Development of a MODIS-Derived Surface Albedo Data Set: An Improved Model Input for Processing the NSRDB

Galen Maclaurin, Manajit Sengupta, Yu Xie, and Nicholas Gilroy
National Renewable Energy Laboratory
Development of a MODIS-Derived Surface Albedo Data Set: An Improved Model Input for Processing the NSRDB

Galen Maclaurin, Manajit Sengupta, Yu Xie, and Nicholas Gilroy

National Renewable Energy Laboratory

Prepared under Task No. ST6P2410
NOTICE

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.

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

Available electronically at SciTech Connect http://www.osti.gov/scitech

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
OSTI http://www.osti.gov
Phone: 865.576.8401
Fax: 865.576.5728
Email: reports@osti.gov

Available for sale to the public, in paper, from:
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312
NTIS http://www.ntis.gov
Phone: 800.553.6847 or 703.605.6000
Fax: 703.605.6900
Email: orders@ntis.gov

Cover Photos by Dennis Schroeder: (left to right) NREL 26173, NREL 18302, NREL 19758, NREL 29642, NREL 19795.

NREL prints on paper that contains recycled content
Acknowledgments

The SunShot National Laboratory Multiyear Partnership (SuNLaMP) program of the U.S. Department of Energy (DOE) funded this project. We thank Andrew Clifton, Sarah Kurtz, and Mike Dooraghi for their input and assistance. We also acknowledge Billy Roberts, who created the figures and maps for this publication. Opinions represented in this report are the authors’ own, and they do not reflect the view of DOE or the U.S. government. This work was supported by DOE under Contract No. DE-AC36-08GO28308 with the National Renewable Energy Laboratory.
## List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRDF</td>
<td>Bidirectional Reflectance Distribution Function</td>
</tr>
<tr>
<td>GHI</td>
<td>Global horizontal irradiance</td>
</tr>
<tr>
<td>IMS</td>
<td>Integrated Multisensor Snow and Ice Mapping System</td>
</tr>
<tr>
<td>MERRA</td>
<td>Modern Era Retrospective-Analysis for Research and Applications</td>
</tr>
<tr>
<td>MODIS</td>
<td>Moderate Resolution Imaging Spectroradiometer</td>
</tr>
<tr>
<td>NREL</td>
<td>National Renewable Energy Laboratory</td>
</tr>
<tr>
<td>NSRDB</td>
<td>National Solar Radiation Database</td>
</tr>
<tr>
<td>POA</td>
<td>Plane-of-array</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root mean square error</td>
</tr>
</tbody>
</table>
Executive Summary

A significant source of bias in the transposition of global horizontal irradiance to plane-of-array (POA) irradiance arises from inaccurate estimations of surface albedo. The current physics-based model used to produce the National Solar Radiation Database (NSRDB) relies on model estimations of surface albedo from a reanalysis climatology produced at relatively coarse spatial resolution compared to that of the NSRDB. As an input to spectral decomposition and transposition models, more accurate surface albedo data from remotely sensed imagery at finer spatial resolutions would improve accuracy in the final product.

The National Renewable Energy Laboratory (NREL) developed an improved white-sky (bi-hemispherical reflectance) broadband (0.3–5.0 $\mu$m) surface albedo data set for processing the NSRDB from two existing data sets: a gap-filled albedo product and a daily snow cover product. The Moderate Resolution Imaging Spectroradiometer (MODIS) sensors onboard the Terra and Aqua satellites have provided high-quality measurements of surface albedo at 30 arc-second spatial resolution and 8-day temporal resolution since 2001. The high spatial and temporal resolutions and the temporal coverage of the MODIS sensor will allow for improved modeling of POA irradiance in the NSRDB. However, cloud and snow cover interfere with MODIS observations of ground surface albedo, and thus they require post-processing. The MODIS production team applied a gap-filling methodology to interpolate observations obscured by clouds or ephemeral snow. This approach filled pixels with ephemeral snow cover because the 8-day temporal resolution is too coarse to accurately capture the variability of snow cover and its impact on albedo estimates. However, for this project, accurate representation of daily snow cover change is important in producing the NSRDB. Therefore, NREL also used the Integrated Multisensor Snow and Ice Mapping System data set, which provides daily snow cover observations of the Northern Hemisphere for the temporal extent of the NSRDB (1998–2015). We provide a review of validation studies conducted on these two products and describe the methodology developed by NREL to remap the data products to the NSRDB grid and integrate them into a seamless daily data set.
Table of Contents

1 Introduction ........................................................................................................................................... 1

2 MODIS Surface Albedo ........................................................................................................................ 3
   2.1 Validation of the BRDF/Albedo Products ..................................................................................... 3
   2.2 MCD43D: BRDF/Albedo Product ................................................................................................. 3
   2.3 MCD43GF: Gap-Filled Albedo Product ....................................................................................... 3

3 Spatial Aggregation of the Gap-Filled Product ................................................................................. 5
   3.1 Assessment of Variability within Aggregation Windows ............................................................. 5

4 IMS Snow Cover .................................................................................................................................... 7
   4.1 Validation of the IMS Snow Cover Data Set ................................................................................ 7
   4.2 Nearest-Neighbor Mapping of the IMS to the NSRDB Grid ........................................................ 7

5 Integration of MODIS Albedo and IMS Snow Cover .......................................................................... 9

6 Discussion and Conclusions ............................................................................................................. 11

7 Future Work ......................................................................................................................................... 12

References ................................................................................................................................................. 13
List of Figures

Figure 1. Conceptual diagram of aggregation method. The blue dots represent the MCD43GF pixel centroids over water, and the red dots represent the centroids of land pixels. The black squares represent NSRDB pixels. ................................................................. 5

Figure 2. Map of the mean and standard deviation within each aggregation window for January 1, 2001. 6

Figure 3. Detailed view of the Great Lakes region from Figure 2. ................................................................. 6

Figure 4. Final data set for January 1, 2001, integrating MODIS-based surface albedo and IMS snow cover. ....................................................................................................................................... 10
1 Introduction

Albedo is the ratio of reflected solar energy (shortwave radiation) from a surface to the incident radiation on that surface. Albedo influences the radiative budget of the Earth, and it is an essential input in modeling solar radiation. Improved quantification of solar irradiance across space and time reduces uncertainty in the estimate of photovoltaic performance. A significant source of bias in the transposition of global horizontal irradiance (GHI) to plane-of-array (POA) irradiance—the latter provides the incident energy on the deployment plane of photovoltaic arrays—arises from inaccurate estimations of surface albedo. The current physics-based model used to produce the National Solar Radiation Database (NSRDB) relies on estimates of surface albedo from the Modern Era Retrospective-Analysis for Research and Applications (MERRA) data set version 1 at $\frac{1}{2}$- by $\frac{2}{3}$-degree spatial resolution (National Aeronautics and Space Administration 2016). Because ground reflections form a larger component of the radiation incident on tilted planes, more accurate surface albedo data from satellite remote sensing at finer spatial resolution will result in improved accuracy when estimating POA irradiance.

The National Renewable Energy Laboratory (NREL) developed an improved white-sky (bi-hemispherical reflectance) broadband (0.3–5.0 μm) surface albedo data set for use in NSRDB applications using two data sources: (1) the Bidirectional Reflectance Distribution Function (BRDF)/albedo product (designated MCD43D) from the Moderate Resolution Imaging Spectroradiometer (MODIS) (National Aeronautics and Space Administration, Land Process Distributed Active Archive Center 2015) and (2) the Integrated Multisensor Snow and Ice Mapping System (IMS) daily snow cover product (National Snow and Ice Data Center 2016). The MCD43D product provides albedo estimates in 7 MODIS bands and 3 broadbands at 30 arc-second spatial resolution for an 8-day compositing period; however, significant cloud contamination and ephemeral snow coverage limit the utility of the data for many applications. Sun (2014) developed an algorithm to fill cloud, ephemeral snow, and poor-quality retrieval pixels from the MCD43D product. The gap-filled product (designated MCD43GF) is provided by the University of Massachusetts, Boston, from 2001–2014 as part of the MODIS Version 5 collection (2015). MCD43GF was produced at the same spatial and temporal resolutions as the MCD43D. Pixels with ephemeral snow cover were removed in the gap-filling procedure because the 8-day temporal resolution is too coarse to accurately capture the variability of snow cover and its impact on surface albedo measurements. The gap-filled albedo product provides cloud- and snow-free global coverage from 2001–2014 (Sun 2014). The IMS data set provides gap-free, daily snow cover observations (as a Boolean variable) across the Northern Hemisphere. From 1997–2004, the data set was produced at 24-km spatial resolution; starting in 2004, it has been produced at a 4-km spatial resolution. Analysts from the National Oceanic and Atmospheric Administration operate the semiautomated IMS system, which integrates visible and microwave satellite imagery with ground station measurements and other ancillary information to provide real-time daily releases of the data set. The IMS data was chosen instead of other snow cover products based on its provision of gap-free spatial coverage. The main alternatives were MODIS snow cover products, which suffer from significant cloud contamination and thus require a procedure to fill in missing observations. A gap-filled MODIS snow cover data set was not available at the time of this project.

The MCD43GF was reprocessed by the National Renewable Energy Laboratory (NREL) to match the 0.04-degree (approximately 4-km) spatial resolution of the NSRDB. MCD43GF pixels
that fell within the spatial extent of an NSRDB pixel were aggregated using the mean. The standard deviation was also calculated on the MCD43GF pixels within each aggregation window to assess the degree of smoothing (i.e., generalization) induced by averaging. Daily snow cover from the IMS data set was mapped to the NSRDB pixels using a nearest-neighbor method.
2 MODIS Surface Albedo

The BRDF/albedo products provide surface albedo estimated from MODIS instruments onboard the Terra and Aqua satellites. A large volume of literature has assessed the accuracy of the MODIS surface albedo products using various methodologies. The studies reviewed in this section found that the MODIS albedo products generally meet an absolute accuracy of 0.02. The BRDF/albedo products are produced using the Terra platform alone for 2001 and 2002, and they are the combined Terra and Aqua platforms from 2003 onward.

2.1 Validation of the BRDF/Albedo Products

Extensive assessments have been conducted on the surface albedo products from MODIS Level 2 and Level 3 data. Absolute accuracy of surface albedo products between 0.02–0.05 is required for climate modeling applications (Schaaf et al. 2011). The spatial resolution (500 m or 1 km) of these products makes validation with surface measurements challenging. A satellite observation represents the area covered by a grid cell, which can differ significantly from a ground measurement at a single location within that grid cell, particularly in heterogeneous landscapes. To overcome this difficulty, many studies rely on scaling up to MODIS resolutions by combining surface measurements with data from Landsat 5 or 7 satellites (30-m resolution). Using this approach, Liang et al. (2002) reported absolute errors less than 0.05 for Terra Level 2 surface reflectance (MOD09) and Level 3 broadband albedo (MOD43B3) products. A subsequent study using measurements from seven ground stations throughout the United States between April 2001 and September 2001 found that the Terra-only MODIS BRDF/albedo products meet an absolute accuracy of 0.02 (Jin et al. 2003). Specific regional studies also found similar levels of accuracy in paddy fields of Japan (Susaki et al. 2007) and a semidesert region of the Tibetan Plateau (Wang et al. 2004) with root mean square errors (RMSEs) of 0.026 and 0.0186, respectively. Salomon et al. (2006) compared “combined” Aqua and Terra products to Terra products and found that their absolute accuracies were both within 0.02. The combined product has a slight improvement, with an RMSE of 0.013, compared to 0.014 of the Terra product.

2.2 MCD43D: BRDF/Albedo Product

The BRDF/albedo algorithm is a kernel-based semiempirical model used to produce the MCD43D, which—among other properties—provides directional hemispherical reflectance (black-sky albedo) and bi-hemispherical reflectance (white-sky albedo) from MODIS surface reflectance observations (Schaaf et al. 2002; Schaaf et al. 2008). MCD43D provides quality assessment (QA) flags indicating two methods of estimation. High-quality albedo estimates relied on the availability of both Terra and Aqua with sufficient observations. Low-quality estimates were based on the BRDFs for that pixel from previous high-quality retrievals and adjusted with the available observations (Strugnell and Lucht 2001; Strugnell, Lucht, and Schaaf 2001). Cloud contamination and the presence of ephemeral snow were significant and limited the utility of the data set for many applications.

2.3 MCD43GF: Gap-Filled Albedo Product

Gap-filling algorithms are commonly used to fill cloud-contaminated pixels using interpolation or smoothing filters (e.g., Gao et al. 2008; Hall et al. 2010). Sun (2014) developed the Version 5 MCD43CF by filling pixels with cloud cover or ephemeral snow in the MCD43D product. The
algorithm is composed of three prioritized methods. The first attempt was a temporal fitting driven by the MCD43D high quality QA flag values. The temporal method used an asymmetric Gaussian fitting with a window of 76 time steps to produce spatially and temporally continuous albedo estimates. The temporal fit filled more than 97% of the missing land values. When insufficient high-quality values were available in the time series for the temporal fit, the second method applied a spatial fitting based on nearby locations with the same type of land cover. This approach generalized the seasonal curve of albedo based on the time series of these neighboring pixels. The missing values were then filled by adjusting the seasonal curve with the available data points in the pixel’s time series. This approach filled less than 1% of the missing land values. Finally, the third method applied a spatial smoothing to fill the remaining gaps when a pixel did not have a single value within the fitting window. Missing values were assigned the mean of nearby pixels with the same type of land cover, filling the remaining 2% of the missing land values. Pixels in the Intertropical Convergence Zone, particularly in parts of South America and Africa, suffered from residual cloud contamination that was an artifact of the aggregation procedure of the Version 5 MCD43D product, and thus they required additional processing. Prior to the gap-filling procedure, an outlier-detection algorithm was applied to remove values in the time series with residual cloud contamination.

Sun (2014) performed an assessment on the MCD43GF product by randomly selecting 2% of all land pixels and then selecting one high-quality value at random from each of the time series. These high-quality values were removed from the time series, and the gap-filling method was applied. This assessment framework was applied to Band 2 (near-infrared) from 2010, which was expected to have the largest errors due to the high variability of vegetation in the near-infrared spectrum. The RMSE for the temporal fit of surface albedo was 0.024 (from 3,462,982 pixels). The spatial fit RMSE was 0.007, calculated from 526 pixels; and the RMSE for the spatial smoothing method was 0.029, from 24,683 pixels. The distribution of residuals (i.e., the difference between the filled and high-quality values) was normal, with a mean of 0.005 and standard deviation of 0.026.
3 Spatial Aggregation of the Gap-Filled Product

From the MCD43GF product, NREL reprocessed white-sky albedo in the shortwave band (0.3–5.0 μm) to match the NSRDB grid. Aggregation was performed using a point-in-polygon approach, wherein MCD43GF pixel centroids that fell within an NSRDB pixel were used to calculate the mean values (Figure 1). Water pixels in the MCD43GF were excluded from the calculation. Within aggregation windows completely on land, between 16 and 25 MCD43GF pixel centroids were aggregated to each NSRDB cell. Approximately 46 million MCD43GF pixels were aggregated to the 2 million NSRDB cells, with an average of 23 MCD43GF pixels per NSRDB cell.

![Conceptual diagram of aggregation method](image)

**Figure 1.** Conceptual diagram of aggregation method. The blue dots represent the MCD43GF pixel centroids over water, and the red dots represent the centroids of land pixels. The black squares represent NSRDB pixels.

3.1 Assessment of Variability within Aggregation Windows

For each 8-day time step, the standard deviation of the MCD43GF values within each aggregation window was assessed for the geographic extent of the NSRDB. Variability in the aggregated data as a whole (i.e., across space and time) was low: 88.3% of all aggregation windows had a standard deviation of less than 0.02, showing that the degree of smoothing (i.e.,
generalization of the spatial structure) was low, and thus the overall spatial distribution of the surface albedo in the MCD43GF was preserved. For any given 8-day time step, the highest variability occurred in aggregation windows with steep terrain, along linear features (such as rivers or shorelines) or in areas with permanent snow and ice (Figure 2 and Figure 3).

Figure 2. Map of the mean and standard deviation within each aggregation window for January 1, 2001

Figure 3. Detailed view of the Great Lakes region from Figure 2
4 IMS Snow Cover

The IMS product provides daily snow cover mapping across the Northern Hemisphere at multiple resolutions from 1997 to the present (National Snow and Ice Data Center 2016). In a semiautomated procedure, system analysts from the Office of Satellite and Product Operations at the National Oceanic and Atmospheric Administration utilize visible and microwave satellite imagery, ground station observations, automated snow mapping algorithms, and other ancillary data to produce the IMS product. Analysts rely primarily on the visible channels from the Polar-Orbiting Operational Environmental Satellites and the Geostationary Orbiting Environmental Satellites to determine the presence of ephemeral snow cover. MODIS visible channels are used to supplement these primary sources. Ground station weather observations over the Northern Hemisphere are used for configuration and verification. Microwave imagery—from the Polar-Orbiting Operational Environmental Satellites’ Advanced Microwave Sounding Unit and the U.S. Department of Defense’s Defense Meteorological Satellite Program—is used to improve snow cover mapping when significant cloud contamination is present (National Snow and Ice Data Center 2016). Daily IMS snow cover observations are made by updating the previous day’s map, adjusting areas where snow has accumulated and melted. Since 1997, the IMS data have been available at a 24-km spatial resolution; starting in 2004, they are available at a 4-km spatial resolution (Helfrich Li, and Kongoli 2012; Helfrich et. al. 2007).

4.1 Validation of the IMS Snow Cover Data Set

Validation studies of the IMS snow cover product primarily used ground station observations to assess the 24-km and the 4-km data sets. A fundamental challenge in such a comparison is that although the IMS data provide Boolean observations for pixels, significant variability in snow cover could be present within a pixel. This is more of a concern during periods of snow accumulation or ablation and in areas with variable snow cover. Brubaker, Pinker, and Deviatova (2005) conducted a comprehensive validation of the IMS using North American ground station measurements as “ground truth” data and for an assessment compared to the MODIS snow product. The methodology assessed IMS pixels containing multiple ground stations and treated the station locations as Bernoulli trials, i.e., random samples from a binary distribution (snow or no snow). Results showed snow cover agreement with the ground stations of approximately 80% in November, 90% in December, and 80% again in the early spring. Comparison to cloud-free pixels from the MODIS snow product showed similar seasonal trends, with agreements of more than 90% during the winter and early spring months, but it was lower in late fall. A study in Turkey compared ground station data to IMS observations and reported an approximate 74% snow cover accuracy across all seasons and approximately 84% in the winter (Sönmez, Tekeli, and Erdi 2014). A positive correlation was suggested between the accuracy of the IMS data sets and the depth of the snow pack measured at ground stations. The seasonal trends in the snow cover accuracy were similar: fall and spring observations had lower accuracy than those during the winter (Brubaker, Pinker, and Deviatova 2005; Sönmez, Tekeli, and Erdi 2014).

4.2 Nearest-Neighbor Mapping of the IMS to the NSRDB Grid

The IMS daily snow cover data were mapped to the NSRDB grid using a nearest-neighbor method. With the coarser 24-km IMS data, multiple NSRDB pixels fell within each IMS pixel and therefore were assigned the same snow cover flag. Starting in February 2004, the 4-km IMS
data provided a one-to-one mapping with the NSRDB, and thus it showed higher fidelity in the representation of snow cover.
5 Integration of MODIS Albedo and IMS Snow Cover

From 1998 through 2015, 25 days were not mapped by the IMS system, and they required gap filling. Single-day gaps in the data set that did not present a change in snow cover (i.e., snow cover before and after or no snow cover before and after) were filled with the persistent value, making the assumption that no change occurred during the 3-day period. Single-day gaps that presented a change in snow cover before and after and multiday gaps were filled using snow cover from the MERRA Version 1 data set (National Aeronautics and Space Administration 2016).

In the final processing step, the aggregated MODIS albedo values and the processed IMS snow cover flags were integrated to produce a daily albedo data set. The aggregated MODIS albedo data were replicated temporally to produce daily values from the 8-day time steps. The temporal variability of land surface albedo is low within an 8-day time period, so the daily replication of each value provides sufficient detail for the NSRDB transposition models. Finally, the IMS data were used to adjust the daily albedo values when snow cover was present. A constant for snow cover albedo was calculated based on the variability of reflectance as a function of snow grain size under clear-sky and overcast conditions. The mean albedo under these two conditions for the spectral region from 0.3–4.0 µm was 0.8669 (Dang, Brandt, and Warren 2015). This value was applied to all snow-covered pixels in the integrated data set.
Figure 4. Final data set for January 1, 2001, integrating MODIS-based surface albedo and IMS snow cover
6 Discussion and Conclusions

The daily albedo data set provides a significant improvement from climatological products in spatial resolution and representation of the temporal variability of albedo. Although the gap-filled MODIS albedo product provides high-quality estimates of surface albedo, the 8-day temporal resolution does not capture the ephemeral nature of snow cover. Changes in snow cover play a major role in the variability of surface albedo, particularly during months of snow accumulation and ablation (Déry and Brown 2007). Integration of the gap-filled MODIS albedo product and the IMS snow cover data set provide high-quality albedo estimates and fine temporal resolution of snow cover.

As an input to spectral decomposition and transposition models, daily representation of snow cover is a required quantity. Representing the large difference between surface albedo with and without snow cover is expected to improve the performance of these models. Although the data set produced here aimed to capture that difference spatially and temporally, uncertainties in the gap-filled MODIS albedo and the snow cover products could have a nontrivial modeling impact. Nonetheless, the temporal extent of the NSRDB (currently from 1998–2015) limits the options for high-quality albedo and snow cover data sets.

Using a snow albedo constant in this data set is a limitation, and representing the variability of snow albedo would improve the final data set. Although a MODIS snow albedo data set is available (MOD10A1), the gap filling of cloud-contaminated pixels has not been planned in the MODIS Collection 6 release. Justification to undertake such a modeling effort would require an examination of potential improvements in the final NSRDB product.
7 Future Work

Future work will assess the impact of uncertainties in the gap-filled MODIS albedo product and the IMS snow cover data set on the modeling of the NSRDB. This will be used to guide the approach for future versions of the surface albedo data set.

The MODIS gap-filled albedo product is available for multiple spectral bands at visible through near-infrared wavelengths (Sun 2014), which can support more detailed assessments of photovoltaic performance. Future work is needed to extend broadband surface albedo to those spectral bands.
References


National Snow and Ice Data Center. 2016. “IMS Daily Northern Hemisphere Snow and Ice Analysis at 1 km, 4 km, and 24 km Resolutions, Version 1” (24-km and 4-km Data File). Boulder, CO. doi: [http://dx.doi.org/10.7265/N52R3PMC](http://dx.doi.org/10.7265/N52R3PMC).


