DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Reference herein to any social initiative (including but not limited to Diversity, Equity, and Inclusion (DEI); Community Benefits Plans (CBP); Justice 40; etc.) is made by the Author independent of any current requirement by the United States Government and does not constitute or imply endorsement, recommendation, or support by the United States Government or any agency thereof.

The National Solar Radiation Database Final Report: Fiscal Years 2022-2024

Manajit Sengupta, Yu Xie, Jaemo Yang, Aron Habte, Grant Buster, Brandon Benton, Galen Maclaurin, Paul Edwards, and Nicholas Gilroy

National Renewable Energy Laboratory

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated under Contract No. DE-AC36-08GO28308

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Technical Report
NREL/TP-5D00-90611
June 2025

The National Solar Radiation Database Final Report: Fiscal Years 2022-2024

Manajit Sengupta, Yu Xie, Jaemo Yang, Aron Habte, Grant Buster, Brandon Benton, Galen Maclaurin, Paul Edwards, and Nicholas Gilroy

National Renewable Energy Laboratory

Suggested Citation
Manajit, Sengupta, Yu Xie, Jaemo Yang, Aron Habte, Grant Buster, Brandon Benton, Galen Maclaurin, Paul Edwards, and Nicholas Gilroy. 2025. *The National Solar Radiation Database Final Report: Fiscal Years 2022–2024*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5D00-90611. https://www.nrel.gov/docs/fy25osti/90611.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated under Contract No. DE-AC36-08GO28308

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Technical Report
NREL/TP-5D00-90611
June 2025

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored by the National Renewable Energy Laboratory for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

Acknowledgments

The authors acknowledge Dr. Tassos Golnas, Dr. Julia Lee, and Dr. Guohui Yuan, the technology managers and the program manager, respectively, of the Systems Integration program of the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office, for their support of this project. The authors also acknowledge Dr. Michael Foster from the University of Wisconsin and Dr. Andrew Heidinger of the National Oceanic and Atmospheric Administration for the development of the satellite algorithms and for producing the cloud properties that are integral to the National Solar Radiation Database (NSRDB). Finally, the authors acknowledge Dr. Christian Gueymard for his contribution to the improvement of aerosols in the NSRDB.

List of Acronyms

ABQ	Albuquerque, New Mexico
AOD	aerosol optical depth
APE	absolute percentage error
ASO	Ashland, Oregon
AWS	Amazon Web Services
BIS	Bismarck, North Dakota
BON	Bondville, Illinois
BRW	Barrow
CONUS	contiguous United States
CSP	concentrating solar power
DHI	diffuse horizontal irradiance
DISC	Direct Insolation Simulation Code
DISC	Dynamic decomposition Improves Scaling Compute
DNI	direct normal irradiance
DRA	Desert Rock, Nevada
EUG	Eugene, Oregon
FARMS-DNI	Fast All-sky Radiation Model for Solar applications with DNI
FPK	Fort Peck, Montana
GHI	global horizontal irradiance
GML	Global Monitoring Laboratory
GOES	Geostationary Operational Environmental Satellite
GWN	Goodwin Creek, Mississippi
HNX	Hanford, California
MAE	mean absolute error
MBE	mean bias error
MERRA-2	Modern-Era Retrospective analysis for Research and Applications, Version 2
MODIS	Moderate Resolution Imaging Spectroradiometer
MSN	Madison, Wisconsin
NOAA	National Oceanic and Atmospheric Administration
NSA C1	North Slope of Alaska central location
NSA C2	North Slope of Alaska
NSRDB	National Solar Radiation Database
OLI	Oliktok Point
PHYGNN	physics-guided neural network
PSM	Physical Solar Model
PSU	Pennsylvania State University
PV	photovoltaics
RMSE	root mean square error
SAM	Systems Advisor Model
SEA	Seattle, Washington
SGP	Southern Great Plains
SIO	Silver Lake, Oregon
SLC	Salt Lake City, Utah
SOLRAD	Solar Radiation

STE	Sterling, Virginia
SURFRAD	the Surface Radiation Budget
SXF	Sioux Falls, South Dakota
TBL	Table Mountain, Colorado
TMY	typical meteorological year
TYP	typical plane-of-array year
UAF	University of Alaska at Fairbanks
UO	University of Oregon
WRF-Solar	Weather Research and Forecasting model with Solar extensions

Executive Summary

The National Solar Radiation Database (NSRDB) is the leading public source of high-resolution solar resource data in the United States, with more than 400,000 users annually. This database represents the state of the art in the satellite-based estimation of solar resource information and uses a unique physics-based modeling approach that enables improvements in accuracy with the deployment of the next-generation geostationary satellites. Making the highest-quality, state-of-the-art, regularly updated datasets available on a timely basis for users reduces costs of solar deployment by providing accurate information for siting studies and system output prediction and thereby reduces project financing costs and risks. Also, high-resolution information from the NSRDB enables moving beyond the levelized cost of energy when valuing the impact of renewables on the grid. Additionally, the NSRDB enables the integration of large amounts of solar on the grid by providing critical information about solar availability and variability that is used to enhance grid reliability and power quality.

In fiscal years 2022–2024, the NSRDB data were updated using the Physical Solar Model (PSM) version 4.0.0, which includes an improved gap-filling algorithm, based on machine learning techniques, for the missing cloud properties; corrections of solar position calculations; an enhanced algorithm for computing land-surface albedo over snow or ice surfaces; and a physics-based model for computing direct normal irradiance (DNI). The updated PSM was used to generate solar radiation data for 2021–2023, including global horizontal irradiance (GHI), DNI, and diffuse horizontal irradiance. This model was also employed to reprocess all previous NSRDB data from 1998–2020. The NSRDB was validated using high-quality measurements from surface sites, including the National Oceanic and Atmospheric Administration's Surface Radiation Budget (SURFRAD). According to data for 2019–2023, the NSRDB processed by the latest PSM has a mean bias error within 5% and 8% for the GHI and DNI, respectively.

In fiscal years 2025–2027, we will update the NSRDB annually to extend the data to cover the period from 2024–2026. The typical meteorological year (TMY) will also be updated annually to include additional years (2024–2026). Research and development efforts will focus on improving the PSM and enhancing the accuracy of the operational NSRDB product. The model and data will be validated using ground measurements to ensure accuracy. Due to limitations in the current geostationary satellites, the Alaska region has not been included in the NSRDB. To respond to this challenge, we have started collecting high-resolution satellite data that can provide continuous coverage over the area, and we are exploring the feasibility and pathway to extend the NSRDB. We are collaborating with the National Aeronautics and Space Administration Langley Research Center to explore available satellite data and cloud products suitable for extending the NSRDB. Long-term solar radiation data over Alaska will be developed and provided to the public in the next NSRDB update. In addition, the NSRDB website and servers will be upgraded to enable the delivery of the new datasets.

Table of Contents

List of Figures

List of Tables

Introduction

This project supports and enables the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office initiative in making solar energy more affordable for all Americans and facilitates the development of tens of gigawatts of solar power plants. Specifically, this project:

- Updates the National Solar Radiation Database (NSRDB) on an annual basis to extend the data to cover the period from 2021–2023
- Develops typical meteorological year (TMY) data to include additional years on an annual basis
- Updates the Physical Solar Model (PSM) to provide data of enhanced accuracy from using the next-generation Geostationary Operational Environmental Satellite (GOES)
- Updates the NSRDB website and servers to enable delivery of the new datasets
- Validates the updated NSRDB datasets with ground measurements to ensure accuracy of the new data
- Delivers complete, long-term, high-resolution data as a public resource for all users
- Conducts research and development to improve the NSRDB and develop advanced products.

The NSRDB, with 26 years of data (1998–2023), improves the accuracy of the prediction of plant output, reduces the cost of the development of solar projects, and enables achieving the goals of reducing the cost of photovoltaics (PV) and concentrating solar power (CSP). In addition, the use of higher-temporal resolution and higher-spatial-resolution data to produce the NSRDB provides improved information about the variability in solar radiation, which greatly reduces uncertainty in the NSRDB and improves plant operations in the long term. Finally, the continued development of 5-minute, 2-km data significantly enhances the awareness of the grid and provides information that is crucial to achieve the grid integration goals of the Solar Energy Technologies Office.

1 Task 1: Acquisition of Satellite and Ancillary Datasets

1.1 Project Approach

- Acquire calibrated and navigated geostationary satellite data for 2021–2023 from two satellites (GOES-16 and GOES-17) and calculate the cloud properties using the satellite information. The cloud properties will be calculated at a 2-km by 2-km resolution and will cover the contiguous United States (CONUS) at a 2-km by 2-km spatial and 5-minute temporal resolution. Full-disk coverage will have the same spatial and 10-minute temporal resolution covering North America, Central America, and South America from 60 N to 60 S latitudes.
- Acquire ancillary data from the National Aeronautics and Space Administration's Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) reanalysis for aerosols, water vapor, temperature, humidity, wind speed, and wind direction. MERRA-2 data are available at a 0.5-degree resolution on an hourly basis.
- Acquire snow-free surface albedo data from the University of Massachusetts, Boston if updates are available.
- Acquire snow data from the National Ice Center's Interactive Multisensor Snow and Ice Mapping System. The snow data are used as an overlay on the snow-free albedo data and replace the albedo by the albedo of snow.

1.2 Project Results and Discussion

Under a subcontract, the University of Wisconsin completed the delivery of cloud property datasets for GOES-16 and GOES-17 for 2021–2023. The calibrated and navigated geostationary data from the two satellites (GOES-16 and GOES-17) were acquired and stored in the National Renewable Energy Laboratory's (NREL's) high-performance computing system. The sample data for cloud optical thickness on July 1, 2021, July 1, 2022, and July 1, 2023, are demonstrated in Figure 1. The spatial and temporal resolutions of the data are 2 km and 5 minutes, respectively. The cloud optical thickness combined using GOES-16 and GOES-17 is displayed in Figure 2.

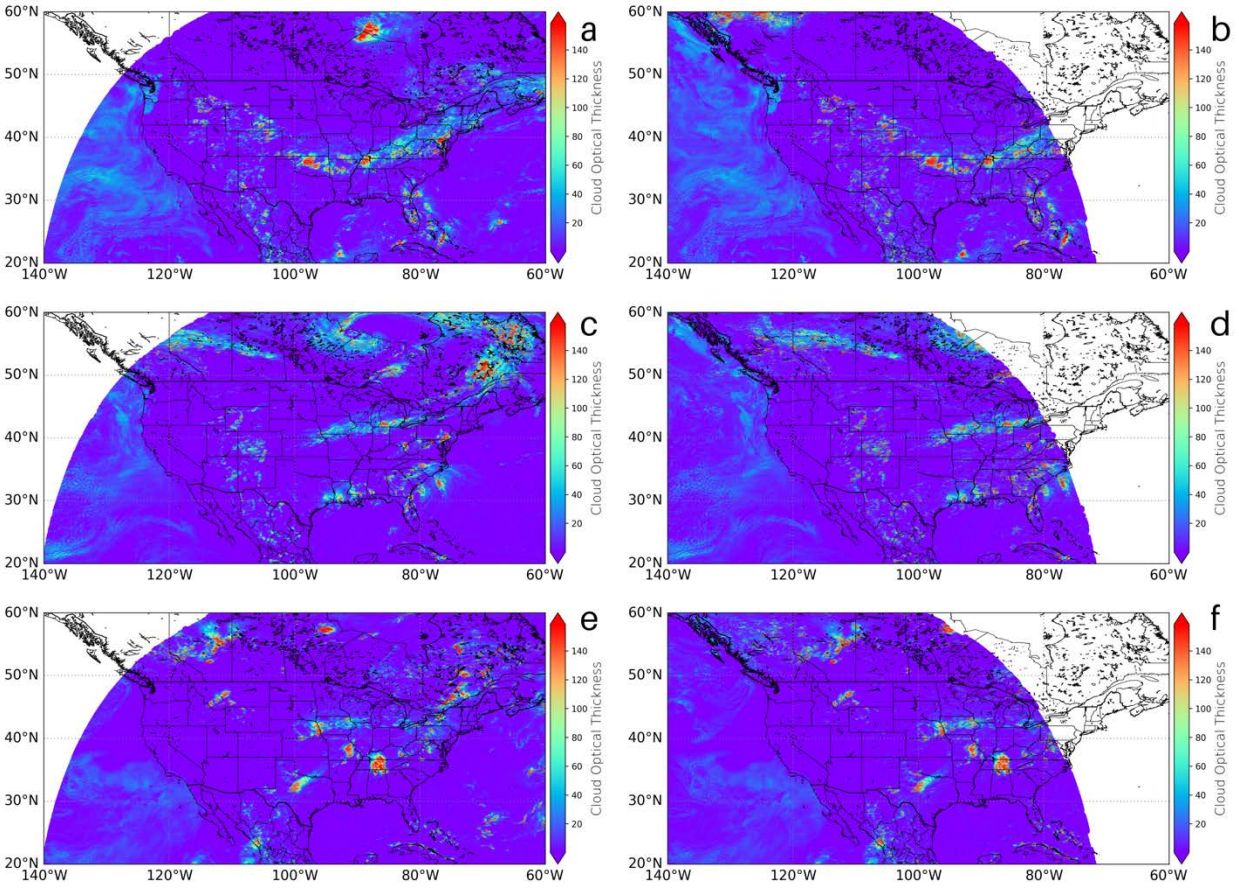

Figure 1. Cloud optical thickness from GOES-16 on (a) July 1, 2021, (c) July 1, 2022, and (e) July 1, 2023, and from GOES-17 on (b) July 1, 2021, (d) July 1, 2022, and (f) July 1, 2023

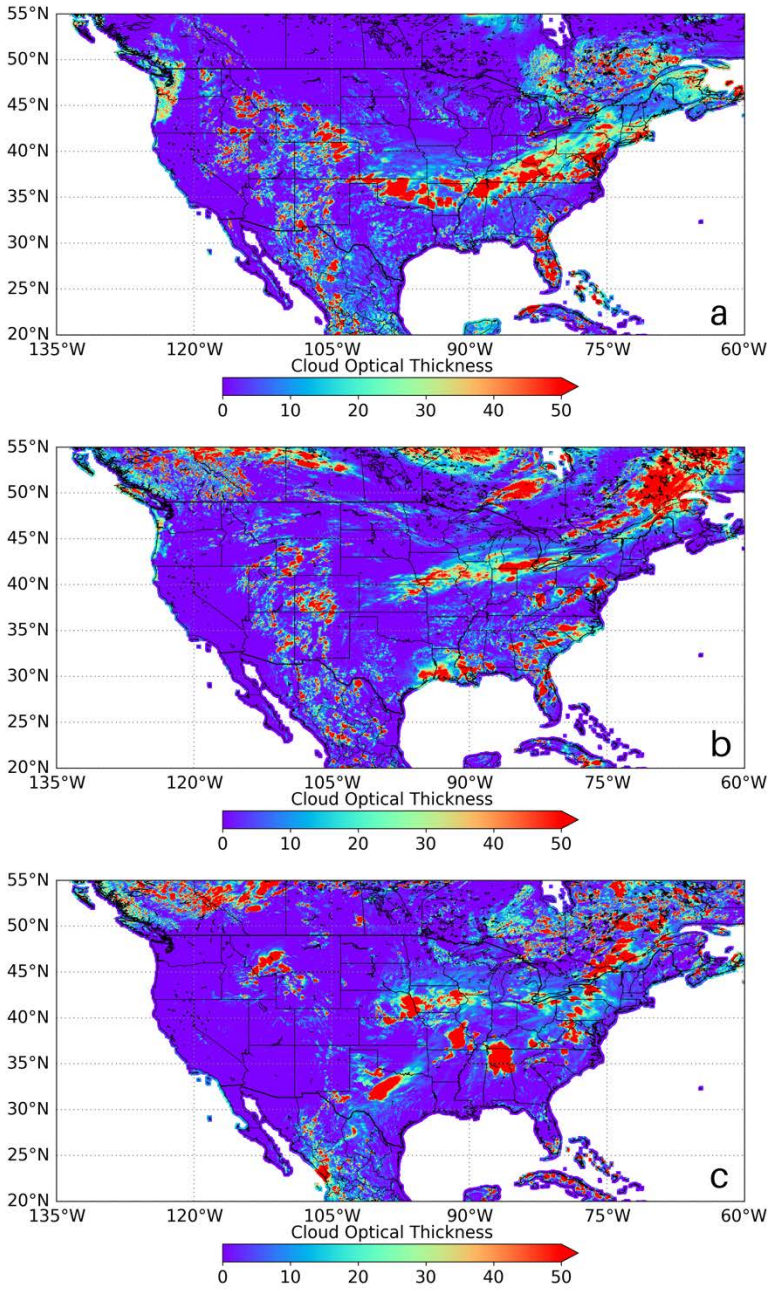

Figure 2. Cloud optical thickness on (a) July 1, 2021, (b) July 1, 2022, and (c) July 1, 2023

The process of acquiring the 2021–2023 MERRA-2 dataset for use in PSM was completed. The ancillary data for 2021–2023—including aerosols, water vapor, temperature, humidity, wind speed, and wind direction—from the MERRA-2 product were acquired and stored in NREL's high-performance computing system. Figure 3 illustrates sample data for aerosol optical depth on July 1, 2021, July 1, 2022, and July 1, 2023. Figure 4 illustrates sample data for precipitable water vapor on July 1, 2021, July 1, 2022, and July 1, 2023. The spatial and temporal resolutions of the data are a half degree and 1 hour, respectively.

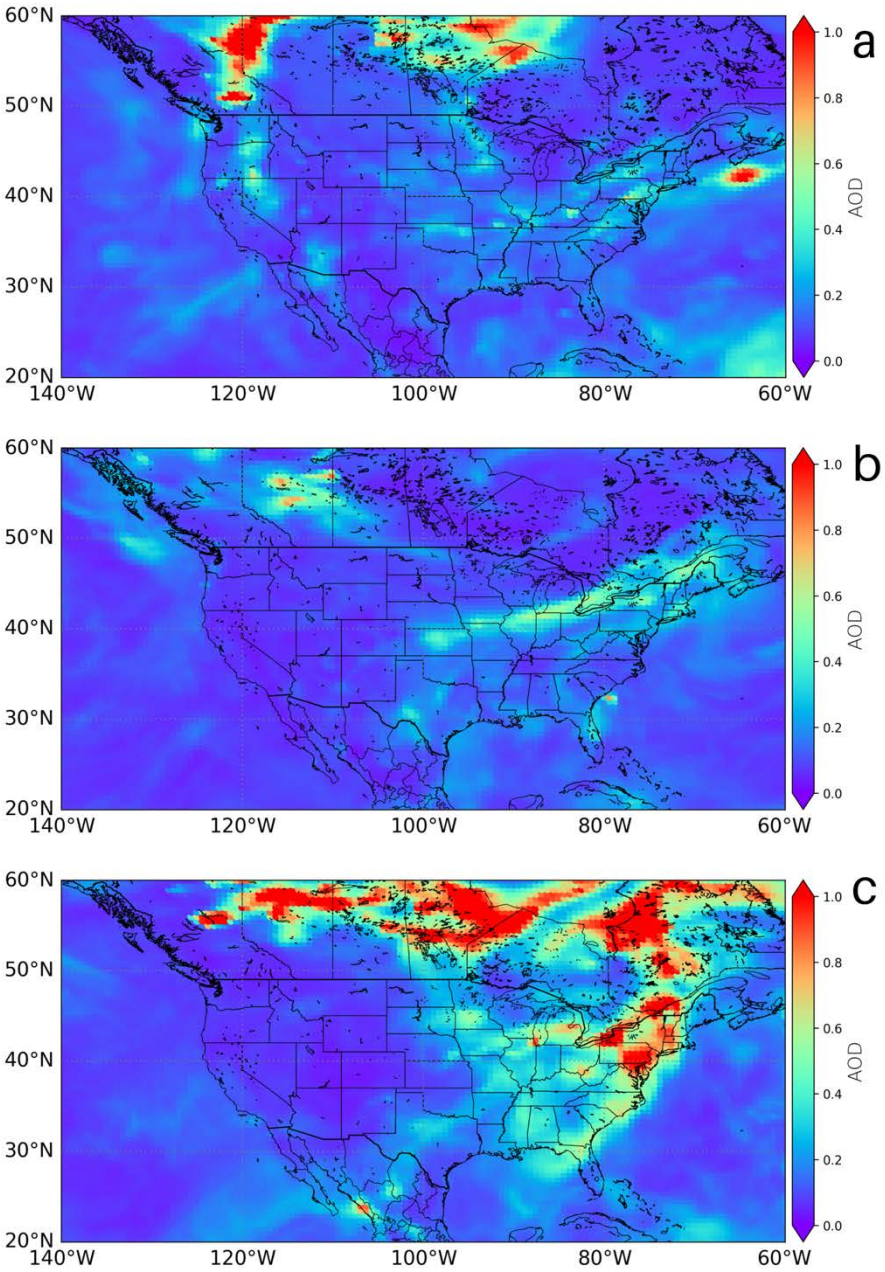

Figure 3. Aerosol optical depth (AOD) on (a) July 1, 2021, (b) July 1, 2022, and (c) July 1, 2023

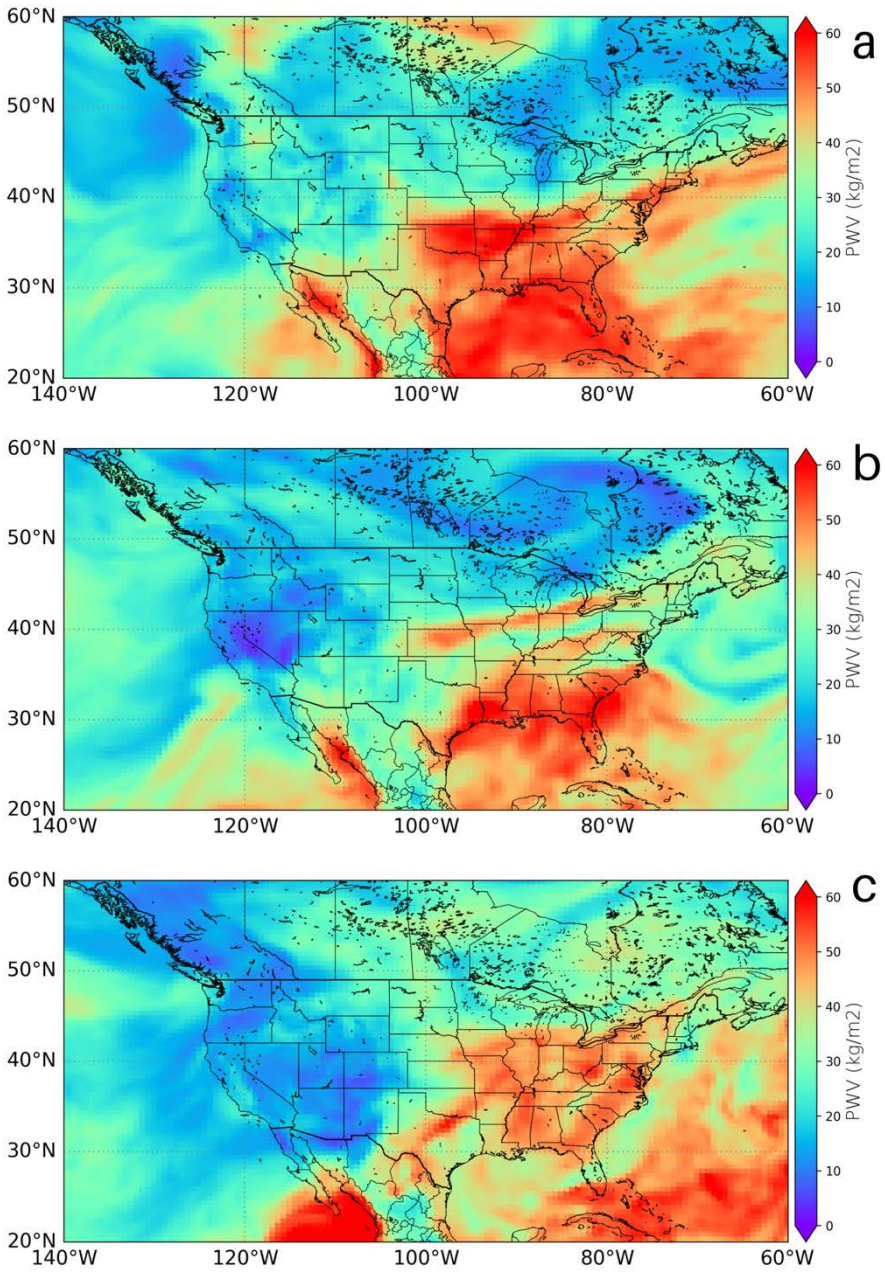

Figure 4. Precipitable water vapor on (a) July 1, 2021, (b) July 1, 2022, and (c) July 1, 2023

The surface albedo data, which contain snow and ice information data for 2021–2023 from the National Ice Center, were updated and stored in NREL's high-performance computing system. The surface albedo of snow and ice is calculated using the surface temperature and the empirical equations from Ross and Walsh (1987). The sample data for the surface albedo on Jan. 1, 2021, Jan. 1, 2022, and Jan. 1, 2023, are demonstrated in Figure 5. A sample of summer surface albedo is given in Figure 6. The spatial resolution of the daily data is approximately 1 km.

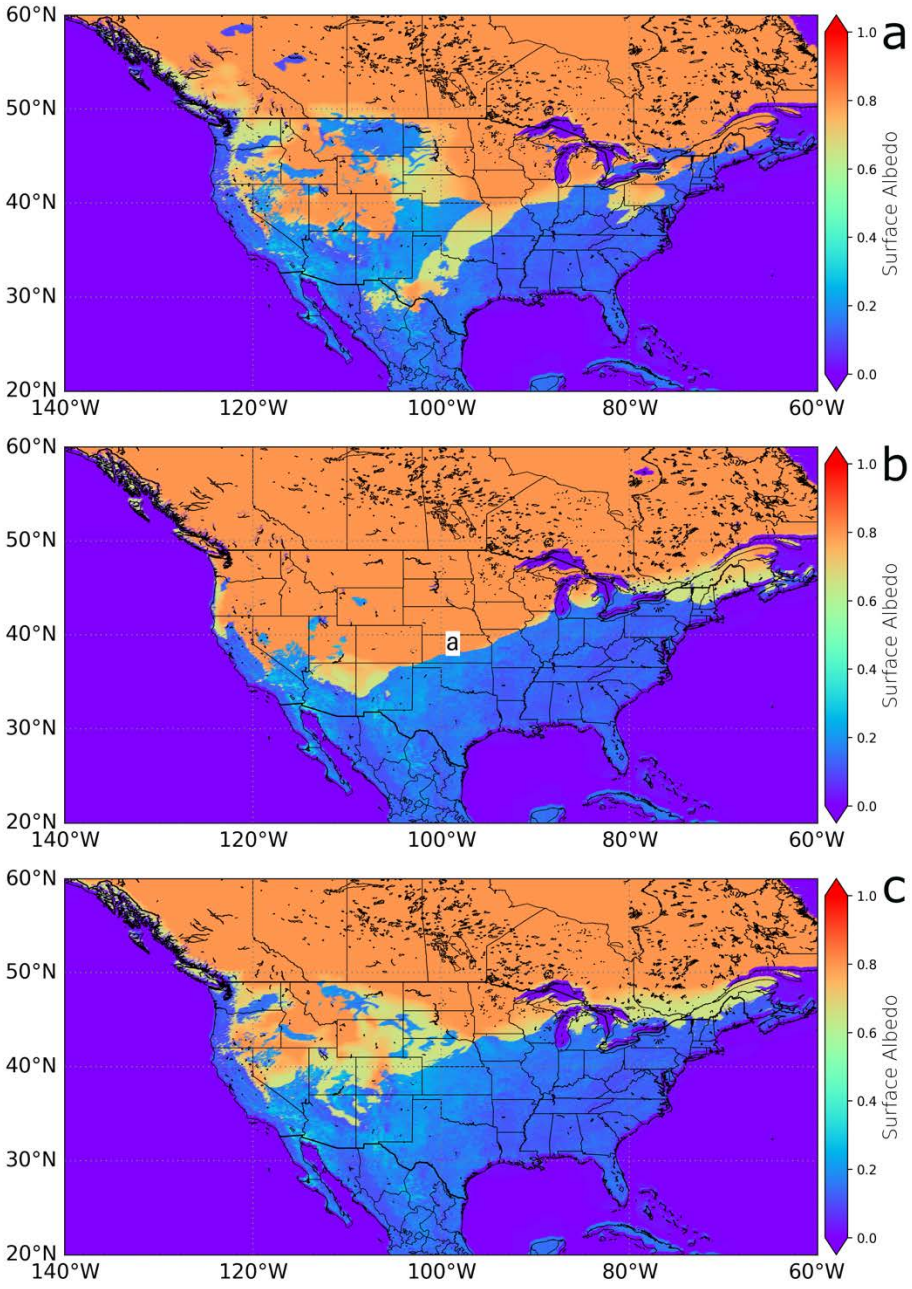

Figure 5. Surface albedo on (a) Jan. 1, 2021, (b) Jan. 1, 2022, and (c) Jan. 1, 2023

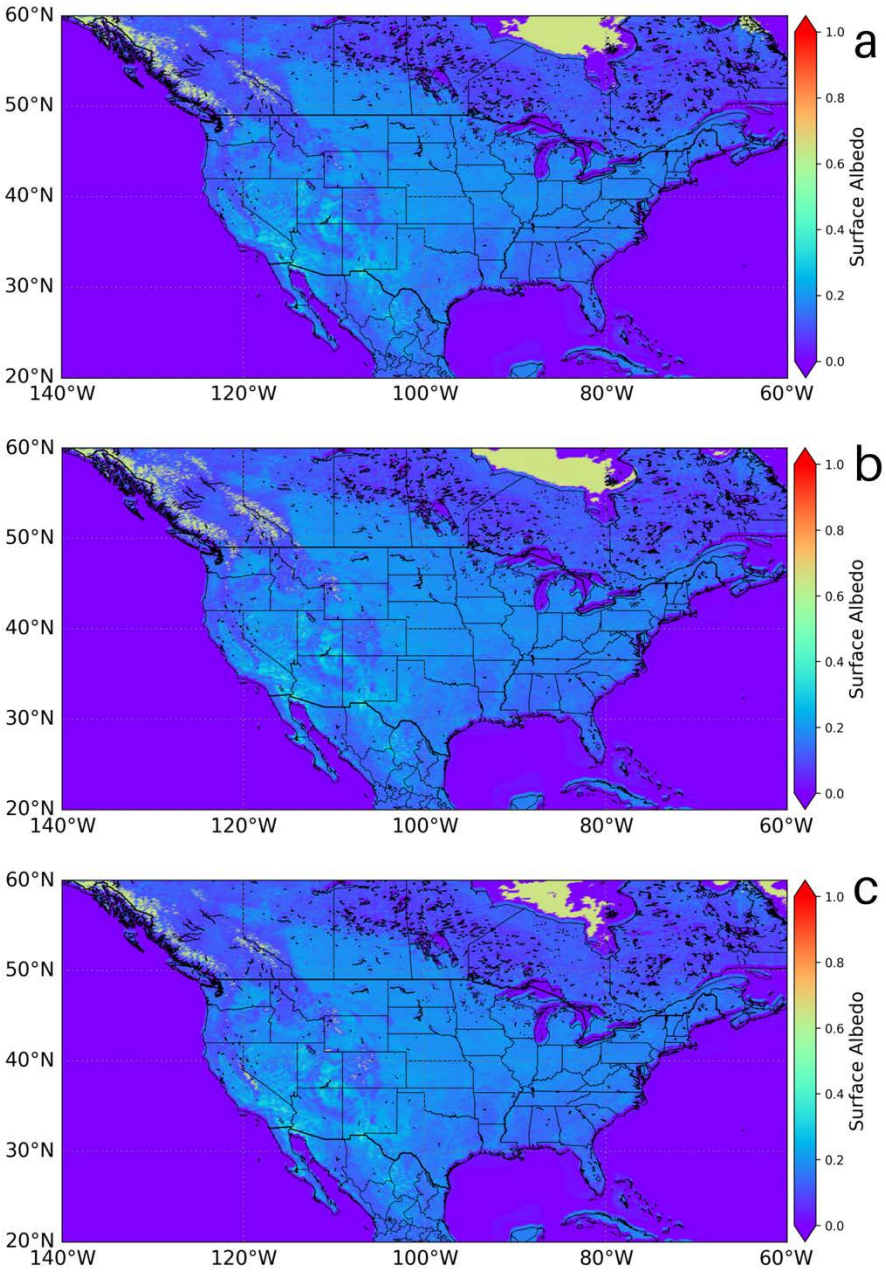

Figure 6. Surface albedo on (a) July 1, 2021, (b) July 1, 2022, and (c) July 1, 2023

2 Task 2: Development, Validation and Dissemination of Solar and Meteorological Data

2.1 Project Approach

2.1.1 Development

- The NSRDB is mapped to a fixed 2-km by 2-km and 4-km by 4-km grid that enables users to access time-series data for a pixel on the grid. The satellite scan can be slightly shifted between scans, and the calculated cloud properties from the satellite data are mapped to a static grid.
- The ancillary data from MERRA-2 will be upscaled to the NSRDB spatial and temporal resolution. Specific processes have been developed to maintain physical consistency while generating the high-resolution ancillary dataset that is mapped to the NSRDB static grid.
- The aerosol data will also be obtained from MERRA-2. These data will be upscaled to the high-resolution NSRDB grid using an elevation-based scaling factor. The AOD will be calibrated to remove bias.
- The physical consistency of the cloud properties and ancillary datasets will be checked to develop solar and meteorological products.
- Missing cloud properties and meteorological data will be filled to produce a serially complete dataset for computing solar radiation.
- The PSM will be run to produce solar radiation, including global horizontal irradiance (GHI), direct normal irradiance (DNI), and diffuse horizontal irradiance (DHI).
- The final output of the solar radiation and meteorological products will be checked for errors, physically consistency, and completeness.
- The GOES-16 and GOES-17 data are generated at a 2-km by 2-km spatial and 5-/10-minute temporal resolution. The resolution of the solar datasets will be averaged to 4 km by 4 km, half-hourly, to match the resolution of the long-term NSRDB data.
- As data are generated using the GOES-East and GOES-West satellites, the data are blended to provide seamless coverage for all of the United States.
- The TMY data will be developed using all years of data from 1998–2023. TMY data are updated every time a new year of data is produced.

2.1.2 Validation

- The satellite data are validated using high-quality ground data from the National Oceanic and Atmospheric Administration's (NOAA's) Surface Radiation Budget (SURFRAD), U.S. Department of Energy Atmospheric Radiation Measurement (ARM), and NREL stations. The 2021–2023 data for all the relevant locations will be acquired.
- A data quality assessment of the ground data will be performed because only high-quality data will be used to evaluate the NSRDB.
- The 2021–2023 NSRDB data will be validated using the ground-based measurement, and a report will be published and made publicly available.

2.1.3 Dissemination

- A capability to provide additional data through the NSRDB website will be developed by enhancing/updating servers.
- The web interface for delivering new datasets will be updated.
- The NSRDB website content will be updated to reflect changes, and new content will be added.
- The operational capability of the NSRDB website and servers will be maintained to ensure that the data are provided on a continuous basis.
- Support for NSRDB users will be provided mainly through responding to email queries. All queries are sent to nsrdb@nrel.gov.
- Data will be processed for upload to Amazon Web Services, and the documentation and methodology for serving data through Amazon Web Services will be updated.
- A webinar will be conducted to inform users of updates.

2.2 Project Results and Discussion

2.2.1 Development of the NSRDB Data

Based on the data from Task 1, we ran the PSM to generate solar radiation data for 2021–2023, including GHI, DNI, and DHI. We also reprocessed all previous NSRDB data from 1998–2020 using PSM v3.2.2. An improved gap-filling algorithm, based on machine learning techniques, for the missing cloud properties and a correction of solar position were used in the update. The NSRDB data were further updated using PSM v4.0.0 based on an improved algorithm to compute the DNI. Figure 7 presents sample GHI data at 20:00 UTC for July 1, 2021, July 1, 2022, and July 1, 2023. Similarly, Figure 8 shows sample DNI data for the same time.

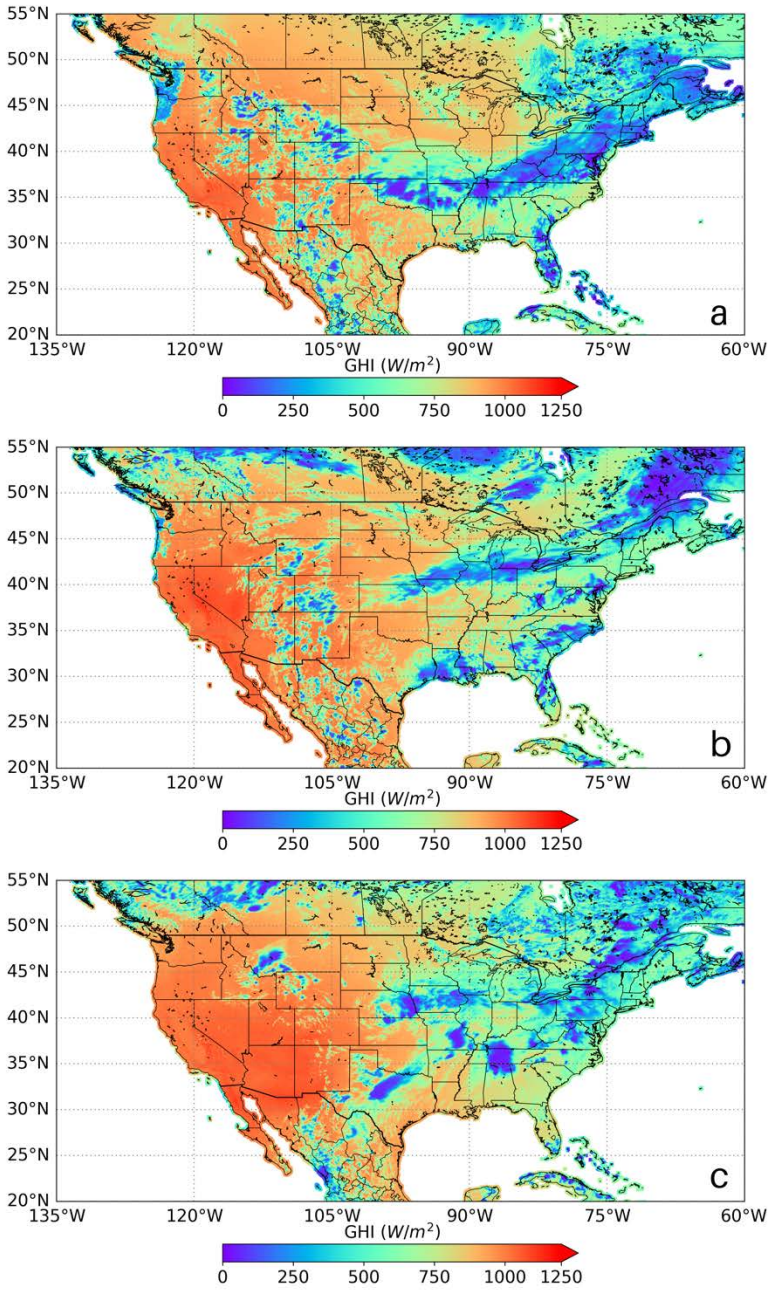

Figure 7. GHI on (a) July 1, 2021, (b) July 1, 2022, and (c) July 1, 2023

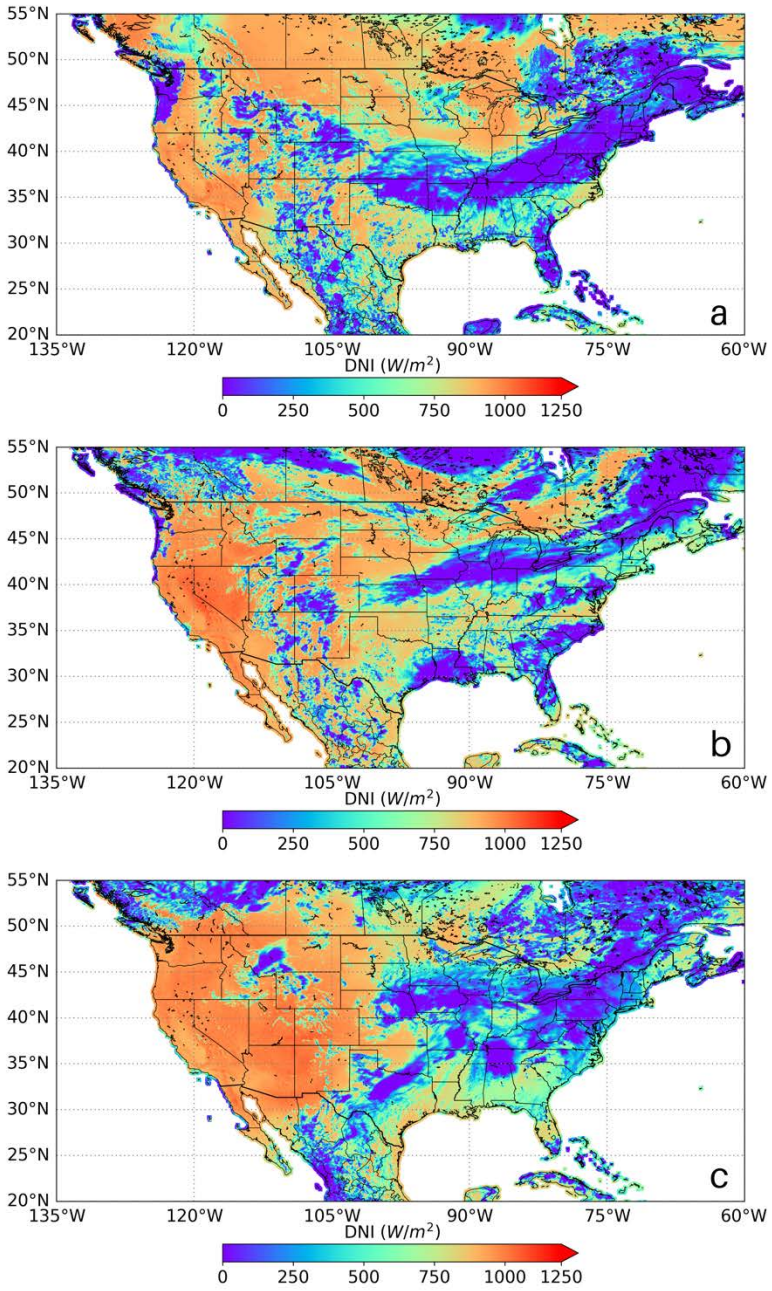

Figure 8. DNI on (a) July 1, 2021, (b) July 1, 2022, and (c) July 1, 2023

2.2.2 Validation of the NSRDB Data

The NSRDB was validated using high-quality measurements from 18 surface sites, including NOAA's Surface Radiation Budget (SURFRAD) (Hicks, DeLuisi, and Matt 1996) and Solar Radiation (SOLRAD) (Augustine and DeLuisi 2000), NREL's Solar Radiation Research Laboratory (SRRL) (Stoffel and Andreas 1981), the Solar Radiation Monitoring Laboratory at the University of Oregon (UO), and the ARM Southern Great Plains (SGP). The 18 surface sites are distributed across CONUS (Figure 9).

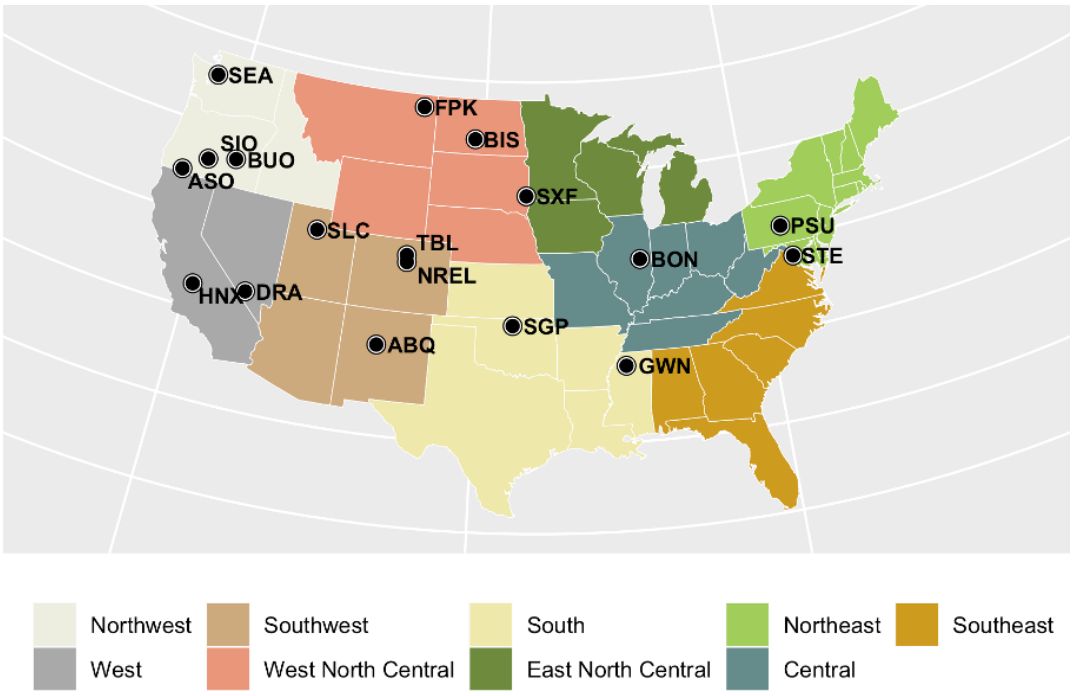

Figure 9. Surface stations used in the validation

To conduct a reasonable validation, data quality assessment and data-filtering techniques are implemented in the surface measurements and the NSRDB data. The following criteria are used to produce a quality-controlled dataset for this validation study: (1) The solar zenith angle is less than 80°; (2) the GHI and DNI are greater than 50 W/m^2; (3) the surface measurements and satellite-derived dataset are excluded when the surface measurements are missing; (4) the cloud types from the satellite data are used to determine the sky conditions; and (5) the surface measurements are averaged around each time step of the satellite data.

Statistics are used to investigate the performance of the NSRDB, including the mean bias error (MBE), the mean absolute error (MAE), and the root mean square error (RMSE), which are defined as follows:

$$\mathrm{MBE} = \frac{1}{n}\sum_{i=1}^{n}(F_{sat} - F_s) \tag{1}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|F_{sat} - F_s| \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(F_{sat} - F_s)^2}{n}} \tag{3}$$

where n is the total number of scenarios, F represents the solar radiation, and the subscripts “sat” and “s” denote the satellite-based solar resource dataset and surface-based observations, respectively.

The NSRDB demonstrated an overestimation (positive bias) compared to the surface data (Figure 10), which is consistent with our previous validation results. Higher RMSE and MAE are seen in some locations. This is partly because the NSRDB pixel represents a 4-km by 4-km area, whereas a ground-based station represents only a small area above the measuring station. For all-sky conditions, the MBEs in most scenes are within 5% and 10% for the GHI and DNI, respectively. The RMSE in the cloudy-sky DNI shows higher spatial differences than the GHI because the ground-based measurements are based on a pyrheliometer, with an approximate 5° field of view, representing a column through the atmosphere; however, the satellite DNI represents the 4-km by 4-km area. The subpixel variability and the misidentification of clouds by the satellites could lead to the higher RMSE and the overestimation in both GHI and DNI (Buster et al. 2021).

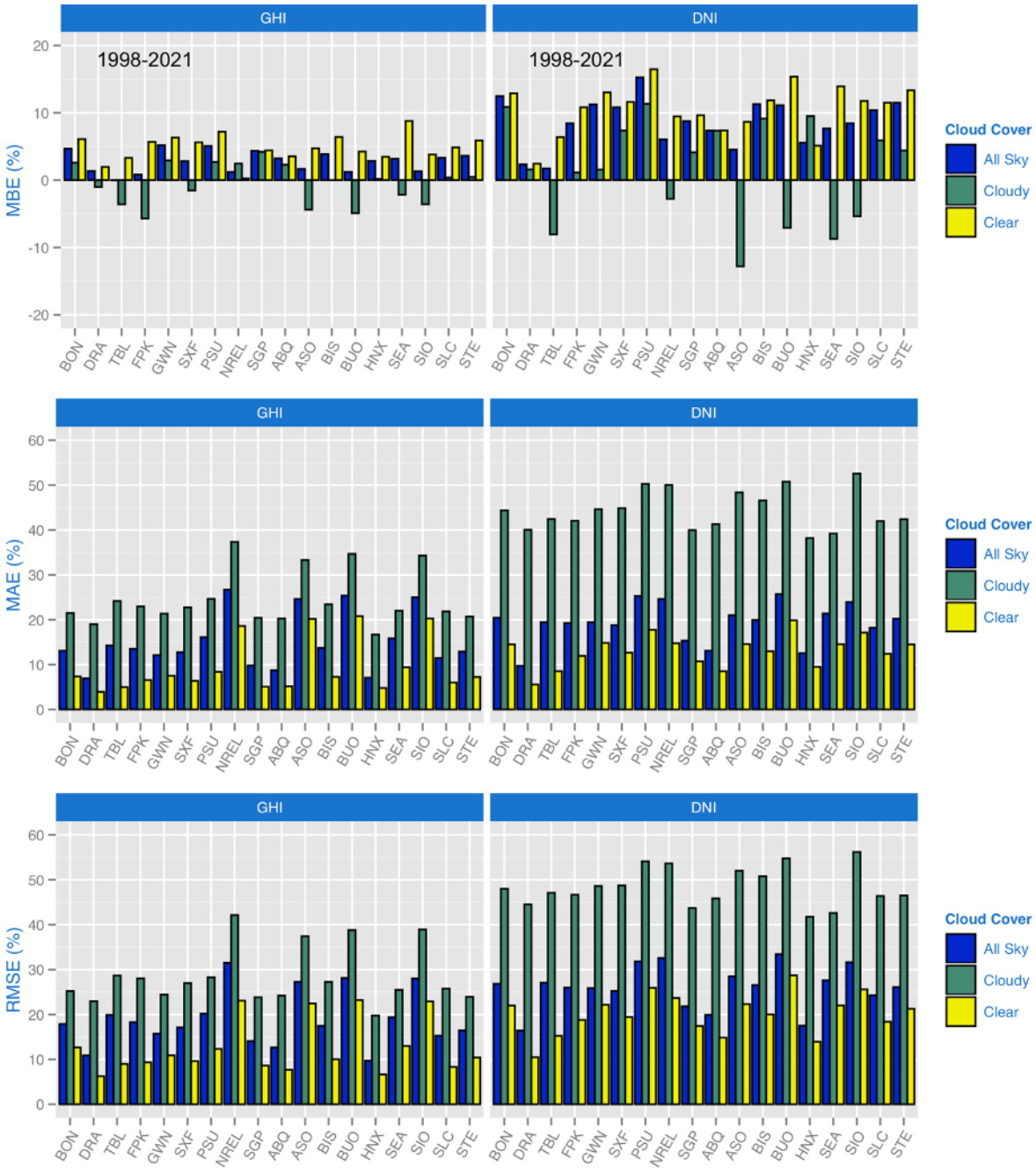

Figure 10. Statistic metrics for hourly averaged data from 1998–2021

Table 1 compares three NSRDB versions involved in the recent update. Figure 11 presents the statistical metrics (MBE, MAE and RMSE) of the GHI and DNI in the all-sky, clear-sky, and cloudy-sky conditions. The NSRDB v3.2.2 shows improvement in most SURFRAD locations and for most statistic metrics. Figure 12 demonstrates a sample time-series GHI plot for two locations.

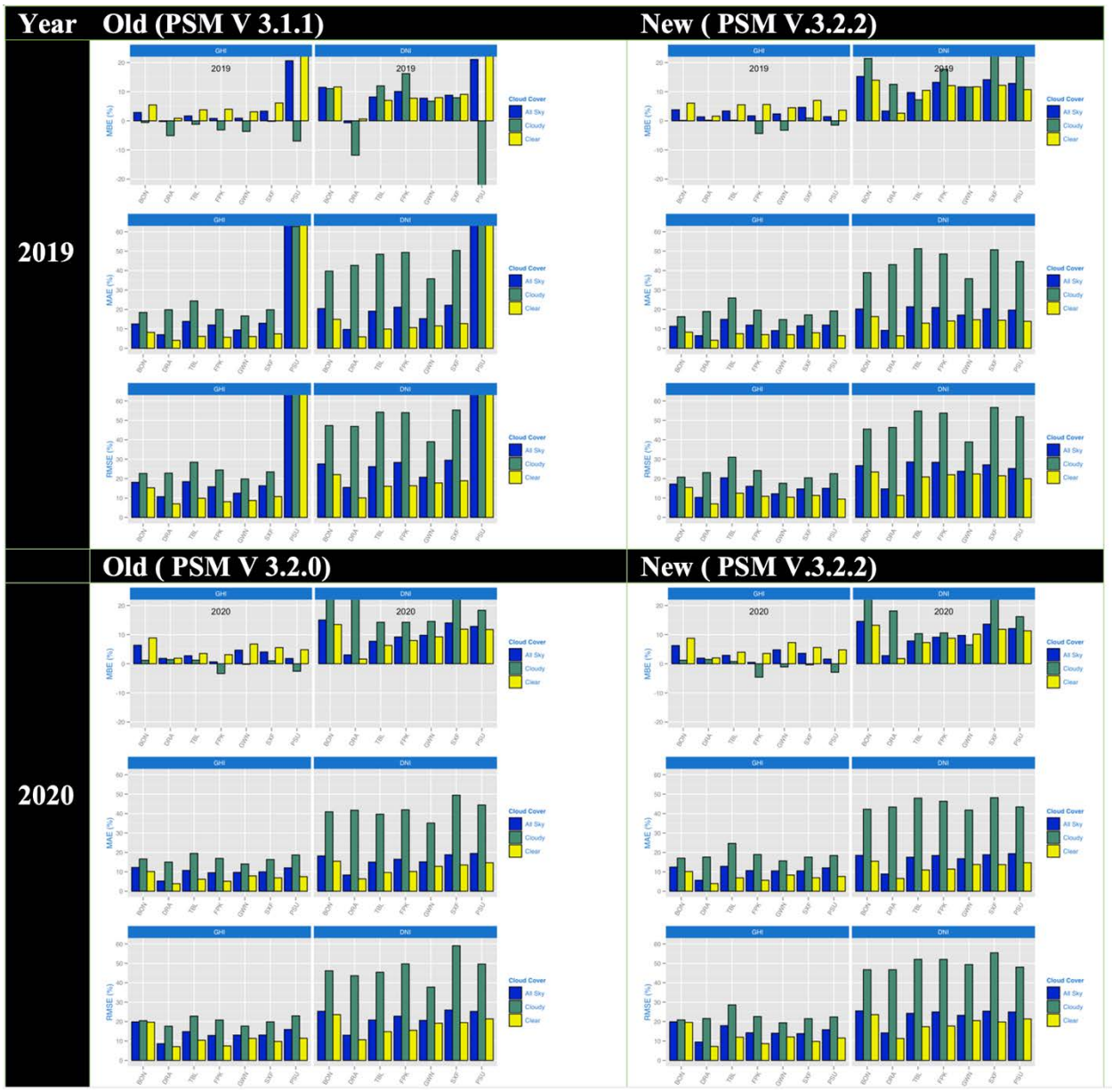

Figure 11. Comparison between the older versions (v3.1.1 and v3.2.0) and the latest version (v3.2.2) for 2019 and 2020

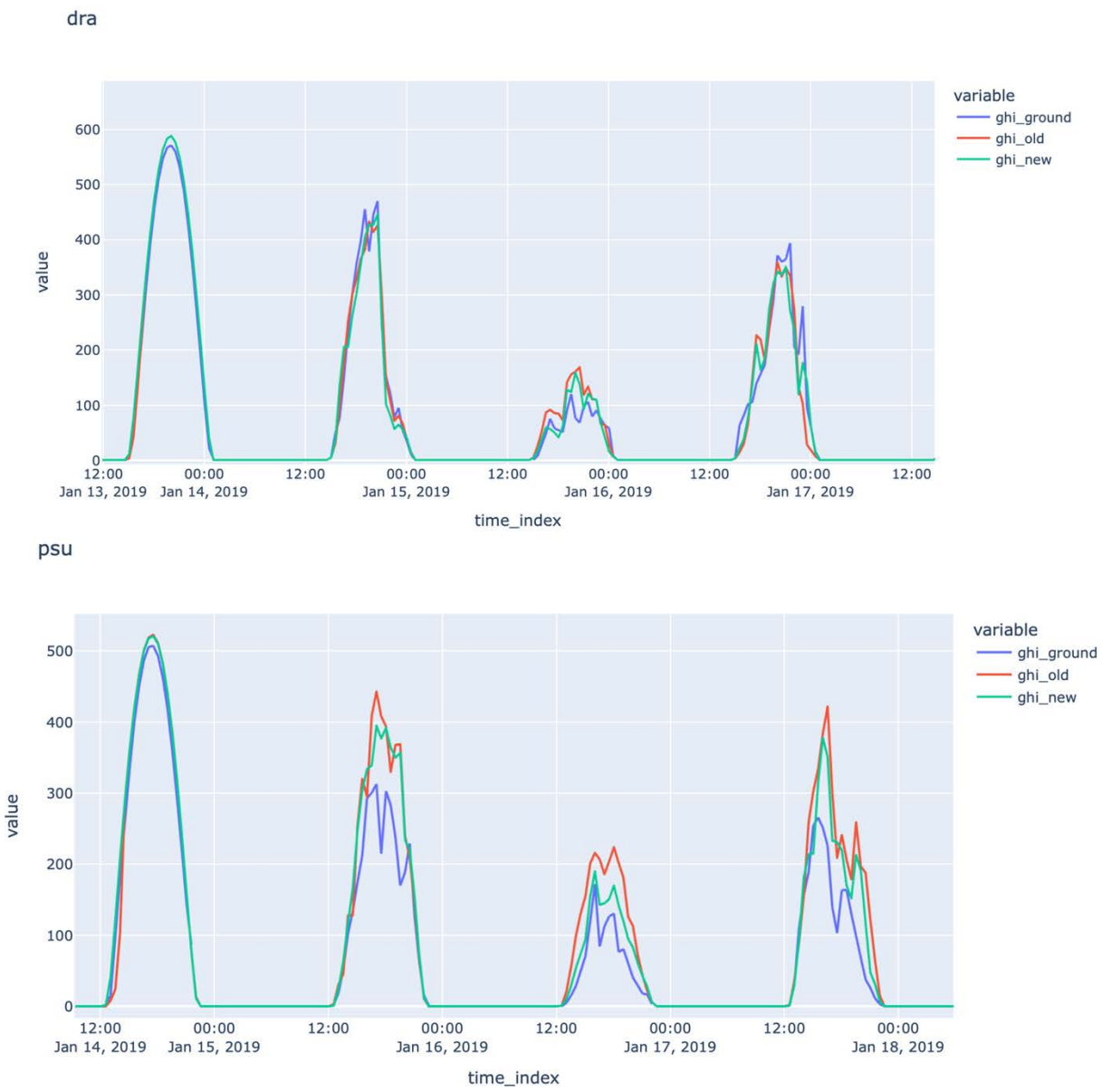

Figure 12. The time series of GHI for V 3.1.1 (ghi_old) and V 3.2.2 (ghi_new) at Desert Rock, NV (upper), and Penn Stat, PA (lower)

Table 1. Recent Versions of PSM Used in the NSRDB Production

Version	Effective Date	Data Years*	Notes
3.1.1	12/5/2019	2018+, TMY/TDY/TGY-2018	Complete refactor of TMY processing code
3.2.0	3/17/2021	2020	Enabled cloud solar shading coordinate adjustment by default, enabled MLClouds machine learning gap-filling method for missing cloud properties (cloud-filling flag #7)
3.2.2	2/25/2022	1998–2021	Implemented a model for snowy albedo as a function of temperature from MERRA-2 based on Ross and Walsh (1987)

Following the integration of the Fast All-sky Radiation Model for Solar applications with DNI (FARMS-DNI), the NSRDB has been further updated to v4.0.0. Figure 13 shows the MBE, MAE, and RMSE of the NSRDB v4.0.0. The NSRDB processed by the latest PSM has an MBE within 5% and 8% for the GHI and DNI, respectively.

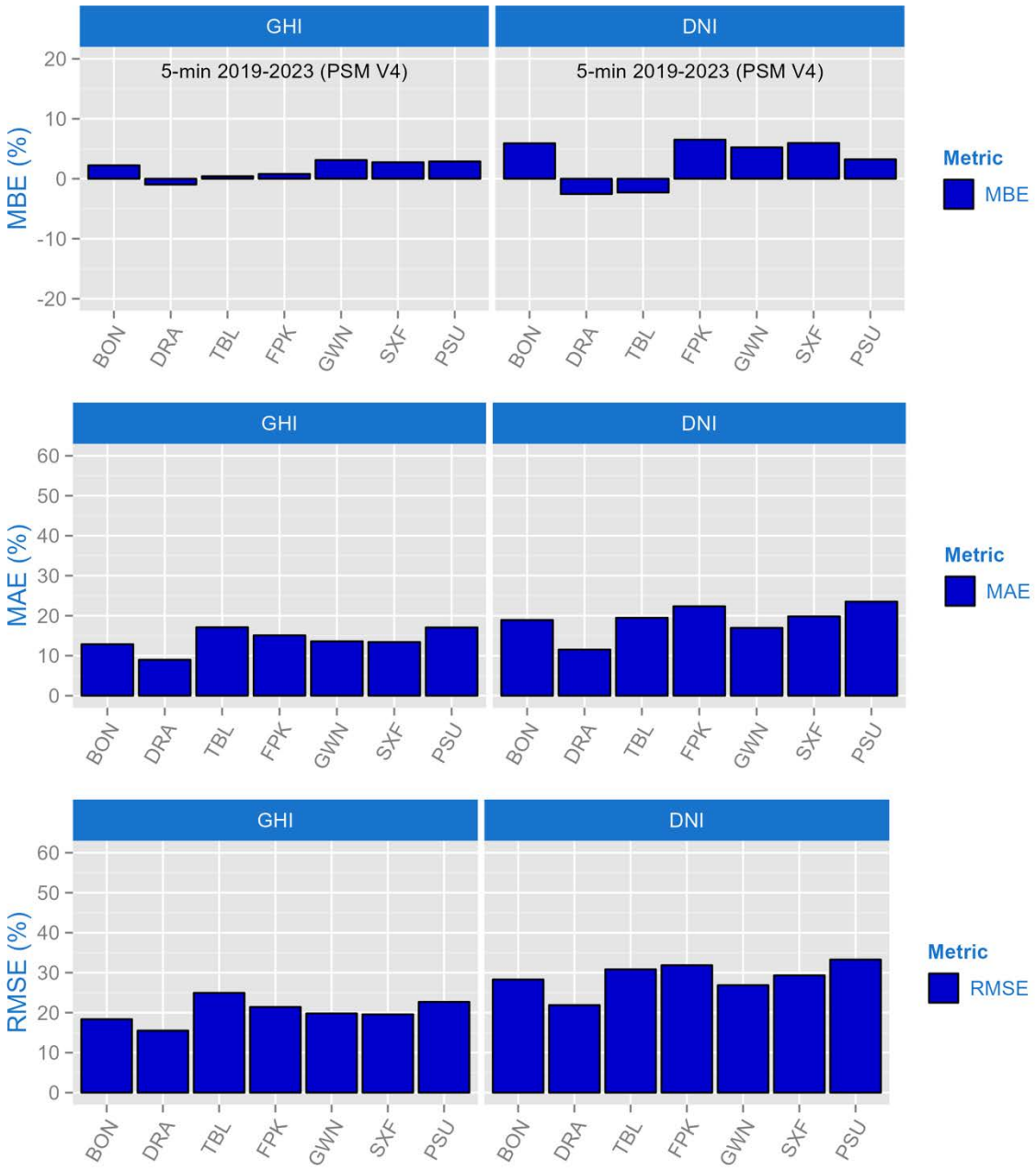

Figure 13. The MBE, MAE, and RMSE of the NSRDB data

2.2.3 Dissemination of the NSRDB Data

The NSRDB data continue to be disseminated through the NSRDB viewer, Amazon Web Services (AWS), an application programming interface, and PVWatts®. The statistics of NSRDB users accessing the data through the NSRDB viewer and application programming interface in 2024 are shown in Figure 14 and Figure 15, respectively.

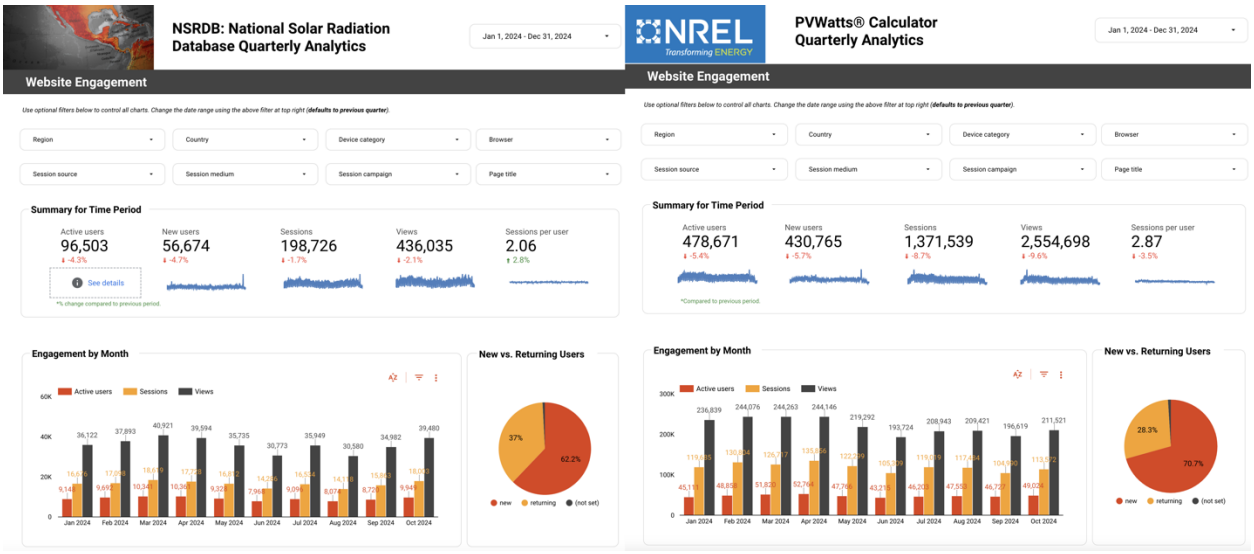

Figure 14. NSRDB users accessing data through the NSRDB viewer in 2024

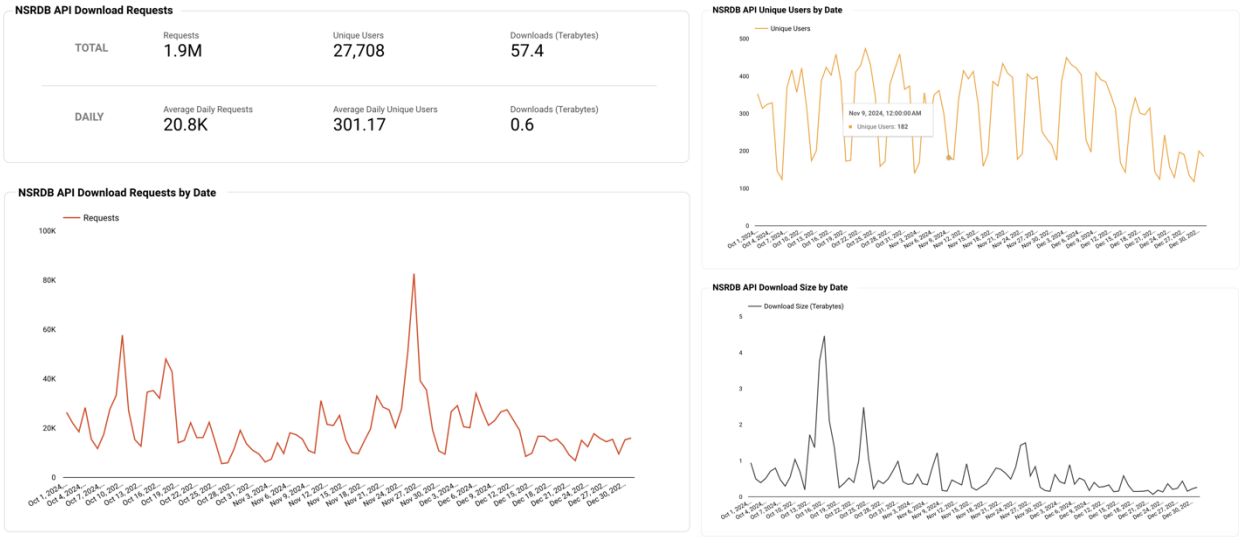

Figure 15. NSRDB users accessing data through the application programming interface in 2024

The NSRDB team successfully held an NSRDB webinar on Nov. 16, 2022 (Figure 16). There were 1,186 Zoom attendees and 119 YouTube attendees, for a total of 1,305 attendees of the 4,070 registrants.[1]

[1] The recording of the webinar is available online at https://www.youtube.com/watch?v=BzNoYACXCOY&ab_channel=NRELLearning. It has been viewed approximately 4,800 times since Dec. 13, 2022.

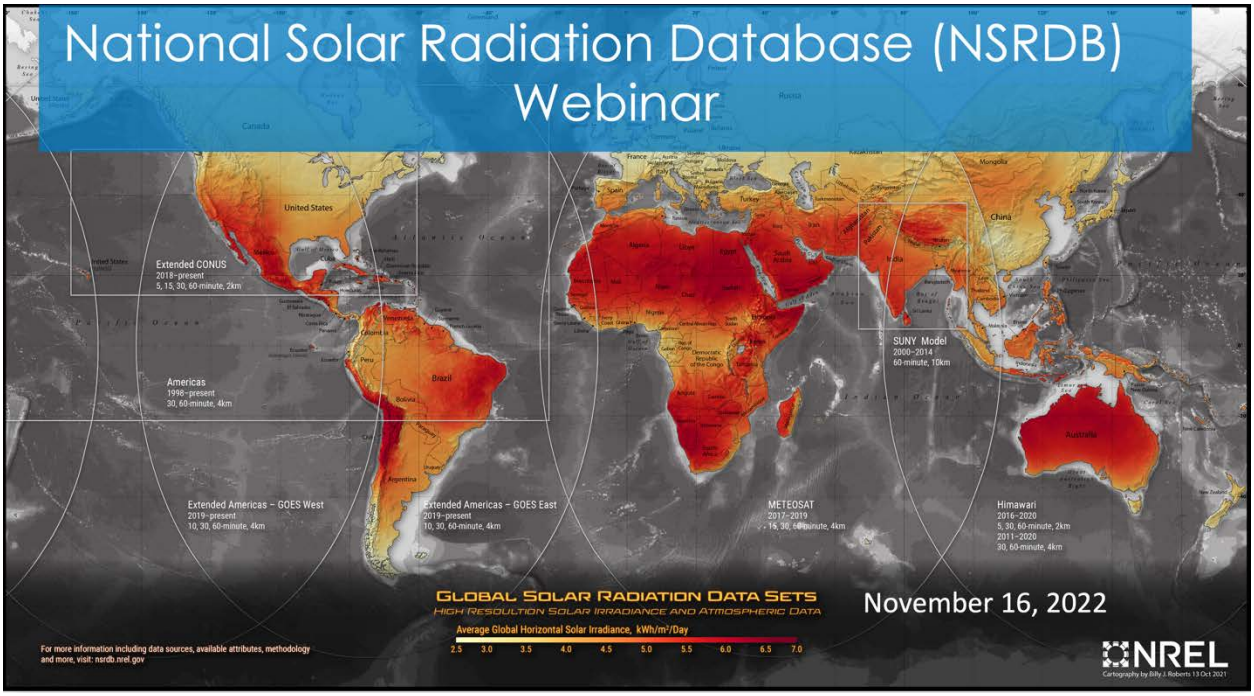

Figure 16. The 2022 NSRDB webinar

3 Task 3: Research and Development to Advance the Physical Solar Model and Develop New Products

3.1 Project Approach

3.1.1 PSM Modeling Advances

- The machine learning-based filling of the missing cloud properties will be incorporated into the PSM to replace the current heuristic methods that are being used, and data from 1998–2020 will be reprocessed. In addition, we will investigate whether machine learning-based methods can detect clouds and generate cloud properties that are comparable to the physical model.
- A new model that produces accurate DNI in clouds (FARMS-DNI) has been developed. This model will be incorporated into the PSM to replace the currently used Direct Insolation Simulation Code (DISC) decomposition model.
- The PSM v3 model will be updated to PSM v4, and data from 1998–2022 will be reprocessed. (PSM v4 includes the machine learning-based gap-filling of cloud properties and the implementation of FARMS-DNI).
- Visible and near-infrared data from the GOES satellites is now available at a 500-m resolution. Our goal is to use information from the 500-m resolution data to improve the accuracy of the NSRDB. A framework consisting of various methods to use 500-m data from GOES-16 and GOES-17 will be developed. Because the development and storage of 500-m data is expensive, we will create 1 year of 500-m data around locations with high-quality ground measurements of solar radiation. The 500-m cloud property dataset will be developed in budget period 1. Also, we will determine various methods that will be applied to extract information from the 500-m satellite imagery. The goal of these methods will be to use the 500-m resolution information to improve the accuracy of the 2-km radiation from the NSRDB.
- Sample 2-km datasets will be generated using the methods developed in budget period 1 to aggregate the 500-m information from GOES-16 and GOES-17. A direct method where radiation is computed for each 500-m pixel will form the baseline dataset for this comparison. Additionally, ground data will also be used for this evaluation.
- We will compare the accuracy of the various methods developed in FY 2023 to aggregate the 500-m cloud information from GOES-16 and GOES-17 to 2 km. This comparison will identify the methodology that provides the best improvements when incorporating information from the 500-m data into the 2-km dataset. The identified method will be implemented in future NSRDB versions.

3.1.2 Operational Improvement

- GOES-16 and GOES-17 satellites have channels similar to the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite that is currently used to compute AOD. Because these satellites have the capability to produce high-resolution AOD, we will investigate the availability of aerosol datasets from these satellites and obtain sample datasets for analysis.

3.1.3 Advanced Products

- The NSRDB contains 23 years of data, and it is now possible to produce summary statistics of these data on a regular basis. We will develop a framework for the summary statistics that will be included in the NSRDB distribution from the NSRDB website.
- Currently, users use the TMY for horizontal radiation when estimating PV generation for various tilt and orientations. This results in site-specific biases that could reach up to 10%; therefore, we will develop a capability to generate typical plane-of-array year (TPY) data (essentially a TMY at any orientation and tilt) based on user inputs. This method will be provided to users through an open-source platform (e.g., pvlib). Additionally, because the TPY can provide more accurate estimates through PVWatts, we will explore methods to use these data through PVWatts in collaboration with the Systems Advisor Model (SAM)/PVWatts team.

3.1.4 High-Latitude NSRDB Development

Based on the available polar-orbiting satellite data and NREL's modeling capabilities, we will explore the feasibility and pathway to extend the NSRDB to provide high-resolution solar resource data over the Alaska region.

3.2 Project Results and Discussion

3.2.1 Machine Learning Approach to Predict Missing Cloud Properties in the NSRDB

The NSRDB uses a two-step PSM that explicitly considers the effects of clouds and other atmospheric variables in the radiative transfer model. High-quality physical and optical cloud properties derived from satellite imagery are the most important data input to the PSM, representing the greatest source of radiation attenuation and scattering; however, conventional methods for cloud property retrieval have their own limitations and are unable to accurately predict cloud properties under all conditions. We have introduced a physics-guided neural network (PHYGNN) that can accurately predict cloud properties when traditional methods fail or are inaccurate. Using this method, we have shown significant reductions in error metrics versus ground truth measurements for time steps that previously had missing or low-quality cloud property data.

We integrated the PHYGNN model into the NSRDB processing software. The software module in the NSRDB containing the PHYGNN model is called MLClouds. Note that although the NSRDB is primarily an open-access data product, it is a complex software package that can process the many terabytes of solar data in a highly efficient pipeline. The NSRDB software has undergone tremendous improvements since the inception of PSM v3, which processed the NSRDB 2017 data. Notably, the NSRDB software underwent a full refactor to process the NSRDB 2018 data using the higher-resolution 5-min, 2-km satellite data from the GOES-16 satellite. As part of the refactor, we were able to gain an approximate 10-times improvement in processing speed; however, the heuristic gap-filling methodology that fills in missing cloud data remained one of the more inefficient parts of the processing code and did not scale well to the 2-km eastern full-disk domain, which is the largest extent we process for the NSRDB. Integrating the MLClouds methodology into the NSRDB software reduces the computational requirements

of the missing cloud property gap-filling and reduces the spatiotemporal dependencies of that piece of code.

Figure 17 and Figure 18 present the MAE for the GHI and DNI, respectively, for the three sky conditions and four NSRDB versions. From the figures, it is immediately apparent that each subsequent version improves the accuracy of the NSRDB irradiance data. Of particular interest is the improvement in cloudy gap-filled sky conditions using the MLClouds model as a gap-filling method (v3.2.0) and the improvement in all cloudy conditions using the MLClouds model as the primary method for cloud property prediction (v3+). Also notable is the ability of the MLClouds model to improve the accuracy of the irradiance data for the 11 sites on which the model is not trained. This indicates that the MLClouds model is trained on sufficiently diverse data to generalize to a wide variety of locations and cloud conditions.

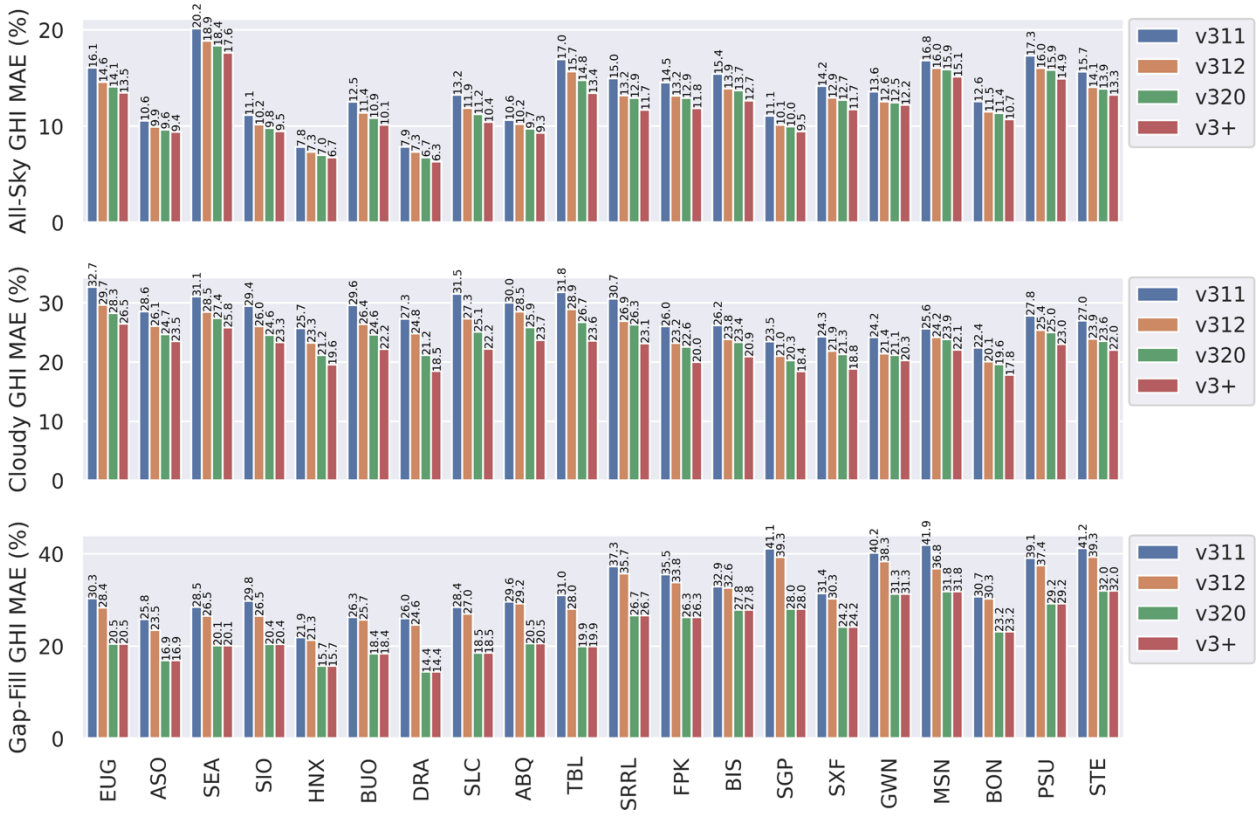

Figure 17. MAE validation results for GHI: all-sky (top), cloudy (middle), and cloudy gap-filled (bottom) conditions

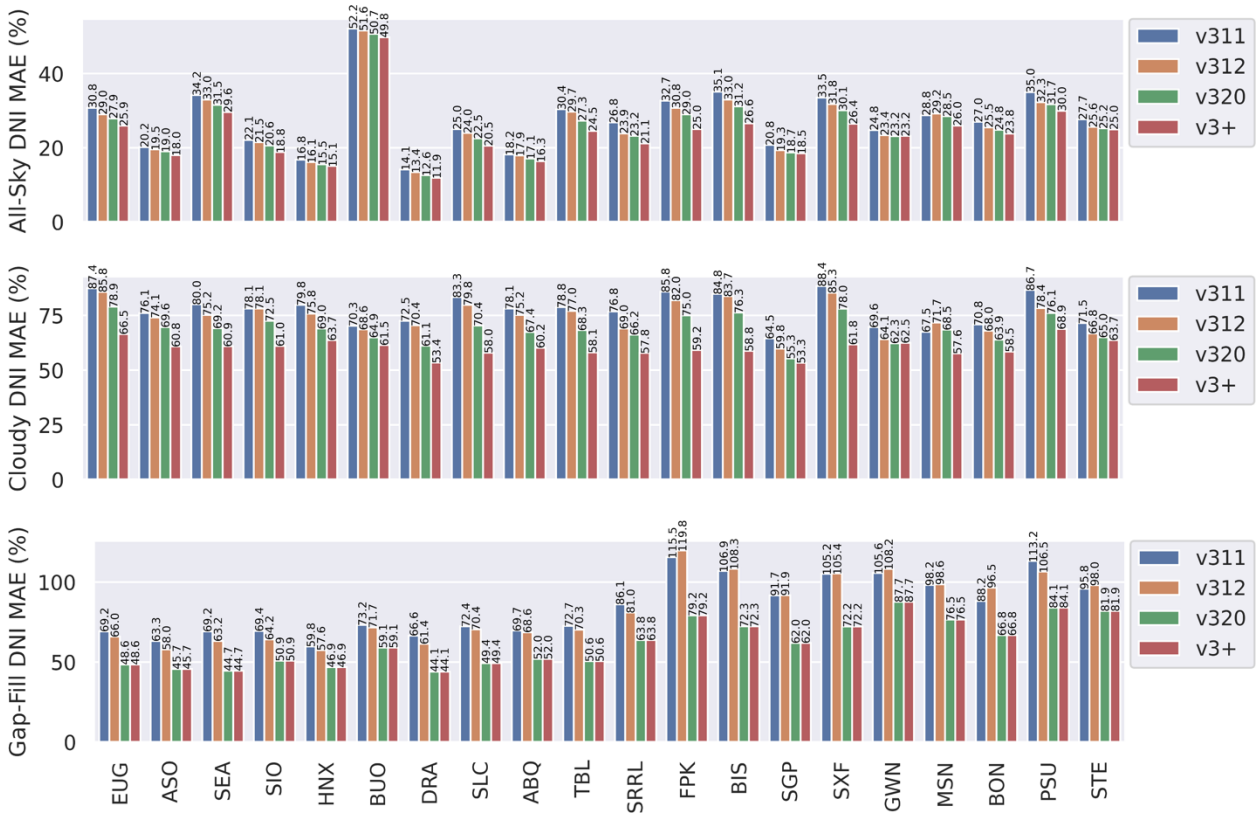

Figure 18. MAE validation results for DNI: all-sky (top), cloudy (middle), and cloudy gap-filled (bottom) conditions

Figure 19 and Figure 20 present the MBE of the GHI and DNI, respectively, for the three sky conditions and four NSRDB versions. The effect of the MLClouds model on the bias error is less significant and monotonic than the improvement in absolute error. Generally, the data do not show dramatic differences across subsequent NSRDB versions. All-sky GHI MBE is typically maintained within +/- 5%, which is consistent with previous NSRDB results (Habte, Sengupta, and Lopez 2017). The most illustrative metrics for the effects of the MLClouds model are the change in MBE from v3.1.2 to v3.2.0 for the cloudy gap-filled conditions and from v3.2.0 to v3+ for all cloudy conditions. These two comparative metrics show the isolated change in MBE due to the application of the MLClouds model to the cloudy gap-filled conditions and all non-gap-filled cloudy conditions, respectively. For the cloudy gap-filled conditions, the MLClouds model results in changes in MBE from -12.19% to -4.41% for GHI and from -8.77% to 2.34% for DNI. Applied to all other cloudy conditions, the MLClouds model results in changes in MBE from -4.29% to -4.48% for GHI and from 11.81% to 11.05% for DNI. The MLClouds model appears to improve the MBE for the cloudy gap-filled conditions, but it does not have a significant impact on the MBE for all other cloudy conditions compared to the DCOMP model. Note that the change from -8.77% to 2.34% MBE for the cloudy gap-filled conditions for DNI from v3.1.2 to v3.2.0 appears to make the MBE significantly worse for all cloudy conditions for DNI (7.04% to

11.81%), but this change is just improving the absolute bias in the gap-filled conditions and making the positive bias more apparent in all other cloudy conditions.

Figure 19. MBE validation results for GHI: all-sky (top), cloudy (middle), and cloudy gap-filled (bottom) conditions

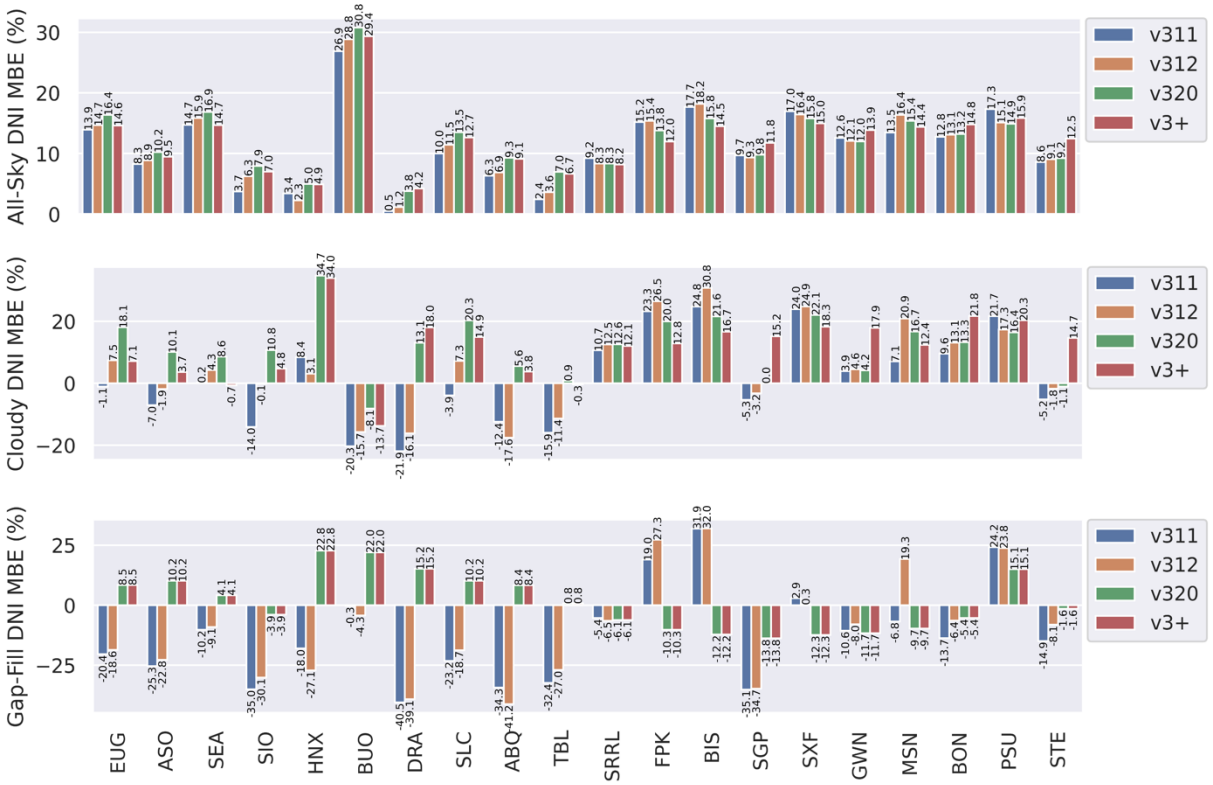

Figure 20. MBE validation results for DNI: all-sky (top), cloudy (middle), and cloudy gap-filled (bottom) conditions

3.2.2 Integration of FARMS-DNI to Enhance the NSRDB

The current state-of-the-art parametric DNI models suffer from imperfections that can result in significant bias in the DNI computation. For instance, the Direct Insolation Simulation Code (DISC) (Maxwell 1987) computes all-sky DNI using empirical relationships between surface observations of GHI and DNI. As a result, independent biases in the empirical relationships and GHI computation can compound and lead to a larger bias in the DNI computation. Additionally, a specific GHI value can correspond to numerous combinations of meteorological properties, each associated with a distinct DNI value. The consistent one-to-one match between GHI and DNI for a specific airmass can result in obvious errors in the DNI computation. The well-known Beer-Bouguer-Lambert law has been used by numerical weather prediction models to compute direct solar radiation (Liou 2002). This algorithm ignores the scattered solar radiation within the circumsolar region and thus often leads to significant underestimation of DNI when used for solar forecasting (Xie et al. 2022). The biases in the current DNI models were explicitly quantified through a comprehensive study using 17 years of surface-based observations conducted at the ARM SGP site (Xie et al. 2020). The findings revealed that DISC exhibited a significant overestimation of cloudy-sky DNI, with a bias of 153.85%. The Beer-Bouguer-Lambert law underestimated the cloudy-sky DNI by 74.98% and 78.85% for clouds composed of water droplets and ice crystals, respectively.

To address the limitations of the current DNI models, Xie et al. (2020) developed a physics-based model, FARMS-DNI, to efficiently infer all-sky solar radiation in the circumsolar region. This model has proved to be beneficial in reducing the bias in DNI forecasting when

incorporated into the Weather Research and Forecasting model with Solar extensions (WRF-Solar) (Xie et al. 2022). FARMS-DNI was also integrated into the algorithm used to create NREL's NSRDB (Sengupta et al. 2018). The long-term DNI is calculated using FARMS-DNI and satellite-based cloud properties and validated using surface observations across CONUS.

The NSRDB algorithm was enhanced with the integration of FARMS-DNI, coupled with the new parameterization of cloud transmittance. The clear-sky DNI is consistent with the conventional NSRDB, where REST2 is used to compute both clear-sky GHI and DNI. For cloudy-sky conditions, a sequential computation by FARMS and DISC is substituted by a concurrent procedure that uses satellite cloud products to infer cloud transmittance and reflectance for both direct and total downwelling solar radiation. The cloud transmittance and reflectance are then incorporated with FARMS, FARMS-DNI, and the clear-sky computation to resolve the cloudy-sky GHI and DNI.

Following the implementation of the algorithm, all-sky DNI near 19 surface sites in the networks of SURFRAD, SOLRAD, UO, the U.S. Department of Energy ARM, and NREL were computed using the NSRDB input data for each 5-minute interval during 2019–2020. Figure 21 displays the locations of the surface sites considered in the study. The SURFRAD network consists of seven sites at Bondville, Illinois (BON); Desert Rock, Nevada (DRA); Goodwin Creek, Mississippi (GWN); Fort Peck, Montana (FPK); the Pennsylvania State University (PSU); Sioux Falls, South Dakota (SXF); and Table Mountain, Colorado (TBL). The SOLRAD network includes sites in Albuquerque, New Mexico (ABQ); Bismarck, North Dakota (BIS); Hanford, California (HNX); Madison, Wisconsin (MSN); Salt Lake City, Utah (SLC); Seattle, Washington (SEA); and Sterling, Virginia (STE). The UO network includes three sites: Ashland, Oregon (ASO); Eugene, Oregon (EUG); and Silver Lake, Oregon (SIO). The ARM and NREL sites are located at the SGP and SRRL, respectively. Following the NSRDB algorithm, the DNI in the west and east of the red line (-113° longitude) in Figure 21 was computed using GOES-17 and GOES-16 data, respectively. For comparison with FARMS-DNI, DNI was also computed using the conventional NSRDB algorithm, where DISC was used to compute DNI from the GHI simulated by FARMS.

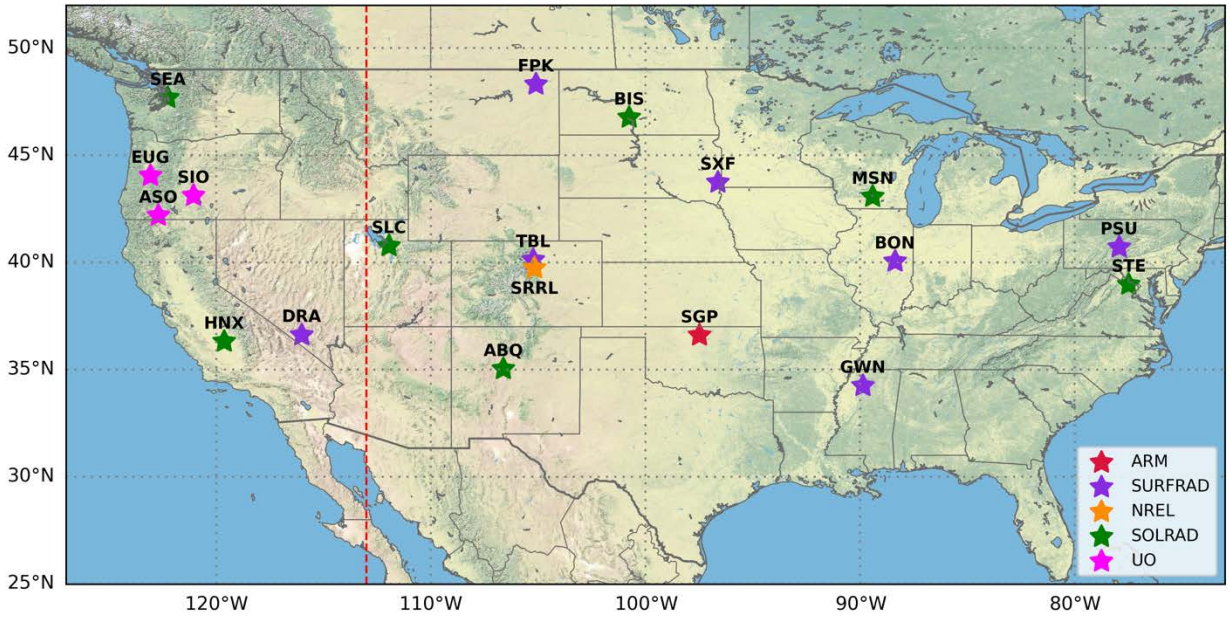

Figure 21. The locations of the surface sites in the evaluation of DNI

Figure 22 illustrates the cloudy-sky DNI computed using FARMS-DNI for the period from 2019–2020 compared to the surface observations from the 19 surface sites in the previously mentioned networks. The distribution of cloudy-sky DNI is skewed toward the lower-intensity region due to the light scattering and absorption by clouds. The cloudy-sky DNI greater than 400 W/m^2 are usually associated with thin clouds where the light scattering around the forward direction is more significant than the other directions. The distribution of cloudy-sky DNI is mostly centered around the 1:1 line, indicating good overall agreement between the computation by FARMS-DNI and the surface observations; however, bias is noticeable in the individual computations. According to the Beer-Bouguer-Lambert law, direct radiation in an infinite-narrow beam is an exponential function of cloud optical thickness, suggesting that the bias in the satellite-based cloud products is usually significantly amplified when they are used to compute DNI; therefore, an accurate cloud retrieval technique is as important as the radiative transfer model in the DNI computation.

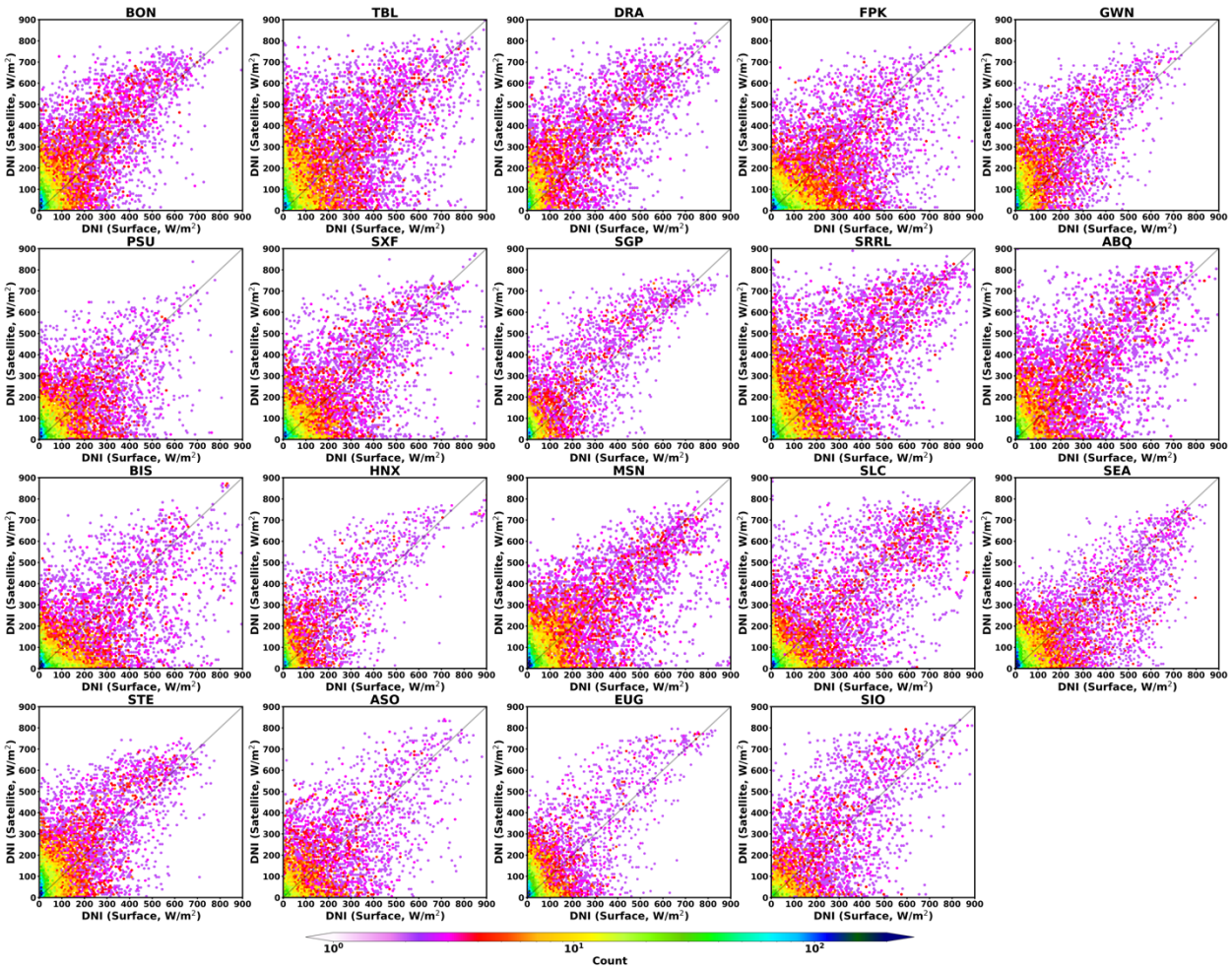

Figure 22. Comparison of the cloudy-sky DNI between those computed by FARMS-DNI and surface observations

Figure 23 (a)–Figure 23 (d) shows the metrics of the cloudy-sky DNI computed by DISC and FARMS-DNI. The MBE and percentage error indicate that DISC overestimates the cloudy-sky DNI at all the sites, which is attributed to both the DISC model and the computation of the cloudy-sky GHI. But FARMS-DNI can moderate this bias, particularly for sites with high occurrences of snow and bright surfaces, such as TBL, SRRL, DRA, and HNX. The overestimation of DNI is converted to a slight underestimation at a few sites, leading to an even lower bias in the average over all 19 sites. In addition, the use of FARMS-DNI reduces the MAE and absolute percentage error (APE) of the cloudy-sky DNI, as shown in Figure 23 (b) and Figure 23 (d). Figure 23 (e)–Figure 23 (h) present the error metrics of the all-sky DNI, including clear-sky and near-clear-sky conditions, which were excluded using the previously mentioned criteria. The cloudy-sky computation is performed by DISC and FARMS-DNI, whereas the clear-sky DNI is given by REST2. The results show that FARMS-DNI outperforms DISC in terms of MBE and percentage error for the all-sky DNI. Moreover, the MAE and APE of DISC and FARMS-DNI are comparable over the 19 surface sites. Based on the improved accuracy in both the cloudy-sky and all-sky conditions, as demonstrated in Figure 23, integrating FARMS-DNI into the NSRDB algorithm is expected to enhance the DNI data.

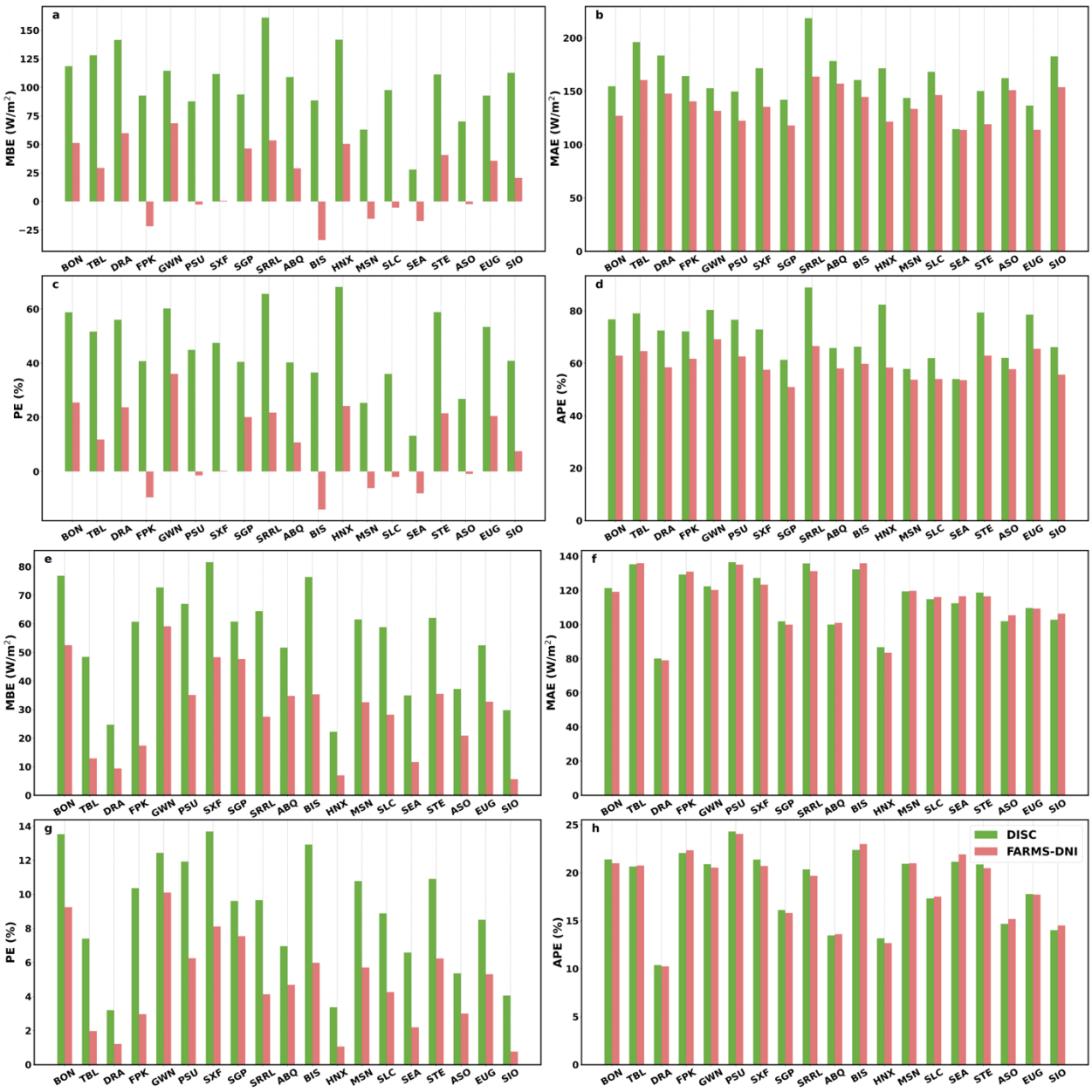

Figure 23. (a–d) The error metrics of the cloudy-sky DNI and (e–h) those of the all-sky DNI computed by DISC and FARMS-DNI

Note that FARMS-DNI still shows substantial bias in the individual cloudy-sky scenarios even though it reduces the overall uncertainty in the long-term averaged data. This is likely a joint effect of the model bias and the bias in the satellite-based cloud retrievals. Because DNI is particularly sensitive to cloud optical thickness, it is crucial to further improve cloud retrieval techniques and use them in future DNI computations. Additionally, the performance of DISC and FARMS-DNI in partially cloudy conditions requires further investigation, even though FARMS-DNI shows better performance in all-sky and confidently cloudy conditions. Accurate cloud coverage data play an important role in computing cloudy-sky DNI because FARMS-DNI is designed for cloudy overcast conditions. The use of precise cloud coverage data from

advanced satellite techniques, such as the improved spatial resolution provided by GOES-16 and GOES-17, should aid in further improvements in DNI assessment.

3.2.3 The Influence of Cloud Cover on the NSRDB

Due to limitations in the current satellite remote-sensing techniques, high-spatial-resolution solar resource datasets derived from satellite observations often categorize retrievals into binary classifications, distinguishing between clear-sky and cloudy-sky scenes. This classification significantly impacts the application of cloud retrievals and the selection of models for computing solar radiation. For instance, when the atmosphere is identified as clear sky by satellite data, FARMS disregards the potential influence of clouds and employs a clear-sky radiative transfer model, REST2, to compute the GHI and DNI. Conversely, in cloudy-sky conditions, FARMS incorporates precomputed cloud transmittances specifically tailored for overcast cloudy conditions (Xie, Sengupta, and Dudhia 2016). Many other solar resource datasets follow a similar classification approach.

Clouds in the natural atmosphere are typically more complicated than a simple binary classification of clear-sky and cloudy-sky scenes. In a study conducted by Kim et al. (2015), cloud fraction was investigated using data from MODIS during the period from May to October 2014. The analysis of the global distribution of cloud fraction revealed that the occurrence frequencies for clear-sky and cloud overcast conditions were approximately 12% and 19%, respectively. Among the substantial instances of scattered and broken clouds, scenes with cloud fractions between 60% and 90% were significantly more prevalent than those with cloud fractions between 10% and 50%.

This study aims to examine whether the current computation of solar radiation is oversimplified by relying on the binary cloud mask data derived from satellite observations. To achieve this goal, satellite-based GHI and DNI data are computed using the NSRDB algorithm and assessed with respect to the cloud fraction obtained from surface observations across 17 sites in the CONUS region. The bias in the satellite-based solar resource data is analyzed across various cloudiness categories that are classified using cloud fraction. The study identifies the physical reason behind the consistent overestimation of clear-sky solar resource data and highlights the advantages of physics-based radiative transfer models in mitigating DNI overestimation under cloudy conditions. It also explores potential biases introduced by the error in satellite-derived cloud products. The findings help diagnose the key sources of uncertainty in current solar resource datasets and suggest potential algorithms for further improvement.

The NSRDB algorithm is employed to compute GHI and DNI at 17 surface sites within the networks of SURFRAD, SOLRAD, UO, ARM, and NREL for each 5-minute interval throughout the years 2019–2020. Details about the NSRDB algorithm have been elucidated by Sengupta et al. (2018). In scenes classified as clear sky by the 2-km GOES cloud mask data, both GHI and DNI are computed using the REST2 model. In cloudy-sky scenes, solar radiation is computed by FARMS-DNI. To diagnose the sources of uncertainty, cloudy-sky DNI is also computed using DISC, which relies on the GHI modeling and empirical regressions of surface observations. To assess the impact of cloudiness on the reliability of a solar resource dataset, cloud fractions are estimated using 1-minute resolution surface observations through two independent methods. The first method uses a physical model proposed by Xie and Liu (2013) (hereafter referred to as XL2013) that simultaneously retrieves cloud fraction and cloud albedo by solving the radiative

transfer equation with a two-stream approximation. The second method involves a temporal average of cloud mask determined by DNI data. The satellite and surface data assembled in the analysis are classified as cloudy-sky scenes based on the following criteria: (1) Cloudy-sky conditions are identified by the GOES cloud mask data; (2) the surface observed GHI is lower than 99% of the clear-sky GHI computed using REST2; and (3) the surface observed DNI is lower than 99% of the clear-sky DNI computed using REST2. However, data flagged as cloudy-sky conditions by the GOES cloud mask data are classified as a "mismatch" if the surface observed GHI exceeds 99% of the clear-sky GHI or if the surface observed DNI exceeds 99% of the clear-sky DNI. These occurrences can be linked to uncertainties in the satellite-based cloud mask data. They could also be associated with strong cloud reflection in pixels leading to substantial diffuse radiation at the observation site. This classification uses surface observations to improve confidence in identifying cloudy-sky scenes, though it might imprecisely classify certain instances of cloudy skies as a mismatch, such as those associated with very thin clouds. The clear-sky classification in this study uses the GOES cloud mask data.

Figure 24 shows the distribution of data and the relative frequencies across different cloudiness categories. To enhance the accuracy of the clear-sky classification, clear-sky data are refined based on cloud fractions estimated by XL2013 with surface observations. Clear-sky conditions are more confidently identified when corroborated by both satellite-based cloud masks and cloud fractions lower than 20%. The satellite-based clear-sky data with cloud fractions exceeding 20% might indicate the presence of scattered clouds, broken clouds, or potential bias in the cloud mask data. Similarly, the cloudy-sky data are further classified into confidently cloudy and partially cloudy conditions based on cloud fractions above and below 80%, respectively. In Figure 24, the surface sites are organized based on the relative frequency of confidently clear-sky conditions with cloud fractions less than 20%. This arrangement facilitates comparisons across surface sites in terms of their cloudiness level, thereby aiding in understanding the NSRDB performance under different climatic conditions. It is observed that the relative frequency of cloudy-sky data generally increases as the frequency of the clear-sky data decreases. Among all the sites, GWN shows a higher occurrence of scenes with scattered clouds despite being classified as clear sky by the satellite data. SRRL and TBL exhibit the highest frequency of mismatch between satellite and surface-based cloud mask data, possibly attributable to specific cloud types and surface terrain characteristics in these regions.

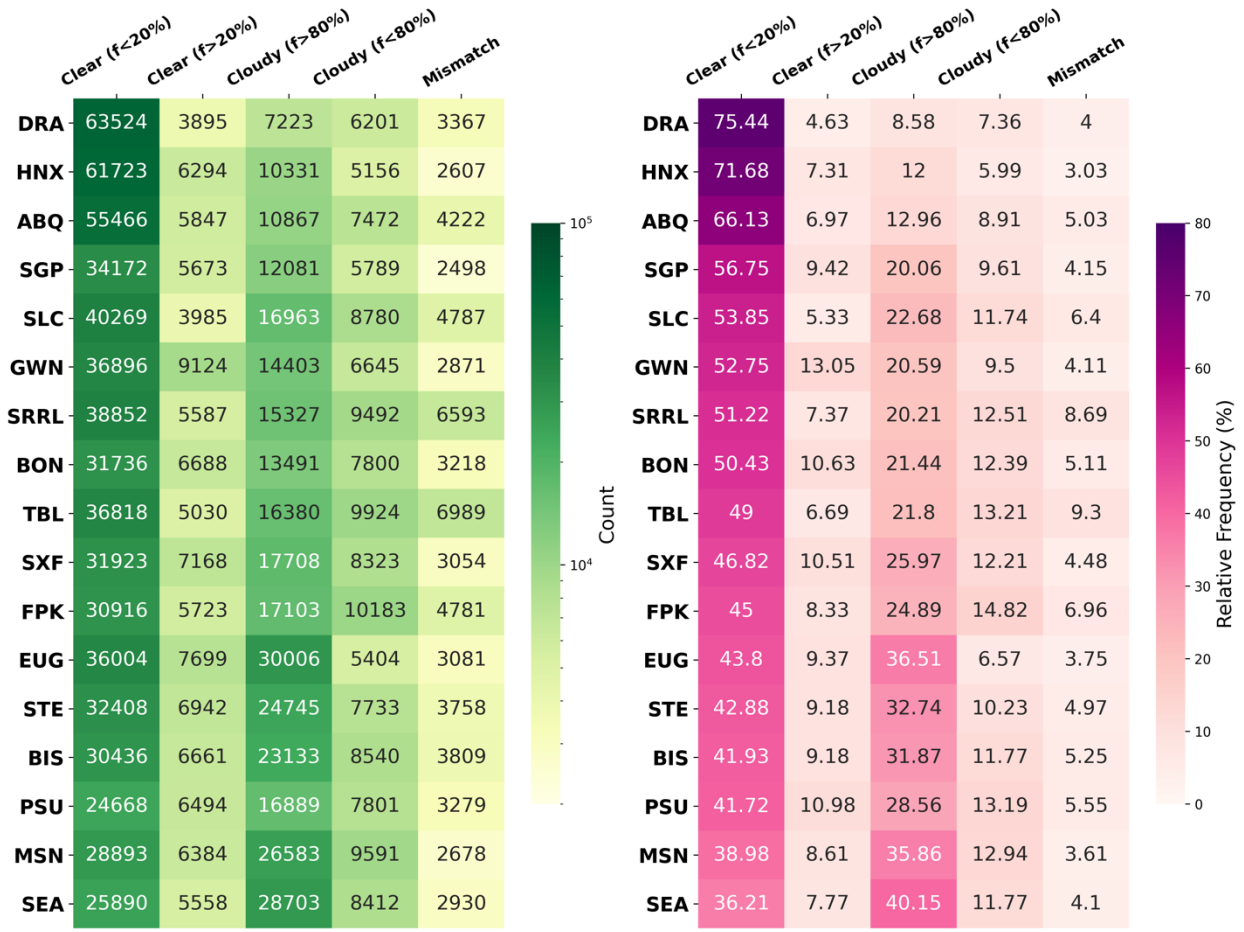

Figure 24. The number of scenes (left) and the relative frequency (right) in each cloudiness category. The surface sites are arranged based on the relative frequency of clear sky with f<20%.

Figure 25 illustrates the MBE for GHI and DNI in clear-sky conditions as classified by the GOES cloud mask data. This bias aligns with that of the NSRDB because it also employs the cloud mask data to determine cloudy conditions. The relative frequency of the data with respect to cloud fraction reveals that most clear-sky instances are associated with a cloud fraction smaller than 20%; however, notable data are observed related to scattered clouds, broken clouds, or even overcast conditions, which suggests relatively low-resolution or potential biases in the GOES cloud mask data. Both schemes for estimating the cloud fraction exhibit similar performance, though XL2013 tends to identify fewer scenes with scattered clouds but more instances of overcast conditions than the temporal average of cloud mask. Across all sites, satellite-based solar resource data tend to overestimate both GHI and DNI for cloud fractions ranging from 0% to 100%, primarily due to the omission of clouds in the computational process. The bias becomes more pronounced with increasing cloud fraction; however, the overall uncertainty related to high cloud fractions could be limited due to the limited sample sizes. Compared to GHI, DNI is more sensitive to the presence of clouds, leading to higher uncertainty in the clear-sky DNI data from the NSRDB. Both GHI and DNI exhibit higher bias when using XL2013 to categorize the sky cloudiness.

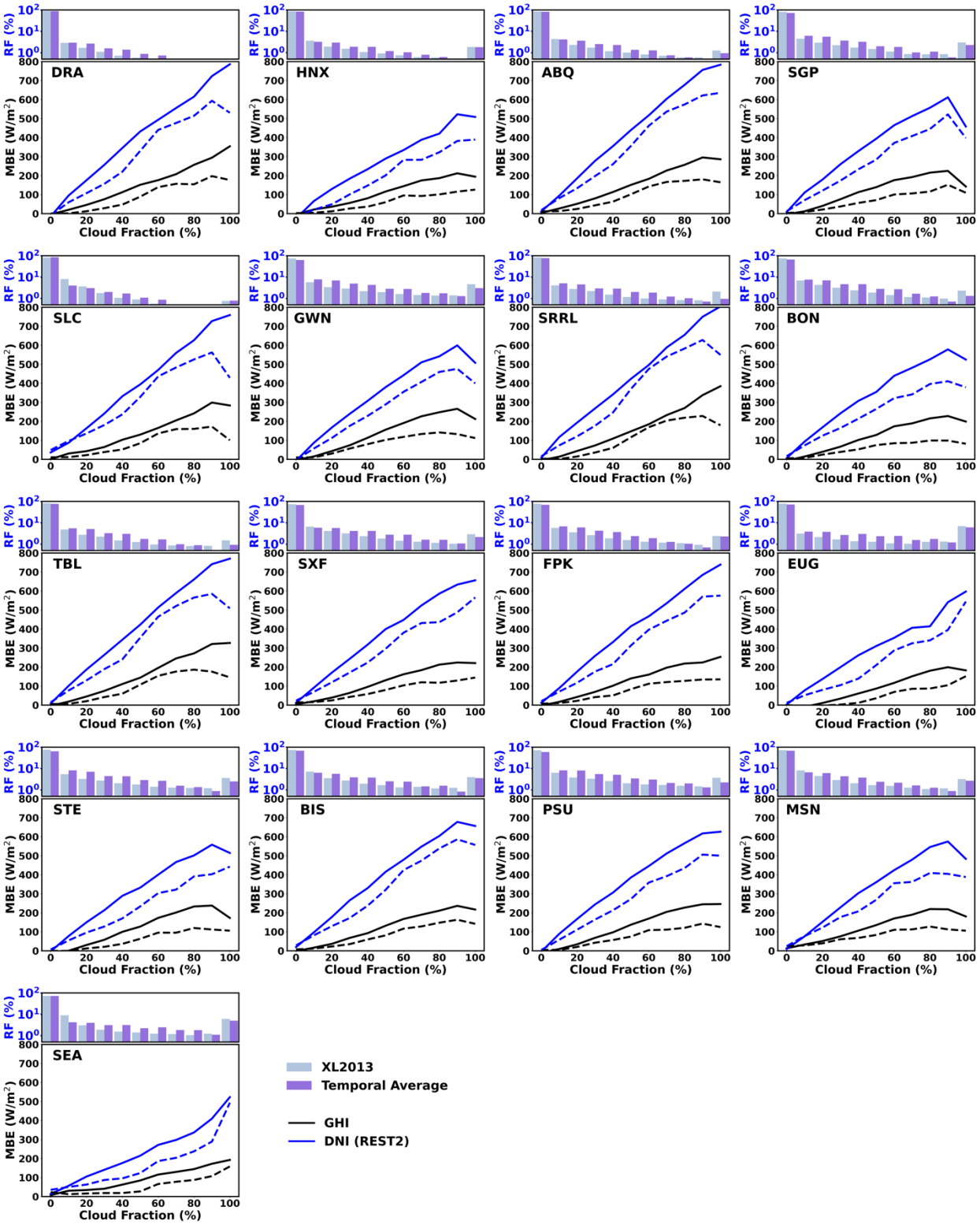

Figure 25. The MBE for GHI and DNI in clear-sky conditions classified by satellite data. The solid and dashed lines denote data associated with the cloud fraction estimated by XL2013 and the temporal average of cloud mask, respectively. The histograms represent the relative frequencies of cloud fraction.

Figure 26 shows the MBE for GHI and DNI in cloudy-sky conditions as classified by the GOES cloud mask data. Most cloudy-sky instances are associated with a cloud fraction larger than 80%, whereas a notable amount of data is related to very low cloud fractions. XL2013 tends to identify more scenes with scattered clouds but fewer instances of overcast conditions than the temporal average of cloud mask. In both schemes, the uncertainty of GHI is much smaller than that of DNI. When the cloud fraction is larger than 80%, DNI computed by DISC exhibits a large positive bias, which decreases with cloud fraction. This suggests a bias of DISC in confidently cloudy scenes that are classified by both satellite data and surface observations. The bias in the cloudy-sky DNI is significantly reduced by using FARMS-DNI, which is a physical model specifically designed for cloud overcast conditions; however, due to conflicts between satellite and surface data, FARMS-DNI underestimates cloudy-sky DNI when the cloud fraction is below 20%. Improvements in satellite-based cloud mask retrieval or the introduction of cloud fraction in the solar radiation computation could potentially reduce this bias in cloudy-sky DNI computed by FARMS-DNI. In contrast, DISC shows better accuracy than FARMS-DNI in those partially cloudy conditions because it is an empirical model that does not directly rely on cloud masks from satellite data. DISC requires input of GHI, which shows only a slight underestimation because it is less sensitive to clouds than DNI.

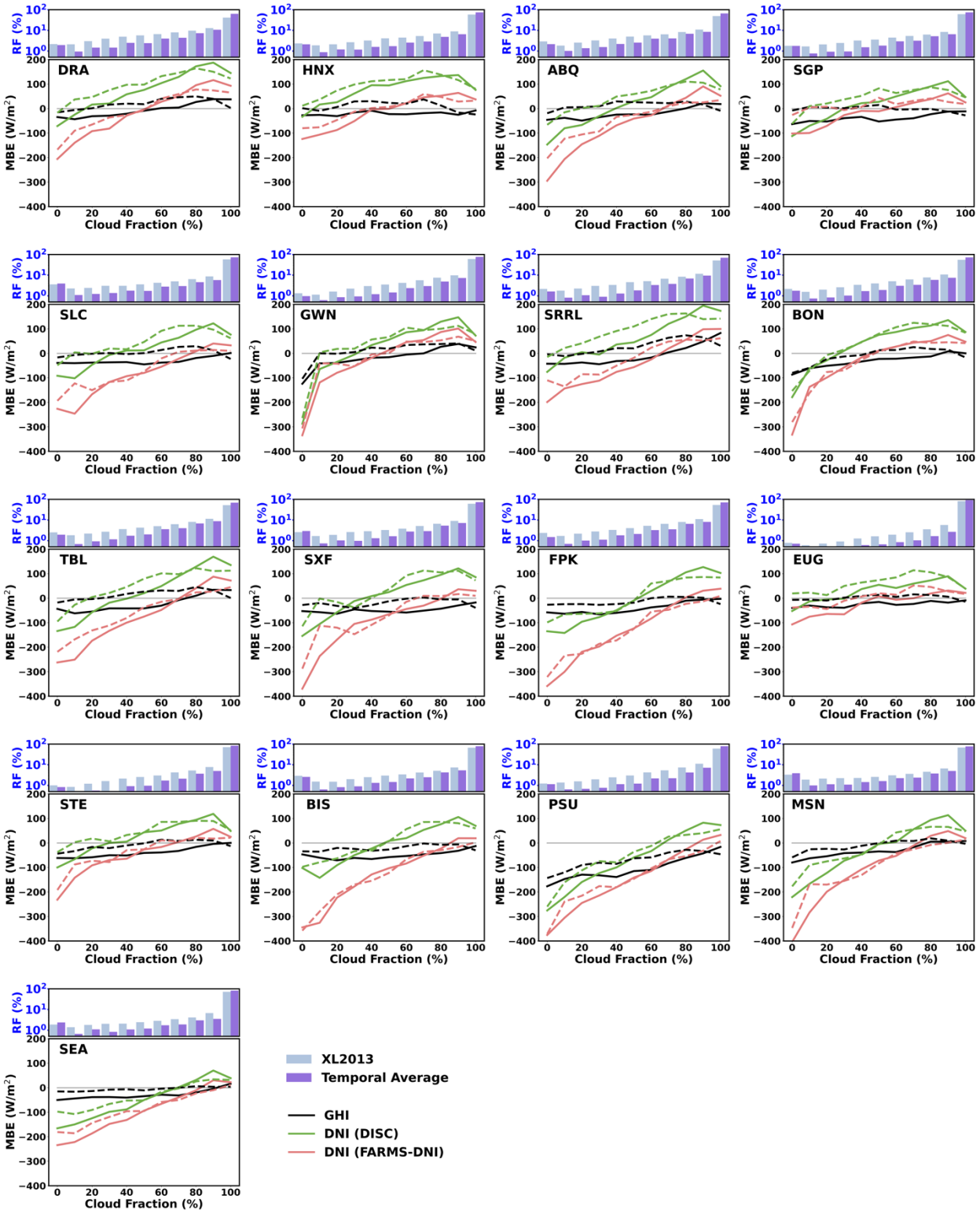

Figure 26. The MBE for GHI and DNI in cloudy-sky conditions classified by satellite data. The solid and dashed lines denote data associated with the cloud fraction estimated by XL2013 and the temporal average of cloud mask, respectively. The histograms represent the relative frequencies of cloud fraction.

Figure 27 illustrates the MBE and MAE for GHI and DNI across each cloudiness category as given in Figure 24. The cloud fractions are given by XL2013 because of the consistent distribution of cloud fractions between the two schemes. XL2013 uses data from 5-minute windows, which aligns with the temporal resolution of the NSRDB. Both GHI and DNI in Figure 27 reveal that the biases in confidently clear-sky conditions are minimal across all surface sites; however, significant biases are observed in solar radiation when scattered clouds are missed by the satellite data. The results clearly indicate that biases in REST2 and AOD data contribute to smaller uncertainties in clear-sky conditions than scattered clouds. It is also evident that the computation of GHI generally performs well in confidently cloudy conditions. The GHI in partially cloudy conditions is nonexclusively underestimated by the NSRDB because FARMS assumes overcast clouds in those scenes. A significant underestimation of GHI is observed in the satellite-surface mismatched data where the surface observed solar radiation is larger than 99% of the computations in the clear-sky conditions. This implies that the satellite retrieval of cloud properties might not accurately represent the characteristics of these scenes. In confidently cloudy conditions, FARMS-DNI mitigates the overestimation of DNI computed by DISC. Similar to GHI, DNI in partially cloudy conditions is underestimated by FARMS-DNI. In contrast, DISC overestimates DNI in most sites for this cloudiness category. The biases of FARMS-DNI in the two cloudiness categories tend to offset each other, whereas those of DISC accumulate. Consequently, FARMS-DNI generally exhibits better performance than DISC in cloudy-sky conditions.

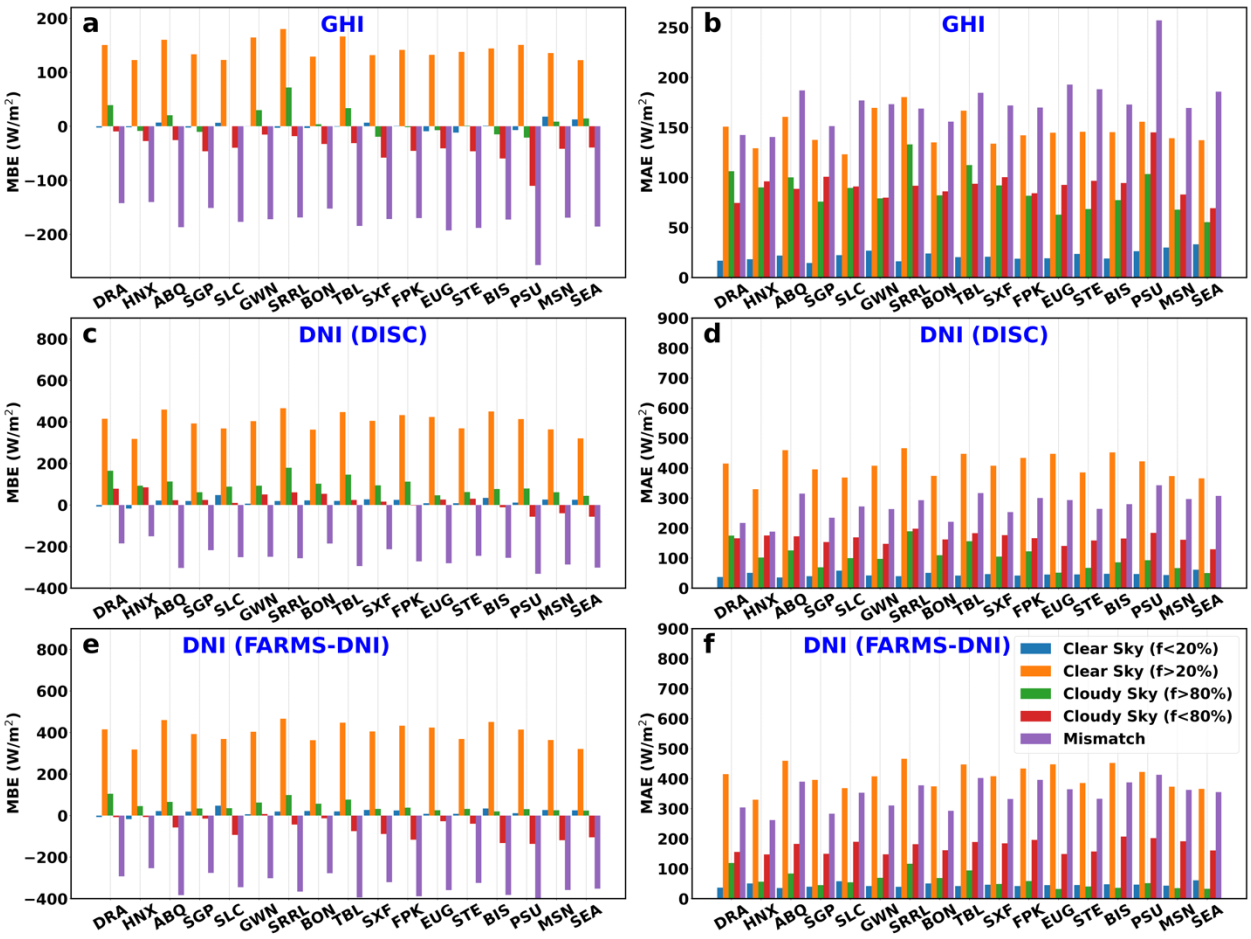

Figure 27. The MBE and MAE for GHI and DNI in each cloudiness category

The MBE of solar radiation faces challenges in discerning the relative significance of the bias contributed by each cloudiness category because it is not weighted by the sample size. To address this issue, we consider the percentage error and APE of solar radiation to represent the total bias from a cloudiness category with respect to the solar radiation across all scenes (Figure 28); thus, the sum of the percentage errors across the cloudiness categories represents the overall percentage error for all-sky conditions. Consistent with the results in Figure 27, Figure 28 reveals that the bias in the partially cloudy conditions and the satellite-surface mismatch data remain the most significant in the current NSRDB. The computations of GHI and DNI in confidently clear conditions are exceptionally accurate even accounting for the substantial sample size in the category. But the bias in the cloudy-sky conditions becomes more pronounced compared to the findings in Figure 27, which suggests that further enhancements in cloud data and cloudy-sky radiative transfer models will be beneficial for improving the NSRDB. For the DNI computed by FARMS-DNI, the overestimation in the confidently cloudy conditions can be counterbalanced by the underestimation in the partially cloudy conditions. Although the underestimation in the satellite-surface mismatch data is more pronounced than that of DISC, it can significantly mitigate the overestimation from the clear-sky DNI when considering the all-sky DNI. These findings collectively explain that FARMS-DNI outperforms DISC in both cloudy-sky conditions and all-sky conditions, as reported in the previous studies.

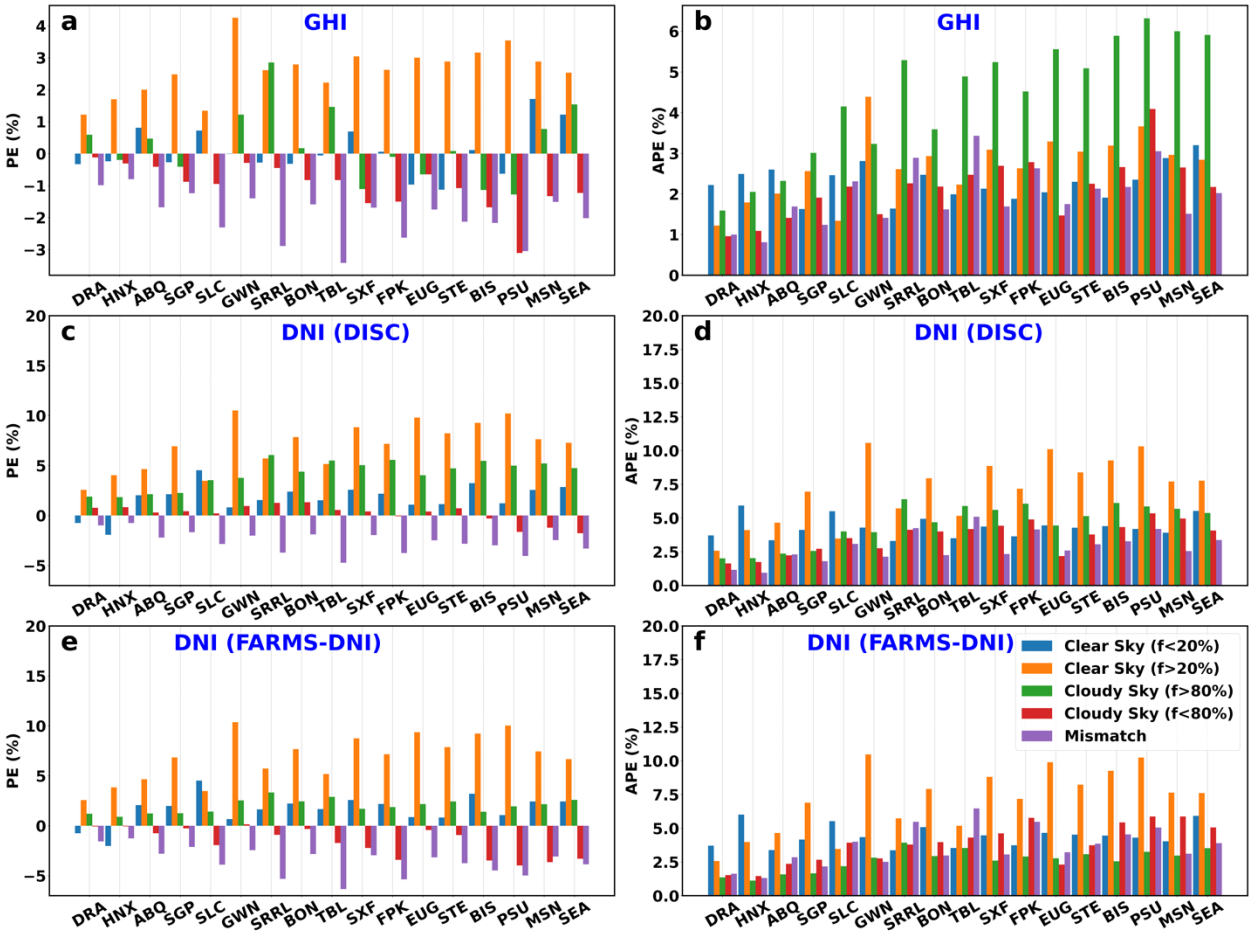

Figure 28. The percentage error and APE for GHI and DNI in each cloudiness category

According to the investigation, implementing precise cloud fraction data in the computations should substantially mitigate the bias caused by scattered clouds in clear-sky conditions. The cloud fraction data could be provided by high-resolution satellite data coupled with an enhanced cloud retrieval algorithm in a future study. A novel radiative transfer model that considers the intricate 3D topography effects and the amplified reflection by clouds could help diminish the bias caused by the satellite-surface mismatch data. Further, the percentage error of the solar radiation indicates that refining the cloudy-sky radiative transfer model and the cloud data could also improve the computations in the confidently cloudy conditions.

3.2.4 High-Latitude NSRDB Development

Alaska has a significant solar resource and high electricity prices, so it can be a favorable area to develop PV projects. Although there is low solar radiation in winter across the whole state, sunlight during the summer months lasts for 18–24 hours per day. The snow reflection in spring and fall also helps increase solar energy production. Moreover, Alaska's electricity prices are almost double the U.S. average, which leads to a great deal of interest in solar energy technologies. Therefore, it is crucial to extend the current NSRDB to provide high-resolution solar resource information using satellite data.

It is essential to acquire high-resolution satellite data that can provide continuous coverage over the Alaska region. This data are then used to develop cloud information that can be used in the NSRDB. We are collaborating with the National Aeronautics and Space Administration's Langley Research Center to explore available satellite data and cloud products that are suitable for extending the NSRDB. Figure 29 shows sample cloud optical thickness over North America based on a combination of polar-orbiting satellites.

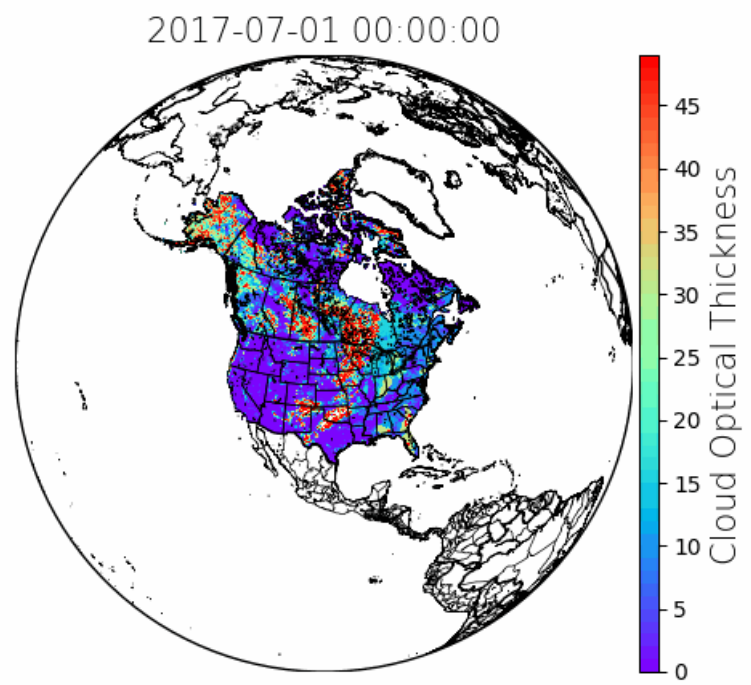

Figure 29. Cloud optical thickness at 0:00 UTC on July 1, 2017

The processing pipeline for the NSRDB contains multiple steps that include both computations and quality control of the data to ensure that (1) the solar data are complete and no gaps are present, (2) ancillary data are developed to meet solar modeling needs, and (3) the dataset is mapped to an appropriate grid to enable solar modeling. We have developed an initial map of the processing pipeline based on the satellite data. Figure 30 shows the associated GHIs computed using the processing pipeline.

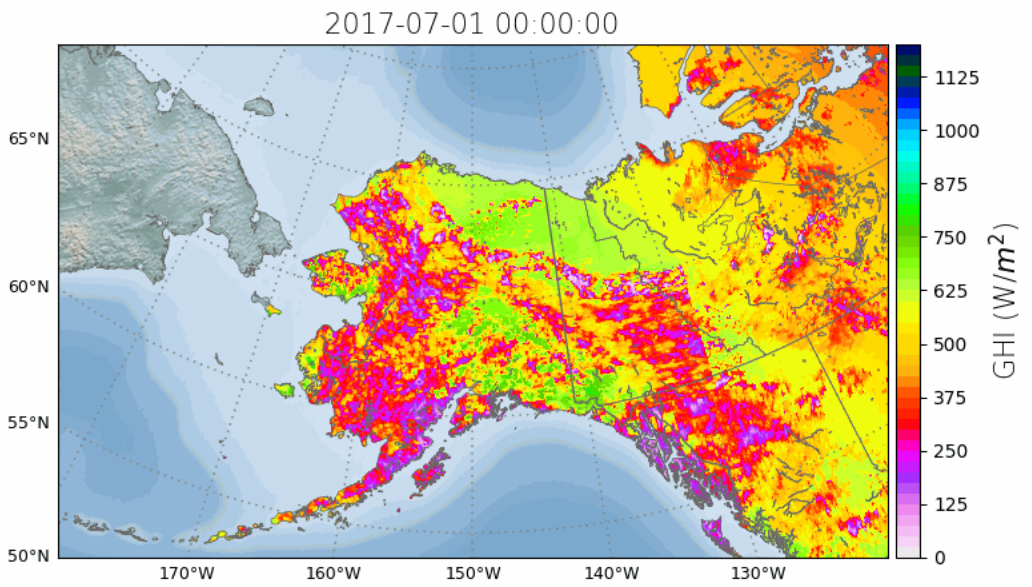

Figure 30. GHI at 0:00 UTC on July 1, 2017

An evaluation study is crucial for developing high-quality solar resource information. We collected the available solar or power data based on surface observations for validating the satellite data. The available surface sites over the Alaska region are given in Figure 31. Five surface sites within the networks of NOAA's Global Monitoring Laboratory (GML), ARM, and the University of Alaska at Fairbanks (UAF) are included. More specifically, NOAA's GML includes one site at Barrow (BRW). ARM includes three sites at Oliktok Point (OLI), the North Slope of Alaska central location (NSA C1), and the North Slope of Alaska (NSA C2). The UAF includes one site at Fairbanks.

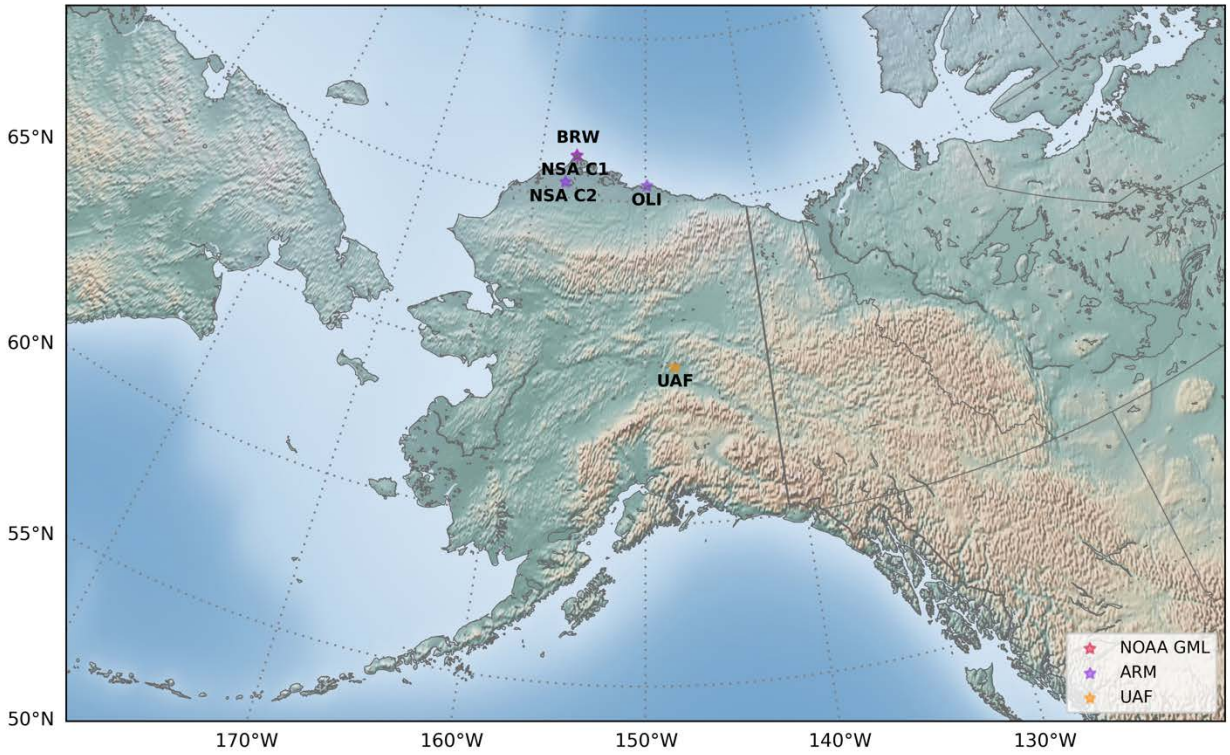

Figure 31. Surface sites over the Alaska region

Following the study, the long-term solar radiation data over Alaska will be developed for 4-km grids.

4 Conclusions

This project supports the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office initiative in making solar energy more affordable for all Americans. Through annual updates to the NSRDB, the data now span 26 years, from 1998–2023. The TMY has been updated to include data in the additional years. By updating the PSM, the accuracy of the data has been significantly improved, as validated using ground observations of solar radiation. Improvements to the NSRDB website and servers ensure seamless access to these datasets for all users. The project's research and development efforts enhance the existing NSRDB data and facilitate the creation of advanced products, including long-term datasets that account for the impact of environmental variability. The availability of high-resolution data with 5-minute temporal and 2-km spatial resolution is expected to improve awareness of the solar resource variability, reduce uncertainty in energy generation, and optimize plant operations.

In FY 2025–2027, we will update the NSRDB annually to extend the data to cover the period from 2024–2026. The TMY will also be updated annually to include additional years (2024–2026). Research and development efforts will focus on improving the PSM and enhancing the accuracy of the operational NSRDB product. The model and data will be validated using ground measurements to ensure accuracy. Due to limitations in the current geostationary satellites, the Alaska region has not been included in the NSRDB. To acquire high-resolution satellite data that can provide continuous coverage over the area, we are collaborating with the National Aeronautics and Space Administration's Langley Research Center to explore available the satellite data and cloud products that are suitable for extending the NSRDB. Long-term solar radiation data over Alaska will be developed and provided to the public in the next NSRDB update. In addition, the NSRDB website and servers will be upgraded to enable the efficient delivery of the new datasets.

References

Augustine, J., and J. DeLuisi. 2000. "SURFRAD—A National Surface Radiation Budget Network for Atmospheric Research." *Bulletin of the American Meteorological Society* 81: 2341–2357.

Buster, G., M. Rossol, G. Maclaurin, Y. Xie, and M. Sengupta. 2021. "A Physical Downscaling Algorithm for the Generation of High-Resolution Spatiotemporal Solar Irradiance Data." *Solar Energy* 216: 508–517.

Habte, A., M. Sengupta, A. Lopez. 2017. *Evaluation of the National Solar Radiation Database (NSRDB): 1998–2015*. Golden, CO: National Renewable Energy Laboratory.

Hicks, B., J. DeLuisi, and D. Matt. 1996. "The NOAA Integrated Surface Irradiance Study (ISIS)—A New Surface Radiation Monitoring Program." *Bulletin of the American Meteorological Society* 77: 2867–2864.

Kim, H., P. Lee, F. Ngan, Y. Tang, H. Yoo, and L. Pan. 2015. "Evaluation of Modeled Surface Ozone Biases as a Function of Cloud Cover Fraction." *Geoscientific Model Development* 8: 2959–2965.

Liou, K. N. 2002. *An Introduction to Atmospheric Radiation, 2nd ed*. Amsterdam; Boston: Academic Press.

Maxwell, E. 1987. *A Quasi-Physical Model for Converting Hourly Global Horizontal to Direct Normal Insolation*. Golden, CO: Solar Energy Research Institute.

Ross, B., and J. Walsh. 1987. "A Comparison of Simulated and Observed Fluctuations in Summertime Arctic Surface Albedo." *Journal of Geophysical Research* 92: 13,115–13,125.

Sengupta, M., Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby. 2018. "The National Solar Radiation Data Base (NSRDB)." *Renewable and Sustainable Energy Reviews* 89: 51–60.

Stoffel, T., and A. Andreas. 1981. NREL Solar Radiation Research Laboratory (SRRL): Baseline Measurement System (BMS). Golden, Colorado (Data).

Xie, Y., and Y. G. Liu. 2013. "A New Approach for Simultaneously Retrieving Cloud Albedo and Cloud Fraction From Surface-Based Shortwave Radiation Measurements." Environ. Res. Lett. 8, doi.org/10.1088/1748-9326/8/4/044023

Xie, Y., M. Sengupta, and J. Dudhia. 2016. A Fast All-sky Radiation Model for Solar Applications (FARMS): Algorithm and Performance Evaluation." *Solar Energy* 135: 435–445.

Xie, Y., M. Sengupta, Y. Liu, H. Long, Q. Min, W. Liu, and A. Habte. 2020. "A Physics-Based DNI Model Assessing All-Sky Circumsolar Radiation." *iScience* 22. doi.org/10.1016/j.isci.2020.100893.

Xie, Y., J. Yang, M. Sengupta, Y. Liu, and X. Zhou. 2022. “Improving the Prediction of DNI With Physics-Based Representation of All-Sky Circumsolar Radiation.” *Solar Energy* 231: 758–766.