Implementing Best Practices for Data Quality Assessment of the National Renewable Energy Laboratory's Solar Resource and Meteorological Assessment Project

Preprint

S. M. Wilcox and P. McCormack

To be presented at SOLAR 2011
Raleigh, North Carolina
May 16–21, 2011

NREL is a national laboratory of the U.S. Department of Energy, Office of Energy Efficiency & Renewable Energy, operated by the Alliance for Sustainable Energy, LLC.

Conference Paper
NREL/CP-5500-50897
April 2011

Contract No. DE-AC36-08GO28308

Available electronically at http://www.osti.gov/bridge

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
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/help/ordermethods.aspx

Cover Photos: (left to right) PIX 16416, PIX 17423, PIX 16560, PIX 17613, PIX 17436, PIX 17721

Printed on paper containing at least 50% wastepaper, including 10% post consumer waste.

IMPLEMENTING BEST PRACTICES FOR DATA QUALITY ASSESSMENT OF THE NATIONAL RENEWABLE ENERGY LABORATORY'S SOLAR RESOURCE AND METEOROLOGICAL ASSESSMENT PROJECT

Stephen Wilcox
Paul McCormack
National Renewable Energy Laboratory
1617 Cole Blvd.
Golden, CO 80401
e-mail: stephen.wilcox@nrel.gov

ABSTRACT

Effective solar radiation measurements for research and economic analyses require a strict protocol for maintenance, calibration, and documentation to minimize station downtime and data corruption. The National Renewable Energy Laboratory's *Concentrating Solar Power: Best Practices Handbook for the Collection and Use of Solar Resource Data* (1) includes guidelines for operating a solar measurement station. This paper describes a suite of automated and semi-automated routines based on the best practices handbook as developed for the National Renewable Energy Laboratory Solar Resource and Meteorological Assessment Project. These routines allow efficient inspection and data flagging to alert operators of conditions that require immediate attention. Although the handbook is targeted for concentrating solar power applications, the quality-assessment procedures described are generic and should benefit many solar measurement applications. The routines use data in one-minute measurement resolution, as suggested by the handbook, but they could be modified for other time scales.

1. INTRODUCTION

The need for economic analyses for deploying large- and small-scale solar power plants has underscored the urgent need for requirements for high-quality solar resource data. Although solar data sets exist for the United States, their uncertainty or geographical coverage may be inadequate for some high-risk projects. The best practices handbook (1), page 106, and (3) note that high-quality measured data may be used to develop a more robust data set and reduce the solar resource uncertainty at targeted geographical locations.

Toward that end, the National Renewable Energy Laboratory (NREL) collaborates with industry partners through the Solar Resource and Meteorological Assessment Project (SOLRMAP) (2). Industry partners underwrite the cost of equipment and station operations, while NREL provides expertise for station design, deployment, operations, and data quality analysis. Much of NREL's obligation is dedicated to daily data inspections, as outlined in the best practices handbook. In summary, the handbook mandates consideration of several parameters for robust station design and provides detail on methods of data quality control and quality assessment. These practices are aimed at averting negative questions about the data and instead instilling confidence about data quality for those performing due diligence or determining a project's financial viability.

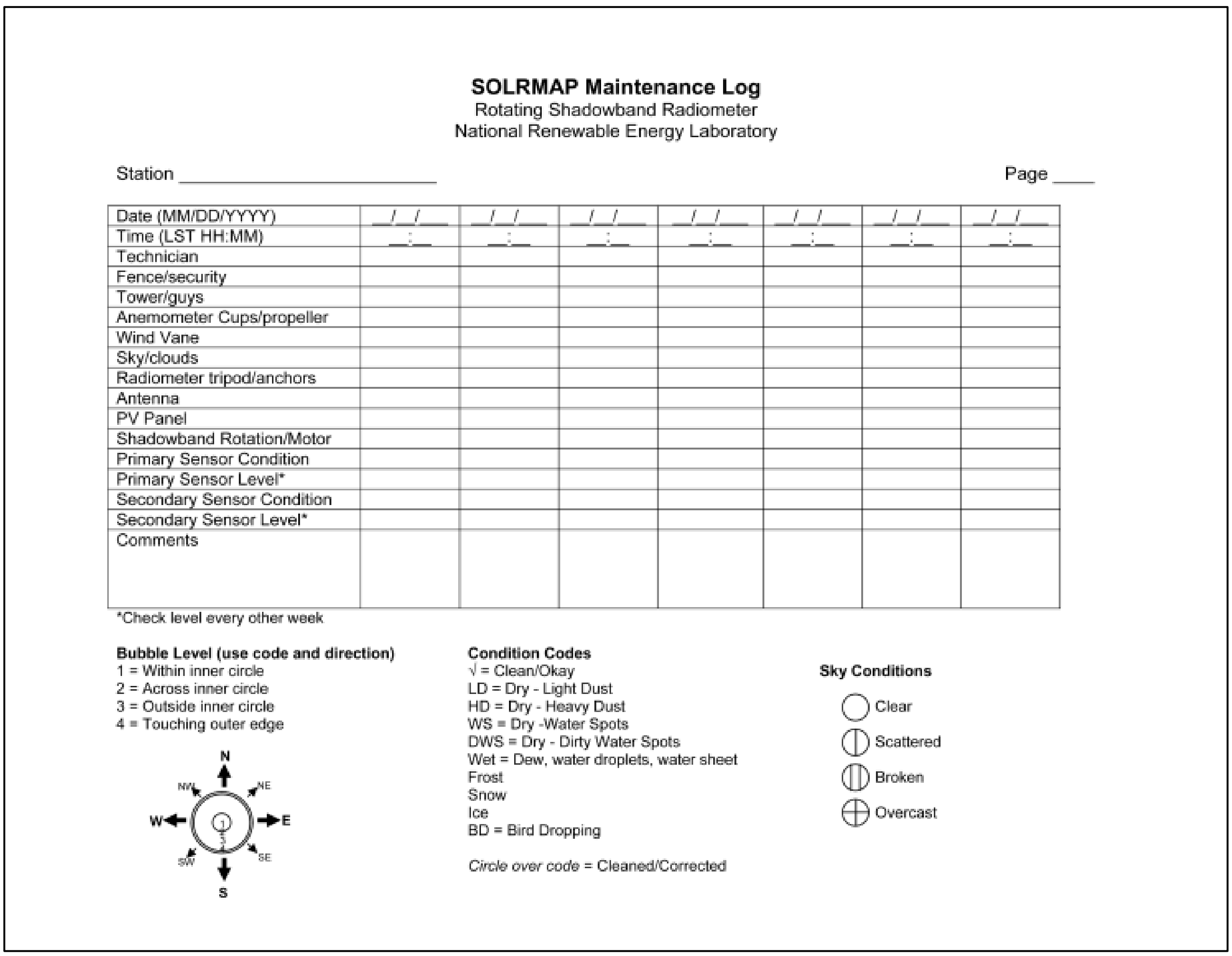
2. QUALITY CONTROL AND DATA QUALITY ASSESSMENT

The quality of a measurement is largely fixed at the time of data acquisition. This means it is difficult or impossible to improve data quality after a measurement is taken. Although some improvements to data seem plausible (e.g. correcting a wrong calibration factor), *knowledge of conditions at the time of the measurement cannot be expanded beyond what is known at that time*. If it is known with high probability that the instrument optics are clean, the tracker is aligned, and a recent calibration was performed, one can make a strong case regarding the uncertainty of the measurement based on instrument specifications. If, however, such checks are not routinely preformed, the measurement conditions are uncertain, as is the magnitude of errors associated with poor maintenance. Hence, worst-case assumptions may be required to properly express the quality and uncertainty of the data. As a rule, it is more efficient and trustworthy to engage a quality-control procedure that helps maintain and monitor measurement conditions. With such a procedure in place, one can more easily assert a defensible argument about the data quality.

The best practices handbook describes quality control as a "well-defined supervisory process" that helps ensure measurements are of expected quality. Part of quality control is based on proper station design, and the handbook describes these parameters:

- Location
- Station security and accessibility
- Power requirements
- Grounding and shielding
- Data acquisition
- Data communications
- Operations and maintenance
- Radiometer calibrations.

Data quality assessment provides a method to quantify and express the quality of data so users understand any errors associated with the measurement. According to the best practices handbook (1), page 39, "Data quality assessment is a method by which data quality can be judged based on criteria for a particular application." Such criteria can include comparison with physical limits determined to be reasonable, redundant or complementary measurements, or models. These comparisons offer some degree of independent measure to perform a judgment on quality. This approach applies not only to primary data (the target measurements) but also to operational parameters such as battery voltage or connectivity uptime.



SOLRMAP Maintenance Log
Rotating Shadowband Radiometer
National Renewable Energy Laboratory

Station ________________________ Page ____

Date (MM/DD/YYYY)	__/__/____	__/__/____	__/__/____	__/__/____	__/__/____	__/__/____	__/__/____
Time (LST HH:MM)	__:__	__:__	__:__	__:__	__:__	__:__	__:__
Technician							
Fence/security							
Tower/guys							
Anemometer Cups/propeller							
Wind Vane							
Sky/clouds							
Radiometer tripod/anchors							
Antenna							
PV Panel							
Shadowband Rotation/Motor							
Primary Sensor Condition							
Primary Sensor Level*							
Secondary Sensor Condition							
Secondary Sensor Level*							
Comments							

*Check level every other week

Bubble Level (use code and direction)
1 = Within inner circle
2 = Across inner circle
3 = Outside inner circle
4 = Touching outer edge

Condition Codes
√ = Clean/Okay
LD = Dry - Light Dust
HD = Dry - Heavy Dust
WS = Dry -Water Spots
DWS = Dry - Dirty Water Spots
Wet = Dew, water droplets, water sheet
Frost
Snow
Ice
BD = Bird Dropping

Circle over code = Cleaned/Corrected

Sky Conditions
Clear
Scattered
Broken
Overcast

Fig. 1: Sample maintenance log.

SOLRMAP quality control begins with station design and an emphasis on the bullet points above. Although industry partners determine station location, NREL offers advice on micro-siting (e.g., away from power lines, poles, and buildings that may shade or reflect light and corrupt measurements). NREL inspects each installation and performs on-site validation to ensure equipment is operating as expected and to generate a commissioning report with data and documentation. This important step demonstrates that the station began operations fully functional.

The quality-control paradigm also includes maintenance protocols specific to the type of solar sensor. Radiometer designs with diffuser elements above the detector (such as the rotating shadowband radiometer, RSR) can be maintained on an every-other week schedule. Pyranometers with domes and pyrheliometers with protective windows are more prone to soiling and require daily inspection to ensure clean optical surfaces. On-site training conveys the maintenance protocol to personnel, and they are provided with a summary document for future reference. Training includes step-by-step procedures for equipment inspection and proper documentation. Maintenance logs correspond with the inspection steps and provide a record of conditions at the time of the visit. A sample log is shown in Fig. 1.

Maintenance technicians are trained onsite by NREL staff or other qualified instructors. The two-hour training includes an overview of the purpose of the measurements and the instrumentation as well as hands-on details of cleaning and maintenance procedures. After training, the technician completes a real-time site inspection visit with the trainer.

With the knowledge of maintenance logs and an inspection protocol, analysts have increased confidence that (1) a rigorous inspection routine exists; (2) problems are discovered, documented, and corrected; and (3) at all times other than during reported errors, *the station was found to be in good working order*. The last item affirms an important nuance that proper equipment operations dominate the station's history.

The written log is the permanent record of station maintenance, but an online log is also used to expedite the transfer of information from the maintenance technician to network operations. Using Google Docs (http://docs.google.com), the maintenance technician accesses an interactive interface to record observations from the maintenance log (Fig. 2), and these observations populate an underlying document viewable by the QA expert. Google Docs can be configured to send an e-mail alert when any changes to the underlying document occur, providing nearly instant f eedback for maintenance activities among many stations.

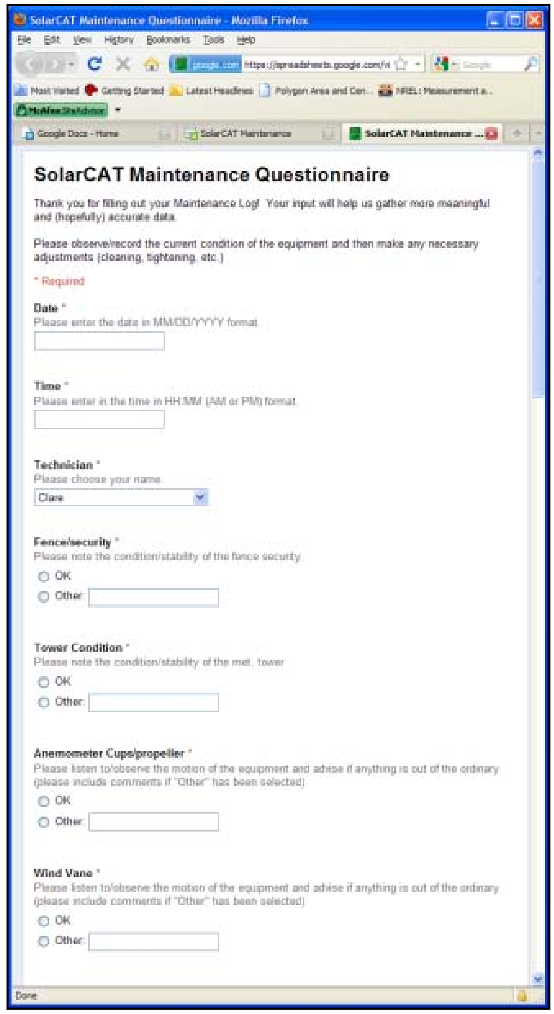


Fig. 2: User interface for online maintenance log.

The best practices handbook discourages using volunteers as maintenance technicians and advises a contracted payment for the service. The importance of regular maintenance cannot be overstated. Although volunteers are an attractive proposition, absent a contract, consistent performance cannot be guaranteed.

These documents and this protocol provide valuable information to analysts regarding the quality of the data they use for critical analyses. The intent is to avert questions about the data and allow data recipients to concentrate on the merits of the analysis.

3. DATA INSPECTION

With a robust station design, expert station deployment, proper startup validation, and a rigorous maintenance protocol in place, confidence in the station and its measurements is bolstered. However, problems can occur that are not

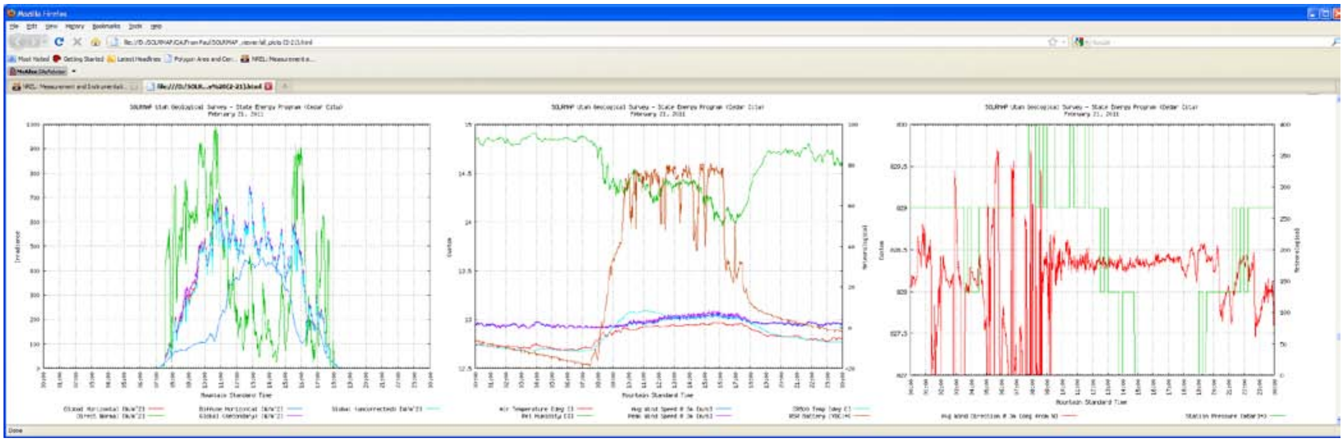


Fig. 3: Multiply-grouped data plots for a single station.

readily apparent from an onsite inspection. A large part of network operations is the routine inspection of data to determine if it is reasonable. NREL has developed several data inspection tools to aid this endeavor. Although these tools are not publically available, the samples and description here should provide the reader with enough information to implement most of the concepts.

The SOLRMAP network generates on the order of a half million data values per day, making detailed scrutiny difficult. Even if it were practical to inspect each data value, viewing the measurement in context is critical to an understanding of the data set as a whole. Data inspection requires significant human involvement, and prospective station operators should be prepared for a labor-intensive effort. However, an efficient and effective data inspection routine facilitates rapid assimilation of data in such a way that makes error conditions conspicuous to the data quality analyst.

3.1 Connectivity and Data Time Series

The first step in the data inspection routine is to determine if the station is operating and online. The NREL Measurement and Instrumentation Data Center (http://www.nrel.gov/midc) automates data assimilation from SOLRMAP and other stations. The Measurement and Instrumentation Data Center issues automated alerts if a station is delinquent in reporting by more than a few hours. (Short delays in communications are common and are generally disregarded.) Data are gathered from the network multiple times each day, providing near real-time access to station measurements for inspection.

The daily overview of each station begins with a time series plot of each parameter. The plots are rapidly generated from the database for quick viewing. This presentation allows the data quality expert to view all measurements in context. For example, data are examined to determine if the three components of solar radiation data look reasonable relative to one another, whether PV battery voltage corresponds with the magnitude of solar irradiance, whether temperature and humidity are correlated; and if wind speed and temperature changes indicate a frontal passage. In Fig. 3, data from the station are grouped and presented in an easily digestible form. With experience, anomalies in the data are conspicuous and can be noted for further investigation. Notes for all parameters are kept in a spreadsheet for permanent record of data inspection results.

In addition, like measurements from nearby sites are superimposed to identify outliers. Data can be grouped geographically to limit the expected variation from site to site, as shown in Fig. 4. In this plot, irradiance data from two relatively nearby sites in Utah, Cedar City and Milford, are plotted together. The cleaning at Cedar City (a), the spike at Cedar City (b), and the later cleaning at Milford (c) are easily distinguishable.

3.2 Long-Term Trends

Long-term trends can reveal problems that may not be evident at shorter time scales. For example, Fig. 5 shows a distinct dip in battery voltage that foretold a loss of station power. In this case, station downtime was avoided by correcting the problem before data were lost.

3.3 Redundant Measurements

For SOLRMAP, NREL mandates that RSR instruments be equipped with a redundant solar measurement. By elegant design, the RSR uses only one sensor to derive the three fundamental components of global, direct, and diffuse irradiance. Although a catastrophic failure of a sensor would be readily apparent, a slight degradation may not be easily discernable. Even a loss of 10% sensitivity could be mistaken for a like attenuation of the solar signal because of dust or

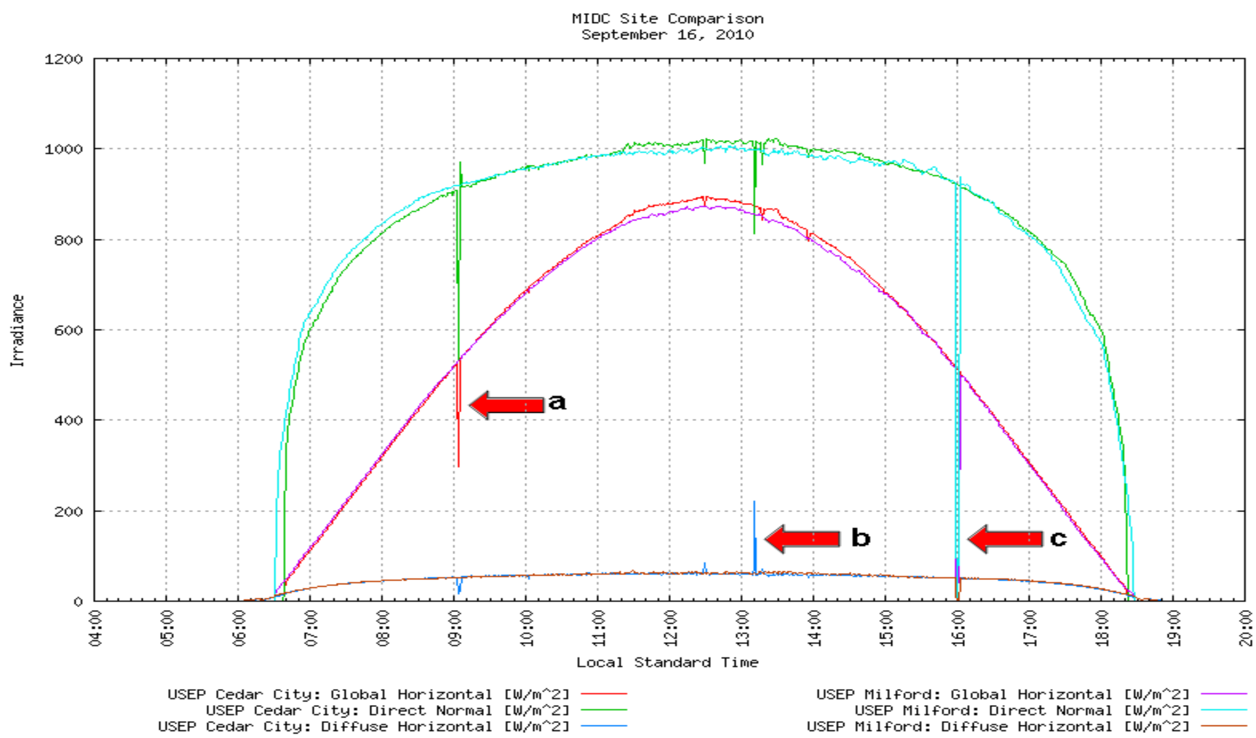


Fig. 4: Comparing the Utah irradiance data at nearby sites reveals anomalies, including (a) a cleaning event at Cedar City, (b) a spurious data sample, and (c) a cleaning event at Milford. The clear sky irradiance is expected to be similar at the two sites.

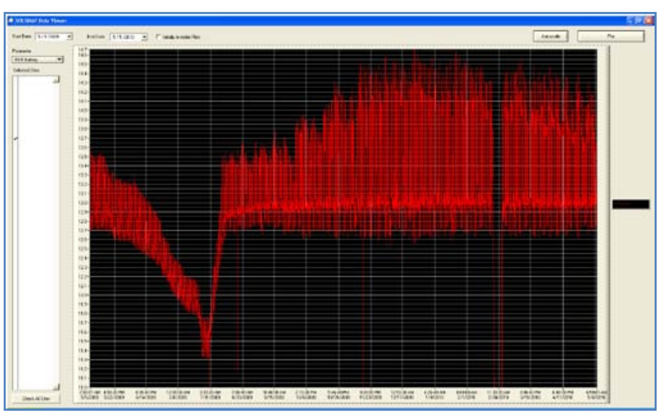


Fig. 5: Battery voltage trend indicating charging problem and correction.

water vapor. In fact, finding true attenuations of that nature can be a primary goal of a measurement campaign and can be expected. So irradiance magnitude by itself is not a reliable indicator of data quality. To help detect such errors, the redundant sensor can be used as an onsite check for sensor degradation, soiling, and unlevel conditions. The cost of the secondary sensor is small compared with the overall cost of the system (about 3%), and the tradeoff in measurement confidence is well worth it.

The secondary sensor is of the same model as the primary sensor (LI-COR LI200), so the two instruments are expected to respond similarly to the solar stimulus and other environmental conditions. This assumption is easily checked by examining the ratio of the two global horizontal measurements on a long-term basis. Fig. 6 shows an example of midday ratio averages plotted over many months. The plot reveals seasonal (or temperature) effects, soiling effects, or leveling problems.

This approach assumes that these problems affect the sensors at different rates and different magnitudes, but as is apparent from the figure, these differences do occur and can be detected. When significant excursions occur, the equipment can be examined to determine the cause. These ratios can also provide valuable insight into the measurement uncertainty of the instrument. As the ratios vary in normal use, one can assume the sensor variability for the measurement is much the same. (An even greater variability could possibly be surmised if both sensors were reacting similarly to some environmental source of error.)

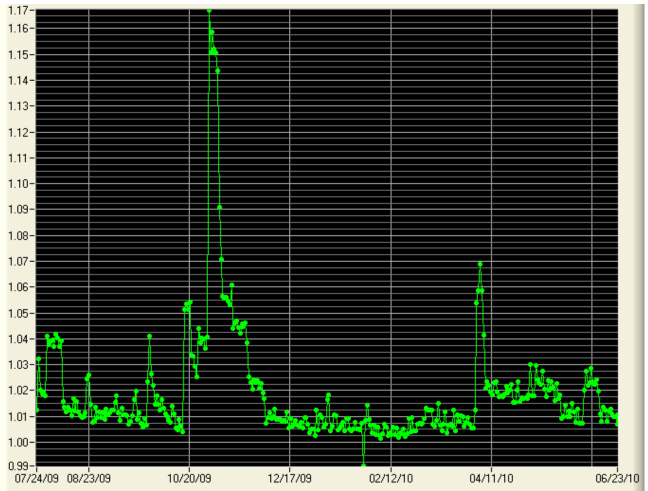


Fig. 6: Ratio of primary and secondary sensor over time showing variations at the La Ola SOLRMAP station in Hawaii.

3.4 Solar Data Quality Assessment

NREL uses the SERI QC data quality software package (4), which relies primarily on the redundancy of three component solar measurements, where the global measurement equals the sum of the diffuse and direct times the cosine of the solar zenith angle. SERI QC operates in K-space, a clearness index derived by normalizing the global, direct, and diffuse irradiance measurements by the extraterrestrial solar radiation at the top of the atmosphere on either a horizontal surface (ETR) or normal to the sun (ETRN). In K-space:

$$K_t = \text{Global} / \text{ETR} \qquad (1)$$
$$K_n = \text{Direct} / \text{ETRN} \qquad (2)$$
$$K_d = \text{Diffuse} / \text{ETR} \qquad (3).$$

In K-space, the three components are related by:

$$K_t = K_n + K_d \qquad (4).$$

And consequently, a residual error component is represented by

$$\varepsilon = K_t - K_n - K_d. \qquad (5).$$

For a further discussion of K-space, see (1), page 9.
SERI QC assigns some data quality flags based on Eq. 5, which represent a departure from the expected perfect coupling of the three components. Although these flags by themselves do not indicate which of the three components is bad, they provide an alert to abnormal conditions that may warrant further investigation. However, certain combinations of flags suggest common errors. For example, a high diffuse measurement or a low direct measurement likely follows from faulty tracker alignment. Further, errors not apparent in a single value may be revealed in a time series of flags in the context of nearby measurements. For example, error conditions based on time of day or solar zenith angle may be conspicuous in a flag-versus-time or flag-versus-zenith angle plot.

To assist in the assimilation of a large amount of data, NREL developed color-coded plots of SERI QC flags to view the results from automated quality analysis (5). Fig. 7 shows such a plot for a month of data. These plots show day of month on the vertical axis and hour of day on the horizontal axis. The left-most plot depicts flag severity, with the least error in dark blue and the greatest error in red. The next three plots show K_t, K_n, and K_d, respectively, with the clearness range represented from dark blue (least) to bright red (greatest). These plots display solar data errors and associated measurements at a glance for an entire month. The plots also reveal patterns in error conditions, such as the frequent morning flags that may indicate dew on the sensors.

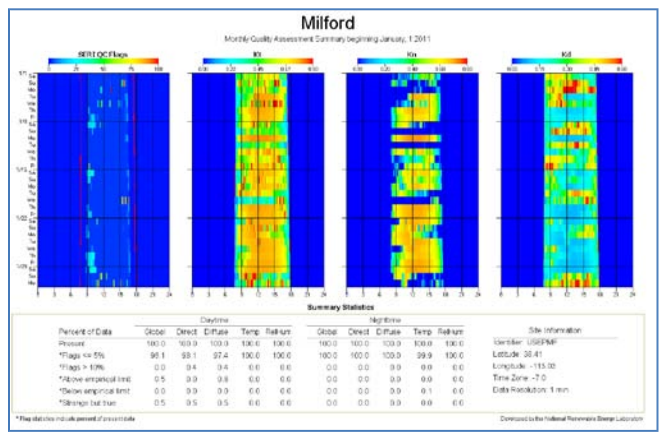


Fig. 7: Monthly SERI QC flag summary.

In the case of the RSR, which produces a direct solar value computed from global and diffuse measurements, SERI QC would consider all data perfectly coupled. To address this, data from the redundant sensor are adjusted for the long-term ratio explained in Section 3.3 and for internal instrument adjustments to act as a surrogate global measurement. This allows SERI QC to operate with at least two measurements, taking advantage of the three-component test.
The SERI QC software package includes the QCFIT tool to plot data as a K_n versus K_t scatter plot (Fig. 8). Among other capabilities, the tool provides an animated view of the data that draws attention to measurements that do not couple well in K-space. The tool allows the user to select questionable data on the plot and extract the exact data record(s) from the input file for further inspection.

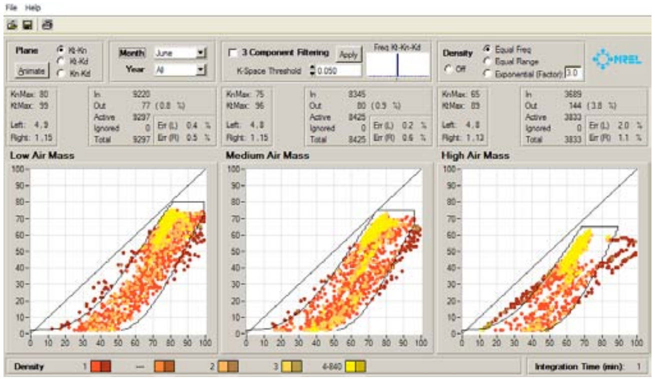


Fig. 8: QCFIT scatter plots.

4. DATA FILLING AND CORRECTION

Producing a measured time series of data for any reasonable length of time without missing records is essentially impossible. Even if equipment failure never happens (which is not out of the question but becomes increasingly rare with longer measurement campaigns), one must consider the disruptive effects of routine instrument cleaning. By its very nature, cleaning briefly blocks solar irradiance to the sensor (e.g., while cleaning fluids are applied or the instrument is rubbed with a cloth). Even if only of a few seconds, such interruptions appear as anomalies in the data.

Data filling for long periods (e.g., hours to days because of operational failure) is an exercise in modeling and beyond the scope of this paper. However, removing brief anomalies caused by maintenance and subsequently filling data over a period of a few minutes is more straightforward. Leaving the data uncorrected could have a significant effect on summary statistics, particularly for short time spans. Conversely, intelligent data filling should have very little adverse effect on the overall statistics of the data set. Hence, mitigating the effects of routine cleaning is a common requirement and worth some effort. For this purpose, NREL developed a semi-automated software tool to detect cleaning and other anomalous episodes and assist in the linear interpolation of data over the target period (typically five minutes or less). A before-and-after view of the data during a cleaning event is shown in Fig. 9.

In this case, the software suggested a proposed interpolation over the anomaly, which the operator accepted. The corrected data set (separate from the original) contains neither the cleaning (morning) nor the RSR spike (afternoon). The operator also has the option of hand-drawing a line with the computer mouse to represent the data, and the software can be configured to use any two components to calculate the third component. The software produces an output file with flags to indicate which data have been filled or estimated by

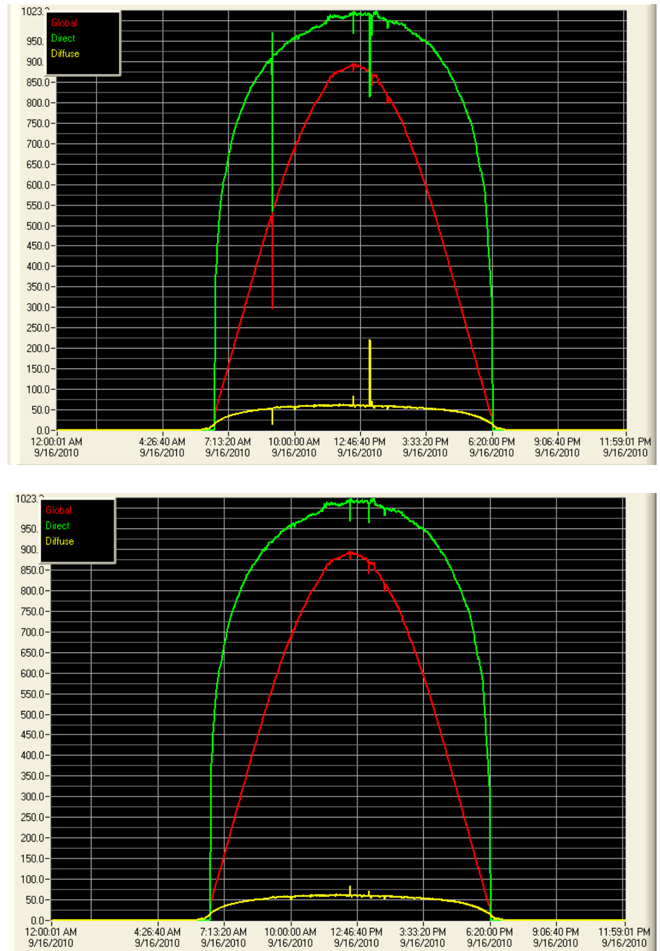


Fig. 9: Plots of the original and corrected irradiance data (in W/m^2).

interpolation. These flags remain in the data set to assist with an overall evaluation of measurement uncertainty in a summary analysis.

5. CALIBRATIONS

Routine instrument calibrations are fundamental to all scientific measurements. The calibration process provides a factor that allows conversion of an instrument's basic response (e.g., voltage) to an accurate representation in the desired engineering units (e.g., W/m^2). The calibration factor must be traceable to a recognized standard to make the measurements comparable in context of other similar measurements. The calibration process, along with manufacturer specifications, help determine the uncertainty of the instrument's measurements. Without valid and current calibration documentation, measurements from an instrument are an easy and legitimate target for skepticism from anyone scrutinizing the data. It is very difficult to declare a favorable measurement uncertainty for a dataset if a satisfactory calibration history cannot be documented.

Station operations should include a well-documented calibration routine. Careful records should be kept of calibration results and application to the network measurements.

6. DISCUSSION

Fundamental to a successful measurement campaign is the concept of a quality assurance cycle, which is depicted in Fig. 10. This introduces the notion of data acquisition, data quality assessment, and operations feedback in a looping, interrelated triad arrangement.

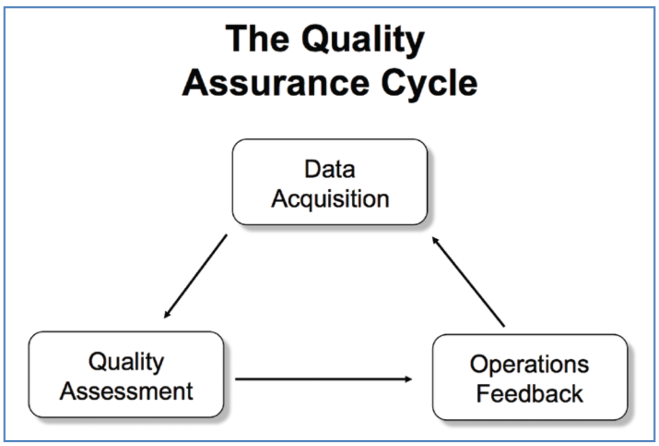


Fig. 10: The quality assurance cycle.

The quality assurance cycle provides quality assessment feedback to the measurement station as problems occur and results in corrections at the site to remove the cause of the errors. Multiple cycles can run simultaneously, each with a different period and scope. For example, daily onsite findings by the maintenance person (e.g., a soiled sensor) result in some rudimentary quality assessment ("This is dirty.") and informal feedback ("I should clean.") that leads to the correction of a substandard condition. Daily, weekly or monthly cycles include quality assessment using off-site data inspection built on the techniques described above, a formulation of feedback, and transmittal of recommendations to the site. Even annual cycles may reveal more subtle long-term problems.

The faster these cycles run and the more immediate the feedback, the quicker problems can be solved. This construct minimizes conditions that adversely affect data quality and ultimately results in a better data set.

7. CONCLUSION

The quality control and quality assessment procedures described here are part of a carefully crafted infrastructure to increase confidence in the quality of solar radiation measurements. A fundamental aspect of this protocol is the rapid assessment of data and the formulation of feedback to station operators to correct deficiencies. With practice, data from dozens of stations can be inspected daily in just an hour or two. Although considerable effort is expended to keep data paths operational, maintenance personnel engaged, and data inspection routines well-practiced, the payoff in fast response to problems is worth the effort. Further, the existence of this infrastructure allows analysts to apply the most favorable uncertainty estimates to the data and provides the background information necessary to defend a stated quality.

8. ACKNOWLEDGEMENTS

This work was supported by the U.S. Department of Energy under Contract No. DE-AC36-08-GO28308 with the National Renewable Energy Laboratory.

9. REFERENCES

(1) Stoffel, T., et al., Concentrating Solar Power Best Practices Handbook for the Collection and Use of Solar Resource Data, National Renewable Energy Laboratory technical report, NREL/TP-550-47465, September 2010

(2) Wilcox, S., and Myers, D., Joint Solar Power Industry and Department of Energy Solar Resource and Meteorological Assessment Project (SOLRMAP), Optical Modeling and Measurements for Solar Energy Systems III Proceedings of SPIE Volume: 7410, 2010

(3) Gueymard, C.A., and Wilcox, S.M., Spatial and Temporal Variability in the Solar Resource: Assessing the Value of Short-Term Measurements at Potential Solar Power Plant Sites, Solar 2009 Conf., Buffalo, NY, ASES, 2009

(4) NREL, User's Manual for SERI_QC Software – Assessing the Quality of Solar Radiation Data, National Renewable Energy Laboratory technical report, NREL/TP-463-5608, 1993

(5) Wilcox, S., A Visual Quality Assessment Tool for Solar Radiation Data, Proceedings of the 1996 Annual Conference of the American Solar Energy Society, Boulder, Colorado, April 1996