortwave radiation ($K_{\downarrow}$) at the surface for **a** NCEP/NCAR and **b** CERES data compared to common GEBA observations for Common Period I (2001-2015). Sub-figures **c** and **d** show mean bias error in incoming diffuse radiation ($K_{\downarrow,d}$) at the surface for NCEP/NCAR and CERES, respectively. Finally, sub-figure **e** shows the grid-wise difference in $K_{\downarrow}$ between NCEP/NCAR and CERES data.

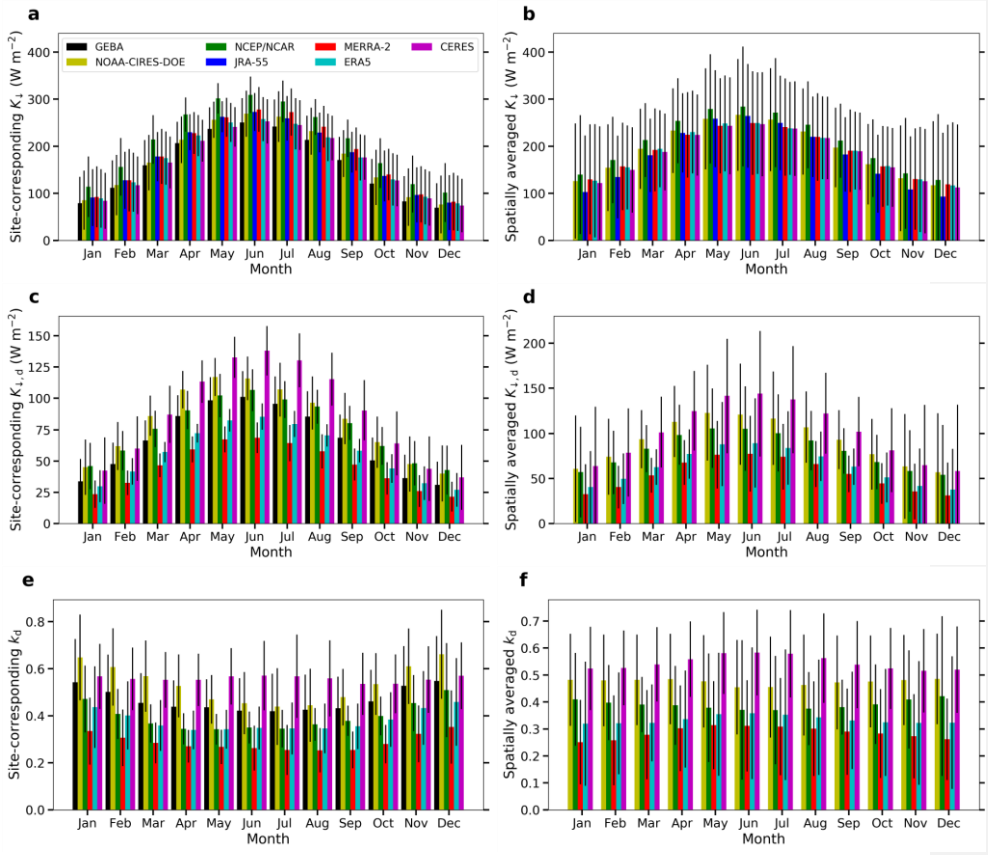

Fig. 3. Seasonal variation in **a** incoming shortwave radiation ($K_{\downarrow}$), **c** incoming diffuse radiation ($K_{\downarrow,d}$), and **e** diffuse fraction (k_d) at the surface for all northern Hemisphere common GEBA sites and the grids overlaying the sites for Common Period I (2001-2015). Sub-figures **b**, **d**, and **f** show the corresponding northern hemisphere means from the gridded products for the same period. In all cases, the black lines show the standard deviations (of the site-level data for **a**, **c**, and **e** and the spatial variability of the grid values for **b**, **d**, and **f**).

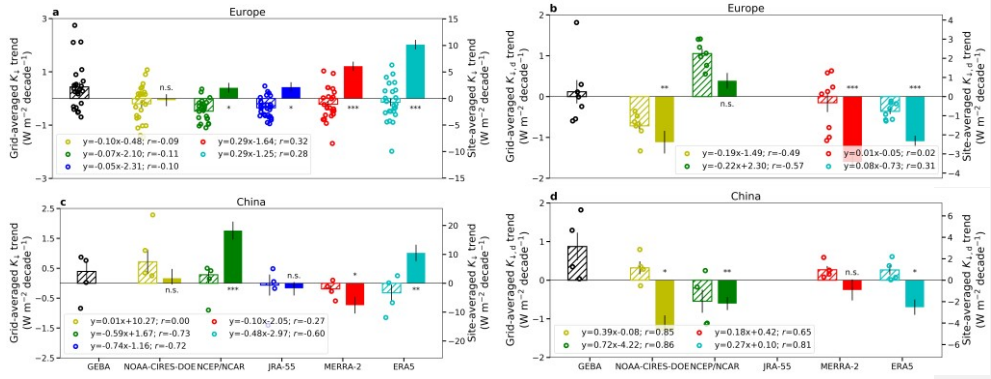

Fig. 4. Long-term trends from gridded and observed data. Sub-figures **a** and **b** show trends in incoming shortwave radiation ($K_{\downarrow}$) and incoming diffuse radiation ($K_{\downarrow,d}$) at the surface over Europe. The long-terms trends in the GEBA observations with at least 10 years of data in Common Period II (1980-2015), as well as the corresponding trends for the overlapping grids from the gridded products, are shown with circles. The hatched bars show the mean (± standard error) of the trends based on these circles. Equations of lines representing the associations between the observed and the corresponding modeled trends are in the legends. The solid bars show the grid-area averaged regional values for the gridded reanalysis products for the entire period (1980-2015), and the error bars represent the standard errors. Sub-figures **c** and **d** are similar, but for China. The p-values of the grid-averaged trends are indicated by asterisks, with three asterisks representing $p<0.0001$, two for $p<0.001$, and one for $p<0.05$.

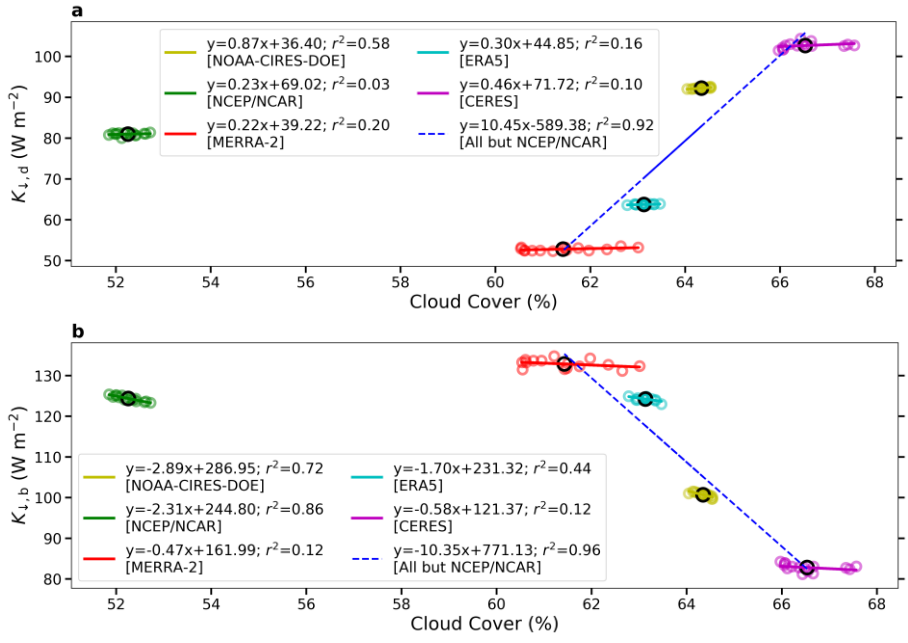

Fig. 5. Associations between percentage cloud cover and **a** incident diffuse radiation ($K_{\downarrow,d}$) and **b** direct beam radiation ($K_{\downarrow,b}$) for Common Period I (2001-2015). Each colored circle represents an annual mean value, while the black circle shows the multi-year average for the gridded product. The lines of best fit and their equations are shown, both for individual gridded products and across products (not considering NCEP/NCAR).

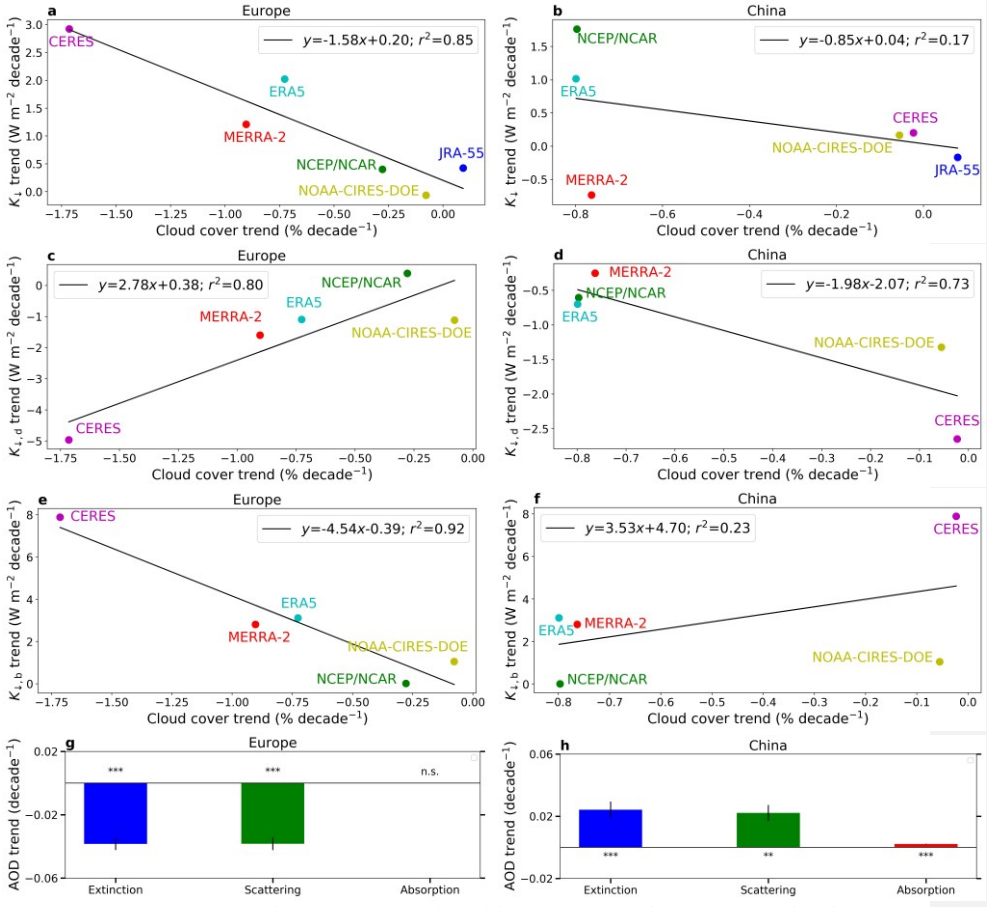

Fig. 6. Linear regressions between trends in grid-area averaged percentage cloud cover and trends in **a** incoming shortwave radiation ($K_{\downarrow}$) over Europe, **b** $K_{\downarrow}$ over China, **c** incoming diffuse radiation ($K_{\downarrow,d}$) over Europe, **d** $K_{\downarrow,d}$ over China, **e** direct beam radiation ($K_{\downarrow,b}$) over Europe, and **f** $K_{\downarrow,b}$ over China for all the gridded products during Common Period II (1980-2015) for the reanalysis products and Common Period I (2001-2015) for CERES. The equations for the lines of best fit are annotated. Sub-figures **g** and **h** show the trends in grid-area averaged aerosol optical depth (AOD), separated into the scattering and absorbing components, for Common Period II as assimilated by MERRA-2. For **g** and **h**, the black lines show the standard errors. The p-values are indicated for **g** and **h** by asterisks, with three asterisks representing $p<0.0001$, two for $p<0.001$, and one for $p<0.05$.