TOWARDS PRODUCTION OF AN UPDATED NATIONAL SOLAR RADIATION DATA BASE

Stephen Wilcox
Mary Anderberg, Ray George
William Marion, Daryl Myers
David Renné
National Renewable Energy Laboratory
1617 Cole Boulevard
Golden, CO 80401
Email: Stephen_wilcox@nrel.gov

William Beckman
University of Wisconsin

Arthur DeGaetano
Northeast Regional Climate Center

Christian Gueymard
Solar Consulting Services

Richard Perez
State University of New York at Albany

Neal Lott
National Climatic Data Center

Paul Stackhouse
National Aeronautics and Space Administration

Frank Vignola
University of Oregon

ABSTRACT

The National Renewable Energy Laboratory and collaborating agencies are working towards an update of the 30-year 1961-1990 National Solar Radiation Data Base (NSRDB). This update will include hourly solar radiation values (almost entirely modeled) for global, direct, and diffuse solar values, as well as accompanying meteorological data. The data set covers the 1991-2000 decade and incremental annual data sets from 2001-2005. An important enhancement relative to the original NSRDB will be an hourly 10-km x 10-km gridded data set for the United States (excluding northern Alaska above 60 deg) produced by the State University of New York at Albany. This satellite-derived data set will not span the entire period back to 1991, but rather will begin in 1998 when the satellite image archive was initiated. The updated NSRDB will be transferred to the National Climatic Data Center for distribution.

1. INTRODUCTION

In 1992, the National Renewable Energy Laboratory (NREL) released the 1961-1990 National Solar Radiation Data Base (NSRDB), a 30-year, 239-station data set of measured and modeled solar radiation and accompanying meteorological data for the United States (1). In 2003, NREL investigated the feasibility of producing a 1991-2000 update of the NSRDB and began collaborative work with several other agencies, including the National Aeronautics and Space Administration (NASA), the National Climatic Data Center (NCDC), the Northeast Regional Climate Center (NRCC), the State University of New York at Albany (SUNYA), the University of Oregon, the University of Wisconsin, and the private firm Solar Consulting Services. The final updated 1991-2005 NSRDB, to include a decade data set for 1991-2000 and five annual updates from 2001-2005, is planned for a midyear 2006 release. Additionally, we will include a 10-km x 10-km gridded data set produced by collaborators at SUNYA for 1998-2005.

The original NSRDB held a serially complete data set for all sun-up hours for 239 stations. Because of expected changes in the roster of National Weather Service sites, as well as the potential for adding additional NSRDB sites, we chose not restrict the list of updated stations to those same 239. Instead, we included as many stations and as much data as possible to increase the usefulness of the data set.

Not all applications require serially complete data, and thus we devised a plan to divide the data set into two subsets, or classes, based on the availability of output data. Class I stations include serially complete hourly solar and select meteorological data for the target decade of 1991-2000. Class II stations fall short of that standard, but have a data density useful in some applications for characterizing the

solar resource. The Class II stations hold at least 75% of possible data for at least three years.
We chose 1504 candidate sites for a preliminary evaluation data set. After processing, an analysis of the output data set demonstrated that about 800 sites qualify as Class I, about 650 as Class II, and about 50 produced insufficient usable data. Fig. 1 shows the distribution of these sites relative to the 1961-1990 NSRDB stations. The 800 Class I sites represent a more than three-fold increase in sites over the 239 sites in the original NSRDB. (These numbers may change in the final version of NSRDB.)

The update data set also includes several meteorological parameters necessary for some solar applications.

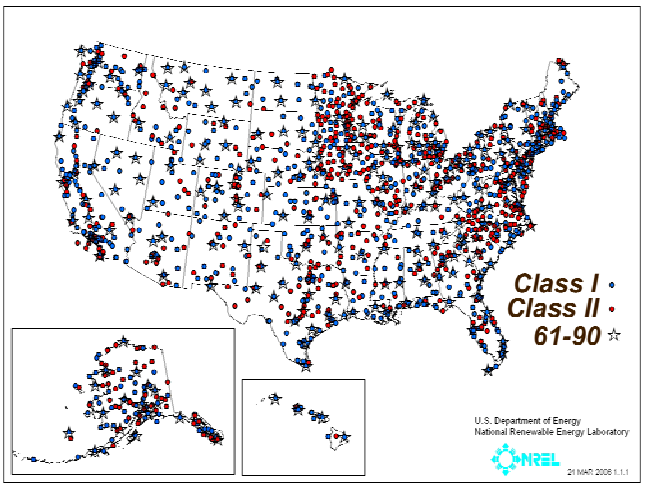

Fig. 1: Locations of NSRDB update sites relative to 1961-1990 NSRDB sites

2. PROJECT STATUS

Work over the past several months toward completion of the update includes:

- Selection of the NSRDB model
- Completion of ten-km gridded input data sets for aerosols, water vapor, and ozone
- Development of a statistical technique for estimating total cloud cover from opaque cloud measurements
- Development of data filling techniques to achieve a complete period of record where possible
- Development of a technique to shift the data from the native time realm of the SUNYA satellite model to the NSRDB hour-ending convention
- Retrieval of measured solar radiation data from various on-line archives or data distribution sites
- Production of a pre-release version of the database for examination by volunteer reviewers.

The sections below discuss each of these items.

2.1 Selection of the NSRDB Model

In earlier work reported by several of the authors (2), three models were evaluated for use in modeling NSRDB data: The NREL Meteorological Statistical (METSTAT) model (3), a model developed by the NRCC for the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) (4), and the SUNYA-developed satellite model (5). The study concluded that each model performed comparably in comparisons with measured ground data, and the decision for which model to use rested less with technical considerations and more with those of convenience.

In that context, we chose the METSTAT model, which is already deployed at NREL and one with which we have considerable experience. Also, because of its use with the original NSRDB, METSTAT may provide greater consistency with the older data set. The SUNYA satellite model will also be used to model 10-km gridded solar radiation data as the first step toward providing users with high-resolution data suitable for Geographic Information System (GIS) applications or on-demand data for any location in the U.S.

The high-resolution gridded data set will be limited to the years 1998-2005, the period for which necessary GOES imagery was archived by the project. Acquiring older imagery is beyond the financial resources of the project, and other input data sets for the SUNYA model are not available prior to 1998.

2.2 Gridded Aerosol, Water Vapor, and Ozone Data

Both the METSTAT and SUNYA models require input aerosols, water vapor, and ozone. The 1961-1990 NSRDB used only DNI measurements to estimate broadband aerosol optical depth (BAOD). Today there is much less measured DNI data available, but much more aerosol data from other sources, including sun photometers and satellite-based estimates. These sources produce data that are spectral rather than broadband. A combination of surface sun photometry, satellite data from the NASA MISR and MODIS satellite-based instruments, and legacy DNI estimates of BAOD allowed creation of monthly mean estimates of BAOD for all locations in the U.S. Spectral AOD data were converted into BAOD using estimated Angstrom Alpha parameters from sun photometry (6). All BAOD data were merged, interpolated and adjusted to local elevation through use of an exponential function (6). These

monthly values were finally smoothly fit to daily values using a mean-preserving interpolation method (7).

For water vapor, NREL used the NASA Water Vapor Project (NVAP) dataset for daily estimates of water vapor on a 1º x 1° grid. NVAP integrates sounding data with satellite measurements of atmospheric water vapor. For the years 1988-1999, NVAP provides once-daily estimates of water vapor on a 1º x 1º worldwide grid. For years 2000 and 2001, NVAP provides the data on a 0.5º x 0.5º grid twice daily. These data were interpolated in space to the location of each of the 1504 NSRDB stations and interpolated in time to provide hourly data for all stations for the years 1988 thru 2001.

The years 1986, 1987 (for the evaluation NSRDB), and 2002-2005 required a new source of data. NREL chose the recently released North American Regional Reanalysis (NARR). This data source provides column water vapor every 3 hours on a nominal 32 km grid. This data product uses a data assimilation model derived from National Center for Environmental Prediction (NCEP) weather prediction models to interpolate weather observations onto a numerical grid. The data observations for precipitable water include radiosondes, dropsondes from airplanes, and the infrared radiance data from the TIROS Operational Vertical Sounder-1B satellite (over water) (8).

Using GIS technology, NREL matched each of the 1504 candidate NSRDB stations with the nearest NARR grid point. Then data from the NARR grids for each 3-hour period was downloaded and interpolated to hourly data files for each station.

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

2.3 Input Cloud Data

Development of the METSTAT model relied on human observations of both total and opaque cloud cover, which were previously a routine part of National Weather Service (NWS) station operations. In the mid 1990s, the NWS embarked on a massive network upgrade to automated observations, which provides only estimates of total cloud cover through use of the Automated Surface Observing System (ASOS). In a previous report on NSRDB work (9), the authors described a method to model both total and opaque cloud cover using ASOS data in conjunction with the NCDC ASOS Supplemental Cloud Product, which provides cloud estimates from satellite imagery.

In subsequent NSRDB tasks, we discovered the ASOS Supplemental Cloud Product was not available for several hundred of the candidate NSRDB sites, including all sites in Alaska and most in Hawaii. This realization placed significant restrictions on the geographical coverage and number of sites in the new NSRDB. To address this situation, we developed a statistical method of estimating cloud cover by computing probability distributions of total cloud cover (oktas) relative to opaque for every site where both total and opaque measurements were available either from the Supplemental Cloud method or human observations. These site-specific monthly distributions derive from the relative number of occurrences of each of the nine total cloud amounts for each of the nine opaque cloud amounts (in oktas as an integer range from 0 to 8).

For this process we considered the ASOS total cloud report as primarily a measurement of opaque clouds, because the ASOS cloud measurement is limited to 3660 m and most translucent clouds occur above that altitude. At those times when only ASOS data are available, a total cloud amount for a particular hour is derived by randomly selecting a total cloud cover value based on the probability distribution derived from the ensemble of reported ASOS values. For sites that have neither the ASOS Supplemental Cloud Product nor manual cloud observations necessary for creating the probability distributions, the process uses the distribution table from the geographically nearest site.

While the effectiveness of this technique may be diminished by inter-annual variability of distributions, we found the difference between total and opaque clouds is typically only one okta or less. Thus, the derivation of total cloud values and possible errors are secondary to the dominant ASOS measurement.

2.4 Data Filling

The NSRDB update specifications for Class I stations include the goal of providing serially complete data for solar radiation and several meteorological parameters important for solar applications (unlike the original NSRDB, which excluded filling of nighttime data). However, any gaps in the model input data prevent model output for those times. Filling *input* metrological data for the model is preferred over filling *output* solar data because of the statistical nature of METSTAT, which imposes natural variability on the output data (while still conforming to long-term means).

Most meteorological data were extracted from the Integrated Surface Hourly (ISH) data set (10). The following parameters were filled in the NSRDB production database: total and opaque cloud cover, dry bulb and dew point temperatures, relative humidity, ceiling height, barometric pressure, wind speed and direction, aerosol optical depth,

precipitable water, and ozone. (Relative humidity was not present in the ISH data set and was calculated using psychrometric relations of dry bulb and dew point temperatures after the completion of all data filling.)

A primary goal of the NSRDB is to provide complete hourly data for ten years (1991-2000) or longer so as to form the basis for decade normals and means. Although short gaps of four or five hours can easily be filled with interpolation, over a ten year period there may exist circumstances that create larger gaps, sometimes of several months. Even one gap violates the goal of serially completeness; therefore, by developing methods to fill both large and small gaps, we greatly increased the number of sites included in the NSRDB.

Data filling methods were categorized in four levels:

1) *Short-term interpolation*. Data gaps of up to five hours were filled with either linear interpolation or, in the case of temperatures, with a site- and month-specific diurnal profile imposed on a linear interpolation. Gaps that extended through the night were also filled with this method.
2) *Medium-term filing*. Gaps of up to 24 hours were filled by building an average profile from the same hours of previous and subsequent days and attaching that profile to the end points of the gap.
3) *Long-term filling*. Gaps of up to one year were filled by characterizing the data before and after the gap and then seeking the most similar data from other years to fill the gap. The first and last 12 hours of the fill data were scaled to fit the end points of the gap, and the rest of the fill data were used without modification. When filling dependent parameters (dry-bulb and dew point temperature, and total and opaque cloud and ceiling height), fill data were pulled from the same source dates when possible. If this could not be accomplished, coupling limits were enforced between the parameters.
4) *Last-ditch filling*. In a few cases, the above methods left gaps because of insufficient data before and after the gap for developing a diurnal profile or characterization. In these cases, data from the same days were examined from other years in random order to find an uninterrupted run of source data to fill the gap. Additionally, if barometric pressure could not otherwise be filled, it was modeled from site elevation. If ceiling height could not be filled, it was modeled from a simple regression with opaque cloud cover.

The data filling process creates log files for methods three and four to document the source year of the fill data. We plan to consider the type and density of data filling when deriving the uncertainty of the modeled data.

2.5 Time-shifting Satellite Model Data

The SUNYA satellite model produces solar estimates based primarily on input from Geostationary Operational Environmental Satellite (GOES) images. These images originate from two satellites – GOES East and GOES West – with the dividing line between the two between 105º-110º west longitude. The GOES East images occur at 15 minutes after the hour and GOES West images at the top of the hour (referred hereafter as :15 or :00 images respectively). These images and resulting solar estimates effectively represent conditions at the moment of image acquisition. However, the convention for NSRDB solar fields is that an hour-ending time stamp represents an integrated hourly value. Although there is no way to transform the instantaneous value to an hourly integrated value, there is some temporal as well as spatial integration inherent in a 100-km^2 pixel as all clouds passing through that pixel are integrated optically at the instant of image acquisition or during subsequent processing. In this regard, we consider the satellite image an hourly-integrated value *centered* on the time stamp. By simple convention, this same value is equivalent to an *hour-ending* integrated value by simply moving the time stamp to the end of the integration period. For example, a 12:00 hour-centered value is the same as a 12:30 hour-ending value, or a 12:15 hour-centered value is the same as a 12:45 hour-ending value.

The primary challenge focuses on the shift in time from the hour-centered :00 or :15 value to the NSRDB convention that always ends exactly on the hour. After achieving the necessary temporal shift, the shift in time-stamp convention is trivial. Accomplishing the temporal shift requires realistic estimates of interim-hour values that take into account not only the measured values, but also known geometric and atmospheric characteristics of solar radiation.

These characteristics can be described under clear sky conditions as a function of the solar zenith angle and the path length through the atmosphere. These parameters are used to normalize measured values. Then interim values at the desired time can be selected and converted back to solar values that follow expected profiles. Simple interpolation does not suffice as it both fails to mimic true measurements and reduces the true variability in measured data.

One drawback of this shifting method applied instantaneously appears as a failure to maintain daily total values, an important measure of the solar resource. In examining annual shifted data sets, we have found any aggregate errors introduced by the method to be typically less than one-half percent. Further study is required to investigate the possibility of a seasonal or geographical bias.

2.6 Measured Solar Radiation

Thirty-three solar measurement sites with nearby meteorological stations were identified and data acquired from various solar networks, including NOAA SURFRAD and ISIS, University of Oregon, University of Texas, the NREL HBCU network, and others. These data were integrated to hourly values, subjected to quality assessment procedures, and added to the production database.

Twelve sites employ rotating shadowband radiometers (RSRs) as a low-cost method of collecting data. These instruments have known deficiencies related to a non-linear spectral sensitivity of the sensor. Part of the work effort toward release of the NSRDB involves refining and applying a correction algorithm at the University of Oregon Solar Monitoring Laboratory for RSR data in the measured solar data fields.

2.7 Pre-release Evaluation Data Set

To form a preliminary NSRDB for expert evaluation, we assembled an input data set for years 1986-2005 and ran all processes to populate all solar and meteorological fields of the production database. The years 1986-1990 were included in this evaluation data set to provide overlap years to compare with the original 1961-1990 NSRDB. These years will not be included in the 1991-2005 update.

To the output data set we added measured solar radiation for 33 sites, the available data from the satellite model (by extracting values from the pixels that represent each station), and the 1986-1990 solar fields from the original NSRDB. Additionally, we created several summary data sets for analysis.

3. ANALYSIS OF NSRDB METHODS

A group of volunteer reviewers was solicited from experts in several fields who had experience with the original NSRDB. This group was tasked with looking at the data and methods for both the quality (correctness) of the data and its usability. Using feedback from this evaluation, we will either modify the process to correct shortcomings (funding permitting) or document the limitations and their implications to users.

As a preliminary analysis, we examined the test case of Miami, Florida for 1988 to evaluate differences between the old and new versions of the NSRDB. We computed the annual and monthly mean global, direct and diffuse irradiances from both versions. Fig. 2 shows the annual means of global, direct and diffuse irradiance from the two versions differed by 10, 9, and 5 W/m^2 respectively. The monthly means differed by at most 25 W/m^2 (6%) in March. For eight of the twelve months, the monthly mean differences were less than 10 Wm/^{2} (less than 5% of the mean value). Since the uncertainty in the original NSRDB is estimated to be at least ±9%, the two independent processes are performing comparably. There is an apparent tendency for the new version to consistently produce a lower monthly and annual mean (positive differences in means). This may be due to the differing sources of and application of cloud, aerosol, and water vapor input data in the two versions.

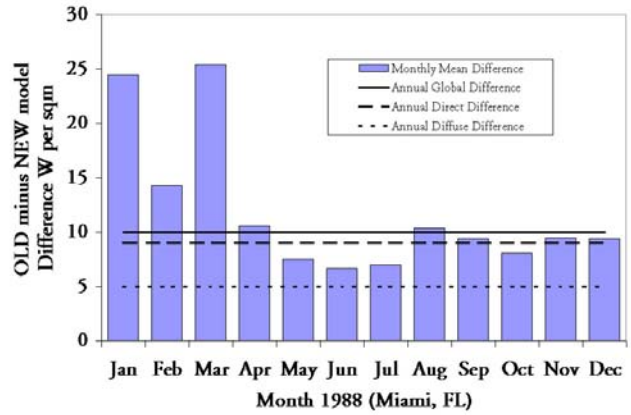

Fig. 2: Monthly (bars) and annual (lines) mean differences in old and new NSRDB versions for Miami, FL.

4. CURRENT AND FUTURE WORK

Several issues remain to prepare for final release of the NSRDB and to look ahead to future enhancements.

4.1 Data Distribution

As with the original NSRDB, the National Climatic Data Center (NCDC) will manage and distribute the updated NSRDB. NCDC distributed the 1961-1990 NSRDB as the *Solar and Meteorological Surface Observation Network* (SAMSON) data set. Ongoing work focuses on identifying the best method of distribution and determining how to integrate the much larger gridded data sets in the NSRDB distribution package. Monthly or annual summary maps and underlying data may also be part of data sets available.

4.2 Future Updates

We expect to release the updated NSRDB in midyear 2006. In addition to the target 1991-2000 decade update, the release will include five incremental annual updates from 2001-2005. Funding permitting, we plan annual NSRDB incremental updates on a yearly basis as input data become available. In 2011, a 2001-2010 decade update is planned.

4.3 A Composite 1971-2000 NSRDB

Most of the 239 sites populating the original 1961-1990 NSRDB appear in the roster of Class I sites in the preliminary updated NSRDB. Justifications exist for producing an updated 30-year 1971-2000 database. Among them are updating 30-year norms, creating a Typical Meteorological Year (TMY), and providing updated input data for applications using the original NSRDB. Work is planned to evaluate the effects of changes in methods and model inputs between the old and new databases to establish the comparability of the old and new data set. In addition, some continuity issues may arise about the smaller number of stations in the proposed merged NSRDB versus the greater number available for the 1991-2000 decade.

4.4 Extending the 10-km Data Grid Backwards in Time

Work continues this year between SUNYA, NASA, and NREL to complete a method for creating 10-km data grids for years prior to 1998 (11). This effort would enhance applications requiring longer-term solar resource assessment information for GIS or other analyses.

4.5 TMY, Data Manuals, and Other Products

After release of the original NSRDB, NREL developed several products based on the 30-year data. These include the TMY2, data manuals, maps, and GIS resource coverages. Each of these products may be updated as demands are identified.

4.6 The Effect of Measured Aerosol Model Inputs

The earlier NSRDB model evaluation (2) revealed the three candidate models performed similarly in comparison with measured ground data. One unexpected revelation of the study was that the variability of the models as a group among sites was much greater than the variability among the models themselves. This led to the possibility that aerosol input data, which are based on long-term monthly means, may inadequately represent actual conditions because of their typically log-normal (rather than normal) distribution. Work is underway to evaluate the effect of measured versus climatological or statistical aerosol data on model performance. This effort may lead to future solar resource estimates with lower uncertainty.

5. CONCLUSION

We have overcome significant obstacles in producing an updated NSRDB, including those based in fundamental changes in the collection of necessary model input data. Additionally, the inclusion of high-resolution gridded data greatly enhances the NSRDB and provides an invaluable basis for state of the art analysis of spatial and temporal characteristics for many resource assessment applications.

6. REFERENCES

(1) Maxwell, E. L., Final Technical Report, National Solar Radiation Data Base (1961-1990), NREL/TP-463-5784 DE9500262, National Renewable Energy Laboratory, Golden, Colorado, USA, 1995

(2) Myers, D., S. Wilcox, W. Marion, R. George, M. Anderberg, Broadband Model Performance for an Updated National Solar Radiation Data Base in the United States of America, Proc. Solar World Congress, International Solar Energy Society, 2005

(3) Maxwell, E. L., METSTAT—The Solar Radiation Model Used in the Production of the National Solar Radiation Data Base (NSRDB), *Solar Energy* 62(4), 263-279, 1998

(4) Belcher, B.N. and A.T. DeGaetano, Integration of ASOS Weather Data into Building Energy Calculations with Emphasis on Model Derived Solar Radiation, Ithaca, NY, Northeast Regional Climate Center, ASHRAE 1226-RP, 2004

(5) Perez, R., P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George and F. Vignola, A New Operational Satellite-to-Irradiance Model, *Solar Energy* 73, 5, 307-317, 2002

(6) Gueymard, C. and R. George, Gridded Aerosol Optical Depth Climatological Datasets over Continents for Solar Radiation Modeling, Proc. Solar World Congress, International Solar Energy Society, 20057

(7) Rymes, M. and D. Myers, Mean Preserving Algorithm for Smoothly Interpolating Averaged Data, *Solar Energy* 71, 4, 225-231, 2001

(8) Shafran, Perry, J. Woollen, W. Ebisuzaki, W. Shi, Y. Fan, R. Grumbine, and M. Fennessy, Observational Data Used for Assimilation in the NCEP North American Regional Reanalysis, 14th Conf. on Applied Climatology, 84th AMS Annual Meeting, Seattle, WA, 2004

(9) Wilcox, S., M. Anderberg, W. Beckman, A. DeGaetano, R. George, C. Gueymard, W. Marion, D. Myers, R. Perez, M. Plantico, D. Renne, P. Stackhouse, F. Vignola, Progress on an Updated National Solar Radiation Data Base for the United States, Proc. 2005 Solar World Congress, Orlando, FL, International Solar Energy Society, 2005

(10) Lott, N., R. Baldwin, and P. Jones, 2001: The FCC integrated surface hourly database, a new resource of global climate data. (National Climatic Data Center technical report 2001-01). Asheville, N.C.: National Climatic Data Center, 42 p.

(11) Perez, R., M. Kmiecik, S. Wilcox, P. Stackhouse, Deriving Long Term High Resolution Solar Irradiances from Low Resolution Archives via Microstructure Patterning (in press)