COMPLETING PRODUCTION OF THE UPDATED NATIONAL SOLAR RADIATION DATABASE FOR THE UNITED STATES

Stephen Wilcox  
Mary Anderberg  
Ray George  
William Marion  
Daryl Myers  
David Renné  
National Renewable Energy Laboratory  
1617 Cole Boulevard  
Golden, CO 80401  
E-mail: stephen_wilcox@nrel.gov  
Neal Lott  
Tom Whitehurst  
National Climatic Data Center

William Beckman  
University of Wisconsin  
Christian Gueymard  
Solar Consulting Services  
Richard Perez  
State University of New York at Albany  
Paul Stackhouse  
National Aeronautics and Space Administration  
Frank Vignola  
University of Oregon

ABSTRACT
The National Renewable Energy Laboratory (NREL) and the National Climatic Data Center (NCDC) have released the 1991–2005 National Solar Radiation Database (NSRDB), an update of the 1961–1990 NSRDB released in 1992. The updated NSRDB data sets are available in several forms, including an hourly ground-based data set of solar and meteorological fields for 1454 stations, an hourly 10-km gridded time series data set of solar data based on satellite modeling (1998–2005), and summary statistical files. Eight hundred fifty eight sites have a complete 15-year period of record, and another 596 sites have data for at least 3 of the 15 years (though not necessarily serially complete). All of the 239 original NSRDB sites are included in the update. Data sets including both solar and meteorological data are available from NCDC, and a research data set with extended solar and related fields is available from NREL. Fees to access data sets that contain meteorological fields may apply, according to the NCDC distribution infrastructure.

1. INTRODUCTION

Shortly after completion of the 1961–1990 NSRDB (1) in 1992, NREL began receiving requests for an NSRDB update from users interested in more recent data or different locations. Because the National Weather Service was switching meteorological observations from human to automated measurements—a move that significantly altered the character of the cloud data used as the primary input to the NSRDB solar model—the update was delayed. In 2004, as demand for an update grew and the nature of the automated meteorological observations became better known, NREL convened an expert committee to investigate the feasibility of an NSRDB update (2).

Committee members recognized the seriousness of the change in cloud data and noted that, although data automation benefited aviation operations, the amount of climatological information characterizing sky cover had been reduced. However, the committee also saw opportunities for an enhanced NSRDB using a recently developed solar model based on satellite imagery as well as improvements in radiometry and modeling.

Work commenced to produce a proof-of-concept database to evaluate the feasibility of an update. Primary funding came from NREL, with support from the Atmospheric Sciences Research Center, State University of New York at Albany; Climate Systems Branch, National Aeronautics and Space Administration; NCDC, U.S. Department of Commerce; the Northeast Regional Climate Center, Cornell University; Solar Consulting Services; and the Solar Radiation Monitoring Laboratory, University of Oregon. From this proof-of-concept project came the production design for a full-scale update to the NSRDB.

Because the quality of the input cloud data varied by time and location, a multi-tiered scheme was developed to cate-
gorize sites by data availability and quality. The three classes of stations allow users to match data to the needs of specific applications or carry through uncertainties to other derived data sets.

2. SIGNIFICANT STEPS IN NSRDB PRODUCTION

2.1 Cloud Observations

The change from human-observed to automated sky cover measurements presented the greatest challenge to an updated NSRDB. The meteorological-statistical (METSTAT) (3) solar model used for the 1961–1990 NSRDB required observations of both total and opaque cloud cover. Although total cloud cover is well correlated with solar irradiance, the inclusion of opaque cloud cover allows the model to distinguish between translucent clouds that scatter and absorb lightly and opaque clouds that more effectively block irradiance. The lack of this distinguishing characteristic hampers the ability to model direct normal and diffuse irradiance—two important parameters for many solar applications.

The National Weather Service began deploying the Automated Surface Observing System (ASOS) in the early 1990s and had largely completed the project by the late 1990s. ASOS equipment uses an up-facing ceilometer to record passing clouds at the zenith, in place of an observer viewing the entire sky dome to estimate sky cover. The ceilometer cannot detect clouds above about 3660 m, which results in underestimation of cloud cover in most climates and a failure to detect high, thin clouds critical for the METSTAT model. Further, in 1996, the National Weather Service changed the units for sky cover observations from tenths of sky to eighths of sky (oktas). And with the deployment of the ASOS, the observations were reported in a clear/few/scattered/broken/overcast scheme that did not translate to completely use the okta range (0 to 8). These developments thus reduced the resolution of sky cover data by more than half.

To mitigate these shortcomings, NCDC's ASOS Cloud Data Set (4) (also called the Supplemental Cloud Product), was used. It is a satellite-based product that holds data for high cloud layers. The ground-based ASOS sky estimates up to 3660 m were combined with the high cloud estimates from the Supplemental Cloud Product, and then general assumptions were made about the frequency of opaque clouds at each layer to derive total and opaque sky cover estimates for input to the METSTAT model.

Meteorological data were supplied by NCDC in the Integrated Surface Database (ISD) (5), formerly known as the Integrated Surface Hourly data set. The Supplemental Cloud Product does not include data for all ISD locations, and among NSRDB sites, they were not available for about half, including all sites in Alaska and most in Hawaii. For sites without the Supplemental Cloud Product, tables of probability relations between total and opaque cloud amounts were developed based on historical manual cloud observations. These relations were used to estimate total and opaque cloud amounts based on the ASOS measurement.

Military installations included in the ISD have continued manual cloud observations to present day as part of routine operations, but they have not reported opaque clouds at any time of their period of record. For these sites, the total and opaque probability tables of nearby sites were used to derive cloud estimates for the METSTAT model.

2.2 Other Atmospheric Components

2.2.1 Aerosols. After clouds and solar position, atmospheric aerosols are the most significant input for solar models. For applications that rely on clear-sky solar irradiance, the quality of aerosol optical depth (AOD) estimates significantly affects the accuracy of modeled solar data.

A combination of surface sun photometry, satellite data from the NASA Multi- Angle Imaging SpectroRadiometer (MISR) and Moderate Resolution Imaging Spectroradiometer (MODIS) satellite-based instruments, and legacy DNI estimates of broadband aerosol optical depth (BAOD) allowed creation of monthly mean estimates of BAOD for all locations in the United States. Spectral AOD data were converted into BAOD using estimated Ångström Alpha parameters from sun photometry (6). These monthly mean values were interpolated to daily values using a mean-preserving algorithm (7).

Using monthly mean BAOD values (or daily interpolated data) rather than specific daily or hourly values is likely to introduce significant errors in direct and diffuse hourly irradiance predictions under clear skies, but the approach used here is conducive to overall better accuracy than the annual sine function method used to estimate aerosols for the 1961–1990 NSRDB.

Aerosol values for the period of July 11, 1991, through Dec. 31, 1994, were adjusted with an additive factor, AODstrat, in a two-part function to represent the effect of additional stratospheric aerosols from the Mount Pinatubo eruption. The first part is a ramping-up function:

\[
AOD_{\text{strat}} = \left( \frac{AOD_{\text{max}}}{\text{Norm}} \right) \times (1.0 - \exp(-T/T1))
\]

where

- \( AOD_{\text{max}} \) is the function limit (set to 0.2 for all sites)
- \( \text{Norm} \) is a normalizing parameter
T is the number of days since Jan. 1, 1991
T1 is the last day of the ramp function (a site-specific parameter as number of days since Jan. 1, 1991).

The second part is a decay function that starts at the end of the ramp-up function (T1):

$$AOD_{\text{strat}} = AOD_{\text{max}} \cdot \exp \left( -\frac{(T-T1)}{T2} \right)$$

where

- T2 is a daily decay constant (250 for all sites).

A sample plot of the AOD additive functions appears in Figure 1.

![Pinatubo AOD function for Lihue, Hawaii](image)

2.2.2 Water Vapor. The NASA Water Vapor Project (NVAP) was used for daily estimates of water vapor on a 1°-x-1° grid. NVAP integrates sounding data with satellite measurements of atmospheric water vapor. These data were interpolated in space to the location of each of the NSRDB stations and interpolated in time to provide hourly data for all stations for the years 1991–2001.

Because of the limited period of record of the NVAP data set, the years 2002–2005 required a new source of data. For this, the recently released North American Regional Re-analysis (NARR) was chosen. Geographic information system analysis methods were used to match each of the NSRDB stations with its nearest NARR grid point. Then, data from the NARR grids for each 3-hour period were downloaded and interpolated to hourly data files for each station.

2.2.3 Ozone. Total column ozone is derived from daily satellite observations from the Total Ozone Mapping Scanner (TOMS). Scans occur once per day on a grid with spatial resolution of 1° in latitude and 1.25° in longitude. The missing data are replaced with long-term mean values for that location.

2.3 Modeled Solar Data

Two models were used for the updated NSRDB: the METSTAT model (used for the original NSRDB), and a satellite-based model developed by the Atmospheric Sciences Research Center, State University of New York at Albany, called the SUNY model.

2.3.1 METSTAT model. Since the production of the original NSRDB, the METSTAT model has been modified several ways:

- The multiple-reflectance algorithm was modified for better performance under overcast skies by using ceiling height as an input rather than present weather (8, 9).
- A software switch that adjusts for the human perception of cloud amounts nearer the horizon being greater than actual was added for the perspective function. This function was turned off when the model was running on ASOS-based data, which is determined by examining only clouds directly overhead.
- When running the model with the statistics “on,” cloud cover amounts were not randomized if the total sky cover was 0 or opaque sky cover was ten-tenths. (If opaque sky cover is ten-tenths, the opaque sky cover should not be randomized to a value less than 10, which results in direct normal values greater than 0 for overcast skies. Similarly, if the total sky cover is zero-tenths, it should not be randomized to a greater value). Also eliminated was the procedure that modifies the random number for consecutive clear or cloudy hours.
- Cloud indexing changed to allow for fractional cloud amounts.

2.3.2 SUNY Model. This model uses images from the East and West Geostationary Operational Environmental Satellites (GOES) to estimate irradiance reaching the ground (10). For the NSRDB update, the SUNY model produced data for 1998–2005. Data for earlier years could not be produced because the project began archiving GOES images only in 1998. The cost of historic satellite images was prohibitive, and the model also requires ancillary data sets that do not exist for earlier years.

Production efforts indicated that the SUNY model performed better and more consistently than the METSTAT model for those times when ASOS or filled cloud data were used for METSTAT input. The period of the SUNY output...
data (1998 and later for most areas) overlapped with many of these suboptimal cloud measurement periods. Thus, the SUNY model output was given preference over METSTAT output for the database of ground stations.

The acquisition times for the GOES images used by the model did not correspond with the hour-ending convention of the NSRDB (solar values at the NSRDB hourly time stamp represent the solar radiation integrated over the previous hour). The GOES East satellite acquires images at 15 minutes after the hour, and the GOES West satellite acquires at the top of the hour. These roughly translate into an instantaneous irradiance value from the model at non-conventional times. To conform to the NSRDB convention, when the SUNY-produced data were included in the ground conventional times. To conform to the NSRDB convention, when the SUNY-produced data were included in the ground station data set, values were shifted in time using a method based on normalized airmass and solar geometry (10). The 10-km gridded data set includes shifted and unshifted data.

### 2.4 Measured Solar Data

Because of the enormous expense of measuring solar radiation, geographic coverage of such data is sparse, and only a small fraction of the NSRDB data sets include measured data. About 7% of the 1961–1990 NSRDB contained measured solar radiation data, and the updated NSRDB contains even less. Forty sites in the NSRDB update include measured solar radiation data, with only a few that cover the entire 15-year period of record (and among those, none with a serially complete data set).

One significant difference between the 1961–1990 NSRDB and the 1991–2005 update is how the measured data are stored in data records. In the original NSRDB, the measured data were merged with the modeled data such that a seamless data set of solar radiation values was produced—with the model essentially filling gaps in the measured data. However, differing characteristics of measured and modeled solar data brought forth some user complaints about this method. The measured data were imperfect because the three components of global, direct, and diffuse—which are mathematically coupled—did not couple because of known and unavoidable measurement uncertainties in radiometers. This was most notable when the diffuse measurement would occasionally slightly exceed the global measurement under overcast skies. Modeled data, on the other hand, although having mathematically coupled components, often fails to accurately track changes in hourly irradiance because of a lack of information about the relation between the sun and clouds.

For this reason, fields were included for both measured and modeled data to allow the user the flexibility of choosing either modeled or (if available) measured data for an application. All measured solar data in the NSRDB update are quality-assessed and flagged. The SERI QC flagging scheme (11) allows the user to filter for a range of quality depending on the requirements of an application. Only data with egregious errors were removed from the data set.

Several sites use rotating shadowband radiometers for measurements. These instruments suffer from a spectral nonlinearity that often results in inaccurate measurements. These data were modified with a spectral correction procedure to minimize errors (12).

### 2.5 Data Filling

One goal of the updated NSRDB was to have serially complete solar and associated meteorological fields for the 15-year period of record. Gaps in data records are inevitable for large meteorological data sets, and when used as input to a model, these gaps map to gaps in the output data sets. Rather than filling gaps in the solar data, gaps were filled in the model input data set to take advantage of the pseudo-natural variability imposed by the METSTAT model (3) and satisfy the needs of some applications for serially complete meteorological data.

A method of short- and long-term data filling was developed for filling the meteorological data (13). Short gaps of up to 5 hours or at night were filled using linear interpolation (or, for temperature fields, by imposing a diurnal profile on the interpolation line). However, during the 15-year period of record, many sites had occasional periods of missing data that ranged from days to months. These gaps were filled by substituting data from other years using a method that attempted to find the most appropriate source year for substitute data. With these larger gaps filled, the NSRDB included a much larger set of stations. The effect of data filling on the accuracy of the model output was handled during the assignment of data uncertainty (Section 2.6).

A large number of sites also had multi-year gaps (for example, stations that started or ceased operations during the 1991–2005 period of record). No effort was made to fill these large gaps, but if a site had at least 3 years of data, it was included in the NSRDB as a potentially valuable resource for some applications. The sites with an incomplete period of record were segregated, as detailed in Section 2.7.

Because the data filling methods were designed to produce realistic solar radiation values and not to maintain climatologically accurate meteorological values, the filled meteorological data in the NSRDB should not be used for climate or meteorological applications.
2.6 Uncertainty

The NSRDB expert committee knew from the project onset that the uncertainties of the modeled data in the update were likely to be greater than those in the original NSRDB because of the change in the quality of cloud observations. Thus, a realistic estimate of uncertainties was necessary to convey the nature of the data quality to the user.

The hourly NSRDB uncertainty flags (±%) are based on model type and quality of input data. The uncertainty is calculated based on model evaluations (14) that compared predictions with measured ground data. The approach is based on the International Bureau of Weights and Measures and International Organization for Standardization Guide to the Expression of Uncertainty in Measurement (15).

Drawing on experience with solar radiation measurements and instrumentation, uncertainties (U_{meas}) were assigned to the measured validation data set as shown in Table 1.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Additional Uncertainty</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short- and medium-term data filling</td>
<td>4</td>
<td>Data derived from close time proximity</td>
</tr>
<tr>
<td>Long-term filling</td>
<td>14</td>
<td>Based on interannual variability; data comes from like dates in other years</td>
</tr>
<tr>
<td>Cloud probability derivation</td>
<td>4</td>
<td>Differences in total and opaque average less than 1 okta</td>
</tr>
<tr>
<td>Cloud probability from nearby site</td>
<td>4</td>
<td>Differences in total and opaque average less than 1 okta</td>
</tr>
<tr>
<td>ASOS-only</td>
<td>22</td>
<td>Limited okta resolution afforded by the coded ASOS-only data; a cloud measurement value could be in error by two oktas</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
<th>Additional Uncertainty</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-shifting</td>
<td>2</td>
<td>Shifting from satellite time to hourly local time causes some error in the hourly values or daily totals</td>
</tr>
</tbody>
</table>

These uncertainties are roughly double optimal instrumentation uncertainties and include additional uncertainties attributable to common measurement errors that occur at even well-run sites.

From the model validation work, model random and bias uncertainty (U_{mod} and U_{bias}) were assigned for each component of each model, as shown in Table 2.

These measurement and validation uncertainties were root-sum squared for an optimal uncertainty (U_{opt}) for each model (Table 3). For the METSTAT model, these optimal uncertainties are based on performance using cloud values derived from the Supplement Cloud Product—the best cloud data available for the evaluation period. Model performance using human-observed cloud values is assumed to be as good or better, but this was not quantified.

<table>
<thead>
<tr>
<th>Model</th>
<th>Glo/Dif (RMS U_{mod})</th>
<th>Glo/Dif (MBE U_{bias})</th>
<th>Dir RMS (U_{mod})</th>
<th>Dir MBE (U_{bias})</th>
</tr>
</thead>
<tbody>
<tr>
<td>METSTAT</td>
<td>8</td>
<td>2</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>SUNY</td>
<td>5</td>
<td>0</td>
<td>14</td>
<td>1</td>
</tr>
</tbody>
</table>

Because many of the cloud observations used for the METSTAT model were sub-optimal (e.g. ASOS, filled, or derived), additional uncertainties were included for these conditions (Table 4). Likewise, additional uncertainty was added for the SUNY satellite model for areas of high albedo (snow cover or light ground) or high latitude—conditions that degrade the performance of the satellite model (Table 5).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Additional Uncertainty</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short- and medium-term data filling</td>
<td>4</td>
<td>Data derived from close time proximity</td>
</tr>
<tr>
<td>Long-term filling</td>
<td>14</td>
<td>Based on interannual variability; data comes from like dates in other years</td>
</tr>
<tr>
<td>Cloud probability derivation</td>
<td>4</td>
<td>Differences in total and opaque average less than 1 okta</td>
</tr>
<tr>
<td>Cloud probability from nearby site</td>
<td>4</td>
<td>Differences in total and opaque average less than 1 okta</td>
</tr>
<tr>
<td>ASOS-only</td>
<td>22</td>
<td>Limited okta resolution afforded by the coded ASOS-only data; a cloud measurement value could be in error by two oktas</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
<th>Additional Uncertainty</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-shifting</td>
<td>2</td>
<td>Shifting from satellite time to hourly local time causes some error in the hourly values or daily totals</td>
</tr>
</tbody>
</table>
Ground snow cover 5

High ground albedo compresses the dynamic range of the model. Periods of snow cover were estimated for 5° latitude and longitude cells on a monthly basis using snow cover probability contours for the United States (16). If the probability of snow for a given location was greater than 25%, the additional uncertainty was added to the data.

High latitude 10

Above 50° (arbitrarily set to affect only Alaska sites), cloud observations are compressed by angle of incidence and curvature of the earth as seen from the geostationary GOES satellite. (The model does not produce data for locations above 60° latitude.)

Final uncertainties (U) for each hourly value were determined by the root-sum square of all applicable uncertainties.

\[ U = (U_{opt}^2 + U_{add1}^2 + U_{add2}^2 \ldots)^{1/2} (\pm \%) \]

where \( U_{add} \) are additional sources of uncertainty from Tables 4 and 5.

In reality, some of these additional uncertainties may apply differently among the global, direct, and diffuse components (e.g., direct beam may be more influenced by uncertainties in cloud cover than global). However, in this simplified yet conservative approach, these additional uncertainties are applied uniformly to all three components. The final uncertainties are included in uncertainty fields in the NSRDB data record. This method of uncertainty calculation applies to all classes of sites, and as detailed in Section 2.7, the uncertainties are used in part to assign site classifications.

2.7 Station Classification

The 1961–1990 NSRDB categorized stations as primary and secondary based on whether a station's data set included measured solar data. Because of changes in the integration of measured solar data, the use of two models, and the varying quality of input cloud data, this method was replaced with a three-tiered classification based on data quality and completeness (13).

- Class I stations have a complete period of record (all hours 1991–2005) for solar and key meteorological fields and have the highest-quality solar modeled data (221 sites).
- Class II stations have a complete period of record but significant periods of interpolated, filled, or otherwise lower-quality input data for the solar models (637 sites).
- Class III stations have some gaps in the period of record but have at least 3 years of data that might be useful for some applications (596 sites).

Completeness in period of record for station classification is based on solar, dry-bulb, and dew-point temperatures; humidity; wind speed and direction; AOD; precipitable water; and station pressure. Other fields in Class I and II stations may not be serially complete. Stations of any class may contain measured solar radiation data. Figure 2 shows the distribution of sites by class and whether they contain measured solar data.

2.8 Data Distribution

The final decisions on data distribution reflected a compromise between completeness of data and the need for proprietary control over the ISD meteorological fields (managed by NCDC). From this need came four fundamental data sets:

- Solar and meteorological data for general solar resource assessment or technology deployment (This data set is further subdivided, as described below.)
- A solar research data set (without meteorological fields but with fields for output from both models, a clear-sky data set from the METSTAT model, aerosols, water vapor, and ozone)
- The 10-km gridded data set for high-resolution resource assessment (without meteorological fields)
- A summary data set of hourly, daily, and threshold statistics.

The NCDC pricing structure allows access to ISD data via the Internet without cost to certain Internet domains (i.e., .gov, .mil, .edu, and .k12). All other domains are fee-restricted, and users must pay for the data. These restrictions also apply to NSRDB data containing ISD fields. To afford the greatest flexibility for data access, several options with and without ISD meteorological data are shown below.

**NSRDB solar and filled meteorological fields.** Available from NCDC. ftp://ftp3.ncdc.noaa.gov/pub/data/NSRDB

No-cost access is domain-restricted to .mil, .gov, .edu, and .k12. A fee-access restriction applies to all other domains.
NSRDB solar and ISD meteorological fields (no data filling). Available from NCDC. At the time of this writing, this data set is planned for a late-2007 release at http://cdo.ncdc.noaa.gov and http://gis.ncdc.noaa.gov. These systems already provide access to the complete ISD archive, with the NSRDB solar parameters to be added. No-cost access is domain-restricted to .mil, .gov, .edu, and .k12. A fee-access restriction applies to all other domains.


The 1991–2005 NSRDB will not be distributed on physical media.

The NSRDB 1991–2005 updated user’s manual (13) is available via these Web sites, and it contains data formats, data quality summaries, and an overview of NSRDB development work.

3. FUTURE WORK

As with the original NSRDB, this updated database will form the basis for additional products and data sets. Future work includes:

Merging with 1961–1990 NSRDB. The 15-year period of record of the updated NSRDB is shorter than the conventional 30-year data sets used for climatological norms. Work is planned to merge the 239 stations in the 1961–1990 NSRDB with corresponding stations in the updated NSRDB.

Annual and decade updates. NREL plans to issue annual updates of the NSRDB as dictated by funding and demand. A two-decade update is possible in 2011 for the period 1991–2010 and may include 10-km gridded data for the entire period based on work in progress (17).

Typical meteorological year (TMY) data sets. NREL is working on new TMYs based on the updated NSRDB. Questions remain about whether the 15-year period of record and anomalous years following the Mount Pinatubo eruption will allow valid typical meteorological year data sets to be produced.

Fig. 2. Distribution of NSRDB update sites by class. 1961–1990 NSRDB sites are superimposed (star symbols).
4. REFERENCES