Utility-Scale Solar Data File for Generation and Market Value

Background

Lawrence Berkeley National Laboratory (Berkeley Lab) estimates hourly project-level generation data for utility-scale solar projects in the seven organized wholesale markets and 18 additional Balancing Areas. The public project-level dataset is updated annually with data from the previous calendar year. To encourage its broader use, Berkeley Lab makes a comprehensive data files public at the Open Energy Data Initiative (OEDI) at https://data.openei.org/submissions/6200 and summary data files at https://emp.lbl.gov/utility-scale-solar.

Annual solar summary statistics by plant (UPV)

We provide project-level (UPV) annual summaries of the solar generation, curtailment, average wholesale energy value, average capacity value (both in $/MWh and $/kW-yr), combined energy and capacity value, and value factor in *Annual_Solar_Value_by_plant.xlsx*. For more information on methods, data, and validation see Appendix A and C in the technical Solar to Grid report: https://emp.lbl.gov/publications/solar-grid-trends-system-impacts-0.

Hourly generation data

In addition to the annual solar summary statistics Berkeley Lab provides hourly generation estimates for 5225 utility-scale solar projects, starting at the project's commercial operation date (or 2012 for older projects) until the end of 2023. A separate .csv file is listed for each UPV project, using the EIA plant ID as its filename. Records are indexed by UTC-Hour-Beginning datetimes. Here we summarize the data by column:

SAM_gen: Modeled generation estimates using NREL's System Advisory Model (SAM) with project-specific system characteristics reported in EIA Form 860 (augmented by data collected for our Utility-Scale Solar Series) and historical irradiance estimates in NREL's National Solar Radiation Database (NSRDB, 2012-2020) and NOAA's High-Resolution Rapid Refresh model (HRRR, 2021 forward).

gen_bias: Modeled generation estimates for a debiasing process that are for the most part identical with *SAM_gen*. Minor deviations occur for projects where system characteristics were updated after the debiasing process was run, resulting in updated *SAM_gen* records.

gen_bias_corrected: Debiased generation estimates where the modeled generation was scaled to fit the (1) project-specific solar generation reported by EIA Form 923 (based on annual generation for the years 2012-2014 and based on monthly generation starting in 2015) and (2) hourly system-wide solar generation for a subset of ISOs/RTOs and Balancing Areas. For a subset of projects in ERCOT, we directly report project-specific hourly generation that is publicly available 60 days after operations day. This is raw data that may contain commissioning data and telemetry errors.

gen_clean: Hourly generation estimates that are used as basis for value and system impact calculations throughout the report. Where feasible, we default to *gen_bias_corrected* estimates. When that data is not available, we use *SAM_gen* estimates. If curtailment is reported in the column *gen_curtailed*, *gen_clean* represents post-curtailment output. The file *UPV_generation_overview_by_plant_year.csv* summarizes which generation estimates are reported in this column by project and year.

gen_curtailed: Estimated hourly curtailment for projects in CAISO and ERCOT.

Who to Contact with Questions?

Questions or comments may be directed to Joachim Seel (jseel@lbl.gov).

Disclaimer and Copyright Notice