Solar 2011 Conf., Raleigh, NC, American Solar Energy Soc., May 2011

A NEW WEB-BASED DATA DELIVERY SYSTEM TO PROVIDE GLOBAL SUPPORT FOR SOLAR SITE SELECTION ANALYSES

John L Keller and Charles A Khuen
Weather Analytics Inc.
6 Lantern Lane
Winchester, MA 01890
John.Keller@wxaglobal.com
Chuck.Khuen@wxaglobal.com

Christian A Gueymard
Solar Consulting Services
P.O. Box 392
Colebrook, NH 03576
Chris@SolarConsultingServices.com

ABSTRACT

We present a new online data access and delivery system to support solar site selection and analysis worldwide at locations where data of this depth and precision was not previously available. Our methodology utilized the newly-released NCEP/Climate Forecast System Reanalysis (CFSR) high quality dataset, which contains hour-by-hour data at about 35-km horizontal resolution across the globe from January 1979 through December 2009 for a total of 35 weather quantities, including both conventional and solar radiation variables.

The CFSR data was pre-processed in a series of steps and the Perez algorithm was employed to generate Direct Normal Irradiance (DNI) from the Global Horizontal Irradiance (GHI) values in the output. A multi-function interactive web-based interface was added to simplify user command and control of site definition and easy data acquisition. Besides gap-free, continuous hourly time-series data for 31 years with global coverage, the interface offers the user options in the special formats (epw, bin, tm2) used as input to popular solar and building system models such as TRNSYS, EnergyPlus, and eQuest/DOE2. The quality of the final output for both conventional and solar radiation variables is verified by comparing the output at various weather stations including sites with multiple GHI and DNI sensors. Historical predictions of 35 weather variables are complemented by their 7-day ahead forecast, based on NOAA's Global Forecast System.

1. INTRODUCTION

Accurately categorizing the climate at any location, essential for reliable simulation and analysis of energy systems, requires at least 30 years of statistically stationary hourly observation history. However, only a limited number of weather sensor stations have been operating for that period – less than 1000 in the world with over 80% complete data and only a small fraction of those include solar radiation. Even high-quality airport surface stations are relocated periodically. This can result in apparent trends and other statistical artifacts in the station climatology.

Early in 2010, after more than five years of research and development, NOAA's National Centers for Environmental Prediction (NCEP) released numerical output from its Climate Forecast System Reanalysis (CFSR) model. This effort yielded a high quality 31-year global reanalysis of the ocean and atmosphere covering the whole world and the period 1979–2009 [1]. Included in this release is an hour-by-hour dataset consisting of over four dozen atmospheric and environmental variables at about 35-km horizontal resolution.

The CFSR dataset is available on the web directly from NOAA. However, converting their GRIB-format files into a useable common format can be a challenging task. This has been successfully undertaken by Weather Analytics (WxA). In the process, a series of interface tools has been added to facilitate access and use.

There is a growing demand for weather-related data in various fields, and particularly in energy applications. Whereas temperature is measured routinely at thousands of weather stations in the world, in situ data for surface solar radiation is not common, especially in developing countries, hence the impetus to develop resources with a global reach at reasonable cost. The CFSR dataset offers a unique potential, since it can be combined with local meteorological measurements for even greater accuracy.

This contribution describes the efforts undertaken at WxA to transform native CFSR data into usable formats with an easily-accessible user-friendly interface. This represents a resource that should appeal to a large community of energy analysts involved in solar, wind and building energy simulations.

2. METHODOLOGY

2.1 Brief Overview of the CFSR Model System

NCEP's CFSR system is a special application of numerical weather prediction (NWP) modeling technology. Modern real-time NWP modeling operations are supported by the sophisticated Global Data Assimilation System (GDAS). GDAS orchestrates the collection, quality control and preprocessing of raw in situ and remote sensor data from a wide array of sensor systems (Fig. 1). Once preprocessed, the resulting state of the atmosphere variables are used as input to a variety of regional- and global-scale NWP models.

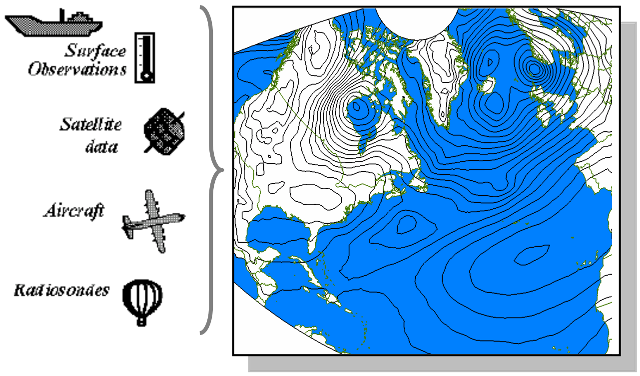


Fig. 1: The Global Data Assimilation System (GDAS) provides the raw sensor data that gives a constantly updated picture of the Earth-Atmosphere system.

NWP models are based on the fundamental physical laws governing the evolution of the earth-atmosphere system. Traditionally, NWP models have been implemented as initial-value applications to provide forecast guidance products for weather forecasters. The initial condition analysis created from GDAS is extrapolated into the future with the future evolution of the earth-atmosphere system constrained by physics.

NWP models can also be run using archived historical GDAS data. The historical GDAS observations are continuously assimilated into the model as the model reaches the times corresponding to the observations. In this case, the *past* evolution of the earth-atmosphere system is constrained both by physics and observations. The CFSR model system was run using this method starting at the end of 1978 through the end of 2009. Figure 2 shows the NWP model moving through time and the various sensor systems (many satellite based) entering the model as they are commissioned.

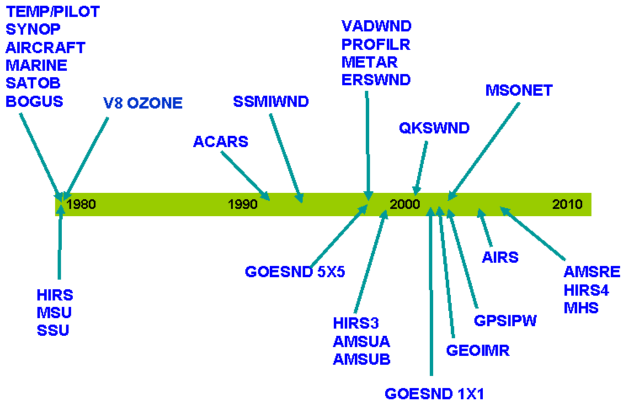


Fig. 2: When observation data types enter the CFSR stream.

2.2 Multi-tiered CFSR Data Processing and Integration

Over 30 terabytes of native CFSR data was downloaded and transferred in its GRIB format. Next the information was peeled from its month by month global structure to one which would support continuous hour-by-hour time-series of weather variables for selected locations. Then each weather variable was converted into its most useful form and the final completed dataset was assembled into a direct-access structure to facilitate rapid response to requests for any given location. This process is schematically visualized in Fig. 3.

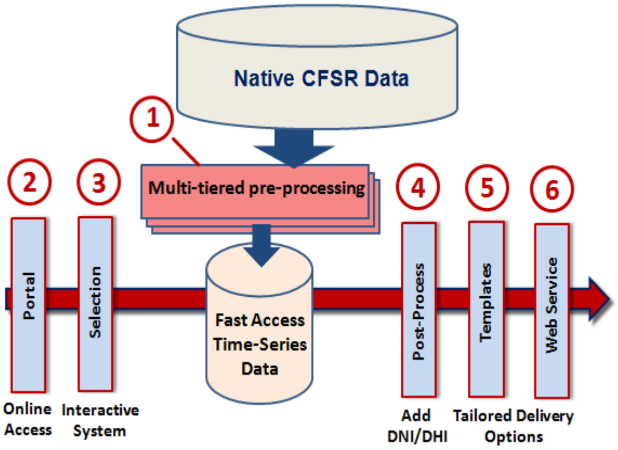


Fig. 3: Schematic of database methodology steps

This dataset was then integrated and geographically aligned with the output from WxA's Sensor Point™ system into a single unified structure consisting of hour-by-hour weather data from the beginning of 1979 through current conditions, as well as a 7-day forecast. This system allows the placement of virtual weather stations – called *Virtual Sensor Stations™ (VSS™)* at any location of interest.

2.3 Solar Irradiance

Among the variables in this CFSR hour-by-hour dataset is Global Horizontal Irradiance (GHI) at the Earth's surface. One particularly important factor in a numerical weather prediction (NWP) model's ability to realistically handle radiative transfer is dependent on its vertical resolution. The CFSR model has 64 vertical levels (Fig. 4), a large increase over the 28 levels used in the previous version of the global reanalysis dataset. By design, this yields a significant improvement in modeled GHI over previous models that were limited by their lower vertical resolution.

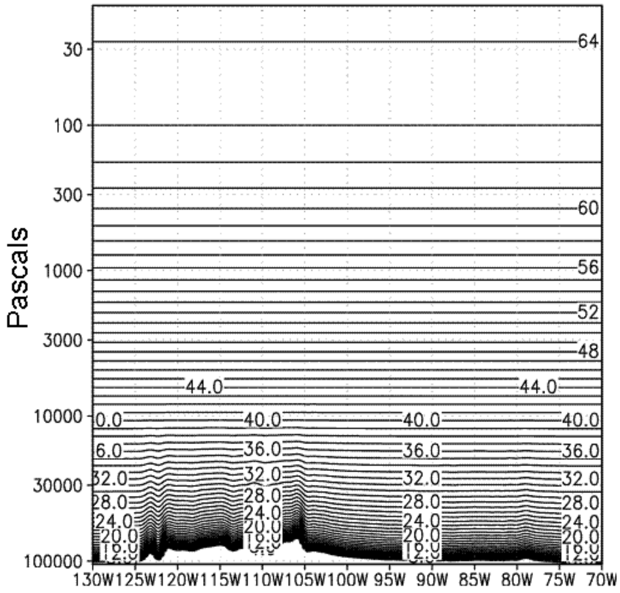


Fig.4: Vertical model levels used in the CFSR system.

3. VERIFICATION

3.1 Primary Variables

To verify the accuracy of the final output for historical data, we positioned a *Virtual Sensor Station™* (*VSS™*) near two co-located weather stations. Comparisons of *VSS™* virtual observations with actual on-site measurements for selected conventional weather variables along with the CFSR-based GHI are shown below.

Surface Temperature: Figure 5 shows the measured observations for two separate weather stations co-located at the airport in Burns, OR (blue and green) and the CFSR-based virtual sensor time series (in red). It can be seen that the virtual sensor time series mimics the station observations during this 10-day period in July 2008.

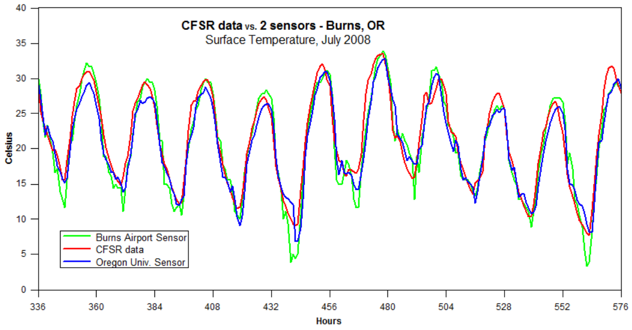


Fig. 5: Virtual sensor temperature output and two station observations at Burns, OR.

Correlations covering the most recent two years of the CFSR dataset are shown in Fig. 6. As is the case for the sample time series, the agreement between the virtual sensor temperature and one of the station sensors is closer than between the two stations, which are approximately 10-km apart and within 14 m elevation.

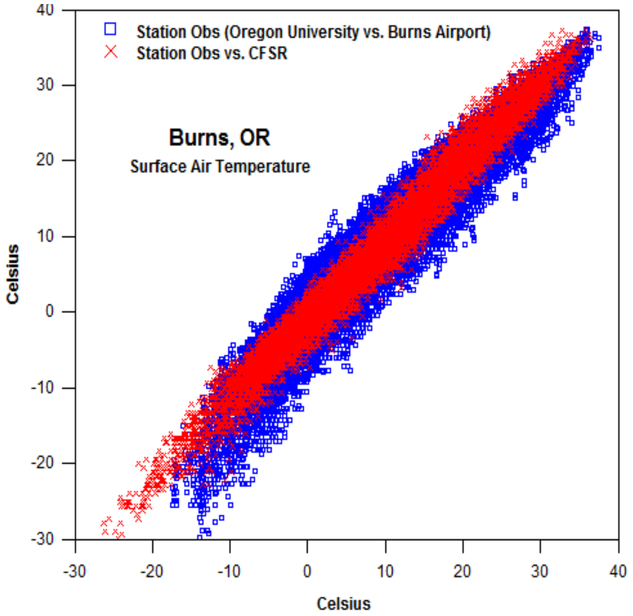


Fig. 6: Comparison of two Burns, OR sensor stations with each other and with 35-km area CFSR output for surface temperature (2008–2009)

<u>Global Horizontal Irradiance (GHI)</u>: Shown below are two representative periods for Eugene, OR (part of University of Oregon's monitoring network in the northwest U.S.) containing several solar sensors during 2008. One period is clear with low humidity (Fig. 7) and the other cloudy with high humidity (Fig. 8). The *VSS*™ output (in red) tracks the measured GHI closely. Statistically, the difference in the total GHI for 2008 between the *VSS*™ and the average of the sensor measurements is 1.6%. Over the period 1998–2005, this mean difference is 4.8% at Eugene, and 2.6% at Burns. This is comparable to differences of 1.3% and -4.1%, respectively, for the same period by NREL's NSRDB data, derived from the SUNY satellite-based model. For the same period, Fig. 9 shows the frequency distribution of the measured hourly GHI at Burns compared to those predicted by the WxA and SUNY models. The two modeled distributions show similar performance, despite the lower spatial resolution of WxA's.

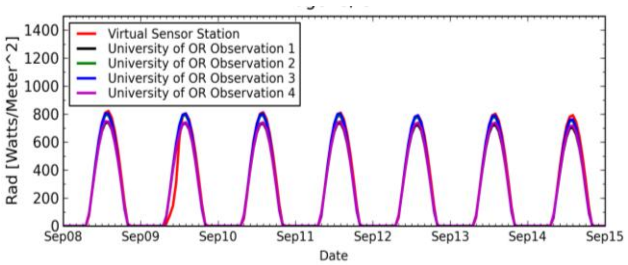


Fig. 7: Virtual sensor output and four station observations at Eugene, OR – GHI in clear, low humidity period

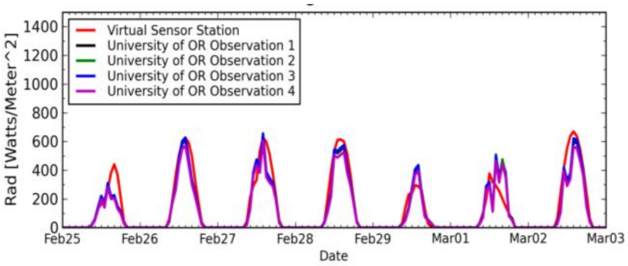


Fig. 8: Virtual sensor output and four station observations at Eugene, OR – GHI in cloudy, high humidity period

Monthly average irradiance values are of the utmost importance for the preliminary design of solar systems, among other applications. For Tamanrasset, southern Algeria, where the solar resource is very high, the average WxA predictions of GHI are compared in Fig. 10 to actual high-quality (BSRN) measurements of the same for 113 complete months during the period 2000–2009. With only one exception, all monthly predictions are within ±10% of the reference measurements, with no significant long-term bias.

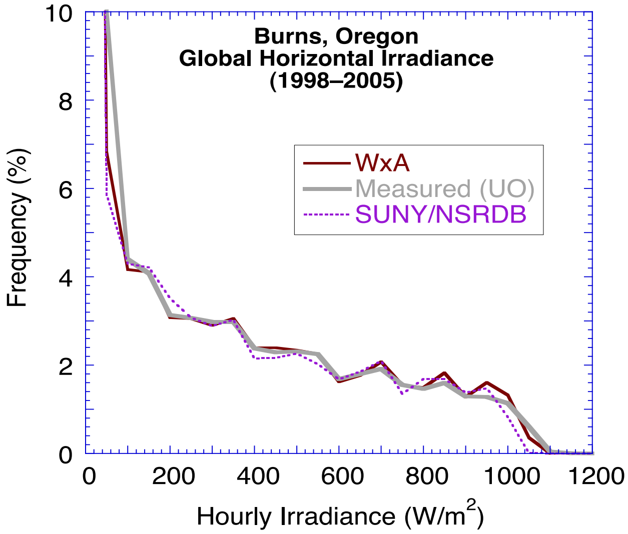


Fig. 9: Frequency distributions of hourly GHI at Burns.

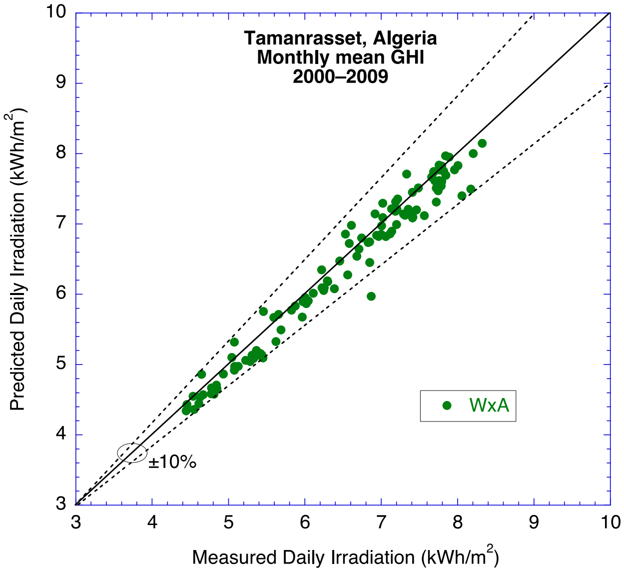


Fig. 10: Predicted vs. measured monthly mean daily GHI at Tamanrasset during the period 2000–2009

3.2 <u>Direct Normal Irradiance (DNI) Validation</u>

Shown below are the DNI comparisons for the same two periods in 2008 as the GHI graphs in Figs. 7 and 8, one clear with low humidity (Fig. 11) and the other cloudy with high humidity (Fig. 12). Note the *VSS*™ output (in red) tracks very well the measured DNI for multiple sensors, particularly on clear days. Statistically, the difference in the total DNI for 2008 between the *VSS*™ and the average of the sensor measurements is 15.8%. For months that were predominately clear this difference is 2.5%.

Over the period 1998–2005, the mean difference between the predicted and measured DNI is 7.9% at Eugene and 0.2% at Burns, as compared to 10.7% and -4.1%, respectively, for the SUNY model's predictions used in the NSRDB. It is stressed that the WxA predictions reported here are not affected by the "Eugene syndrome" that overestimates the winter NSRDB data at sites with significant cloudiness and snow cover [2].

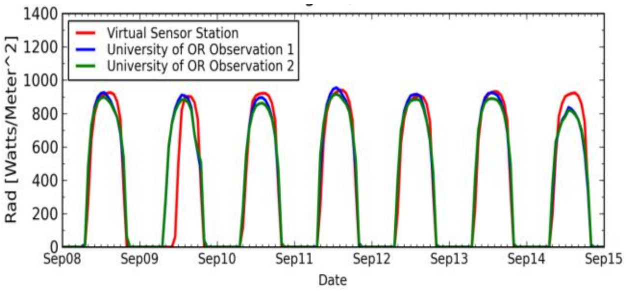


Fig. 11: Virtual sensor output and two station observations at Eugene, OR – DNI in clear, low humidity period

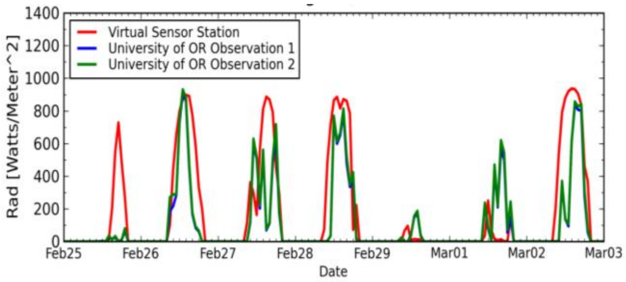


Fig. 12: Virtual sensor output and two station observations at Eugene, OR – DNI in clear, low humidity period

Figure 13 shows the inter-annual variations of DNI at Burns over the 30-year period 1980–2009, compared to high-quality measured data from the University of Oregon network. During that period, the measured mean daily DNI varied between 4.3 and 6.0 kWh/m^2, for a 30-year mean of 5.4 kWh/m^2. This long-term mean, and ±5% deviations from it, are indicated by the horizontal line and shaded area, respectively. Natural fluctuations in cloudiness, which strongly affect DNI, are correctly captured by the reanalysis scheme and subsequent derivation of GHI and DNI. Exceptions (marked by arrows) do occur due to the presence of large amounts of volcanic aerosols, which was the case during the El Chichon and Pinatubo years. Although the predicted GHI is still relatively accurate, the unusually high aerosol load makes the Perez algorithm overestimate DNI. In contrast, the observed overestimation during 2007–2009 might be due to a change in CFSR observing satellite sensors, but this issue is still under investigation.

3.3 TMY Data Validation

To extend the value of this output for solar radiation analyses Weather Analytics calculated DNI from GHI using an appropriate adaptation of the Perez separation algorithm [3]. This methodology also produces Diffuse Horizontal Irradiance, which further enables the construction of Typical Meteorological Year (TMY), as well as single Actual Meteorological Year (AMY) EnergyPlus weather (epw) and .bin format files to support solar simulation in multiple engineering models in common use, such as EnergyPlus, eQuest or TRNSYS.

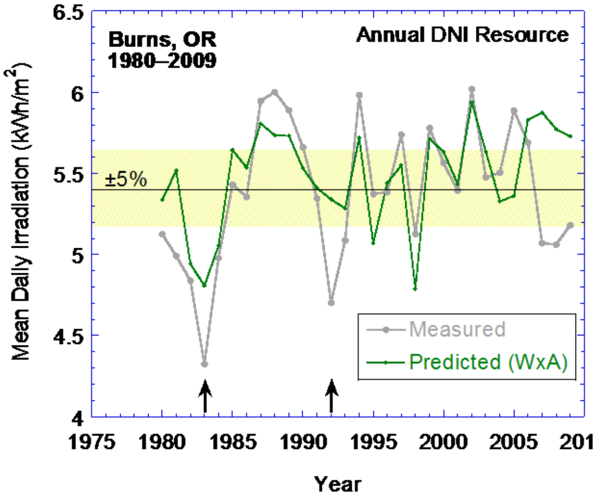


Fig. 13: Predicted vs. modeled inter-annual variability of DNI at Burns.

WxA's TMY files are constructed from basic 30-year data, following the NREL methodology [4]. Although this statistical methodology is the same as followed by NREL to develop its TMY2 and TMY3 files, the underlying radiative models and basic data periods are different. These are 1961–1990 for TMY2, 1980–2009 for WxA, and (normally) 1976–2005 for TMY3. In the case of Burns, Fig. 14 shows the frequency distribution of GHI over the 8760-hour period in these various TMY products to that from actual measurements over the period 1979–2009. The WxA distribution is remarkably close to those of TMY3 and of the measured data.

4. ACCESS, DISPLAY & DELIVERY LAYER

4.1 Users Access Portal

To further facilitate the usefulness of the data an online user access portal with an interactive interface has been created. A user can specify the services, data period (Fig. 15), and weather variables required (Figs. 16, 17).

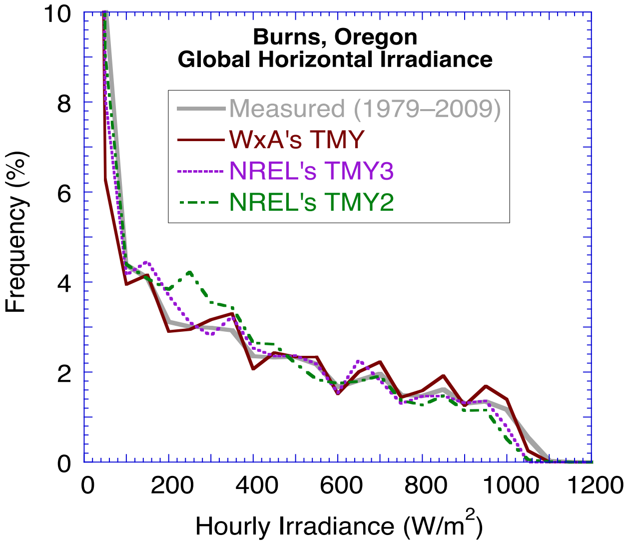
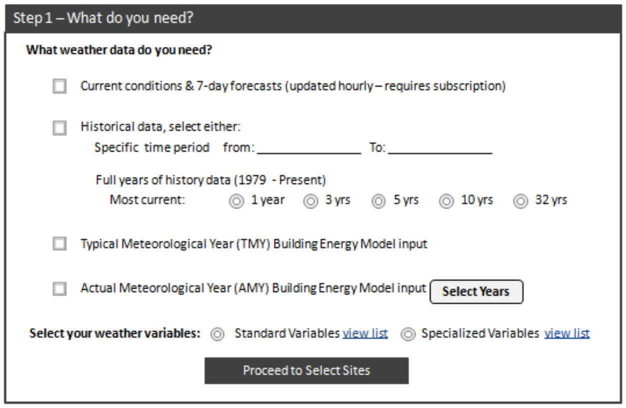


Fig. 14: Frequency distributions of GHI measured at Burns and extracted from various TMY constructions.



Step 1 – What do you need?

What weather data do you need?

☐ Current conditions & 7-day forecasts (updated hourly – requires subscription)

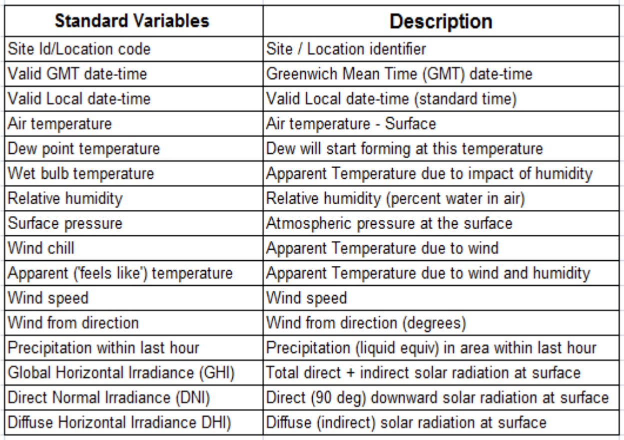
☐ Historical data, select either:
Specific time period from: ______________ To: ______________
Full years of history data (1979 - Present)
Most current: ○ 1 year ○ 3 yrs ○ 5 yrs ○ 10 yrs ○ 32 yrs

☐ Typical Meteorological Year (TMY) Building Energy Model input

☐ Actual Meteorological Year (AMY) Building Energy Model input

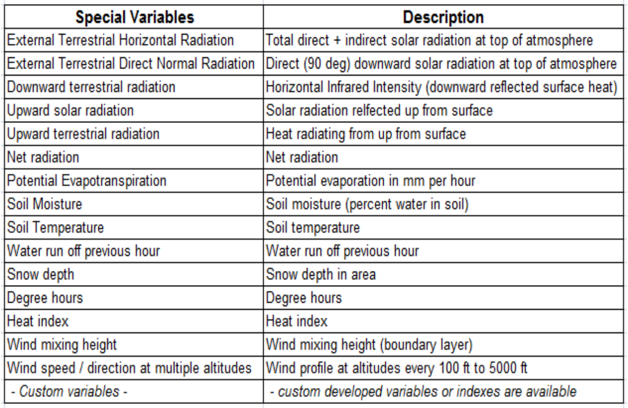
Select your weather variables: ○ Standard Variables view list ○ Specialized Variables view list

Fig. 15: Specify services, time period and variables



Standard Variables	Description
Site Id/Location code	Site / Location identifier
Valid GMT date-time	Greenwich Mean Time (GMT) date-time
Valid Local date-time	Valid Local date-time (standard time)
Air temperature	Air temperature - Surface
Dew point temperature	Dew will start forming at this temperature
Wet bulb temperature	Apparent Temperature due to impact of humidity
Relative humidity	Relative humidity (percent water in air)
Surface pressure	Atmospheric pressure at the surface
Wind chill	Apparent Temperature due to wind
Apparent ('feels like') temperature	Apparent Temperature due to wind and humidity
Wind speed	Wind speed
Wind from direction	Wind from direction (degrees)
Precipitation within last hour	Precipitation (liquid equiv) in area within last hour
Global Horizontal Irradiance (GHI)	Total direct + indirect solar radiation at surface
Direct Normal Irradiance (DNI)	Direct (90 deg) downward solar radiation at surface
Diffuse Horizontal Irradiance DHI)	Diffuse (indirect) solar radiation at surface

Fig. 16: Standard variables with all deliveries



Special Variables	Description
External Terrestrial Horizontal Radiation	Total direct + indirect solar radiation at top of atmosphere
External Terrestrial Direct Normal Radiation	Direct (90 deg) downward solar radiation at top of atmosphere
Downward terrestrial radiation	Horizontal Infrared Intensity (downward reflected surface heat)
Upward solar radiation	Solar radiation relfected up from surface
Upward terrestrial radiation	Heat radiating from up from surface
Net radiation	Net radiation
Potential Evapotranspiration	Potential evaporation in mm per hour
Soil Moisture	Soil moisture (percent water in soil)
Soil Temperature	Soil temperature
Water run off previous hour	Water run off previous hour
Snow depth	Snow depth in area
Degree hours	Degree hours
Heat index	Heat index
Wind mixing height	Wind mixing height (boundary layer)
Wind speed / direction at multiple altitudes	Wind profile at altitudes every 100 ft to 5000 ft
- Custom variables -	*- custom developed variables or indexes are available*

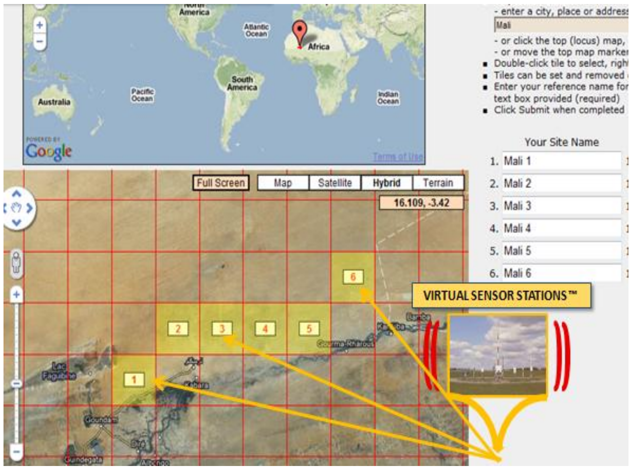
Fig 17: Additional specialized variables

The user then proceeds to locate the site or sites of interest either as specific points or 35-km areas (Figs. 18, 19).

- To position either:
  - enter a city, place or address:
  Taj MaHal, Agra India
  - or click the top (locus) map,
  - or move the top map marker
- Double-click lower map to select
  to remove
- Sites can be moved but are fixed
- Enter your reference name for ea
  text box provided (required)
- Click Submit when completed

Your Site Name
1. Taj MaHal

Fig. 18: Specify one or more point locations

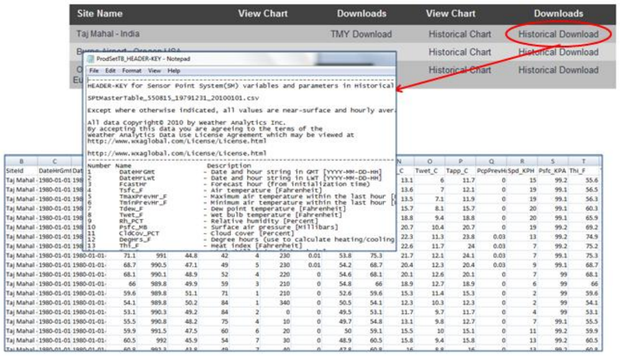


- enter a city, place or address
Mali
- or click the top (locus) map,
- or move the top map marker
- Double-click tile to select, righ
- Tiles can be set and removed
- Enter your reference name for
  text box provided (required)
- Click Submit when completed

Your Site Name
1. Mali 1
2. Mali 2
3. Mali 3
4. Mali 4
5. Mali 5
6. Mali 6

Fig. 19: Specify one or more area locations

A 30+ year hourly history or TMY file is finally completed within 30 minutes and the user notified of its availability for direct download. A typical file structure is shown in Fig. 20.



Site Name	View Chart	Downloads	View Chart	Downloads
Taj Mahal - India		TMY Download	Historical Chart	Historical Download
Burns Airport - Oregon USA			Historical Chart	Historical Download
			Historical Chart	Historical Download

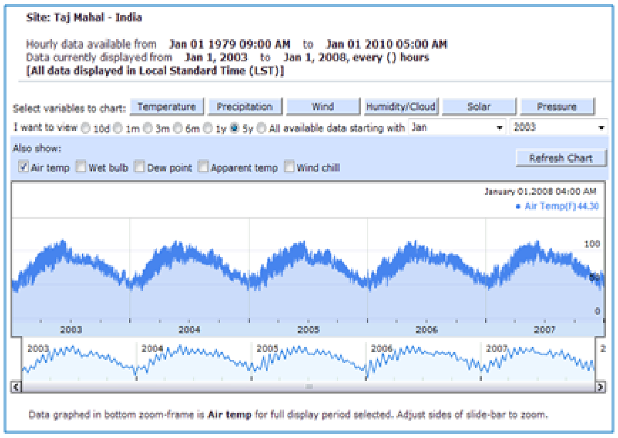
ProdSetTB_HEADER-KEY - Notepad

```
----------------------------------------------------------------
HEADER-KEY for Sensor Point System(SM) variables and parameters in Historical

SPtMasterTable_550815_19791231_20100101.csv

Except where otherwise indicated, all values are near-surface and hourly aver

All data Copyright© 2010 by Weather Analytics Inc.
By accepting this data you are agreeing to the terms of the
Weather Analytics Data Use License Agreement which may be viewed at
http://www.wxaglobal.com/License/License.html

http://www.wxaglobal.com/License/License.html
----------------------------------------------------------------
Number Name                Description
1      DateHrGmt           - Date and hour string in GMT [YYYY-MM-DD-HH]
2      DateHrLwt           - Date and hour string in LWT [YYYY-MM-DD-HH]
3      FcastHr             - Forecast hour (from initialization time)
4      Tsfc_F              - Air temperature [Fahrenheit]
5      TmaxPrevHr_F        - Maximum air temperature within the last hour [
6      TminPrevHr_F        - Minimum air temperature within the last hour [
7      Tdew_F              - Dew point temperature [Fahrenheit]
8      Twet_F              - Wet bulb temperature [Fahrenheit]
9      Rh_PCT              - Relative humidity [Percent]
10     Psfc_MB             - Surface air pressure [Millibars]
11     CldCov_PCT          - Cloud cover [Percent]
12     DegHrs_F            - Degree hours (use to calculate heating/cooling
13     Thi_F               - Heat index [Fahrenheit]
```

Fig. 20: Detailed hourly time-series downloads

4.2 Data display and Graphical Review

Once a user's site is activated, the user may review all variables with a graphics tool that supports zooming from the full period down to a few days or hours (Fig. 21).



Site: Taj Mahal - India

Hourly data available from **Jan 01 1979 09:00 AM** to **Jan 01 2010 05:00 AM**
Data currently displayed from **Jan 1, 2003** to **Jan 1, 2008, every () hours**
[All data displayed in Local Standard Time (LST)]

Select variables to chart: Temperature Precipitation Wind Humidity/Cloud Solar Pressure

I want to view ○ 10d ○ 1m ○ 3m ○ 6m ○ 1y ● 5y ○ All available data starting with Jan 2003

Also show:

☑ Air temp ☐ Wet bulb ☐ Dew point ☐ Apparent temp ☐ Wind chill

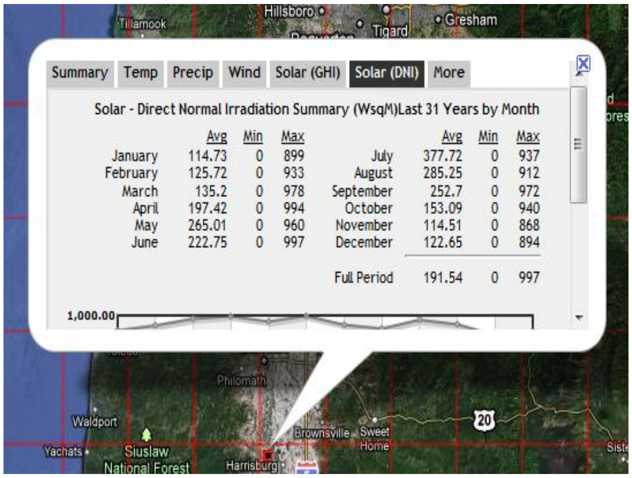
Data graphed in bottom zoom-frame is **Air temp** for full display period selected. Adjust sides of slide-bar to zoom.

Fig. 21: Graphical review with zooming

Quick review of potential sites of interest before requesting the full hourly history is facilitated by the availability of 31-year summary data for over 600,000 Virtual Sensor Stations™ sites across the globe (Fig. 22).

4.3 Data Delivery Options

Users may request the full complement of variables in their output or may select a subset of interest. To support a variety of modeling and simulation systems the historical output data is made available in multiple formats, including .CSV, .EPW, .TM2 and .BIN. Users may also request individual years in .EPW or .BIN formats to support targeted and comparative analysis of years with suspected highs and lows in GHI and DNI. For automatic systems full Web Data Services access is supported.



Solar - Direct Normal Irradiation Summary (WsqM)Last 31 Years by Month

	Avg	Min	Max		Avg	Min	Max
January	114.73	0	899	July	377.72	0	937
February	125.72	0	933	August	285.25	0	912
March	135.2	0	978	September	252.7	0	972
April	197.42	0	994	October	153.09	0	940
May	265.01	0	960	November	114.51	0	868
June	222.75	0	997	December	122.65	0	894
				Full Period	191.54	0	997

Fig. 22: 31-year monthly summary data for GHI and DNI

5. POTENTIAL SOLAR APPLICATIONS

5.1 Site Selection and Analysis

Potentially, to those wanting to perform solar site selection analyses, the results of these efforts based on the CFSR dataset represent an accessible network of "virtual weather stations" placed one every 35 km around the globe that have been in continuous operation since January 1979, reporting with 100 % reliability each hour. The interactive and direct access to the numerical data output of these virtual stations can be used to support solar site selection and analysis at multiple levels, including:

- Initial identification of potential sites of interest using an interactive map-based interface with 30-year climate-summary data
- On-location TMY files for modeling and first-level analysis of site potential
- Detailed on-site 31-year hour-by-hour time series data and 30 single-year EnergyPlus input files to support more detailed site analyses.

Under development are techniques for de-biasing the 31-year CFSR-based output utilizing short-term on-site sensor readings. These will exploit model output statistics (MOS) and full-physics column model technologies. The resulting dataset effectively extends a measured data record of as little as one year, and the reach of statistical analyses, to 31 years. The ability to select individual years from the 31-year record enables a more accurate characterization of the potential short-term or long-term variability, particularly when TMY-type data are not enough.

5.2 High Precision Forecasting

While the historical data output discussed above tracks the 35-km spatial resolution of the CFSR data, the WxA Sensor Point™ system is able to provide far higher precision for forecasts, particularly in the first 24 hours. Sensor Point™ uses a column model approach to integrate the roughly 35-km resolution NOAA Global Forecast System (GFS) output with nearby sensor readings, thereby increasing the GFS spatial accuracy and forecast frequency to that of the sensor output, and localizing the forecast to footprints as small as 1 km.

Sensor Point™ can use private sensors, including radiometers, with real-time remote data access. These on-site sensors enable short-term solar radiation and other key variable forecasts specific to that site. Since there is complete control of the observational data, frequent updates are an option. In this way, the 35 meteorological variables that are available for the last 31-year period can also be part of 7-day forecast time series. Among other important applications, such forecasts are an efficient tool to help electric utilities and independent service operators manage and dispatch the variable production of solar electricity in their grid. Similar integrated sensor-driven column model systems have been successfully deployed to support other weather-sensitive operations, such as those controlling airport acceptance rates [5].

6. SUMMARY

The global network of over 600,000 reliable and accurate Virtual Sensor Stations™ to support site research and analysis has become a convenient reality. Time series data are now readily available via the WxA web site portal (http://www.wxaglobal.com) for over 35 weather variables, including GHI and DNI. Both historical and forecast time series data are available.

The WxA historical database is based on the hour-by-hour, 31-year, 35-km resolution CFSR model output data. The CFSR model uses raw sensor data and physics to provide for this period a detailed and accurate picture of the earth-atmosphere system. The WxA hour-by-hour, 7-day, 35-km resolution forecast database is built on the GFS model output data localized with Sensor Point column model technology. Both WxA historical and forecast databases hold the same variables.

Uses for the WxA data include first-pass analyses of candidate solar sites, simulations of building energy systems and other environmentally sensitive systems. Input data files are available in formats that can be immediately used as input to popular spreadsheet tools and software applications, such as the EnergyPlus ‘epw’ format. Data file content can either be hour-by-hour time series for periods up to 31 years or based on Typical Meteorological Year (TMY) statistical criteria. Single-year time series data can be provided as “Actual Meteorological Year” (AMY) epw and .bin format files.

6. ACKNOWLEDGEMENTS

The CFSR data was developed by NOAA's National Centers for Environmental Prediction (NCEP). The data for this study are from NOAA's National Operational Model Archive and Distribution System (NOMADS) which is maintained at NOAA's National Climatic Data Center (NCDC).

The University of Oregon Solar Radiation Monitoring Laboratory and their network of monitoring stations provide sound solar resource data for planning, design, deployment, and operation of solar electric facilities and for improved tuning of satellite solar radiation data. The Baseline Surface Radiation Network (BSRN) is part of the World Radiation Monitoring Center. Both networks are commended for the quality of their data.

7. REFERENCES

1. Saha S., et al., The NCEP Climate Forecast System Reanalysis. Bull. Amer. Meteor. Soc., 91, 1015–1057, 2010.
2. Gueymard C.A. and Wilcox S.M., Assessment of Spatial and Temporal Variability in the U.S. Solar Resource from Radiometric Measurements and Predictions from Models Using Ground-Based or Satellite Data. Solar Energy, in press, 2011.
3. Perez R.R. et al., Dynamic global-to-direct irradiance conversion models. ASHRAE Trans. 98(1): 354-369, 1992.
4. Marion W. and Urban K. User’s manual for TMY2s Typical Meteorological Years. NREL, 1995.
5. Keller, J. L., Smith, C. B., Wilson, F. W., Applications of Column Models for Terminal Weather Nowcasts, Proc. 6th Conf. on Aviation Weather Systems, Dallas, TX, Amer. Meteor. Soc., pp. 66-71, 1995.