

## Mapping an Observation-Based Global Solar Irradiance Climatology across the Conterminous United States

DAVID E. RUPP,<sup>a</sup> CHRISTOPHER DALY,<sup>b</sup> MATTHEW K. DOGGETT,<sup>b</sup> JOSEPH I. SMITH,<sup>b</sup> AND BEN STEINBERG<sup>b</sup>

<sup>a</sup> Oregon Climate Change Research Institute, College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon

<sup>b</sup> PRISM Climate Group, Northwest Alliance for Computational Science and Engineering, College of Engineering, Oregon State University, Corvallis, Oregon

(Manuscript received 10 November 2021, in final form 20 March 2022)

**ABSTRACT:** The exponential growth in solar radiation measuring stations across the conterminous United States permits the generation of gridded solar irradiance data that capture the spatiotemporal variability of solar irradiance far more accurately than was previously possible from ground-based observations. Taking advantage of these observations, we generated a 30-yr climatology (1991–2020) of mean monthly global irradiance at a resolution of 30 arc s (~800 m) on both a horizontal surface and a sloped ground surface. This paper describes the methods used to generate the gridded data, which include extensive quality control of station data, spatial interpolation of effective cloud transmittance using the “PRISM” method, and simulation of the effects of elevation, shading, and reflection from nearby terrain on solar irradiance. A comparison of the new dataset with several other solar radiation products reveals some spatial features in solar radiation that are either lacking or underresolved in some or all of the other datasets. Examples of these features include strong gradients near foggy coastlines and along mountain ranges where there is persistent orographically driven cloud formation. The workflow developed to create the long-term means will be used as a template for generating time series of monthly and daily solar radiation grids up to the present.

**KEYWORDS:** Atmosphere; North America; Shortwave radiation; Seasonal cycle

### 1. Introduction

Solar radiation is a major component of the radiation budget at Earth’s surface and drives many hydrological, biological, and chemical processes. As such, downwelling surface shortwave radiation (surface solar irradiance) is a key input variable used in models of land surface processes. These land surface models (a term we use in a general sense to include models of, e.g., hydrological processes, snow and ice dynamics, vegetation dynamics, and carbon and nutrient fluxes) often use gridded spatial meteorological data as inputs.

Reanalysis delivers gridded datasets of meteorological variables, including global (direct + diffuse) horizontal irradiance (GHI), derived from retrospective runs with weather forecast models that assimilate historical data of various types (Dee et al. 2013). Being both long-duration (multidecadal) and high frequency (typically 1 or 3 h), reanalysis provides a basis for climatological studies where observations are not available or are of

short duration. Refinements in reanalysis such as the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al. 2017), and the fifth major global reanalysis produced by ECMWF (ERA5; Hersbach et al. 2020) include improved representation of radiative fluxes through the assimilation of satellite-derived aerosol concentrations and cloud cover. The spatial resolution has also increased; for example, ERA5 is on a  $0.5^\circ \times 0.5^\circ$  horizontal grid and the land component of ERA5 (ERA5-Land) driven by downscaled meteorological forcing from ERA5 is on a  $0.1^\circ \times 0.1^\circ$  horizontal grid (Muñoz-Sabater et al. 2021). These reanalyses, however, do not assimilate ground-based solar irradiance measurements.

In contrast, WorldClim 2 (Fick and Hijmans 2017; herein simply referred to as WorldClim) provides gridded, high-resolution ( $30 \text{ arc s} \times 30 \text{ arc s}$ , or  $\sim 1 \text{ km}$ ) GHI over land derived from interpolated ground-based measurements using longitude, latitude, elevation, and satellite-derived cloud cover as covariates. Covering the globe, the intent of WorldClim is not necessarily to maximize information over the United States. Also, the GHI grids are limited to climatological monthly averages representing the period 1970–2000, so do not take advantage of the proliferation of stations measuring solar radiation in the United States during the last two decades (see Fig. 1).

A literature review of other gridded GHI datasets of relatively high spatial resolution ( $\leq 1/4^\circ$ ) over the conterminous United States (CONUS) reveals that ground-based GHI observations are seldom incorporated into their development, and when they are, it is typically through using a relatively small number of stations to help to reduce biases in the gridded product. Moreover, the underlying GHI information can be traced to a few sources for most datasets. Several gridded datasets rely on the North

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/JAMC-D-21-0236.s1>.

Rupp’s ORCID: 0000-0003-3562-2072.

Corresponding author: David E. Rupp, david.rupp@oregonstate.edu

DOI: 10.1175/JAMC-D-21-0236.1

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

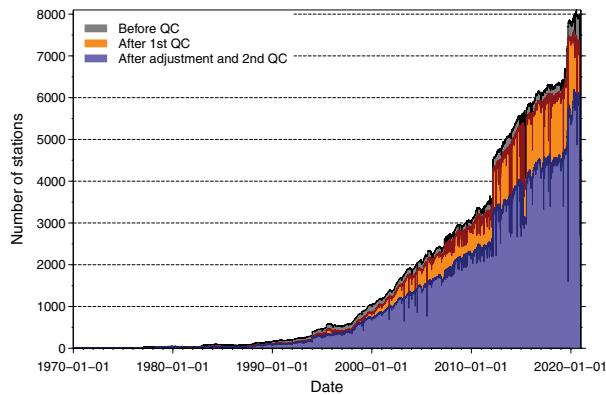


FIG. 1. Number of stations with GHI measurements in CONUS acquired for this study: all stations with GHI data ingested into the database (slate blue plus orange plus gray shading); stations with accepted daily GHI after the first QC stage, which culls outliers or days with incomplete subhourly data (slate blue plus orange shading); stations with accepted daily GHI after adjustment and second QC stage (slate blue shading).

American Regional Reanalysis (NARR; Mesinger et al. 2006), which provides simulated GHI at 3-hourly and 32-km resolution. NARR assimilates multiple types of data but not ground-based GHI. The North American Land Data Assimilation System (NLDAS; Cosgrove et al. 2003; Xia et al. 2012) regrid NARR GHI to  $1/8^\circ$  and disaggregates to an hourly time step while using the *Geostationary Operational Environmental Satellite-8* (GOES-8) to reduce the CONUS-wide mean monthly bias. GridMET (Abatzoglou 2013) aggregates hourly GHI from NLDAS (whose source is NARR) to a daily time step and bilinearly interpolates to  $1/24^\circ$ . Holden et al. (2018) generate daily GHI by simulating clear-sky irradiance at 8 arc s that is then adjusted for cloud cover, where the cloud cover adjustment is derived from NLDAS (again, whose source is NARR). Holden et al. (2018) bias correct NLDAS solar irradiance data using stations from the National Solar Radiation Database (NSRDB; Wilcox 2012) and the U.S. Climate Reference Network (USCRN; Diamond et al. 2013).

Other gridded datasets rely on the Mountain Microclimate Simulation Model (MTCLIM; Thornton and Running 1999; Bohn et al. 2013) to estimate daily irradiance indirectly as a function of the daily maximum and minimum temperature and daily precipitation. Daymet (Thornton et al. 1997, 2021) provides daily global irradiance at  $1 \text{ km} \times 1 \text{ km}$  resolution using MTCLIM wherein the temperature and precipitation observations from the Global Historical Climatology Network (GHCN; Menne et al. 2012) are spatially interpolated (Thornton et al. 1997, 2000). Livneh et al. (2013) calculate GHI using MTCLIM at  $1/16^\circ$  wherein temperature and daily precipitation observations from NOAA Cooperative Observer (COOP) stations are spatially interpolated. The Multiscale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP; Huntzinger et al. 2013) uses both MTCLIM and NARR. Daily GHI is calculated at  $1/4^\circ$  using MTCLIM with temperature and precipitation taken from NARR after correcting precipitation biases (Wei et al. 2014).

Yet other gridded datasets are based on models used in NSRDB, including the “SUNY” model incorporating geostationary satellite data (Perez et al. 2002). SolarAnywhere provides GHI at a spatial and temporal resolution as fine as  $0.01^\circ$  and 1 min, respectively, based on Perez et al. (2002) with modifications summarized on the SolarAnywhere website.

Not all the above datasets are freely or readily available for researchers. The higher-resolution SolarAnywhere products must be purchased and have restrictive licensing agreements, and other datasets were created for a specific research project and are not publicly available for download (Holden et al. 2018).

The above datasets contain errors that may be significant depending on the application. For example, Slater (2016) found that datasets derived from NLDAS and MTCLIM (Daymet; MsTMIP) had mean summer GHI errors exceeding  $\pm 10\%$  of the observed mean and the errors varied strongly across CONUS (spatial correlations  $< 0.7$  between observed and modeled values of the ratio of actual to clear-sky GHI at the surface). Slater (2016) also noted particularly large positive biases in NARR over all CONUS. GOES products can have strong east–west variability in bias over CONUS (Slater 2016) and both Jepsen et al. (2012) and Slater et al. (2013) found periods of erratic values and large systematic biases in parts of the western United States.

Although having an observation-based dataset as an alternative or addition to the above datasets is desirable, numerous challenges exist in using solar irradiance observations to make a quality, gridded, and up-to-date CONUS-wide product, which may explain why none exist, particularly given the availability of reanalysis products that are of both finer resolution and higher quality with each generation. One challenge is that no single solar radiation network has sufficient spatial and temporal coverage over CONUS for reliable high-resolution mapping (Kafka and Miller 2019). Achieving dense coverage requires acquiring data from many networks with different protocols of data curation and standardizing the data. However, such pooling and standardization of data has been accomplished for various meteorological variables (Daly et al. 2008) including solar radiation (Slater 2016). A second challenge is filling the spatial gaps between solar radiation measurements in a way that is computationally feasible but still emulates the physical processes that drive spatial variability in solar irradiance. While a variety of interpolation methods exist and many are readily available as packages for commonly used software (e.g., Kafka and Miller 2019), we might expect that greater accuracy can be achieved when the influence of elevation, coastal proximity, vertical atmospheric layer (boundary layer and free atmosphere), and topographic position, for example, are considered, such as has been done for variables like temperature, precipitation, dewpoint, and vapor pressure deficit (Daly et al. 1994, 2008, 2015) and, to some extent, for solar radiation (Fick and Hijmans 2017).

The challenge of coping with radiometer measurement error may be the primary reason no observation-based gridded datasets have been developed using the large number of records currently available. Research-grade observations provide the highest accuracy but are relatively rare, especially those of

long duration (Gueymard and Myers 2009), and most solar radiation networks use one of a few pyranometers on the market. Accuracy varies among types and models of radiometers, with different instruments having different sensitivities in accuracy to changes in solar and atmospheric radiation, spectral radiation distribution, incidence angle of the incoming radiation (e.g., pyranometer cosine response), thermal offset [a difference in temperature between detector and dome(s)] and temperature (Habte et al. 2016). The factory calibration errors of  $\pm 5\%$  for GHI reported for most pyranometers are generally supported by independent evaluations (Cronin and McPhaden 1997; Stoffel et al. 2000; Gueymard and Myers 2009; Habte et al. 2015) but errors vary with environmental conditions. For example, Gueymard and Myers (2009) found errors of only  $\pm 2\%$  averaged over a year, yet mean monthly errors were as large as  $-8\%$  in winter. Habte et al. (2015) found errors of  $\pm 5\%$  at zenith angles  $< 60^\circ$  but error increased considerably for some instruments (up to  $17\%$ ) at large zenith angles ( $70^\circ$ – $80^\circ$ ).

The measurement errors given above are for carefully maintained instruments, while in practice instruments can degrade for years before being cleaned, recalibrated, repaired or replaced (e.g., Slater 2016) such that systematic error is nonstationary and often exceeds factory specifications. It is a major undertaking to identify nonstationary error in thousands of instruments to filter out bad data and adjust biases in salvageable measurements in order to homogenize observations prior to interpolating to a grid.

The lack of a high-resolution gridded observation-based solar irradiance dataset remains a data gap for evaluating climate models and reanalysis, for driving land surface models, and for general understanding of spatiotemporal variability in solar radiation across CONUS. In response, this paper presents a method for generating a gridded dataset of global irradiance over CONUS using primarily ground-based solar radiation measurements, combined with cloud-cover observations and modeling of clear-sky irradiance. The mapping borrows from techniques in Daly et al. (1994, 2008, 2015, 2021) with modifications particular to solar radiation data. This paper also describes our first product: a 30-yr climatology (1991–2020) of daily global irradiance averaged by calendar month at a resolution of 30 arc s ( $\sim 800$  m). We compare this new “PRISM” climatological dataset with other datasets that are based on ground-based observations (WorldClim), ground-based observations of covarying environmental variables (Daymet), and reanalysis (ERA5-Land, MERRA-2, and NLDAS). Our comparisons are largely qualitative and meant to highlight unique features in the PRISM dataset, leaving a more comprehensive and quantitative evaluation of solar irradiance datasets (Slater 2016) for further study.

## 2. Data

### a. Station data

Observations of GHI were acquired from station networks that had data at any time during the years 1961 through 2020 (see Table S1 in the online supplemental material). These

data came in time steps ranging from 5 min to hourly. Sub-hourly data were first averaged to hourly. For consistency with other PRISM datasets (Daly et al. 2008, 2015), hourly data were then aggregated to create daily values ( $\text{MJ m}^{-2} \text{day}^{-1}$ ) corresponding to 1200:00–1159:59 coordinated universal time (UTC). Effective daily cloud transmittance  $T_c$  was calculated as daily GHI ( $S_{\text{surf}}$ ) divided by daily clear-sky GHI ( $S_{\text{clear\_sky}}$ ) after applying quality control criteria and an algorithm to reduce measurement error in  $S_{\text{surf}}$  [see section 3a(1)]. For each day of the year, we used a climatological average for  $S_{\text{clear\_sky}}$  [see section 3b(1)], which meant that variability in  $T_c$  was not only due to cloud cover variability but also partly to anomalies in water vapor and aerosol concentrations. Still, we refer to  $T_c$  as “cloud transmittance” for brevity.

Although not used for making the gridded solar radiation datasets, daily GHI data from the National Renewable Energy Laboratory (NREL) database were acquired for 16 locations (Table 1, along with Table S2 in the online supplemental material). These station data were used to examine the accuracy of the gridded solar radiation data, assuming the NREL instrumentation were among the more carefully maintained and calibrated across networks (Gueymard 2012).

We also obtained modeled hourly solar radiation at over 1300 locations in CONUS from the National Solar Radiation Database 1961–90 (Maxwell et al. 1995) and 1991–2005 (Wilcox 2007). NSRDB estimates GHI from other environmental variables using models that are both empirically and physically based. Where GHI was available from more than one model, we used the value with the lowest assigned error estimate. We treated NSRDB modeled station data the same as we treated observed data, though additional adjustments were made to account for detected regional biases [see section 3b(2)]. NSRDB modeled data were not used where irradiance observations from a collocated station were available for estimating the climatology [see section 3a(3)].

We additionally estimated station global irradiance from cloud height and sky coverage measured by ceilometers available from Aviation Routine Weather Report (METAR) records taken at airports using the National Weather Service’s automated observation systems, which include the Automated Surface Observation System (ASOS), Automated Weather Observation System (AWOS), and Automated Meteorological Observation Systems. (For brevity, we use “ASOS” to include all airports generating METAR records, regardless of the particular system.) It is worth noting that ceilometers measure cloud properties directly overhead via a narrow beam of light, so do not scan the entire sky. We used ASOS data beginning 1 January 1996 because by that time most U.S. airports had transitioned from visual observation to automated systems; the method we used for estimating  $T_c$  was calibrated to these automated observations (Belcher and DeGaetano 2007). Hourly and subhourly values of  $T_c$  were calculated using METAR cloud base height and coverage following Belcher and DeGaetano (2007), though we extended their method to accommodate cloud bases  $> 3659$  m (see section S1 in the online supplemental material). Daily  $T_c$  was calculated as a weighted average of the subdaily values of cloud transmittance as

TABLE 1. NREL solar radiation measuring stations with at least 5 years of valid observations per calendar month. Here and in Table 3, below, ID indicates identifier.

Station ID	Station name	Lon (°E)	Lat (°N)	Elev (m)
BMS	NREL Solar Radiation Research Laboratory [Baseline Measurement System (BMS)], CO	-105.179	39.742	1829
BS	Bluefield State College, WV	-81.240	37.265	803
EC	Elizabeth City State University, NC	-76.216	36.282	26
HSU	California State Polytechnic University, Humboldt [Solar Radiation Monitoring Station (SoRMS)], CA	-124.080	40.876	36
IRRSP	NREL Solar Radiation Research Laboratory [Irradiance, Inc., rotating shadowband pyranometer (RSP), version 2], CO	-105.179	39.742	1829
LMU	SOLRMAP Loyola Marymount University [rotating shadowband radiometer (RSR)], CA	-118.423	33.967	27
LRSS	Lowry Range Solar Station (RSR), CO	-104.580	39.607	1860
NPC	Nevada Power Clark Station, NV	-115.052	36.086	523
NWTC	NREL Flatirons Campus (M2), CO	-105.235	39.911	1855
SMUDA	Sacramento Municipal Utility District (Anatolia), CA	-121.240	38.546	51
SPMD	South Park Mountain Data, CO	-105.625	39.273	2944
STAC	Solar Technology Acceleration Center (SolarTAC), CO	-104.620	39.757	1674
UAT	SOLRMAP University of Arizona [Observed Atmospheric and Solar Information System (OASIS)], AZ	-110.955	32.230	786
UNLV	University of Nevada, Las Vegas, NV	-115.143	36.107	615
UOSRML	University of Oregon [Solar Radiation Monitoring Laboratory (SRML)], OR	-123.074	44.047	133.8
VTIF	NREL Vehicle Testing and Integration Facility RSR, CO	-105.176	39.742	1793

$$T_c = \sum_{i=1}^N T_{c,i} \cos Z_i \Big/ \sum_{i=1}^N \cos Z_i, \quad (1)$$

where  $i$  is the  $i$ th of  $N$  observations in a day (using same definition of a “day” as above), and  $Z$  is the solar zenith angle. Detection of regionally systematic biases in ASOS  $T_c$  prompted an additional adjustment to reduce biases [see section 3b(2)]. Last,  $S_{\text{surf}}$  was calculated by multiplying  $T_c$  by a modeled  $S_{\text{clear\_sky}}$ :

$$S_{\text{surf}} = T_c S_{\text{clear\_sky}}. \quad (2)$$

Section 3b(1) describes the modeling of  $S_{\text{clear\_sky}}$ .

### b. MERRA-2 reanalysis

MERRA-2 surface elevation and hourly surface shortwave flux data at 0.625° longitude by 0.5° latitude resolution were used to estimate clear-sky atmospheric extinction parameters for modeling clear-sky GHI climatology at higher (30 arc s) horizontal resolution over CONUS [see section 3b(1)]. Hourly values of surface albedo  $\alpha_s$ , top-of-atmosphere incoming shortwave flux  $S_{\text{toa}}$ ,  $S_{\text{clear\_sky}}$ , and  $S_{\text{surf}}$  were averaged to daily values (1200–1200 UTC). Effective daily clear-sky transmittance  $T_{\text{clear\_sky}}$  was calculated as  $S_{\text{clear\_sky}}/S_{\text{toa}}$  using the daily values for each variable.

### c. Other gridded solar radiation data

Gridded global irradiance data from ERA5, NLDAS (specifically, the NLDAS-2 forcing data), WorldClim, and Daymet were acquired for comparison with our product. Thirty-year climatologies of monthly values were computed from ERA5 (1991–2020), NLDAS (1991–2020), and Daymet (1990–2019;

2020 was not available at the time of this analysis). WorldClim was only available as the climatology of monthly values for the period 1970–2000.

## 3. Methods

The generation of the mean monthly gridded global irradiance datasets required multiple steps, including station data quality control, bias reduction, spatial interpolation, and solar radiation modeling. This section describes these steps, which are summarized in a workflow diagram in Fig. 2. In the interest of length, some details have been made available in the online supplemental materials, which also includes a more detailed workflow diagram (Figs. S1 and S2 in the online supplemental material).

### a. Station data quality control and bias reduction

#### 1) SOLAR RADIATION OBSERVATIONS

Previous studies have used various quality-control (QC) methods for ground-based solar radiation measurements (e.g., Younes et al. 2005; Shi et al. 2008; Journée and Bertrand 2011; Longman et al. 2013; Slater 2016). As Slater (2016) points out, however, many methods are for hourly data, require both the direct and diffuse components of radiation, use a lengthy (many years) time series to identify outliers, or are limited to identifying erroneous data. Like some studies (e.g., Longman et al. 2013; Slater 2016), we wanted to not only identify erroneous data but also “salvage” data that appeared to be erroneous in an absolute sense but may be acceptable in a relative sense and therefore amenable to relatively simple bias correction.

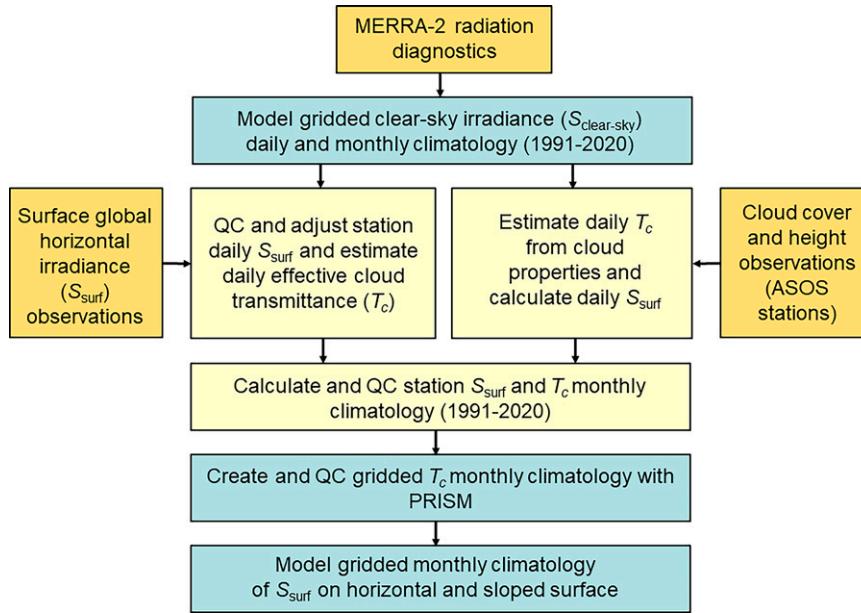


FIG. 2. Workflow diagram for creating the gridded global irradiance dataset. Orange boxes indicate data sources, yellow boxes indicate station data processing, and blue boxes indicate gridded data processing.

Similar to others (e.g., Riihimatki et al. 2006; Slater 2016) our bias-correction method rescales observations based on relative difference between the off-calibration clear-sky values and assumed “true” clear-sky values. Following Slater (2016), we defined a clear-sky ratio (CSR):

$$CSR = \frac{E_{clear\_sky}}{S_{clear\_sky}} \quad (3)$$

where  $E_{clear\_sky}$  is an upper-bound envelope of a measured, and possibly biased, daily value and  $S_{clear\_sky}$  is the “true” daily clear-sky GHI. The upper-bound envelope is estimated from observations over a predetermined time interval (183 or 91 days). For each station,  $S_{clear\_sky}$  came from gridded daily clear-sky climatological values simulated at a 30-arc-s resolution [see section 3b(1)] with values from the nearest grid point assigned to each station. After determining CSR, a rescaled, or “bias corrected,” global radiation  $S_{surf}^*$  was calculated as

$$S_{surf}^* = \frac{S_{surf}}{CSR} \quad (4)$$

for each measured  $S_{surf}$ . The adjustment in Eq. (4) assumes that clear- and cloudy-sky error can be equally rescaled.

The key to successfully applying Eqs. (3) and (4) is defining  $E_{clear\_sky}$ . We found that the algorithm in Slater (2016) for defining  $E_{clear\_sky}$  did not always produce desirable results, especially when the CSR was relatively far from 1. Though Slater (2016) limited rescaling with values of CSR in the range of 0.95–1.05, we wanted to salvage a larger proportion of data. We also wanted an algorithm that struck a balance between robustness and simplicity and did not require coding many “special” cases. Our method applies Eq. (3) twice, once after

each pass through the daily time series with sliding windows of 183 then 91 days. Data that do not meet quality control criteria are flagged and excluded from further analysis. See section S2 in the online supplemental material for a complete description of the method.

There are too many different ways in which bad data can present themselves in a time series to discuss here. We illustrate the results of our adjustment procedure on just two examples of common situations. In the first example, measurements at a site in Tucson, Arizona (USCRN site 53131), degraded over a period of a few years (2014–18) until observed values systematically increased in the spring of 2018 due presumably to maintenance of the instrumentation (Fig. 3a). The adjustment procedure rescales all the observations so that the upper envelope of data approaches the expected clear-sky GHI (Fig. 3b) but flags the first few months of 2018 as bad because the method does not identify the precise date of maintenance. In the second example, winter observations at a site in Bend, Oregon [Pacific Northwest Cooperative Agricultural Weather Network (AgriMet) site BEWO], never reached the expected clear-sky values, and the negative bias progressively worsened over a period of at least 15 years (Fig. 3c). A site visit by the authors in October 2019 revealed trees as obstacles to direct radiation even near midday, as well as shadow from nearby buildings at low sun angles. Progressively degrading measurements over many winters (when sun angle is low) is consistent with an increase in shade resulting from upward-growing vegetation. Our rescaling does not fully compensate for this seasonally varying bias and part of winter and spring are flagged as bad every year beginning winter 2007–08. The data quality worsens sufficiently over time such that entire years are flagged as bad (Fig. 3d).

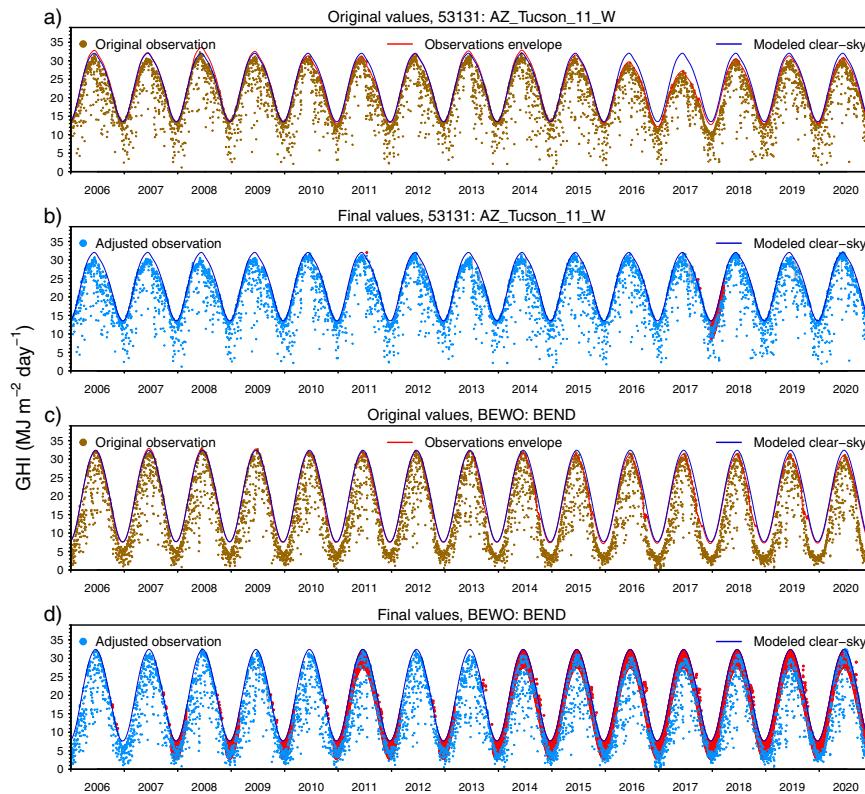


FIG. 3. Examples of (a) original daily observed GHI ( $S_{\text{surf}}$ ) with modeled clear-sky GHI ( $S_{\text{clear\_sky}}$ ) and first pass at fitting envelope of observed clear-sky days ( $E_{\text{clear\_sky}}$ ) and (b) adjusted data after second pass at USCRN station 53131 (Tucson, Arizona; 32.24°N, 111.17°W). (c),(d) As in (a),(b), respectively, for AgriMet station BEWO (Bend, Oregon; 44.05°N, 121.32°W). Data identified as unreliable are marked in red in (b) and (d).

For all stations in all networks, our adjustment procedure rejected 17% of the daily data that passed initial daily completeness and quality criteria. Of the daily data that were accepted, 41% of values were adjusted by an amount less than, or equal to,  $\pm 5\%$ , while 23% of the values were adjusted by more than  $\pm 10\%$ .

## 2) ASOS CLOUD TRANSMITTANCE

Belcher and DeGaetano (2007) developed their algorithm for estimating  $T_c$  from ASOS METAR when few ASOS ceilometers detected clouds above 3840 m (and actually reported cloud height no higher than 3659 m, or 12 000 ft). Their algorithm had to compensate for this instrument limitation, which caused days with clouds above 3659 m to be reported as clear. Because many airports now have ceilometers that detect clouds above 3840 m, we reevaluated Belcher and DeGaetano's (2007) algorithm using a larger number of records both in terms of record length and number of stations, including stations that do not ("type I" stations), and do ("type II" stations), regularly report clouds above 3659 m.

Like Belcher and DeGaetano (2007), we paired ASOS stations with proximal stations from other networks measuring solar radiation. For brevity we refer to stations with solar

irradiance measurements as "solar" stations. Restricting the sample to solar stations within 20 km of an ASOS station, requiring concurrent records of at least 730 days, and applying other criteria (section S3 in the online supplemental material), resulted in 971 pairs with an average ASOS-solar station distance of 10.2 km and average of 3135 concurrent daily observations.

Inspection of the station pairings revealed systematic biases in ASOS  $T_c$  (see section S3 and Fig. S3 in the online supplemental material). The systematic biases in ASOS  $T_c$  motivated us to correct biases in ASOS  $T_c$  that were a function of ASOS  $T_c$  itself. These biases differed between stations that do not (type I), and do (type II), report clouds above 3659 m, so we applied a different bias-correction function depending on whether a station was classified as type I or II, and, if it was type II, whether it was reporting clouds above 3659 m on the day in question (Fig. S4 in the online supplemental material). No regional or seasonal variations in the bias-correction functions were made at this stage.

A consequence of the bias correction was a tendency to bring daily values toward the mean, so errors in low and high daily values of  $T_c$  often increased. Because our goal was to produce the best estimates of climatological monthly values, however, this shortcoming was not severe. We recommend

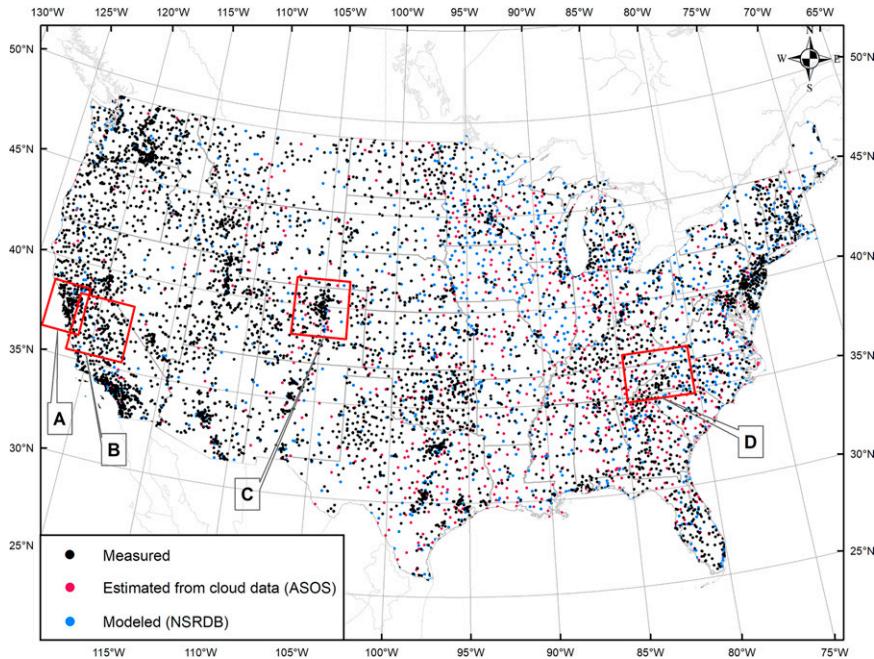


FIG. 4. Locations of stations with solar radiation data used to generate long-term (1991–2020) mean monthly global irradiance. Red-outlined boxes show areas highlighted in text and later in Fig. 8 (area “A”), Fig. 10 (“B”), Fig. 11 (“C”), and Fig. 12 (“D”).

applying other methods for reducing bias when the intent is to estimate daily time series of  $T_c$ .

### 3) STATION-BASED CLIMATOLOGY

The daily GHI values were averaged to create monthly means for each year of record. A maximum of two invalid (missing values were also considered invalid) daily values per month were allowed for a monthly value to be considered valid. Monthly mean clear-sky GHI was also calculated for each valid month of solar radiation. The monthly mean GHI was converted to  $T_c$  in preparation for mapping with PRISM by dividing by the clear-sky GHI.

Monthly average  $T_c$  were tested for spatial consistency using the Assay QC system, a version of PRISM that estimates station values in their absence and compares them with the observed values (Daly et al. 2008). Monthly values failing the Assay QC check (i.e., prediction differs from observation by more than 10%) were set to missing. The remaining monthly values were averaged over 1991–2020, if they had five or more years of data during this period, otherwise they were averaged over their historical period of record. A 1991–2020 monthly climatology estimated using data from at least 23 of the 30 years (75% data coverage) was considered sufficiently representative of the 1991–2020 period and was termed a “long term” station. However, monthly climatologies calculated from fewer than 23 years in 1991–2020, or calculated from data outside this period, were still considered for inclusion in order to increase station density. These were termed “short term” stations and were adjusted using nearby long-term

stations to reduce possible short-term deviations from the 1991–2020 mean (see section S4 in the online supplemental material).

### b. Mapping methods

#### 1) CLEAR-SKY SOLAR RADIATION

Daily clear-sky GHI climatology was modeled using the USDA Agricultural Research Service (ARS)–USGS, version 2.4.1, of the Image Processing Workbench (IPW; Frew 1990; Marks et al. 2018). IPW uses a two-stream approximation to the radiative transfer equation and simulates the effects of elevation, shading, and reflection from nearby terrain on irradiance on a horizontal or inclined surface (Dubayah et al. 1990; Dubayah 1994). We used recommended values of time-invariant atmospheric parameters in the model (Daly et al. 2007).

IPW requires a surface albedo  $\alpha$ , and clear-sky atmospheric vertical optical depth  $\tau_{\text{clear\_sky}}$  at mean sea level as input. A daily  $\alpha$ , climatology was derived directly from MERRA-2 as follows: We smoothed the time series of daily  $\alpha$ , at each MERRA-2 grid cell with a Gaussian filter (standard deviation = 7 days) and then calculated a 30-yr (1991–2020) average of smoothed daily  $\alpha$ , for each grid cell and each of 365 calendar days (29 February excluded).

To estimate  $\tau_{\text{clear\_sky}}$  consistent with IPW parameterizations but using MERRA-2 data, we used an inverse-modeling approach to map MERRA-2  $T_{\text{clear\_sky}}$  to IPW  $\tau_{\text{clear\_sky}}$ . First, we regridded MERRA-2 data from its native resolution ( $0.625^\circ$  longitude by  $0.5^\circ$  latitude) to  $0.5^\circ \times 0.5^\circ$  because IPW expects longitude and

latitude resolution to be identical. For each MERRA-2 grid cell and each day of one calendar year (2006), IPW was then used to calculate  $T_{\text{clear\_sky}}$  for 22 values of  $\tau_{\text{clear\_sky}}$  ranging from 0.01 to 1.6. We fitted the following empirical equation relating IPW  $T_{\text{clear\_sky}}$  to  $\tau_{\text{clear\_sky}}$ :

$$\ln T_{\text{clear\_sky}} = b_1 \tau_{\text{clear\_sky}}^{b_2} \quad (5)$$

Using nonlinear regression, the parameters  $b_1$  and  $b_2$  were estimated for each one of 365 days and for each MERRA-2 grid cell (see examples in Fig. S5 in the online supplemental material). To greatly reduce the total number of parameters, both  $b_1$  and  $b_2$  were subsequently modeled as a function of the day of year using an eighth-order polynomial equation with parameters estimated using standard linear regression (see examples in Fig. S6 in the online supplemental material).

A  $T_{\text{clear\_sky}}$  climatology was generated by smoothing the time series of daily MERRA-2  $T_{\text{clear\_sky}}$  at each grid cell with a Gaussian filter (standard deviation = 21 days) and then averaging the smoothed  $T_{\text{clear\_sky}}$  over 30 years (1991–2020) for each grid cell and each of 365 calendar days (leap day excluded). Climatological daily  $\tau_{\text{clear\_sky}}$  for each grid cell was calculated from the climatological daily MERRA-2  $T_{\text{clear\_sky}}$  by inverting Eq. (5). Both climatological  $\tau_{\text{clear\_sky}}$  and  $\alpha_s$  were regridded to the 30-arc-s resolution grid using a Gaussian filter.

Last, daily clear-sky GHI was simulated for one representative year (2006) using IPW with the 30-arc-s topography and 30-arc-s daily climatological values of  $\tau_{\text{clear\_sky}}$  and  $\alpha_s$ . The same clear-sky GHI values for the representative year were used for all other years.

## 2) EFFECTIVE CLOUD TRANSMITTANCE AND GLOBAL IRRADIANCE

Mean monthly  $T_c$  was interpolated to a regular grid at 30-arc-s resolution with the PRISM climate mapping system (Daly et al. 1994, 2002, 2003, 2008). For each grid cell, PRISM calculates a local regression function between a climate element and an explanatory grid such as a digital elevation model (DEM) or an existing climate grid. Nearby stations entering the regression are assigned weights based primarily on the physiographic similarity of the station to the grid cell. Physiographic factors relevant to this study are distance, elevation, coastal proximity, vertical atmospheric layer (boundary layer and free atmosphere), and topographic position (relative to surrounding terrain). We used a process called climatologically aided interpolation (CAI; Willmott and Robeson 1995) to perform the interpolation. CAI is effective at mapping climate variables for which there are relatively few stations, and for which there is an existing grid (called the predictor grid) that is spatially correlated on a local level with the interpolated element (Daly et al. 2012, 2015).

We considered three candidate climatological predictor grids for the interpolation of mean monthly  $T_c$ , all part of the PRISM suite of gridded monthly climatologies: mean daily temperature range (DTR), mean daily minimum relative humidity ( $\text{RH}_{\text{min}}$ ), and mean daily maximum vapor pressure

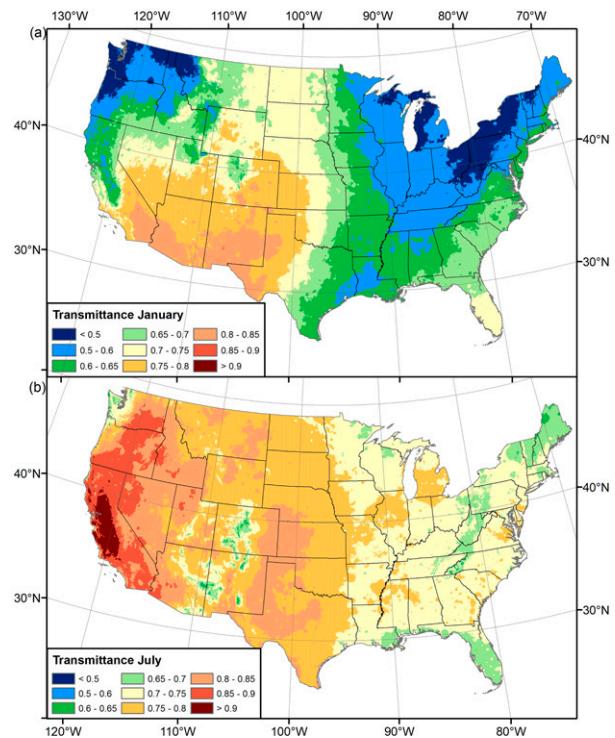


FIG. 5. Mean (a) January and (b) July effective cloud transmittance  $T_c$  from PRISM.

deficit ( $\text{VPD}_{\text{max}}$ ). DTR has been shown to be correlated with cloudiness; cloudy days tend to have depressed maximum temperature due to the attenuation of direct solar radiation during the daylight hours, and the presence of clouds in the morning hours limits surface cooling through upwelling longwave radiation, raising the minimum temperature (Thornton and Running 1999).  $\text{RH}_{\text{min}}$  has also been found to be associated with cloudiness; days with high afternoon RH values tend to occur during cloudy conditions, while those with low values are often cloud-free (Cenzig et al. 1981).  $\text{RH}_{\text{min}}$  was estimated with 1991–2020 grids of mean monthly maximum daily temperature and  $\text{VPD}_{\text{max}}$ , after Daly et al. (2015), their Eq. (6).  $\text{VPD}_{\text{max}}$ , as a standalone measure of daytime moisture deficit, was also considered as a possible predictor grid.

Of the three potential predictor grids considered,  $\text{RH}_{\text{min}}$  was found to be the most effective in the interpolation of  $T_c$ . DTR exhibited excessive fine scale variation in differing topographic positions (e.g., low DTR on ridgetops versus high DTR in valley bottoms) under relatively constant solar transmittance conditions, causing noise in the local relationships. The relationship between  $\text{VPD}_{\text{max}}$  and  $T_c$  was found to vary with temperature, resulting in cooler, high-elevation areas appearing cloudier than warmer, low-elevation areas, under relatively constant solar transmittance conditions. In contrast,  $\text{RH}_{\text{min}}$  varied in a relatively conservative fashion, showing little variation with elevation or topographic position under relatively constant solar transmittance conditions. In addition,  $\text{RH}_{\text{min}}$  was effective at delineating strong gradients in summertime transmittance caused by fog and

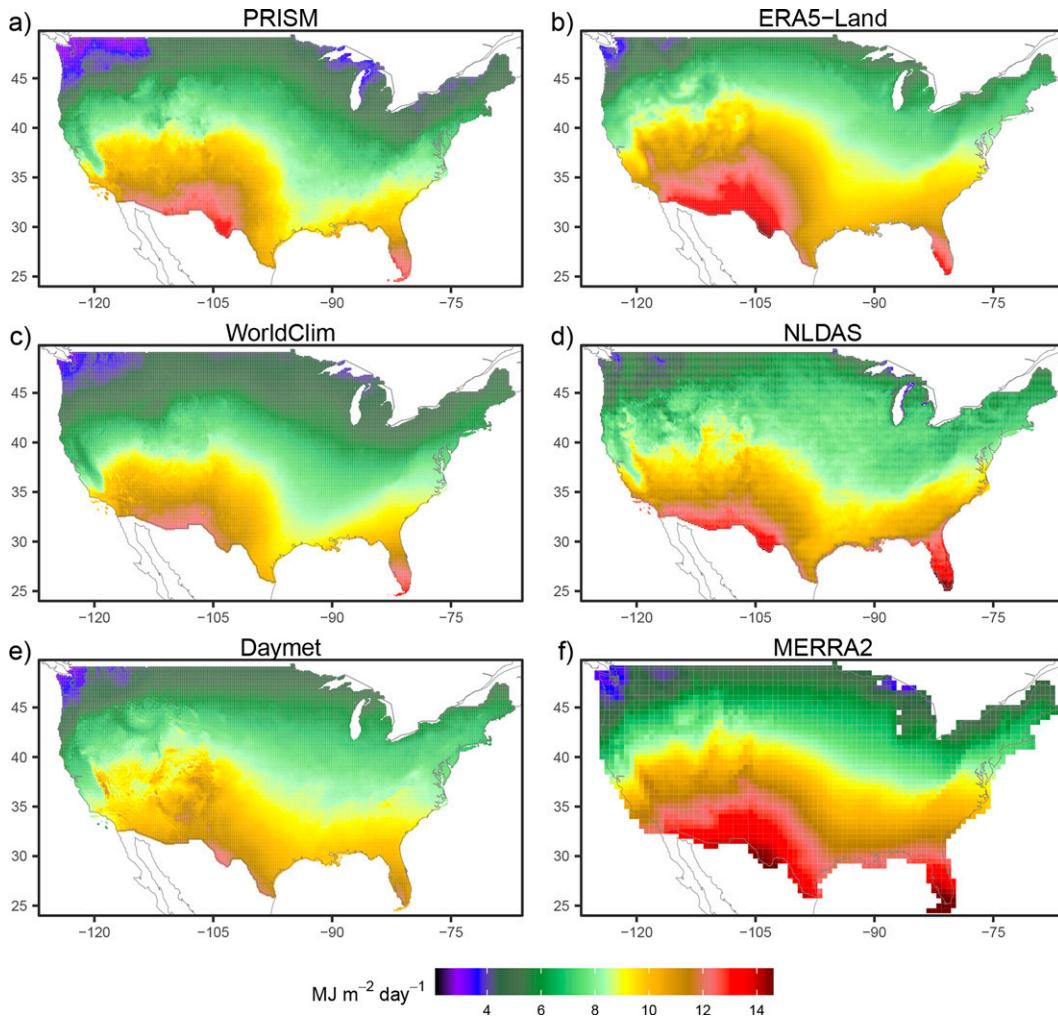


FIG. 6. Mean January GHI from (a) PRISM, (b) ERA5-Land, (c) WorldClim, (d) NLDAS, and (f) MERRA-2, and mean January global irradiance on a sloped surface from (e) Daymet. The effect of slope in Daymet is not visually discernable at the resolution shown here. Spatial resolution for PRISM, WorldClim, and Daymet was regridded to  $0.1^\circ \times 0.1^\circ$  to facilitate plotting. All other datasets are at their native resolution. Higher values of radiation in some panels exceed the upper limit of the color scale.

low stratus along the West Coast, and persistent winter fog in inland valleys such as the California’s Central Valley and Idaho’s Snake River Plain (see, e.g., maps of January and July  $\text{RH}_{\min}$  in Fig. S7 in the online supplemental material).

A preliminary inspection of mapped mean monthly  $T_c$  suggested there were some regionally varying differences in  $T_c$  between stations with measured solar radiation and stations with modeled solar radiation, namely ASOS and NSRDB. We confirmed this by making monthly grids of ASOS and NSRDB biases in mean monthly  $T_c$  following the mapping procedure described in section 3b(2) above but without using  $\text{RH}_{\min}$  as a predictor. We used bias estimates at those ASOS and NSRDB stations that had solar stations within a 20-km radius (see section S5 in the online supplemental material for details). Separate bias maps were made for ASOS and NSRDB (see, e.g., maps of January and July bias in Fig. S8 in

the online supplemental material). Mean monthly  $T_c$  values at all ASOS and NSRDB stations were adjusted by subtracting the mapped biases at the station locations from the original station values.

Revised maps of mean monthly  $T_c$  were made that incorporated the bias-adjusted ASOS and NSRDB values. The maps were visually inspected for patterns (e.g., bullseyes) that appeared erroneous. Stations whose values differed greatly from our expectations based on surrounding stations and our knowledge of local climate geography were then flagged and excluded from a final round of mapping. The majority of the excluded stations were from the ASOS and NSRDB networks. Figure 4 shows the locations of the stations used in the final mapping.

While ASOS and NSRDB stations were treated the same as solar observation stations in the above mapping process, it

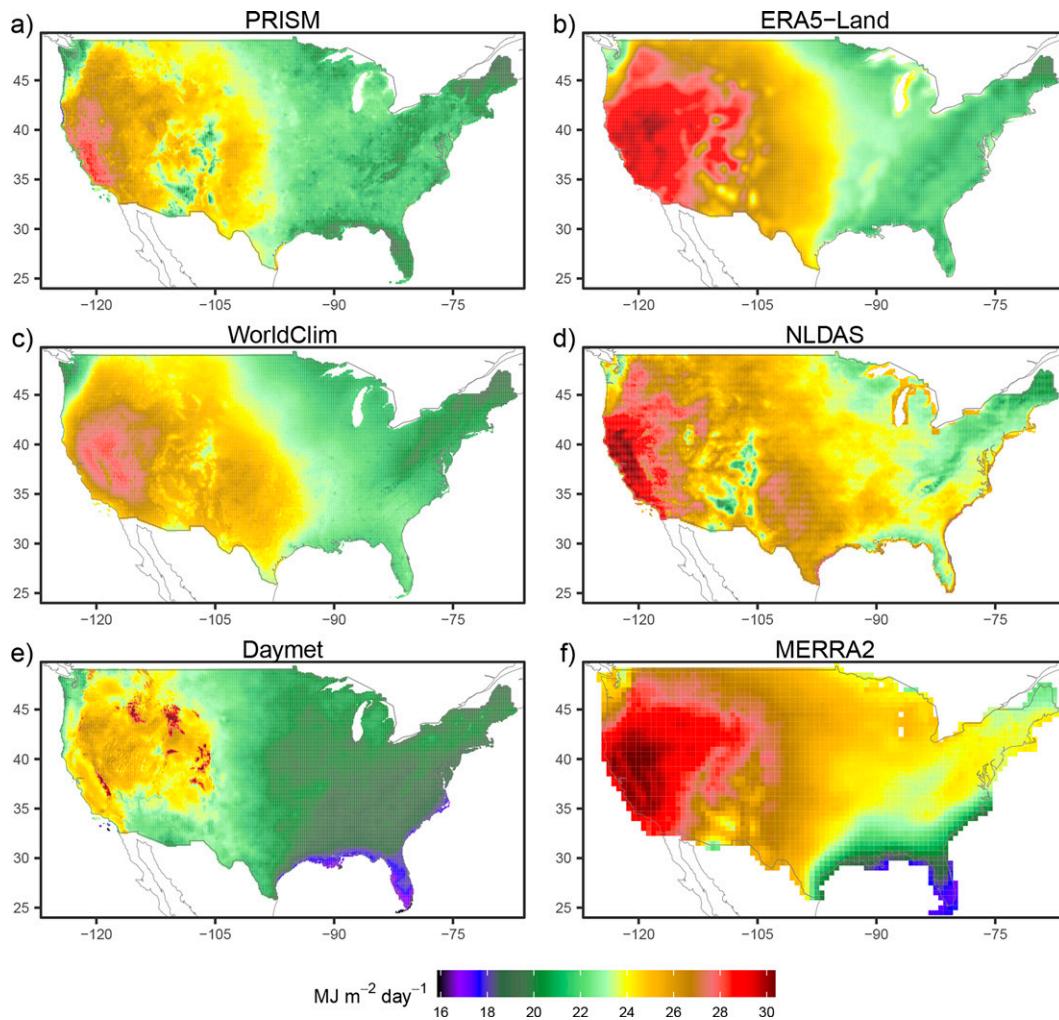


FIG. 7. As in Fig. 6, but for July.

is worth noting that many ASOS and NSRDB stations are actually the same station. Where records of two collocated stations overlap in time, records for the two stations were not merged into single time series. Instead, the collocated stations were treated as distinct stations while calculating normals and the mapping process effectively averaged the ASOS and NSRDB values where they were collocated.

As the last step, we ran IPW again for each day in one calendar year (2006) using the gridded climatological monthly  $T_c$  from above as input. Model results consisted of daily global irradiance on both horizontal and sloped ground surfaces, where the slope and aspect of the surface are calculated at the 30-arc-s resolution. Daily values were aggregated to monthly means as the final output.

#### 4. Results and discussion

Our primary results are the PRISM 30-arc-s gridded datasets of 1991–2020 climatological monthly global irradiance on horizontal and sloped ground surfaces. Using the PRISM

data, we briefly discuss the climatology of global irradiance over CONUS and make comparisons of PRISM with some commonly used gridded solar radiation datasets. We also provide four illustrative examples of where differences among datasets are notable and where PRISM brings new information: the central coast and southern Sierras of California, the Rocky Mountains of Colorado, and the Appalachian Mountains of North Carolina.

##### a. CONUS seasonal climatology

Once sun angle and elevation are accounted for, which together determine the thickness of the atmosphere normal to the sun's rays, the spatial pattern of global irradiance across CONUS is largely driven by variability in cloud cover, whose direct effect here is parameterized by the effective cloud transmittance  $T_c$ . Aerosols and water vapor also play roles, although in our method their impact is subsumed in the clear-sky irradiance (see, e.g., maps for January and July clear-sky GHI in Fig. S9 in the online supplemental material).

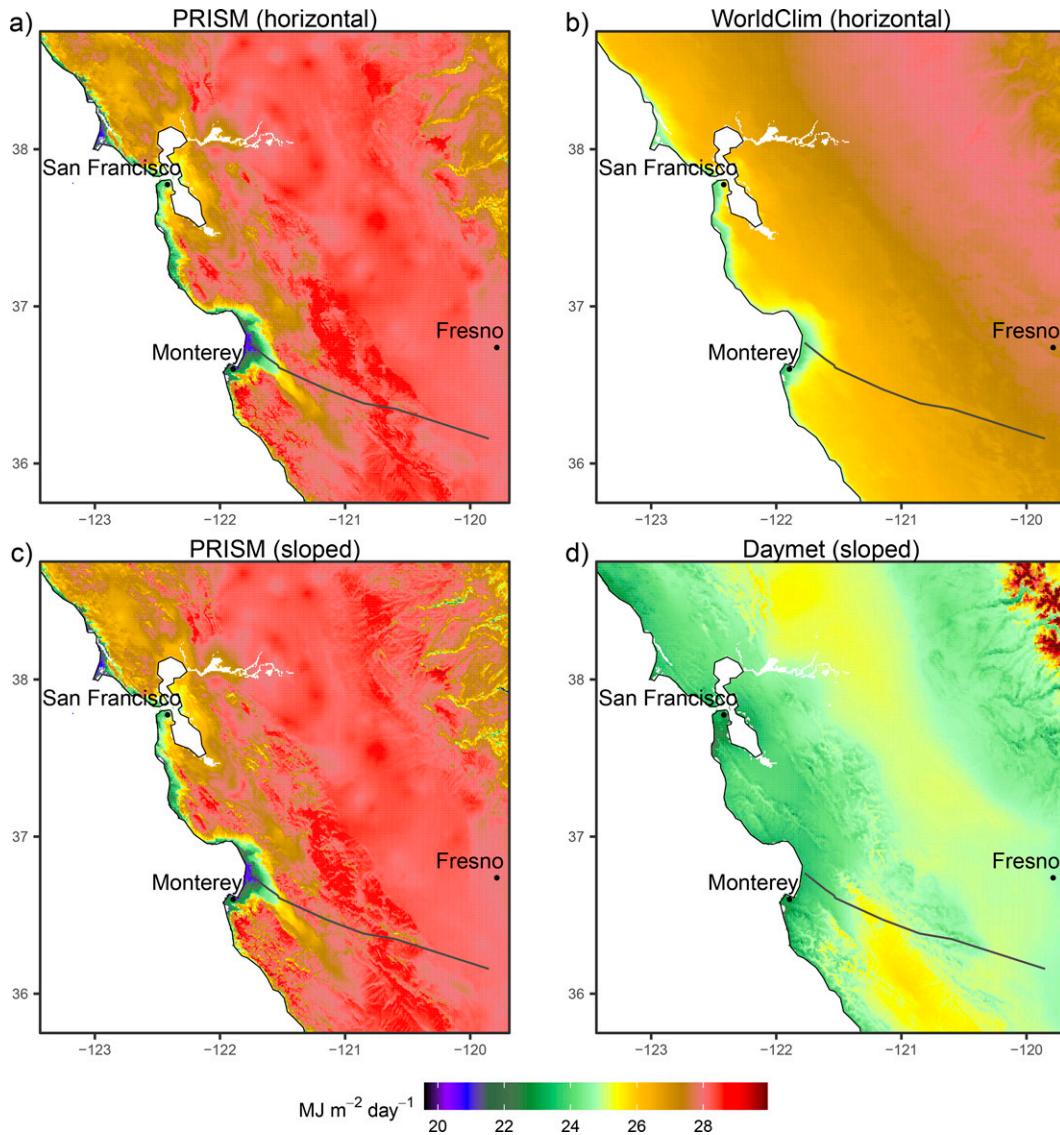


FIG. 8. Mean July GHI from (a) PRISM and (b) WorldClim, and global irradiance on a sloped surface from (c) PRISM and (d) Daymet, highlighting the effect of low clouds and fog along the coast of central California. The dark-gray line shows the location of transect 1 used in Fig. 9a, below. The mapped region corresponds to area “A” in Fig. 4.

Patterns of climatological  $T_c$  from PRISM show strong seasonal variations (Fig. 5). In winter,  $T_c$  is lowest in the Pacific Northwest, especially across the Cascades and Northern Rocky Mountains and in the upper Midwest, with local minima in regions downwind of the Great Lakes and west of the Appalachians (See Fig. 5a for January). Local minima are also seen in western valleys that experience persistent inversions, such as the Central Valley of California. Large scale patterns of winter  $T_c$  generally follow those of precipitation, with maxima in the dry southwestern United States, decreasing as one moves into the wetter eastern United States. In summer, climatological drought is reflected in very high  $T_c$  in the western United States (Fig. 5b), except for immediate coastal areas subject to frequent marine layer intrusions (see section 4b below). Also

evident is increased cloudiness in mountain areas of the southwestern United States during the North American monsoon, which is typically at its height in July and August.  $T_c$  is generally lower in the eastern United States, due to cloudiness associated with frequent convective showers and thunderstorms. Minima are seen in the southern Appalachians, which receive substantial moisture from the Gulf of Mexico in summer.

As would be expected, global irradiance across CONUS exhibits strong seasonal variations, reaching a minimum in winter and a maximum in summer (Figs. 6 and 7 for January and July, respectively). Spatial patterns of global irradiance also vary seasonally. In winter, a strong north–south latitudinal gradient, controlled by changes in sun angle and day length, is modulated by patterns of  $T_c$  described earlier. Here we

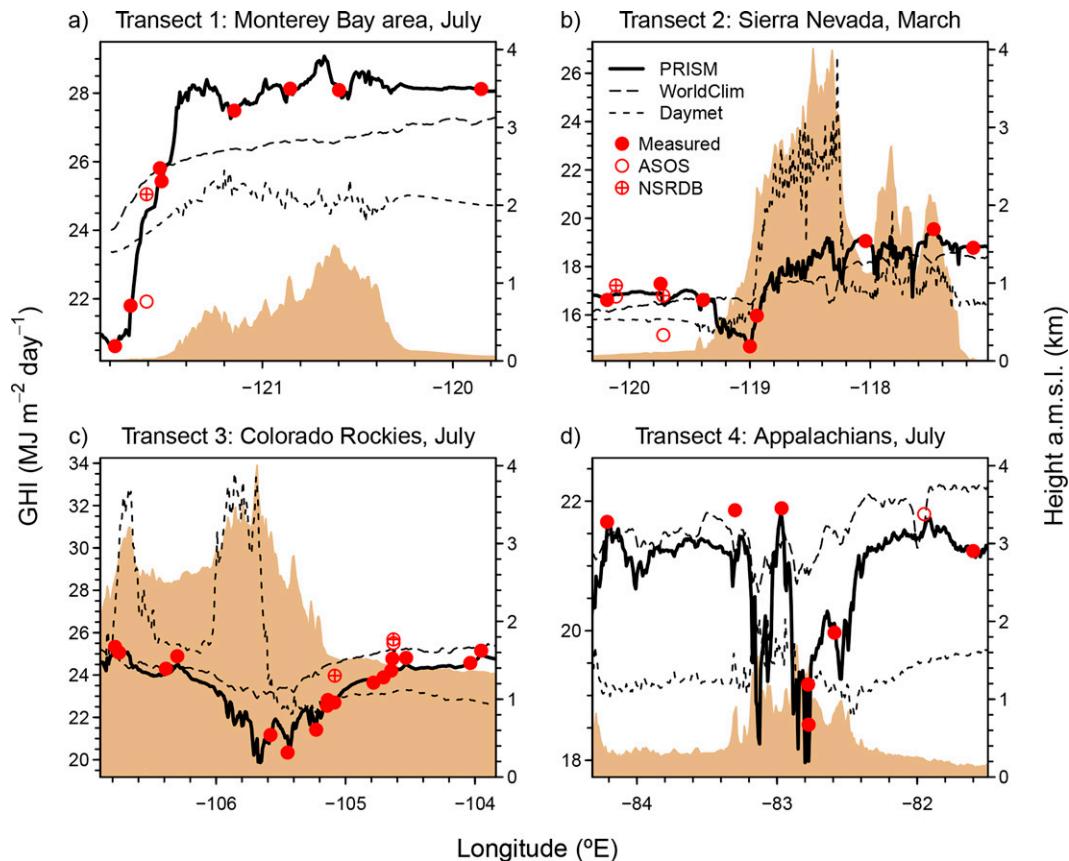


FIG. 9. Mean monthly GHI at stations (symbols) and from PRISM, WorldClim, and Daymet (solid and dashed lines) along (a) transect 1 in July, (b) transect 2 in March, (c) transect 3 in July, and (d) transect 4 in July. Transect locations are shown in Figs. 8, 10, 11, and 12, respectively. Light-brown shading shows surface elevation. Station values were adjusted for biases as described in section 3. Daymet values were converted from a sloped surface to a horizontal surface using the ratio of the horizontal to sloped surface radiation from the PRISM normals at the same grid resolution. The conversion only had a very minor effect relative to the full range of Daymet values shown.

compare global irradiance products from PRISM and WorldClim, which incorporate surface observations directly; ERA5-Land, NLDAS, and MERRA-2, which are derived from modeling and remote sensing products; and Daymet, which derives total atmospheric transmittance from daily temperature range and precipitation. In winter, all exhibit similar overall patterns, but the magnitudes of the gradients differ somewhat. MERRA-2, NLDAS, and ERA5-Land show slightly higher values than the others in the southern tier of states. All capture the irradiance minimum in the Pacific Northwest to some extent, with PRISM and WorldClim exhibiting deeper and more extensive minima than the others. The intensity and southern extent of the “trough” of lower irradiance in the Midwest varies from product to product; MERRA-2, ERA5-Land, and Daymet show a relatively limited southern extent of this trough in comparison with PRISM and WorldClim.

Differences in global irradiance among products are more dramatic in summer (Fig. 7). The north–south gradient is much reduced at this time of year, leaving  $T_c$  as the primary controlling influence on patterns of global irradiance. PRISM

and WorldClim are the most similar overall, albeit with PRISM showing greater spatial detail and a greater range of values. Both show the western United States as exhibiting relatively high irradiance values except for the Pacific Northwest and southwestern mountains, and the eastern United States as somewhat darker due to higher rainfall at this time of year. ERA5-Land also shares this pattern, with less detail due to limited spatial resolution. NLDAS shows a relatively bright eastern United States, but the southwest mountains are highly accentuated and have the lowest values in the CONUS. MERRA-2 has extremely high irradiance loadings in the western United States, and the lowest values are focused on the southeastern United States along the Gulf Coast. Daymet shows even lower values in the southeast and extends low values across the entire eastern United States. Daymet also has extremely high values over the higher terrain in the western United States. The July spatial patterns of NLDAS and Daymet (Figs. 7d,e) are qualitatively similar to summer (June–August) patterns reported in Slater (2016), though Slater (2016) used an earlier version of Daymet (the exact version number was not given).

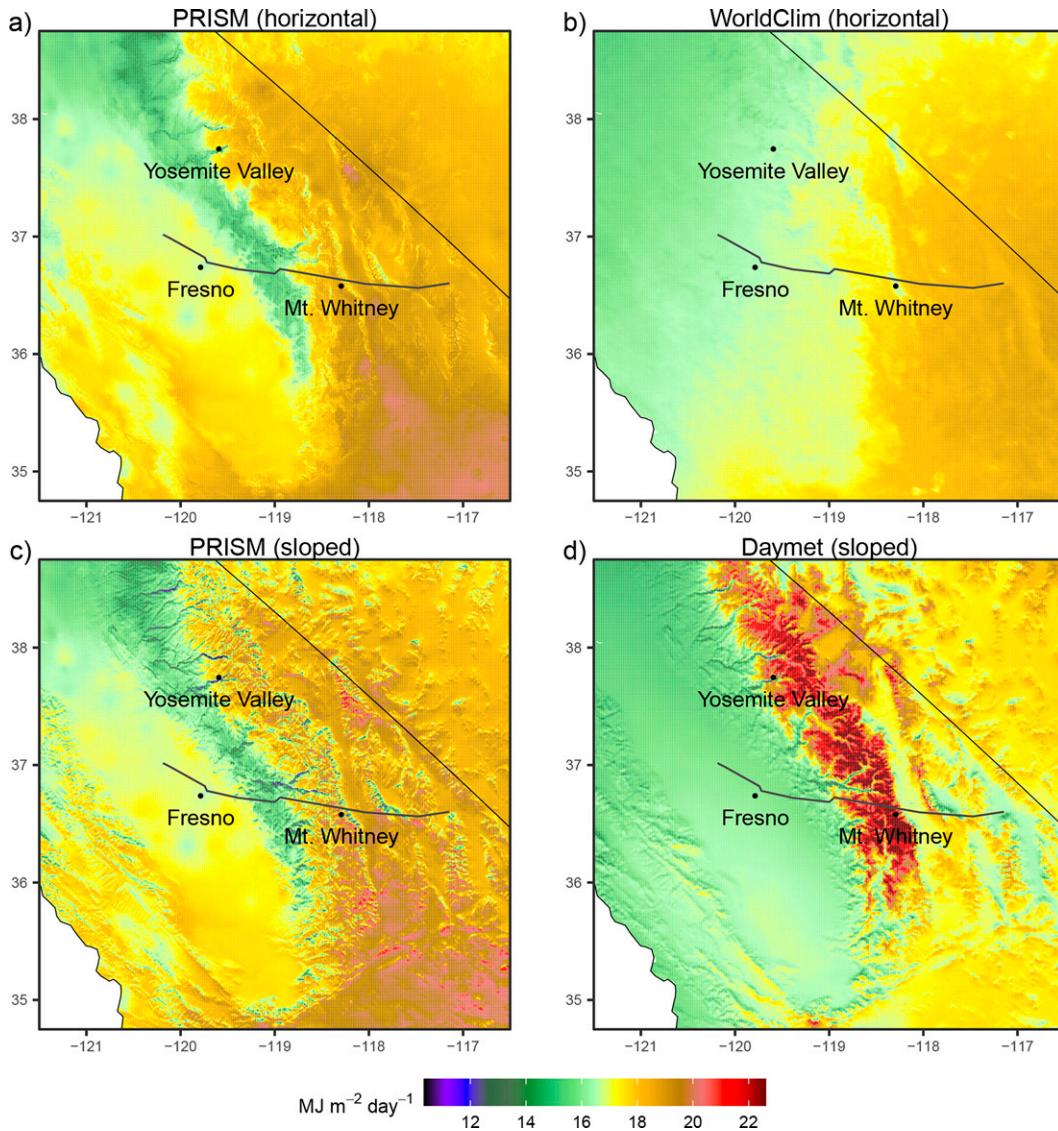


FIG. 10. As in Fig. 8, but for March and highlighting the effect of cloud cover along the western slopes of the Sierra Nevada in southern California. Higher values of radiation from Daymet exceed the upper limit of the color scale. The dark-gray line shows the location of transect 2 used in Fig. 9b. The mapped region corresponds to area “B” in Fig. 4.

*b. Central California coast*

PRISM clearly shows the effect of clouds and fog on global irradiance in July around San Francisco Bay and Monterey Bay and along the coastline (Fig. 8), an effect too spatially fine to be reproduced by the coarser reanalysis products (Fig. 7). WorldClim and Daymet show much weaker gradients than PRISM in global irradiance from the coast to higher inland elevations above the marine layer (Fig. 9a). For example, over a distance of about 10 km from the city of Monterey, California, to the nearby hills of the Santa Lucia Range to the south, global irradiance increases from about 20 to 30  $\text{MJ m}^{-2} \text{ day}^{-1}$  in PRISM, but only from 24 to 25  $\text{MJ m}^{-2} \text{ day}^{-1}$  in WorldClim and from 23 to 24  $\text{MJ m}^{-2} \text{ day}^{-1}$  in Daymet.

*c. Southern Sierra Nevada*

During the months of February to April, PRISM shows a band of diminished global irradiance along the western foothills of the southern Sierra Nevada Range, bounded by higher radiation in the central Valley to the west and higher radiation in the high Sierra Nevada and their leeward side to the east (see Fig. 10 for March; February and April are not shown). This darker band results presumably from clouds that form along the windward slopes of the Sierra Nevada. The band is absent from both WorldClim and Daymet in March (Fig. 9b). WorldClim simply shows a generally monotonic increase in global irradiance from the coast to the east of the Sierra Nevada. Daymet shows a large

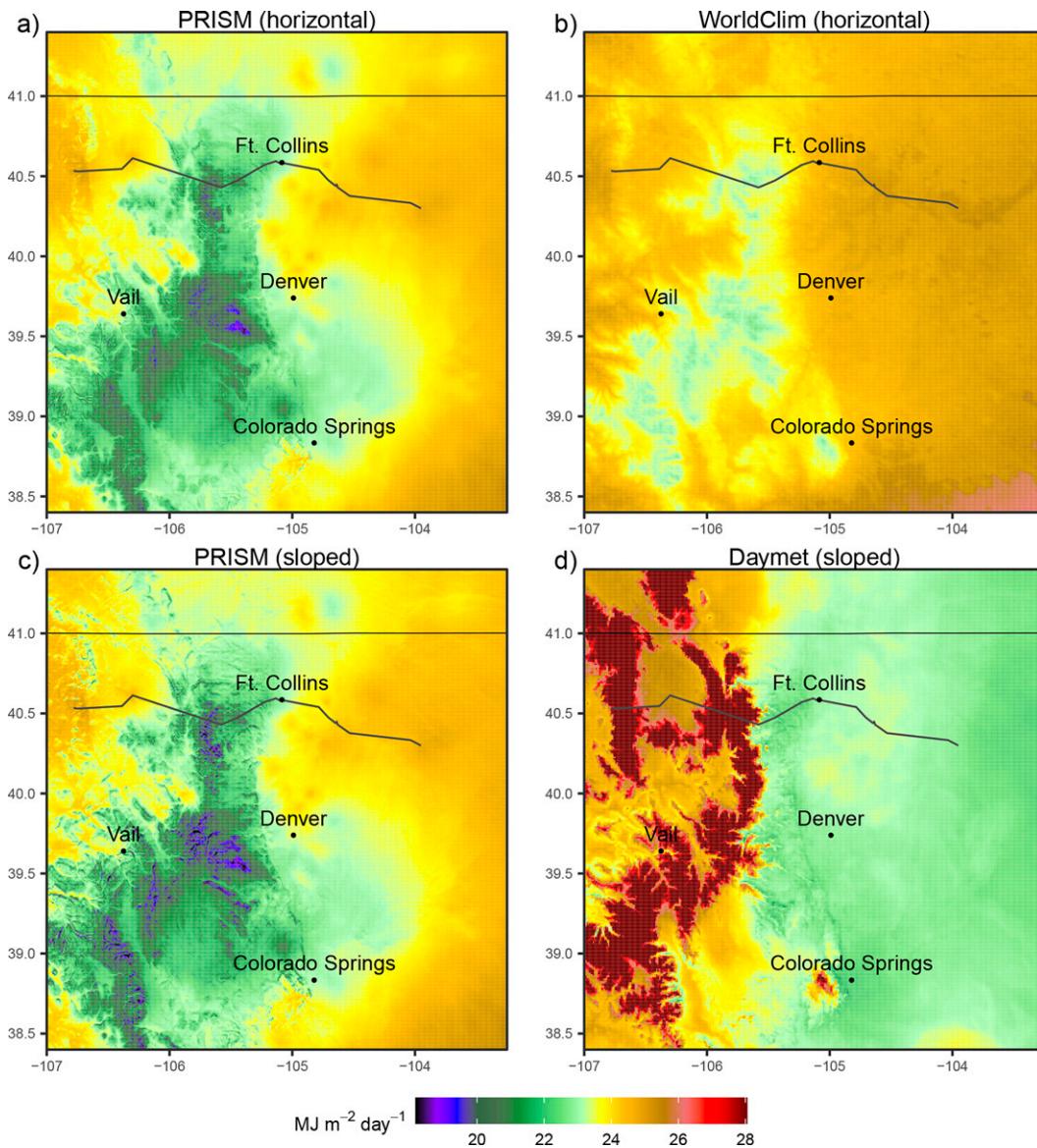


FIG. 11. As in Fig. 8, but for July and highlighting the effect of monsoonal moisture over the Rocky Mountains of central Colorado. The dark-gray line shows the location of transect 3 used in Fig. 9c. Higher values of radiation from Daymet exceed the upper limit of the color scale. The mapped region corresponds to area “C” in Fig. 4.

increase in solar radiation from the Central Valley to the peaks of Sierra Nevada, with values at the highest elevations approximately 20% higher than those in PRISM at the same locations. The darker band is evident in NLDAS and ERA5-Land (not shown), albeit with lower granularity.

For this region, Lapo et al. (2017) concluded that MTCLIM, the algorithm used in Daymet, provided the best of four methods they examined for estimating global irradiance. However, we find notable biases in Daymet (Figs. 9b and 10). In all but winter months, Daymet overestimates global irradiance at high elevations but underestimates it in the Central Valley. Our results are more consistent with Slater (2016), although note that all three studies used a different

version of Daymet (or MTCLIM) and applied different QC procedures.

#### d. Colorado Rockies

PRISM shows a strong gradient in global irradiance between the Rocky Mountains and the surrounding plateaus and western Great Plains in Colorado during the monsoonal months of July and August (see Figs. 9c and 11 for July). WorldClim shows a similar spatial pattern, although with higher global irradiance values in the mountains (~23 MJ m<sup>-2</sup> day<sup>-1</sup>) relative to PRISM (19–21 MJ m<sup>-2</sup> day<sup>-1</sup>). Daymet, curiously, shows an opposite pattern: very high global

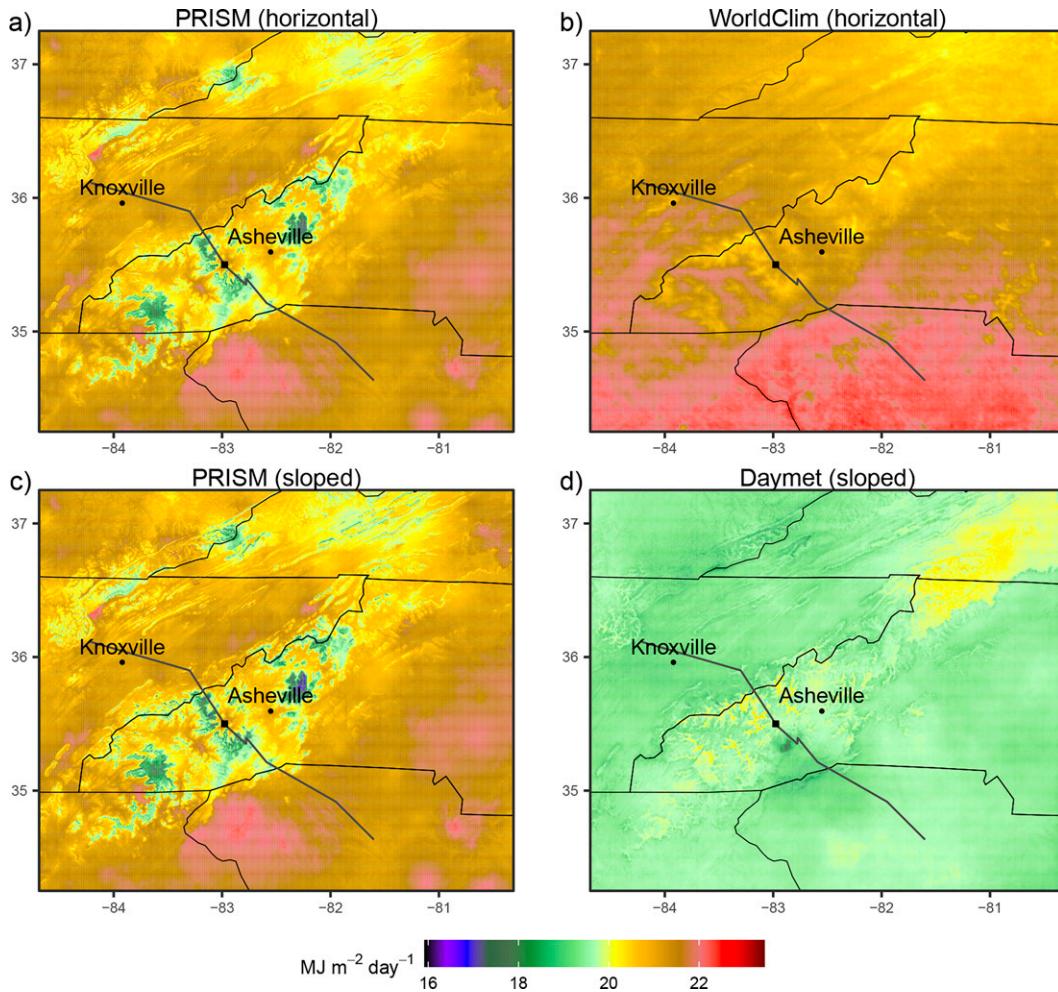


FIG. 12. As in Fig. 8, but for July and highlighting the orographic influence of the Appalachian Mountains of North Carolina. The dark-gray line shows the location of transect 4 used in Fig. 9d. The black square marks the location of the interpolated cell discussed in the text (see also Fig. 13, below). The mapped region corresponds to area “D” in Fig. 4.

irradiance in the mountains ( $24\text{--}33 \text{ MJ m}^{-2} \text{ day}^{-1}$ ) and lower global irradiance in the Great Plains. This opposite pattern is consistent with the summer biases shown by Slater (2016). NLDAS is similar to PRISM, albeit with slightly higher values in general (Fig. 7).

*e. North Carolina Appalachians*

In the summer months, PRISM shows distinctly lower global irradiance over the orographically favored areas of the Appalachian Mountains in North Carolina than on the Piedmont to the southeast and to the Ridge and Valley region to the northwest (see Figs. 9d and 12 for July). Within the mountain range, there is also a clear contrast between the drier interior valleys (higher radiation) and wetter mountains (lower radiation). By October, the pattern has been replaced by one with lower global irradiance north and west of the mountains as the predominant wind direction becomes northwesterly (not shown). The July pattern in global irradiance mimics the

pattern in  $RH_{\min}$  (see Fig. S7 in the online supplemental material) and is a consequence of the distinctly negative correlation between  $T_c$  and  $RH_{\min}$  in this region. As an example, Fig. 13 shows mean July  $T_c$  against mean July  $RH_{\min}$  for the 26 stations used to estimate  $T_c$  at  $82.873^\circ\text{W}$  and  $35.489^\circ\text{N}$ . The coefficient of determination  $R^2$  of the linear regression is 0.59. All 26 stations are less than 79 km from the interpolated location. WorldClim shows a similar but muted spatial pattern, with only about one-half of the range between the low and high values of solar radiation relative to PRISM. Curiously, this summer pattern is evident in both PRISM and WorldClim, and even in the lower-resolution NLDAS and ERA5-Land (not shown, but is visible for July in Fig. 7), but is largely absent from Daymet.

*f. Error analysis*

The total error at a given location in the gridded data arises from error in the station data and error from the interpolation

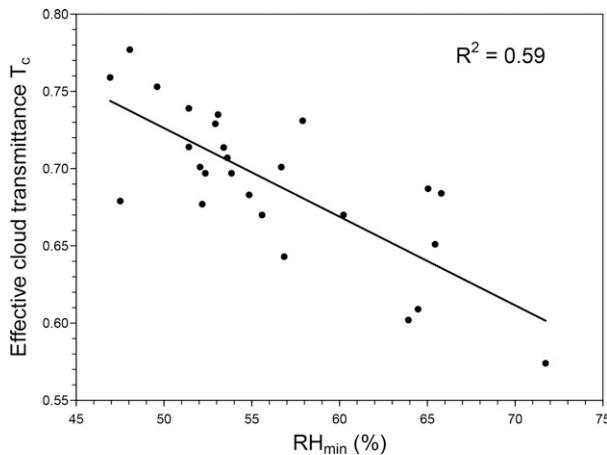


FIG. 13. Mean July effective cloud transmittance  $T_c$  against the predictor variable mean July daily minimum relative humidity  $RH_{\min}$  for the 26 stations (black circles) used to estimate  $T_c$  at 82.873°W and 35.489°N, a valley location in the Appalachians of North Carolina (see Fig. 12). The black line is the result of the linear regression.

method. Errors contributed from the latter source can be estimated using omit-one jackknife cross validation. Overall, mean absolute interpolation error of mean monthly effective cloud transmittance  $T_c$  is relatively small (2.4%); lower than typical factory calibration errors of pyranometers. Errors vary seasonally and are highest in midwinter (3.1%) and lowest in summer (2.1%) (Table 2). Higher error in midwinter can be expected given that relative errors in effective cloud transmittance will tend to be higher when solar radiation is low.

The error contributed from observations can be estimated by leaving out some high quality observations completely from the mapping processes and comparing those data with gridded data at the same locations. We compared PRISM gridded GHI monthly normals with monthly means calculated from 16 NREL stations not used in the mapping process. The period of record for the NREL means ranged from 5 to 34 years and spans 1985 through 2021. We applied the same QC procedure to the NREL daily data as we did to the other daily observations to exclude questionable data. However, the NREL data that passed QC were left unadjusted, the assumption being that they were already very accurate.

We estimated error as the mean of the absolute value of the percent difference in GHI between the gridded monthly normals and the NREL station monthly means. We acknowledge that errors at NREL station locations will not represent the complete distribution of errors across CONUS. Given that, PRISM errors ranged from as little as 1.9% to as high as

10.6% across stations, with a mean of 5.2% (Table 3). Because means based on shorter records will include shorter-term fluctuations, we also recalculated the mean error across stations after increasing the minimum acceptable record length progressively by one year. In general, mean error decreased as record length increased and number of stations decreased. At 14 years, eight stations remained, and the mean error was 4.2%. Error varied by month (Table 4), with overall higher error in winter (e.g., 8.5% in January) than in summer (e.g., 3.7% in June).

PRISM had the least overall error (5.2%) among the six gridded solar radiation datasets examined; mean errors for the other datasets ranged from 5.8% (WorldClim) to 10.5% (MERRA-2). Although the other datasets showed less error for some stations and some months, PRISM had the least error for a plurality of stations (Table 3) and a plurality of months (Table 4).

#### g. Limitations and uncertainty

The quality of the gridded solar radiation data is ultimately limited by the quality, length, and density of solar observations over CONUS. As previously noted (e.g., Slater 2016), solar observations across the United States are frequently degraded as a result of poor calibration, poor maintenance, and inappropriate site conditions. Our QC process excluded approximately 25% of the observations we initially acquired, and of the daily values we retained, 88% were adjusted by at least 1%. When estimating the long term monthly means, our minimum allowable record length (5 years) further excluded 22% of the stations that still had some “valid” solar observations.

Uncertainty is undoubtedly large in areas of CONUS with low coverage density (e.g., Maine, northeast Arizona and northwest New Mexico, and western Montana). In some areas, interpolation relied heavily on ASOS and NSRDB stations where solar observations were relatively scarce (e.g., much of the Midwest). ASOS and NSRDB stations have wide coverage over CONUS, and ASOS stations tend to have long records, but, as we showed, sizable biases can result from the algorithms used to estimate GHI from cloud properties (ASOS) or other environmental variables (NSRDB). Although errors in daily estimates of global solar radiation using ASOS cloud data can be large, we might expect long-term biases to be much smaller, and Belcher and DeGaetano (2007) do report relatively small mean seasonal biases ranging from 1.1% in winter to  $-2.5\%$  in summer with no apparent regional patterns in bias. In contrast, we found distinct regional patterns of bias across CONUS with mean monthly bias ranging from about  $-10\%$  to  $10\%$ . Similar to NSRDB, Wilcox (2012) report

TABLE 2. CONUS-wide PRISM-interpolation cross-validation mean absolute error (%) for mean monthly effective cloud transmittance  $T_c$ . Errors are reported as the average error over two months.

Statistic	Dec–Jan	Feb–Mar	Apr–May	Jun–Jul	Aug–Sep	Oct–Nov	Avg
Mean $T_c$	0.65	0.68	0.72	0.77	0.77	0.71	0.72
Error (%)	3.1	2.7	2.1	2.1	2.1	2.3	2.4

TABLE 3. Station comparison of PRISM, WorldClim, Daymet, ERA5-Land, NLDAS, and MERRA-2 gridded 30-yr normals with NREL observations: mean of the absolute value of the percent difference in GHI between the gridded monthly normals and the NREL station monthly means. Values were extracted from the gridded normals at the stations' coordinates using bilinear interpolation. Daymet values were converted from a sloped surface to a horizontal surface using the ratio of the horizontal to sloped surface radiation from the PRISM normals at the same grid resolution. The conversion only had a minor effect on the tabulated results. The first column gives the average NREL station GHI for all months for reference. Boldface font indicates the dataset with the least error.

Station ID	NREL (MJ m <sup>-2</sup> day <sup>-1</sup> )	PRISM (%)	WorldClim (%)	Daymet (%)	ERA5-land (%)	NLDAS (%)	MERRA-2 (%)
BMS <sup>a</sup>	16.4	<b>3.1</b>	4.0	5.2	13.4	5.2	13.9
BS <sup>a</sup>	13.5	5.2	<b>3.4</b>	10.4	13.0	13.3	25.0
EC <sup>a</sup>	15.3	3.1	<b>2.4</b>	6.8	5.4	10.3	11.6
HSU <sup>a</sup>	13.9	<b>2.0</b>	5.6	5.6	17.6	24.7	17.8
IRRS <sup>a</sup>	16.5	<b>3.9</b>	5.0	5.7	12.3	6.3	12.8
LMU	18.7	<b>5.0</b>	8.0	14.4	8.1	8.5	11.3
LRSS	17.5	7.1	6.4	5.5	5.9	<b>4.1</b>	7.9
NPC	20.4	8.3	5.5	18.5	3.0	4.3	<b>3.6</b>
NWTC <sup>a</sup>	16.0	<b>1.9</b>	3.8	3.2	16.2	5.2	16.5
SMUDA	18.0	4.3	7.6	11.8	<b>3.9</b>	<b>3.9</b>	4.4
SPMD <sup>a</sup>	17.1	5.6	9.0	7.8	14.0	<b>4.5</b>	11.7
STAC	17.8	10.6	8.1	8.8	<b>3.2</b>	6.1	5.3
UAT	21.0	5.0	8.0	17.7	<b>1.8</b>	3.3	2.6
UNLV <sup>a</sup>	20.4	8.6	4.4	18.2	<b>2.6</b>	4.6	3.2
UOSRML	14.4	7.1	7.5	6.5	5.5	8.9	<b>4.9</b>
VTIF	16.2	<b>2.9</b>	4.9	5.1	14.4	6.4	14.9
Mean	17.1	<b>5.2</b>	5.8	9.4	8.8	7.5	10.5
Mean <sup>b</sup>	16.2	<b>4.2</b>	4.7	7.9	11.8	9.3	14.1

<sup>a</sup> Station with at least 14 years of data per calendar month.

<sup>b</sup> Only stations with at least 14 years of data are used in the mean.

small mean monthly biases (from -0.07 to 1.73%), yet we found distinct regional variability in mean monthly bias also ranging from about -10% to 10%. Although we attempted to estimate and reduce these ASOS and NSRDB biases across CONUS, the actual biases were poorly known where reliable nearby observations were scarce.

With station density increasing greatly from 1991 through 2020, most locations have less than 10 years of record and include only the last decade of the analysis period. Though steps were taken to reduce the deviations at these stations

from the long-term mean [section 3a(3)], the final climatologies are still weighted toward the later years.

Last, our estimation of cloud transmittance and the steps taken to reduce errors from degraded instruments rely on modeled clear-sky GHI with optical parameters derived from MERRA-2. Mean absolute biases in modeled clear-sky GHI of roughly one percent, if not more, should be expected (Gueymard 2012; Sun et al. 2019) and although we have not done so here, future work should quantify the biases in our modeled clear-sky GHI against benchmark data.

TABLE 4. Monthly comparison of PRISM, WorldClim, Daymet, ERA5-Land, NLDAS, and MERRA-2 gridded normals with NREL observations: mean of the absolute value of the percent difference in GHI between gridded monthly normals and NREL station monthly means. The first column gives the monthly average global solar radiation for all NREL stations for reference. See Table 3 caption for additional details.

Month	NREL (MJ m <sup>-2</sup> day <sup>-1</sup> )	PRISM (%)	WorldClim (%)	Daymet (%)	ERA5-land (%)	NLDAS (%)	MERRA-2 (%)
Jan	9.2	8.4	9.3	10.1	8.5	8.7	<b>6.4</b>
Feb	12.5	7.7	9.2	9.9	9.3	7.6	<b>6.3</b>
Mar	17.0	<b>5.4</b>	6.8	7.1	8.4	7.0	7.1
Apr	21.0	4.8	5.7	7.4	9.0	<b>4.7</b>	9.1
May	23.3	<b>4.0</b>	5.4	9.7	9.6	9.1	12.2
Jun	25.5	3.7	<b>2.9</b>	7.1	7.8	6.1	12.4
Jul	23.9	<b>4.1</b>	5.3	7.6	11.4	9.1	13.6
Aug	21.8	4.8	<b>4.2</b>	8.4	9.9	8.9	12.5
Sep	18.4	3.9	<b>3.1</b>	8.9	8.8	7.1	14.0
Oct	14.0	<b>3.5</b>	<b>3.5</b>	11.6	8.0	7.0	14.1
Nov	10.2	<b>5.4</b>	8.5	14.6	6.3	6.1	9.3
Dec	8.1	6.9	<b>6.3</b>	10.8	8.2	8.3	8.4
Mean	17.1	<b>5.2</b>	5.8	9.4	8.8	7.5	10.4

## 5. Conclusions

High-resolution (30 arc s) grids of long-term (1991–2020) mean monthly global irradiance were developed for CONUS. To meet multiple user needs, we calculated both irradiance normal to a horizontal surface and to a sloped ground surface. Recognizing that complementary datasets used to generate solar irradiance could also aid researchers, we have also provided global clear-sky GHI and effective cloud transmittance  $T_c$  in the PRISM solar radiation dataset.

To generate these datasets, we took advantage of the exponential growth in solar radiation measurement locations across CONUS over the last several decades (Fig. 1). Even with the expanded networks of observations, there exist large resources still untapped by our work, namely the radiation or power output measured at photovoltaic systems. With 2.7 million residential photovoltaic systems installed as of 2020 (Feldman et al. 2021) and large growth expected over the next few decades (U.S. Energy Information Office 2021), there exists the potential for additional solar radiation measurements, mostly in private hands, to be acquired at vastly more locations than what is currently and publicly available from existing networks. High-resolution atmospheric optical properties derived from recently deployed satellites (e.g., Heidinger et al. 2020) also offer the potential to improve the accuracy of spatial interpolation of surface station data.

The creation of a gridded monthly climatology of global irradiance is a first step toward generating gridded monthly and daily time series datasets at both 30 arc s and  $1/24^\circ$  ( $\sim 4$  km) resolutions to complement the larger PRISM suite of gridded meteorological data. As with other meteorological variables, the climatologies have the potential to serve as predictors in the climatologically aided interpolation of the monthly and daily values.

*Acknowledgments.* We thank Michael Halbleib for making multiple figures (Figs. 4 and 5, along with Figs. S7, S8, and S9 in the online supplemental material), Dylan Keon and Eileen Kaspar for reading an earlier version of this paper, and Rebecca Loiselle and Keith Olsen for assisting with data acquisition. All authors were supported by the USDA Risk Management Agency (Cooperative Agreement 2019-2363).

*Data availability statement.* MERRA-2 data from the NASA Global Model and Assimilation Office (GMAO) are openly available from the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) at <https://doi.org/10.5067/ME5QX6Q5IGGU> (elevation grid) and <https://doi.org/10.5067/Q9QMY5PBNV1T> (radiation diagnostics). NLDAS-2 data are openly available from the NASA GES DISC at <https://doi.org/10.5067/THUF4J1RLSYG>. ERA5 data from ECWMF are openly available from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) at <https://doi.org/10.24381/cds.e2161bac>. Daymet data are openly available from the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) at <https://doi.org/10.3334/ORNLDAAC/1840>. WorldClim, version 2.1, data are openly available at <https://worldclim.org/data/worldclim21>.

[html](#) (accessed 24 March 2021). NSRDB 1961–1990 and 1991–2005 archive data are openly available from the National Renewable Energy Laboratory at <https://nsrdb.nrel.gov/datasets/archives.html> (accessed 9 September 2020). Station solar and cloud observations used in this study are openly available and were compiled from many networks. See the online supplemental material for the full list of station networks (supplemental Table S1) and the URLs where the data were accessed (supplemental Table S2). PRISM climatologies of meteorological data are openly available at <https://prism.oregonstate.edu/normals>.

## REFERENCES

- Abatzoglou, J. T., 2013: Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.*, **33**, 121–131, <https://doi.org/10.1002/joc.3413>.
- Belcher, B. N., and A. T. DeGaetano, 2007: A revised empirical model to estimate solar radiation using automated surface weather observations. *Sol. Energy*, **81**, 329–345, <https://doi.org/10.1016/j.solener.2006.07.003>.
- Bohn, T. J., B. Livneh, J. W. Oyler, S. W. Running, B. Nijssen, and D. P. Lettenmaier, 2013: Global evaluation of MTCLIM and related algorithms for forcing of ecological and hydrological models. *Agric. For. Meteorol.*, **176**, 38–49, <https://doi.org/10.1016/j.agrformet.2013.03.003>.
- Cenzig, H. S., J. M. Gregory, and J. L. Sebaugh, 1981: Solar radiation prediction from other climatic variables. *Trans. ASAE*, **24**, 1269–1272, <https://doi.org/10.13031/2013.34431>.
- Cosgrove, B. A., and Coauthors, 2003: Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res.*, **108**, 8842, <https://doi.org/10.1029/2002JD003118>.
- Cronin, M. F., and M. J. McPhaden, 1997: The upper ocean heat balance in the western equatorial Pacific warm pool during September–December 1992. *J. Geophys. Res.*, **102**, 8533–8553, <https://doi.org/10.1029/97JC00020>.
- Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteorol.*, **33**, 140–158, [https://doi.org/10.1175/1520-0450\(1994\)033<0140:ASTMFM>2.0.CO;2](https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2).
- , W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. *Climate Res.*, **22**, 99–113, <https://doi.org/10.3354/cr022099>.
- , E. H. Helmer, and M. Quiñones, 2003: Mapping the climate of Puerto Rico, Vieques and Culebra. *Int. J. Climatol.*, **23**, 1359–1381, <https://doi.org/10.1002/joc.937>.
- , J. W. Smith, J. I. Smith, and R. B. McKane, 2007: High-resolution spatial modeling of daily weather elements for a catchment in the Oregon Cascade Mountains, United States. *J. Appl. Meteorol. Climatol.*, **46**, 1565–1586, <https://doi.org/10.1175/JAM2548.1>.
- , M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.*, **28**, 2031–2064, <https://doi.org/10.1002/joc.1688>.
- , M. P. Widrlechner, M. D. Halbleib, J. I. Smith, and W. P. Gibson, 2012: Development of a new USDA plant

- hardiness zone map for the United States. *J. Appl. Meteor. Climatol.*, **51**, 242–264, <https://doi.org/10.1175/2010JAMC2536.1>.
- , J. I. Smith, and K. V. Olson, 2015: Mapping atmospheric moisture climatologies across the conterminous United States. *PLOS ONE*, **10**, e0141140, <https://doi.org/10.1371/journal.pone.0141140>.
- , and Coauthors, 2021: Challenges in observation-based mapping of daily precipitation across the conterminous United States. *J. Atmos. Oceanic Technol.*, **38**, 1979–1992, <https://doi.org/10.1175/JTECH-D-21-0054.1>.
- Dee, D. P., M. Balmaseda, G. Balsamo, R. Engelen, A. J. Simmons, and J.-N. Thépaut, 2013: Toward a consistent reanalysis of the climate system. *Bull. Amer. Meteor. Soc.*, **95**, 1235–1248, <https://doi.org/10.1175/BAMS-D-13-00043.1>.
- Diamond, H. J., and Coauthors, 2013: U.S. Climate Reference Network after one decade of operations: Status and assessment. *Bull. Amer. Meteor. Soc.*, **94**, 485–498, <https://doi.org/10.1175/BAMS-D-12-00170.1>.
- Dubayah, R. C., 1994: Modeling a solar radiation topoclimatology for the Rio Grande River Basin. *J. Veg. Sci.*, **5**, 627–640, <https://doi.org/10.2307/3235879>.
- Dubayah, R., J. Dozier, and F. W. Davis, 1990: Topographic distribution of clear-sky radiation over the Konza Prairie, Kansas. *Water Resour. Res.*, **26**, 679–690, <https://doi.org/10.1029/WR026i004p00679>.
- Feldman, D., K. Wu, and R. Margolis, 2021: H1 2021 solar industry update. National Renewable Energy Laboratory Tech. Rep. NREL/PR-7A40-80427, 65 pp., <https://doi.org/10.2172/1808491>.
- Fick, S. E., and R. J. Hijmans, 2017: WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.*, **37**, 4302–4315, <https://doi.org/10.1002/joc.5086>.
- Frew, J., 1990: The image processing workbench. Ph.D. dissertation, University of California, Santa Barbara, 305 pp., [https://frew.eri.ucsb.edu/local\\_pubs/1990\\_IPW.pdf](https://frew.eri.ucsb.edu/local_pubs/1990_IPW.pdf).
- Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- Gueymard, C. A., 2012: Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models. *Sol. Energy*, **86**, 2145–2169, <https://doi.org/10.1016/j.solener.2011.11.011>.
- , and D. R. Myers, 2009: Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling. *Sol. Energy*, **83**, 171–185, <https://doi.org/10.1016/j.solener.2008.07.015>.
- Habte, A., S. Wilcox, and T. Stoffel, 2015: Evaluation of radiometers deployed at the National Renewable Energy Laboratory's Solar Radiation Research Laboratory. National Renewable Energy Laboratory Tech. Rep. NREL/TP-5D00-60896, 188 pp., <https://doi.org/10.2172/1126287>.
- , M. Sengupta, A. Andreas, S. Wilcox, and T. Stoffel, 2016: Intercomparison of 51 radiometers for determining global horizontal irradiance and direct normal irradiance measurements. *Sol. Energy*, **133**, 372–393, <https://doi.org/10.1016/j.solener.2016.03.065>.
- Heidinger, A. K., M. J. Pavolonis, C. Calvert, J. Hoffman, S. Nebuda, W. Straka, A. Walther, and S. Wanzong, 2020: ABI cloud products from the GOES-R Series. *The GOES-R Series*, S. J. Goodman et al., Eds., Elsevier, 43–62.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Holden, Z. A., and Coauthors, 2018: Decreasing fire season precipitation increased recent western US forest wildfire activity. *Proc. Natl. Acad. Sci. USA*, **115**, E8349–E8357, <https://doi.org/10.1073/pnas.1802316115>.
- Huntzinger, D. N., and Coauthors, 2013: The North American carbon program multi-scale synthesis and terrestrial model intercomparison project—Part 1: Overview and experimental design. *Geosci. Model Dev.*, **6**, 2121–2133, <https://doi.org/10.5194/gmd-6-2121-2013>.
- Jepsen, S. M., N. P. Molotch, M. W. Williams, K. E. Rittger, and J. O. Sickman, 2012: Interannual variability of snowmelt in the Sierra Nevada and Rocky Mountains, United States: Examples from two alpine watersheds. *Water Resour. Res.*, **48**, W02529, <https://doi.org/10.1029/2011WR011006>.
- Journée, M., and C. Bertrand, 2011: Quality control of solar radiation data within the RMIB solar measurements network. *Sol. Energy*, **85**, 72–86, <https://doi.org/10.1016/j.solener.2010.10.021>.
- Kafka, J. L., and M. A. Miller, 2019: A climatology of solar irradiance and its controls across the United States: Implications for solar panel orientation. *Renewable Energy*, **135**, 897–907, <https://doi.org/10.1016/j.renene.2018.12.057>.
- Lapo, K. E., L. M. Hinkelman, E. Sumargo, M. Hughes, and J. D. Lundquist, 2017: A critical evaluation of modeled solar irradiance over California for hydrologic and land surface modeling. *J. Geophys. Res. Atmos.*, **122**, 299–317, <https://doi.org/10.1002/2016JD025527>.
- Livneh, B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier, 2013: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *J. Climate*, **26**, 9384–9392, <https://doi.org/10.1175/JCLI-D-12-00508.1>.
- Longman, R. J., T. W. Giambelluca, and M. A. Nullet, 2013: Use of a clear-day solar radiation model to homogenize solar radiation measurements in Hawai'i. *Sol. Energy*, **91**, 102–110, <https://doi.org/10.1016/j.solener.2013.02.006>.
- Marks, D. G., S. C. Havens, and M. J. Johnson, 2018: The Image Processing Workbench (computer code). USDA-ARS-NWRC/ipw, Zenodo, <http://doi.org/10.5281/zenodo.1301290>.
- Maxwell, E. L., W. F. Marion, D. R. Myers, M. D. Rymes, and S. M. Wilcox, 1995: National Solar Radiation Data Base (1961–1990). National Renewable Energy Laboratory Final Tech. Rep. NREL/TP-463-5784, 318 pp., <https://doi.org/10.2172/70747>.
- Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily database. *J. Atmos. Oceanic Technol.*, **29**, 897–910, <https://doi.org/10.1175/JTECH-D-11-00103.1>.
- Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, <https://doi.org/10.1175/BAMS-87-3-343>.
- Muñoz-Sabater, J., and Coauthors, 2021: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data*, **13**, 4349–4383, <https://doi.org/10.5194/essd-13-4349-2021>.
- Perez, R., P. Ineichen, K. Moore, M. Kmieciak, C. Chain, R. George, and F. Vignola, 2002: A new operational model for satellite-derived irradiances: Description and validation. *Sol. Energy*, **73**, 307–317, [https://doi.org/10.1016/S0038-092X\(02\)00122-6](https://doi.org/10.1016/S0038-092X(02)00122-6).

- Riihimatki, L., L. S. Lohmann, R. Meyers, R. Perez, and F. Vignola, 2006: Long-term variability of global and beam irradiance in the Pacific Northwest. *Proc. Solar 2006 Annual Conf.*, Denver, CO, American Solar Energy Society, 54–59.
- Shi, G.-Y., T. Hayasaka, A. Ohmura, Z.-H. Chen, B. Wang, J.-Q. Zhao, H.-Z. Che, and L. Xu, 2008: Data quality assessment and the long-term trend of ground solar radiation in China. *J. Appl. Meteor. Climatol.*, **47**, 1006–1016, <https://doi.org/10.1175/2007JAMC1493.1>.
- Slater, A. G., 2016: Surface solar radiation in North America: A comparison of observations, reanalyses, satellite, and derived products. *J. Hydrometeorol.*, **17**, 401–420, <https://doi.org/10.1175/JHM-D-15-0087.1>.
- , A. P. Barrett, M. P. Clark, J. D. Lundquist, and M. S. Raleigh, 2013: Uncertainty in seasonal snow reconstruction: Relative impacts of model forcing and image availability. *Adv. Water Resour.*, **55**, 165–177, <https://doi.org/10.1016/j.advwatres.2012.07.006>.
- Stoffel, T. L., I. Reda, D. R. Myers, D. Renne, S. Wilcox, and J. Treadwell, 2000: Current issues in terrestrial solar radiation instrumentation for energy, climate, and space applications. *Metrologia*, **37**, 399–402, <https://doi.org/10.1088/0026-1394/37/5/11>.
- Sun, X., J. M. Bright, C. A. Gueymard, B. Acord, P. Wang, and N. A. Engerer, 2019: Worldwide performance assessment of 75 global clear-sky irradiance models using principal component analysis. *Renewable Sustainable Energy Rev.*, **111**, 550–570, <https://doi.org/10.1016/j.rser.2019.04.006>.
- Thornton, P. E., and S. W. Running, 1999: An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agric. For. Meteorol.*, **93**, 211–228, [https://doi.org/10.1016/S0168-1923\(98\)00126-9](https://doi.org/10.1016/S0168-1923(98)00126-9).
- , —, and M. A. White, 1997: Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.*, **190**, 214–251, [https://doi.org/10.1016/S0022-1694\(96\)03128-9](https://doi.org/10.1016/S0022-1694(96)03128-9).
- , H. Hasenauer, and M. A. White, 2000: Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: An application over complex terrain in Austria. *Agric. For. Meteorol.*, **104**, 255–271, [https://doi.org/10.1016/S0168-1923\(00\)00170-2](https://doi.org/10.1016/S0168-1923(00)00170-2).
- , R. Shrestha, M. Thornton, S.-C. Kao, Y. Wei, and B. E. Wilson, 2021: Gridded daily weather data for North America with comprehensive uncertainty quantification. *Sci. Data*, **8**, 190, <https://doi.org/10.1038/s41597-021-00973-0>.
- U.S. Energy Information Office, Ed., 2021: Annual Energy Outlook 2021: With Projections to 2050. Government Printing Office, 56 pp., <https://www.eia.gov/outlooks/aeo/pdf/00%20AEO2021%20Chart%20Library.pdf>.
- Wei, Y., and Coauthors, 2014: The North American Carbon Program Multi-Scale Synthesis And Terrestrial Model Intercomparison Project—Part 2: Environmental driver data. *Geosci. Model Dev.*, **7**, 2875–2893, <https://doi.org/10.5194/gmd-7-2875-2014>.
- Wilcox, S. M., 2007: National Solar Radiation Database 1991–2005 update: User’s manual. National Renewable Energy Laboratory Tech. Rep. NREL/TP-581-41364, 472 pp., <https://doi.org/10.2172/901864>.
- , 2012: National Solar Radiation Database 1991–2010 update: User’s manual. National Renewable Energy Laboratory Tech. Rep. NREL/TP-5500-54824, 479 pp., <https://doi.org/10.2172/1054832>.
- Willmott, C. J., and S. M. Robeson, 1995: Climatologically aided interpolation (CAI) of terrestrial air temperature. *Int. J. Climatol.*, **15**, 221–229, <https://doi.org/10.1002/joc.3370150207>.
- Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.*, **117**, D03109, <https://doi.org/10.1029/2011JD016048>.
- Younes, S., R. Claywell, and T. Muneer, 2005: Quality control of solar radiation data: Present status and proposed new approaches. *Energy*, **30**, 1533–1549, <https://doi.org/10.1016/j.energy.2004.04.031>.



**AMS**  
American Meteorological Society

## Supplemental Material

© [Copyright 2022 American Meteorological Society](#) (AMS)

For permission to reuse any portion of this work, please contact [permissions@ametsoc.org](mailto: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 USC §107) or that satisfies the conditions specified in Section 108 of the U.S. Copyright Act (17 USC §108) does not require 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 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>).

**Supplementary Material for**

**Mapping an observation-based global solar irradiance climatology across  
the conterminous United States**

David E. Rupp<sup>a</sup>, Christopher Daly<sup>b</sup>, Matt K. Doggett<sup>b</sup>, Joseph I. Smith<sup>b</sup>, and Ben Steinberg<sup>b</sup>

<sup>a</sup>Oregon Climate Change Research Institute, College of Earth, Ocean, and Atmospheric Sciences, Oregon State  
University, Corvallis, OR

<sup>b</sup>PRISM Climate Group, Northwest Alliance for Computational Science and Engineering, College of  
Engineering, Oregon State University, Corvallis, OR

*Corresponding author:* David E. Rupp, david.rupp@oregonstate.edu, <https://orcid.org/0000-0003-3562-2072>

## S1. Estimating subdaily cloud transmittance from ASOS (METAR)

Following Belcher and DeGaetano (2007), total cloud transmittance at a given time of observation was calculated as the product of partial transmittance values for each cloud layer observed. Partial transmittance values depend on cloud base height and coverage (see Table S3). The total transmittance  $T_{c,i}$  at time indexed by  $i$  is calculated as

$$T_{c,i} = \prod_{j=1}^n T_{c,i,j} / (1 - \alpha_s \alpha_c) \quad (\text{S1})$$

where  $j$  indexes a cloud layer,  $n$  is the total number of layers,  $\alpha_s$  is the surface albedo and  $\alpha_c$  is the cloud albedo. Total transmittance was further reduced by additional factors when certain weather phenomena were observed (see Table S3). Like Belcher and DeGaetano (2007), we set  $\alpha_s = 0.2$  and did not adjust for presence of snow because we did not consistently have snow cover data at each location. Also like Belcher and DeGaetano (2007), when the lowest cloud base was below 3659 m,  $\alpha_c = 0.5$ , else  $\alpha_c = 0$ .

Belcher and DeGaetano (2007) assumed ASOS data would contain up to three cloud layers and no cloud bases above 3659 m. However, since they first developed their algorithm, many ASOS stations with newer ceilometers began reporting more than three layers and report cloud bases above 3659 m. To use this additional information, we simply repeated the values for cloud layer 3 for cloud layers 4 and up. For cloud base heights above 3659, we simply used the partial transmittance values for 3049 – 3659 m range. Partial transmittance values produced by Belcher and DeGaetano (2004) for manual observations of cloud height and coverage suggest higher transmittance for cloud bases above 3659 than the values we used. Ideally, new partial transmittance tables would be generated for the newer ceilometers. In the meantime, however, we attempted to reduce systematic biases in ASOS total transmittance based on the existing partial transmittance values in Table S3 (see Section S3).

Obtaining records on the precise instrument used at each ASOS station and when it was installed or discontinued was infeasible. Therefore, we relied on the data themselves to estimate the history of reporting methods and/or instrumentation. For each calendar year we calculated the fraction of days with recorded cloud heights  $> 3659$  m during daylight hours. If the frequency of days in a year with cloud heights above 3659 m was less  $< 5\%$ , we assumed the ceilometer used in that year could not detect clouds above 3840 m (“Type I” station), otherwise we assumed it could (“Type II” station). We did not attempt to determine when, within a year, a change in instrumentation or reporting methods was made.

We allowed stations to change type from year to year because we found cases where a station historically did not (Type I) and then did (Type II) frequently report cloud heights  $> 3659$  m, and cases where a station historically did (Type II) and then did not (Type I) frequently report cloud heights  $> 3659$  m. The former cases suggest an upgrade in instrumentation. The latter cases could be explained by the discontinuation of frequent manual observations, but we did not attempt to determine how often a station used manual observations.

## S2. Quality control and bias reduction in global horizontal irradiance (GHI) measurements

First, range and completeness tests were applied to the hourly and daily GHI ( $S_{\text{surf}}$ ) in the following sequence:

- 1) General hourly range check: Hourly  $S_{\text{surf}}$  outside the range  $-1.0$  to  $1,500 \text{ Wm}^{-2}$  were flagged as invalid; hourly  $S_{\text{surf}}$  between  $-1.0$  and  $0 \text{ Wm}^{-2}$  were set to 0 and retained. An exception was made for the SURFRAD network, where hourly values between  $-6.0$  and  $0 \text{ Wm}^{-2}$  were set to 0 and retained.
- 2) General daily completeness check: If fewer than 18 (out of 24) hourly  $S_{\text{surf}}$  values were valid, the day was flagged as invalid.
- 3) Daytime range and completeness check: If any hourly  $S_{\text{surf}} = 0$  or was missing or invalid between one hour after sunrise and one hour before sunset, the day was flagged as invalid
- 4) Nighttime range and completeness check: If more than 20% of the hourly  $S_{\text{surf}} \neq 0$  or were missing or invalid between one hour after sunset and one hour before sunrise (nighttime hours), the day was flagged as invalid
- 5) General daily range check: Once hourly data were aggregated to daily totals, daily  $S_{\text{surf}}$  outside the range  $0 - 45 \text{ MJ m}^{-2}$  were flagged as invalid

Next, we used the following algorithm to adjust observed  $S_{\text{surf}}$  and identify values that did not mean quality criteria after adjustment:

- 1) Assign a modeled clear-sky GHI ( $S_{\text{clear\_sky}}$ ) to each day that has an observed  $S_{\text{surf}}$ . For leap days, use the average of  $S_{\text{clear\_sky}}$  on 28 February and 1 March. Set prior guess of CSR = 1.

First pass:

- 2) Define a 183-day analysis window starting on day  $i$  (e.g., 1 January 1990 for  $i = 1$ ) and do steps 3 through 8 within this window. (A range of window size was tested. 183 days struck a balance between over-correcting data but not missing brief periods (sub-annual) when the instrument was out of calibration or faulty).
- 3) Calculate  $T_c = S_{\text{surf}} / S_{\text{clear\_sky}}$  for day in window. Exclude days with  $T_c \geq 1.3$  from the following calculations.
- 4) Get the  $n = 5$  highest values of  $T_c$  ( $T_{c,n}$ ), calculate their mean ( $\overline{T_{c,n}}$ ) and their anomalies (anom) as ratios with respect to their mean:  $T_{c,n\_anom} = T_{c,n} / \overline{T_{c,n}}$ .
- 5) Exclude low outliers (those with  $T_{c,n\_anom} < 0.01$ ) if they exist and recalculate  $T_{c,n\_anom} = T_{c,n} / \overline{T_{c,n}}$  for  $n < 5$ .
- 6) Exclude high outliers (those with  $T_{c,n\_anom} > 1.03$ ) if they exist and recalculate  $T_{c,n\_anom} = T_{c,n} / \overline{T_{c,n}}$  for  $n < 5$ .
- 7) Let  $\overline{T_{c,n}}$  be the estimate of CSR (i.e.,  $E_{\text{clear\_sky}} = \overline{T_{c,n}} S_{\text{clear\_sky}}$ ). If  $n = 0$ , use prior guess of CSR (at iteration  $i - 1$ ).
- 8) Adjust all observations in window using Eq. (4) and replace  $S_{\text{surf}}$  with  $S_{\text{surf}}^*$  within the window.
- 9) Advance window forward 1 day ( $i = i + 1$ ) and repeat steps 2 through 8 until end of record reached.

Second pass:

- 10) Repeat steps 2 through 9 but with a window that varies in size sinusoidally from 91 days in mid-winter to 183 days in mid-summer. (Narrower windows were used in winter to attempt to correct for situations where obstacles, such as trees, impeded direct radiation at low sun angles, which would have more impact on daily GHI in winter).

Third pass:

- 11) Calculate  $T_c$  for entire record using adjusted values of  $S_{\text{surf}}$ .
- 12) Flag days with  $T_c > 1.05$  as bad.
- 13) Find periods of 60 consecutive days with  $T_c < 0.75$ . Flag all 60 days as bad.
- 14) Find periods of 90 consecutive days with  $T_c < 0.9$ . Flag all 90 days as bad.

Fourth pass:

- 15) Flag an entire calendar year as bad if more than 33% of days were previously flagged as bad.

### **S3. Reducing bias in estimates of daily cloud transmittance from ASOS (METAR)**

As discussed in Section 3.a.2, we paired ASOS stations with proximal stations measuring GHI to estimate errors in the estimation of ASOS  $T_c$  using the method of Belcher and DeGaetano (2007) with our minor modifications. For brevity we refer to non-ASOS stations with GHI measurements as “solar” stations. Restricting to solar stations within 20 km of an ASOS station, excluding SNOTEL stations because of the generally low quality of SNOTEL solar data and poor siting within forested locations, and requiring concurrent records of at least 730 days, resulted in 2562 pairs. There were many cases of an ASOS station paired with multiple solar stations, which allowed us to see when patterns were, or were not, consistent among solar stations.

Next, we generated a filtered set of station pairs that include no more than one station pair per ASOS station. When filtering, we first excluded pairs that met the following criteria:

- 1) Pairs with elevation differences  $> 300$  m;
- 2) Pairs where one station’s elevation was  $\leq 25$  m but the other station’s elevation was  $> 40$  m, because of complications from coastal fog;
- 3) Pairs where the root mean squared difference (RMSD) of daily  $T_c$  between stations was  $> 0.2$ .

When there were still multiple pairings for an ASOS station, selection was based on proximity, length of concurrent records, and whether the solar station was in the Remote Automatic Weather Station (RAWS) network, which tend to be of poorer quality. To get a dimensionless score for each pairing with a given ASOS station, the distance between stations and the number of concurrent records were first divided by the maximum value of each metric, respectively, of all the pairings for that ASOS station. A score was then assigned to each pair that was a sum of

the unitless concurrent record length and the inverse of the dimensionless distance. The score of each pair that included a RAWS station was then divided by 2. Selecting the pair with the highest score for each ASOS station resulted in 971 pairs.

We checked which ASOS stations regularly reported cloud base heights above 3659 m. “Regularly reported” meant that a station reported clouds above 3659 m on at least 5% of days within a calendar year. Whether a station was classified as not regularly reporting clouds > 3659 m (“Type I”) or as regularly reporting clouds > 3659 m (“Type II”) could change from one calendar year to the next, but not within a calendar year.

For each of the 2562 station pairs, we calculated the mean of the difference (MD) of ASOS  $T_c$  minus solar station  $T_c$  distinctly within 0.05-wide bins of solar station  $T_c$ . (Note that this is different from what is described in the main text for determining the bias adjustment, where we calculated the mean of the difference of solar station  $T_c$  minus ASOS  $T_c$  as function of ASOS  $T_c$ .) MD was also calculated distinctly for three cases: 1) no clouds above 3659 m reported and the station was Type I, 2) no clouds above 3659 m reported and the station was Type II, and 3) clouds above 3659 m reported at either station Type I or II. Finally, the mean of MD over all stations was calculated, giving each station equal weight.

We found that the bias (MD) was dependent on observed  $T_c$  and most so when clouds above 3659 m were reported (Fig. S3). Bias tended to be positive for Type I stations when they did not report clouds > 3659 meters (case 1). Positive biases for all three cases were highest at low values of observed  $T_c$  and this occurred because the Belcher and DeGaetano method very rarely resulted in ASOS  $T_c < 0.15$  though such low values of  $T_c$  were often observed.

We repeated the analysis for the subset of 971 station pairs. This reduced the range of MD across stations, but had negligible impact of the mean result across stations. We also calculated MD separately for each calendar month, but did not detect sufficient intra-annual variability in MD to justify stratifying our bias analysis by month (results not shown).

We reduced systematic biases in ASOS  $T_c$  through bias correction functions that were a function of ASOS  $T_c$  itself. Bias correction functions were applied daily to each of 3 distinct cases: 1) no clouds above 3659 m reported and the station was Type I, 2) no clouds above 3659 m reported and the station was Type II, and 3) clouds above 3659 m reported at either station Type I or II.

For each station pair, we calculated the mean of the difference (MD) of solar station  $T_c$  minus ASOS  $T_c$  distinctly within 0.05-wide bins of ASOS  $T_c$ . MD was also calculated distinctly depending on how the days were classified (Type I, Type II, with or without clouds > 3659 m). The mean of MD over all stations was then calculated, giving each station equal weight. Third-order polynomials were fitted to mean MD versus bin-averaged ASOS  $T_c$  to parameterize the adjustment we applied to daily ASOS  $T_c$  (Fig. S4). We applied the bias corrections uniformly CONUS-wide. Though we saw some indication of variability in bias by region, we did not at this stage parameterize regional variability in the bias correction. Seasonal variability appeared relatively small so we applied the same equation year-round. Table S4 shows the reduction in mean bias, mean absolute error, and root mean square error of daily ASOS  $T_c$  after bias-correction.

#### **S4. Removing short-term deviations from the long-term mean for stations with shorter records**

This section describes the procedure of removing “short-term” deviations in mean transmittance  $T_c$  from the long-term (1991-2020) mean for stations with records shorter than 23 years.

The short-term (< 23 years), or “target,” station means were adjusted to 1991-2020 using nearby long-term ( $\geq 23$  years) “anchor” stations. Anchor stations were chosen based on proximity to the target station along with the number of years of data they both had in common for a given month. First, anchor stations were ranked by distance from the target station using normalized inverse distance weighting. Next, anchor stations were ranked by the percentage of years of data they had in common with the target station during 1991-2020. Finally, anchor stations were ranked by the percentage of years they had in common with the target station outside 1991-2020. Anchor stations with more years in common within 1991-2020 were ranked higher than those with years outside 1991-2020. The desired number of years in common was capped at 23 since that is the number of years considered sufficient to represent 1991-2020. The final ordering of anchor stations was based on a combined score of the three separate rankings with proximity weighted highest, then percent of years within 1991-2020, and lastly the percent of years outside 1991-2020.

After selecting an anchor station, monthly means for the target and anchor stations were calculated using their respective data from the set of common years. We allowed this set of common years to extend outside the period 1991-2020 if data were available. An adjustment factor was then calculated by dividing the anchor station 1991-2020 mean by the anchor station extended mean. The target station extended mean was multiplied by this factor to determine the target station adjusted mean for 1991-2020. Arithmetically, the target adjusted mean ( $\overline{X}_t'$ ) was calculated as follows:

$$(\overline{X}_t') = \overline{X}_{tc}(\overline{X}_a/\overline{X}_{ac}) \quad (S2)$$

where  $\overline{X}_a$  is the anchor 1991-2020 mean,  $\overline{X}_{ac}$  is the common-period anchor mean, and  $\overline{X}_{tc}$  is the common-period target mean (in our case  $X$  represents  $T_c$ ).  $\overline{X}_t'$  was calculated for each target station using the three anchor stations with the highest scores. These three adjusted means were then averaged to obtain the final 1991-2020 mean for the target station. The choice of using three anchor stations is described in Appendix A of Daly *et al.* (2008).

The final 1991-2020 monthly average values were tested again for spatial consistency using the ASSAY QC system, flagging those that failed as invalid. Finally, 1991-2020 monthly mean GHI values were calculated by multiplying the 1991-2020 average  $T_c$  values by the monthly mean clear-sky GHI values.

## **S5. Mapping regionally consistent biases in estimates of mean monthly cloud transmittance from ASOS and NSRDB stations**

As discussed in Section 3.b.2, we mapped regionally consistent biases in mean monthly cloud transmittance  $T_c$  modeled at ASOS and NSRDB stations and then used the maps to adjust the ASOS and NSRDB values.

Biases were estimated by subtracting the estimate of the “true” mean monthly  $T_c$  from the mean monthly  $T_c$  modeled at each ASOS and NSRDB station. The “true” value was taken as the average  $T_c$  from all stations with valid observations of mean monthly  $T_c$  within a 20 km radius of each ASOS or NSRDB station. Biases were not estimated for ASOS or NSRDB stations with no measured values within 20 km distance.

Mean monthly biases were mapped separately from ASOS and NSRDB stations using CAI as described in Section 3.b.2, except that mean monthly daily minimum relative humidity ( $RH_{min}$ ) was not used as the predictor.  $RH_{min}$  generally provided little explanatory power, so the interpolation was dominated by inverse distance weighting, while smoothing parameters were chosen to avoid very fine-scale variability. As examples, Fig. S8 shows bias maps for two months for ASOS and NSRDB.

## **References**

- Belcher BN, DeGaetano AT. 2004. Integration of ASOS weather data into building energy calculations with emphasis on model derived solar radiation. Ithica NY, Northeast Regional Climate Center Pub. ASHRAE 1226-RP.
- Belcher BN, DeGaetano AT. 2007. A revised empirical model to estimate solar radiation using automated surface weather observations. *Solar Energy*, 81(3): 329–345.  
<https://doi.org/10.1016/j.solener.2006.07.003>.
- Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, 28(15): 2031–2064.  
<https://doi.org/10.1002/joc.1688>.

**Table S1.** Meteorological monitoring networks and number of stations used to generate gridded 1991-2020 mean monthly global irradiance.

	<b>Code</b>	<b>Provider</b>	<b>No. of stations</b>	<b>Name</b>
1	AGRIMET	AGRIMET	158	U.S. Bureau of Reclamation Cooperative Agricultural Weather Network
2	AGWXNET	AGWXNET	175	AgWeatherNet
3	AIRNOW	MESOWEST	136	EPA AirNow
4	AMOS	COOP	2	National Weather Service Cooperative Observer Program
5	APRSWXNET CWOP	MESOWEST	1601	Citizen Weather Observers Program
6	ARLFRD	MESOWEST	35	NOAA Air Resources Laboratory/Field Research Division
7	ARLSORD	MESOWEST	22	NOAA Air Resources Laboratory/Special Operations And Research Division
8	ASOS	ASOS	2266	Automated Surface/Weather Observation System
9	AZMET	MESOWEST	29	University of Arizona Meteorological Network
10	CEMP	MESOWEST	27	DOE/DRI Community Environmental Monitoring Program
11	CIMIS	CIMIS	215	California Irrigation Management Information System
12	COAGMET	COAGMET	92	Colorado Agricultural Meteorological Network
13	DEOS	DEOS	92	Delaware Environmental Observing System
14	DRI	MESOWEST	97	Desert Research Institute
15	FAWN	FAWN	44	Florida Automated Weather Network
16	FGNET	MESOWEST	23	Utah Fruit Grower's Weather Monitoring Network
17	HADS	MESOWEST	96	NOAA Hydrometeorological Automated Data System
18	HJA	HJA	4	H. J. Andrews Experimental Forest
19	KSTATE	KSTATE	79	Kansas Mesonet
20	KYMESONET	KYMESONET	71	Kentucky Mesonet
21	MAWN	MESOWEST	77	Enviroweather
22	NCECONET	NCECONET	40	North Carolina ECONet
23	NDAWN	HPRCC	167	North Dakota Automated Weather Network
24	NEMESO	HPRCC	58	Nebraska Mesonet
25	NEVCAN	NEVCAN	9	Nevada Climate-ecohydrological Assessment Network
26	NJWXNET	NJWXNET	46	New Jersey Climate and Weather Network
27	NMAQ	MESOWEST	17	New Mexico Environment Department
28	NMCC	MESOWEST	10	New Mexico Climate Center
29	NSRDB	NSRDB	1306	National Solar Radiation Database

30	NWT	LTER	3	Niwot Ridge LTER
31	OKMESONET	OKMESONET	142	Oklahoma Mesonet
32	RAWS	RAWS	1368	Remote Automatic Weather Stations
33	SCAN	SNOTEL	192	Soil Climate Analysis Network
34	SD_MESONET	SD_MESONET	37	South Dakota Mesonet
35	SEV	LTER	10	Sevilleta LTER
36	SNTL	SNOTEL	34	Snowpack Telemetry
37	SRP	MESOWEST	29	Salt River Project
38	SURFRAD	SURFRAD	7	Surface Radiation Budget Network
39	TWDB	MESOWEST	11	Texas Water Development Board
40	UCC-AGNET	MESOWEST	41	Utah State University Agricultural Weather Network
41	UGA	UGA	81	University of Georgia Weather Network
42	UOREGON	UOREGON	31	University of Oregon Solar Radiation Monitoring Laboratory
43	USA	USA	27	South Alabama Mesonet
44	USCRN	USCRN	117	U.S. Climate Reference Network
45	UTAHCLIMATE CENTER	MESOWEST	42	Utah Climate Center
46	UUNET	MESOWEST	21	University of Utah MesoWest Group
47	VCAPCD	MESOWEST	6	Ventura County Air Pollution Control District
48	VTWAC	MESOWEST	17	Vermont Weather Analytics Center
49	WACNET	MESOWEST	7	Wyoming Agricultural Climate Network
50	WTEXAS	MESOWEST	95	West Texas Mesonet
		<i>Total</i>	<i>9312</i>	
		<i>Solar obs only<sup>1</sup></i>	<i>5740</i>	

<sup>1</sup>Excludes ASOS and NSRDB

**Table S2.** Website URLs for solar radiation data.

<b>Code</b>	<b>URL</b>	<b>Name<sup>1</sup></b>
AGRIMET	<a href="https://www.usbr.gov/gp/agrimet">https://www.usbr.gov/gp/agrimet</a> <a href="https://www.usbr.gov/pn/agrimet">https://www.usbr.gov/pn/agrimet</a>	U.S. Bureau of Reclamation Cooperative Agricultural Weather Network
AGWXNET	<a href="https://weather.wsu.edu">https://weather.wsu.edu</a>	AgWeatherNet
CIMIS	<a href="https://cimis.water.ca.gov">https://cimis.water.ca.gov</a>	California Irrigation Management Information System
COAGMET	<a href="http://www.coagmet.colostate.edu">http://www.coagmet.colostate.edu</a>	Colorado Agricultural Meteorological Network
COOP	<a href="https://www.weather.gov/coop">https://www.weather.gov/coop</a>	National Weather Service Cooperative Observer Program
DEOS	<a href="http://www.deos.udel.edu">http://www.deos.udel.edu</a>	Delaware Environmental Observing System
FAWN	<a href="https://fawn.ifas.ufl.edu">https://fawn.ifas.ufl.edu</a>	Florida Automated Weather Network
HJA	<a href="https://andrewsforest.oregonstate.edu">https://andrewsforest.oregonstate.edu</a>	H. J. Andrews Experimental Forest
HPRCC <sup>1</sup>	<a href="https://hprcc.unl.edu">https://hprcc.unl.edu</a>	High Plains Regional Climate Network
KSTATE	<a href="http://mesonet.k-state.edu">http://mesonet.k-state.edu</a>	Kansas Mesonet
KYMESONET	<a href="http://www.kymesonet.org">http://www.kymesonet.org</a>	Kentucky Mesonet
LTER <sup>1</sup>	<a href="https://lternet.edu">https://lternet.edu</a>	Long-Term Ecological Research Network
MADIS <sup>1</sup>	<a href="https://madis.noaa.gov">https://madis.noaa.gov</a>	NOAA Meteorological Assimilation Data Ingest System
MESOWEST <sup>1</sup>	<a href="https://mesowest.utah.edu">https://mesowest.utah.edu</a>	University of Utah MesoWest
NCECONET	<a href="https://climate.ncsu.edu">https://climate.ncsu.edu</a>	North Carolina ECONet
NDAWN	<a href="https://ndawn.ndsu.nodak.edu">https://ndawn.ndsu.nodak.edu</a>	North Dakota Agricultural Weather Network
NEVCAN	<a href="https://nevcan.dri.edu">https://nevcan.dri.edu</a>	Nevada Climate-ecohydrological Assessment Network
NJWXNET	<a href="https://www.njweather.org">https://www.njweather.org</a>	New Jersey Climate and Weather Network
NREL <sup>2</sup>	<a href="https://midcdmz.nrel.gov/">https://midcdmz.nrel.gov/</a>	National Renewable Energy Laboratory
OKMESONET	<a href="https://www.mesonet.org">https://www.mesonet.org</a>	Oklahoma Mesonet
RAWS	<a href="https://raws.nifc.gov">https://raws.nifc.gov</a>	Remote Automatic Weather Stations
SCAN	<a href="https://www.wcc.nrcs.usda.gov/scan">https://www.wcc.nrcs.usda.gov/scan</a>	Soil Climate Analysis Network
SD_MESONET	<a href="https://climate.sdstate.edu">https://climate.sdstate.edu</a>	South Dakota Mesonet
SNOTEL	<a href="https://www.wcc.nrcs.usda.gov/snow">https://www.wcc.nrcs.usda.gov/snow</a>	Snowpack Telemetry
SURFRAD	<a href="https://gml.noaa.gov/grad/surfrad">https://gml.noaa.gov/grad/surfrad</a>	Surface Radiation Budget Network
UGA	<a href="http://www.weather.uga.edu">http://www.weather.uga.edu</a>	University of Georgia Weather Network
UOREGON	<a href="http://solardat.uoregon.edu">http://solardat.uoregon.edu</a>	University of Oregon Solar Radiation Monitoring Laboratory
USA (aka CHILI)	<a href="http://chiliweb.southalabama.edu">http://chiliweb.southalabama.edu</a>	South Alabama Mesonet
USCRN	<a href="https://www.ncdc.noaa.gov/crn">https://www.ncdc.noaa.gov/crn</a>	U.S. Climate Reference Network

<sup>1</sup>See Table 1 for all networks accessed via this source. <sup>2</sup>NREL data were not used in the making of the gridded datasets.

**Table S3.** Partial transmittance  $T_{i(n)}$  by cloud base height and coverage<sup>1</sup>. Adapted from Belcher and DeGaetano (2007).

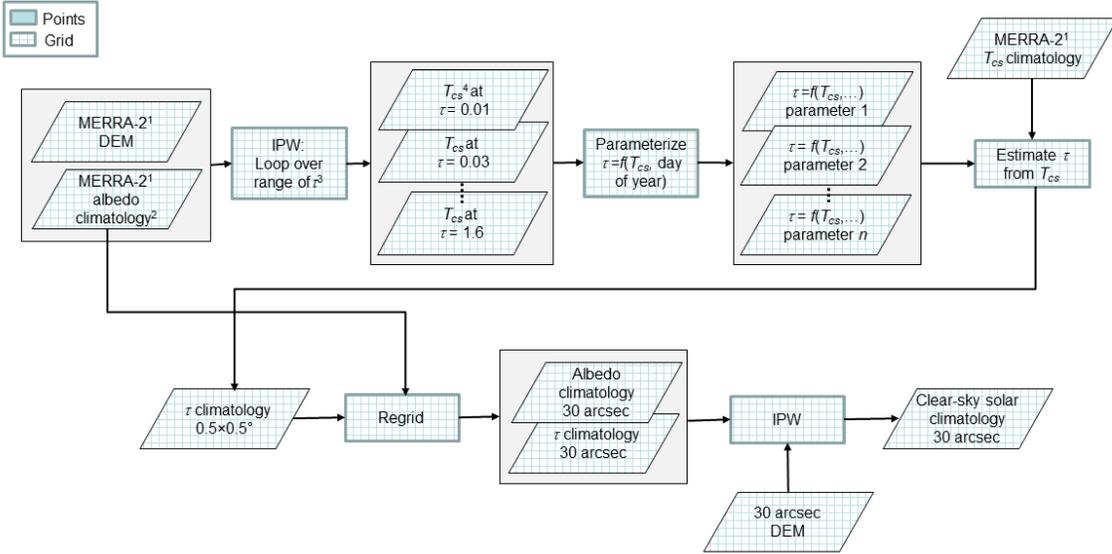
<b>Height (m)</b>	<b>FEW</b>	<b>SCT</b>	<b>BKN</b>	<b>OVC</b>
<i>Layer 1</i>				
0 – 610	0.79	0.73	0.64	0.30
610 – 1220	0.85	0.81	0.70	0.37
1220 – 1830	0.86	0.82	0.69	0.40
1830 – 2440	0.85	0.78	0.64	0.45
2440 – 3049	0.84	0.73	0.59	0.48
> 3049	0.77	0.68	0.57	0.53
<i>Layer 2</i>				
0 – 610	0.83	0.76	0.80	0.42
610 – 1220	0.87	0.90	0.84	0.50
1220 – 1830	0.87	0.90	0.81	0.54
1830 – 2440	0.84	0.86	0.75	0.56
2440 – 3049	0.87	0.87	0.74	0.59
> 3049	0.89	0.88	0.71	0.58
<i>Layer 3</i>				
0 – 610	0.92	0.92	0.67	0.49
610 – 1220	0.92	0.92	0.84	0.55
1220 – 1830	0.88	0.90	0.81	0.58
1830 – 2440	0.90	0.90	0.78	0.61
2440 – 3049	0.88	0.88	0.77	0.60
> 3049	0.87	0.87	0.74	0.60
<i>Layer 4 and up</i>				
Use values for Layer 3				
<i>Additional weather</i>				
Obscured	0.27			
Fog	0.86			
Haze	0.91			
Visibility < 16.1 km	0.95			

<sup>1</sup>FEW = few, SCT = scattered, BKN = broken, OVC = overcast

**Table S4.** Mean over 971 stations of the station mean difference (MD), station mean absolute difference (MAD), and station root mean squared difference (RMSD) between effective daily cloud transmittance  $T_c$  estimated at ASOS and  $T_c$  calculated from nearby GHI measurements before and after bias correction (BC) of ASOS  $T_c$ .

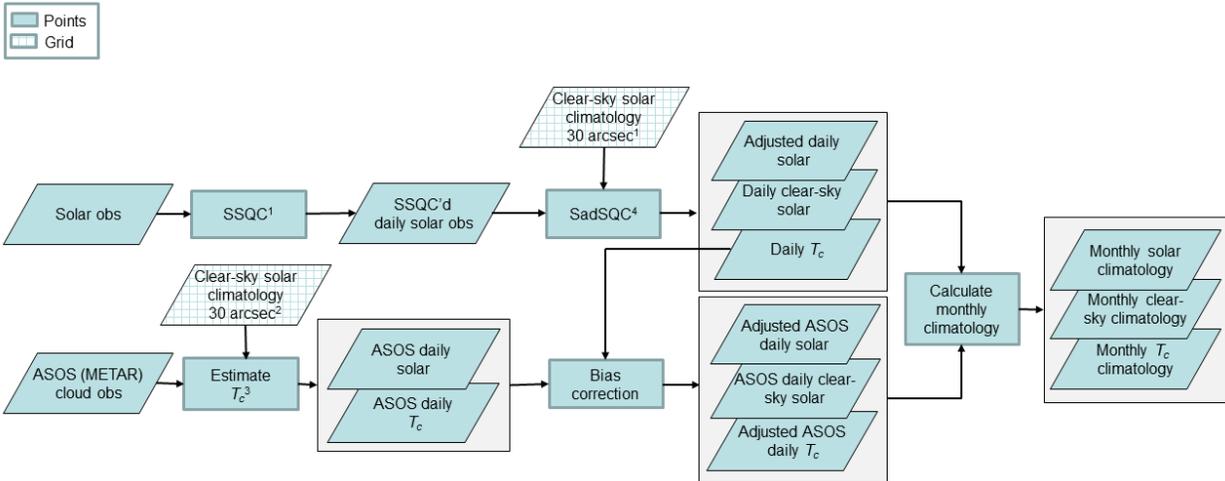
	<b>Mean MD</b>	<b>Mean MAD</b>	<b>Mean RMSD</b>
Before ASOS BC	0.028	0.118	0.160
After ASOS BC	-0.002	0.104	0.135

Part 1: Generating gridded 30 arc-sec clear-sky solar radiation daily climatology



<sup>1</sup>Regridded from 0.625°x0.5° to 0.5°x0.5°; <sup>2</sup>climatology = 1991-2020 daily climatology; <sup>3</sup> $\tau$  = effective clear-sky optical depth <sup>4</sup> $T_{cs}$  = clear-sky transmittance

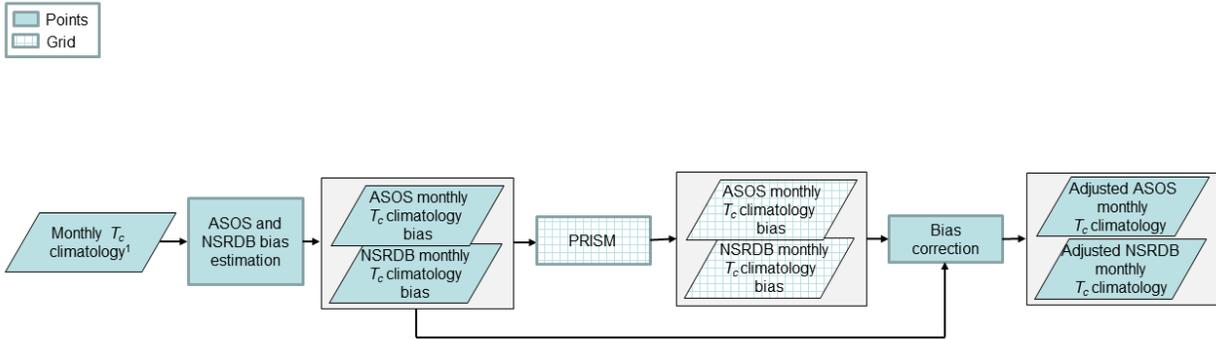
Part 2: Generating station solar radiation daily climatology



<sup>1</sup>SSQC = single-station quality control; <sup>2</sup>From Part 1; <sup>3</sup> $T_c$  = effective cloud transmissivity; <sup>4</sup>SadSQC = solar adjustment and solar QC

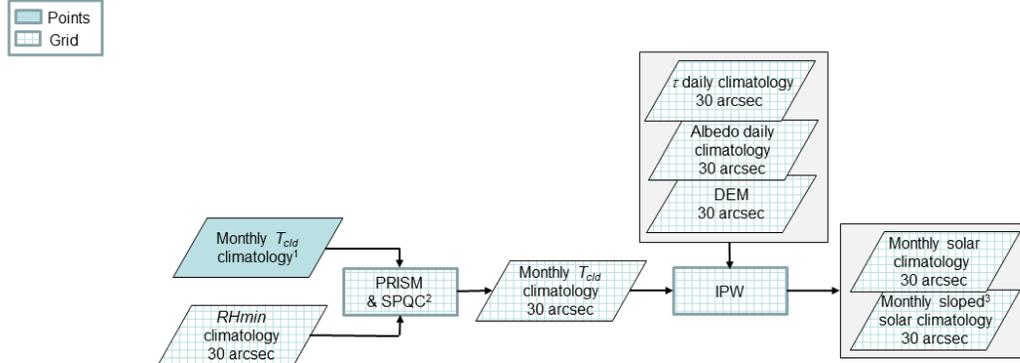
**Fig. S1.** Expanded workflow diagram for data processing and mapping of global irradiance, Parts 1 and 2.

Part 3: Reducing regional biases in ASOS and NSRDB monthly  $T_{cid}$  climatology



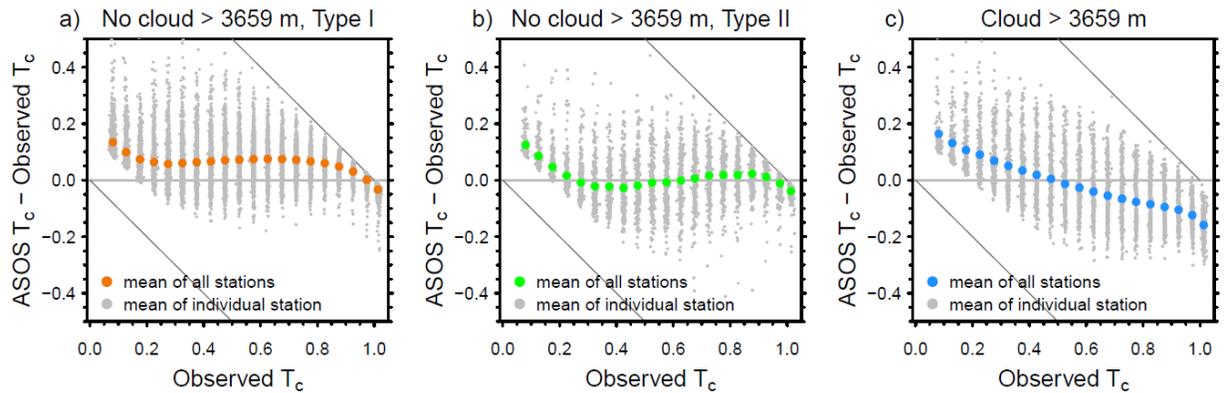
<sup>1</sup>From Part 2, <sup>2</sup> $RH_{min}$  = daily minimum relative humidity

Part 4: Generating gridded 30 arc-sec solar radiation daily climatology

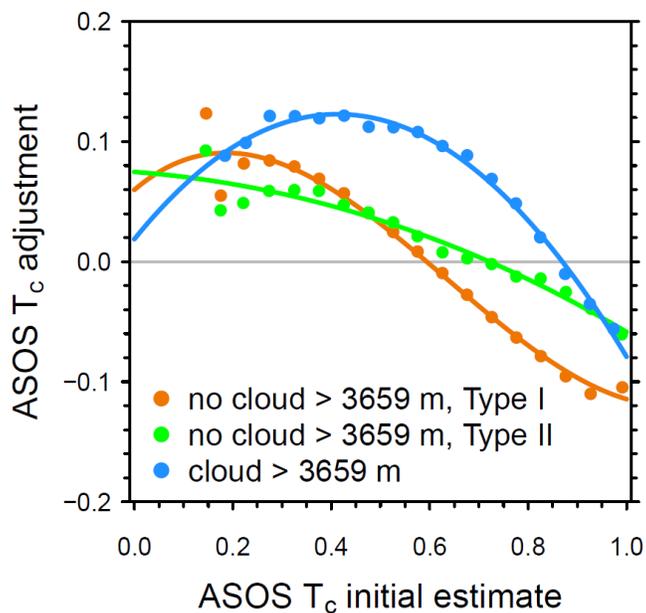


<sup>1</sup>From Part 2, includes ASOS and NSRDB adjusted in Part 3; <sup>2</sup>SPQC = spatial QC; <sup>3</sup>sloped = normal to ground surface

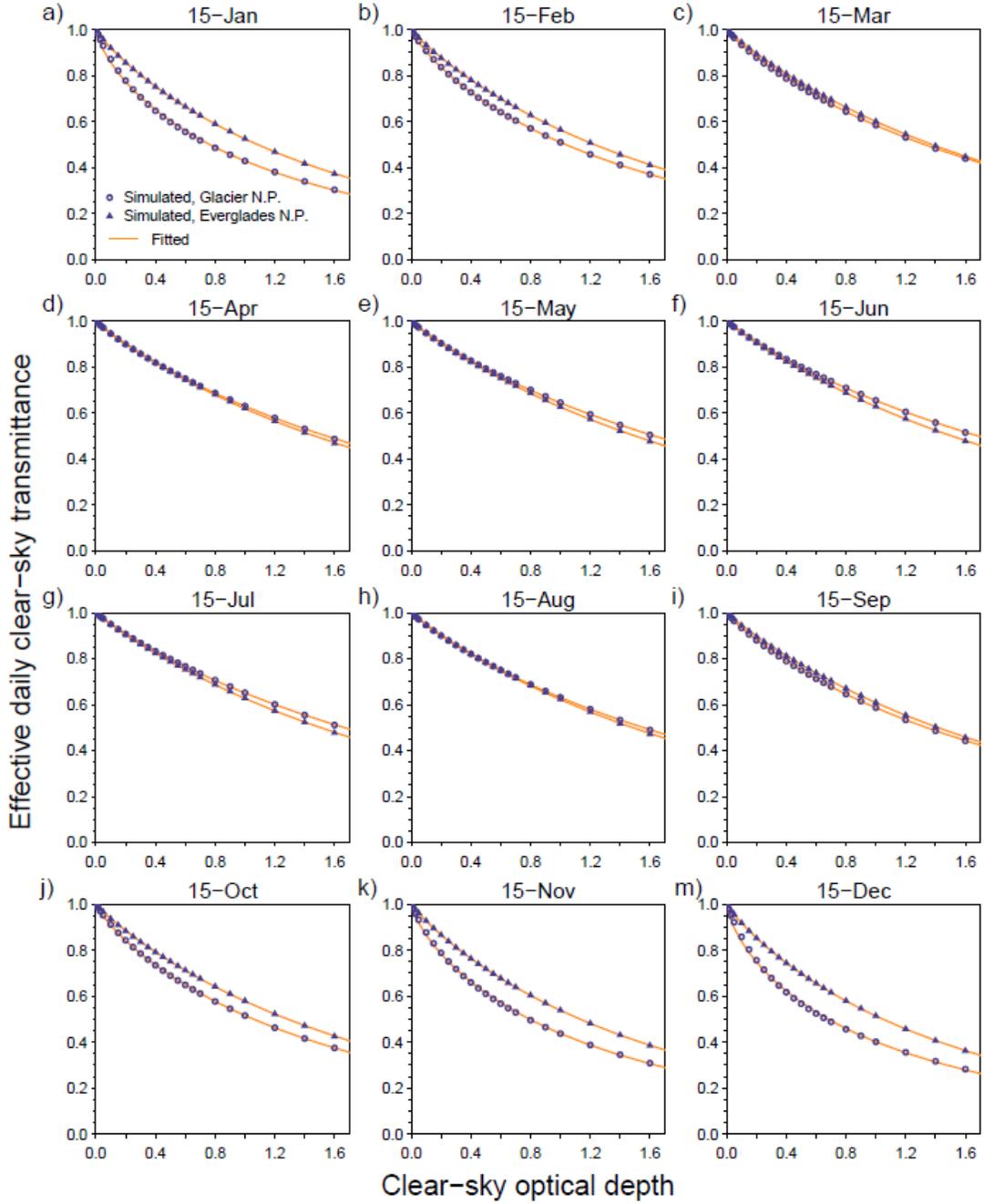
**Fig. S2.** Expanded workflow diagram for data processing and mapping of global irradiance, Parts 3 and 4.



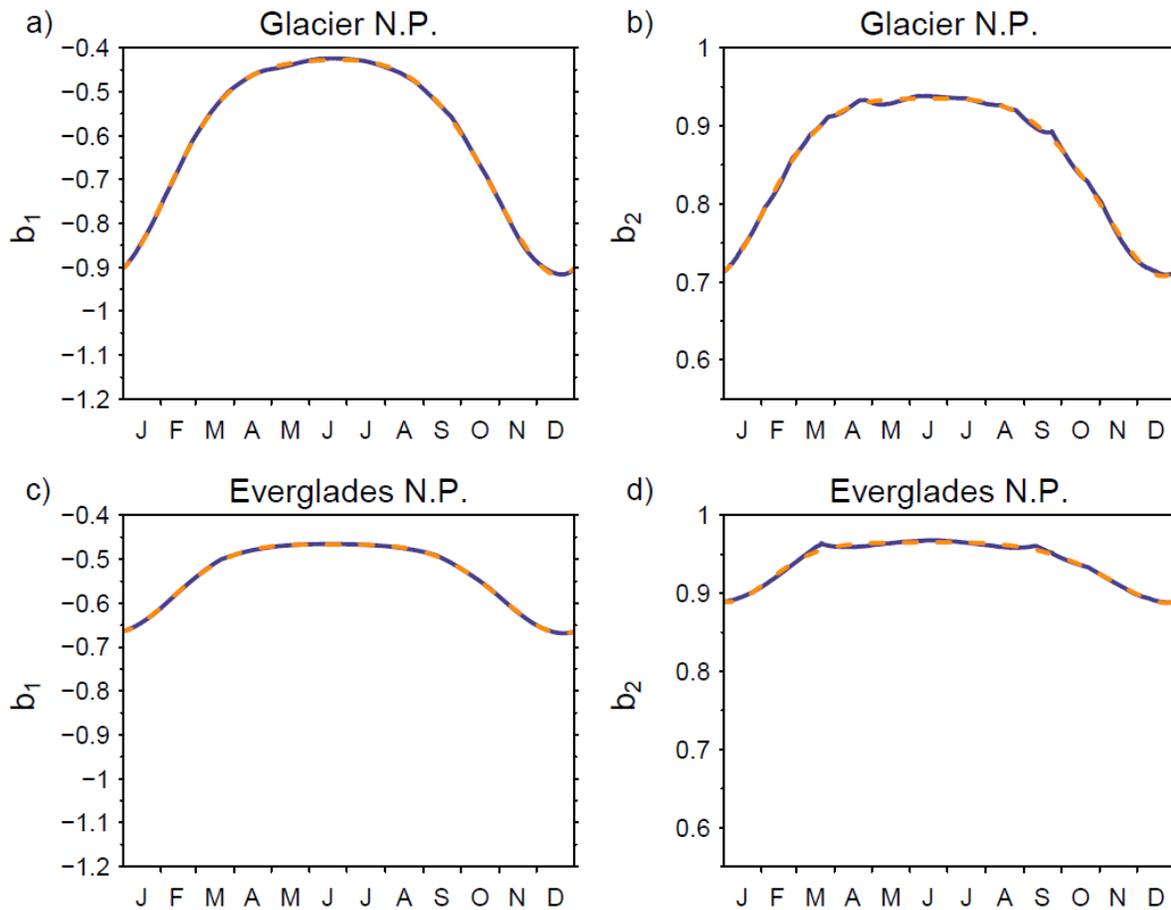
**Fig. S3.** Station bias in estimated ASOS cloud transmittance  $T_c$  as a function of observed  $T_c$  on **(a)** days when no clouds above 3659 m were reported during daylight hours and the station did not regularly report clouds above 3659 m (Type I station), **(b)** on days when no clouds above 3659 m were reported during daylight hours and the station did regularly report clouds above 3659 m (Type II station), **(c)** and on days when clouds above 3659 m were reported during daylight hours. Bias estimates were based on comparison of ASOS  $T_c$  with  $T_c$  calculated from nearby stations with GHI observations (i.e., solar stations). Filled gray circles show ASOS  $T_c$  minus solar station  $T_c$  against solar station  $T_c$  averaged in 0.05-wide bins of solar station  $T_c$  at each of 2,868 station pairs. Filled colored circles show the means of all 2,868 station means. Left and right diagonal lines show lower and upper bounds, respectively, of possible values.



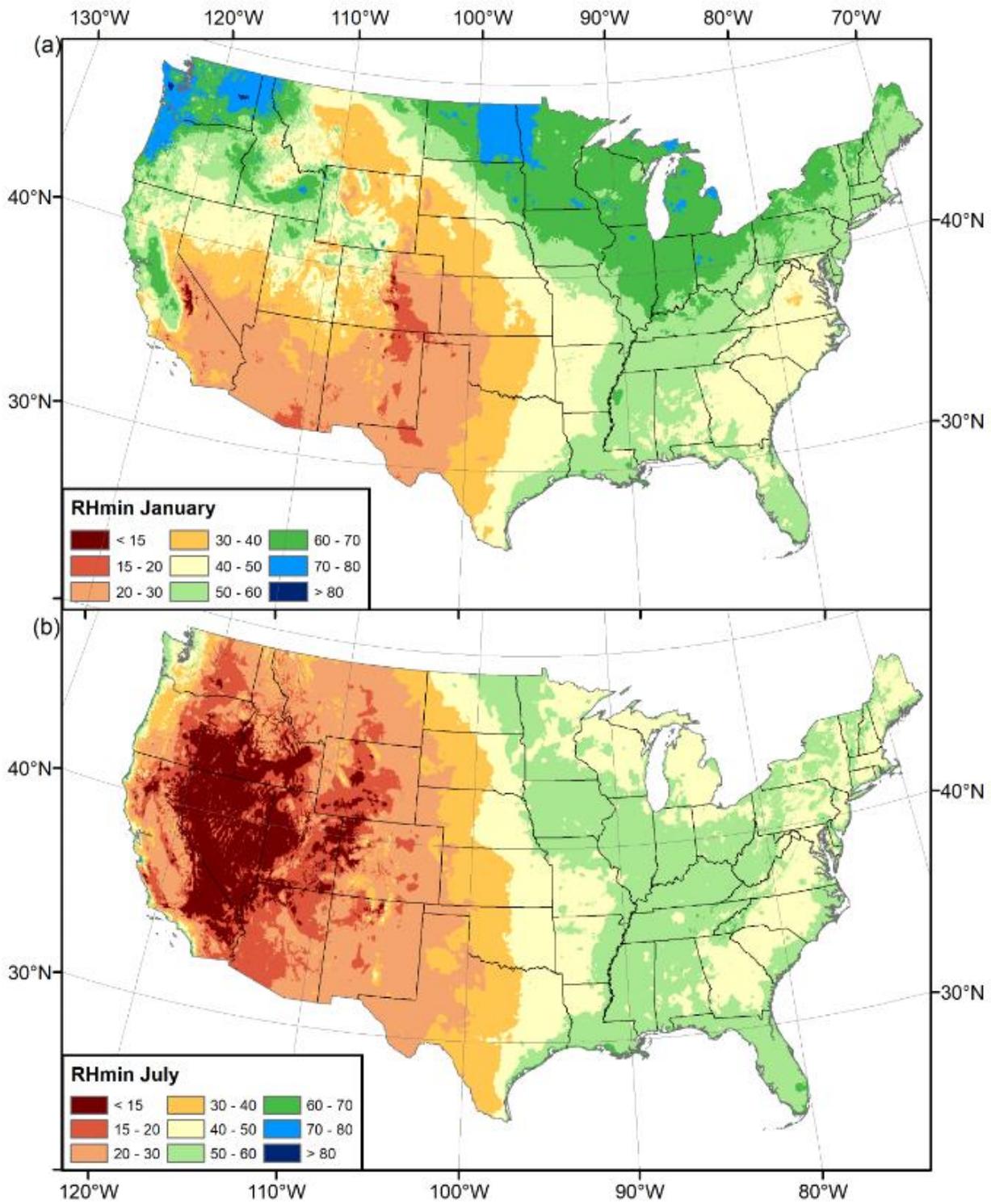
**Fig. S4.** Additive adjustment to initial estimate of ASOS cloud transmittance  $T_c$  as a function of the initial estimate of ASOS  $T_c$  (colored lines) on days when no clouds above 3659 m were reported during daylight hours and the station did not regularly report clouds above 3659 m (Type I station; orange), on days when no clouds above 3659 m were reported during daylight hours and the station did regularly report clouds above 3659 m (Type II station; green), and on days when clouds above 3659 m were reported during daylight hours (blue). Adjustment was based on comparison of ASOS  $T_c$  with  $T_c$  calculated from nearby stations with GHI observations (i.e., solar stations). Filled circles show solar station  $T_c$  minus ASOS  $T_c$  against ASOS  $T_c$  averaged in 0.05-wide bins of ASOS  $T_c$  and averaged over 971 station pairs.



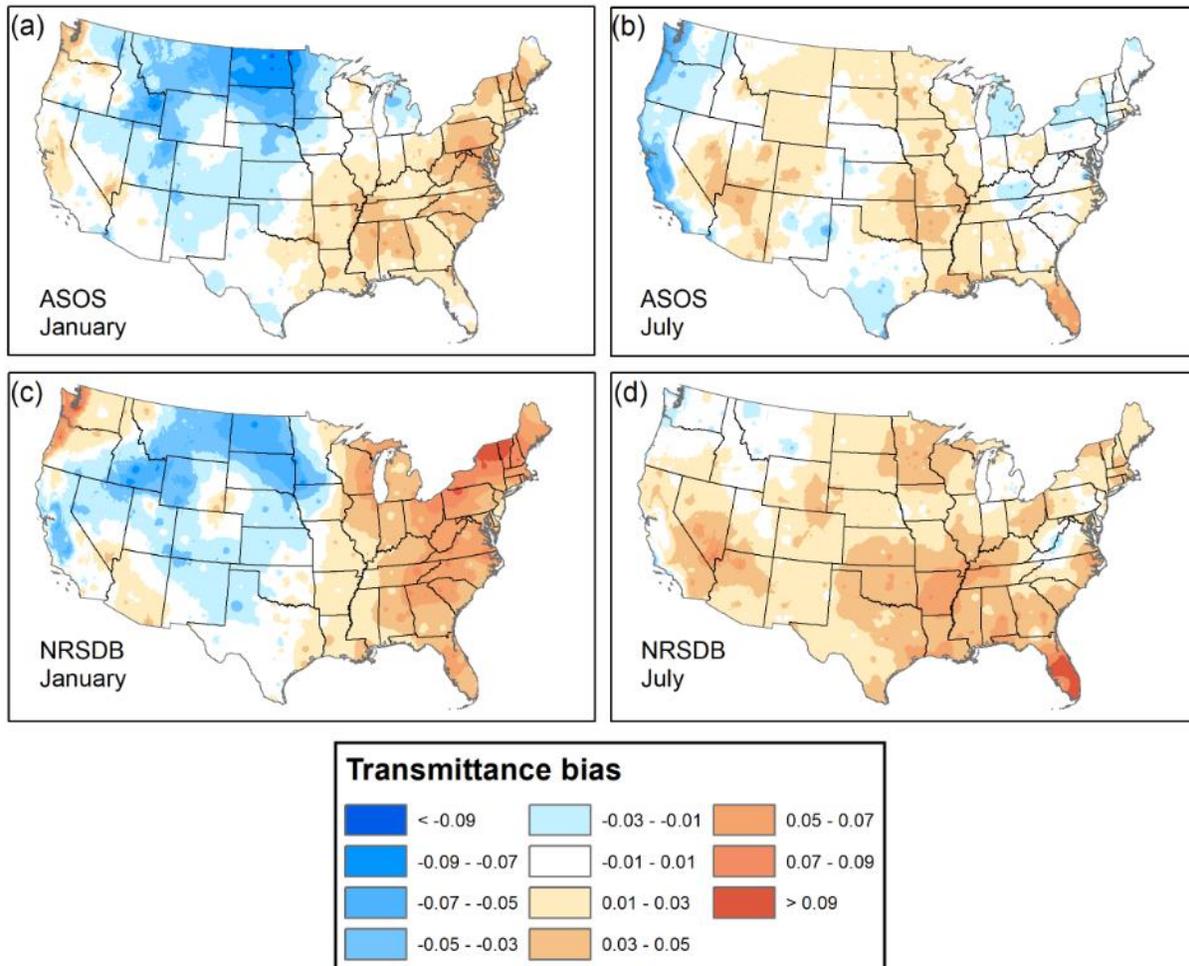
**Fig. S5. (a-m)** Effective daily clear-sky transmittance  $T_{\text{clear\_sky}}$  on the 15<sup>th</sup> day of the month simulated with IPW as a function of prescribed clear-sky vertical optical depth  $\tau_{\text{clear\_sky}}$  on the MERRA-2 grid at  $49^\circ$  N,  $114^\circ$  W (near Glacier National Park, Montana) and  $25.5^\circ$  N,  $81^\circ$  W (near Everglades National Park, Florida). The fitted line is Eq. (5).



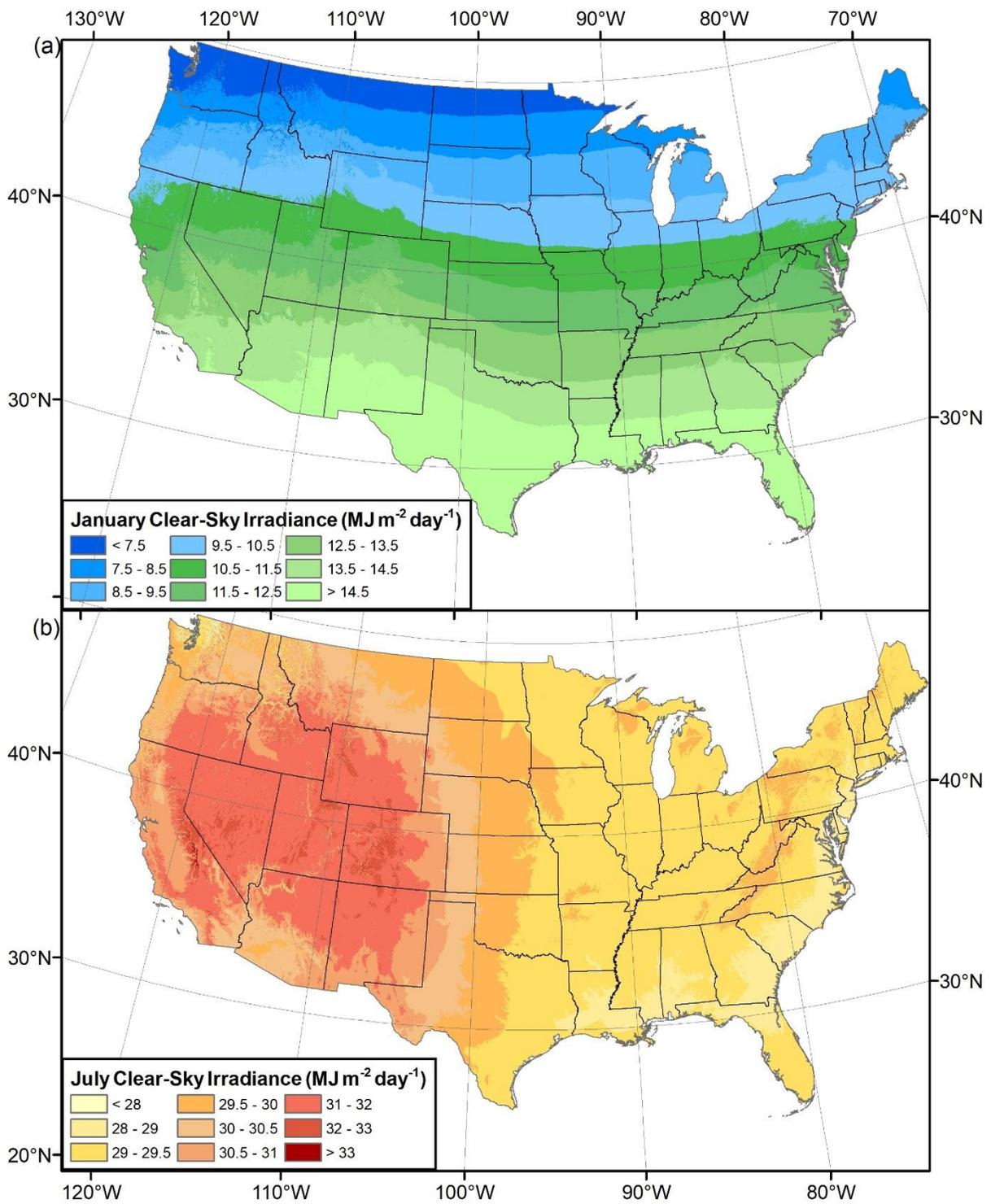
**Fig. S6.** Coefficients **(a, c)**  $b_1$  and **(c, d)**  $b_2$  from Eq. 5 fitted to IPW results (solid line) as a function of day of year for the MERRA-2 grid cells at **(a, b)** 49° N, 114° W (near Glacier National Park, Montana) and **(c, d)** 25.5° N, 81° W (near Everglades National Park, Florida). The dashed line shows an 8<sup>th</sup>-order polynomial fit for each coefficient as a function of day the year.



**Fig. S7.** Mean (a) January and (b) July daily minimum relative humidity  $RH_{min}$  (%) from PRISM.



**Fig. S8.** Spatially interpolated mean (a, c) January and (b, d) July effective cloud transmittance  $T_c$  bias in (a, b) ASOS and (c, d) NRSDB point estimates.



**Fig. S9.** Modeled mean (a) January and (b) July clear-sky GHI.