Solar Resource Parameters Verification Guidelines

We have performed extensive validation of the source, GSIP V1 database and some limited-scope verification (so far) of the derived database of solar resource parameters (SRDB). It is only natural if users will do their own verification of the database. In this document, we intend to share our experience in order to help users to avoid known pitfalls.

1. Compare comparable:
   a) SURFRAD provides frequent measurements with 1 or 3 minute repetition rate. They may be drastically different from the SRDB data because of rapidly changing cloudiness at the ground station. SRDB data describe average of a large 0.5deg x 0.5deg cell (about 50x50 km). Local station may see the clouds of this cell only as a result of wind transporting the cell clouds over the station during some period of time. That is why the averages, at least one-hour long, are natural subject of comparison in this case.
   b) Try to use the same formulas for ground data as we do in SRDB database. In these formulas, operate with hourly averages of ground data, not with instantaneous measurements.
   c) Again, insolation is critically dependent on clouds. Comparison should give the best results when the cloud fraction in vicinity of the station mimics that of the GSIP cell containing the station. Unfortunately, the cloud fraction estimate does not belong to the standard, verified parameter set of the ground stations. However, one can use fraction of direct to total flux as a proxy for the local clear fraction (or diffuse to total for the local cloud fraction). There is no direct equality though, so regression of the SRDB cell cloud fraction to local diffuse-to-total ratio can help substantially.
2. Ground measurements may have problems, too:
   a) Total (global) solar radiation sensors can have problems, so watch for quality flags in the ground data; in a typical ground arrangement, you can use direct and diffuse fluxes to reconstruct total one.
   b) Local surroundings can greatly influence diffuse flux.
   c) Low-flux measurements may be affected by dark currents of the sensors, so the use of a low-flux threshold is advisable (in order of 80 W/m^2).
3. Timing issues may be of importance. Though GOES CONUS data were taken once an hour with reported time of hour + 15 minutes, for any given cell, they are rather snapshots taken at different time. That is why the hourly average of the ground station should be taken with the central time equal to the moment of the GOES scan of the cell containing the station.
4. Seasonal factors. Detection of clouds over the snow cover is always challenging, so GSIP insolation data for snowy regions may show larger errors. However, our validation shows that this segment of data has little influence on overall performance of GSIP/SRDB.
5. Use proper measure of errors. For the solar energy parameters, the high-flux performance is most important because it is related to the highest output of the solar-energy factories. If we use the standard measure of errors, CV (Coefficient of Variation) approach, it gives equal weight to the errors in strong and weak fluxes. Therefore we find more suitable the estimate of the average value (AV) error, which weigh the error with the relative strength of the flux. The contribution of errors associated with low fluxes is suppressed in favor of those of large fluxes.
6. Do not fully trust viewer routines, always check numbers. For example, popular HDFviewer checks for maximal and minimal values of 2D array before plotting it. Then it rounds the minimal value (for example, 186.59 to 187) and compares it with the table values. Naturally, 186.59 < 187, and this point is displayed as a fill value though it is a physically-legitimate value for flux. This happens on almost all 2D plots. Moreover, in multipage data, HDFviewer keeps the bounding values unchanged from page to page. That results in clipping the images.