

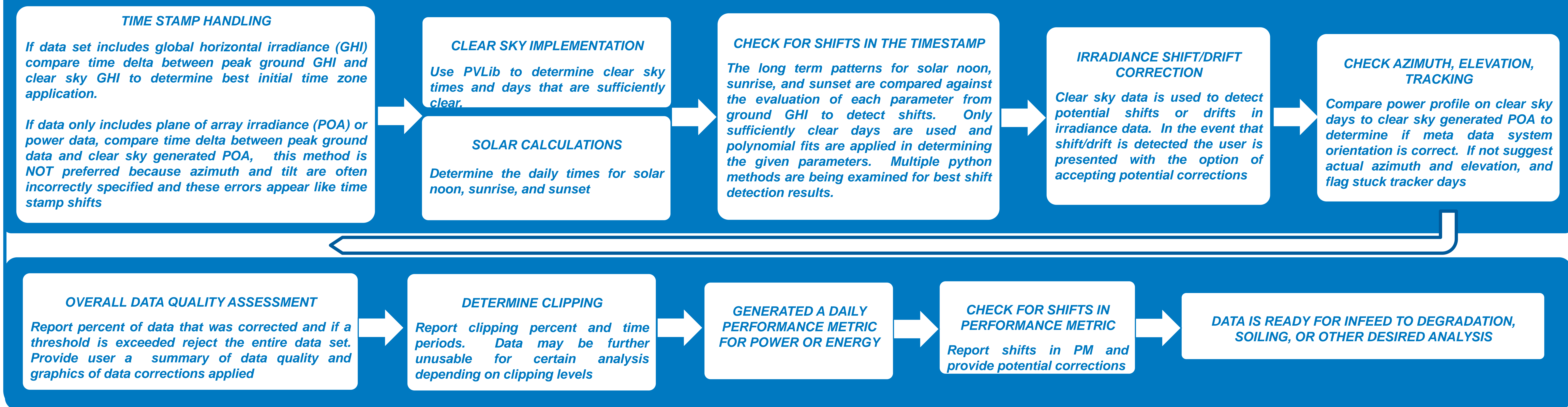
Development of Automated Methods to Determine the Quality of Long-term PV Data Sets

Matthew Muller, National Renewable Energy Laboratory

PROBLEM: analysis of PV degradation, soiling, and other performance trends benefits from analyzing 100s of time series data sets. These data are often provided with quality issues such as incorrect time stamps, drifting or miscalibrated sensors, incorrect field names, or missing periods of data. PV experts can often manually identify and correct the issues but this method is too time consuming for large scale analysis.

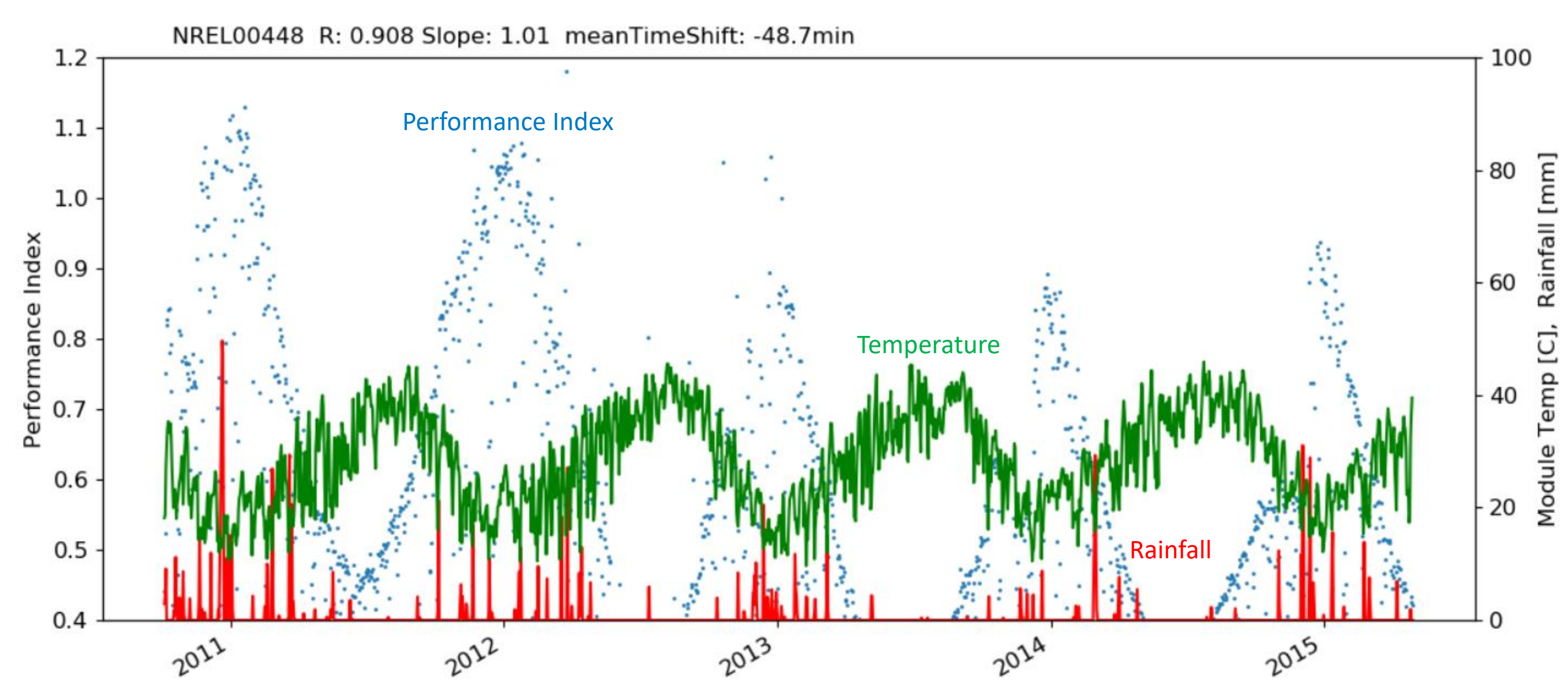
GOAL: open source algorithms to automatically flag suspect data and present the user with potential corrections

DATA HANDLING AND FLOW

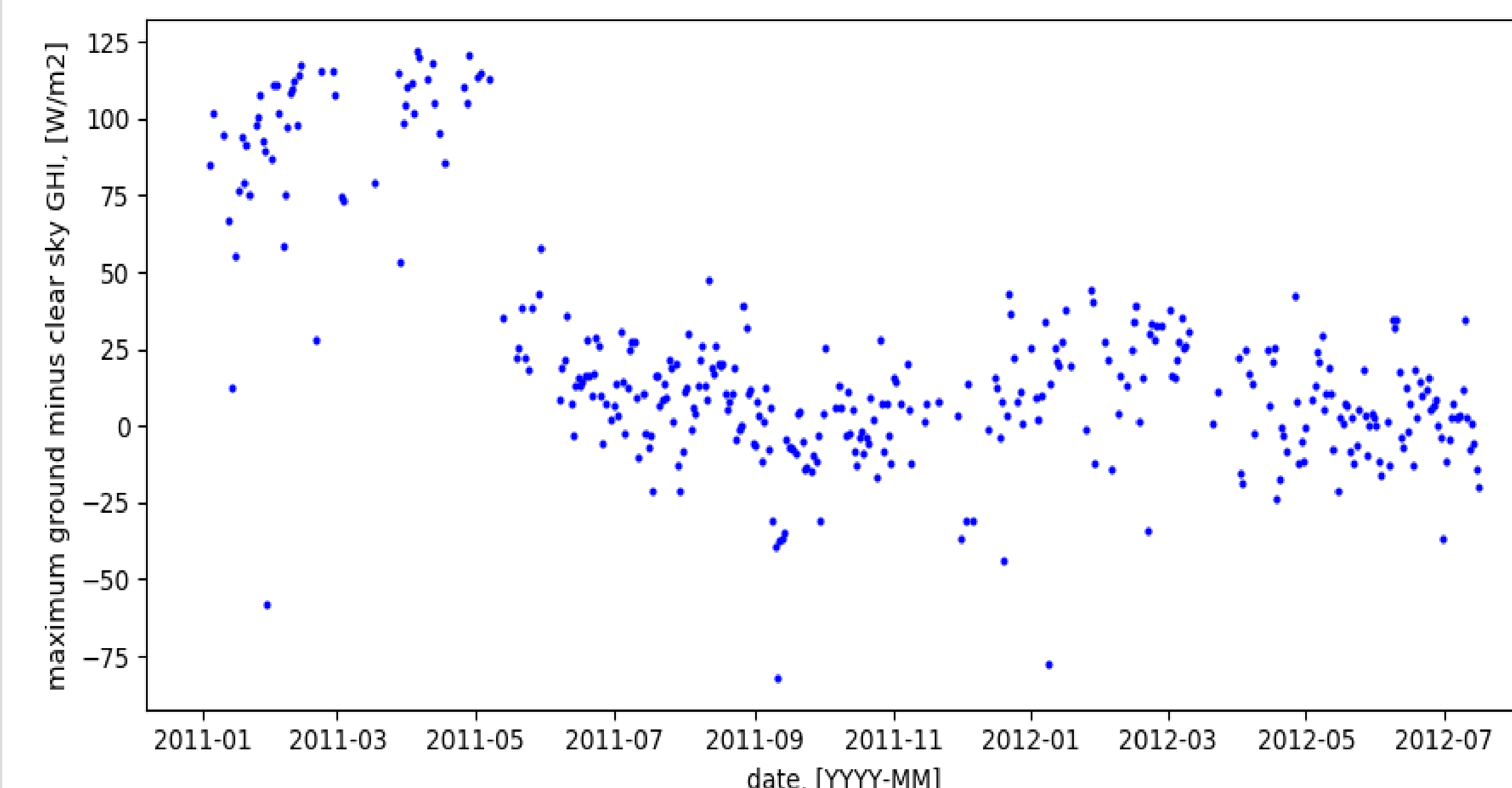
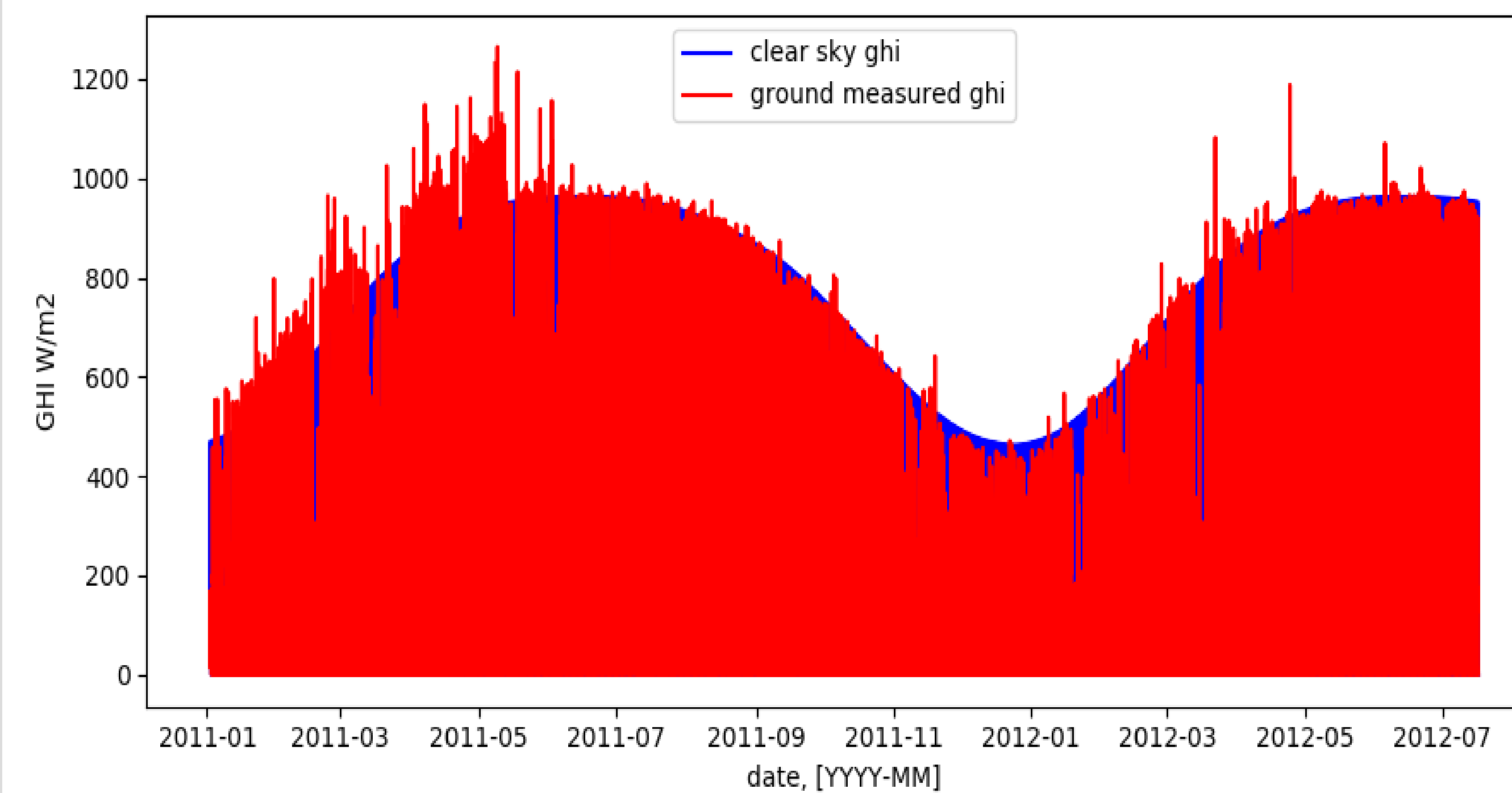
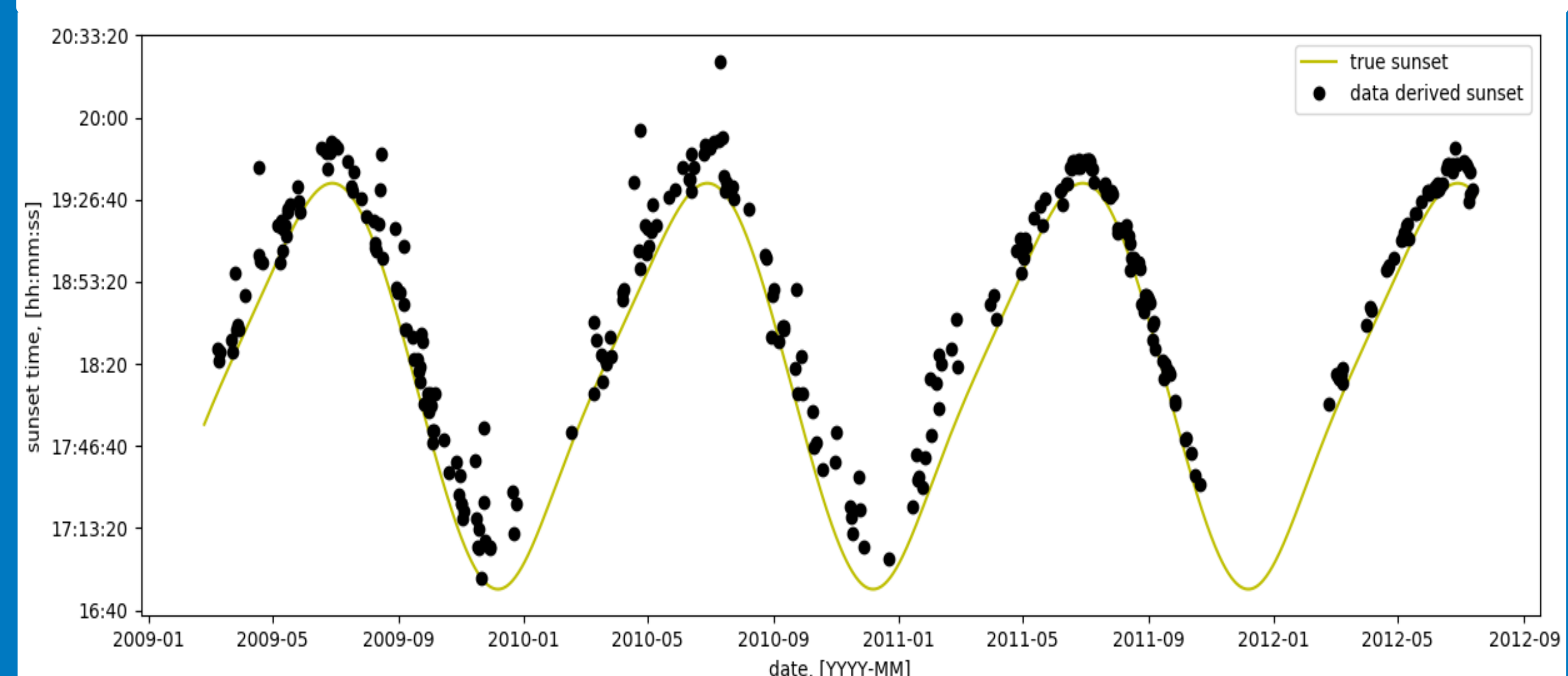


DATA SET WITH MULTIPLE PROBLEMS IDENTIFIED

- Initial timestamp handling identified a mean time shift between the peak clear sky POA and peak power of 49 minutes.
- Tilt error was identified
- Performance shift between 2012 and 2013 was identified
- Large seasonal swings in performance index due to both time error and tilt error



USING SUNRISE, SUNSET, SOLAR NOON TO DETECT TIME SHIFTS



PYTHON RUPTURES() IDENTIFIES SHIFT DETECTION IN GHI DEVIATION

RESULTS AND ONGOING WORK

- Algorithms are being tested on initial data sets
- There are 101 potential quality issues with data and the algorithms are not yet generalized enough to handle a wide range of variation in such problems
- Filtering levels are being tested to determine what settings best handle various data sets and time stamps
- Ground GHI data provide a much stronger starting point for quality analysis than POA or power data alone.
- Various python shift detection algorithms are being tested for robustness to solar data
- EVENTUAL SHARING THROUGH RTools