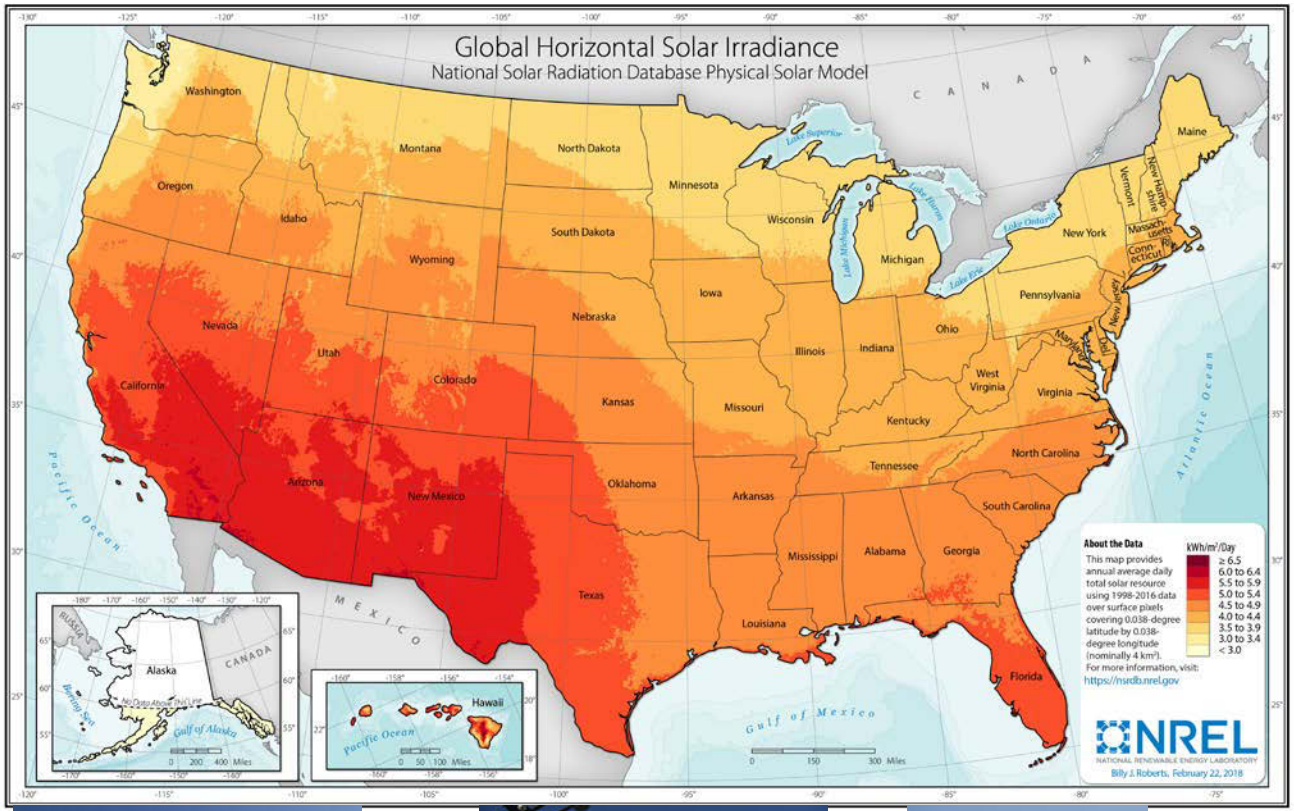
Data Quality Assessment Using SERI-QC

Manajit Sengupta



What Is Data Quality?

Data quality is a function of one's knowledge of the measurement environment and infrastructure.

1. Data quality is fixed (unchangeable) at the time a measurement is taken.
2. No amount of "quality control" after the fact can improve the fundamental data quality.
3. Data sets without good documentation are of *unknown quality*.

What Is Data Quality Assessment?

"A judgment of how well measurements represent the physical world."

Requires:

- A standard of measurement
- A criterion for judgment.

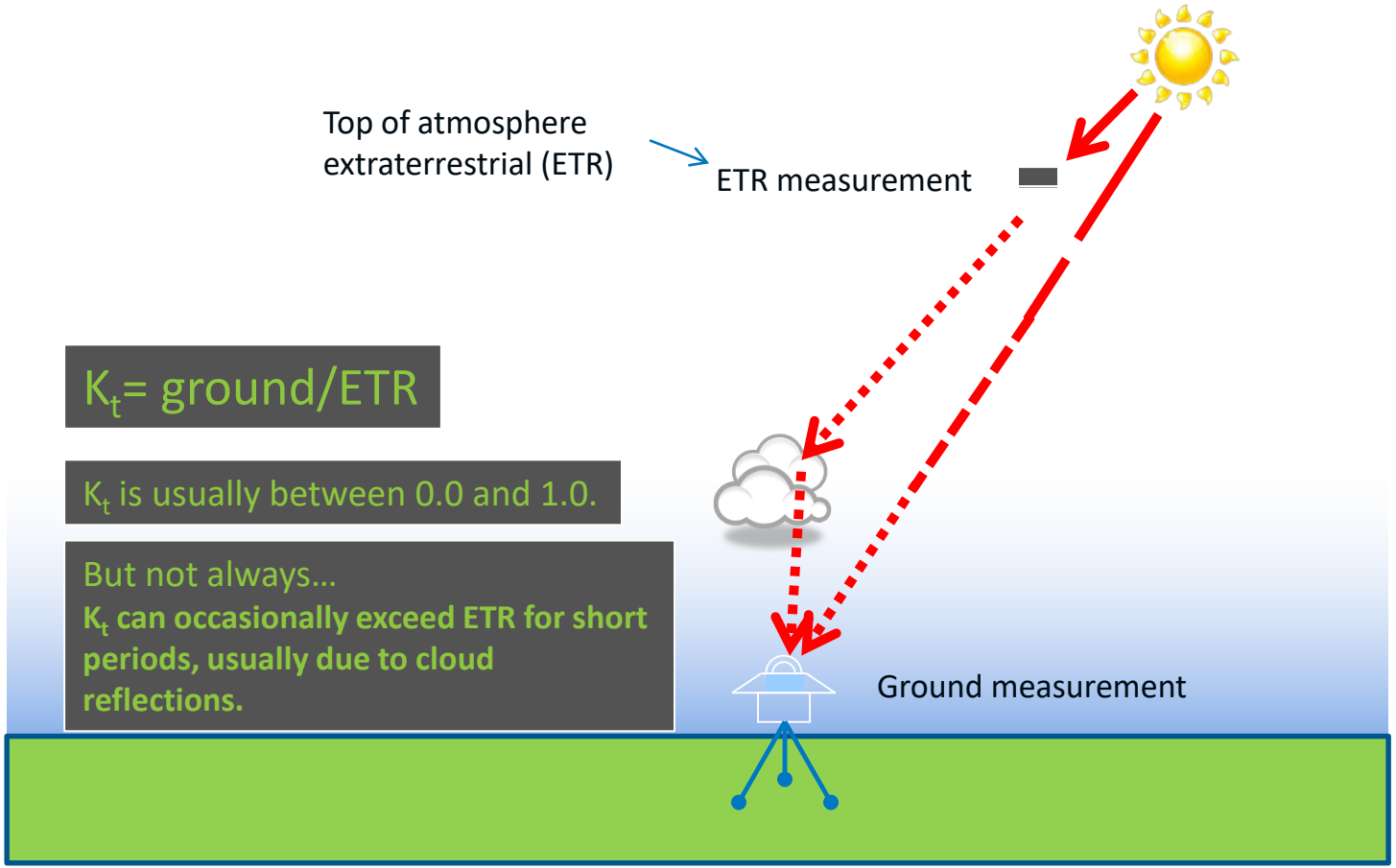
What Is Data Quality Assessment?

Quality assessment is not quality control.

- Quality assessment requires judgment and analysis. *This happens after the measurements*.

- Quality control is a supervisory process. *This happens before and during the measurements*.

Defining Clearness Index (K_t)



SERI-QC: Defining the K-Space

Operates in K-space: fraction of possible irradiance

Variable	Definition
K_t	Global / (ETRN * cos (Z))
K_n	Direct / ETRN
K_d	Diffuse / (ETRN * cos(Z))

ETRN = extraterrestrial radiation normal to the sun (DNI above the atmosphere)
K-space subscripts: t = total (global), n = normal (DNI), d = diffuse

$$K_t = K_n + K_d$$

https://www.nrel.gov/docs/legosti/old/5608.pdf

Data Quality Assessment Using NREL's SERI-QC

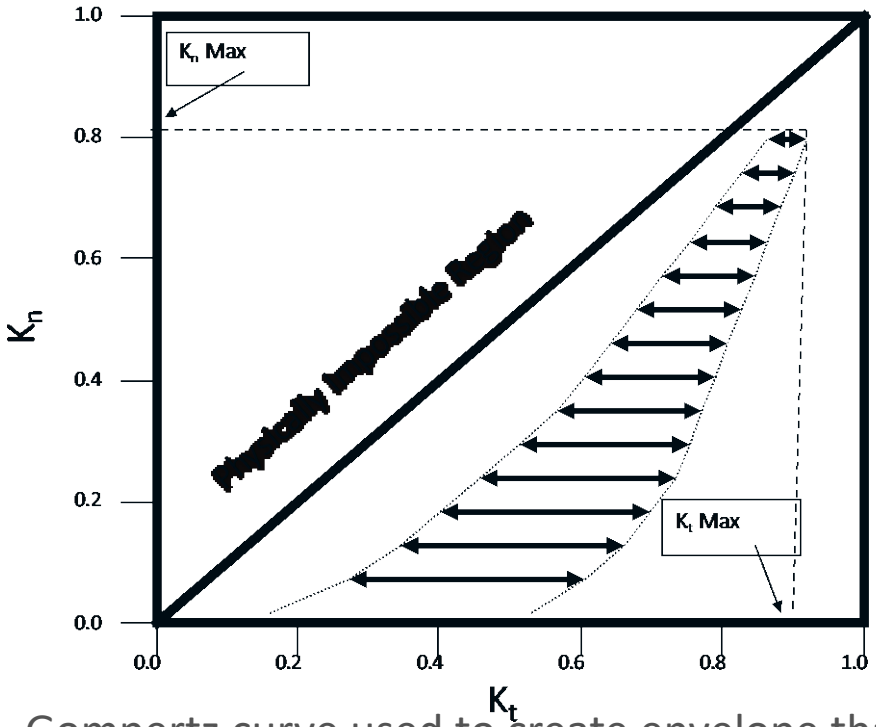
Depending on the available data, SERI-QC performs one-element, two-element, or three-element tests.

1. One-element test by defining a range of acceptable values between minimum and maximum values of K_t, K_d, or K_n, based on three air mass regimes and the month of the year.
2. Zenith angle (at the middle of the hour) < 80°and all three of the elements are present: Three-element test performed using a range of acceptable values fulfilling $K_t = K_d + K_n$ within arbitrary error limit of ± 0.03, which accounts for measurement uncertainties.
3. Data passes the three-element test (or at least two elements passed the one-element test): Two-element test performed by defining a range of acceptable values within boundaries.

SERI-QC publication: https://www.nrel.gov/docs/legosti/old/5608.pdf

QCFIT: Reducing the Area of Expected Values

Maximum K_t and K_n

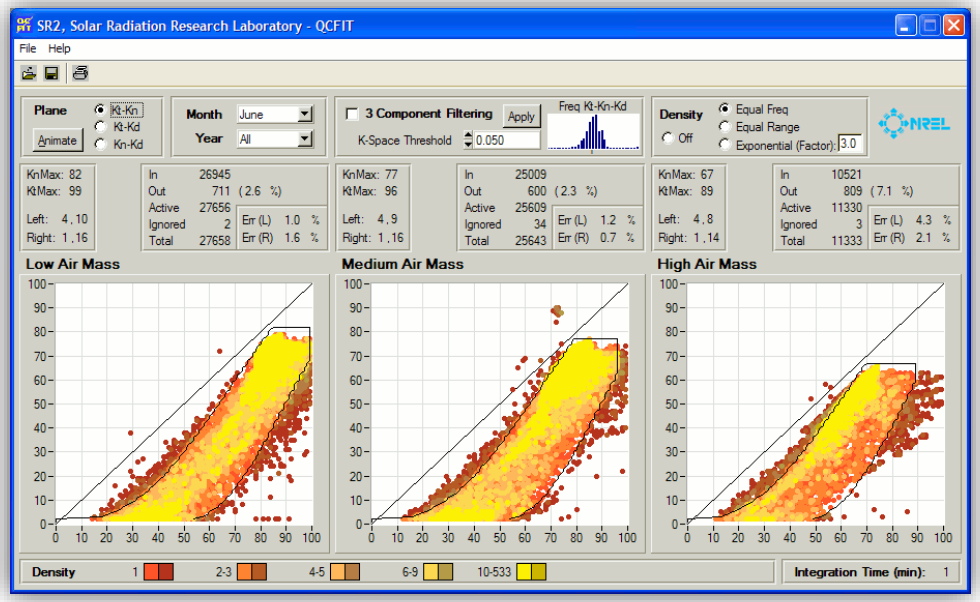


Gompertz curve used to create envelope that fits to multiyear ground measurements.

Air Mass / Zenith Angle Ranges

Range	Air Mass	Zenith Angle
Low	1.00–1.25	0.00–36.96
Medium	1.25–2.50	36.96–66.57
High	2.50–5.76	66.57–80.00

QCFIT: Reducing the Area of Expected Values



- The two-component analysis is useful for checking three-component data.
- An envelope is developed for application to the quality assessment process.

Data Quality Assessment Using NREL's SERI-QC

Flag	Description
00	Untested (raw data)
01	Passed one-component test; data fall within min-max limits of K_t, K_n, or K_d
02	Passed two-component test; data fall within 0.03 of the Gompertz boundaries
03	Passed three-component test; data come within 0.03 of satisfying $K_t = K_n + K_d$
04	Passed visual inspection; not used by SERI-QC
05	Failed visual inspection; not used by SERI-QC
06	Value estimated; passes all pertinent SERI-QC test
07	Failed one-component test; lower than allowed minimum
08	Failed one-component test; higher than allowed maximum
09	Passed three-component test but failed two-component test by >0.05
10-93	Failed two- or three-component tests in one of four ways. To determine the test failed and the manner of failure (high or low), examine the remainder of the calculation (flag + 2)/4. REM — Failure 0 — Parameter too low by three-component test ($K_t = K_n + K_d$) 1 — Parameter too high by three-component test ($K_t = K_n + K_d$) 2 — Parameter too low by two-component test (Gompertz boundaries) 3 — Parameter too high by two-component test (Gompertz boundaries) The magnitude of the test failure (distance in K-units) is determined from: d = (INT (flag +2)/4)/100.
94-97	Data fall into a physically impossible region where $K_n > K_t$ by K-space distances of 0.05–0.10 (94), 0.10–0.15 (95), 0.15–0.20 (96), or ±0.20 (97).
98	Not used
99	Missing data (associated data field is filled with -9900)

SERI-QC returns a two-digit quality flag:

- The magnitude of the error
- The direction of the error (high, low)
- The test that reported the error.

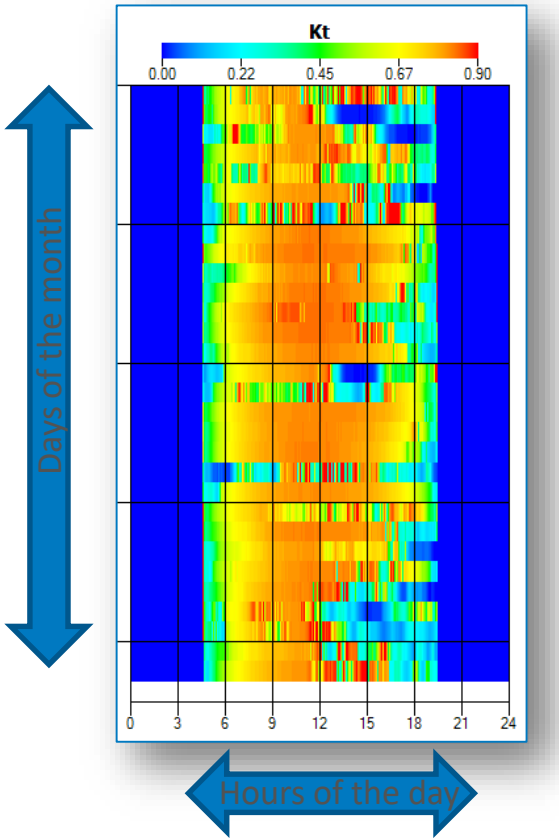
SERI-QC is currently a programming function, not an application. (It is currently being packaged as an application.)

Data Quality Assessment Using NREL's SERI-QC

Daily Quality Checks

SERI-QC cylinder plots:

- A full month of data and quality flags can be seen at a glance.
- Shows data for each of the three solar components.
- Errors become instantly evident.
- You can correlate flags with irradiance values
- You can view the three components in context with each other.



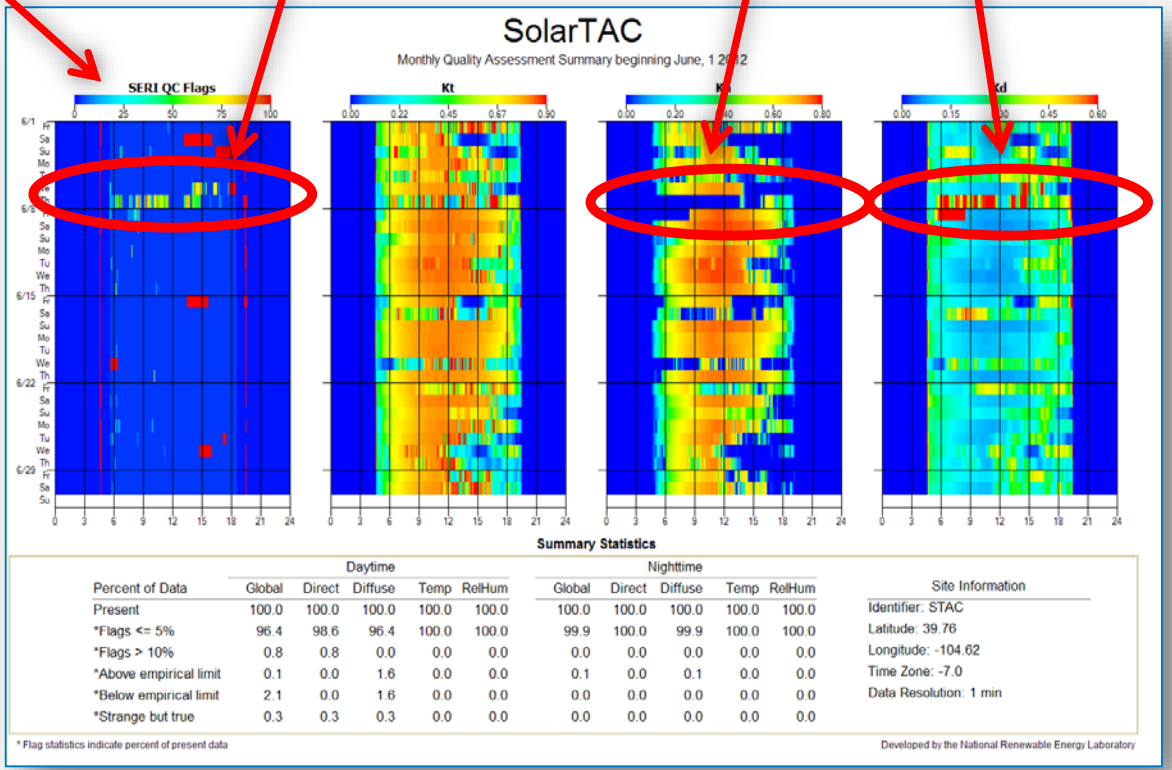
Data Quality Assessment Using NREL's SERI-QC

$\varepsilon = K_t - K_n - K_d$

High SERI-QC flags

Low K_n

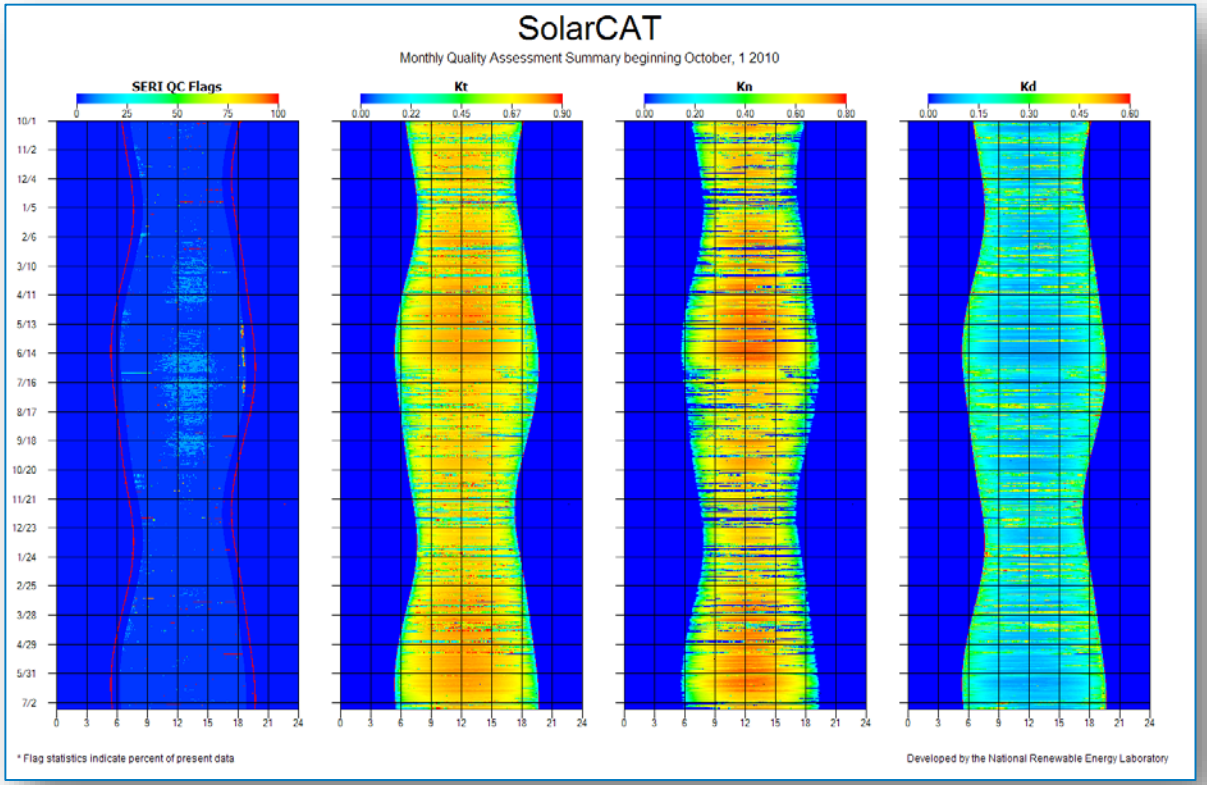
High K_d



Example: tracker failure June 6 and 7

Data Quality Assessment Using NREL's SERI-QC

A seasonal look at flags: data for 1.5 years



NREL's SERI-QC Limitations

- If the data set contains few data points, or points clustered in a small region, or many questionable points, the Gompertz curves needs to be edited.
- Great number of values for which the direct normal measurement is near zero and skews the automatic curve fitting algorithm.
- Data points with a Kn component <0.02 are excluded from consideration by QCFIT.

Future Plan for NREL's SERI-QC

- Make it available in other programming languages, such as Python.
- Remove limitations previously mentioned by making the QCFIT curve-fitting process more autonomous.
- Refine the partitioning atmospheric combinations to improve quality assessment.

Thank You

https://www.nrel.gov/midc

Contact: Manajit.Sengupta@nrel.gov

NREL/PR-5D00-74203

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

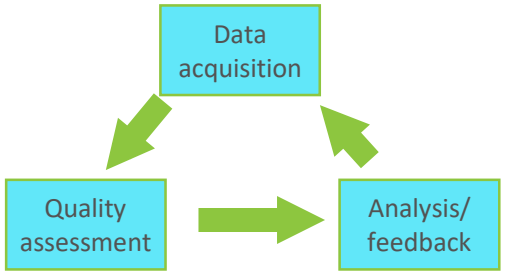
NREL
Transforming ENERGY

Additional Slides

Data Quality Assessment

Data quality analysis procedure:

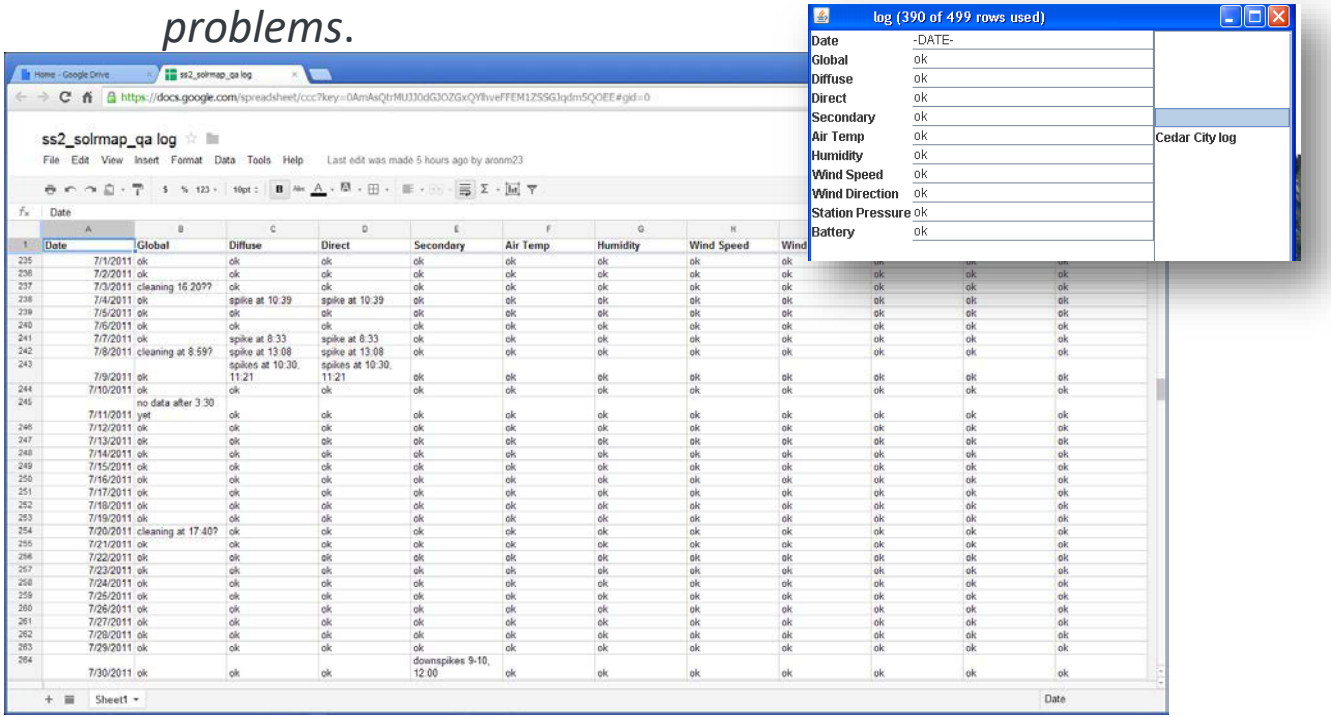
- **View all data as frequently as possible (daily is best).**
  - The longer the delay, the longer error conditions will persist.
  - The more frequent your data checks, the more in tune you are with your stations.
- View in context of other measurements.
  - Measurements by themselves can be deceiving.
- Automate the data plots as much as possible.
  - Spend your time *analyzing* data, not assembling data sets.
- Set up a feedback infrastructure.
  - Communicate findings back to the station.
  - Good results should be communicated also.



Data Quality Assessment

Documenting data quality:

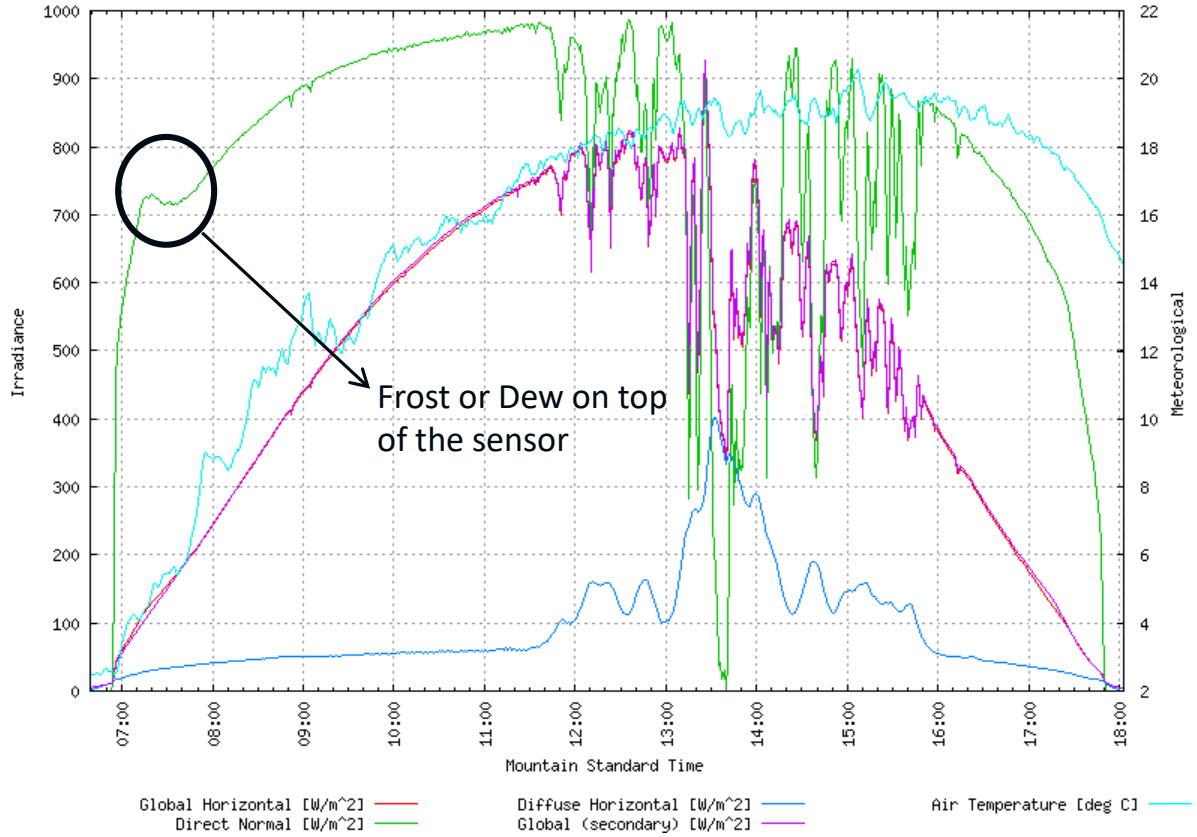
- Document all findings, *including findings of no problems*.



A Java script interfaces with Google Docs

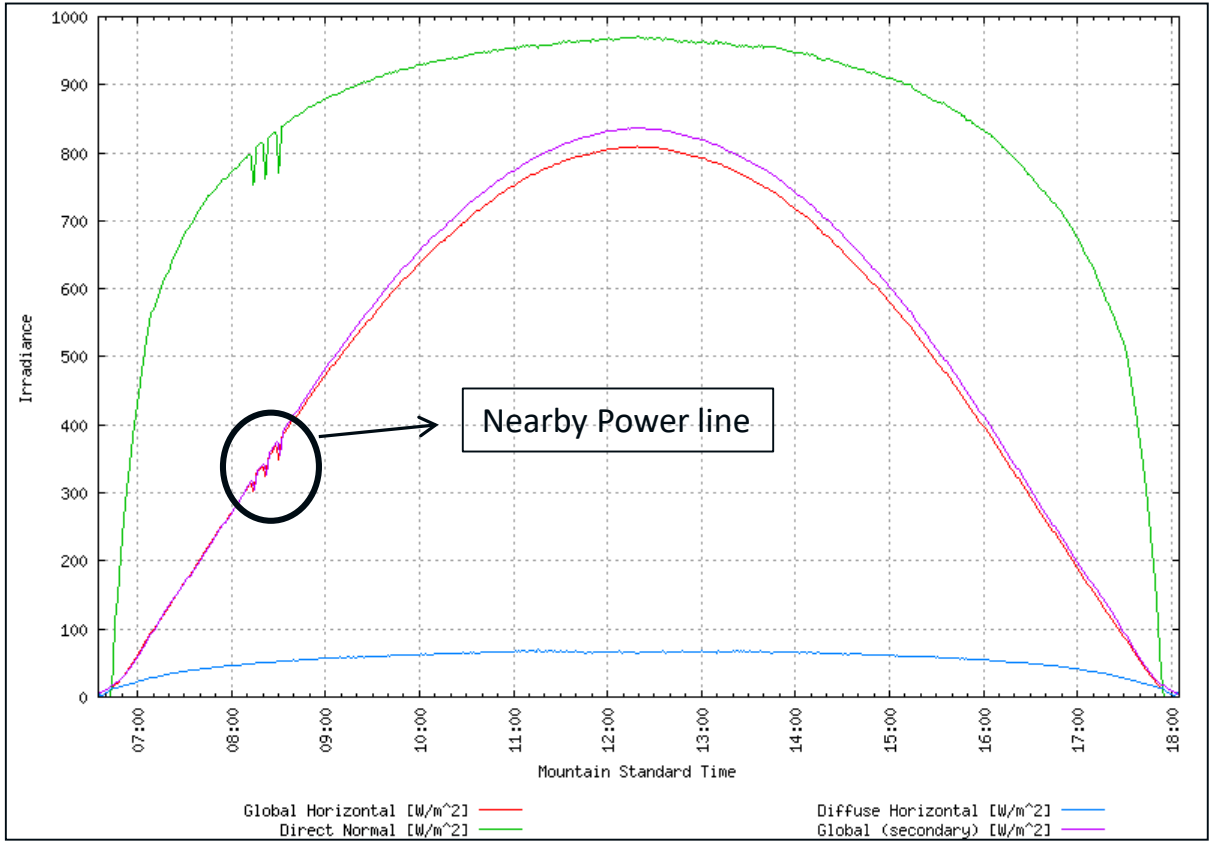
Data Quality Issues

Anomalous data



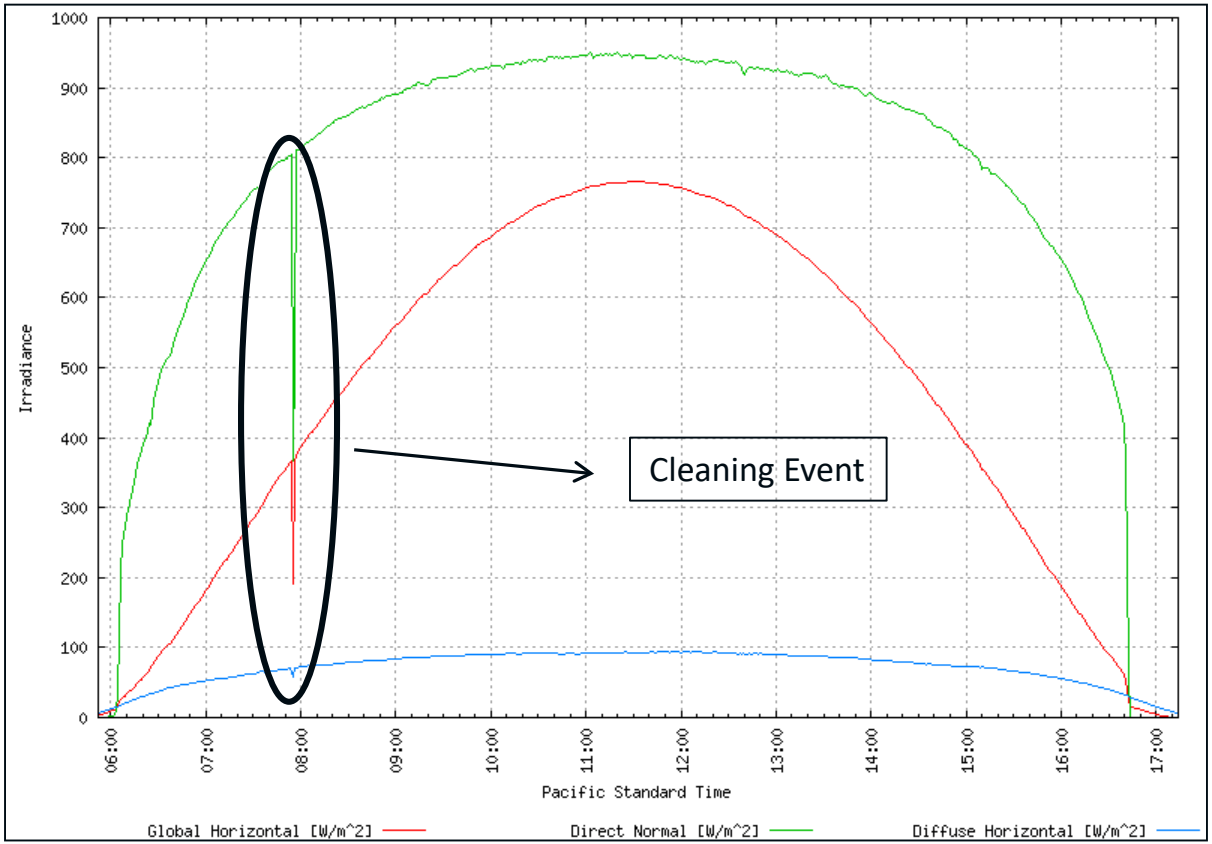
Data Quality Issues

Anomalous data



Data Quality Issues

Anomalous data



Data Quality Issues

Anomalous data